Rakibul Hasan

Managing Artificial Intelligence Across Borders

ACTA WASAENSIA 591

University of Vaasa
VAASAN YLIOPISTO

ISBN 978-952-395-282-9 (print)
978-952-395-283-6 (online)

ISSN 0355-2667 (Acta Wasaensia 591, print)
2323-9123 (Acta Wasaensia 591, online)

URN https://urn.fi/URN:ISBN:978-952-395-283-6

PunaMusta Oy, Joensuu, 2026.

ACADEMIC DISSERTATION

To be presented, with the permission of the Board of the School of Marketing and Communication of the University of Vaasa, for public examination on the 4th of September, 2026, at noon.

Article-based dissertation, School of Marketing and Communication, International Business

Author	Rakibul Hasan https://orcid.org/0000-0003-2088-0490
Supervisor(s)	Professor Arto Ojala University of Vaasa, School of Marketing and Communication, International Business
	Professor Heidi Kuusniemi Tampere University, Faculty of Information Technology and Communication Sciences and University of Vaasa, School of Technology and Innovations
Custos	Professor Arto Ojala University of Vaasa, School of Marketing and Communication
Reviewers	Professor Markus Salo University of Jyväskylä, Faculty of Information Technology
	Associate Professor Man Yang Hanken School of Economics, Department of Management and Organizing
Opponent	Professor Markus Salo University of Jyväskylä, Faculty of Information Technology

Tiivistelmä

Tekoälyn johtaminen kansallisten rajojen yli poikkeaa perinteisistä tarkasteluista, jotka keskittyvät taloudelliseen integraatioon ja sopeutumiseen yhteisen hyvän saavuttamiseksi. Tekoäly on laskennallisen kehityksen kärjessä ja muuttaa yritysten rajat ylittävän toiminnan organisointia. Digitaalisilla globaaleilla markkinoilla johtamiskäytännöt mahdollistavat nykyaikaisen tekoälyn hyödyntämisen sekä ihmisen ja tekoälyn vuorovaikutuksen tukemisen antamalla tekoälylle inhimillisiä piirteitä. Tekoäly luo kilpailuetuja mutta myös riskejä sosiaaliselle hyvinvoinnille. Tekoäly herättää sääntelyviranomaisissa huolta harhaanjohtamisesta, manipuloinnista ja käyttäjien hyvinvoinnista. Nykyiset sääntelytoimet eivät kuitenkaan riitä luomaan yhteistä viitekehystä haittojen ehkäisemiseksi ja vastuullisen tekoälyn tukemiseksi eri institutionaalisissa ympäristöissä. Tästä syystä tieteellistä keskustelua tulee arvioida uudelleen sen osalta, miten tekoälyä organisoidaan rajat ylittävästi yhteistä hyvää tukevalla tavalla. Väitöskirja tarkastelee tekoälyn strategista organisointia ja kansainvälistä johtamista taloudellisten ja sosiaalisten tavoitteiden integroimiseksi. Tutkimus yhdistää kansainvälisen liiketoiminnan (IB) strategiatutkimuksen ja tietojärjestelmätieteen (IS) tietojohtamisen keskustelut. Työ nojaa teoreettiseen pluralismiin, refleksiivisyyteen ja järjestelmälähtöiseen metodologiseen täydentävyyteen säilyttäen ankkuroinnin molempiin aloihin. Tarkastelu on monitasoinen ja kattaa johtajat, käyttäjät sekä nykyaikaisen tekoälyn yhteisen hyvän kontekstissa. Keskeinen kontribuutio on globaalisti reagoivan tekoälyn organisoinnin viitekehys, joka jäsentää tekoälyn johtamista autonomian, vuorovaikutteisuuden, oppimisen ja ulkoasun kautta. Nämä ulottuvuudet toteutuvat kolmena keskinäisriippuvaisena sitoumuksena: instituutioihin vaikuttamisena, haitallisten keinojen estämisenä ja vastuullisuuden varmistamisena. Nämä sitoumukset integroidaan oletusarvoisesti tekoälyn eri ulottuvuuksiin globaalisti, mikä edellyttää jatkuvaa johtamista rajat ylittävän reagointikyvyn varmistamiseksi. Väitöskirja tarjoaa käytännön ohjeita johtajille ja sääntelyviranomaisille käyttäjien hyvinvoinnin turvaamiseksi ja siirtää keskustelua ”luotettavasta tekoälystä” kohti ”tekoälyn luottamusvastuuseen perustuvaa organisointia”, mikä lisää johdonmukaisuutta ja vahvistaa vastuullisuutta eri markkinoilla. Kokonaisuudessaan tutkimus edistää IB- ja IS-alojen teoriaa ja käytäntöä viidennen teollisen vallankumouksen ja YK:n kestävän kehityksen tavoitteiden mukaisesti osoittaen, kuinka tekoäly voi tukea globaalia kilpailukykyä ja edistää yhteiskunnallista hyvinvointia.

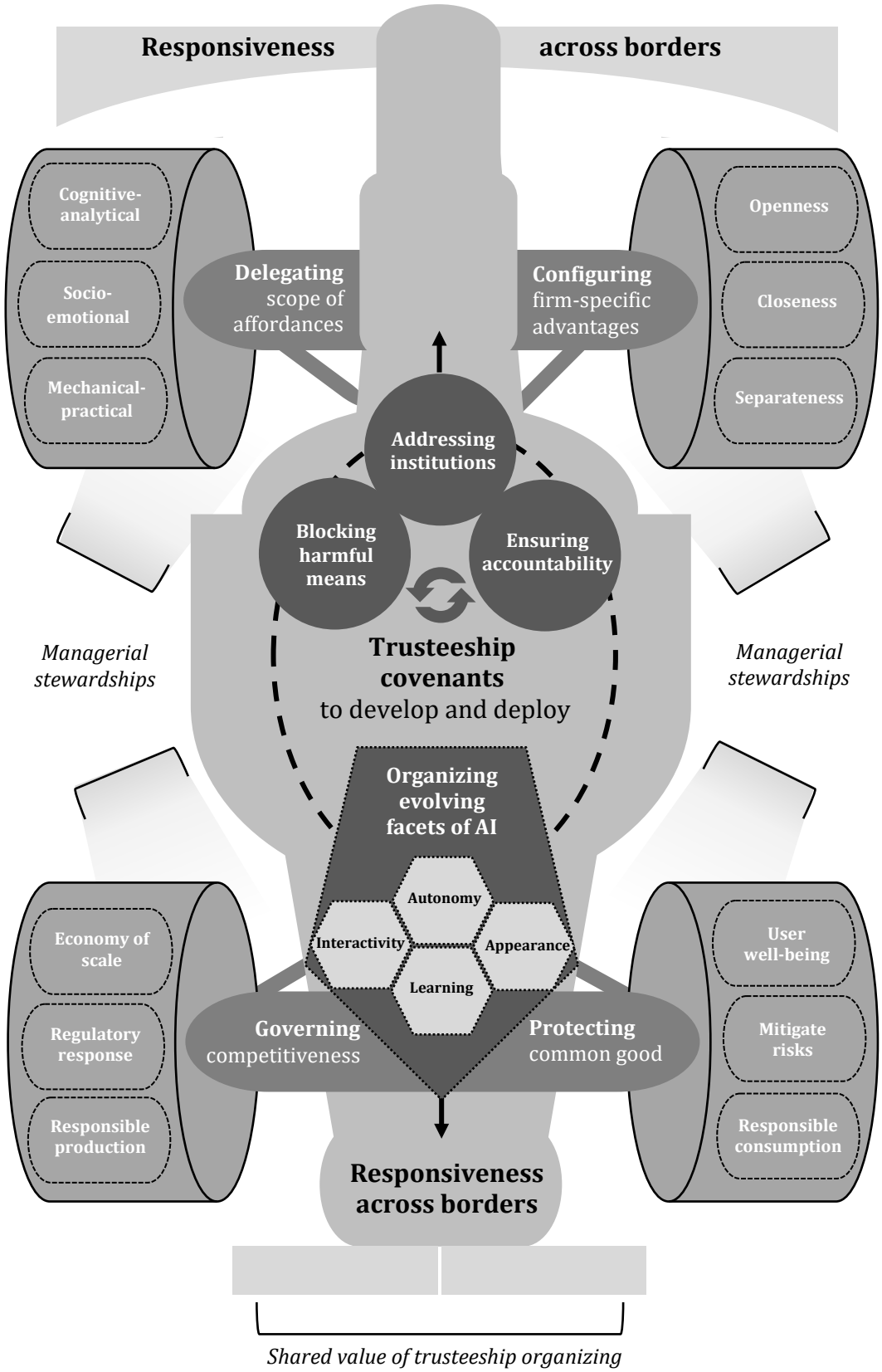
Asiasanat: kansainvälinen liiketoiminta; tiedonhallinta; tekoälyjärjestelmät; globaalisti reagoiva tekoälyn organisointi; vastuullisuus; affordanssi; yrityskohtaiset kilpailuedut.

Abstract

Managing artificial intelligence (AI) across national borders differs from traditional approaches that focus on economic integration and responsiveness in pursuit of the common good. AI is the frontier of computational advancements and is transforming how firms organize cross-border activities. In digital global markets, managerial actions increasingly enable contemporary AI in different forms to facilitate human-AI interaction through anthropomorphization, or the attribution of human-like features. While AI offers managers substantial opportunities to create firm-specific advantages, it also poses challenges by creating intended and unintended affordances that may jeopardize social well-being. AI thus raises regulatory concerns about deception, manipulation, and user well-being. However, regulatory responses remain insufficient to establish a shared framework for preventing harm and supporting responsible AI use across diverse institutional contexts, such as norms and cultures. As a result, existing scholarly conversations need to reconsider how AI can be organized across borders in ways that ensure responsiveness to the common good. Accordingly, this dissertation examines the strategic organizing of AI and, more specifically, how AI can be managed across borders in a globally responsive manner to integrate economic and social goals. It brings together international strategy research in international business (IB) and information management discourse in information systems (IS). It embraces theoretical pluralism, reflexivity, and systems-oriented methodological complementarity while remaining grounded in both IB and IS. The unit of observation is multilevel, encompassing managers, users, and contemporary AI in the context of the common good. The central contribution of this dissertation is a novel framework, Globally Responsive AI Organizing (GRAO), which proposes managing AI through four interrelated facets: (1) autonomy, (2) interactivity, (3) learning, and (4) appearance. These facets are enacted through trusteeship organizing to constitute three interdependent covenants that intentionally prioritize proactiveness in (a) addressing institutions, (b) blocking harmful means, and (c) ensuring accountability. These trusteeship covenants are integrated into the facets of AI globally by default, which may require constant managerial action to ensure responsiveness across borders. The dissertation provides practical guidance for managers and regulators to safeguard user well-being, shifting the existing discourse from “trustworthy AI” to “trusteeship organizing of AI” to achieve coherence and ensure accountability across markets. Overall, it advances theory and practice in IB and IS in alignment with the Fifth Industrial Revolution and the United Nations Sustainable Development Goals by showing how AI can support global competitiveness while promoting social welfare.

Keywords: international business; information management; AI systems; Globally Responsive AI Organizing; trusteeship; affordance; firm-specific advantages.

Visual Abstract



ACKNOWLEDGEMENT

I begin with God's name, the Kind, the Caring.

I express my deepest gratitude to the individuals and organizations who have supported my doctoral journey. This dissertation reflects their collective contributions through intellectual guidance, funding, opportunities, sacrifice, and support. While many deserve acknowledgment, I highlight only those most essential.

I am profoundly grateful to my supervisor, lifelong mentor, and co-author, Prof. Arto Ojala. Your intellectual guidance, inspiration, trust, and openness allowed me the freedom to grow as a student of science and develop independence as a researcher. Your kindness and simplicity helped me a lot in every aspect of my life, not just research. I wish you good health and peace be upon you.

I extend my sincere thanks to my current and former second supervisors, Prof. Heidi Kuusniemi and Dr. Tiina Leposky, for your support. I am also thankful to Prof. William Baber for hosting me and facilitating my research exchange at Kyoto University, Japan. In addition, I honor the memory of Prof. Jorma Larimo, whose inspiration, trust, and encouragement enabled me to begin this doctoral journey here in Vaasa.

It is my honor to have two distinguished scholars, Prof. Markus Salo from the University of Jyväskylä and A/Prof. Man Yang from Hanken School of Economics, as pre-examiners of my dissertation. I sincerely thank you for taking time to read and examine my dissertation. I am also grateful for your constructive recommendations that have significantly improved this dissertation. In addition, I am particularly thankful to Prof. Markus for agreeing to be the opponent for my dissertation defense.

I would like to thank Prof. Peter Gabrielsson for being the opponent in my dissertation pre-defense and discussant of my research proposal and articles in internal seminars throughout the doctoral journey. I deeply appreciate all those valuable comments you provided and look forward to continuing to learn from you.

I am grateful to all senior and fellow colleagues in the School of Marketing and Communication. I want to thank EVERYONE from the International Business and Marketing Strategy research group. I thank Dean Prof. Merja Koskela for providing all necessary support to carry out my work. I am especially grateful to Prof. Hannu Makkonen, Prof. Zaheer Khan, Prof. Rebekah Rousi, A/Prof. Anisur Faroque, A/Prof. Emilene Leite, A/Prof. Hanna Leipämaa-Leskinen, Dr. Minnie Kontkanen, Dr. Tahir Ali, Dr. Feroj Mahmood, and Tiina Shen for your continuous guidance and support. I especially want to thank Nicholas Rowe for your copyediting assistance.

I extend my sincere gratitude to my mentors, former supervisors, and co-authors for your contributions to the development of my research and publication skills before, during, and beyond my doctoral journey. I am especially grateful to A/Prof. Mohammad Bakhtiar Rana, Dr. Vlad Stefan Lichtenthal, Dr. Mark Avis, Dr. Ausma Bernot, A/Prof. Park Thaichon, and Prof. Scott Weaven.

I gratefully acknowledge the University of Vaasa for employment opportunities, the Finnish Ministry of Education and Culture for the Finland Fellowship, and the AuroraSpace project, funded by the EU-Interreg Aurora. I sincerely appreciate the Finnish Foundation for Economic Education for the work, research exchange, and conference grant. Thanks to OpenInnoTrain, funded by the EU, and the AIB/Sheth doctoral stipend to support my impact-oriented training and academic engagement. Additionally, I acknowledge the Griffith University (International) Postgraduate Research Scholarship of Australia, whose support of my earlier doctoral work contributed to one article included in this dissertation.

Last but not least, I owe my deepest appreciation to family and close relatives. The highest gratitude goes to my beloved parents (Alep Hossain and Rokeya Begum). Your countless sacrifices enabled me to pursue knowledge abroad. My greatest thanks belong to my dear wife, Foujia Sultana Nipa, and my children. Your unwavering love, patience, sacrifices, and constant encouragement sustained me through every challenge of this journey and gave me strength in moments of uncertainty. Your presence made this endeavor both meaningful and possible. I extend thanks to my elder brother, Rashedul Hassan, and my cousin, Tarek Hasan, for your ongoing support. I am also thankful to many friends, collaborators, and well-wishers from Bangladesh, Denmark, Australia, Finland, and around the world for their intellectual, social, and moral support.

A particular note is that like any scientific work, this dissertation is neither perfect nor beyond criticism, particularly when undertaken by a somewhat naive student of science like myself. However, it was undertaken in a sincere effort to pursue knowledge for the benefit of both worldly life and the hereafter. I acknowledge all limitations with humility, and accept that errors or misinterpretations may remain. I welcome constructive criticism, continued learning, and future improvement, and acknowledge any shortcomings as entirely my own. I warmly invite readers to reach out and share their insights or suggestions.

Peace be upon you!

Rakibul Hasan
August 2026, Vaasa

To those who
engage with the world thoughtfully,
nurture cooperation for peace,
do good for the hereafter, or
contribute towards the common good.

… … ..

Reckoning

(A profound metaphysical expression rooted in the reading of the Qur'an)

In the ever-expanding universe,
I am on a tiny planet called Earth,
Where mountains as stakes and the sky as a roof.
Rain from the sky, giving life to the earth after its death.

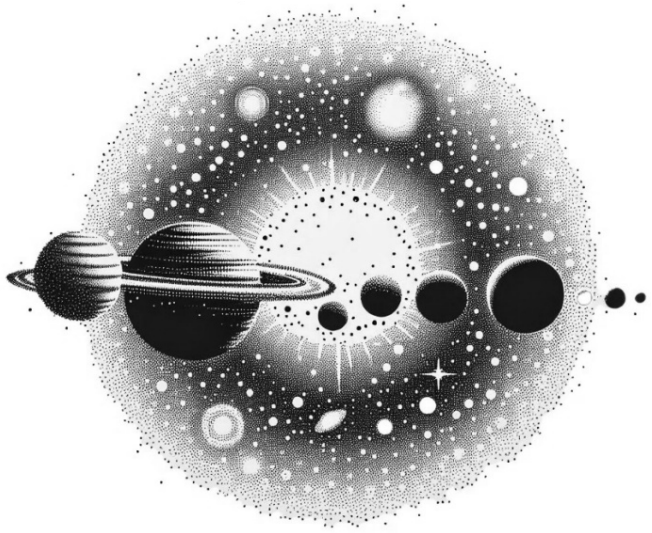


Illustration: Author, developed with Copilot

To comprehend how small the Earth is in our solar system, visit solartoscale.com

I crawled out from the womb by single pairs.
Suddenly, non-existence exists for a short period of time.
Amazed by differences, distances, and various colors.

Alone among the unseen, light, river, fire, and all that exists?

Above and within the earth,
There are entities, places, and events.
Some are observed by the law of nature.
Some are unseen and unknown, unable to explain.
But they seem to exist.

Everything seems organized and purposeful.
Interconnected to a single unifying structure.

Respect, responsibility, and humility for unity.
Justice for peace and everlasting happiness.

The constant reckoning of trusteeship for goodness.

Return to Omniscience, to the state of All-Knowing.

Contents

TIIVISTELMÄ V

ABSTRACT VI

VISUAL ABSTRACT VII

ACKNOWLEDGEMENT IX

1 INTRODUCTION 1

2 THEORETICAL BACKGROUND 9
2.1 Contemporary Understanding of AI in International Business 9
2.1.1 Contemporary Research on AI in International Business 9
2.1.2 Contextualizing AI for the Common Good within Firms' Cross-Border Activities 11
2.2 Foundations for Theorizing AI in Cross-border Activities 13
2.2.1 AI as an evolving phenomenon of computation 13
2.2.2 Contemporary AI Systems and their Representations 15
2.2.3 Theorizing AI by Drawing on Socio-technical Systems 20
2.3 Strategic Organizing of AI Across Borders 24
2.3.1 Responsiveness of AI as a Source of Firm-specific Advantage 24
2.3.2 Foundational Focus on Human Actions to Organize AI 27
2.4 Existing Theoretical Discourse in Cross-border Strategic Organizing 30
2.4.1 Global Integration-Local Responsiveness Perspective 30
2.4.2 Loose Coupling Systems Perspective 32
2.4.3 Rethinking Strategic Organizing for AI 35
2.5 Imperatives for Trusteeship Organizing of AI Across Borders 37
2.5.1 Trusteeship Framework 37
2.5.2 Changing Nature of Production and Facets of AI 40
2.5.3 Changing Nature of Consumption and AI Anthropomorphism 42
2.5.4 Growing Risks of Harm and Insufficient Response from Regulators 43
2.6 Towards a Globally Responsive Trusteeship Organizing of AI 45

3 METHODOLOGY 49

3.1 A Systems-oriented View of Methodological Complementarity 49
3.2 Ultimate Presumptions: A System-oriented Paradigmatic Reflection 52
3.2.1 A Conception of Existence 52
3.2.2 A Conception of Reality 53
3.2.3 A Conception of Knowledge 54
3.2.4 A Conception of Science 55
3.2.5 A Scientific Ideal 56
3.2.6 Responsibility and Aesthetic Aspects 57
3.3 A Multi-method Qualitative Research Strategy: The System-oriented Operatives 58
3.3.1 The Summary Dissertation: An Abstracted Configurational Approach to Understand and Explain a Management System 62
3.3.2 Article I: Directed Content Analysis to Understand AI Systems from a Socio-technical Perspective ... 67
3.3.3 Article II: A Contextualized Explanation Case Study to Explain the Dark Side of AI Systems 68
3.3.4 Article III: A Mechanism-based Explanation Case Study to Understand the Risks of Harm of AI Systems 69
3.4 Quality Assessment 70
3.5 Limitations 72

4 SUMMARY OF DISSERTATION ARTICLES 74
4.1 Article I: Understanding AI Systems from a Socio-technical Systems perspective 74
4.1.1 Interesting Issues and Perspectives 74
4.1.2 Conceptualization 74
4.1.3 Key Findings 75
4.1.4 Contributions 76
4.2 Article II: Explaining the Dark Side of AI Systems 76
4.1.5 Interesting Issues and Perspectives 76
4.1.6 Conceptualization 77
4.1.7 Key Findings 77
4.1.8 Contributions 77
4.2 Article III: Explaining the Risks of Harm of AI Systems 78
4.2.1 Interesting Issues and Perspectives 78
4.2.2 Conceptualization 78
4.2.3 Key Findings 78
4.2.4 Contributions 79

5 DISCUSSION AND CONCLUSION 80
5.1 Globally Responsive AI Organizing (GRAO): A Proposed Theoretical Framework 80
5.2 A Typology of Evolving Facets of AI into Cross-border Activities 82

5.2.1 Autonomy 83
5.2.2 Learning 84
5.2.3 Interactivity....84
5.2.4 Appearance 85
5.2.5 The Interplay of Scope of Affordances and Performance of AI Across Facets 86
5.3 Instituting Trusteeship Covenants for Responsiveness Across Borders 89
5.3.1 Addressing Institutions....91
5.3.2 Blocking Harmful Means 92
5.3.3 Ensuring Accountability 94
5.4 Implications for Managers with Different Strategic Stewardships 95
5.4.1 Possible Implementation of the GRAO framework.95
5.4.2 Delegating Scope of Affordances 96
5.4.3 Configuring Firm-specific Advantages....98
5.4.4 Governing Competitiveness....100
5.4.5 Protecting the Common Good 100
5.5 Implications for Regulators on Emerging Accountability ... 101
5.5.1 Call for a New Industry Standard: 'Global Psychological Safety in Autonomous and Intelligent Systems'....102
5.5.2 Shifting Discourse from 'Trustworthy AI' to 'Trusteeship Organizing of AI'....102
5.6 Future Research Agenda 103
5.6.1 Echeloned Design Science Research Method for Robotic AI System and Responsiveness....106
5.6.2 Interpretive Ethnography on the Performance of Virtual AI Systems....106
5.6.3 A Multiple Explanation Case Study Design on the Scope of Affordances: Robotic vs Virtual AI Systems....107
5.6.4 Construct Measurement of the GRAO Framework: Scale Development and Validation for Different Contemporary Forms....108
5.6.5 Quasi-experimental Design on Facets of AI in International Expansion 111
5.7 Concluding Remarks....112

REFERENCES....114

APPENDICES....132
Appendix 1. Conducting Qualitative Interviews....132

ARTICLES....135

Figures

Figure 1. Theoretical engagement and contribution..... 7
Figure 2. Examples of given computing capabilities of AI 14
Figure 3. Definition and contemporary representations of AI 17
Figure 4. Interpreted relevant assumptions of the AI frontier perspective 21
Figure 5. Interpreted relevant assumptions of the I-R perspective 31
Figure 6. Interpreted relevant assumptions of the loose coupling perspective 35
Figure 7. Rethinking for strategic organizing of AI for worldwide effectiveness 46
Figure 8. Managing AIs' characteristics in cross-border activities75
Figure 9. Evolving facets of AI modulated into its contemporary representations 83
Figure 10. Organizing facets of AI on the global stage..... 87
Figure 11. The processual perspective of the GRAO framework... 89
Figure 12. The globally responsive AI organizing (GRAO) framework 97
Figure 13. An adaptation of the framework for possible testing..109

Tables

Table 1. Possible types of contemporary AI systems..... 19
Table 2. Different characteristics of AI systems in firm-level settings..... 23
Table 3. Presentation of different methodological views 50
Table 4. A methodological overview of the dissertation 59
Table 5. Overview of all data sources..... 63
Table 6. Description of interview data 64
Table 7. Illustrative reflections of trusteeship covenants to organize facets..... 90
Table 8. Topical agenda for future research 104

Abbreviations

IB	International Business
IS	Information Systems
AI	Artificial Intelligence
AIs	AI Systems
I-R	Global integration-local responsiveness
LCT	Loose coupling theory
FSAs	Firm-specific advantages
GRAO	Globally responsive AI organizing

List of Publications and Author's Contribution*

The summary dissertation consists of three articles.

[Article I] **Hasan, R.,** & Ojala, A. (2024). Managing artificial intelligence in international business: Toward a research agenda on sustainable production and consumption. Thunderbird International Business Review, 66(2), 151-170. https://doi.org/10.1002/tie.22369

Rakibul Hasan: Conceptualization, data curation, methodology, formal analysis, visualization, writing – original draft, and writing – review and editing. **Arto Ojala:** Supervision and writing – review and editing.

[Article II] **Hasan, R.,** Ojala, A., Quach, S., Thaichon, P., & Weaven, S. (2025). The dark side of AI anthropomorphism: A case of misplaced trustworthiness in service provisions. Proceedings of the 58th Hawaii International Conference on System Sciences. https://hdl.handle.net/10125/109530

Rakibul Hasan: Conceptualization, data curation, methodology, investigation, formal analysis, visualization, writing – original draft, and writing – review and editing. **Arto Ojala, Sara Quach, Park Thaichon, and Scott Weaven:** Supervision and writing – review and editing.

[Article III] **Hasan, R.** (2026). Artificial intelligence and anthropomorphism: A case of risks of harm on the global stage. Unpublished manuscript.

Sole authorship

Note: *Authors' contribution statement using CRediT (see https://casrai.org/credit)

1 INTRODUCTION

Development and deployment of artificial intelligence (AI) across national borders is transforming organizing at the firm level, creating space for novel theorizing in international business (IB) and information systems (IS) scholarship. AI-driven transformation is widely recognized in existing conversations in both IB (Buckley et al., 2026; Ratten et al., 2024) and IS (Asatiani et al., 2021; Lyytinen et al., 2021). Based on various advanced computing approaches such as machine learning, computer vision, and generative pre-trained transformer, contemporary AI is increasingly being embedded into international business activities. Defined as the frontier of computational advancements to address decision-making problems with or without human intervention, AI is an evolving frontier of organizing systems rather than a discrete tool (Berente et al., 2021). However, AI increasingly manifests through diverse contemporary forms broadly encompassing robotic, virtual, embedded, and algorithmic representations. These forms vary according to physical (dis)embodiment, unique identity, and anthropomorphic (human-like) features (Glikson & Woolley, 2020; Hasan, Thaichon, et al., 2021). For example, robotic AI navigates the complex physical world, such as in autonomous vehicles equipped with computer vision. Virtual AI, in contrast, operates within digital environments through applications such as conversational interfaces that leverage natural language processing. Embedded AI is integrated into devices like facial recognition cameras, enabling local data processing, while algorithmic AI includes systems like recommendation engines that analyze user behavior behind the scenes for personalization. These representations highlight the diverse ways in which AI is manifested across physical and digital environments (Hasan, Weaven, et al., 2021). However, contemporary AI in these forms is increasingly prevalent across national borders. The rising development and deployment of AI reflect its contemporary forms and global investment, projected to reach $632 billion by 2028 (International Data Corporation, 2024). Statista (2024) estimates an annual AI growth of 27.67%, reaching $826.73 billion by 2030. This widespread growth underlies actions within firms to address economic reasoning (Parkes & Wellman, 2015) and social well-being (Rahwan et al., 2019). This growth is important because AI offers enormous opportunities and critical challenges, depending on how managers strategically organize and employ it in international business activities. Therefore, any scholarly conversation of the strategic management of AI, and more specifically, the question of how AI can be strategically organized across borders in a globally responsive manner is particularly impactful.

Among existing conversations, strategic rethinking that emphasizes human-centric aspects of managing a system (Ali et al., 2025) and responsible organizing of evolving

foundational properties of AI in firm-level settings is significant (Berente et al., 2021). While differentiating contemporary AI is critical, focusing on its facets that are ingrained across various representations is more strategically important. Contemporary AI poses facets like learning and autonomy, distinguishing them from earlier technological advancements. These facets of AI provide managers with immense possibilities to design products or services and to innovate new capabilities, gaining firm-specific advantages (Ameen et al., 2024; Stelmaszak et al., 2025). Examining facets of AI is therefore paramount because they create different opportunities and institutional constraints in international markets. For example, virtual AI with an interactivity facet is often developed and deployed globally by default, based on an integrated architecture and trained on globally sourced data. In contrast, robotic AI with an appearance facet may embody local institutional value judgments and be trained on locally sourced, less representative datasets that subsequently shape system behavior. Similarly, algorithmic AI with a learning facet may rely on globally integrated models trained using synthetic data produced by another computational system. Facets of AI are highly relevant for both IB and IS. In IB, they influence how managerial actions within firms respond to uncertainty across countries stemming from environmental heterogeneity and institutional differences. In IS, they constitute managerial rationale for different socio-technical arrangements in response to firms' environmental and strategic choices. Emerging pervasive dynamics of facets of AI challenge existing schemas in theorizing, and there are novel opportunities to join existing theoretical conversations to achieve the common good.

The context of the common good can be understood as the collective well-being of society, aligning with justice and responsibility, while also generating rents and making profits for survival and growth (Ali et al., 2013; Fuso Nerini, 2026). Understanding how to organize facets of AI across borders in the context of the common good is paramount for information management discourse in IS (Berente et al., 2021; Zheng & Jarvenpaa, 2021), where both opportunities and challenges are simultaneously internationalized. Explaining strategic organizing of AI in cross-border activities at the firm level is also particularly relevant for international strategy discourse in IB (Bartlett & Ghoshal, 1989; Nambisan & Luo, 2021). Managing AI across borders, therefore, requires moving beyond country-level adaptation or compliance, and more towards a responsible, globally coordinated strategic organizing for responsiveness (Nambisan & Luo, 2021). Accordingly, firm-specific actors like top and middle managers may need to rethink how to organize AI strategically to be globally responsive to integrate economic reasons with social imperatives. At the same time, regulators may want to enable responsibility to develop and deploy contemporary AI across markets to serve the common good (Ali et al., 2025; Bartneck et al., 2021). For example, managers need to develop AI-based assets and deploy them in user-facing services while remaining responsive to local

institutions, and regulators need to institute a shared framework to evaluate emerging facets of AI and prevent risks of harm to ensure user well-being (the AI Act, European Commission, 2024). While managers navigate such tensions and complexity across borders, their actions in assuming greater responsibility are crucial and cannot be ignored in internationalizing activities (Tatarinov et al., 2023). Therefore, conversations that conceive AI as an evolving frontier of organizing capability rather than a single entity provide a more precise analytical basis for understanding and explaining (Chalmers et al., 2026).

As ongoing managerial capabilities, embedded into products, or offering services by itself, contemporary AI underpins economic reasoning in cross-border business activities (Brynjolfsson et al., 2019; Messner, 2022a). In digital globalized markets, managerial actions are embedding facets of AI into user-facing activities, enabling it to increasingly mediate interaction between firms and users. As a result, AI-enabled services and experiences are becoming more pervasive across geographic and national boundaries. However, AI does not facilitate utopia. Accompanied by opportunities, facets of AI create many challenges for managers, who face challenges from AI's expanding scope of affordance as well as (un)intended consequences stemming from human-like anthropomorphic features (Mori et al., 2012; Schleidgen et al., 2023; Zheng & Jarvenpaa, 2021). To enhance user experience, managers may enable advanced computing to pursue AI anthropomorphism, in which users tend to attribute human qualities to computing systems. In turn, managerial actions around contemporary AI pose risks of harm to users (Bartneck et al., 2021; Carroll et al., 2023; Danaher, 2020). Accordingly, AI raises concerns for regulators about deception, accountability, and user well-being that underscore social imperatives, regardless of national boundaries (European Commission, 2024; Wood, 2021). Contemporary AI in its different forms is not necessarily confined to a national boundary and can be deployed across borders. These forms of contemporary AI with evolving facets may then be deployed in markets across borders with heterogeneous institutions such as norms, culture, and regulations, with minimal adaptation or little regard to national borders. Accordingly, there is a growing concern about misalignment and user vulnerabilities, highlighting negative consequences that extend beyond national borders (Bartneck et al., 2021; Jobin et al., 2019). Collectively, both economic reasoning and the social imperatives of AI in cross-border activities underpin the context of the common good (Ali et al., 2025; Cowls et al., 2021).

Managers are responsible actors within firms for all major decisions regarding the development and deployment of AI at home and abroad. They coordinate for AI-based capabilities that are less location-bound and transferable across borders, and supervise capabilities to be embedded into products or services, enabling consumption convergence in digital globalized markets. Managers also use AI for

decision-making and customer targeting (Messner, 2022c, 2022a). They monitor and adjust decisions, processes, and routines appropriate for development and deployment (Lyytinen et al., 2021; Sobolev, 2023), as well as allocating resources, overseeing projects, and governing AI (Berente et al., 2021). Collectively, this range of managing activities underpins managerial actions around AI, which in turn shape the future and the common good. For managerial actions, AI introduces a variety of challenges in internationalizing activities that are socio-technical in nature. Many challenges are technical, including solving issues related to safety, black-box problems (difficulties in explaining and interpreting), and implementing measures to prevent negative outcomes (Rahwan et al., 2019; Rai, 2020). However, many challenges are linked to responsible issues, particularly around the deployment, usage, and consumption of products and services. These include concerns about deception, algorithmic fairness, and the risk of harm (Danaher, 2020; Dolata et al., 2022; van den Broek et al., 2025). Recent reports in popular media suggest that firms using AI may be involved in predatory and exploitative practices that result in significant harm. For example, some incidents involve sexual intimacy, violence, self-harm, death, and murder, including the exploitation of vulnerable users (Burga, 2025; Horwitz, 2025; Jargon & Kessler, 2025; Marsden, 2025). Therefore, there is an increasing call for accountability where firms are deemed responsible for the possible dire consequences of AI in many instances (Buolamwini & Gebru, 2018; Tóth et al., 2022; Wood et al., 2025). However, accountability also lies with managers. Managerial actions are fundamental to communicating, leading, coordinating, supervising, and controlling firm-level efforts to navigate complexity that arises from diverse institutions in international markets (Aguinis & Gabriel, 2022; Eden & Nielsen, 2020). Simultaneously, it is critical for managers and firms to be reflective and careful to avoid inflicting harm and negative consequences.

Against this backdrop, advancing conversations on AI at the intersection of international strategy and information management discourse may help managers gain a systemic understanding to promote the common good. Research on strategic organizing of AI in IB and IS remains limited, despite growing relevance. A relatively new conversation on strategic organizing of AI is emerging in the general strategic management discourse (Hutzschenreuter & Lämmermann, 2025; Stelmaszak et al., 2025). However, existing discourse maintains a focus on economic goals in its underlying reasoning while overlooking social imperatives and cross-border business activities, both of which are major obstacles to advancing scholarly conversation to protect the common good. Uncertainty thus remains whether managerial actions on AI within firms' cross-border activities will support or hinder global sustainable development. Despite its growing relevance to both theory and practice, research on how to strategically organize AI across borders remains limited, particularly that which integrates economic reasoning with social imperatives. In

response, this summary dissertation asks an impact-oriented, overarching research question to advance scientific understanding, which is further explored in two interrelated sub-questions: ***How can AI be managed across borders in a globally responsive manner to integrate economic and social goals?***

> **Sub-question 1:** *How can AI development and deployment be managed to achieve responsiveness in international business activities?*
>
> **Sub-question 2:** *How can AI be managed to safeguard users' well-being through enabling responsible development and deployment across borders?*

An understanding of managing AI across borders in a globally responsive manner is crucial for both international strategy and information management discourse. Such an understanding requires attention to facets of AI that are ingrained across different contemporary representations, including robotic, virtual, embedded, and algorithmic forms. The rationale is that facets of contemporary AI are increasingly a core enabler of value creation, competitiveness, or cross-border coordination. Addressing the first sub-question advances international strategy discourse to explain a possible processual mechanism to integrate economic reasoning with societal imperatives across diverse institutional environments. It also advances information management discourse by explaining facets of AI to ensure adaptive, trustworthy, and locally sensitive development and deployment. The second sub-question directly addresses governance challenges posed by AI in the global marketplace, thereby enriching both discourses. It outlines potential regulatory interventions to develop and deploy AI that can safeguard user well-being, minimize the risk of harm, and facilitate accountability across markets. Therefore, answering these questions collectively constructs an organizing framework for responsiveness across borders. The framework aims to bring economic competitiveness and social welfare, thereby fully addressing the central research question. To achieve this, this dissertation embraces reflexivity, theoretical pluralism, and methodological complementarity to pursue knowledge. It is also situated in transdisciplinary AI discourse while remaining grounded in IB and IS.

The framing draws on theoretical pluralism that couples relevant underlying assumptions from established theoretical and contemporary conversations in both international strategy and information management discourses. In IB, conversation includes global integration-local responsiveness (Bartlett & Ghoshal, 1989), and the loose coupling systems involved (Nambisan & Luo, 2021). In IS, it includes the AI frontier (Berente et al., 2021) and technology anthropomorphism (Zheng & Jarvenpaa, 2021). However, it also engages with transdisciplinary AI discourse that includes technology harm (Wood, 2021), the AI Act (European Commission, 2024), and the trusteeship framework on AI ethics (Ali et al., 2025). To corroborate and align

these theoretical conversations, this dissertation adapts a systems-oriented view of methodological complementarity to seek knowledge (Arbnor & Bjerke, 2009). The contextual boundary involves managerial actions within firms' cross-border activities and responsibility for consumption or usage to achieve the common good. The empirical unit of observation is multilevel, encompassing managers, users, and contemporary forms of AI. The unit of analysis is the organizing facets of AI used in development and deployment to enable globally integrated responsible organizing for responsiveness across national borders. Subsequently, this dissertation constructs and explains a systems-oriented framework to organize facets of AI in a globally responsive manner, namely the 'Globally Responsive AI Organizing (GRAO)'.

The framework offers ongoing managerial capabilities for responsiveness across borders, underpinning firm-specific advantages. It construes four evolving and interrelated facets centering on (1) *autonomy*, (2) *interactivity*, (3) *learning,* and (4) *appearance* to develop and deploy contemporary AI. It then enacts these facets with a shared value of trusteeship organizing to achieve ongoing managerial capabilities for responsiveness across borders. It institutes three mutually managerial trusteeship covenants that are global by default, that intentionally prioritize proactiveness in (a) *addressing institutions*, (b) *blocking harmful means*, and (c) *ensuring accountability*. Accordingly, trust is baked into managing the scope of affordance and performance of AI. Therefore, the summary dissertation contributes to knowledge on theory, practice, and regulations, and advances science at the nexus of international strategy and information management discourse. Theoretically, it constructs and explains an emerging framework for managing AI across borders to achieve responsiveness. Accordingly, it broadly engages with two classical streams of conversation within IB and IS, which have inherited a multidisciplinary nature. In IB, it contributes to international strategy discourse that focuses on worldwide effectiveness (Bartlett & Ghoshal, 1989; Birkinshaw, 2022; Nambisan & Luo, 2021). In IS, it contributes to information management discourse that focuses on organizing advanced technological artifacts in organizational settings (Berente et al., 2021; Zheng & Jarvenpaa, 2021). Figure 1 illustrates the theoretical engagement and contributions to both IB and IS. However, the proposed theorization is action-oriented, making the understanding highly relevant for managers and regulators.

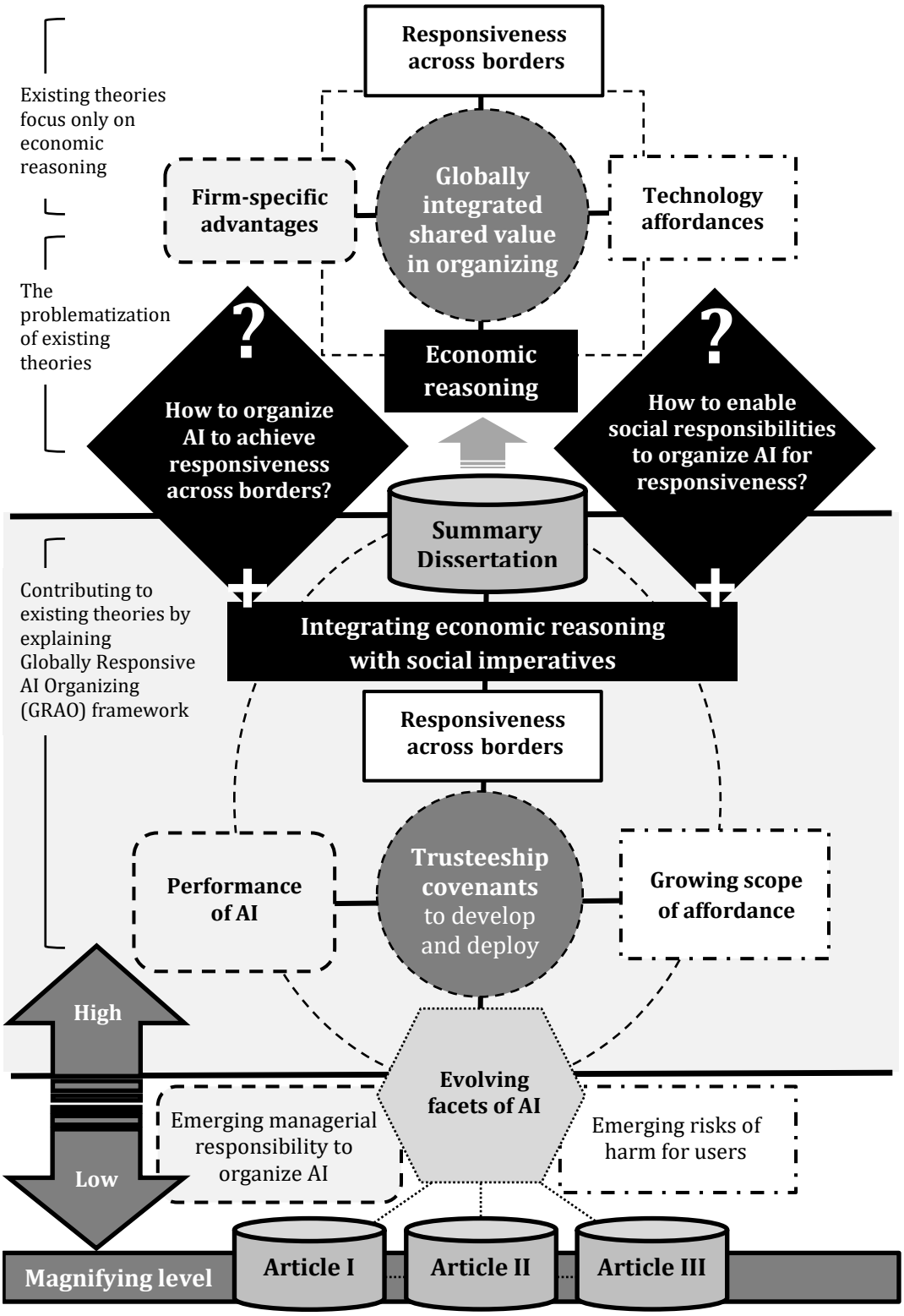


Figure 1. Theoretical engagement and contribution

The summary dissertation has important implications for top and middle managers across firms to use AI in international activities, and to configure AI-enabled solutions, including new ventures, small and medium-sized enterprises, and multinational enterprises. The framework guides managers with concrete stewardship to develop and deploy AI that includes delegating affordances, configuring firm-specific advantages, governing for competitiveness, and protecting the common good. Additionally, it provides implications for regulations to address emerging accountability and psychological safety, and recommends how to reduce the risk of harm and safeguard user well-being, thereby supporting managers – especially those overseeing user-facing applications. Collectively, the summary dissertation calls for a cognitive shift in existing discourse from 'trustworthy AI' to 'trusteeship organizing of AI' to achieve coherence and ensure accountability across markets. Finally, it identifies several avenues for future research to emphasize methodological complementarity and outline new issues and perspectives to advance research in IB and IS. Besides creating impact for theory and practice, it also has wider relevance. Particularly, it aligns with the Fifth Industrial Revolution, which promotes human-centric, sustainable technology (European Commission, 2022). Unlike the Fourth Industrial Revolution, the Fifth emphasizes ethical and responsible AI development and deployment. The strategic integration of economic and social reasoning supports the United Nations' (2023) Sustainable Development Goals (SDGs), particularly SDGs #12 and #16. These goals are vital for achieving global sustainability through firms' cross-border activities, emphasizing responsible consumption and production, as well as peace, justice, and strong institutions (Marano et al., 2024; Montiel et al., 2021; Vinuesa et al., 2020). This collective effort hopes to achieve common good in these contexts.

This summary dissertation is structured as follows. After the introduction in Chapter 1, Chapter 2 provides an overview of the theoretical underpinnings and conceptualization of the dissertation. Next, Chapter 3 presents the research methodology, including paradigmatic standpoints, research design, and data collection, analysis, and presentation. After, Chapter 4 summarizes three articles included in the dissertation. Finally, Chapter 5 explains the constructed framework in relation to the central research question, and provides understandings to guide theory, practice, regulations, and future research opportunities.

2 THEORETICAL BACKGROUND

Positing on firms' activities across national borders, this summary dissertation engages in a theoretical conversation on strategic management at the intersection of IB and IS. In IB, this chapter broadly engages with strategic organizing discourse in international strategy that focuses on worldwide effectiveness (Bartlett & Ghoshal, 1989; Birkinshaw, 2022; Nambisan & Luo, 2021). In IS, it engages with the strategic organizing of advanced information technology artifacts in information management (Berente et al., 2021; Zheng & Jarvenpaa, 2021), and also engages with wider discourse on responsibility to embrace transdisciplinary research on AI (Ali et al., 2025; Wood, 2021). Instead of gap-filling, however, this chapter engages in *problematization* (Alvesson & Sandberg, 2011). This approach is well-established across disciplines, including IS, IB, and management studies (see Dolata et al., 2022; Kriz et al., 2023; Ramaul et al., 2025). Through problematization, it identifies assumptions for integration and politely challenges their underlying reasoning through contextualizing them. Accordingly, it engages with problematization within the contextual boundary of achieving the common good in the Fifth Industrial Revolution. The objective of this chapter is thus not to provide an exhaustive or systematic review of existing literature. Rather, it aims to emphasize the emergent nature of strategic organizing of AI that rethinks economic reasoning with considerations of societal well-being. However, it also serves as a baseline for conceptualization.

2.1 Contemporary Understanding of AI in International Business

2.1.1 Contemporary Research on AI in International Business

Although theoretical conversations of AI in international strategy discourse in IB are limited, the topic is not entirely new (Cavusgil & Evirgen, 1997; Veiga et al., 2000). In recent years, there have been growing efforts and renewed interest seen in various cross-border settings (Cerar et al., 2026; Messner, 2022a, 2022c; Yoon et al., 2025). Collectively, however, extant discourse mainly focuses on economic goals in its reasoning, with little regard for social well-being (Ciulli & Kolk, 2023; Tatarinov et al., 2023), which poses a challenge to achieve the common good in the Fifth Industrial Revolution. Hence, there are increasing opportunities to engage in novel theorizing in IB (Buckley et al., 2026; Ratten et al., 2024; Verbeke, 2026). However, there are also growing challenges in theorizing about possible negative consequences, and some new challenges include strategic concerns regarding violations of end-user

privacy, human development, and monopolistic competitive advantage (Benito et al., 2022; Ghauri et al., 2021), and emerging risks associated with consumption, responsibility, and regulation (Birkinshaw, 2022, 2024; Furr et al., 2022). While examining AI, there is also minimal conceptual clarity about what one means by AI and the types of contemporary AI one deals with (e.g., dos Santos & Williamson, 2024; Verbeke, 2026; Yoon et al., 2025). Thus, without clarity, reader (including scholars, managers, or regulators) may not find actionable insights in organizing. However, continuing such a conversation requires a clear conceptualization of what one means by AI.

Existing scholarly discourse in IB tends to assume that AI possesses human qualities in its conceptualizations, and similar observations have also been found in general management and organizational research (Ramaul et al., 2025). To illustrate, Ghauri et al. (2021, p. 6) view that "AI can mimic the human psychology, including empathy". Ciulli and Kolk (2023, p. 6) define AI as "the ability of machines to perform cognitive functions that mimic or exceed human cognitive skills". Hemphill and Kelley (2021, p. 589) refer to AI as "a system that makes autonomous decisions". Veiga et al. (2000, p. 225) assert that AI is "capable of sophisticated, perhaps intelligent, computation and pattern recognition similar to those that the human brain routinely performs". Buckley et al. (2026) settle on a machine's ability to perform cognitive functions associated with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem-solving, decision-making, and even creativity. However, these conceptualizations of AI, even from experts and scientists, suggest that humans are inherently inclined to anthropomorphize (i.e., attribute human traits, emotions, intentions, or behaviors to non-human entities, giving something non-human a human form or personality) as they tend to relate everything to their existing knowledge about themselves (Zheng & Jarvenpaa, 2021). This dissertation argues that anthropomorphic explanations in theorizing overlook the differences in actions and arrangements between AI and human agents (Ramaul et al., 2025). Nevertheless, existing discourse has been tackling AI across international borders implicitly or explicitly for more than two decades (Cavusgil & Evirgen, 1997; Grant & Phene, 2022; Nair et al., 2007; Veiga et al., 2000). While there is minimal research on strategic organizing in international strategy discourse, three broad streams explicitly and implicitly relate to AI within firms across borders.

The *first* stream conventionally considers AI an effective analytical or methodological tool for doing research and theorizing (Aguinis, 2026). Such a method-driven aspect of AI seems effective for pattern recognition in cross-border activities, compared with traditional statistical methods (Veiga et al., 2000). Across borders, managers may use AI to uncover hidden aspects of national cultures (Messner, 2022b), select international markets (Brouthers et al., 2009), evaluate global expansion decisions

(Nair et al., 2007), and analyze consumer behavior (Messner, 2022a; Schlegelmilch et al., 2023).

The *second* stream focuses on the economic scale and scope of generating rents and maximizing profit. Some economic opportunities include minimizing transaction costs through augmentation and maximizing efficiencies in global operations and the global value chain through automation (Gooderham et al., 2022; Menzies et al., 2024; Oliva et al., 2022). Additional evidence indicates that managers deploy AI to gain complementary advantages in internationalization activities (Brynjolfsson et al., 2019; Huo & Chaudhry, 2021; Yu et al., 2022). Deploying AI aids managerial decision-making, particularly in seamless transactions across national and cultural borders (Araghi et al., 2024). It also aids in selecting cooperative venture partners (Cavusgil & Evirgen, 1997), facilitates knowledge transfer (Grant & Phene, 2022), and manages emerging political risk (Hemphill & Kelley, 2021).

The *third* stream points to managerial opportunities and challenges for societal goals that intersect global sustainable development. There is growing concern about the potential for harm arising from increased user interaction, implicitly suggesting AI anthropomorphism to attribute human qualities to AI (Ghauri et al., 2021; Kopalle et al., 2022). Simultaneously, there are opportunities to develop AI to configure transferable and scalable firm-specific advantages to tackle wicked problems and social good (Tatarinov et al., 2023). However, while a grave uncertainty remains about AI on international strategy across issues of social and environmental sustainability (Ciulli & Kolk, 2023; Denicolai et al., 2021), the extant discourse remains almost silent on social imperatives, although a few exceptions can be seen. For example, scholarly efforts have focused on social goals that also consider economic reasoning (Ciulli & Kolk, 2023; Tatarinov et al., 2023). But extant conversations indicate that pervasive technological advancement like AI challenges existing schemas for theorizing in cross-border activities (Birkinshaw, 2022; Ghauri et al., 2021).

2.1.2 Contextualizing AI for the Common Good within Firms' Cross-Border Activities

This dissertation engages in a theoretical conversation on the strategic organizing of AI within firms' boundaries. Managerial actions enable boundary spanning within firms to enact international strategy in cross-border activities (Schotter et al., 2017). In IB, managerial actions for strategic organizing are driven by external and internal boundaries. These actions address external boundaries that arise from the liability of foreignness (Zaheer, 1995), as well as from formal and informal institutions such as culture, norms, and regulations (Dau et al., 2020, 2022). Additionally, these actions

address internal operational activities in internationalization (Ojala et al., 2018, 2023). In IS, however, managerial actions embody the nature of a socio-technical system within firms to address both external and internal boundaries (Fuenfschilling & Binz, 2018; Sarker et al., 2019). Within a firm's boundaries, managing AI in international business activities thus requires dual emphasis on instrumental and humanistic elements. Instrumental emphasis includes managerial capabilities in systems and processual activities, and the humanistic focus lies on social well-being, while being competitive (Montiel et al., 2021; Vinuesa et al., 2020). Therefore, this dissertation considers societal imperatives alongside economic goals on a global scale to contextualize its underpinnings to achieve the common good. This focus necessarily aligns with the Fifth Industrial Revolution and the global sustainable development agenda (European Commission, 2022; United Nations, 2023).

Within firms' cross-border activities, the common good can be understood as the collective well-being of society that aligns with justice, fairness, and moral responsibility, as well as generating rents and making profits for survival and growth (Ali et al., 2013; Fuso Nerini, 2026). The act of boundary spanning for the common good is highly relevant to the Fifth Industrial Revolution, which steers toward human-centric, sustainable development in technology production and consumption (European Commission, 2022). Unlike the Fourth Industrial Revolution, it emphasizes advanced computing, such as AI, that aligns with ethics, social well-being, and responsibility in technological development and deployment. Such a shift to achieve economic performance by safeguarding social reasoning underpins the common good across borders (Ali et al., 2025; Cowls et al., 2021; Vinuesa et al., 2020). In the context of the common good, strategic organizing of AI may be understood as the boundary spanning of managerial actions within firms involved in internationalization activities. It aims to establish conditions, practices, or policies that enable all stakeholders in society to thrive, including corporate entities that are developing and deploying AI across national borders, rather than prioritizing only economic reasoning for corporate gains and profit maximization, and argues for the integration of societal well-being.

Contextualizing the common good underpins managers' actions in international business activities that embody moral responsibilities to maximize overall goodness and minimize negative consequences (Montiel et al., 2021). Managers achieve this by reconceiving firm-specific competencies that address local social needs, leading to internationalization advantages. They also redefine productivity in the global value chain, which addresses social and environmental issues (Ali et al., 2025; Porter & Kramer, 2011). In turn, they support corporate integrity and safeguard users' well-being, including the vulnerable. Common good urges a shift in the contemporary discourse on advanced technology in IB, which has a sole focus on the economic

performance and well-being of internationalizing firms (Luo, 2022; Verbeke & Hutzschenreuter, 2021). In this shift, managerial actions around AI focus on empowering users and improving their well-being, not to replace humans or cause harm. Reflecting on this context, developing and deploying AI enhances human rights and dignity, and ensures equitable property across national borders. But managing AI in cross-border activities demands new organizational forms and practices for responsible stewardship, in order to build trust and social justice.

In summary, the theoretical discourse on AI in IB is not new, and is productive and growing. While it provides little knowledge on strategic organizing of AI, it emphasizes economic reasoning that includes production efficiency, cross-market learning, capability building, and economies of scale in its international expansion activities. However, the sole focus on economic motives, without considering social imperatives, underscores problematic trends to advance IB as a scientific field. There is a growing need for social imperatives to achieve the common good, including privacy protection, regulatory frameworks, accountability, and addressing emerging inequalities across institutional settings. Accordingly, this dissertation focuses on social imperatives alongside economic reasoning, and theorizing AI within socio-technical systems. Doing so requires drawing an understanding from information management discourse in IS (Berente et al., 2021; Zheng & Jarvenpaa, 2021), and it is also essential to conceptualize AI, identify its type and temporality, and contextualize it to make a meaningful engagement.

2.2 Foundations for Theorizing AI in Cross-border Activities

2.2.1 AI as an evolving phenomenon of computation

AI is considered a general-purpose technology similar to electricity, steam engines, and internal combustion engines (Berente et al., 2021; Lyytinen et al., 2021). One must comprehend that the terminology of ’artificial intelligence’ or AI has diverse meanings and definitions (Russell & Norvig, 2021; Skilton & Hovsepian, 2017). As per the Oxford English Dictionary definition (n.d.), it means "*the capacity of computers or other machines to exhibit or simulate intelligent behavior; the field of study concerned with this*". Consequently, it redirects two broad conceptions. AI relates to a multidisciplinary scientific field in its own right that emerged in the mid-1950s (McCarthy et al., 1955), and examines learning or any other feature of intelligence simulated in machines. As a field of scientific research, it combines principles from computer science, mathematics, cognitive science, linguistics, neuroscience, and

other related disciplines. It encompasses various areas, including machine behavior, perception and control, machine learning, machine cognition, anthropomorphism, AI ethics, and human-machine interaction (to name a few). However, AI also relates to computational systems, which is the focus of this dissertation. AI is related to the advanced computing that emerged in the early 1950s (Turing, 1950), yet is an evolving computational advancement, similar to other human-devised general-purpose technologies (Berente et al., 2021; Stelmaszak et al., 2025). The contemporary form of AI may consist of various computational capabilities (see Figure 2 for some examples), such as computer vision, natural language processing, large language models, and machine learning (Hasan, Thaichon, et al., 2021; Hasan, Weaven, et al., 2021). As a general-purpose technology, AI can also be applied in research methods, which is not the primary focus of this dissertation (e.g., Messner, 2022a; Veiga et al., 2000). Such a view shapes how one views and theorizes AI within cross-border activities in the context of the common good.

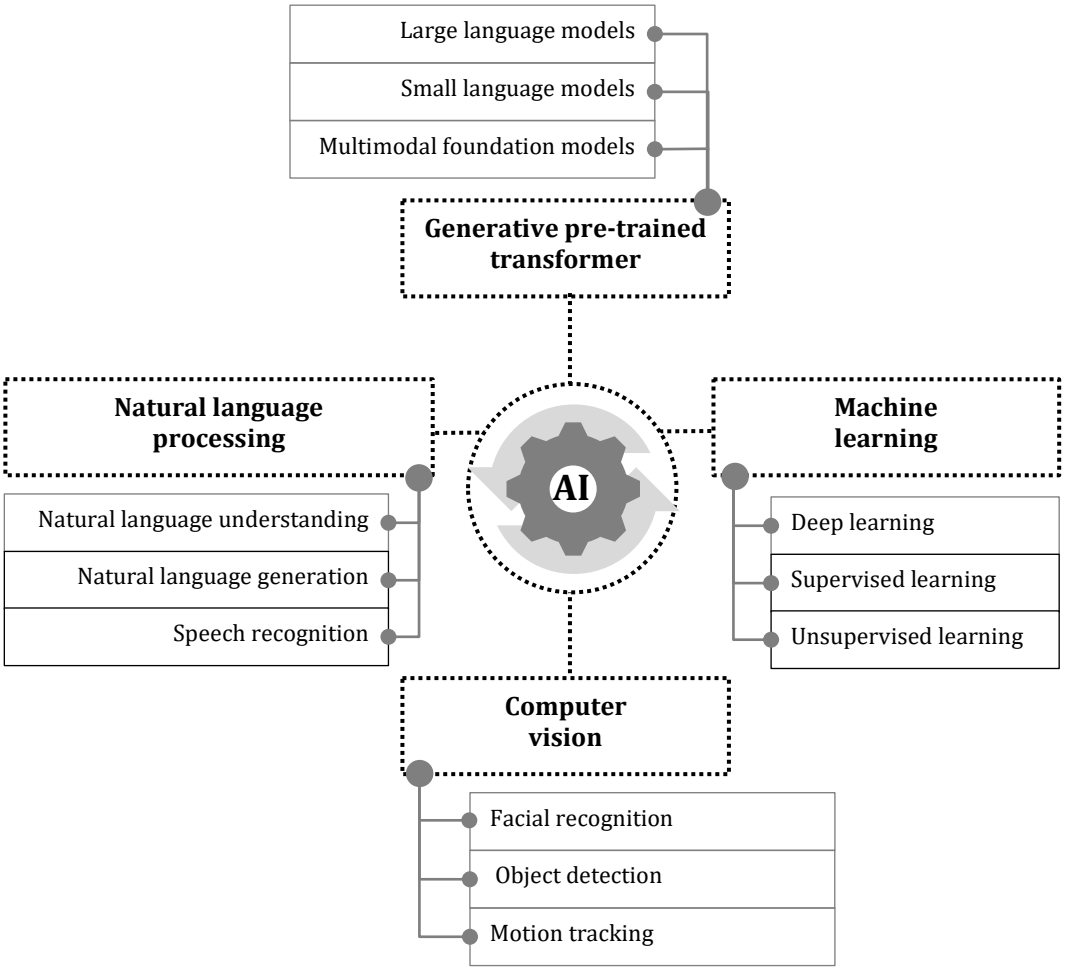


Figure 2. Examples of given computing capabilities of AI

This dissertation defines AI *as a frontier of computational advancements to address decision-making problems in both physical and digital environments, with or without human intervention.* Being at the frontier of computational advancements, AI is not a single entity or a discrete tool, and is not confined to a device, program, or algorithm (Berente et al., 2021). Instead, AI is a dynamic concept and may be ideated as a moving object. The notion of frontier focuses on an evolving conception within the narrow AI, also known as weak AI (Duffy, 2003; Hasan & Rana, 2018). Unlike general AI which aims to emulate human intelligence across various activities, narrow AI is focused, task-oriented, and operates within a given context (Bostrom, 2016; Zheng & Jarvenpaa, 2021). Computing enables machines to perform tasks and solve both simple and complex problems. Such functions can be engineered to learn, evolve, and develop with extensive, minimal, or no human intervention (Rahwan et al., 2019; Russell & Norvig, 2021). AI is a socio-technical system with an evolving computational process, and forms an example of the fluid nature of computational advancement and constant technological improvement. It advances towards moving objectives of evolving phenomena through computing-based learning capabilities, such as machine learning (Hutzschenreuter & Lämmermann, 2025; Leavitt et al., 2021). What most would consider AI today may not be considered as such in the future. For instance, Alan Turing broke the Enigma code with his machine in the 1940s—an almost impossible task for a single human being. While his machine performed the calculation, he did not consider it to be 'intelligent'. However, he pointed to the evolving development of digital computing and learning machines toward an ambiguous target of intelligence (Turing, 1950). Similarly, in an IB setting, rule-based expert systems were considered as AI in the 1900s to tackle internationalization decision-making problems (Cavusgil & Evirgen, 1997). But many today would not consider them as AI, even though at that time they were. While AI is laminar, there is always some form or representation that temporarily emerges in global markets.

2.2.2 Contemporary AI Systems and their Representations

Given their inherited socio-technical nature, narrow capabilities of AI enable advanced technological artifacts to perform goal-directed or a specific set of tasks, such as information exchange, facial recognition, or language translation (Hasan, Thaichon, et al., 2021). These AI-enabled information technology artifacts (which this dissertation refers to as AI systems) represent a contemporary form available in the global marketplace. A system is a constellation of interconnected components organized into subsystems governed by interactions and relationships among them (Orton & Weick, 1990; Simon, 1962). It may consist of subsystems and components that can be open, closed, or both simultaneously governed by modularity (Banalieva

& Dhanaraj, 2019; Ojala et al., 2018; Yoo et al., 2010). This perspective on AI shifts focus from mere tools to systems that learn and maintain context-preserving state across various tasks, and act autonomously over time. Essentially, such systems not only respond, but they maintain continuity to evolve through data iteration and recalling previous instructions, appearing less like traditional information technology artifacts. However, AI systems are developed and deployed within a firm's boundary to operate in both physical and digital environments (Hasan, Weaven, et al., 2021; Lyytinen et al., 2021). Such a view aligns with the European Union AI Act (2024), wherein: "*AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*" (see Article 3).

Hence, this dissertation conceives AI systems (hereafter, AIs; see Figure 3) as *contemporary forms of AI that perform goal-directed tasks and exhibit given human-like characteristics to simulate responses and influence user perceptions and behaviors through (dis)embodied representations.* This conception takes a user-centric approach and focuses on physical embodiment (Glikson & Woolley, 2020) that addresses representations of distinctive identity and operational embeddedness, whether within digital and/or physical environments (Hasan, Weaven, et al., 2021). AIs vary in their representations based on whether they have a physical shape, (dis)embodied forms, distinct identity, and anthropomorphic (human-like) cues (Hasan, Thaichon, et al., 2021). The current conception yields four interrelated contemporary representations of AIs, namely *robotic*, *virtual*, *embedded*, and *algorithmic*.

Robotic AIs. Reflecting on materiality and hardware, *robotic* AIs navigate the complex physical world. For example, physically embodied robotic AIs may appear like humans, including humanoid robots with affective computing. Robotic AIs can also be formed into non-humanlike physical artifacts like autonomous vehicles equipped with computer vision and vacuum-cleaning bots with machine perception. These artifacts interact with the world using AIs to perform tasks autonomously or semi-autonomously. To illustrate, 1X's NEO, featuring a humanoid representation, and Baidu Apollo Go, featuring a driverless autonomous vehicle, are examples of robotic AIs that perform service tasks in a physical environment. Common examples also include ABB's robotic arm GoFa, Unitree's G1, Boston Dynamics' Altas, and Tesla's self-driving cars.

Virtual AIs. Reflecting on algorithms and software, *virtual* AIs operate in the digital world and pose a distinguished identity, either in embodied or disembodied forms.

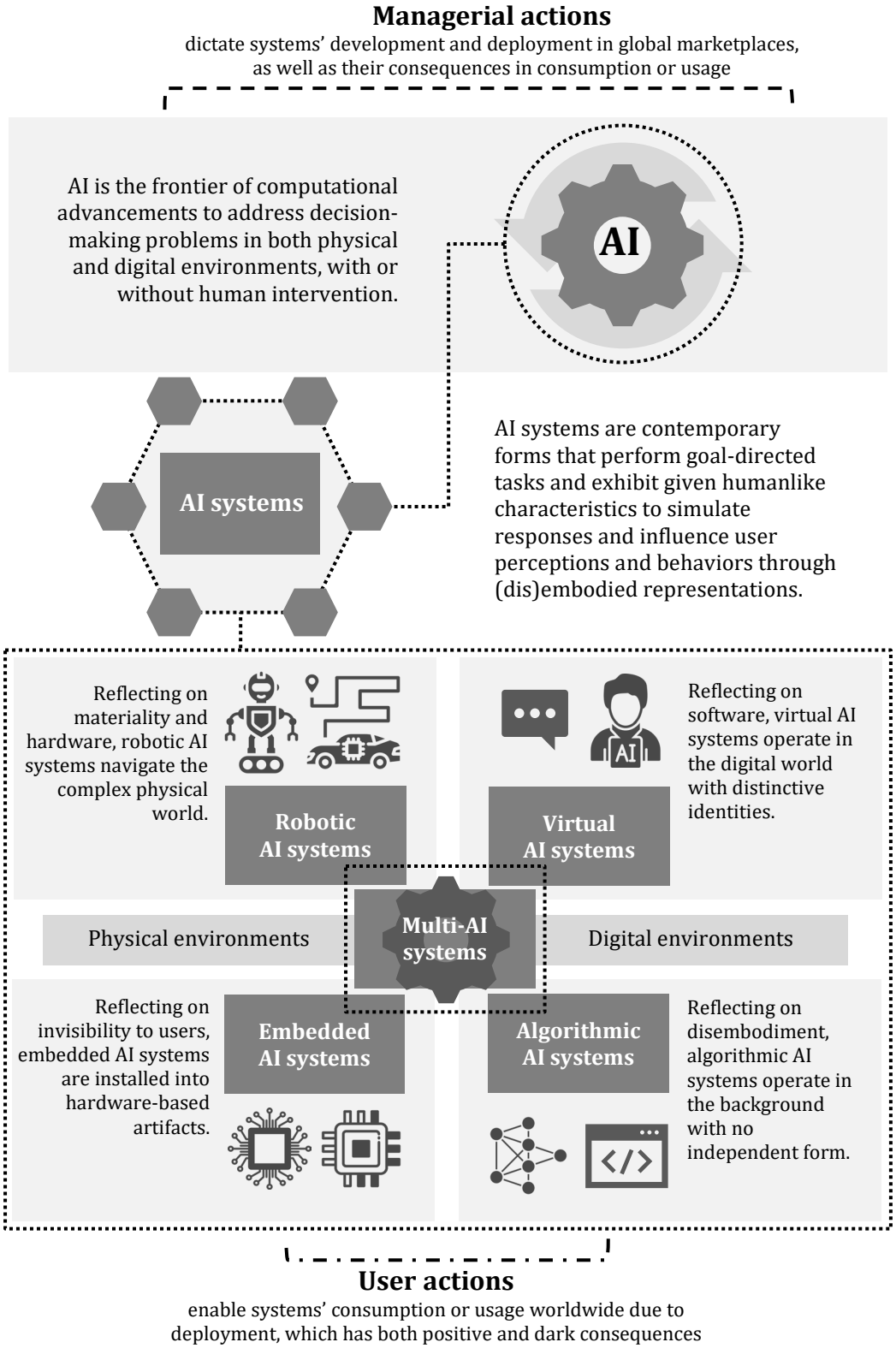


Figure 3. Definition and contemporary representations of AI

For example, virtually embodied AIs can be designed to appear like humans, including avatars with voice and facial recognition capabilities. Virtual AIs can also be disembodied but may appear humanlike through conversations and autonomous responses, powered by large language models and natural language processing that generate statistically likely continuations of word sequences. These artifacts often simulate human interaction or perform tasks within virtual environments and may operate through cloud computing. To illustrate, Soul Machines' avatar with embodied representations and Moonshot AI's Kimi K3 with no embodied representations are forms of virtual AIs that are reasonably adept at facilitating interactive information exchange in digital environments. Common examples also include DeepSeek's reasoning model R1, Apple's Siri, Amazon's Alexa, customer service bots, and the Replika chatbot.

Embedded AIs. Reflecting on both hardware and software, embedded AIs do not necessarily need a visual representation or a distinguished identity and are often invisible to users. For example, edge computing on low-Earth-orbit satellites processes data in orbit using onboard AIs and hardware to analyze images, filter useless data, and make real-time decisions instead of sending raw files to Earth. In such a contemporary form, embedded AIs often operate behind the scenes to enhance functionality without being visible and do not necessarily have an independent form. To illustrate, Apple Watch uses AIs locally to instantly detect irregular heart rhythms and track sleep stages. Common examples also include Volvo's driver understanding system and Apple's Face ID (facial recognition), where AIs run on dedicated chips within hardware systems.

Algorithmic AIs. Similar to embedded AIs, algorithmic AIs may pose neither visual representation nor an independent form but are often integrated into software systems. For example, Uber's dynamic pricing is an algorithmic decision-making system that statistically adjusts prices using real-time usage (demand and supply) data, especially during peak hours, major local events, or after significant weather events. Common examples also include IBM Watson, AI-enabled robotic process automation, credit scoring systems, Allianz's automatic claim processing, Zalando's product recommendation systems, Spotify's recommendation systems, and Alphabet's Google Search, which assist users in retrieving and processing information through machine learning-driven algorithms.

Robotic AIs, virtual AIs, embedded AIs, and algorithmic AIs represent distinctive forms, and differentiating them is important. Each form operates through distinct reasoning and within machine-machine or human-machine interaction, triggering different machine behavior and outcomes. Contemporary AIs are also known as AI agents (Rahwan et al., 2019). or agentic AI (Stryker, 2025) in academia and industry.

Table 1 provides a brief understanding of different types of contemporary AIs in firm-level settings, conceptualized in extant transdisciplinary discourse.

Table 1. Possible types of contemporary AI systems

Reference	Focus	Different types of contemporary AI systems
(Parkes & Wellman, 2015)	Rationality and design rules	*AI agent/systems:* A moving conception of synthetic rational entities that perceive the world and act in it, focusing mainly on disembodied and algorithmic forms. *Multi-agent AI:* Agents perform specific actions and tasks, and their outputs aggregate into overall outputs.
(Rahwan et al., 2019)	Machine behavior	*Virtual AI:* A form of representation with no physical presence. *Embodied AI:* Machine embodies physical properties to perform mechanical or socially oriented tasks.
(Glikson & Woolley, 2020)	User-centered and physical embodiment	*Robot AI:* Different mechanical or human-like representations to perform socially oriented tasks. *Virtual AI:* A form of representation with no physical presence but has a distinguished identity. *Embedded AI:* Invisible to users but does not have a visual representation or a distinguished identity.
(Huang & Rust, 2021, 2024)	Customer-centric and service tasks	*Mechanical AI:* Learning and adapting only to a minimal degree but aiming to maximize efficiency. *Thinking AI:* Learning and adapting from data, which can be analytical or intuitive. *Feeling AI:* Learning and adapting from data, relating to contextual- and individual-specific experience.
(European Commission, 2024)	Risk-based approach	*AI systems:* A machine-based system with varying levels of autonomy and may exhibit adaptiveness for explicit or implicit objectives that can influence physical or virtual environments.
(Pan et al., 2025)	Design and use	*AI agents:* Software agents that are capable of perceiving the world, possess reasoning abilities, and act autonomously to follow given instructions.
(Chalmers et al., 2026)	Core capabilities and distinctive affordances	*Predictive AI:* Pattern recognition and efficiency in routine analysis. *Generative AI:* Synthetic production of text, images, code, audio, multimodal artifacts, and content abundance. *Agentic AI:* Multi-step reasoning, task decomposition, partial autonomy, and simulation of cognitive workflows. *Embodied AI:* Physical manipulation, mobility, physical task execution, and embodied coordination.
(Sapkota et al., 2026)	Autonomy, task complexity, collaboration, adaptation	*AI agents:* Autonomous software programs that perform specific tasks with autonomy, task-specificity, and reactivity. *Agentic AI:* Systems of multiple AI agents collaborating to achieve complex goals.
Current conception	User-centric, physical embodiment, distinct identity, and anthropomorphic features	*Robotic AIs:* Navigate in the physical world, may have both embodied humanlike and non-humanlike representation. *Virtual AIs:* Operate in digital environments and pose a distinctive identity. *Embedded AIs:* Invisible to users and integrated into hardware and software systems that process data locally. *Algorithmic AIs:* Operate in digital environments, often working in the background with no independent form.

Development and deployment of these contemporary forms often involve integrating multiple computing capabilities and AIs to perform goal-directed tasks with minimal or no human intervention. For example, robotic AIs (e.g., autonomous vehicles) with computer vision can be combined with virtual AIs with generative pre-trained transformer to enhance support machine perception and voice-based interaction during driving. Similarly, different AIs can be integrated into larger configurations, analogous to satellite constellations in which multiple satellites work collectively to provide global coverage and services. When several AIs are coordinated to achieve a specific objective with limited human supervision, they form a computational framework known as multi-AI systems (also referred to as multiagent systems or agentic AI systems) (Chalmers et al., 2026; Parkes & Wellman, 2015; Sapkota et al., 2026). In this representation, each form functions as a specialized subsystem assigned for particular tasks while collaborating with others to address a broader problem.

A notable point is that, although the terms 'agent' and 'agentic' are often associated with the notion of computational agency (Stryker, 2025), this summary dissertation refrains from using such labels. Anthropomorphic vocabulary such as 'agent', 'agency', 'think', and 'feel' imply characteristics of living, sentient beings (Shanahan, 2024), which contemporary AIs do not possess, per conceptualization (see Table 1 for a brief comparison). Avoiding anthropomorphic language is also important for reducing hype and preventing conceptual fallacies around AIs (Placani, 2024). For AIs, conceptual explanations require careful attention to the possibility that a term may be misaligned with its intended meaning or be misunderstood or misinterpreted in the context in which it is used (Grimm, 2021). Rather than conceiving AIs as a singular entity, this dissertation emphasizes the evolving nature of computational advancements. Accordingly, the labels used to describe contemporary AIs are intended to capture systems that exhibit increasing adaptability and responsiveness. These contemporary AIs embody facets that can be embedded within firms and deployed across firm-level activities.

2.2.3 Theorizing AI by Drawing on Socio-technical Systems

While differentiating AIs based on physicality, (dis)embodiment, distinct identity, and anthropomorphic features is important, this summary dissertation focuses on characteristics that may be ingrained across the various representations discussed above. It conceptualizes AIs not as a single, discrete entity but as an evolving frontier of computational capabilities. This perspective provides a more precise analytical foundation for understanding and explaining the phenomenon under investigation (Berente et al., 2021; Chalmers et al., 2026). Accordingly, this moving idea of AIs

makes two relevant points. *Firstly,* the logic of frontier posits that AIs are always liminal and emerging, thereby continually redefining its boundaries. Hence, managing AIs is *"a constant process of emergence and performativity"* (Berente et al., 2021, p. 1437). This understanding integrates two dimensions—*performance* and *scope* (see Figure 4). The dimension of *performance* reflects the recent ability of machine learning to tackle increasingly complex tasks. The dimension of *scope* reflects the growing ubiquity of AIs across an ever-expanding range of contexts. Therefore, as increased performance is accompanied by increased scope, managers need to manage both performance and scope of AIs as the frontier of advanced computing. Management of AIs involves developing, governing, deploying, and controlling, thereby involving decision-making at the firm level. As AIs advance and solve increasingly complex decision-making problems, their contemporary forms differ qualitatively from previous generations. For example, while past approaches to AIs have relied on rule-based decision-making, current approaches center on self-learning and transformer models that may support uses in many aspects with minimum human intervention. Particularly, AIs relate to managerial aspects of vast information processing in cross-border activities (Luo, 2022; Verbeke, 2026).

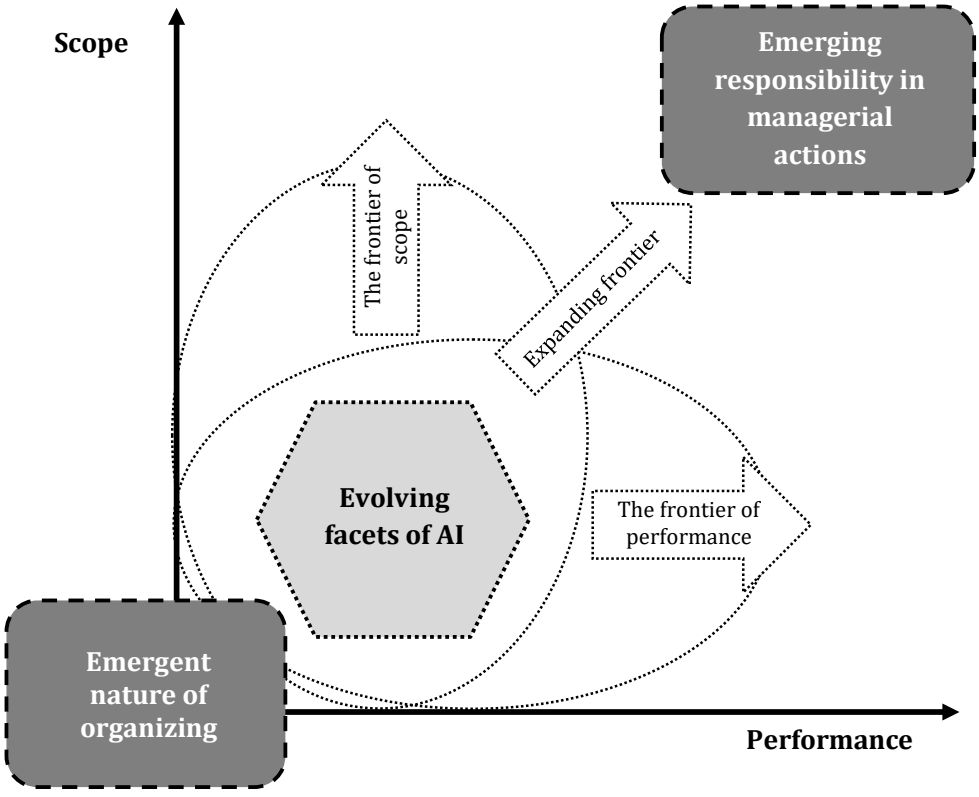


Figure 4. Interpreted relevant assumptions of the AI frontier perspective

Secondly, AI is a frontier for decision-making that constantly improves its performance and scope, thereby exposing underlying facets (Berente et al., 2021). These facets represent an ongoing effort to push the boundaries of scope and performance of AIs. They underscore a foundational property or capability of contemporary AIs that describes its functions. Facets are core characteristics that define a system's behavioral, technical, and experiential qualities, and also dictate the types of actions AIs enable. Managers need to manage these interrelated facets of AI simultaneously across development and deployment, in both production- and consumption-facing activities. However, this dissertation considers that facets of AI help decision-making for both managers and users. For instance, a particular facet like autonomy can be deployed in consumption that works as a distributed decision-making system to delegate information processing capability to represent firms.

Facets of AI are primarily characterized by data-driven learning and computational decision-making. Through computing, these facets enable automation and augmentation of a wide range of activities at the firm level. Facets of AI certainly offer economic benefits for firms, including scale, superior features, and a competitive advantage. However, they also involve dynamics of interactions and relationships between AIs and human agency and behavior (Murray et al., 2021; Zheng & Jarvenpaa, 2021). This may relate to difficulties of human rationality due to an absence of complete information, the presence of anthropomorphic cues (humaneness), or emerging uncertainty stemming from cross-border activities. In turn, facets of AI may promote codified discrimination, ego-centric biases, and an instrumental-rational approach to problem-solving rather than value-based reasoning. Therefore, constant interaction across facets raises critical ethical issues that must be managed responsibly, such as unintended side effects and violations of privacy. Table 2 presents different characteristics of AIs in firm-level settings that underlie facets.

In summary, AI is an emerging phenomenon with different contemporary forms within firms, particularly robotic AIs, virtual AIs, embedded AIs, and algorithmic AIs. Robotic AIs enable physical task execution, whereas virtual AIs facilitate knowledge-intensive interactions and service delivery across geographic distances. Embedded AIs support localized and real-time decision-making within products and services, while algorithmic AIs underpin optimization through performing analytical tasks. These forms can also be integrated into multi-AI systems that combine complementary computational capabilities to achieve broader goals. These contemporary AIs operate and respond to increasingly diverse institutional stimuli across national borders. Theorizing AIs based on socio-technical systems emphasizes the emergence and evolving facets of AI in the context of common good. It points to constant managerial actions across various facets of contemporary AIs, such as

autonomy and learning, as their performance and scope grow. It implicitly advocates managerial intentionality to organize AI within firms, and highlights social responsibility to organize AIs responsibly, especially considering increasing ethical concerns such as privacy violations and unintended negative effects. Therefore, this dissertation focuses on organizing facets of contemporary AIs for responsiveness across borders, where the actions of humans play central roles in transmitting capabilities within firms.

Table 2. Different characteristics of AI systems in firm-level settings

Reference	Evolving aspects	Characteristics of AI systems
(Davenport & Ronaki, 2018)	Tasks and solving business problems	*Process automation:* Automating digital and physical tasks. *Cognitive insight:* Detecting patterns in vast volumes of data and interpreting their meaning. *Cognitive engagement:* Engaging with users for support activities and customization.
(Berente et al., 2021)	Emerging computational advancement	*Autonomy:* Increasing capacity to act on their own, without human intervention. *Learning:* Inductively improve automatically through data and experience. *Inscrutability:* Increasingly opaque algorithmic models and outputs to human understanding.
(Hutzschenreuter & Lämmermann, 2025)	Functional focus and deployment context	*Potential task superiority:* Acquiring powerful decision rules. *Black box perception:* Lack of human understandability of complex algorithmic behavior. *Dynamic nature:* Constantly changing nature of self-learning systems.
(Ramaul et al., 2025)	Rationality and anthropomorphism	*Computational superiority:* Features and outcomes concerning performance. *Decisional objectivity:* Complementarity to humans' bounded rationality. *Anthropomorphic agency:* Performing humanlike functions. *Organizational actorhood:* Distributed decision-making systems within organizations.
(Stelmaszak et al., 2025)	Properties for organizing capability	*Connectivity:* Enacting systematic relationships and coupling with human actors. *Codependence:* Enacting relations among humans and algorithms to bring their mixed, varied, and specialized contributions. *Emergence:* Change over time in practice as relations are enacted in a system
(Sapkota et al., 2026)	Design, goal orientation, and adaptation	*Autonomy:* Given ability to function with minimal or no human intervention. *Task specificity:* Purposefully built for perform tasks. *Reactivity:* Interacting with dynamic inputs to respond to real-time stimuli, leading to responsiveness to changes.

2.3 Strategic Organizing of AI Across Borders

2.3.1 Responsiveness of AI as a Source of Firm-specific Advantage

Actions within firms around AIs may turn into managerial capabilities that underpin a source of firm-specific advantage. Firm-specific advantages (FSAs) refer to knowledge bundles of firms that provide a competitive advantage (Adarkwah & Malonæs, 2022; Rugman & Verbeke, 2003). Recent IB studies have acknowledged the central role of information technology artifacts in configuring FSAs (Banalieva & Dhanaraj, 2019; Lee et al., 2021). AI-driven FSAs can be non-location-bound and transferable across borders, and underpin possible location agonistics of advanced technological artifacts (Ojala et al., 2023; Tatarinov et al., 2023). They imply traditional asset- and transaction-based, as well as contemporary recombinant-based and network-based advantages (see Banalieva & Dhanaraj, 2019; Lee et al., 2021 for more detailed accounts). AIs as a source of FSAs thus relate to (in)tangible assets, learning capabilities, and tools for relationship development. AIs can serve as core capabilities for firms, such as a foundational model or unified model architecture that can be used for multiple types of products and services. However, AIs can also be embedded to combine with other capabilities, such as self-driving abilities in cars. Furthermore, AIs can be standalone products or services of a firm, such as a humanoid robot, and as relationship-building tools. However, recent computational capabilities imbued in FSAs have a strong influence on how users buy and consume services, including products (Meyer et al., 2023). Consequently, there are increasing concerns about protecting users' well-being across borders through formal regulatory institutions. Therefore, it becomes imperative to integrate societal welfare into a theorization of AIs as a source of FSAs (Bartneck et al., 2021; Cowls et al., 2021; Fuso Nerini, 2026), rather than focusing solely on rent generation and profit maximization (Banalieva & Dhanaraj, 2019; Lee et al., 2021). One way to respond to this predicament is to organize AIs for FSAs that may ensure global integrity, while at the same time ensuring responsiveness across borders.

Responsiveness across borders is generally understood as a deliberate organizing ability to respond to local sensitivities and changes across geographically dispersed institutions. For FSAs, actions within firms necessitate arrangements that are institutionally appropriate and adaptable across heterogeneous institutional (e.g., regulatory, cultural) environments. Existing discourse on strategic organizing in both IB and IS argues for focusing on technology as a source of advantages (Banalieva & Dhanaraj, 2019; Faraj & Leonardi, 2022; Yoo et al., 2010). It also implies using both technology development and deployment to configure FSAs (Tatarinov et al., 2023). Therefore, this dissertation proposes the responsiveness of AIs as a source of FSAs.

The responsiveness of AIs in the context of the common good can be understood as processual activities that are jointly shaped by two analytically distinct episodes that include development and deployment. Collectively, they form a feedback loop of managers' development and deployment, users' adoption and usage, emerging consequences, and managerial responses with new decisions. Therefore, responsiveness may necessitate a consideration of how AIs are developed and deployed, which underpins the nature of FSAs.

Development refers to the process of designing, building, training, and configuring features, including algorithms, model architecture, data infrastructure, testing, and decision logic. During development, responsiveness is largely ex ante and architectural, as managers embed their assumptions about international dimensions into problem formulations, data choices, and model architectures. These design decisions determine the latent potential of AIs to accommodate cross-national variation by enabling—or constraining—local adaptation, regulatory compliance, and culturally sensitive interpretation (Rahwan et al., 2019; Tatarinov et al., 2023). In contrast, responsiveness during deployment is in situ and enacted, emerging through the configuration, consumption, and governance of AIs within specific national or institutional settings. Deployment refers to the introduction, implementation, and integration of developed artifacts into real-world contexts, including international markets, organizations, and everyday life. At this stage, responsiveness depends on actions within firms, which determine how, where, and under what conditions AIs are deployed. However, development happens primarily before they are made available in markets for consumption.

Regarding actions within firms, the development of contemporary AIs is strategically intended and planned. Such actions during development include setting objectives and priorities (e.g., efficiency, profit, social impact), allocating resources (e.g., capital, talent), and selecting architects and data sources that shape systems' capabilities and limitations. Actions also include defining ethical, legal, and governance frameworks (e.g., bias mitigation, transparency, compliance) and design choices surrounding contemporary forms (e.g., robotic AIs, algorithmic AIs). However, user action also plays a limited but important role during development. Potentially, development influences responsiveness that may occur through anticipatory consumption, where management imagines future use and learns through a feedback loop such as prior versions, testing, or pilot studies. Influences may also occur through data generation, as when user behavior is used as training data. In some cases, systems may be co-developed or customized using participatory or platform-based initiatives. However, development implies a cyclical socio-technical process that also includes deployment.

Deployment also depends on how outputs of AIs are aligned and adjusted in response to institutional differences, user expectations, and unforeseen consequences. Decisions that lead to action in deployment may include market selection and scaling (local vs. global), integration into the product or service, pricing, and customer access (open vs. closed vs. localized). They may also include risk management, accountability, regulatory compliance, and market positioning. However, user action becomes central and decisive at deployment. User actions during deployment include adoption, resistance, or rejection of AIs, as well as patterns of use, including intended use, unintended use, and accidental misuse. Additionally, these actions relate to interaction and relationships with contemporary AIs (e.g., virtual AIs), leading to social, ethical, and behavioral consequences. Notably, while development shapes organizing boundaries of how AIs can respond across countries, deployment determines how these systems are realized and contested in consumption. Conceptualizing responsiveness that spans both development and deployment thus highlights the socio-technical nature of AIs as a source of FSAs, and underscores processual dynamics to organize AIs for responsiveness across borders that embody upstream design choices as well as downstream adoption processes.

In the context of the common good, responsiveness is thus a multifaceted construct that may not be understood primarily as country-level adaptation. Instead, it is embedded in system development, governance choices, deployment, and the ongoing evolution of facets of AI. However, this view zooms in on foundational aspects of organizing a system with a firm-level setting (Faraj & Leonardi, 2022; Habib et al., 2020), and not firms as an organizing system (Orton & Weick, 1990). Accordingly, it does not necessarily involve the hierarchical control of firms, or the dual pressure of global integration-local responsiveness, in the traditional sense (Bartlett & Ghoshal, 1989). In contrast, responsiveness articulates how to organize AIs by design (through a unified model architecture) and consistently manages their properties (e.g., location-specific learning) across countries during consumption. The responsiveness of AIs relates to developing calibrated trustworthiness through baking trust in service systems among users (Bartneck et al., 2021). It is also related to user satisfaction and well-being which does not cause a risk of harm, thereby enhancing regulatory response (Wood, 2021). In turn, systems become responsive across institutions, thereby enhancing competitiveness in cross-border activities. Hence, responsiveness becomes an ongoing managerial capability to organize system properties to achieve the common good.

This framing presents management actions to organize AIs and their facets in a way that they are adaptive to emerging location-specific dispersed stimuli (Ali et al., 2025; Ciulli & Kolk, 2023). Organizing for responsiveness focuses on how individual managers organize different facets of AIs to achieve the common good in the Fifth

Industrial Revolution. Hence, it is an evolving process that may take a responsible strategic view with a global-by-default approach to address both global and localized factors across national borders. It focuses on different aspects of relationships between technological parts (or properties) within a system. Hence, the responsiveness of AIs reflects the socio-technical nature of organizing and underpins firms' capabilities (Lyytinen et al., 2021; Stelmaszak et al., 2025). Responsiveness of AIs across borders that imbues into FSAs addresses uncertainty, ambiguity, or complexity that arises from the liability of foreignness and diverse institutions in international markets (Meyer et al., 2023; Zaheer, 1995). Such indeterminacy issues in internationalizing activities arise in numerous ways within the given operating environment. They may rise due to the inclusion of the different entities—such as institutions, actors, and technological factors—and the new interactions among these entities (Aguinis & Gabriel, 2022; Eden & Nielsen, 2020). Therefore, managerial actions involve multiple levels of indeterminacy as firms are embedded in complex economic, social, technological, and regulatory dimensions (Bartlett & Ghoshal, 1989; Nambisan & Luo, 2021). However, managerial actions are key within firms to shape AIs, which in turn have a significant impact on users during consumption. Accordingly, a focus on human agents in strategic organizing of FSAs is paramount.

2.3.2 Foundational Focus on Human Actions to Organize AI

Strategic organizing typically involves developing and selecting a specific course of action, followed by implementing it through communication, interpretation, and execution of the plans (Raes et al., 2011). For AIs, strategic organizing involves actions that relate to leading, monitoring, delegating, communicating, cultivating, coordinating, and reflecting within firms (Asatiani et al., 2021; Berente et al., 2021; Lyytinen et al., 2021). This dissertation views firms as corporate entities where actors such as top and middle management shape strategies through their actions, including FSAs in internationalization (Aguinis et al., 2022; Tatarinov et al., 2023). Within firms, it is assumed that the interface between actors of top and middle management is key to effective strategic organizing (Raes et al., 2011). Top management actors are individuals who serve on the executive board or hold at least the highest level of strategic decision-making (e.g., chairman, vice chairman, chief executive officer, chief technology officer) (Belderbos et al., 2022). Collectively, they formulate, communicate, and carry out the strategic and tactical initiatives of firms in international business activities. Middle managers are individuals who work directly beneath top management and above first-level supervisors who support initiatives from the operating level, combine resources, accelerate strategy implementation, and conceptualize new strategic organizing (Floyd & Wooldridge, 1992). At the intersection of these actors, trustworthiness in their role behavior through

information exchange and mutual influences is considered central for the failure or success of firms' performance (Raes et al., 2011). Therefore, collective efforts of both top and middle management actors provide better economic reasoning for firms in international markets (Boyett & Currie, 2004).

Managers are responsible for taking actions based on their intentions and decisions, which in turn shape firms' strategies across national borders and ultimately FSAs. They furthermore directly and indirectly influence patterns of consumption by setting incentives, constraints, and technical properties into AIs (Kopalle et al., 2022; Verbeke, 2026). They are thus key to organizing firm-level efforts to address the liability of foreignness in international markets, which emerge from uncertainty or a limited understanding of foreign markets (Zaheer, 1995). They also need to navigate the complexity that arises from diverse institutions in international markets (Aguinis & Gabriel, 2022; Dau et al., 2022; Eden & Nielsen, 2020). Accordingly, their strategic actions enable operations across multiple countries while integrating global efficiency with responsiveness across borders (Bartlett & Ghoshal, 1989; Nambisan & Luo, 2021).

For AIs to be responsive as a basis for FSAs, actions of managers need to embody critical reflection and be careful to avoid inflicting harm and negative consequences (Ali et al., 2025; Hasan, Weaven, et al., 2021; Lyytinen et al., 2021). Managers are responsible for all major decisions regarding the development and deployment of AIs at home and abroad (Berente et al., 2021; Ratten et al., 2024). They coordinate for AI-based capabilities that are less location-bound and transferable across borders, and supervise capabilities to be embedded into products or services, enabling consumption convergence in digital globalized markets. Managers also use AIs for decision-making and customer targeting. They monitor, delegate, cultivate, and reflect on tasks (Lyytinen et al., 2021), and decide and allocate resources, oversee projects, and govern AIs. They also adjust decisions, processes, and routines appropriate for development and deployment (Asatiani et al., 2021). Collectively, this range of managing activities underpins managerial actions around AIs within firms, which in turn shape ongoing production and consumption and future trajectories.

Managerial actions underpin organizational, strategic, and governance decisions that shape how contemporary AIs are developed and deployed in global and local markets. These actions occur at the firm level, intersect with institutions (e.g., norms, culture, regulations), and determine where, how, and for what purpose AIs are introduced. They influence market structure, ethical standards, access, and long-term social implications. Put simply, managerial actions govern AIs from the top down, deciding what type of AIs representation exists, how it is designed, how it is deployed,

and for what purposes it is used. Therefore, managerial actions influence the scope of affordances triggered by user actions.

Affordances can be generally understood as behavioral possibilities or opportunities (Gibson, 1979). However, the actualization of affordances is multifaceted along with given technological features and psychological needs (Anderson & Robey, 2017; Karahanna et al., 2018). Affordances facilitate action through actors' (e.g., users) perceived and behavioral possibilities or opportunities, which can be either intended or unintended (Wood, 2021; Zheng & Jarvenpaa, 2021). Therefore, extant IB studies consider affordances as one of the central means of AI for user actions in international business activities (Ciulli & Kolk, 2023; Tatarinov et al., 2023). User actions refer to how individuals or consumers interact with, adopt, and intend to use, or actually use products and services in everyday contexts, both physical and digital. User actions enable systems' consumption and use worldwide, and occur through interaction with contemporary AIs (e.g., robotic AIs like humanoid robots, autonomous vehicles and virtual AIs like chatbots and avatar). These actions may include acceptance, resistance, misuse, adaptation, and non-anticipatory uses. User actions produce both positive, but also negative (dark) consequences, such as efficacy gains, dependence, bias reinforcement, the spread of misinformation, and death. Put simply, users' actions operationalize AIs from the bottom up, turning deployed systems into live experiences with real social, ethical, and behavioral outcomes. While user action shapes the possibilities of AIs in the global marketplace, managerial action determines FSAs to achieve the common good.

For managerial actions, AIs at the intersection of affordances introduce a variety of challenges in internationalizing activities that are socio-technical in nature. Many challenges are instrumental, including solving issues related to safety, black-box problems, and implementing measures to prevent negative outcomes (Rahwan et al., 2019; Rai, 2020). However, many challenges are humanistic and linked to moral and ethical issues, particularly around the usage and consumption of products and services. These include concerns about deception, privacy, algorithmic fairness, social justice, and the risk of harm (Danaher, 2020; Dolata et al., 2022; van den Broek et al., 2025). Recent reports in popular media suggest that firms using AIs may be involved in predatory and exploitative practices that result in significant harm. For example, some incidents involve sexual intimacy, violence, self-harm, death, and murder, including the exploitation of vulnerable users (Burga, 2025; Horwitz, 2025; Jargon & Kessler, 2025; Marsden, 2025). Therefore, there is an increasing call for accountability where firms are responsible for the possible dire consequences of AIs in many instances (Buolamwini & Gebru, 2018; Tóth et al., 2022; Wood et al., 2025). Accountability also lies with managers. But nevertheless, existing theoretical lenses of FSAs only tend to consider rent generation and profit maximization (Adarkwah &

Malonæs, 2022; Banalieva & Dhanaraj, 2019; Lee et al., 2021), which is problematic for the common good.

To summarize, the responsiveness of AIs across borders poses a processual mechanism for ongoing managerial capabilities that underpin FSAs. While managerial actions govern development and deployment of AIs in configuring FSAs, user action becomes decisive during deployment, which also affects the possible consequences, either positive or negative. Collectively, they form a socio-technical feedback loop comprising intention, development, deployment, consumption, consequences, and managerial responses. For responsiveness, the development of AIs precedes their initial deployment. However, deployment enables user actions that generate feedback and consequences, which in turn, shape subsequent cycles of development. Therefore, these two episodes impact responsiveness. However, there is an established discourse on strategic organizing in international business activities which provides a significant understanding of orchestrating FSAs and responsiveness across borders.

2.4 Existing Theoretical Discourse in Cross-border Strategic Organizing

This sub-chapter now engages with two key theoretical lenses in IB, namely the global integration-local responsiveness theory and the loose coupling theory. They collectively focus on internal and external governance for organizing, which seems most relevant for coordinated actions and responsiveness. The dissertation integrates relevant assumptions into managerial actions, while politely challenging their underlying reasoning in the context of the common good. While both theories have great relevance to strategic organizing, their focus lies on firms, not managers. Additionally, their reasoning heavily relies on economic reasoning with no regard for social well-being, which may not necessarily align with the Fifth Industrial Revolution and the global sustainable development agenda (European Commission, 2022; United Nations, 2023).

2.4.1 Global Integration-Local Responsiveness Perspective

Global integration-local responsiveness (I-R) theory posits that strategic organizing for worldwide effectiveness depends on the dual pressure of global integration and local responsiveness (Bartlett & Ghoshal, 1989; Birkinshaw et al., 1995; Prahalad & Doz, 1987). Beyond international strategy literature in IB, this theory has proliferated significant research, including information management in IS (Tractinsky & Jarvenpaa, 1995). While the theory focuses on the structural organizing of large

multinational firms with subsidiaries, it can also relate to the foundational level of managerial actions. The reasoning of this theory is rooted in two relevant underlying assumptions that provide actionable guidance (see Figure 5).

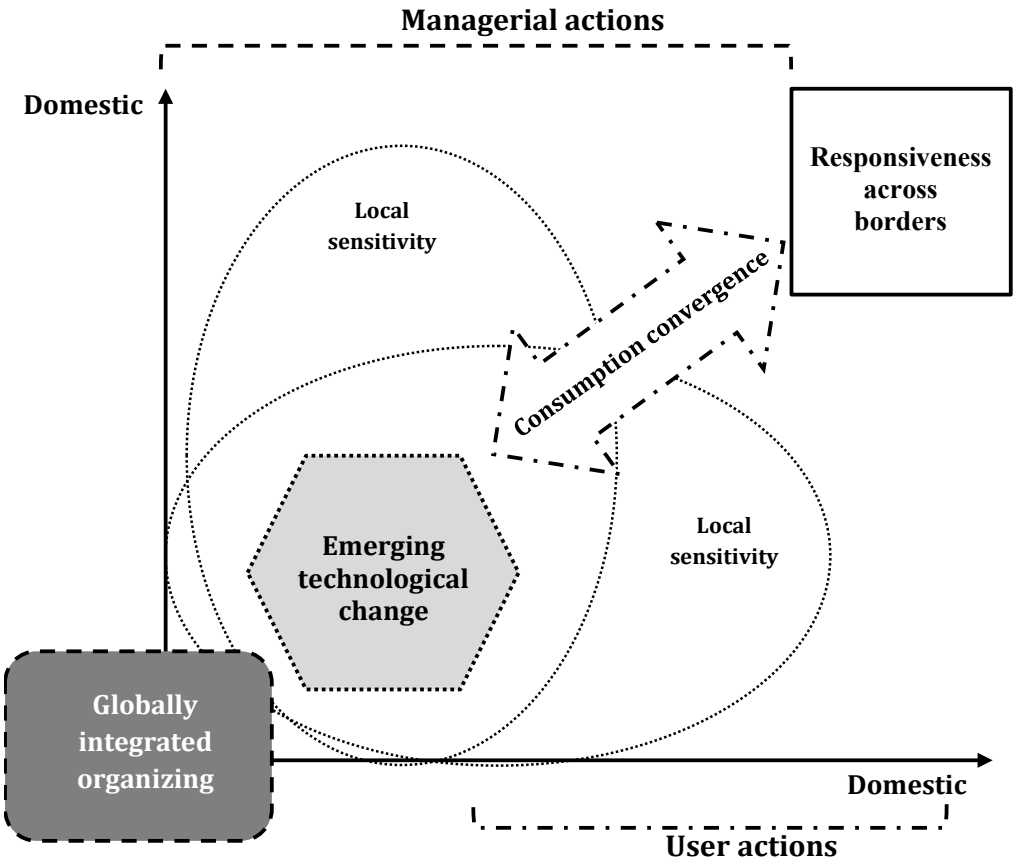


Figure 5. Interpreted relevant assumptions of the I-R perspective

The *first* assumption concerns intended managerial actions for strategic organizing across borders within a dual dichotomy (Prahalad & Doz, 1987). An organizing of global integration implies that resource-committing decisions are centralized in line with the overarching corporate vision. It involves central structural configurations to coordinate international business activities across countries to achieve efficiency and leverage similarities across locations (Bartlett & Ghoshal, 1989). However, strategic organizing of local responsiveness implies a decentralized decision structure that enables local units to accommodate domestic market conditions through responsive actions (Birkinshaw et al., 1995). It involves responding to specific needs across a variety of domestic markets with different institutions, including regulations, ethnicity, religions, and national culture. The I-R theory institutes distributed decision-making capabilities in international business activities. For instance, in digitalization as a strategic orientation for multinational firms with physically bounded manufacturing, the balancing of I-R seems necessary (Du et al., 2025).

Managers can navigate this dual pressure by marketing technological innovation in customer-facing activities and by external collaborative governance under industry pressure and data policy. Such dual pressure enables managers of internationalizing firms to navigate a hierarchical organizational structure into global, international, multi-domestic (multinational), and transnational strategies (Bartlett & Ghoshal, 1989). Therein, managers may choose to emphasize one over the other, combine them, or adopt a multifocal strategy (Andersen & Hallin, 2017).

The *second* assumption holds that emerging technologies have a significant impact on consumption across markets that dictate the dual pressure of I-R. Consumption includes symbolic, contextual, and experiential elements of use, shaped by social interactions embedded into institutions (i.e., norms, culture) (Arnould & Thompson, 2005). A recent IB study acknowledges that the emergent nature of digital technological change significantly shapes consumption worldwide (Meyer et al., 2023). While significant institutional distances (such as ethnic diversity) persist in the global marketplace, technological advancement may enable consumption convergence across borders (Dau et al., 2020, 2022; Ozturk et al., 2021). Convergence in consumption refers to the extent to which consumers self-associate with global institutions (i.e., beliefs and attitudes) rather than with local ones. Recent technological advancements powered by contemporary AIs (i.e., robotic, virtual, algorithmic) may follow similar trajectories (Baldwin, 2016, 2019). This underscores the importance of users' institutions in deploying AIs. However, the I-R theory points to a need to institute loosely coupled organizational systems (Bartlett & Ghoshal, 1989). Therefore, this sub-chapter now turns to a loose coupling theory with a focus on digital globalization.

2.4.2 Loose Coupling Systems Perspective

Loose coupling theory (LCT) focuses on a systems perspective, which intersects with institutional theory in organization settings (Orton & Weick, 1990). Recently, IB has engaged with LCT in digital globalization (Nambisan & Luo, 2021). With a relevant focus on technological characteristics, LCT emphasizes both linkages and separateness to argue for responsiveness across borders. It suggests that the coexistence of coupling (linkages) maintains responsiveness, and looseness (separateness) retains distinctiveness. Therefore, LCT can provide valuable insights to address governance-related challenges. Similar to the I-R theory, it focuses on large multinational firms. Nevertheless, LCT can also translate to foundational understandings of managerial actions, as it relates to organizing socio-technical artifacts in a digitally fragmented world. Theses of LCT are rooted in three relevant assumptions.

The *first* assumption concerns actors' intentionality on actions to shape the nature of a system for responsiveness, underpinned by adaptation (Nambisan & Luo, 2021; Orton & Weick, 1990). Coupling reflects a need for different parts of a system to be linked, while looseness reflects a need for them to be separate and flexible to adapt to spontaneous external changes. In turn, the whole system achieves responsiveness as parts of it can be modified for necessary adaptation. Hence, loose coupling implies addressing two opposing forces and allows them to coexist in the same strategic design. In international business activities, it is an organizing way to bring the technical core of systems close to eliminating uncertainty, where parts of said systems are open to embracing differences from location-specific diversity and institutions (Nambisan & Luo, 2021). As a result, a system is arranged in a way that is open and closed simultaneously, and this arrangement is also spontaneous and deliberate. This system perspective suggests that different parts of a system, such as modules, are linked to one another while retaining a degree of independence and indeterminacy. Accordingly, LCT emphasizes the role of actors' intentions on actions in management practices, underscoring how emerging cognitive institutions shape loosely coupled organizing (Orton & Weick, 1990). Loose coupling can anchor in several competing ideas, including planning and implementation, official structures and negotiated orders, and structural facades and technical cores. Such an assertion implicitly indicates the socio-technical nature of a system, as argued in IS, which has both humanistic and instrumental elements (Sarker et al., 2019).

Reflecting on digitalization, the *second* assumption concerns affordances as a key characteristic of digital artifacts that configure FSAs as loosely coupled (Nambisan & Luo, 2021). Affordances facilitate action through actors' (e.g., users) perceived and behavioral possibilities or opportunities, which can be either intended or unintended, shaping FSAs. Such characteristics may shape the simultaneous exercise of looseness and coupling for FSAs. Affordances enable the recombination of FSAs in development and deployment through openness, closeness, and separateness. In turn, FSAs facilitate asset orchestration and competitive positioning in operating markets. Recent IB studies underscore perspectives on affordances that have a focus on AI (Ciulli & Kolk, 2023; Tatarinov et al., 2023).

The *third* assumption concerns instituting a shared value on a global scale as a central mechanism of organizing loose coupling for responsiveness (Nambisan & Luo, 2021; Orton & Weick, 1990). It argues for a shared worldview or a standard set of principles in governance, implying the need for cognitive institutions to compensate for the system's looseness. Such a mechanism is needed to accommodate looseness in some dimensions, while coupling them with other dimensions. Having a shared value is critical because it enacts three determinants of loose coupling: causal indeterminacy, external fragmentation, and internal fragmentation. Causal indeterminacy may result

from unclear means-ends links caused by bounded rationality, biased perception, uncertainty, ambiguity, and intangibility of complex processes and knowledge (Nambisan & Luo, 2021; Orton & Weick, 1990; Simon, 2000). In other words, cognitive limitations inherent to human information processing stem from uncertainty, ambiguity, and the absence of tangible knowledge. For instance, an IB study by Lindner et al. (2025) suggests that algorithmic AIs affect existing biases to shape managerial actions through distorted decision-making outcomes. Such effects arise due to unfamiliarity arising from the liability of foreignness and institutions of foreign countries. However, another source that arises from external fragmentation is driven by location-specific factors and dispersed stimuli (Nambisan & Luo, 2021). Dispersed stimuli are related to geographical dispersion, culturally sensitive cues, and contradictory external pressures in foreign markets, and can be a source of incompatible expectations arising from institutional and technical pressures (Orton & Weick, 1990). This can be seen, for example, in the competitiveness and race for AIs in the global marketplace (Lundvall & Rikap, 2022), and growing market-based regulatory institutions such as the AI Act (European Commission, 2024). Hence, external fragmentation may arise from geopolitics, techno-nationalism, the absence of shared values among partners, and regulatory institutions. Additionally, internal fragmentation is a third recurring issue that may arise from misalignment between actions and arrangements within managerial units. For example, an IB study suggests that misalignment due to the absence of clear principles of responsible use obstructs scaling of virtual AIs across borders, even in closely proximate countries (Tatarinov et al., 2023). Collectively, these issues demand having common cognitive institutions as a shared value for organizing things cohesively.

To summarize, all of the relevant assumptions presented above make one key point: there is a need for a globally integrated share value to manage technological characteristics embedded in FSAs that address local sensitivities across diverse institutions to enable responsiveness (see Figure 6). These assumptions are relevant, but pose a fundamental problem for the common good. Therefore, this sub-chapter calls for a rethinking of existing views of strategic organizing for responsiveness.

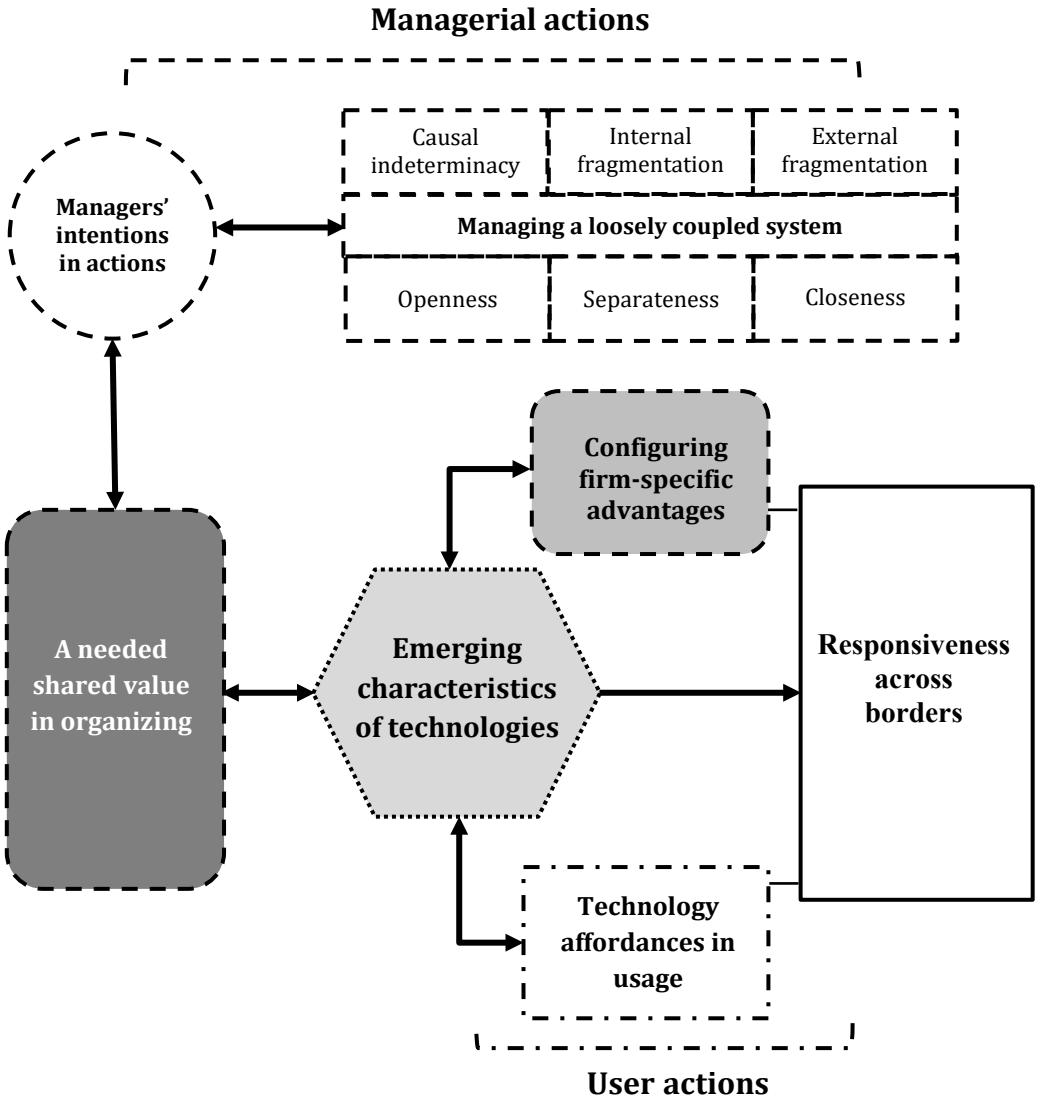


Figure 6. Interpreted relevant assumptions of the loose coupling perspective

2.4.3 Rethinking Strategic Organizing for AI

While forehead assumptions in strategic management discourse in international strategy are important, a substantial distinction persists in achieving the common good. While both I-R and LCT theories are useful, they suffer from their underlying reasoning. The former I-R theory focuses on a coordinated approach that emphasizes technological change and a convergence of consumption in international markets. The two assumptions of the I-R theory are rooted in reasoning that global production and consumption are organized around tangible, physically bounded processes. Examples include assembling cars, shipping textiles, or manufacturing electronics.

Additionally, it emphasizes location-bound subsidiaries and coordination among different business units across diverse geographic locations. However, it has a core focus on economic reasoning for responsiveness, and has little regard for social imperatives and emerging technical properties. Recent technological advancements enable managers to organize less through sequential adaptation, and more through continuous interaction with regulatory regimes and users on the global stage (Birkinshaw, 2022, 2024). The latter LCT theory focuses on a systems perspective to emphasize technological properties. While being relevant in digital globalization to achieving responsiveness across foreign markets, its three assumptions show no regard for societal welfare in configuring FSAs. Similar to the formal I-R theory, it focuses only on economic reasoning. Yet it also does not underline the multifaceted actualization of affordances that may pose negative consequences (Anderson & Robey, 2017). Notably, the social imperatives of emerging affordances of AIs remain largely unanswered (Ciulli & Kolk, 2023; Tatarinov et al., 2023), and do not consider ethical challenges and responsibilities in instituting shared value. With AIs, managers may cause harm (intentionally or non-intentionally) through enabling technological properties and affordances (Wood, 2021).

Existing assumptions are relevant, but may not explain responsiveness in the context of the common good. AIs reshape managerial actions on scaling and coordination. AI differs qualitatively from earlier technological changes, and even from prior waves of digitalization (Berente et al., 2021; Brynjolfsson et al., 2019; Lyytinen et al., 2021). Therefore, rethinking is necessary for several reasons. First, AI is emergent and inherently socio-technical, as managers utilize its contemporary form across borders to facilitate affordance and FSAs, with its performance and scope expanding. Second, managerial intent and actions shape the nature of a system, whether open, closed, or with a particular technological property or facets. Third, the facets or technological properties of AIs, such as autonomy, influence human rationality, as humans are inherently limited in their cognitive capabilities. Decision-making complexity arises from the dynamic nature of the international marketplace and the AI frontier. Fourth, developing and deploying AIs in international business activities raises emerging ethical considerations and responsibilities across facets or technological properties. Managers need to address the impact of AIs utilized in international business activities on the societies in which they operate, which underscores the global challenge of responsible production and consumption. Finally, there is an urgent need for a globally integrated, shared value to responsibly organize AIs to maintain responsiveness, and this underpins responsible technology production and consumption for the common good that integrates reasonings of both economic and social goals.

Taken together, all five assumptions of the two theories emphasize only economic reasoning, including production efficiency, profit maximization, and economies of scale. The extant assumptions offer little insight into social reasoning. However, focusing solely on economic reasoning is problematic in ensuring the common good, particularly in the era of the Fifth Industrial Revolution. The strategic organizations are very different from those of the physically bound industrial era (Leavitt et al., 2021; Zuboff, 2015). Managerial actions on AIs point to emerging inequality and monopolistic competitive advantages (Ahi et al., 2022; Buckley & Enderwick, 2025; Ghauri et al., 2021), as well as to the distributed nature of decision-making embodied in the temporal form of AIs that increasingly interact with users in international markets (Kopalle et al., 2022). However, there remains uncertainty about the responsible use of AIs that hinders global sustainable development, including dishonest anthropomorphic properties and mass surveillance (Leong & Selinger, 2019; Zuboff, 2015). Managers from internationalizing firms are responsible actors to develop and deploy AIs as capabilities to extract, predict, and influence users' behavior, underpinning FSAs. They focus on economies of scale for profit maximization, often at the expense of individual autonomy and in violation of fundamental human rights (Bartneck et al., 2021). AIs, as a contemporary form, can serve as a distributed decision-making system that interacts with users on behalf of internationalizing firms in a humanlike manner (Norwegian Consumer Council, 2023). However, AIs may also pose dire risks of harm (Wood et al., 2025). Therefore, the dissertation now engages with a trusteeship framework to complement existing assumptions.

2.5 Imperatives for Trusteeship Organizing of AI Across Borders

2.5.1 Trusteeship Framework

A trusteeship framework has evolved based on ancient knowledge and contemporary contributions in firm-level settings (Ali et al., 2025; Rice, 1999). Managers, as human agents, are viewed as trustees of capabilities that need to be developed and deployed to protect everyone's well-being. In strategic organizing discourse, trusteeship holds that managers have a moral responsibility to their key stakeholders, including users, business partners, and society (Beekun & Badawi, 2005; Rice, 1999). However, trusteeship is deeply rooted in ancient knowledge. For example, trusteeship relates to *Amanah,* which derives from the Arabic root *'a-m-n*, which conveys trust, faithfulness, and security. In moral reasoning derived from ancient knowledge, trusteeship institutes a proactive governance concept that calls human agents to act

ethically and justly as part of a sacred trust placed upon them (Qur'an: Surah Al-Ahzab 33:72). They are custodians of trust and enact trustworthiness in business relationships through coordinated covenants. In the context of the common good, covenants underline self-commitment and governing acts to engage in responsibility and refrain from harmful actions.

For contemporary AIs, the trusteeship framework demands three foundational covenants, namely ontological, epistemological, and existential (Ali et al., 2025). Collectively, these covenants provide a foundation of responsible organizing of AI that integrates intentionality, intellectual integrity, and practical stewardship. Each covenant reflects a distinct aspect of managerial action to ensure human responsibilities. These three covenants are well-suited to manage the socio-technical nature of AIs in international business activities: (a) accountability (ontological), (b) data and knowledge bundles (epistemological), and (c) real-world impact (existential). The first one is the ontological covenant. It is grounded in humans' ultimate moral reasoning and innate nature for pursuing truth, which creates a sense of duty tied to a higher purpose. Such a sense works as a basis for accountability, which indicates innate self-consciousness to ensure just governance and contribute positively to societal well-being. The second one is the epistemological covenant. It emphasizes honesty in knowledge acquisition and application, which indicates proactive measures to block possible harm. It is actioned to avoid distortions for personal or collective gain through technological means, prioritizing truthfulness and accessibility. The third one is the existential covenant. It underlies universal stewardship toward the external environment, which indicates institutions in dispersed geographical locations around the world. It is aimed to ensure justice and compassion within human societies, and address inequalities. It also ensures that technological advancements benefit all members of society equitably. Collectively, these covenants entail accountability, deliberate action, and responsible stewardship, which are fundamental for trusteeship organizing of AIs.

Trusteeship organizing proposes that managers act as trustees to serve the common good beyond economic performance, and that they prioritize moral responsibility over law, and stewardship over ownership in their actions. Hence, organizing AIs under "global trusteeship by default" seems to be a suitable approach. Trusteeship organizing promotes responsibility in an individual's actions while protecting social interests, particularly by avoiding harm to the community through internationalization activities (Ali et al., 2013). Hence, contemporary thinking on AIs and responsibility naturally follows its underlying expectations (Bartneck et al., 2021; European Commission, 2019; Jobin et al., 2019). Accordingly, managers act as trustees dedicated to promoting social equity and justice, ensuring user protection for collective and social good. Their actions align with core principles like justice,

compassion, and balance, demonstrating genuine intent to maintain moral integrity. Nevertheless, trusteeship organizing emphasizes proactivity and a responsibility for the common good, including socio-economic matters. These insights can be translated into guiding managerial actions that focus on accountability, blocking harmful means, and addressing institutions by default. They bring together a triad focus on the development and deployment of AIs, and may inform processual mechanisms to realize responsiveness across borders. Therefore, trusteeship organizing refers to *the act of responsible development choices across institutional settings to block harmful means and ensure accountability through necessary responsive actions during deployment and usage.*

Trusteeship organizing is somewhat distinct from the established theorization of trust, while mirroring its underlying outlook (Wijaya et al., 2023). The existing discourse on trust builds on the assumption that actors are willing to be vulnerable to other actions (Mayer et al., 1995). It relies on the expectation that others behave reliably and beneficially, even without control or monitoring. This concept helped to explain trust in several areas, supporting research on social trust (Gaur et al., 2022), customer trust (Schumann et al., 2010), relational trust-building (Johanson & Vahlne, 2009; Katsikeas et al., 2009), and human trust in contemporary AIs (Benbasat & Wang, 2005; Blut et al., 2021; Glikson & Woolley, 2020). These contributions are valuable. However, this established discourse focuses on trust as perceived or sensed by human observers, and emphasizes beliefs and expectations among human actors and entities (Glikson & Woolley, 2020; Mayer et al., 1995). In contrast, the trusteeship organizing framework offers a distinct but complementary perspective.

In the context of the common good, trusteeship organizing proposes that managers act as trustees of capabilities enabled by knowledge within the firm's boundary, including firm-specific internationalizing advantages. It aims to guide responsible actions by default, not just through interpersonal relations or reciprocity, and argues that managers act as trustees to serve the common good. This includes social, environmental, and economic well-being. It focuses on the moral responsibility of human actors to organize AIs, as custodians of trust, to utilize capabilities acquired by knowledge bundles. Therefore, integrating trusteeship into the existing discourse offers a complementary perspective. While the established view of trust emphasizes perceived reliability and vulnerability, trusteeship highlights a moral responsibility and proactive governance in organizing. For AIs, trusteeship relates to managers in firms who are engaged in decision-making, preparation, motivation, planning, scheduling, execution, and evaluation that lead to outcomes (Wood, 2021). However, concrete outcomes can be both intentional and non-intentional (Wood et al., 2025). Managers enable action to shape organizing cognition and the existence of technological properties. They also shape attitudes towards something that are built

around previous actions that have a means-end relationship within a system. Reflecting on shaping the nature of a system, managerial actions consist of intended acts to lead, monitor, delegate, communicate, cultivate, coordinate, and reflect for the development and deployment of AIs. Therefore, managers may want to proactively not only strive to block harm emerging from intentional acts, but also to engage in experiments to block possible unforeseeable acts. Unintended acts may emerge from affordances due to given technological properties (i.e., facets of AI) as means. Within firms, trusteeship organizing enables internal organizing mechanisms where managers constantly reflect on information and (in)directly adjust their actions through engaging in external environments. The ability to enact effective internal coordination mechanisms that consider risk exposure and management is crucial for developing firm-specific competencies that lead to a competitive advantage in international markets (Luo, 2022). However, there are three imperatives for trusteeship organizing for competitive advantage.

2.5.2 Changing Nature of Production and Facets of AI

The first imperative arises from the rise of globally oriented algorithmic production. This shift in production challenges assumptions of both I-R and LCT theory. Value creation across borders occurs through computational advancement that does not necessarily involve physically bounded production processes (Faraj & Leonardi, 2022; Thompson et al., 2021). For AIs, production processes are shifting toward more computational and algorithmic systems instead of physically bound production processes (Baldwin, 2019). Such a change in value creation triggers the emerging strategic organizing, namely 'global by default' (Birkinshaw, 2022). The activities in global value chains are less capital-intensive (in terms of physical assets) and less location-bound (Autio et al., 2021; Rikap, 2022). Consequently, some traditional domains of inquiry, like subsidiary role typologies, are becoming less important. Value creation has thus transformed, which differs from the traditional sense, and in turn challenges the first assumption of the I-R theory. Contrary to assumption, AIs are increasingly developed and deployed globally by default, with limited dependence on location-bound production or market structures. Managerial action around contemporary AI plays a central role in enabling this shift.

Managers from many emerging firms are able to expand across borders almost instantaneously after launching products or services, yielding consumption convergence. Instances may consider the managerial actions at the Mistral AI, where managers operate with a largely unified systems architecture without conventional foreign subsidiaries or geographically fragmented production, apart from the physical constraints associated with data centers and computing infrastructure.

Similar are managerial actions around Anthropic's Claude, which represents a virtual AIs whose development and deployment are largely happening globally in a default responsible manner across jurisdictions.

Of course, recent technological advancements and techno-geopolitical uncertainty demand managers' deliberate actions to shape the nature of organizing systems, as per the first assumption of LCT theory. For example, Baidu's Apollo is a scalable AIs for autonomous driving—an FSA—that reflects its global ambitions. The managers concerned with deliverability integrate a full-stack driving solution that includes vehicle situational awareness (e.g., computer vision, LiDAR processing) for driving-related decision-making (e.g., path planning, prediction). The managers at Baidu use unified architecture, shared simulation, and safety testing protocols, but enable compliance with location or industry-specific regulations and safety standards. Here, managers leverage distinctive facets of AI—such as learning and autonomy—in algorithmic production (Berente et al., 2021; Murray et al., 2021). However, problematic managerial actions are arising where facets of AI scale without regard for human experiences across borders. To illustrate this, it is worth considering the performance of learning facets under model collapse in algorithmic production (Shumailov et al., 2024). When managers deliberately permit large language models to statistically learn from synthetic outputs without restraint, it erodes data quality and causes irreversible defects in models whose inputs and outputs lack genuine human experience. Such as a global learning shift from experience to reproducing model outputs, undoubtedly does not encode human meaning or local sensitivity by default, and the irresponsible utilization of the facet of AI generates systemic risk that undermines responsiveness. So while managers may achieve loose coupling, the model collapse illustrates how a facet of AI can progressively erase human experience and distort reality globally. These failures expose the limits of efficiency-driven organizing in algorithmic production, and the existing assumption does not consider moral reasoning in intentions to shape a system.

Trusteeship organizing responds by recognizing that human meaning is embedded in locally situated institutions such as cognition and judgment. Managers oversee facets of AIs in algorithmic production that enable interaction and shape user perception and behavior across borders. Yet, as the performance and scope of AIs expand, persistent challenges of responsiveness emerge across these facets. Development of AIs, therefore, requires human-centric approaches and addresses location-specific features to reflect institutional differences. Because human experiences used for facets of AI are morally significant and jurisdictionally embedded, they cannot be treated as exploitable inputs. Instead, they constitute rights that are entrusted. Trusteeship thus becomes a necessary intended organizing logic to align globally responsible development of AIs with locally grounded meanings in deployment.

Therefore, organizing AIs through globally coordinated trusteeship seems a responsible action for addressing institutions by default, instead of the dual pressure of I-R. As algorithmic production grows, so does its consumption, and trusteeship organizing is needed to prevent harm, pointing to the second imperative.

2.5.3 Changing Nature of Consumption and AI Anthropomorphism

Managers are increasingly deploying AIs across borders in consumption. A recent McKinsey global survey found that spending on development and deployment primarily focuses on AI-enabled products or services for user- or customer-facing activities (Singla et al., 2025). As per the second assumption of I-R theory, technological changes may promote convergence in consumption patterns worldwide. AIs have become much more widespread, which may facilitate convergence in consumption, and indicates the role of users' institutions in AI deployment and usage. Nevertheless, managers face growing responsibilities as there are growing risks of harm stemming from AI anthropomorphism (Bartneck et al., 2021; Claypool, 2023). AI anthropomorphism occurs when users attribute human-like qualities, intentions, or mental states to non-human agents, such as contemporary AIs (Zheng & Jarvenpaa, 2021). However, anthropomorphic reasoning neither occurs in objects nor in the minds of human observers in isolation (Duffy, 2003). Rather, it is a deliberate act by managers who make decisions on technological features and means imbued into AIs (Leong & Selinger, 2019). The tendency of AI anthropomorphism is enabled by technological features accompanied by affordances.

Such characterization underlies the second assumption of the LCT theory. For instance, OpenAI's ChatGPT operates on a globally connected infrastructure with minimal local adjustments. Through conversational interaction and responsiveness, it creates socio-emotional affordances that encourage human-like interpretations. Managers typically oversee the design of centralized models and training data in the firm's home country, and deploy them globally via applications, web interfaces, or local computing devices. This approach enables immediate consumption across countries, languages, and cultures, thereby amplifying anthropomorphism. Under AI anthropomorphism, affordances in consumption are multifaceted and often unpredictable, which poses a dire risk of harm. However, existing assumptions from both I-R and LCT theory on consumption coverage and technological affordances have little regard for social well-being.

Technological features create affordances under AI anthropomorphism (Zheng & Jarvenpaa, 2021), and such affordances intensify users' anchoring with their existing human knowledge and psychological needs while weakening the adjustment process

in AI anthropomorphism under cognitive load. The increased anchoring and decreased adjustment processes enhance egocentric biases that lead to negative consequences. Cognitive load reinforces the human mind's automatic responses and limits information-processing capabilities, thereby affecting perceptual and behavioral outcomes. As a result, users may overly rely on and place excessive trust in AIs, which increases the risks of privacy violations, manipulation, and other harms (Bartneck et al., 2021; Claypool, 2023). Users interact with AIs, which are perceived as trusted social actors, while responsibility remains fragmented and unclear across firms and maker-based regulatory institutions. In response, trusteeship organizing aims to proactively block harmful means and to organize AIs responsibly for consumption across borders. Trusteeship frames managers of deploying firms as trustees responsible for safeguarding users' rights and mitigating the risk of harm. It enables responsiveness because negative consequences are not simply confined to single markets, but arise across national borders. However, growing concerns about the risk of harm among international regulators and managers alike point to the third imperative.

2.5.4 Growing Risks of Harm and Insufficient Response from Regulators

As the nature of AI-driven production and consumption changes, the final imperative emerged due to a growing interface between managerial action within firms and regulators (Birkinshaw, 2022; Furr et al., 2022). The third assumption of LCT demands shared value as cognitive institutions to promote cohesiveness to address external environments, including market-based regulatory institutions. For AIs, however, managers seem to pay very little attention to international borders – in fact, they tend to bypass market-based institutional differences to achieve economic goals (Birkinshaw, 2024; Hemphill & Kelley, 2021). Regulators often respond slowly. When managers enable AIs to configure FSAs that operate globally, accountability remains fragmented. However, managers learn to exploit both markets and users through exploiting weak regulatory institutions. In turn, they gain monopolistic-competitive advantages through intellectual intangible assets like AIs, which impede innovation and jeopardize user well-being (Rikap, 2022; Weko, 2025). However, existing assumptions of LCT do not consider emerging accountability in instituting shared value, and in turn, the risk of harm is rising. Risk of harm is generally understood as a possible event or circumstance that negatively impacts individuals' well-being (Wood, 2021). Regarding AIs, negative impacts on well-being may arise from managerial actions that physically, psychologically, or spiritually injure individuals or cause financial loss. Such impacts can also emerge from the utilization of AIs in a given context that thwarts one's needs, development, potential, health, and dignity.

This view of harm extends beyond what is legally contestable and focuses more on social justice across institutional settings. However, the mechanism of inflicting harm through AIs can manifest in various ways, depending on actors' intentionality, affordances, and interaction outcomes (Wood, 2021; Wood et al., 2025). It may also emerge from the stratigraphy of harm concerning technological affordance, due to the presence of technological properties, such as anthropomorphic features in strategic design choices (Danaher, 2020; Wood, 2021). However, one can still assume that regulatory institutions play a vital role for minimizing risks of harm in line with the second assumption of I-R theory (Meyer et al., 2023).

Although regulatory responses are inadequate, significant progress has been made. For instance, the AI Act is the first-ever supranational and toughest market-based regulatory framework that adopts a risk-based approach (European Commission, 2024). It prohibits the development and deployment of AIs that may distort users' perception and behavior through deception, manipulation, and the exploitation of vulnerability. However, it remains inadequate when it comes to the risk of harm under AI anthropomorphism (Franklin et al., 2023). Recent allegations and lawsuits underscore the severity of these risks. Reported incidents linked to both virtual and robotic AIs include sexual intimacy, violence, self-harm, and death, often exploiting vulnerable users (Burga, 2025; Horwitz, 2025; Jargon & Kessler, 2025; Marsden, 2025). These troubling incidents raise significant concerns among regulators about potential harm that arises intentionally or non-intentionally through technological means (e.g., through anthropomorphic properties) and affordance (Wood et al., 2025). Therefore, there is a dire need for trusteeship organizing that may work as a shared value that aims to ensure accountability against the risks of harm, as implied in the third assumption in LCT. However, the view of trusteeship organizing as a shared value can be seen as a cognitive institution that guides managerial actions, and helps to enhance competitiveness and regulatory response that integrates both economic and social goals. In short, this view links competitiveness to the ability to perform not only through scale or profit, but also by addressing local communities. Taken together, these three growing and interrelated imperatives call for a globally responsive organizing of AI through trusteeship.

In a nutshell, the forehead assumptions presented three imperatives that highlight a shift in strategy around AI towards trusteeship, viewing managers as custodians of user welfare rather than only aiming at profit maximization. Collectively, they emphasize to enact a trusteeship framework for responsiveness through responsible AI governance across borders, integrating economic goals with social welfare to support the common good.

2.6 Towards a Globally Responsive Trusteeship Organizing of AI

The globally responsive trusteeship approach helps to create trust-based yet systems-oriented shared values. Simultaneously, it also addresses local sensitivity, given the diverse normative and cognitive institutions across national borders and even within a single nation (Messner, 2022c; Yoon et al., 2025). This approach has the potential to tackle governance challenges faced by managers. These challenges arise from the fragmented nature of digital globalization, shaped by geopolitics and regulatory institutions (Nambisan & Luo, 2021). Accordingly, this cognitive framing also enables regulators to govern AIs to promote social justice to underpin the common good. The trusteeship view of globally responsive AI organizing also complements broader discussions on AI-related responsibility and governance. Trusteeship organizing particularly relates to market-based regulatory frameworks, such as the AI Act (European Commission, 2024). It aims to achieve the common good and support global sustainable development (Jobin et al., 2019; Vinuesa et al., 2020), and complements cross-border discourse on AI that focuses on responsibility and social good (Cowls et al., 2021; Floridi et al., 2020). Therefore, it seeks to integrate economic goals such as economies of scale, with social goals including human values and institutional sensitivity. It does so through responsible development and deployment. This view responds to key assumptions in I-R theory and LCT (see Figure 7).

Assumptions of I-R theory remain grounded in reasoning derived from large multinational firms, location-bound production, and a pressure of adaptation across markets. Assumptions of LCT provide limited guidance on how managers should organize facets of AI to remain responsive when the nature of these facets is emergent and often extends beyond firm-level outcomes. Additionally, technological affordances may create convergence in perceptions and behavior among users in consumption-facing activities (Tatarinov et al., 2023; Zheng & Jarvenpaa, 2021). Collectively, they emphasize only economic reasoning, including production efficiency, profit maximization, and economies of scale. These dominant economic reasons leave little room for societal welfare considerations, and largely overlook distinctive technological properties of AI. In response, the trusteeship framework addresses convergence of consumption that requires global coordination and a shared value for responsiveness (Bartlett & Ghoshal, 1989; Nambisan & Luo, 2021). Integrating trusteeship may universally ensure responsiveness across diverse institutions through embedding responsibility into the development and deployment of AIs. In turn, managerial action driven by trusteeship framework may achieve the common good. Trusteeship organizing integrates reasoning on both economic and social goals in responsible production and consumption. Since the dissertation

addresses the socio-technical nature of AI, which is emergent, there is an increase in managerial responsibility in firms' internationalizing activities. Trusteeship organizing may thus help them create shared value for the common good while remaining globally responsive, configuring FSAs. Therefore, a global trusteeship by default framework as a cognitive institution seems critical for managing AI across borders.

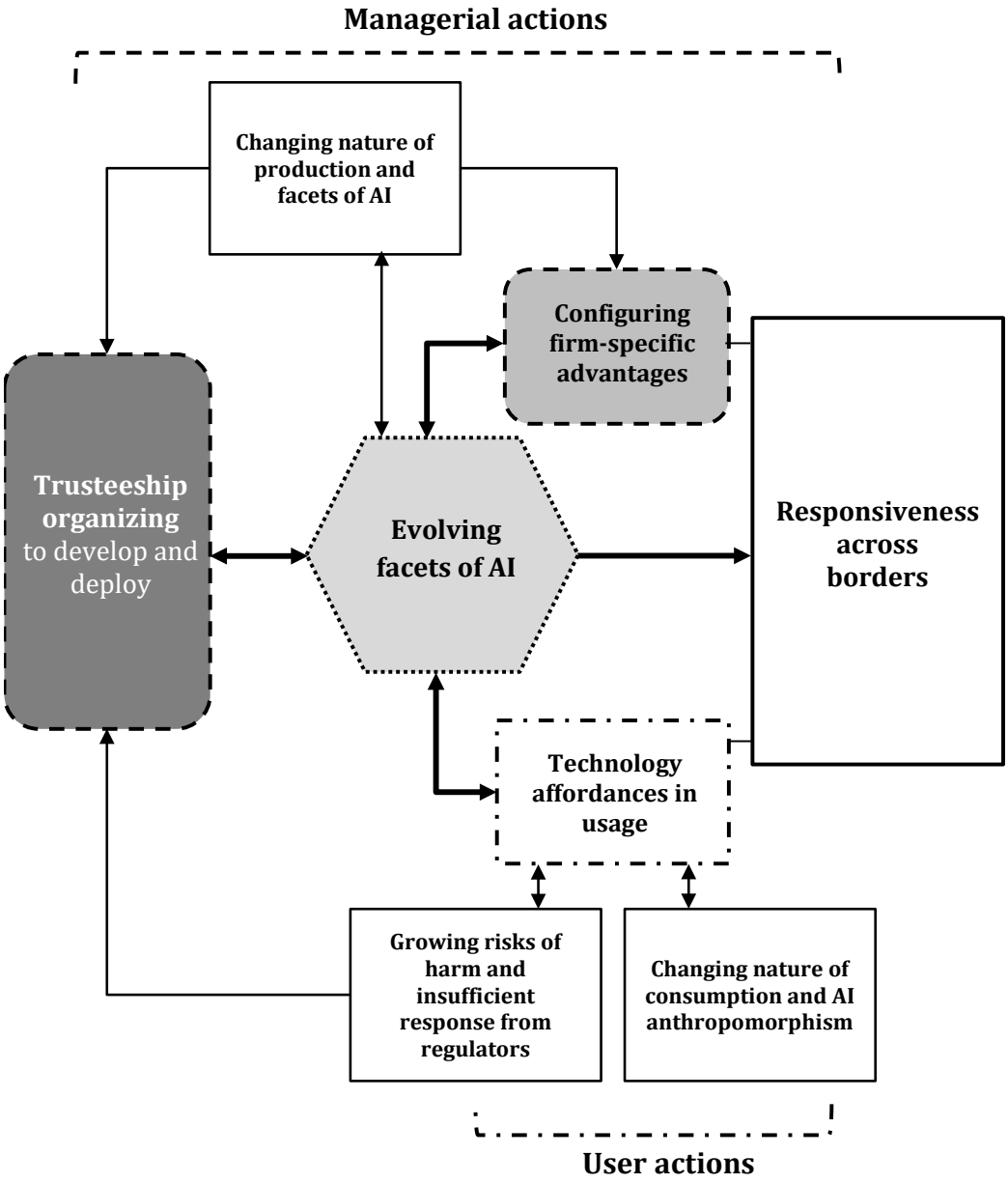


Figure 7. Rethinking for strategic organizing of AI for worldwide effectiveness

Importantly, AIs are increasingly accompanied by the risk of unintended negative consequences, including bias, misinformation, and misalignment with societal values.

These risks foreground the social imperative of responsibility to organize AIs for responsiveness. Responsiveness, therefore, may no longer be reduced to efficiency gains or market adaptation alone. Instead, it also encompasses intentionality with which managers oversee development, govern, and steer AIs to address firm-specific competencies such as decision-making and value creation, while simultaneously engaging with broader societal and regulatory expectations. This reconceptualization calls for a new theorizing of responsiveness that places AIs as a source of FSAs. It also embodies managerial intentionality and cross-border societal implications in its theorizing. Organizing AIs across borders, therefore, may require a shared value of trusteeship that provides a basis for responsible cross-border governance. From LCT, the dissertation conceives the trusteeship framework as a shared value for responsiveness, integrating moral intentionality, human dignity, and accountability. Reflecting on socio-technical systems, it points to managerial activities as custodians of trust in the development and deployment of contemporary AIs across borders. Accordingly, it emphasizes human agency to act justly and responsibly, and to be accountable for any misuse or misconduct in the common good. From I-R theory, trusteeship may serve as a globally coordinated governance mechanism that enables responsible consumption and addresses local sensitivities during deployment.

Against this backdrop, internal strategic organizing of AIs in the Fifth Industrial Revolution differs from the physically bound industrial era (Bartlett & Ghoshal, 1989; Birkinshaw, 2022). The emergent nature of facets such as autonomy and learning makes the external organization of AIs distinct from that of broader digital artifacts (Berente et al., 2021; Nambisan & Luo, 2021). This shift necessitates rethinking the process of responsiveness. These facets of AI fundamentally transform how managers organize and scale their internationalization activities, and introduce new strategic tensions that existing theories do not fully explain. However, enacting trusteeship logic as a shared value in worldwide effectiveness offers a valuable path for international managers and regulators (see Figure 7). The existing discourse needs to move beyond economic goals such as economies of scale and profit maximization (Hutzschenreuter & Lämmermann, 2025). To address global challenges, organizing includes concerns related to fairness, manipulation, and sustainable consumption (Dolata et al., 2022; Jobin et al., 2019). These concerns are relevant not only to international managers, but also to regulators. Hence, rethinking responsibility in globally integrated AI strategies is critical for international strategy and information management discourse. At the same time, it is also important for international regulators to develop or amend regulations to safeguard users' well-being. As AIs increasingly influence global production and consumption patterns, these advanced artifacts pose risks of harm to users and raise ethical challenges for strategic organizing (Bartneck et al., 2021; Berente et al., 2021). Therefore, there is a pressing need to rethink strategic organizing to include both the globally relevant facets of AI

and a responsibility for common good. It is on this integrated understanding that this dissertation develops its framework for organizing AIs responsibly across borders by employing a systems-oriented methodology.

3 METHODOLOGY

Methodology is a mode of systematic thinking and reasoning in science (Arbnor & Bjerke, 2009; Furnari et al., 2021; Welch et al., 2022). For research processes, it sheds light on ways of reflection for seeking knowledge of realities and for being coherent in intellectual judgment (Al-Ghazālī, 2024/1095). In social science, methodology is not just about choosing among different methods of data collection, production, and analysis (Archer et al., 1998). It is also about reflecting on choices or the development of suitable methodological approaches for theorizing (Arbnor & Bjerke, 2009; Köhler et al., 2026; Welch et al., 2011). Hence, methodological pluralism is paramount to advance science. Positioned within methodological pluralism, this dissertation acknowledges and embraces diversity in scientific inquiry and multiple ways of thinking and acting (Avital et al., 2017; Miles et al., 2014). This openness enabled a flexibility to innovate and adapt multiple views, established reporting styles, research methods, and analytical approaches. Such a position is considered necessary to continue the advancement of IB and IS as a field of science, particularly given their inherent multidisciplinary domains (Avital et al., 2017; Furnari et al., 2021; Welch et al., 2011). Instead of merely following prescriptive cookbooks, method templates, or common writing genres, this dissertation adopted a critical and reflexive approach (Al-Ghazālī, 2024/1095; Braun & Clarke, 2025; Cunliffe, 2003). However, the present language and methodological reflections are somewhat distinct, as they mingle the researcher's self-consciousness that aligns with chosen methodological view (Arbnor & Bjerke, 2009). Doing so extends exciting method-related cognitive institutions (i.e., schema, reporting style) that are not only appreciated, but also desired in both IS and IB (Grover & Lyytinen, 2015; Welch et al., 2022).

3.1 A Systems-oriented View of Methodological Complementarity

Methodological view is a form of reflection that involves self-awareness and criticality of method and related decisions throughout the research process. In turn, it guides thinking and acting (Arbnor & Bjerke, 2009). Reflection in research guides the step-by-step examination of a question that moves from clear concepts to sound judgments. This examination is not necessarily linear and can be recursive, and when possible, it demonstrates a degree of certainty by understanding relations among conceptual entities (Al-Ghazālī, 2024/1095). This dissertation emphasizes a strategic management of AI, and more specifically, the question of how AI can be strategically organized across borders in a globally responsive manner to integrate economic and social goals. It emphasizes strategic rethinking that emphasizes foundational aspects of organizing a system in a firm-level setting. Additionally, its theoretical

underpinnings also underlie systems thinking, particularly the social-technical aspects of AIs and loose coupling organizing (Ali et al., 2025; Berente et al., 2021; Nambisan & Luo, 2021). Simultaneously, it embodies a systems-oriented view as its focal methodological mode of reflection for seeking knowledge. Therefore, the dissertation's methodological view may be best described as a systems-oriented view of methodological complementarity (Arbnor & Bjerke, 2009). However, there are three methodological views: the analytical, systems, and actor views that are used to seek knowledge (see Table 3 for a brief comparison and Arbnor and Bjerke (2009) for a detailed account).

Table 3. Presentation of different methodological views

Focal view	Core focus	What is reality? (Ontology)	How to acquire knowledge of reality? (Epistemology)	How to study reality? (Methods)	Root paradigm and theory
The analytical view	Generalizable explanations	Filled with facts, independent of individual perceptions, and summative in nature.	Regularities and common opinions.	Laws of cause and effect, hypothesis-driven inquiry, surveys, publicly available data, statistical testing, reliability, and dominant opinions.	Positivism, empiricism, rationalism, materialism, and analytical philosophy.
The systems view	Explanation and/or understanding of a complex phenomenon	Consists of interrelated structures and patterns, and parts that co-influence each other.	Design, irregular aspects of the context, regular patterns, interactions, relations, and situations of structures.	Existing literature, case studies, system mapping, historical/longitudinal analysis, secondary information, interviews, trustworthiness, and illustrations.	Holism, pragmatism, structuralism, systems theory, sensemaking, metaphorical thinking, and materialism.
The actors view	Deep understanding and insightful action	Reality is socially constructed, through meanings and context-dependent.	Co-creation of meaning, non-linear process, unique, chaotic, and specific, and self-reflection,	Dialectical-driven inquiry, participatory approaches, iteration, metaphor, credibility, and interpretations.	Social constructionism, pragmatism, idealism, sensemaking, metaphorical thinking, grounded theory, critical theory, and hermeneutics.

The systems view has various practices. From a methodological point of departure, this dissertation reflected on systems in terms of paradigmatic stances or ultimate presumptions, interesting issues and perspectives, conceptualization, methods, and results (Arbnor & Bjerke, 2009). It conceived the systems view as a holistic approach to studying reality, aiming to explain and understand knowledge. It also provided a coherent stance about what reality is, how it can be known, and which methods and practices are suited to studying and improving business knowledge. Such a methodological approach served a dual function that encompasses certain ultimate presumptions and serves as the prerequisite for the design of practical procedures. It further clarified ultimate assumptions underlying a paradigmatic stance in subject areas such as international strategy in IB and information management in IS. However, these assumptions or paradigmatic stances also shaped research process, practical procedures, instruments, and reporting. Hence, the system-oriented complementarity view has been a coherent stance of reflections that articulated the researcher's ultimate presumptions about reality (ontology) and the nature of knowledge (epistemology), and enabled a translation of those reflections into a structured logic of inquiry (methodology) with operative practices and methods for seeking knowledge.

Such a systems-oriented view is grounded in five principles (Arbnor & Bjerke, 2009). First, reality is a network of interdependent components rather than isolated parts, which stresses the principle of totality – the whole is more than the sum of its parts. Second, the view recognizes complexity, meaning that any model or interpretation of a system is necessarily limited and shaped by practical delimitation, and there are possibilities for multiple perspectives and magnifying levels. Third, it embraces relativity, acknowledging that system representation depends on framing and conceptualization. It accepts that every representation of empirical reality is approximate, and relative to the given perspective and purpose. Fourth, elements in a system influence each other reciprocally, which points to the principle of mutuality. Thus, a cause in one subject may become a predicate in another. Finally, there is a fundamental limit to predicting a system's future behavior, which underlies the fifth principle of unpredictability. This arises since the constructed system constantly interacts with its environment, potentially leading to (un)intended consequences. Therefore, there may be a need for a sincere effort to improve intellectual judgment, in order to achieve universal meanings and their certainty (Al-Ghazālī, 2024/1095).

According to the systems-oriented view, reflecting requires a holistic approach from ontological (nature of reality), epistemological (nature of knowledge), and methodological perspectives (strategy for generating knowledge). As most readers in social sciences observe, established presentations use these concepts to classify different paradigms (Burrell & Morgan, 1979; Guba, 1990). However, the present

presentation is consistent with these concepts yet distinct from such well-known classifications, which follow. Particularly, it embodied various ontologically and epistemologically classified paradigmatic stances (ultimate presumptions) concerning methodology (see Arbnor et al., 2009 for a more detailed account). However, this dissertation has not taken the existing systems view for granted, and has consulted and corroborated with wider scholarly works on methodology (Al-Ghazālī, 2024/1095; Archer et al., 1998; Mingers et al., 2013). Hence, constant reflectivity enabled adaptation even within ultimate presumptions, leaning toward a methodology of complementarity (Arbnor & Bjerke, 2009). In other words, it employed the systems view as the primary mode of reflection, while embracing pluralism (Welch & Piekkari, 2017) and incorporating complementary ideas from other views, namely the analytical and actor views. Such a pluralistic approach with complementarity involved adapting existing concepts and theories, as well as collecting, framing, and interpreting data. Therefore, complementary procedures were configured against the ultimate presumptions of the primary systems view. However, the reported view evolved over time until closure of this dissertation. Therefore, the presentation style employed here is reflexive, which may not look like other approaches, yet still relate to established paradigmatic stances. In this dissertation, such constant reflection is deemed necessary to create an impact in society and achieve the common good through research in IB and IS.

3.2 Ultimate Presumptions: A System-oriented Paradigmatic Reflection

Reflecting on the systems-oriented complementarity, this dissertation adapted conscious assumptions about reality, accompanied by a clear understanding of what knowledge is and how it emerges. Such an emerging paradigmatic stance held certain underlying assumptions, indicating a particular way of reflecting on study areas such as international strategy or information management. It extended to embody moral inclusiveness and other relevant paradigmatic stances in knowledge-seeking. Therefore, the dissertation emphasized ultimate presumptions that comprise a conception of existence, a conception of reality, a conception of knowledge, a conception of science, a scientific ideal, as well as responsibility and aesthetic aspects.

3.2.1 A Conception of Existence

Existence means that a thing is, as opposed to not being at all (Al-Ghazālī, 2024/1095). It concerns what kinds of things actually exist and how they exist. Things differ in their existence, whether they stand on their own or depend on

something else. A thing is real by its substance, which exists by itself – for example, different forms of contemporary AIs exist in international markets. Additionally, something may be present not on its own, characterized by the existence of essences in a particular thing. For example, the essence of learning or autonomy is embedded in contemporary AIs (Berente et al., 2021). This understanding of existence provides an important perspective on the distinction between properties and independent objects, like the characteristics of contemporary AIs and the phenomenon of AI itself. However, knowing what something is does not necessarily mean it exists (Al-Ghazālī, 2024/1095).

A thing's essence is thus understood even when its existence is unknown, doubtful, or denied. For example, one might know what artificial superintelligence is and its characteristics, but it does not yet exist (Bostrom, 2016). Therefore, there are three fundamental types of existence that exist in some way, independent of human observation (Al-Ghazālī, 2024/1095). First, the actual existence that could exist by its own right or that depends on something else. For example, firms, managers, users, virtual AIs, events (human-AI interaction), and a regional cross-border market. Second, the possible existence of things that could exist by their own right, or that depend on something else, or not at all. For example, the potential of artificial general intelligence in the near future, and responsible localized deployment of AIs in a given market. Third, the necessary existence of a being that must exist and depends on nothing else. Only God has necessary existence for a believer, which is outside the scope of this dissertation. However, this understanding now points to the existence of reality.

3.2.2 A Conception of Reality

Reality consists of fact-filled system structures in objective reality, and of subjective opinions of such structures (Arbnor & Bjerke, 2009). Reality (what truly is) is whatever exists, either within the human mind or inside or outside of the external world (Al-Ghazālī, 2024/1095). Realities arise from the subject matter of knowledge and reasoning, and the existence of reality is arranged in such a way that the whole differs from the sum of its parts. It emphasizes objects and entities, as well as their interactions and relations within a system, and arguably, the behaviors, tendencies, or ways of acting of those individual parts. In the context of the common good, facets of AI evolve as the scope of affordance and performance expands. In turn, differing realities may arise, with both positive and dire consequences (intended or unintended), triggered by managerial and user actions. Therefore, reality about AIs is conceived as emergent and relational, where outcomes arise from the interplay between actors and entities. The same circumstances around AIs may simultaneously

generate economic outcomes through FSAs and risk of harm for users, resulting in multiple yet interconnected realities that cannot be fully understood in isolation. However, there could be different modes of existence for realities, independent of human observation, which can be unknown, doubtful, or denied (Al-Ghazālī, 2024/1095).

At the systems level, there are relevant presumptions in the conception of reality (Archer et al., 1998; Mingers et al., 2013; Welch et al., 2011). Reality is thus conceived as a stratified yet dynamic system. The basic assumption about reality is that there is a real world independent of human knowledge, observation, or experimentation. As a whole system, reality consists of objects, entities, mechanisms, events, and experiences. However, reality has three domains: the real, the actual, and the empirical, seen as stratified systems (Archer et al., 1998; Mingers et al., 2013). The real domains consist of objects or entities, events, and experiences, arguing for a whole reality. The actual consists of events that occur (or not) independently of humans. In the empirical domain, those events may (or may not) be observed or experienced. One should not reduce all events to those that are observed or experienced. Constructing knowledge in business, as in other social sciences, is more difficult than in natural sciences. Thus, clarity about reality is, for the most part, nonexistent. Therefore, this dissertation strives to strike a balance between deep data engagement and imaginative theorizing (Köhler et al., 2026), leading to the concept of knowledge.

3.2.3 A Conception of Knowledge

Knowledge is an organized form of information, understanding, and theories, characterized by approximations, evidence, and improvement as new evidence emerges (Archer et al., 1998; Furnari et al., 2021). It provides an objective and a subjective understanding, which are both universal and contextual (Welch & Piekkari, 2017). Knowledge may also relate to absolute certainty about the existence of reality. Hence, knowledge can be divided into two categories (Al-Ghazālī, 2024/1095). First, knowledge of the essence of things, objects, or entities, such as a knowledge of managers, users, or a particular facets of contemporary AIs. This kind of knowledge which is called a conception, and concerns parts of systems. Second is a knowledge of the relationship of conceived essences to each other, either by means of mechanism, event, and/or experience. Therefore, one must understand how to organize AIs in firms by conceptualizing their essence (characteristics or facets) at intersection of managers and/or users. However, it also concerns whole system. Reflection on a singular component of parts then relates to the compound of entire systemic behavior. In total, it aims at a full or approximate knowledge of realities

through understanding and explaining structure, based on empirical evidence or proofs (Al-Ghazālī, 2024/1095; Archer et al., 1998; Mingers et al., 2013). However, a potential structure that causes or emerges from these events at the intersection of objects and entities can exist in unseen and accidental forms (Arbnor & Bjerke, 2009).

As argued earlier, underlying actions regarding AIs may have unforeseen and non-intentional negative consequences, indicating possible risks of harm across national borders. The mode of analogical reasoning thus may include abduction (Dubois & Gadde, 2002; Welch et al., 2011) and generative doubt with rational control and irrational free-play for theorizing (Köhler et al., 2026). To explain and understand the knowledge of reality, one must tap into the realm of objects and structures themselves, in order to understand underlying mechanisms. The reality thus consists of structures of objects and entities, both material and social, that give rise to particular behaviors, tendencies, or ways of acting, called mechanisms. These mechanisms may (or may not) trigger events, and the mechanism of events at one level can be seen as being generated by those of the lower-level objects (Bechtel, 2009; Machamer et al., 2000). Consequently, the systems-oriented methodological complementarity view may explain parts in terms of characteristics (facets of AI) of the whole of which they are part (trusteeship organizing of AI), so as to capture an approximate reality to achieve responsiveness. However, knowledge about the existence of reality is directly relevant to science.

3.2.4 A Conception of Science

The system view conceives many ways of gaining and producing knowledge about the existence of realities (Arbnor & Bjerke, 2009). For AIs, one prominent way is termed as scientific, which enables an explanation and understanding of knowledge that emerges from empirical reality, as argued above. Other ways could be through consultancy and policy debates, for instance (Claypool, 2023; European Commission, 2019). In pursuit of knowledge, science is a systematic process of inquiry into realities in natural and social worlds. Science may also relate to outer space, the universe, and beyond, like other worlds, the multiverse, or the hereafter, which are outside the scope of this dissertation. Consequently, science enables a better understanding and explanation of phenomena under investigation, and delves into conceptualization, modeling, and frameworks that underlie theorizing processes. However, the scientific approach has two necessary yet interconnected characteristics (Arbnor & Bjerke, 2009).

First, it must be an engaged and sincere attempt to explicate the relation between conception and empirical observation, and also be relevant to reality (empirical reality) (Mingers et al., 2013; Welch et al., 2022). This view holds that knowledge of

empirical reality arises from inquiry, reflection, reasoning, and observation, which is subject to critical scrutiny and intellectual challenge (Welch et al., 2011). Accordingly, through the scientific inquiry undertaken across Articles I-III, this dissertation recognizes that strategic organizing AI for responsiveness across borders is associated with both opportunities and challenges. For managers, AIs may create opportunities for developing firm-specific advantages while posing challenges for responsible organizing. For users, however, AIs may give rise to risks of harm, misplaced trust, and unintended consequences stemming from their evolving affordances. For regulators, AIs introduce concerns related to data privacy, accountability, and deceptive practices. These perspectives represent different yet interconnected knowledge of realities that emerge from scientific endeavor undertaken across Articles I-III.

Second, using a scientific methodology necessarily requires a conscious application of reasonably clearly formulated rules and analogical reasoning (Al-Ghazālī, 2024; Furnari et al., 2021). Procedures and results must also be explicitly articulated for other members of the scientific community and for critical scrutiny. Accordingly, this dissertation engages in scientific inquiry to gain knowledge about strategic organizing of AIs that integrate economic and social imperatives, enabling responsiveness across borders. Consistent with this aim, operatives and analogical reasoning embrace reflexive openness to emphasize transparency regarding the researcher's actions and methodological integrity to ensure coherence among ultimate assumptions, research design, methods, and theorizing (Braun & Clarke, 2025). Together, these characteristics clarify that this summary dissertation is oriented toward understanding, interpretation, and explanation. However, scientific inquiry also requires a scientific ideal concerning the researcher as a person.

3.2.5 A Scientific Ideal

The author of this dissertation presumes two things in his activities in business as a science. First, the systems-oriented methodological complementarity view implies a flexibility to develop new and emerging concepts that better capture systematic (re)interpretation of responsiveness and AIs in international strategy and information management reality. The whole and patterns provide a processual perspective that requires continuous effort to improve knowledge to underpin the strategic organizing of AIs, and so achieve better explanations and a better understanding of the various types of organizing systems. As argued earlier, organizing facets of AIs is emergent and adaptive, in which behavior can change over time in response to different internal and external circumstances (Berente et al., 2021).

Second, the researcher views himself as a seeker of knowledge. This rationale draws on a firm personal belief in the unity of knowledge, rooted in an unseen, singular source that structures reality as an integrated whole. While humans have limited knowledge, they also suffer from bounded rationality and cognitive load (Simon, 2000; Zheng & Jarvenpaa, 2021). However, they are still accountable agents who must be sincere and responsible in their thinking and actions. Many aspects of reality as a whole may not be comprehensible for a single human being to interpret, understand, and explain. Therefore, any knowledge derived from a scientific way may still be fallible, subject to temporality, and confined to institutions. But within the scientific way, this dissertation seeks knowledge to explain and understand through an intentional and critically engaged effort, which is deliberate, purposeful, and ethically bound to acquire, extend, or challenge existing knowledge on international strategy and information management (Arbnor & Bjerke, 2009). However, knowledge is also subject to intellectual judgment within responsible and aesthetic limits through questioning, humility, debate, and innovation.

3.2.6 Responsibility and Aesthetic Aspects

This dissertation presumes that the responsibility and ethical consequences of knowledge gained and disseminated through systems views are increasingly important in science. Business knowledge is directly linked to society, as research examines the relationships between firms and their surroundings, including employees, users or consumers, environments, and technology. For example, the way contemporary AIs are developed and deployed significantly impacts how AI-based solutions or artifacts behave in the real world, and underlies economic reasoning (Parkes & Wellman, 2015; Rahwan et al., 2019). Hence, focusing on responsible organizing, reflecting on truth, and serving humanity to achieve the common good is key to science in seeking knowledge about AIs. Moral inclusiveness is paramount to seeking knowledge and constructing understanding through science. Hence, knowledge of strategy and AIs that derives from the scientific way needs to be helpful in achieving good for oneself and for society as a whole (Ali et al., 2025; Floridi et al., 2020; Fuso Nerini, 2026).

Aesthetically, the dissertation aims to provide an approximate yet holistic description of the business reality of AIs in IB and IS. These descriptions are presented in the form of illustrations and meaningful visual representations of knowledge, whenever possible, such as examples, tables, and figures. However, they should also provide an overview to facilitate practical action to create impact. It is hoped that these descriptions may inspire other members of broader society and scientific communities.

In summary, the ultimate presumptions outlined above underpin the dissertation's systems-oriented methodological complementarity view. It allows for exploring dynamic interactions and systemic patterns to achieve a deeper explanation and understanding of contemporary AIs. However, when a particular methodological view is applied to pursue knowledge, it is called a methodological approach. In turn, it contemplates research design, including data sources and analysis. Such paradigmatic stances guide the identification of interesting issues and perspectives, conceptualization, the development of coherent methods, and the presentation of results (Arbnor & Bjerke, 2009). Thus, the relationship between these presumptions, accompanied by the systems view, is linked to the operative paradigm that enables the research to be conducted. Accordingly, it addresses key concepts that describe steps and relations involved in seeking, understanding, and explaining knowledge. This sub-chapter now turns to practicalities.

3.3 A Multi-method Qualitative Research Strategy: The System-oriented Operatives

This dissertation employs a multi-method qualitative research strategy that involves operatives (operational procedures) from two or more methods, such as design and/or techniques. Unlike mixed methods, which blend operatives from both qualitative and quantitative traditions, a multi-method strategy remains within one tradition or the other. One particular note on complementarity is that any operative can be configured from any methodological view, including analytical, systems, and/or actor views, to complement one another when needed. Hence, this dissertation remains within the qualitative tradition in line with the paradigmatic stances (ultimate assumptions) presented earlier. System-oriented operatives often focus on existing literature, case studies, secondary sources, interviews, transparency, and illustrations. Consequently, the practice employs multiple operatives, including an abstracted configurational approach, qualitative directed content analysis, a single contextualized explanation case study, and a single mechanism-based explanation case study.

All of these operatives utilize empirical insights from primary and/or secondary sources for both conceptualization and theorization. This multi-method qualitative research enables the overall dissertation to consist of multiple parts. The dissertation comprises three articles that collectively address the central research question, and their collective understandings go beyond the sum of their individual contributions. Therefore, this summary dissertation serves as a whole. Accordingly, there are four outcomes in this pursuit of knowledge: the summary dissertation (whole) and three articles (parts). Please see Table 4 for a more detailed account. The system-oriented

research design and operatives employed in this dissertation can be divided into two groups: those used in the summary dissertation and the others used in three articles.

Table 4. A methodological overview of the dissertation

	Whole	**Part 1**	**Part 2**	**Part 3**
		Knowledge outputs		
Item	**Summary dissertation**	**Article I**	**Article II**	**Article III**
Title	Managing artificial intelligence across borders	Managing artificial intelligence in international business: Toward a research agenda on sustainable production and consumption	The dark side of AI anthropomorphism: A case of misplaced trustworthiness in service provision	Artificial intelligence and anthropomorphism: A case of risk of harm on the global stage
		Ultimate presumptions		
Paradigmatic stances	The systems-oriented methodological complementarity	Critical realism	Critical realism	The systems-oriented methodological complementarity
		Interesting issues and perspectives		
Conceptualized type of narrow AI systems	Robotic, virtual, embedded, and algorithmic	Virtual, robotic, and embedded	Virtual, robotic, and embedded	Robotic and virtual
Tensions and conflict	Urgent need for a responsible shared value in organizing that is global by default, but locally responsive	Balancing economic and social goals	Dark pattern of technology usage	Emerging AI-related harm from development and deployment
Magnifying level	*High*: Key facets of AI systems and their relations to globally responsive organizing	*Balance*: Key characteristics of AI systems and their relations to international business management	*Low*: Key anthropomorphic features of AI systems and their relations to the development of misplaced trustworthiness	*Low*: Key facets of AI and their relations to potential harm across borders via deception and manipulation
		Conceptualization		
Conceptual underpinnings	The AI frontier perspective; Global integration-local responsiveness theory; Loose coupling theory; The trusteeship perspective to AI ethics	The AI frontier perspective	The 'computers are social actors' paradigm; The three-factor theory of anthropomorphism	Global integration-local responsiveness theory; The trusteeship perspective to AI ethics
Research context	Globally responsive information management	Responsibility in managing international business activities	Service provision	Service tasks across borders

	Whole	Part 1	Part 2	Part 3
Unit of observation	Managerial actions, user actions, and AI systems	Managerial actions and AI systems	User actions and AI systems	Managerial actions, user actions, and AI systems
Unit of analysis	Organizing facets of AI in development and deployment to enable globally integrated, responsible organizing for responsiveness across borders	Characteristics of AI and their management in international business activities to achieve responsible production and consumption	Utilization of anthropomorphic features in AI that enables misplaced trustworthiness among consumers	Potential risk of harm under AI anthropomorphism
		Methods		
Research design	Qualitative, abstracted configurational approach	Qualitative, directed content analysis	Qualitative, single contextualized explanation case study	Qualitative, single mechanism-based explanation case study
Data sources	Data from three articles	Published empirical research (n=85)	Interview (n=14)	Interview (n=14), review research (n=44), and archival media data (n=210)
Analysis	Configurational through systems interpretation	Directed content analysis	Configurational theorizing approach	Three-step analytical approach
		Results		
Key findings	- Four key facets of AI—autonomy, learning, interactivity, and appearance. - Trusteeship covenants, including addressing institutions, blocking harmful means, and ensuring accountability to manage those facets, enable responsiveness across borders,	- Managing three characteristics of AI (autonomy, learning, and combinative) has profound implications to balance social and economic goals in international business operations. - Managers can utilize these characteristics in their international expansion activities at the intersection of responsible consumption and production.	- Theorize that physical and behavioral anthropomorphic features imbued in AI agents enable consumers to develop misplaced trustworthiness. - Highlight the interplay of machine and consumer behavior factors that amplify the undesired outcome.	- Construct three facets of AI related to anthropomorphism (appearance, interactivity, and autonomy) that lead to risks of harm. - Theorize the mechanism of AI anthropomorphism that leads to harm via deception and manipulation.
		Other practicalities		
Tools for aesthetic illustration	Microsoft Word: Drawing Canvas and Table function	None	Adobe Illustrator and Microsoft Word: Table function	Microsoft Word: Drawing Canvas and Table function
(Un)paid tools for language improvement	A human editor and Microsoft Copilot (university-paid), and Grammarly (self-paid)	A human editor (university-paid) and LanguageTool (self-paid)	Microsoft Copilot (university-paid) and LanguageTool (self-paid)	A human editor and Microsoft Copilot (university-paid), and Grammarly (self-paid)

	Whole	**Part 1**	**Part 2**	**Part 3**
Responsible use of AI in writing	Microsoft Copilot (university-paid): Restricted to brainstorming and only directed to writing tasks for clarity and improvement; the author reviewed, edited, and reconstructed sentences to make further sense based on or to align with his original input.	None	None	Microsoft Copilot (university-paid): Restricted to brainstorming and only directed writing tasks for clarity and improvement; the author reviewed, edited, and reconstructed sentences to make further sense based on or to align with his original input.
Author(s)	Rakibul Hasan	Rakibul Hasan and Arto Ojala	Rakibul Hasan, Arto Ojala, Sara Quach, Park Thaichon, and Scott Weaven	Rakibul Hasan
Possible publication	Unpublished: Tentatively targeting a respected journal in IB by transforming into a conceptual paper	Already published: Thunderbird International Business Review (2024)	Already published: Proceedings of the 58th Hawaii International Conference on System Sciences (2025)	Unpublished: Tentatively targeting a respected journal in IS by further improvement
Transparency on author(s) contributions *(using CrediT: https://casrai.org/credit)*	Sole authorship	Rakibul Hasan: Conceptualization, data curation, methodology, formal analysis, visualization, writing – original draft, and writing – review and editing. Arto Ojala: Supervision and writing – review and editing.	Rakibul Hasan: Conceptualization, data curation, methodology, investigation, formal analysis, visualization, writing – original draft, and writing – review and editing. Arto Ojala, Sara Quach, Park Thaichon, and Scott Weaven: Supervision and writing – review and editing.	Sole authorship
Targeted SDGs	SDG#12 on responsible consumption and production, and SDG#16 on peace, justice, and strong institutions	SDG#12 on responsible consumption and production	SDG#12 on responsible consumption and production	SDG#12 on responsible consumption and production

3.3.1 The Summary Dissertation: An Abstracted Configurational Approach to Understand and Explain a Management System

The research design of the summary dissertation may best be described as an 'abstracted configuration approach' to understanding and explaining through system interpretation (Arbnor & Bjerke, 2009). The approach is well-suited to studying strategic organizing of contemporary AIs, as it enables the analysis of interrelated components and their evolving relationships within a broader management system. The outcome of this approach is a processual understanding of strategic organizing of AIs within a strategic management system, which itself is a subsystem of a firm. Firm-specific human actors like top and middle managers operate it, and it requires continuous managerial effort. The core focus of the summary dissertation is therefore to explain how managers organize contemporary AIs to achieve responsiveness across borders, while integrating reasoning into economic and social goals. Managing AIs is thus a socio-technical system comprising managers, end users, and regulators. Managers are responsible for development and deployment in domestic and foreign markets. Any possible positive or negative consequences of such development and deployment not only affect a firm's economic reasoning, but also have a real-life impact on end users or consumers, which also raises tensions and conflicts concerning social reasoning. Hence, regulations and responsibility go hand-in-hand in such a configurational interpretation of AIs in international strategy. This directly links information management and its strategic organizing in the international marketplace, and in turn, the summary dissertation becomes a theoretical endeavor to configure an organizing framework.

3.3.1.1 Data Source

Data from three articles serve as the empirical basis for this summary dissertation. Using a combination of purposeful sampling, the study employed both a maximum variation strategy and a criterion-based strategy for data triangulation (Patton, 2002). This combined sampling approach enabled the researcher to identify distinctive and diverse variations across data sources while also uncovering common systemic patterns that emerged from multiple data sources, including interviews and archival data. The maximum variation strategy enriches understanding by capturing diverse experiences and features that reflect different countries, types of AIs, and viewpoints on risk of harm or managerial responsibilities. Simultaneously, the criterion-based strategy ensures the inclusion of data that met purposive quality-assurance criteria, particularly for secondary sources such as review research. Corroboratively, these strategies help achieve data triangulation that allows systematic integration of empirical insights derived from different types of data. This dissertation is based on three main sources of data (see Table 5).

Table 5. Overview of all data sources

Category	Data source	Purpose
	Main sources	
Research review	*N=85 empirical research on AI and international business activities (completed in 2023), providing empirical insights from 1993 to 2023.	(i) Developing an understanding of characteristics of AI systems in international business activities. (ii) Tracing managerial opportunities at the firm level and related responsibilities. (iii) Tracing managerial challenges for responsibilities across national borders. (iv) Contextualizing common good by reflecting on economic reasoning and societal imperatives to develop and deploy AI systems.
	**N=44 empirical research on AI anthropomorphisms in service settings (completed in 2024), providing empirical insights from 2014 to 2024.	(i) Tracing possible human-like features imbued into AI systems. (ii) Capturing and reinterpreting possible immediate negative outcomes due to the presence of anthropomorphic features across countries. (iii) Finding possible responsible ways to develop and deploy AI systems across markets.
Interview	N=14 interviews (7 users and 7 experts) conducted in 2021, yielding approximately 225 double-spaced typed pages of interview transcripts.	(i) Developing an in-depth understanding of AI systems and anthropomorphism. (ii) Tracing responsibilities around AI and possible reasons behind negative consequences and risk of harm. (iii) Understanding deliberate actions within firms around AI development and deployment.
Archival data	***N=210 archival media, including newspaper (122 articles), magazine (47 articles), web-based news (14 sources), blogs (20 posts), and transcripts of podcasts (7 episodes), providing insights from 2005 to 2026. Dataset completed in 2026.	(i) Identifying strategic actions by firm-specific actors (e.g., top management) and associated accountability. (ii) Tracing the evolving discourse of AI systems in diverse institutional settings and the associated risks and responsibilities. (iii) Gathering public accounts of usage or consumption of pervasiveness and the experiential nature of AI systems. (iv) Contextualizing common good by reflecting experts' views and users' live experiences and opinions on harm and responsibility around AI systems.
	Additional sources	
Online observation	N=12 subreddits from Reddit forums (e.g., r/Artificial Intelligence: Why 'relationships' with AI aren't really relationships)	(i) Understanding naturally occurring contemporary conversations about the possible negative consequences of AI systems. (ii) Understanding different contemporary AI systems and how firms shape their functionalities.
	N=2 documentaries (The Age of A.I. on YouTube and The Social Dilemma on Netflix)	
Documents	N=2 system details (GPT-4o from OpenAI and Claude from Anthropic)	(i) Understanding deliberate actions within firms and the associated risks of contemporary AI systems. (ii) Understanding the possible negative consequences of AI systems and anthropomorphism across countries (e.g., the UK, Norway) and users (e.g., vulnerability due to age).
	N=4 reports by governmental agencies and non-governmental organizations (e.g., reports by the Norwegian Consumer Council and Internet Matters)	

Note: Constructing the dataset based on evolving inclusion and exclusion criteria. See *(i) Methodology of Article I, **(ii) Appendix A of Article III, and ***(iii) Appendix B of Article III.

First, a research review of N=85 empirical research articles with a focus on AI and international business activities and another research review of N=44 for empirical insights into AI anthropomorphisms in service settings. Second, semi-structured interviews of N=14 (users and experts), with a focus on AI and anthropomorphism (see Table 6). Conducting interviews focuses on wider and emerging aspects of AIs, anthropomorphism, and responsibilities. Two separate yet interrelated question guides were used to elicit interviewees' views and experiences (see Appendix 1). Third, archival media data of N=210 on AI development and deployment in the global marketplace. Furthermore, the dissertation also utilizes additional sources related to online observation and document analysis (see Table 4).

Table 6. Description of interview data

Pseudonym	Sex	Age	Types of experiences or expertise	Occupation	Country perspective
			Interview type: Users		
Ricky	M	24	Customer queries with a humanoid robot (robotic AIs); Automated algorithmic offerings in information search and social media (algorithmic AIs); Smart home through digital assistants (virtual AIs)	Research assistant	Anonymous /Australia
Xiang	F	36	Child education and learning by a small robot (robotic AIs); Information search and smart home through digital assistants (virtual AIs)	Doctoral candidate	China /Australia
Darya	F	37	Navigation in a car and smart home through digital assistants (algorithmic and virtual AIs); Cleaning through a robot (robotic AIs)	Postgraduate student	Iran /Australia
Shahid	M	30	Service inquiry through chatbots (virtual AIs)	Operations coordinator	Pakistan/United Arab Emirates
Cooper	M	22	Information search through digital assistants (virtual AIs); Service encounters of algorithmic recruitment (algorithmic AIs)		Australia
Mei	F	27	Service order and delivery in restaurants through humanoid robots (robotic AIs); Cleaning through robots (robotic AIs); Smart home through digital assistants (virtual AIs)	Postgraduate student	China
Roger	M	25	Service inquiries through digital assistants (virtual AIs); Mobility through autonomous driving (robotic AIs)	Postgraduate student	United States
			Interview type: Experts		
Adriana*	F	35	Data privacy consultancy (algorithmic AIs); Smart home through digital assistants (virtual AIs)	Consultant	Brazil /Australia
Josephine	F	26	AI ethics and online gaming (robotic and virtual AIs)	Early career researcher	France /Germany
Ibrahim	M	37	AI and service marketing (virtual AIs)	Early career researcher	Singapore /Australia

Pseudonym	Sex	Age	Types of experiences or expertise	Occupation	Country perspective
Jacinta*	F	25	Developing AI systems and tech philosophy (robotic and algorithmic AIs); Smart home through digital assistants (virtual AIs)	Tech philosopher	Mexico
Harry**	M	54	Developing AI systems (avatar) with various service settings (virtual AIs)	Professor, AI scientist	The UK /New Zealand
Aryan**	M	26	Developing AI systems (prediction model) in healthcare service (algorithmic AIs)	Company owner/founder	Germany
Pramarn	M	59	Healthcare service for the elderly through humanoid robots (robotic AIs) and automated algorithmic offerings (algorithmic AIs)	Professor, Tech philosopher	Thailand

Note: Shared insights as (i)* a user, as well as an expert, and (ii)** a firm-specific actor, as well as an expert.

Consequently, the question presented in this summery dissertation does not necessarily mirror the original research questions of each article. Instead, it reflects a higher-order framing that answers an overarching research question, extending beyond the sum of the individual contributions. Put simply, it moves beyond the individual contexts and original findings of each article to mold their insights into a unified perspective that creates new meaning. However, it also accommodates the researcher's own reflection and personal observations that may not be reported in any of the articles. It further integrates both objective factual reality and subjectivity, including creativity. However, it is always anchored to theory and bounded in the context of the common good, and the analysis revolves around responsible development and the deployment of AIs in international markets.

3.3.1.2 Analysis

The goal of the analysis was to represent an approximate reality through systems interpretation (Arbnor & Bjerke, 2009). The unit of analysis was the organizing facets of AI used in development and deployment to achieve responsiveness that integrates reasoning on both economic and social goals. The analysis focused on managerial actions that consciously enable the organization of these facets across borders. Analytical processes were iterative, which allowed theory and data interpretation to evolve together by using abduction (Dubois & Gadde, 2002). Accompanied by abductive reasoning, theorization emerged from empirical observation and creativity to balance deep data engagement with imaginative theorizing (Köhler et al., 2026). Collectively, the analysis followed a recursive three-step process: 1) identifying through seeing, 2) mapping through challenging, and 3) abstracting through articulating.

Identifying through seeing. The first step involved identifying system components through a process of seeing (Köhler et al., 2026). For seeing, both lower and higher

magnifying levels were used to identify system components (Arbnor & Bjerke, 2009). This step identified core components of managing AI that arose from different contemporary forms (e.g., robotic, virtual, and algorithmic) and their interconnected facets, as well as affordances that emerged from AIs. This step focused on the whole and its relationships, and not on isolated elements (Arbnor & Bjerke, 2009). Key issues involved how facets of AI operate as parts of a larger system involving managers and users across borders. Identifying considered patterns, contextual influences, as well as the open, dynamic, and adaptive nature of AIs. Understanding thus required attention to environmental complexity and interdependencies among facets of AI. However, the identifying step was also theoretically anchored to develop initial constructs (Dubois & Gadde, 2002). Throughout component identification, the analysis moved back and forth between already theoretically driven conceptualization and evolving system interpretation. This step uncovered hidden or missing patterns related to trusteeship and responsiveness across borders. Emerging constructs were continuously anchored to existing theory.

Mapping through challenging. The second step involved mapping to interpret system relations. It challenged existing theoretical assumptions through engaging with problematization (Alvesson & Sandberg, 2011), as presented in the theoretical background section. This step examined how components and constructs relate to one another to develop possible or alternative explanations for underlying mechanisms (Arbnor & Bjerke, 2009). Emerging relationships were then confronted with surprising observations at the intersection of various forms of contemporary AIs and their evolving characteristics. They were confronted with managerial opportunities, users' (misplaced) trustworthiness, and possible negative consequences of AIs, regardless of national borders. Arising from both managerial and users' actions, anomalies in systems interpretation led to doubt about the explanatory utility of existing theories about responsiveness in relation to contemporary AIs, as an evolving frontier of organizing capabilities at the firm level. This generative doubt thus produced a state of 'not knowing', subsequently motivating a search for alternative understanding for theoretical explanations (Köhler et al., 2026).

This search process for understanding and interpreting strategic organizing of AIs in the context of commod good has evolved over time. The researcher iterated through possible explanations, engaged with more data and analysis, and explored literature and data further to reach a series of conceptual leaps to construct relationships. In addition to systematic engagement with data, this step moved beyond data to freely explore possibilities and hunches, activating a dual mode of analytical thinking in abduction (Köhler et al., 2026). One is rational control to systematically observe and deliberately pursue objective understanding, and the other is irrational free-play to

imagine possibilities and follow hunches to pursue subjective explanation in the context of common good. The analysis also compared common patterns and outliers across articles and existing theory, and reflected the absence of elements to identify shared connectors that may shape these relations, as per conceptualization. It connected a theoretical understanding beyond that employed in the included three articles, which elevated context-specific findings into broader abstracted conceptual insights about managing AIs across borders, and formed a unified perspective that yields emergent meaning.

Abstracting through articulating. Finally, the refined components and relationships were integrated into an explanatory model through abstraction (Bygstad et al., 2016; Furnari et al., 2021). At this stage, components and relationships were named and refined to a level that is understandable to improve conceptual clarity and coherence. This stage shifted the analysis from descriptive detail to a higher level of conceptual expression through writing, visualizing, and modeling (Arbnor & Bjerke, 2009). Through articulating (Köhler et al., 2026), the purpose was to transform theoretical observations and preliminary interpretations into clear and coherent abstracted framing for presentation and communications. Ultimately, the goal was to generate a plausible explanation that enhances understanding of strategic organizing AIs for responsiveness across borders to integrate both economic and social imperatives. This step thus ensured that abstract concepts reflect underlying system logic, including interdependencies among facets of AI that intersect with managerial and user actions. Through this iterative refinement, mapped components moved from loosely connected insights toward a coherent conceptual language. This synthesis translated abstracted insights into a model of a processual management system that serves as the mainframe of the emerging Globally Responsive AI Organizing (GRAO) framework. The model development involved assembling components into a configuration that reflects their system-level interactions. In this step, the analysis also drew on metaphors and conceptual illustrations to make the emerging system more intuitive (Arbnor & Bjerke, 2009). In doing so, the GRAO framework becomes both a mode of explanation and a practical guide for relevant stakeholders.

3.3.2 Article I: Directed Content Analysis to Understand AI Systems from a Socio-technical Perspective

3.3.2.1 Research Design

The article employs directed qualitative content analysis as its research design. Such a design is particularly suitable when existing knowledge on a phenomenon is

limited. This operative extends theoretical understanding, and systematically analyzes textual data from prior empirical studies. Drawing on the systems view, it emphasizes reflexivity and context-specific interpretation, rather than rigid methodological templates.

3.3.2.2 Data Source

In line with systems views, the article constructed a dataset with already published secondary information. The dataset consists of N=85 empirical research articles on AIs published across multiple disciplines, including international business, management, information systems, computer science, and engineering. Articles were selected through purposive sampling based on relevance to AI utilization in international business operations and sustainable development. The time frame spans from 1950 to 2023 to capture both foundational and contemporary insights.

3.3.2.3 Analysis

The analysis involved iterative intra- and inter-textual comparisons between theory and data. These emerging insights are directed toward responsible consumption and production. It explained emerging themes through constructed key characteristics of AIs—autonomy, learning, and combinative. The process included coding empirical findings, identifying patterns, and relating them to socio-technical perspectives to construct a conceptual understanding of responsibilities around AIs in international business management contexts.

3.3.3 Article II: A Contextualized Explanation Case Study to Explain the Dark Side of AI Systems

3.3.3.1 Research Design

The article adopted a qualitative single contextualized explanation case study design. It explained how anthropomorphic features (characteristics) of AIs lead to misplaced trustworthiness among consumers or end-users. The empirical context was service provision. The unit of analysis was how anthropomorphic features are utilized in contemporary AIs and how consumers respond to them, which leads to misplaced trustworthiness.

3.3.3.2 Data Source

Primary data were constructed through N=14 semi-structured qualitative interviews with two groups. There were two types of interview. One targeted experts in AIs, either in industry or academia, or both. The other type targeted experienced users or consumers who regularly interact with AIs. Participants differed in gender, nationality, and experiences/actual services. The interviews were conducted via Teams or Zoom or in person between May and June 2021, lasting 20–70 minutes.

3.3.3.3 Analysis

The article applied a configurational theorizing approach, using scoping, linking, and naming techniques. Through abductive reasoning, the analysis combined insights from the literature and interview data to identify anthropomorphic features. It then constructed higher-order interactions and uncovered interrelated patterns explaining how these features lead to misplaced trustworthiness.

3.3.4 Article III: A Mechanism-based Explanation Case Study to Understand the Risks of Harm of AI Systems

3.3.4.1 Research Design

This article employed a single mechanism-based explanation case study design, which can be seen as a new variant in case-based research. It engaged a case of risk of harm that emerged from AI anthropomorphism in the context of service tasks. However, the 'case' in this study is neither a firm nor a predetermined analytical frame, but rather the case that emerged from contextualized data during the analysis.

3.3.4.2 Data Source

This article used three sources of qualitative data: interviewing, review research, and archival media. It employed purposive sampling with a maximum variation strategy to construct information-rich insights into the given context (which focuses on service tasks), and to capture diversity in anthropomorphic reasoning and cues. The study conducted 14 semi-structured interviews with users and experts via Teams, Zoom, and face-to-face between May and June 2021, each lasting 20 to 70 minutes. Additionally, a review research dataset was constructed from empirical studies from several sources in early 2024. The dataset construed N=44 studies that contain only empirical elements, including peer-reviewed conference proceedings and academic

journals published in multiple disciplines. It provided insights into firm-specific actors, users, and experts from over 30 countries spanning from 2014 to 2024. Furthermore, the article constructed an archival media dataset to capture live experiences and public discourse surrounding AI anthropomorphism, harm, and well-being. The study further included text-based records from various publicly available sources, and the dataset was completed in June 2026, including N=210 records that consist of five categories: newspaper (122 articles), magazine (47 articles), web-based news (14 sources), blogs (20 posts), and transcripts of podcasts (7 episodes). It provided insight into users and experts from over 15 countries spanning from 2005 to 2026. In addition to these three data sources, further exploration was made based on observations on online forums and documentaries, and consultation with document analysis, including system details and reports.

3.3.4.3 Analysis

This paper does not follow any established analytical templates or sequential multimethod analysis. Instead, it took a three-step approach to analyze all three data types simultaneously, using abductive reasoning in casing. The recursive three steps in casing include the identification of key entities of risk of harm under AI anthropomorphisms, the identification of key mechanisms and their immediate outcomes, and explaining via abstraction. It identified main anthropomorphic features across the datasets to explain the risks of harm under AI anthropomorphism.

3.4 Quality Assessment

Qualitative research can be done in many ways (see, Alvesson & Gabriel, 2013; Avital et al., 2017; Braun & Clarke, 2006; Bygstad et al., 2016; Furnari et al., 2021; Greckhamer et al., 2018; Köhler et al., 2026; Silverman, 1998; Welch et al., 2022, for some examples), along with its quality assessment (Bonache, 2021; Lincoln & Guba, 1985; Miles et al., 2014). Existing methodological discourse cautions against proceduralism, arguing that there is no gold standard to assess (Braun & Clarke, 2025), and also how not to judge the quality of qualitative research (Busetto et al., 2020; Welch & Piekkari, 2017). However, this dissertation points to its quality assessment through reflexive openness and methodological congruence (Braun & Clarke, 2025; Welch & Piekkari, 2017). Readers may interpret these actions through analytical view and actors view (Arbnor & Bjerke, 2009), which may relate them to rigor and credibility, respectively.

Firstly, reflexive openness emphasizes transparency about the researcher's actions (Braun & Clarke, 2025), including underlying theoretical assumptions, research

objectives, methods of producing evidence, analytical procedures, researcher positionality, and subjectivity (Arbnor & Bjerke, 2009). Quality assessment through reflexive openness differs somewhat from established standards used for qualitative research, whose ultimate assumptions are rooted either in interpretivism or positivism paradigms (e.g., Lincoln & Guba, 1985; Yin, 2014). In contrast, reflexivity-based quality criteria are used in this dissertation, in which the assessment centers on the researcher's conscious actions throughout the journey and ability to explain them. These abilities may need the researcher to reflect on the interactions in a given field, ultimate assumptions, theoretical preconceptions, disciplined imagination, and contextuality that influence interpretations and findings (Köhler et al., 2026; Welch & Piekkari, 2017). This dissertation maintains reflexive openness by clearly outlining its paradigmatic stance, methodological view, operative methods, and context dependencies. This transparency shows how the researcher makes sense of and interprets data, information, and insights within the empirical reality of knowledge construction. It also acknowledges how one's own assumptions and scientific ideals may influence the research process, emphasizing issues of methodological congruence or integrity.

Secondly, methodological congruence emphasizes integrity, and whether different elements of research fit together (Braun & Clarke, 2025). It mainly focuses on the alignment between underlying philosophical assumptions, the research questions, the research design, methods, the treatment of data, and so on. Assessing congruence involves knowingness as a researcher who strives to own his or her perspectives, both theoretical and personal. As articulated earlier, this research is built on a clearly articulated system-oriented paradigmatic reflection, enabling deliberation and thoughtful decision making. However, scholarly feedback and internal dialogue play a paramount role in maintaining integrity in research practice and reporting. Such a feedback loop occurred both formally and informally in this work, at the home university and beyond. The dissertation has been strengthened through ongoing discussions with co-author(s), colleagues, and senior researchers on its outputs (the summary dissertation and three articles). These conversations helped with conceptualization, refined emerging ideas, and improved clarity. It also enabled a coherence of the methodological and theoretical choices. However, the formal and institutionalized review process was also helpful, with peer review within the scientific community, specifically in IB and IS disciplines. Each output of this dissertation benefited from constructive criticism and scrutiny through journal review, conferences, doctoral colloquia, and research seminars. Feedback from reviewers, opponents, and participants was carefully assessed, and relevant feedback was sincerely incorporated to make informed revisions to the study design, theory, conceptualization, and ideation, reflecting on contextuality. These conscious and

deliberate actions enhanced the quality of this dissertation. However, like any scientific work, this dissertation has some limitations.

3.5 Limitations

The systems-oriented methodological view (Arbnor & Bjerke, 2009) posits that knowledge-seeking efforts are subject to continuous reflection, invocation, and development. There may always be inherent limitations in closing a particular scientific effort, and this dissertation is no exception. However, this sub-chapter focuses on limitations from a theoretical and methodological point of departure, highlighting each research output.

Although *the summary dissertation* constructed a novel framework for globally responsive AI organizing based on diverse data sources and reflecting the doctoral journey, it has two limitations. First, there are limited primary insights directly drawn from firm-specific actors (e.g., managers) through interviews, whereas the constructed archival media dataset and review research provided a more nuanced understanding. There is also a lack of a clearly defined sample of managers responsible for managing AI across national borders, particularly in regard to development and deployment. Therefore, the dissertation would have benefited largely from having additional data from interviews with managers. A clear purposive sampling would be top and middle managers who are involved in the development and deployment of AIs across national borders. Managers would further be involved in either a large multinational or a small and medium-sized firm in the traditional sense. Alternatively, the size of the firm may be defined by an evaluation of its financial scale, usage, valuation, and organizational structure. Second, the summary dissertation brought one entirely new theoretical perspective from IB, namely, the loose coupling theory (Nambisan & Luo, 2021) that had not been utilized in any of the included articles. It was a necessary element to make sense of the whole (the summary dissertation), which is greater than the sum of its parts (Articles I-III). However, each article and the summary dissertation that make up the whole dissertation are aligned with a socio-technical systems view and its underlying facet of the AI frontier. For further development, the dissertation would have benefited greatly from building a more coherent theoretical integration by using the loose coupling in one of the included articles. Such inherited limitations invite readers to interpret the constructed framework as a proposed organizing logic, which is restricted by the theoretical and methodological issues noted above.

Article I is intriguing for its theoretical framing of managing AI in international business activities, which is based on a socio-technical system perspective from IS.

However, the dissertation would have benefited if the article had integrated the assumptions of the new theory brought into the summary dissertation (the loose coupling system theory), and combined them. Additionally, adding a summary table of all of the reviewed articles would enhance the readability and the emerging insights in the current discourse.

For *Article II*, current interview data provided a nuanced understanding of the developing issue of misplaced trustworthiness. Nevertheless, the article may have two limitations. Firstly, the sample size was relatively smaller, and the interviewees represented various countries and sociocultural contexts. Although the current data may provide diverse understandings of lived experiences and expert opinions across national borders, it may also limit coherence. Therefore, employing purposive sampling with a homogeneous approach (Patton, 2002), interviewing a specific AI-related subgroup can yield detailed insights into a targeted subgroup. For example, participants may include users with different age groups and genders from the Nordic region who use only virtual AIs, powered by generative pre-trained transformer (e.g., large language models). This more focused user group, yet a larger sample size would have been beneficial for further development. Secondly, theoretical and empirical analyses do not explicitly deal with international dimensions such as the convergence of consumption across national borders. Of course, the phenomenon of developing misplaced trustworthiness is not necessarily confined to a specific national border. However, institutional differences (i.e., cultural expectations) may vary, which may shape tendencies in response to various levels of institutional features in AIs. An explicit focus on institutions in Article II would have benefited the wider dissertation.

Article III theorized mechanisms of risk of harm under AI anthropomorphisms that provide a novel understanding by using different data, involving both primary and secondary sources. However, the article may have two limitations on the data. First, the interviews were conducted in 2021, before the emergence of the widely popular AIs enabled by large language models. Accordingly, interviewing users with a larger sample size would have been beneficial for further development. Second, as part of the secondary sources, it constructed a dataset based on existing scholarly published empirical research and innovatively integrated insights into the analysis and reporting of findings. However, it is not an established approach. For future development, the article would benefit from isolating and treating them as part of the literature review for conceptualization, and reflecting them in the discussion chapter against the findings.

4 SUMMARY OF DISSERTATION ARTICLES

4.1 Article I: Understanding AI Systems from a Socio-technical Systems perspective

Hasan, R., & Ojala, A. (2024). Managing artificial intelligence in international business: Toward a research agenda on sustainable production and consumption. Thunderbird International Business Review, 66(2), 151-170.

Keywords: Artificial intelligence (AI), global sustainable development, international business, international expansion, sociotechnical theory

4.1.1 Interesting Issues and Perspectives

Article I aimed to construct characteristics of AIs to manage them in cross-border activities. It is an empirically informed conceptual study that embodies a socio-technical systems perspective on the AI frontier. Even though research on AIs is not a recent phenomenon in IB, the article engages with whether contemporary AIs will enable or hinder responsible consumption and production worldwide. Engagement is aligned with the Fifth Industrial Revolution, which emphasizes advanced technologies (including AIs) to support human-centric sustainable development. The article is directly related to managerial actions within diverse institutional environments. It also implies a need for responsible development and deployment of AIs in international business activities.

4.1.2 Conceptualization

Existing international strategy discourse in IB timely points toward AIs at the intersection of global sustainable development. However, it provides only limited knowledge of AIs' characteristics and how managers can leverage them in international business activities. Therefore, this knowledge gap is problematic for managing AI across borders that intersect with sustainable development. Without effective management strategies, policies, and institutions, there can be opposite, deleterious effects that hinder responsible consumption and production worldwide. In response, this article drew firm-level management perspectives that embraced a socio-technical systems view of the AI frontier, aiming to integrate social and economic goals. It employed a qualitative content analysis of published empirical research on AIs.

4.1.3 Key Findings

The article constructed three characteristics of AIs to underpin their facets, namely autonomy, learning, and combinative (see Figure 8). First, autonomy relates to an increasing capacity to act without human intervention and behave without human knowledge. International managers deploy AI-driven automation to inform decision-making and manage global value chains and international operations. However, AI-driven autonomy also raises concerns about control problems in internationalization decision-making, the complexity of attributing responsibility and liability, and emerging international inequalities.

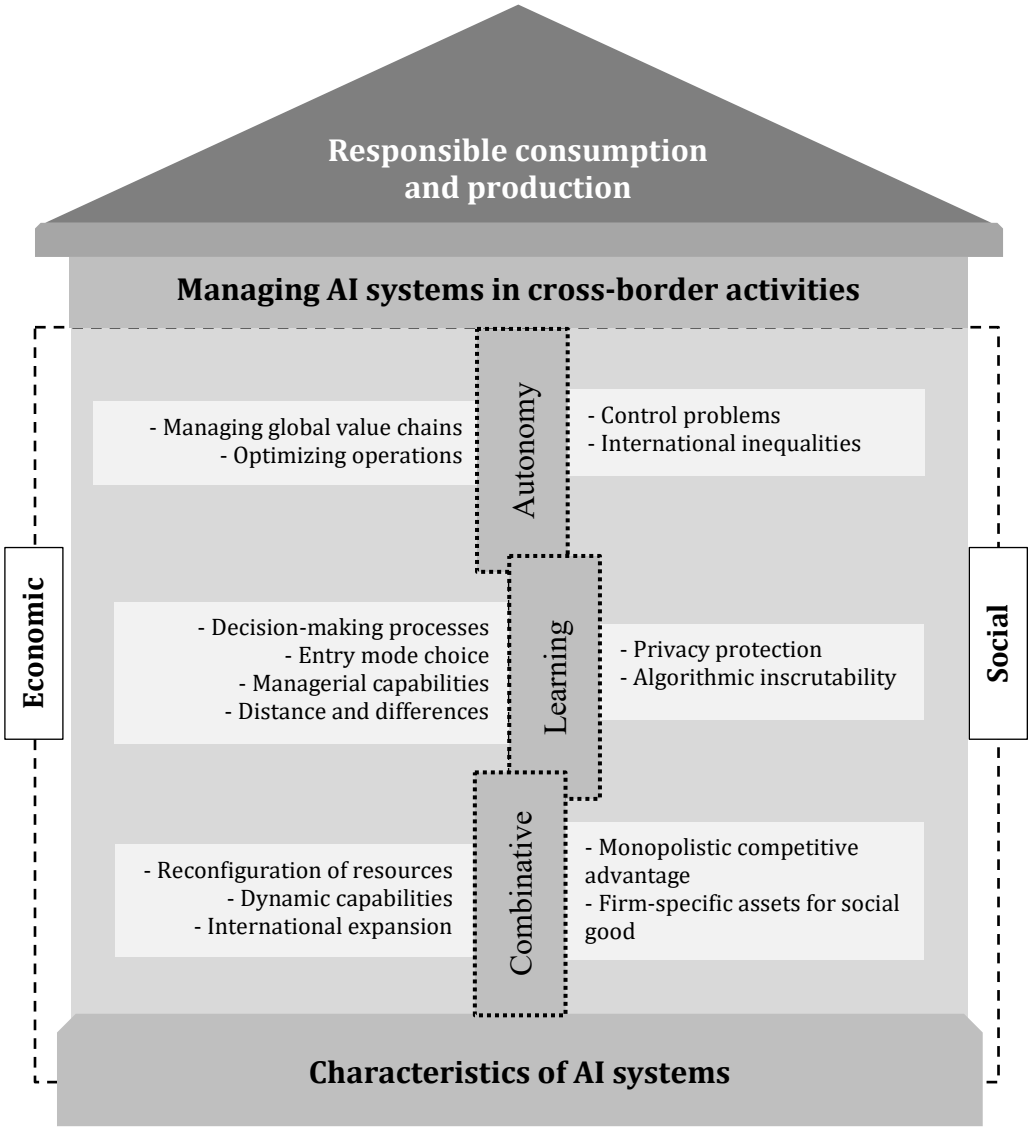


Figure 8. Managing AIs' characteristics in cross-border activities

Second, learning relates to machines' ability to inductively improve automatically through data and experience. Managers can use AI-driven learning in internationalization decision-making processes and to measure the performance of entry-mode choices. AI-based learning helps to develop managerial capabilities through gaining significant knowledge about distances and differences. However, firms' extensive use of AI-based learning across borders raises questions of potential privacy violations and algorithmic inscrutability in decision-making. Third, combinative involves a hybrid interaction between humans and machines. Managers utilize such characteristics to reconfigure resources and develop dynamic capabilities. Therefore, combinative enables bundling FSAs for social good, which facilitates scaling a solution across borders and international expansion of firms. However, AI-enabled FSAs also create tensions over firms' monopolistic competitive advantage.

4.1.4 Contributions

Article I contributed to the growing research on AIs in IB in two main ways. First, it constructed three characteristics of AIs and showed how managers can utilize those characteristics in international business activities. Second, it proposed some critical future research directions, aiming to guide IB research toward integrating economic and social goals to facilitate responsible production and consumption across global markets.

4.2 Article II: Explaining the Dark Side of AI Systems

Hasan, R., Ojala, A., Quach, S., Thaichon, P., & Weaven, S. (2025). The dark side of AI anthropomorphism: A case of misplaced trustworthiness in service provisions. Proceedings of the 58th Hawaii International Conference on System Sciences. https://hdl.handle.net/10125/109530

Keywords: Artificial intelligence, service, anthropomorphism, misplaced trustworthiness, AI ethics

4.1.5 Interesting Issues and Perspectives

Article II focused on the dark side of AI anthropomorphism and problematizes trustworthiness. It is an empirical study that adopts a socio-technical systems perspective and employs a case study design. It explained a mechanism of AI

anthropomorphism that enables consumers to develop misplaced trustworthiness in AIs during service provision.

4.1.6 Conceptualization

AI anthropomorphism enables firms to facilitate favorable human-machine interaction. However, contemporary AIs that have been developed with anthropomorphic features have a dark side that negatively impacts service consumption and trustworthiness. Such concerns (mis)align with the AI Act, which is designed to mitigate risks. However, systematic investigations that intersect anthropomorphic features and misplaced trustworthiness are limited. This is worrisome, as managers may cause significant harm to users through deceptive practices and the exploitation of vulnerabilities. Against such a backdrop, this article took a user perspective and problematizes trustworthiness in consumption to protect user well-being. It employed a qualitative single contextualized explanation case study design.

4.1.7 Key Findings

The article constructed a framework that explains the dark side of AI anthropomorphism in service provision. It theorizes that anthropomorphic features imbued in AIs enable users to develop misplaced trustworthiness. Anthropomorphic features broadly divide into two categories: physical (appearance, embodiment) and behavioral (interactivity, autonomy). These characteristics distort users' perceptions and expectations, and create illusions of competence or empathy, thereby leading to misplaced trustworthiness. They enable misplaced cognitive and affective trustworthiness. However, misplaced trustworthiness is amplified by machine-related deceptive techniques (hidden or superficial cues) and user-related vulnerabilities (age, limited cognitive capacity). These factors lead to a risk of harm such as privacy invasion, emotional dependency, and reduced autonomy.

4.1.8 Contributions

Article II made three significant contributions to IS. First, it constructs a framework to explain how machine- and user-related factors in anthropomorphic design features can develop misplaced trustworthiness. Second, it outlined practical implications for managers and policymakers to mitigate risks that align with the AI Act. Finally, it proposed pressing future research opportunities on the responsible development and deployment of AIs that emphasize transparency and institutions.

4.2 Article III: Explaining the Risks of Harm of AI Systems

Hasan, R. (2026). Artificial intelligence and anthropomorphism: A case of risk of harm on the global stage. Unpublished manuscript.

Keywords: Artificial intelligence, AI anthropomorphism, risk of harm, affordance, deception, manipulation, the AI Act

4.2.1 Interesting Issues and Perspectives

Article III engages with AI anthropomorphism and raises the question of whether it poses a risk of harm to users, and if so, how. It shows that computational advancement enables problematic consumption convergence patterns and urges the need for policies to protect users across borders. As an empirical study that builds on international strategy and information management discourse, it explains the mechanism by which the risk of harm emerges under AI anthropomorphism.

4.2.2 Conceptualization

AIs increasingly influence consumption patterns in service provision that exhibit humanlike characteristics, such as appearance or autonomy. Ingesting such anthropomorphic features enables successful human-AI interaction and serves as a source of competitiveness. However, AI anthropomorphism also leads users to misidentify the capabilities and limitations of AIs, thereby posing a risk of harm. Such risks align with the unacceptable risk of deception and manipulation in the transnational regulatory institution of the AI Act. Despite the inherited risk, managers are increasingly developing and deploying systems with anthropomorphic features. However, the current literature provides a limited understanding of AI anthropomorphism and its associated risks of harm. Such a vacuum is problematic as AIs are equipped with hidden capabilities, and rendered anthropomorphism are rising in the global marketplace. Therefore, research on AI anthropomorphism and the associated risk of harm is paramount to safeguard users' well-being. It employs a single mechanism-based explanation case study.

4.2.3 Key Findings

The article theorizes that irresponsible AI anthropomorphism poses a risk of harm through deception and manipulation. It constructed three facets of AI regarding anthropomorphism, namely (a) appearance, (b) interactivity, and (c) autonomy, to

explain the mechanism that leads to a risk of harm. These facets trigger user vulnerability through the affordances that they create. Accordingly, this article theorizes that the interplay between these facets and affordance leads to the risk of harm via deception and manipulation.

4.2.4 Contributions

Article III contributes to the existing conversation in two significant ways. First, it constructs key facets of AI regarding anthropomorphism and explains how they create risks of harm through deception and manipulation in information management discourse. Second, it suggests that consumption convergence enabled by AIs may generate systemic vulnerabilities for users, leading to harmful outcomes through anthropomorphism that may propagate across national borders. Besides contributing to theoretical conversations, this study has relevant managerial and regulatory implications. Overall, this article promotes AI-driven responsible production and consumption patterns to achieve global sustainable development.

5 DISCUSSION AND CONCLUSION

This chapter presents the dissertation's knowledge contributions to theory, practice, regulations, and future research in the context of the common good. Within such a context, the explanations may be understood as a boundary spanning of managerial actions within firms involved in international business activities. Within firms, understandings are bound by external and internal boundaries. Externally, the explanations embody formal and informal institutions (i.e., culture, norms, regulations), and internally, they relate to internationalization activities. However, the explanations also embody the socio-technical nature of organizing, with a dual emphasis on economic reasoning and societal well-being, emphasizing managerial capabilities in systems and processual activities. By doing so, this chapter explains understandings of knowledge bases in relation to the central research question, thereby advancing the understanding of managing AI across borders.

The chapter concludes its discussion by presenting four key knowledge bases. First, it explains a constructed system-oriented organizing framework for *theoretical contributions*. The framework theorizes whole systemic mechanisms as a central outcome, which is framed as ongoing capabilities. Doing so makes a novel theoretical contribution in both IB and IS. Accordingly, this dissertation contributes to two classical streams of literature within these two disciplines. In IB, it contributes to strategic organizing literature in international strategy that focuses on worldwide effectiveness (Bartlett & Ghoshal, 1989; Birkinshaw, 2022; Nambisan & Luo, 2021). In IS, it contributes to the organization of technological artifacts in information management literature (Berente et al., 2021; Zheng & Jarvenpaa, 2021). Second, the theoretical framework is action-oriented, and the necessary stewardship for its application underscores its *managerial implications*. Third, it explains an emerging accountability in regulations in international markets for *regulatory implications*. Finally, it concludes with *opportunities for future research* based on the constructed theoretical framework, with a core focus on the methodology of complementarity (Arbnor & Bjerke, 2009) and some topical illustration. The chapter closes with concluding remarks.

5.1 Globally Responsive AI Organizing (GRAO): A Proposed Theoretical Framework

The proposed theoretical framework may be best named as the *Globally Responsive AI Organizing (GRAO)*, and is hereafter referred to as *the GRAO framework*. Grão, is a Portuguese word meaning 'seed', and metaphorically it reflects deliberate, generative actions of planting an organizing seed in the hearts of human actors and in day-to-

day operations, so that responsible practices and outcomes can grow over time to integrate economic reasoning with societal well-being. The framework uses systems language to explain processual activities to underpin ongoing managerial capabilities for responsiveness. The GRAO framework puts trusteeship covenants at the heart of managerial actions, embedded as a socio-technical core to constitute how AIs are developed and deployed for scaling in internationalization activities. In its theorizing, responsibility is seen as a core of organizing activities rather than an external governance tool or an abstract managerial value. For responsiveness across borders, the theorization proposes three interrelated covenants that intentionally prioritize proactiveness in (a) *addressing institutions*, (b) *blocking harmful means*, and (c) *ensuring accountability*. Managers may want to utilize them for governing AIs in international business activities, and can conceive them collectively as a shared value that institutes proactive commitments to embed moral, regulatory, and societal considerations directly into the development and deployment of AIs.

Trusteeship covenants are inseparable from organizing the evolving and interrelated four facets of AI—(1) *autonomy*, (2) *learning,* (3) *interactivity,* and (4) *appearance*. They may serve as a central mainframe for capacity building to organize facets of AI and enact responsible practices for configuring FSAs. However, facets of AI are shaped not only by technical optimization, but also by values related to transparency, human vulnerability, and oversight. As these facets of AI evolve, trusteeship may thus become the purpose of reflection through system behavior and human–AI interaction. Instituting trusteeship to organize facets positively influences the growing scope of affordances, enabled by users' actions. However, facets of AI are influenced by how users interact with and perceive systems. This feedback loop of affordances enhances AIs through interaction and relationships. In turn, human-AI interaction provides further data for learning that is transmitted into a firm's dynamic economic resources. Therefore, the responsiveness of AIs may require addressing institutional differences, blocking any means that pose a risk of harm, and embedding accountability. Any FSAs that involve facets of AI, such as interactivity or appearance, are emergent and socio-technical by nature. Managerial actions thus involve constant efforts to mitigate risks of harm through anthropomorphization. Actions also require mitigating potential regulatory and reputational risks. Affordance expands through consumption, as users engage with AIs and interpret their trustworthiness and satisfaction, underpinning their wellbeing. In turn, it also affects the performance of facets modulated into the responsiveness of AIs.

The GRAO framework proposes responsiveness across borders as an ongoing managerial capability, and as part of the processual outcome to underpin FSAs. It reflects managerial ability to translate embedded trusteeship across diverse institutional (e.g., cultural and regulatory) environments, while sustaining calibrated

trustworthiness, user well-being, competitiveness, and rapid regulatory responses. This capability, in turn, positions managers of internationalizing firms at the forefront of competitiveness, baking trustworthiness into their AI-enabled service systems. By proactively taking accountability for misconduct or unintended harm, managers can adjust systems, procedures, and protocols to mitigate the risk of harm. Consequently, this ensures users' well-being and satisfaction, and also strengthens regulatory compliance and enables swift adaptation, underpinning resilience of FSAs. It further enables the system to respond to changes or shocks in market-based regulatory institutions, geopolitics, pandemics, or adversity.

After briefly introducing the GRAO framework, this sub-chapter now turns to detailed explanations through three parts. The first part presents a typology of facets of AI. These facets are distinguished by the scope of their affordances and performance, including autonomy, learning, interactivity, and appearance. The second part introduces the mainframe of organizing. It brings trusteeship as a shared value to govern these facets and achieve cross-border responsiveness. The mainframe consists of three interrelated trusteeship covenants that intentionally emphasize proactiveness in addressing institutions, blocking harmful means, and ensuring accountability. Finally, the framework offers a practice-oriented explanation of stewardship, highlighting its managerial and regulatory implications. In particular, it proposes how to possibly operationalize the GRAO framework in practice.

5.2 A Typology of Evolving Facets of AI into Cross-border Activities

One of the central theses of the GRAO framework is that contemporary AIs have foundational properties or capabilities that are ever-changing in international environments. These properties can be modulated into AIs in either physical or digital environments that may provide a source of FSAs. Such system properties underscore particular facets of AI. Contemporary AIs may have distinct facets in a given context (Berente et al., 2021; Hutzschenreuter & Lämmermann, 2025; Stelmaszak et al., 2025). However, the GRAO framework proposes four facets of contemporary AIs: (1) *autonomy*, (2) *learning*, (3) *interactivity*, and (4) *appearance*, in order to provide its foundational understandings. The constructed facets of AI, namely autonomy, learning, interactivity, and appearance, may be system-agnostic and not necessarily bound to any forms of AIs (see Figure 9).

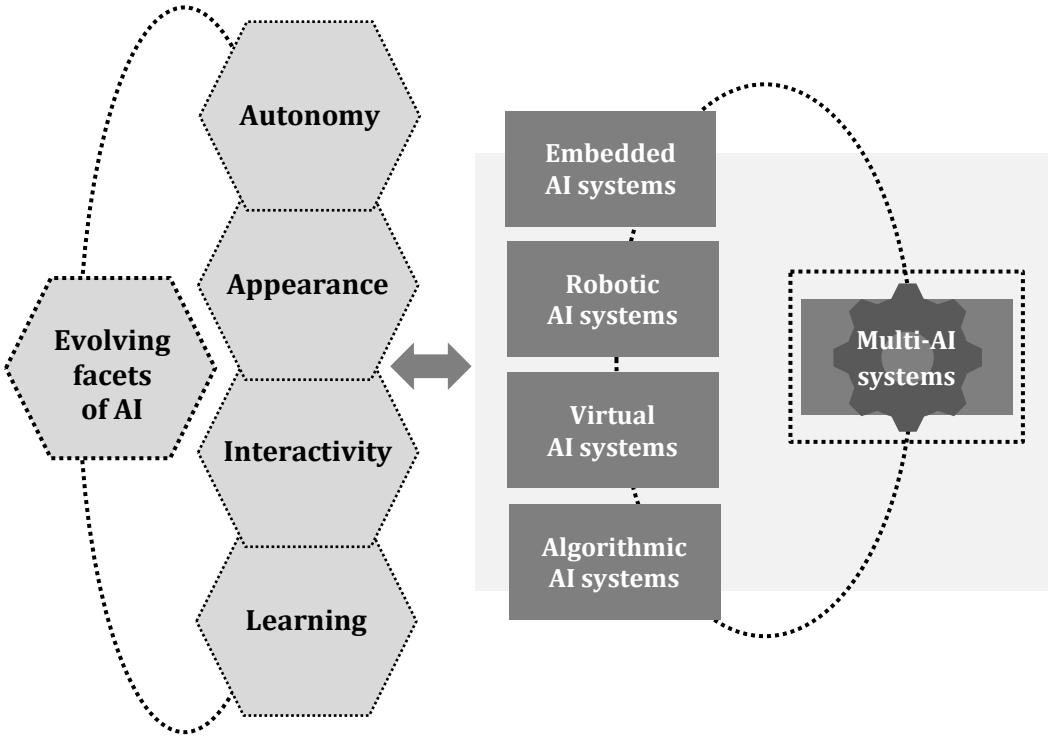


Figure 9. Evolving facets of AI modulated into its contemporary representations

5.2.1 Autonomy

Contemporary AIs pose an autonomy facet, which delineates their increasing capacity to act without human intervention and behave without human knowledge (Berente et al., 2021). Unlike past information technology artifacts powered by rule-based automation that required humans to explicitly codify and program tasks, AIs are based on machine learning that learns from data. Managers can oversee the development of AIs in a way that enables them to reprogram themselves automatically. This self-development aspect allows AIs to automate routine tasks across borders. Such a facet of autonomy is also related to perceptions of autonomy (perceived autonomy) among human users, driven by the presence of anthropomorphic cues (Gray et al., 2007). Recent IB studies suggest that such autonomous capabilities can be expanded across borders by feeding fine-grained data from diverse national contexts (Autio et al., 2021; Ciulli & Kolk, 2023). For example, deploying AI-enabled machine translation enables the reconfiguration of FSAs to automatically reduce institutional distance (i.e., language) in foreign markets, thereby facilitating international expansion (Brynjolfsson et al., 2019). The autonomy facet may allow managers to utilize autonomous machine intelligence to complement human decision-making and knowledge transfer across borders (Grant & Phene, 2022). With such proliferation in multinational settings, autonomy may further

enable FSAs through less monitoring and coordination, thereby reducing transaction costs. Thus, the autonomy facet may become the frontier of transaction-based FSAs.

5.2.2 Learning

The facet of learning has been one of the central concepts of AIs from inception (Turing, 1950). Learning refers to machines' "ability to inductively improve automatically through data and experience" (Berente et al., 2021, p. 1437). Machines can learn to recognize patterns through supervised or unsupervised learning techniques, and other large-scale advanced computational methods such as deep or reinforcement learning and neural networks. For instance, the neural network model is inspired by the human brain, with extensive interconnected units as neurons that form a vast network to recognize complex patterns. IB literature explains that automatic self-learning mechanisms are modeled using many computational elements that function in parallel and are arranged in layers (Veiga et al., 2000). However, deploying AI with self-learning capabilities offers excellent learning opportunities for managers during their internationalization process (Huo & Chaudhry, 2021). This learning facet enabled by machines may aid managerial learning in cross-border activities. Such learning is paramount for contemporary FSAs, as reflected in existing IB theories (Luostarinen, 1994; Vahlne & Johanson, 2017), and accordingly, the learning facet may thus become the frontier of asset-based FSAs.

5.2.3 Interactivity

With advances in autonomy and learning, temporary AIs are increasingly being deployed to render humanlike interaction, delineating interactivity as an emerging facet. Interactivity is a multidimensional aspect understood as a two-way process involving reciprocity, information exchange, and real-time synchronicity (Burgoon et al., 2000; Schleidgen et al., 2023). With recent advances in large language models enabled by generative pre-trained transformer, managers are delegating AIs to communicate naturally with users. For instance, users may interact with AIs by giving instructions (gesturing, typing, voice), entering dialogs (conversation style), or responding to an AI-initiated interaction. In turn, this interactivity relates to communication or conversational style, language, and voice. Managers may turn AIs into a relationship-building tool in internationalization activities. While research on interactivity facets of AI in IB activities is limited, deploying AIs with interactivity is growing in practice (Claypool, 2023; Delouya, 2023). However, relevant insights suggest that such a facet of AI influences users' decision-making and perceptions across various country-specific settings (Brendel et al., 2023; Nakanishi et al., 2019;

Zhou et al., 2022). The interactivity facet may thus become the frontier of relationship-based FSAs.

5.2.4 Appearance

As interactive artifacts, contemporary AIs can render actual or imagined morphological forms in both physical and digital environments. The concept of appearance as a facet emerged from the uncanny valley phenomenon (Mori et al., 2012). Nevertheless, it comprises visual, auditory, and behavioral cues that make AIs familiar to living beings, including issues of gender and race. Deployed contemporary AIs may resemble or show no resemblance to humans. For example, AIs with a given face, form, and body in terms of visual perspective (mechanical, humanoid, and android) represent low, medium, and high levels of human-like appearance, respectively. However, appearance may also relate to auditory cues such as a feminine voice, and behavioral cues such as an authoritative tone, which may give an impression of a specific gender and neutrality. In turn, appearance facilitates human-AI interaction to develop relationships with users under AI anthropomorphism. Although research on AIs' appearance in IB is rare, existing insights indicate high local sensitivity in user responses (Złotowski et al., 2020). Hence, with facial structure and skin color as geographical dispersal stimuli, AIs can be developed to give the impression of a resemblance to Nordic, East Asian, South Asian, or Middle Eastern countries, for example. Accordingly, the appearance facet may also become the frontier of relationship-based FSAs.

To summarize, the autonomy facet implies increasing capabilities to act independently. AIs make autonomous decisions without human intervention or even without human knowledge. The learning facet denotes the ability to learn inductively from data. The interactivity facet delineates a two-way process interaction, involving reciprocity, information exchange, and real-time synchronicity. The appearance facet implies actual or imagined morphological forms comprise visual, auditory, and behavioral cues that make AIs seemingly familiar to living beings, including aspects of gender and race. Collectively, they represent ongoing managerial efforts to expand AI scope and performance, highlighting fundamental properties or capabilities of contemporary AIs. Managers may utilize these facets through recombinants that may include asset-based, transaction-based, or relationship-based FSAs. At the intersection of these facets, AIs are intentionally developed and deployed to emulate human capabilities (the ability to act humanely), including decision-making. Therefore, these facets of AI may become a core competency and facilitate FSAs in internationalization activities. These insights introduce the first tenet of the GRAO framework and its subsequent proposition:

Proposed tenet 1 (Facets across contemporary AI) Evolving facets of AI can be system-agnostic and not necessarily bound to a specific contemporary form, including robotic, virtual, embedded, and algorithmic representations.

Proposition 1 Facets of AI represent fundamental properties that can be modulated in contemporary AI depending on intended development and deployment purposes in international business activities.

5.2.5 The Interplay of Scope of Affordances and Performance of AI Across Facets

With these facets, contemporary AIs create ever-expanding opportunities for actions, enabling managers to enhance their information technology-oriented knowledge bundles over time. The GRAO framework proposes to distinguish them based on the scope of their affordances and the performance of AIs (see Figure 10). Scope of affordance refers to the range of possible actions that become available through AIs upon deployment. It is emergent and related to what users can do with AIs, given its underlying facets. Scope of affordance emerges directly from given underlying facets embedded into systems. These facets may influence how users perceive, engage with, and rely on systems, leading to an increased scope of affordance. However, the growth can be known or unpredictable due to given anthropomorphic properties. AIs can be differentiated based on their technological properties to determine whether they render anthropomorphism. Whereas AIs are designed with minimal anthropomorphic properties to provide their internal capabilities, relatively higher anthropomorphic properties are used to facilitate user interaction.

The deployment of contemporary AIs in the global marketplace creates a range of anticipated affordances spanning cognitive-analytical to emotional-social tasks, as well as mechanical-practical tasks (Wirtz et al., 2018; Zheng & Jarvenpaa, 2021). Cognitive-analytical task affordance involves information processing and decision support, where AIs can serve as a safety component or enabler for service task functions (e.g., autonomous driving capabilities and information validation). Emotional-social task affordance involves the use of AIs to address social and relational aspects of a service (e.g., caregiving, emotional support). Mechanical-practical tasks involve using AIs to address routine-based, practical tasks (e.g., cleaning the floor, exchanging pre-defined information). However, the scope can go far beyond what managers may have initially intended during development and even

during deployment. Since the possibility of performing tasks is also dependent on the users, in addition to the given technological properties, and whether they have a minimal or strong anthropomorphic presence. In contrast (yet related), performance refers to how effectively AIs are developed, integrated, restricted, and enhanced, with underlying evolving facets of AI. AIs may be sources of FSAs that include a firm's technical assets (e.g., proprietary training data, domain-specific model architectures), managerial routines (e.g., governance structures), and relational capabilities (e.g., user trust-building, cultural adaptation). In many cases, AIs themselves (such as Grammarly's adaptive writing assistant or Volvo's Pilot Assist) are an FSA because they embed the firm's accumulated managerial knowledge and technological sophistication. However, there is a reciprocal, evolving feedback loop between the scope of affordances and the performance of AIs.

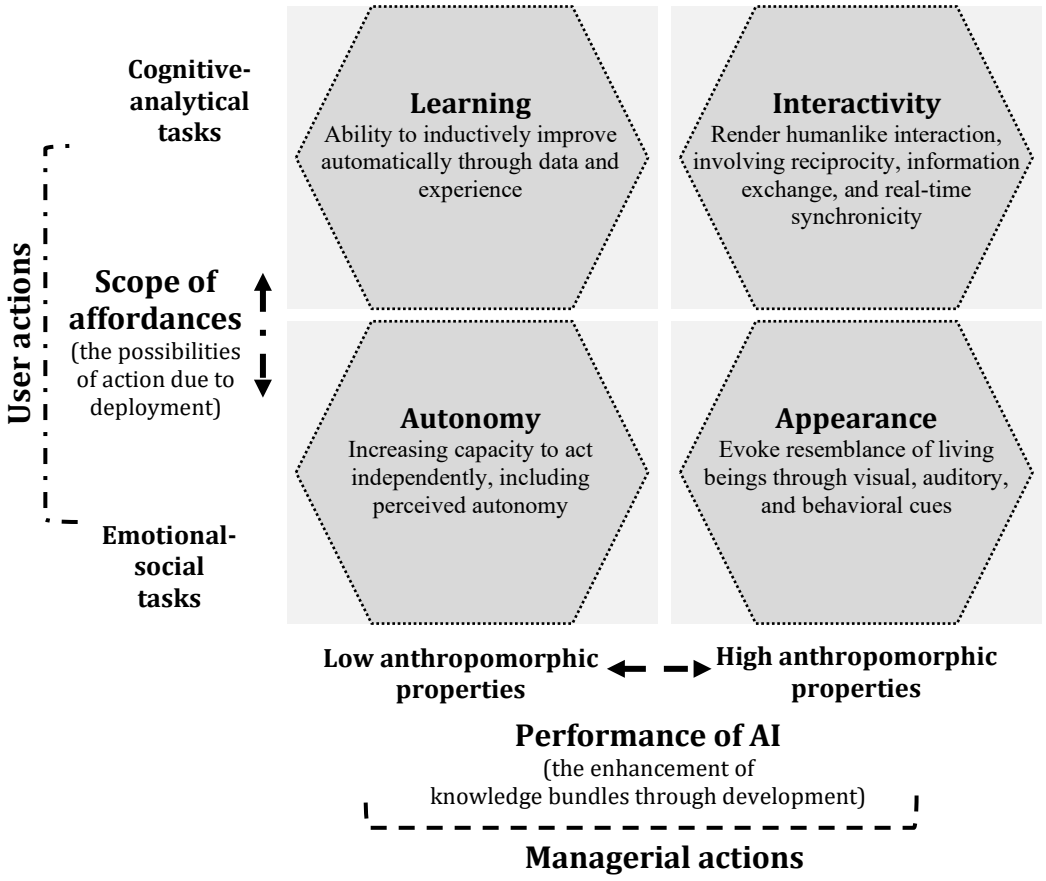


Figure 10. Organizing facets of AI on the global stage

As managers are responsible for enhancing facets (for instance, refining learning algorithms or improving interactivity), new affordances may emerge for users upon deployment. These user interactions, in turn, generate data, insights, and behavioral patterns that feed back into AIs and inform managerial decisions, thereby

strengthening or weakening performance. However, this interplay is dynamic and sometimes unpredictable. Affordances may evolve in unintended directions, amplified by boundaries of confined institutional contexts (i.e., culture) or user expectations. Positive patterns of interaction can improve calibrated trustworthiness, expand adoption, and reinforce ongoing managerial capabilities for competitiveness. Conversely, dark patterns of interaction, such as an over-reliance on autonomous features, deceptive signaling via humanlike appearance, or biased learning feedback, may degrade user well-being. In turn, the scope of affordances negatively impacts the performance of AIs that may trigger misplaced trustworthiness and expose firms to reputational or regulatory risks. The interplay between the growing scope of affordances and performance of AIs is thus both a driver of innovation and a potential source of risk. The GRAO framework proposes that managers deliberately shape these four facets of AI to expand desirable affordances, while blocking harmful ones, and so ensure that the evolution of systems aligns with user well-being, cultural expectations, and regulatory norms. Therefore, managing the aforementioned four facets of AI for responsiveness across borders is a constant process of emergence and performance. These insights propose the second tenet of the GRAO framework and following proposition:

Proposed tenet 2 (Underlying facets for organizing) Enabled by underlying facets, namely autonomy, learning, interactivity, and appearance, AI may become a source of firm-specific advantages. When facets of AI perform in conjunction with the scope of affordances, they may both drive competencies and pose risks for responsiveness across borders.

Proposition 2 Facets at the interplay of the performance of AI and scope of affordances may positively influence responsiveness or change it through posing risk for competitiveness and regulatory response, as well as for trustworthiness and user well-being.

5.3 Instituting Trusteeship Covenants for Responsiveness Across Borders

Trusteeship covenants enable the management of forehead dynamics on facets that may enable the responsiveness of AIs across borders, underpinning ongoing competencies. The GRAO framework proposes three mutual covenants as its central mainframe (see Figure 11). The trusteeship mainframe may serve as a cognitive institution (organizing logic within management culture) to enact facets of AI, transmitted into FSAs with an understanding of shared rituals, norms, and routines. Such shared values may be necessary to make contemporary AIs responsive in cross-border activities. They may serve as a glue of responsibilities for managerial actions as responsiveness relates to the ever-expanding scope of affordances and performance of AIs. They may enable regulators to promote coherent policies and actions to enact responsiveness of AIs as sources of FSAs. However, they may also serve as a guardrail even amid potential uncertainty such as geopolitical tensions and conflicts.

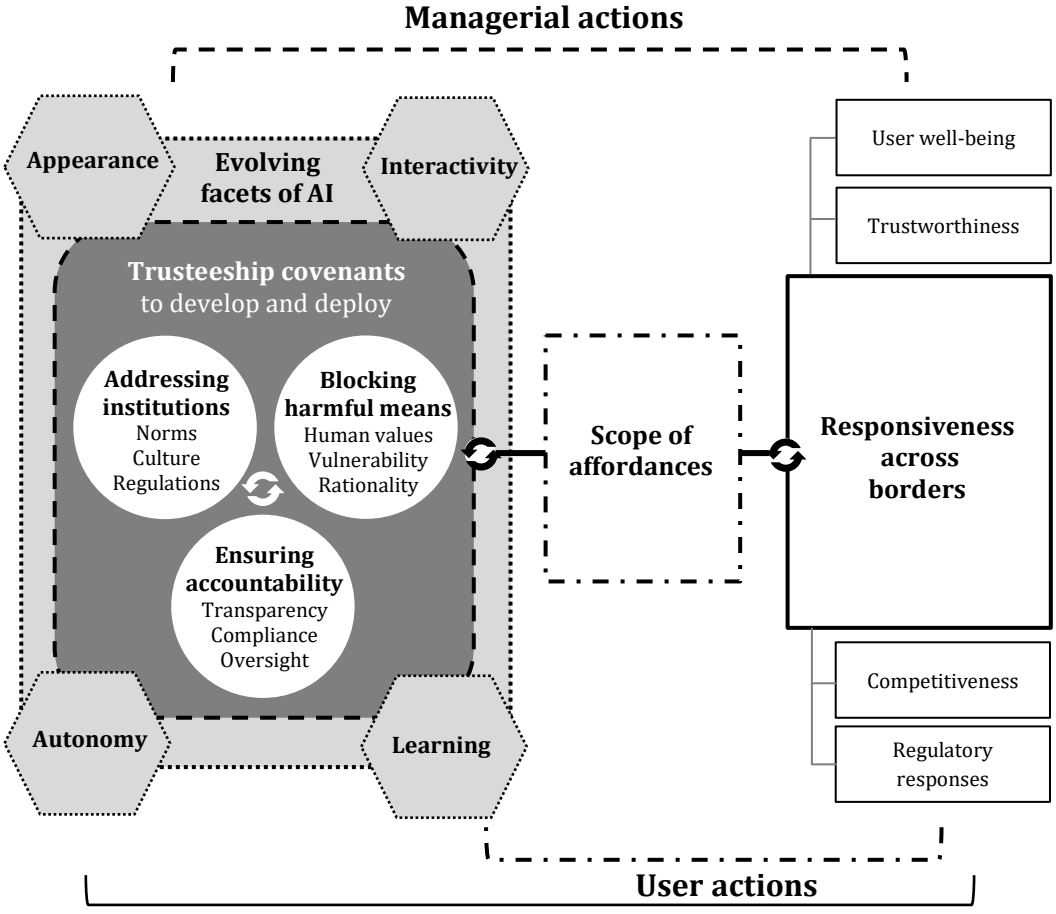


Figure 11. The processual perspective of the GRAO framework

Circumstances of uncertainty of managerial actions may arise from causal indeterminacy (e.g., bounded rationality). They may also emerge from internal or external fragmentation due to the evolving nature of facets of AIs, regulation, regionalization, and nationalization. Collectively, trusteeship covenants may institute AIs as a loosely coupled socio-technical system through coordinated actions to open, close, or separate systems that ensure responsiveness. Therefore, they share an underlying assumption of intentionality to shape a management processual activity across two theoretical lenses, namely I-R theory and LCT. However, such a globally responsive approach is distinct and may ensure social well-being that goes beyond legal compliance and economies of scale. Consequently, the trusteeship organizing perspective requires constant reflection on managerial actions to organize each facet of AI (see Table 7 for illustrative reflections). The sub-chapter now explains three core covenants as action mechanisms to develop and deploy AIs across borders while remaining responsive.

Table 7. Illustrative reflections of trusteeship covenants to organize facets

Facets of AI systems	**Covenant 1: Addressing institutions**	**Covenant 2: Blocking harmful means**	**Covenant 3: Ensuring accountability**
Autonomy	Embed institutional awareness (e.g., legal, culture) across counties to guide decision-making boundaries and swift regulatory responses. Assess the performance of autonomous FSAs (machine behavior outcomes) across borders to ensure alignment with diverse institutional expectations.	Prioritize human agency, fairness, and dignity. Provide user control options via override options, layered content, and an emergency stop, tailored to diverse expectations and sensitivities.	Integrate traceability modules, decision logics, logs, and options with decision reversibility. Use real-time monitoring and an incident response mechanism with human-in-the-loop oversights.
Learning	Integrate institutional constraints into learning, including data type limitations, dynamic consent, and the global representation of diverse institutions in the training dataset. Enable adaptation for local sensitivity, including locally run models, region-specific data-handling protocols, and adaptive learning boundaries.	Prevent bias and discrimination when learning occurs across markets, and safeguard data and physical privacy. Employ localized federated learning and continuous monitoring outcomes to detect and correct violations of human values (e.g., discriminatory patterns).	Ensure auditability of model updates and decision evolution through versioning, explainable learning paths, and human oversight. Assign clear accountability roles within firms for learning outcomes with evolving autonomy.
Interactivity	Assess how users across different institutional contexts perceive and respond to	Respect user dignity, emotional safety, and cognitive limitations to ensure calibrated trust	Ensure clarity in interaction boundaries through disclaimers or visual

Facets of AI systems	Covenant 1: Addressing institutions	Covenant 2: Blocking harmful means	Covenant 3: Ensuring accountability
	anthropomorphic cues (e.g., body language and facial expressions). Localize the degree of anthropomorphic features (e.g., tone of voice, gestures) in the user interface to align with regional norms.	that does not mislead users or distort behavior. Monitor interactions to prevent deception, manipulation, or dependencies, especially among vulnerable groups, and provide an opt-out option.	cues to signal non-human nature. Assign responsibility when anthropomorphic reasoning is used, especially when influencing trust or decision-making behavior.
Appearance	Evaluate features such as facial traits, gestures, and clothing to align with institutional norms across regions. Localize if needed to avoid expectation mismatch and misinterpretation.	Avoid deceptive signaling of the presence or absence of internal capabilities. Monitor for deception or discomfort.	Clarify design intentions to prevent user misinterpretation. Collect feedback on negative effects and misplaced trust, and conduct cross-cultural sensitivity audits to identify needed adaptations.

5.3.1 Addressing Institutions

Addressing institutions refers to the proactive alignment of facets of AI with diverse institutional settings across international markets. Diversity in institutions can be both normative (such as norms, values, and expectations) and cognitive (such as shared understanding, culture, and decision-making logics). Managers may want to oversee the development of AIs to reflect the plurality of values, beliefs, religions, and traditions across regions. They may also want to ensure that deployment respects that cultural sensitivity is essential for how users interact, communicate, and interpret contemporary AIs. For instance, anthropomorphic features for the interactivity facet, such as language and interaction styles, are designed in a way that is culturally appropriate, so as to avoid offense and cause harm. By addressing institutions by default, organizing may emphasize that facets of AI are developed and deployed to reflect local expectations, and ensure institutional embeddedness in service systems across dispersed stimuli and geographic areas. It may also enable managerial competency to align facets of AI to enable adaptabilities with given institutions, thereby addressing external fragmentation in internationalization activities. Being locally responsive yet globally responsible when addressing institutions serves as a baseline for facets in FSAs, thereby enhancing their competitiveness. However, addressing regulatory institutions is also paramount.

Human actors (e.g., managers) are entrusted with developing and deploying AIs in accordance with existing and emerging local regulations (e.g., the AI Act). Globally responsive organizing requires deliberate looseness in design to adapt to regulatory frameworks across jurisdictions. Managers would avoid a one-size-fits-all approach,

and instead allow flexibility in facets of AI within deployed institutional settings to ensure responsiveness. Put simply, managers would isolate targeted facets for deployment and make necessary modifications to adapt swiftly to evolving regulations. Such flexibility in deployment results in internationalizing firms' regulatory responses. These reflections offer the following tenet and propositions:

Proposed tenet 3 (Addressing institutions) Enacting to address institutions globally by default into the development of facets may enable responsiveness across borders in deployment, with continuous interaction between the scope of affordances and performance of AI.

Proposition 3 Addressing institutions may positively interact with the performance of AI, leading to responsiveness to market-specific sensitivities. The interaction may further intensify through (a) institutional alignment of facets (e.g., culturally appropriate appearance) and (b) scope of affordances delegated in a given market (e.g., socio-emotional tasks).

Proposition 4 Addressing institutions may enable responsiveness for remaining competitive amid rapid international regulatory responses. This relationship may be further enacted through facets (e.g., learning) as loose systems (e.g., separateness) to configure firm-specific advantages, particularly for heterogeneous (tighter vs weaker) regulated markets (e.g., data privacy laws).

5.3.2 Blocking Harmful Means

Blocking harmful means refers to the proactive prevention of harm by identifying and restricting affordances and facets that could compromise human values, exploit user vulnerabilities, or undermine human rationality. Such a compromise could lead to harmful outcomes, whether intentional or non-intentional. Means may arise from intentionally provided technical properties in facets of AI to configure products or services. Means may also arise due to the user's egocentric biases, cognitive load, tendency to anthropomorphize AIs, or to unpredictable affordances. Blocking any harmful means requires constant reflection, reckoning, and anticipating possible risks of harm. Hence, measures may need to be extended to uphold human values such as dignity, privacy, fairness, and human autonomy. Users may also be emotionally, cognitively, or socially vulnerable in their interactions with and in their

anthropomorphization of AIs. Managerial anticipation of how anthropomorphic or persuasive features in AIs might invite misplaced trustworthiness, emotional dependency, or exploitation is critical. Facets of AI modulated into contemporary forms (e.g., virtual) may enable deception and manipulation to inflict harm, such as bias amplification, continuous surveillance, and violations of data and physical privacy. Hence, the development and deployment of AIs that provide FSAs may need to be arranged so that harmful affordances do not arise as indirect means that result in harmful ends. In some cases, restricting the performance of AIs through adjusting facets to block their harmful affordance may be needed to neutralize risks of harm. Furthermore, AIs deployment may also need to preserve human rationality.

Deployed facets of AI cannot distort users' abilities to make informed and rational choices, and would prevent the use of nudging, framing, or deceptive signaling. Accordingly, blocking harmful means also includes adaptive controls over facets to prevent harm across contexts and user groups. However, this covenant has strategic outcomes. By proactively identifying and restricting affordance and facets in FSAs that could enable harm, managers demonstrate a commitment to ethical integrity. Such integrity includes upholding human values and avoiding exploitative practices that facilitates calibrated trustworthiness across borders, thereby deployment becomes responsive. However, blocking harmful means also directly contributes to user psychological safety by protecting vulnerability and rationality, which in turn leads to user well-being. Such a processual perspective of organizing AIs offers the following tenet and propositions:

Proposed tenet 4 (Blocking harmful means) Enacting to proactively block harmful means in the development and deployment of facets to configure firm-specific advantages may support responsiveness across borders, with continuous interaction between the scope of affordances and performance of AI.

Proposition 5 Blocking harmful means in facets of AI may reduce the emergence of harmful affordances that increase user psychological safety and preserve human rationality (e.g., guardrails against manipulation in interactivity, disclosure cues in autonomy), thereby achieving responsiveness for improving user well-being and calibrated trustworthiness across markets. This relationship may intensify when anthropomorphic features are high, and users are more prone to develop misplaced trustworthiness in a given market.

Proposition 6 Blocking harmful means in facets of AI may enable cross-border responsiveness for swift regulatory responses and competitiveness. This relationship may intensify when employing adaptive harm-control mechanisms in place in facets to configure firm-specific advantages, allowing the swift neutralization of risky affordances without compromising performance.

5.3.3 Ensuring Accountability

Ensuring accountability involves proactive arrangements to attribute responsibility throughout the lifecycle of contemporary AIs, necessitating continuous monitoring and adaptive intervention. It may emphasize human centrism for actions and arrangements that underpin the logic of how one organizes things, acts, and holds oneself accountable. It may also resonate with technology-related means-end rationality, which explores how actors inflict harm and jeopardize well-being, considering both intentional and non-intentional consequences. Therefore, facet-specific transparency may ensure truthfulness, enabling users to understand a system's boundaries, internal capabilities, and limitations. However, managers would pursue compliance beyond existing regulations and standards, as facets of AI are dynamic and may yield harmful outcomes that deviate from originally intended usages. Additionally, human oversight is required to ensure accountability through continuous monitoring of the performance of AIs. Managers are responsible for shaping contemporary AIs as much as they shape usage through the growing scope of affordances. Hence, they might be responsible for any negative consequences as well as positive ones. However, contemporary AIs are characterized by their emergent nature, and so too are their facets. It is somewhat challenging to predict or control such loosely coupled evolving systems. Consequently, managing AIs across borders requires ongoing, deliberate development efforts to address emergent negative issues that arise during deployment from diverse institutions. Here, managers would create a human-in-the-loop protocol with clear responsibility assignments should any harmful events or circumstances occur. These efforts may result in responsiveness across borders while balancing regulatory response and trustworthiness in international markets. Such a responsible perspective on organizing AIs offers the following tenet and propositions:

Proposed tenet 5 (Ensuring accountability) Enacting to ensure accountability by default in the development and deployment of facets of AI, where transparency, compliance, and human oversight enact traceable

actions and correction of emergent harms may lead to responsiveness across borders, necessitating continuous monitoring and adaptive intervention.

Proposition 7 Ensuring accountability through transparency relevant to facets (e.g., disclosing system capabilities and limitations in autonomy) may reduce the emergence of harmful affordances, in turn achieving responsiveness for calibrated trustworthiness and regulatory response across borders. This relationship may intensify when proactive arrangements extend beyond compliance, necessitating continuous monitoring and adaptive intervention.

Proposition 8 Ensuring accountability through human oversight with clearly assigned responsibility may improve detection and correction of emergent risks of harm in affordances, in turn, achieving responsiveness for competitiveness and user wellbeing across borders. This relationship may intensify when the performance of AI exhibits a higher level of emergence and unpredictability due to the presence of particular facets like learning, necessitating continuous monitoring and adaptive intervention.

5.4 Implications for Managers with Different Strategic Stewardships

5.4.1 Possible Implementation of the GRAO framework

Implementing the GRAO framework may be achieved through a digital tower operating model, similar to digital tower systems for managing multiple airports from a central workstation. A digital tower could be a globally integrated workspace (both physical and remote), in which managing the scope, emergence, and performance of multiple AIs happens simultaneously. It may comprise dedicated, cross-functional teams that work together for responsiveness in a fast-moving environment. In this way, relevant human employees are grouped to perform specific tasks, utilizing reusable tools and repeatable processes. Teams are small and may typically consist of five to nine people, working closely with business units that may function as a start-up accelerator. Each team is dedicated full-time to developing and deploying AIs for the given purpose or business unit. Managing these teams would be mission-specific with clear goals, and the working autonomy is governed by a central

steering center that comprises middle management. This center may function as a stewardship to facilitate development and deployment with training, oversight, expertise, and guidance. Finally, an operating committee, consisting of members of top management (executives), sets the mission for each team, reviews progress, secures and allocates the budget, and removes barriers. The operating committee would enact shared trusteeship and strategic vision in close collaboration with middle managers serving in the central steering center. Alternatively, implementation may also follow other operating models suitable for firms' structural settings.

Drawing on the GRAO framework, top managers of the operating committee and middle managers of the steering center collectively may want to enact four strategic stewardships for AIs, as illustrated in Figure 12. They may enable trusteeship organizing through simultaneously (i) *delegating the scope affordances to users, (ii) configuring firm-specific assets, (iii) governing for competitiveness, and (iv) protecting the common good*. These stewardships may offer strategic benefits for both internal and external governance, which is considered interrelated rather than seen in isolation. Managers would utilize trusteeship covenants at the core (as shared value or cognitive institutions), and evolving facets of AI are necessary components of a system that links strategic stewardships. In turn, such processual activities may lead to achieving responsiveness across borders. Depending on the institutional settings of internationalization activities, they may reconfigure relevant rolling benefits from each stewardship and recombine them as needed.

5.4.2 Delegating Scope of Affordances

Foremost, managers may want to delegate affordances to contemporary AIs during development, and constantly adapt them upon deployment. Delegating the scope of affordances is about designing boundaries alongside capabilities for possible user actions for responsiveness. This strategic stewardship is about deciding what users can and cannot do with AIs during adoption. Managers would constantly be reckoning with possible intended or unintended user actions, given AI's inherent technical properties (e.g., appearance and learning). Managers would delegate which tasks (e.g., cognitive vs. mechanical vs. emotional) their deployable AIs enable. For simplicity, managers can think of cognitive-analytical (e.g., decision support), socio-emotional (e.g., relationship development), and mechanical-practical (e.g., automating floor cleaning) affordances.

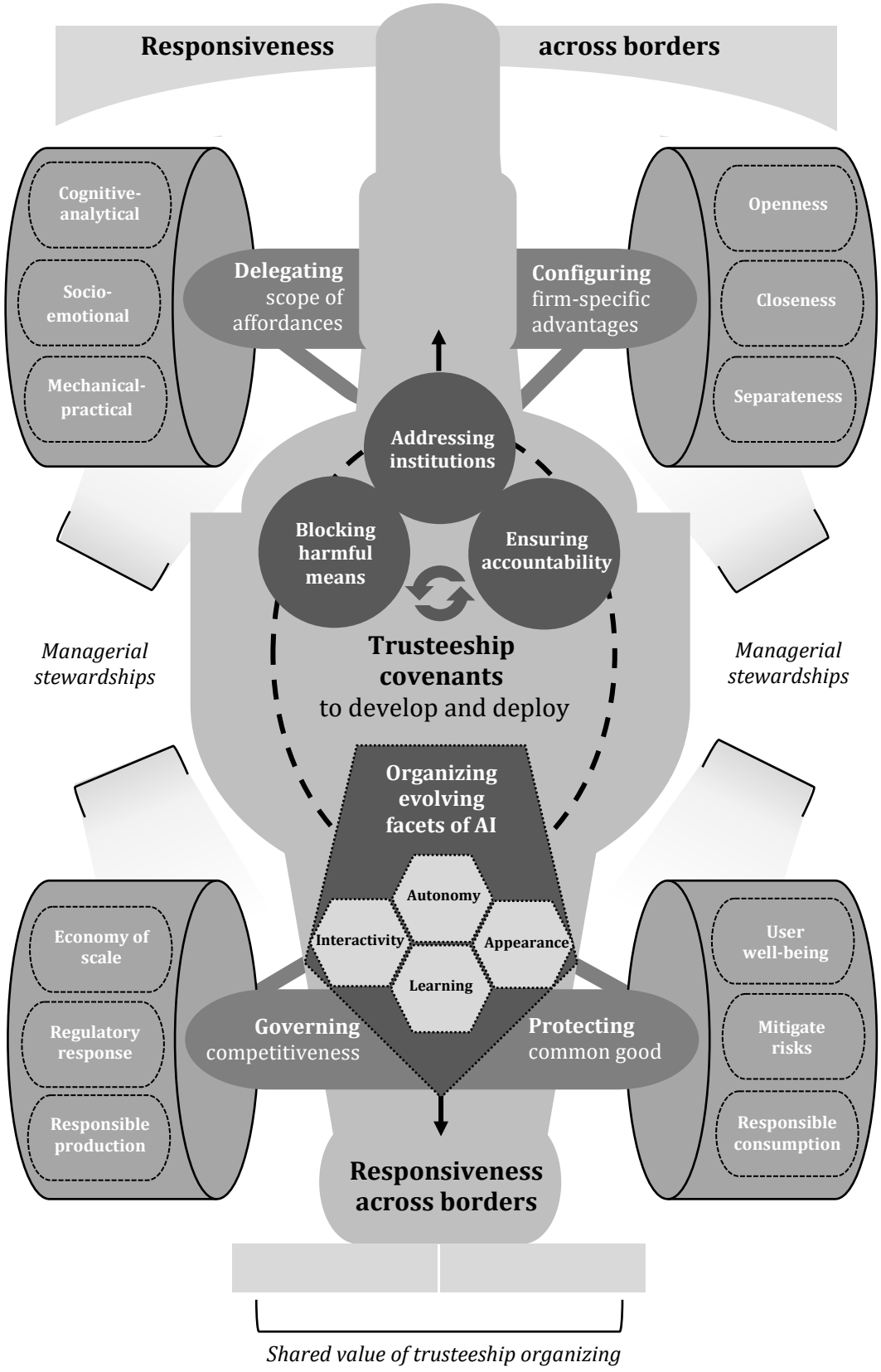


Figure 12. The globally responsive AI organizing (GRAO) framework

Delegating cognitive-analytical affordances relates to reasoning, analysis, or pattern detection, whereas socio-emotional affordances relate to human-like social or emotional interaction (e.g., simulated empathy). Additionally, mechanical-practical affordances relate to physical or routine practical tasks (e.g., vacuum cleaning). Hence, managers would delegate the scope of affordances by intentionally providing capabilities for specific tasks or functions that users can achieve through their actions. These achieved tasks or functions can be either partially or fully predicated. However, they can also be unanticipated, leading to unintended affordances and unforeseen risks of harm or outcomes. Therefore, managers would take preventive measures in their actions that may have direct consequences for firms across institutions, including those which are cultural and regulatory. Similarly, direct consequences for users include satisfaction, well-being, daily life, and even death. In turn, this affects managers by way of rewards, penalties, job loss, and legal action. Therefore, the strategic stewardship of delegating the scope of affordances is paramount for trusteeship organizing to enable responsiveness across markets. However, delegating affordances responsibly is inseparable from firm-specific advantages.

5.4.3 Configuring Firm-specific Advantages

In line with delegating affordances, managers may want to focus on a recombinant perspective of assets, transactions, and relationships of AIs in development and deployment. In turn, they may configure firm-specific advantages (FSAs) through achieving responsiveness across borders. Configuring FSAs for AIs through recombinant perspectives may involve organizing knowledge bundles of firms that provide a competitive advantage. They would utilize trusteeship covenants at the core of processual activities for developing a competitive advantage, accompanied by evolving facets of AI. They would address institutions, block harmful means, and ensure accountability globally by default. These covenants organize evolving facets of AIs (autonomy, learning, interactivity, and appearance), which are socio-technical and situated between their actions in processual activities and outcomes. Managers may want to take them as a shared cognitive logic across working units to govern how facets of AI are developed and deployed, including embedding them into products and services. In practice, facets of AI as modules can be intentionally modified or adapted to improve responsiveness against the aforementioned emerging affordances. Additionally, they can be distinctly reconfigured and combined with other modules (facets) to configure AIs that are part of FSAs. Contemporary AIs can be developed and deployed to combine two or more facets as modules and enable coupling while maintaining a degree of interdependence. Simultaneously, they can achieve looseness by making those modules independent. Hence, those facets can be

responsively modified to prevent harm due to affordance, while maintaining performance. Some practical illustrations may shed light on this process.

For example, managers would develop robotic AIs, such as a humanoid robot with a culturally sensitive appearance facet, to help care services emulate empathy and relationships. They would primarily focus on emotional-social affordances with high anthropomorphism and transparency. Such a system can also be recombined with a learning facet that includes computer vision and local temporal data storage and processing (with embedded AIs) to achieve secondary cognitive-analytical affordances. Additionally, managers may want to develop globally safe, robotic AIs, such as autonomous cars, that utilize an autonomy facet, which considers diverse road signs, markings, day, and weather conditions. Such systems may fall within the scope of mechanical-practical affordances. Managers may also recombine the learning facet through a subscription model to integrate public transportation services (with algorithmic AIs), helping users lower their car ownership costs and reduce their carbon footprint. Managers would ensure flexibility to comply with local regulatory frameworks such as the General Data Protection Regulation (GDPR) for learning and the AI Act for appearance or autonomy.

In practice, the foundational architecture of AI-enabled solutions can be governed via openness (e.g., open source, open weights), closeness (e.g., proprietary secrecy), or separateness (e.g., localized computing). While openness involves making parts of AIs modifiable or publicly auditable, closeness involves restricting access to model details or system logic. For instance, solutions could follow an open-source cocreation model that connects with other systems beyond focal internationalizing firms. However, they can also be entirely closed, yet released through a specific facet on the internet or a native mobile app, such as interactivity. AI-based solutions can also open and close simultaneously, which underpins the separateness mechanism that organizes architecture so that parts of the system operate independently of cloud services or external information technology networks. It also underscores the localization of systems that operate on local hardware or computation power. For example, managers deploy a set of totally separate, localized virtual AIs, such as chatbots powered by large language reasoning models, to protect users' data privacy with learning facets. Such a localized, moderate-level anthropomorphic solution can run on a local drive, a personal computer, or an isolated information technology environment, even without an internet connection. They also enable the interactivity facet to protect against age-related vulnerability and enhance information exchange services. Therefore, the strategic stewardship of configuring FSAs based on trusteeship organizing enables managers to govern AIs for competitiveness.

5.4.4 Governing Competitiveness

For governing competitiveness, managers may want to deliberately employ mechanisms that enable AI-driven solutions to compete, while remaining aligned with diverse institutional expectations on human values and accountability. This strategic stewardship would focus on the performance of AIs, but be responsible. Managers would ensure that AIs can scale across markets while maintaining user satisfaction, calibrated trustworthiness, and regulatory readiness. This means using trusteeship covenants as a strategic resource to make AIs globally deployable through institutional fit. In practice, this involves setting clear, shared internal rules for how systems are trained, updated, and monitored, so responsible teams can scale it across every country or business unit. For example, managers from a global service firm would deploy embodied virtual AIs for customer assistants like avatars, where they standardize its core design logic while allowing local business units to adapt its language, tone, and facial expressions. Concerning the interactivity facet, for instance, managers would enable the adaptation of communication style or imbue culture- or age-specific dialogue rules to avoid inappropriate conversations. Related to the appearance facet, for example, managers would enable localized accents in voice, facial features, and clothing to respect cultural sensitivities (e.g., gender representation) and avoid conflict with religious norms (e.g., modesty). Therefore, these processual activities reduce operational costs to achieve economies of scale, while maintaining higher user satisfaction. Managers who invest resources in covenants may face fewer disruptions when regulations change or when techno-political conflict arises, because their systems are already transparent, auditable, and adaptable. In turn, they may achieve social legitimacy that aligns with local value judgments and be trustworthy to service systems.

5.4.5 Protecting the Common Good

Finally, managers may want to protect the common good while ensuring competitiveness by delegating affordances and configuring FSAs. Protecting the common good requires managers to ask how AIs could unintentionally harm users or society. This stewardship thus emphasizes everyday managerial decisions about risk mitigation and user wellbeing, rather than abstract ethical principles. Creating economic value that integrates social welfare across both production and consumption thereby underscores the complexity of producer–product relations. In turn, it will enable managers to protect their internationalizing efforts or firms by avoiding reputational damage and related costs, as well as legal or regulatory issues. Managers would ensure that any facets of AIs are not deceptive or manipulative. They would also ensure that AIs do not exploit user vulnerabilities or amplify bias. They would constantly calibrate the performance of their value-creating AI-based assets

so as not to create over-reliance, unsafe dependencies, or misplaced trustworthiness. Reflecting on the interactivity facet, for instance, when deploying algorithmic AIs for personalization, managers may want to deliberately limit persuasive nudging or emotional cues that could manipulate users, for example, to block impulsive purchases, overuse, overspending, or addictive engagement. Similarly, they may want to restrict unnecessary data accumulation for the learning facet during human-AI interaction, even in the face of weak regulatory institutions that allow it to some extent, and so reduce the risk of privacy violations and protect firms' reputation.

In summary, implementing these forehead strategic stewardships is intended to move beyond economic reasoning and compliance. Instead, managers in the given operating model would make globally integrated strategic choices with long-term social and economic outcomes, while partitioning for institutional differences. It enables them to configure solutions for long-term competitiveness. Hence, it strikes a balance between economic and social impact. From this perspective, managerial responsibility is not peripheral – it is core to organizing facets of AI. Accordingly, strategic choices would promote responsible development and deployment in both production and consumption-facing activities, underpinning human-centered processual activities. Therefore, this perspective creates a fundamental shift in strategic organizing to scale contemporary AIs globally, thereby protecting the common good and simultaneously achieving responsiveness across borders. Action-oriented implications drawn from GRAO framework are relevant for managers, as much as for regulators.

5.5 Implications for Regulators on Emerging Accountability

The GRAO framework is directly relevant to regulation. In particular, its focus on ensuring accountability implies the core purpose of recent regulatory institutions undertaken worldwide, such as the AI Act (European Commission, 2024; IEEE, 2019; Jobin et al., 2019). Since the scope of affordance constantly evolves, along with the anthropomorphic properties of AI-based assets, products, or services in domestic and international markets, regulators may want to ensure users' psychological safety to block harmful means. Therefore, the two proposed implications below focus primarily on ensuring accountability, with a peripheral emphasis on blocking harmful means and addressing institutions.

5.5.1 Call for a New Industry Standard: 'Global Psychological Safety in Autonomous and Intelligent Systems'

The first proposal calls for international standard-setting bodies to develop an industry standard which can be named as the 'global psychological safety in autonomous and intelligent systems' standard. The call closely relates to the existing IEEE P7000 series of standards, particularly P7014 on emulated empathy (IEEE, 2024; Shneiderman, 2020). It also directly relates to the International Standards Organization's technical committee on robotics. This call focuses on two fundamentals – the GRAO framework and psychological impact testing. *Firstly*, regulators would develop a unified 'global trusteeship by default' framework for system development and deployment. To ensure accountability, it may include reinforcing self-disclosure and transparency, as well as contextual usage restrictions on emotive computation. For example, it may involve external transparency cues and internal trustworthy systems to ensure accountability. *Secondly,* the standard would include longitudinal psychological impact testing before deploying to contemporary AIs, similar to testing conducted during the development of medications. Its purpose is to block harmful means and address institutions (please see Table 7 for critical reflections, which can help to develop measures and dimensions). This testing would assess whether AIs contribute to unhealthy or unintended behavioral patterns among users, such as emotional dependence or misplaced trustworthiness. AIs developed for high-risk applications like mental health support or caregiving would respect human dignity. Deployments would undergo rigorous evaluations to ensure that they do not replace essential human touch and relationships, or exploit emotional vulnerabilities. The standard may accompany the labeling of potentially unacceptable risks.

5.5.2 Shifting Discourse from 'Trustworthy AI' to 'Trusteeship Organizing of AI'

The second proposal recommends using the label 'trusteeship organizing of AI' in regulations and policymaking to achieve coherence and ensure accountability, rather than 'trustworthy AI'. If there is a crash or accident involving a commercial public aircraft due to the producer's poor actions in quality control or the presence of technical issues, who should bear the responsibility? Is this the airline, the pilot, the passengers, or the producer? Can we call an aircraft 'trustworthy', or should human trust be baked into producer management systems that develop and deploy them in global markets? Similar questions arise amid the growing race to develop and deploy AIs worldwide, leading to mass usage and consumption. However, we still call it 'trustworthy AI'. Therefore, the GRAO framework shifts the focus of trustworthiness to actual human actors (managers who are responsible for development or

deployment), business entities (e.g., internationalizing firms or multinational corporations), and service systems (e.g., the offerings from firms), rather than to AIs as individual entities. It also shifts the problematic focus of the 'trustworthy AI' discourse, as used by regulators such as the European Commission (2019) and in policies such as the OECD Global Partnership on Artificial Intelligence (OECD, n.d.).

'Trustworthy AI' is conceived as a human-centric, safe, and secure approach to AI development and deployment that respects human rights. However, it is linguistically misleading. It implies that contemporary AIs are agentic entities that can inherently be trustworthy or worthy of trust. However, contemporary AIs cannot be held accountable, even though humans may develop misplaced trustworthiness in them, which can lead to dire consequences, even such as death (Bartneck et al., 2021; Burga, 2025). Instead, trusteeship organizing, as a shared value, positions trust or trustworthiness as being baked into the strategic organizing of AIs and the performance of AI-based solutions. Trustworthiness goes to human agents and the firms who own and control AIs, and not models or algorithms. They are responsible for developing and deploying AIs to act ethically and justly. Therefore, the labeling of 'trustworthy AI' risks naturalizing the idea that responsibility can be solved through technical fixes. This misattribution unfortunately undermines the socio-technical nature of facets of AI, thereby weakening accountability for emerging intangible assets, and shifts attention away from the risks of structural harm, such as the spread of disinformation, surveillance capitalism, labor exploitation, and the monopolistic control of computation and data. Thus, using the label 'trusteeship organizing of AI' may convey a true meaning.

5.6 Future Research Agenda

The above discussion provides insights into key theoretical explanations of the GRAO framework and its managerial implications, as well as well-developed regulatory recommendations. However, this sub-chapter now presents new issues and perspectives for a future research agenda that emphasizes methodological complementarity (Arbnor & Bjerke, 2009). Additionally, Table 8 presents some illustrative topical avenues for future research, derived from the proposed theoretical framework.

Table 8. Topical agenda for future research

Evolving facets of AI	Key illustrative recommendations		
	New topical issues and perspectives arising from the GRAO framework		
	Addressing institutions	**Blocking harmful means**	**Ensuring accountability**
Autonomy	***Development*** Examine how to develop calibrated degrees of autonomy in firm-specific advantages that match institutional expectations of human controllability. Examine how system architectures incorporate institutionally required human-override functions in delegating affordances (e.g., mechanical-practical). ***Deployment*** Explain why responsiveness differs when local norms discourage machine-led decision authority in affordances (e.g., cognitive-analytical or socio-emotional).	***Development*** Explore how autonomy constraints can be engineered to prevent harmful self-initiated actions in consumption. Examine how autonomy helps mitigate the risks of harm that emerge from anthropomorphization across national cultures. ***Deployment*** Assess to what extent existing ethical frameworks (e.g., OECD AI principles) detect harmful autonomous behaviors post-deployment. Assess the impact of responsive emergency shutdown mechanisms on ensuring users' complete control over a system and on developing calibrated trust in internationalizing firms.	***Development*** Explore how the globally integrated auditability mechanism of autonomous decisions can be configured in systems. ***Deployment*** Explain why managers develop an organizing framework to clarify accountability for potential malfunctions or harm across markets. Assess the effectiveness of deploying a workflow logging procedure to enhance transparency in a given foreign market.
Learning	***Development*** Assess how the learning facet can be designed to offer flexibility to respect local data-protection norms and/or laws. Understand how to embed national culturally sensitive feedback loops to prevent data-driven discrimination. ***Deployment*** Examine how managers enable the learning facet to adapt to emerging regional regulatory regimes or to adjust to geopolitical conflicts. Assess why internationalizing firms enable closeness, openness, or separateness in learning-facet-oriented firm-	***Development*** Explore how guardrails can be built to prevent harmful learning (e.g., hallucinations or the spread of disinformation) from data and user interactions. Assess which design approach (e.g., value-based) ensures learning-enabled firm-specific assets that do not exploit vulnerabilities in user-centric applications. ***Deployment*** Explain why managers ensure safe model adaptation to protect vulnerable groups across markets (e.g., regulatory response,	***Development*** Test the extent to which explainability can be embedded into the learning facet in international markets. Examine the mechanisms by which system architectures allow role-based responsibility tracking in a given foreign market. ***Deployment*** Assess the extent to which managers ensure the emergent behavior of learning-based firm-specific advantage remains auditable. Explain mechanisms in which managers detect harm-amplifying

Evolving facets of AI	Key illustrative recommendations		
	specific advantages to enhance resilience (e.g., in the face of geopolitical conflicts).	organizational cultures, long-term competitiveness).	learning in firm-specific advantages.
Interactivity	***Development*** Understand how interactivity can be designed to match culturally appropriate communication norms or interaction boundaries. Examine how interactivity can be developed to ensure real-time reciprocity that respects culturally specific interaction cues (e.g., expression, head tilting, head nodding, eye gazing). ***Deployment*** Explain how managers adapt interactivity in real time when the interaction style conflicts with market-specific norms and value judgments. Examine how managers localize interaction styles across markets while maintaining their firms' value propositions or brand identity.	***Development*** Explore how systems avoid deceptive or manipulative conversational patterns. Explore what design approaches limit exploitative personalization in international markets. Examine how interactivity can be developed to ensure real-time reciprocity without deception or manipulation. ***Deployment*** Examine the mechanisms for monitoring risk of harm across national cultures (or specific ethnic cultures with a national border). Examine how managers enable rapid adjustments to interactive features when the risk of harm (e.g., deception or manipulation, privacy violations) becomes evident.	***Development*** Examine how an interactive loop in AI anthropomorphism remains traceable in human-AI interaction. Examine how interaction-related process tracing logs support accountability while protecting user privacy. ***Deployment*** Explore what standardized mechanisms trace the interaction-based risk of harm in AI anthropomorphism via affective computing.
Appearance	***Development*** Examine how morphological cues avoid violating institutional norms, such as those regarding gender or religious representation. Explore the necessary guardrails to prevent appearance-based cultural misinterpretation. ***Deployment*** Explain why users' religious or cultural beliefs affect the intention to use or purchase in embodied systems in given foreign markets. Examine how visual or auditory cues are adapted to evolving cultural	***Development*** Explore the design choices to prevent reinforcing harmful stereotypes via appearance. ***Deployment*** Examine the mechanism by which appearance can be adapted to prevent risks of harm (e.g., overreliance, misplaced trust, over-purchase) due to anthropomorphism. Examine how appearance be adapted in identity-sensitive foreign markets.	***Development*** Examine how managers justify morphological design choices for legal or regulatory accountability in host vs home markets. Explore how managers develop embodied systems (e.g., humanoid robots) to avoid deceptive signaling (e.g., hidden capabilities, false impressions). ***Deployment*** Examine why appearance enables users to attribute moral responsibility to AIs (vs. firms) due to cultural sensitivities (e.g., China, Korea), making

Evolving facets of AI	Key illustrative recommendations	
	sensitivities in response to local political shifts (e.g., conservatism, nationalism). Explore how internationalizing firms localize appearance without fragmenting global deployment (e.g., virtual AIs).	accountability for potential harm problematic.

5.6.1 Echeloned Design Science Research Method for Robotic AI System and Responsiveness

Future research may employ the echeloned design science research (eDSR) method, combined with the GRAO framework, to develop robotic AIs for responsiveness (Tuunanen et al., 2024). With such a research design, studies can conceptualize AIs as layered socio-technical architectures. Each layer shapes trusteeship covenants, and the facets, affordances, and performance of AIs. This approach aligns with the framework's focus on ongoing managerial capabilities. At the *highest echelon*, research can examine a trusteeship covenant layer as foundational limits, which serves to define a globally stable design principle across markets while allowing adaptation during deployment. A *second echelon* can focus on the configuration of facets of AI, such as autonomy and interactivity. These facets are modular and adjustable. Future studies might investigate how managers enhance or restrict these facets during international expansion. For example, they can limit humanlike appearance within a given institutional context, while enhancing interactivity to align with local cultural norms. An echeloned design allows such reconfiguration without compromising the core performance of AIs. A *third echelon* may relate to the affordances delegated to users, in which managerial decisions determine how the capabilities of robotic AIs translate into cognitive, mechanical, or socio-emotional actions. Research could examine how affordances are gradually expanded or restricted across markets based on user vulnerability and expectations. Balancing these echelons over time may help robotic AIs to maintain calibrated trustworthiness and competitiveness under institutional differences.

5.6.2 Interpretive Ethnography on the Performance of Virtual AI Systems

Future research can employ an interpretive ethnography research design (Mahadevan & Moore, 2023) to study managers inside a global digital firm whose core FSAs rely on virtual AIs. The firm's virtual AI-as-a-service might be delivered to

users via the cloud, a web interface, or a mobile app. Researchers can observe work as an insider in a specific institutional setting such as India, China, Australia, or Denmark, while paying close attention to the firm's own organizational culture, and implicitly examine how managers apply three trusteeship covenants when configuring FSAs through AIs, achieving responsiveness. However, the unit of analysis may want to reflect on the conflicts and tensions surrounding these covenants, including how managers confront, negotiate, or resist them in the context of their organizational culture. Against this backdrop, researchers may produce a thick description and a reflexive account as insiders, and uncover hidden managerial actions and determine whether alternative logics dominantly exist. Such logics may entirely oppose the understanding of the GRAO framework, such as growth imperatives, competitive positioning, and investor expectations, and are often accompanied by minimal or no commitment to harm prevention, institutional attentiveness, or accountability. However, studies may also unpack why managers decide to take particular actions as they navigate competing logics, power asymmetries, and commercial pressures. Therefore, the empirical unit of observation may include multiple management actors within different units. Researchers can further reflect on how managers of relevant units detect risks of harm, negotiate the framing of accountability in relation to geopolitical or regulatory pressures, and interpret institutional concerns. For example, researchers may observe whether managers localize interactivity or appearance facets to respect cultural norms, or instead prioritize speed-to-market and standardize global deployment.

Alternatively, ethnographic studies can examine how 'blocking harmful means' is understood within fast-paced development environments: for example whether it is treated as a design priority or a challenge that slows performance metrics. Similarly, accountability may be enacted through managerial narratives of 'innovation freedom' or outsourced to compliance teams with minimal authority. If possible, however, longitudinal future studies can trace how facets of AI are selectively configured, restricted, or enhanced across markets, and interpretive ethnography is uniquely capable of capturing these subtle managerial actions within a given organizational culture. Alternatively, the contextualized explanation case study design might be used for examining such a context-laden phenomenon.

5.6.3 A Multiple Explanation Case Study Design on the Scope of Affordances: Robotic vs Virtual AI Systems

Future research can employ a multiple mechanism-based explanation case study design (Väyrynen et al., 2025; Welch et al., 2022), emphasizing the scope of affordances across markets. They can construct cases based on different robotic and virtual AIs and, for instance, contextualize them in service-related settings. They can

study how each system enables or distorts consumers' or users' possible actions, and also analyze the underlying, perhaps hidden, socio-technical means that pose a risk of harm. A configurational theorizing or mechanism-based approach can support this endeavor (Furnari et al., 2021; Siponen et al., 2025), focusing on responsiveness as the outcome of interest. Here, the scope of affordances serves as a key anchor for understanding its mechanism. Besides primary data such as expert interviews, studies may also use secondary data, including news articles, reports, Reddit threads, consumer forums, and user reviews.

Affordances arise from the interaction between given facets of AI, user actions, and institutions such as culture and norms. Robotic AIs might trigger mechanical-practical and socio-emotional affordances through physical presence. Simultaneously, however, they can also create a risk of proximity-based harms, such as surveillance or perceived threat. In contrast, virtual AIs may yield cognitive-analytical and socio-emotional affordances, which may enable decision support and companionship, which may or may not pose a risk of harm, depending on the users' expectations, shaped by their institutions (e.g., value judgment). Therefore, future studies can compare different institutional settings, including national cultures, to explain how affordances expand or new actions emerge. Particularly, a new affordance may emerge that is neither predicted nor anticipated by managers when the system is developed, leading to unknown or unintended risks of harm. These unintended effects may hamper the performance of AIs that managers originally intended, undermining the strategic goal of responsiveness. Therefore, such multiple casing may reveal how risks of harm vary across affordances, across types of AIs, and across countries. Configuration analysis can map these risks for robotic AIs and virtual AIs and show which affordances can be standardized globally, and which require local adaptation. This comparison, in turn, helps to explain why managers need to adjust relevant facets of AI into FSAs through trusteeship mechanisms. It may also explain how to remain responsive across borders while protecting user well-being and enabling calibrated trustworthiness.

5.6.4 Construct Measurement of the GRAO Framework: Scale Development and Validation for Different Contemporary Forms

A critical avenue for future research lies in developing and validating measures of construct to operationalize the GRAO framework in both IB and IS (see Cadogan et al., 1999; MacKenzie et al., 2011). This sub-chapter conceptualizes and defines distinct constructs identified earlier, and formally specifies a measurement model with a causal path (see Figure 13 for an adapted model of the framework for possible testing, and Table 7 for additional insights). These understandings and explanations may serve as a starting point to develop items to represent the construct and assess its

content validity. Translating such systems-oriented framing into testable model constructs will bring complementarity across methodological views. However, separate multi-level measurement may be useful as the interaction effects among constructs are processual, relational, and evolving for each form of contemporary AIs, namely robotic AIs, virtual AIs, embedded AIs, and algorithmic AIs.

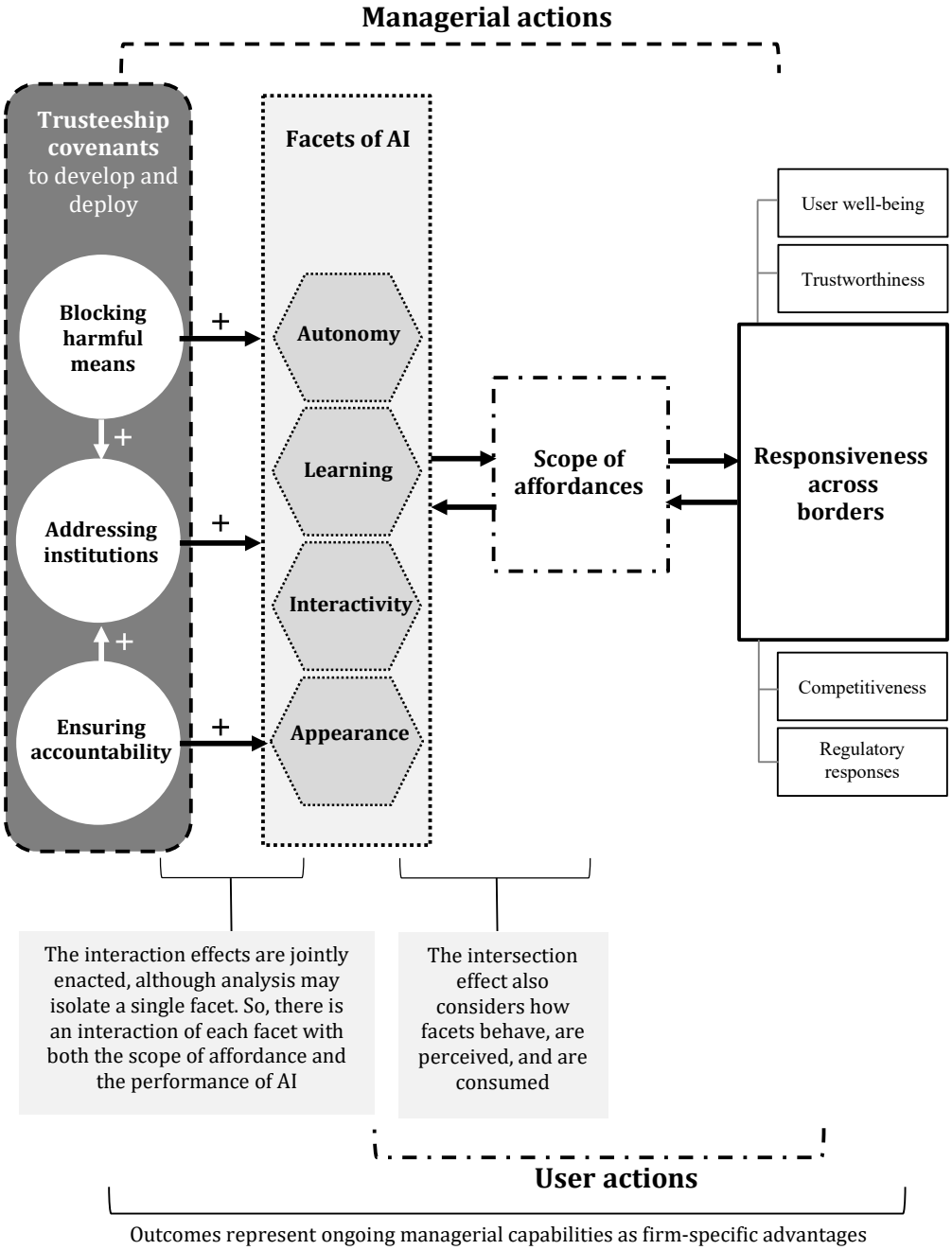


Figure 13. An adaptation of the framework for possible testing

First, a trusteeship covenant-level measure that comprises three key constructs. These should be treated not as abstract ethical intentions, but more as observable organizing practices for developing and deploying AIs in international markets. Measures should capture joint covenant effects, as the framework suggests that trusteeship covenants reinforce one another even though managers may operationalize them through distinct interventions. It further reflects bundles of practices rather than isolated actions.

Second, facets of AI-level measures, consisting of autonomy, learning, interactivity, and appearance constructs. Since each facet has its own dynamics and may be embedded within AIs alone, depending on the nature or type of its contemporality, each facet must be treated as a unique construct, making different yet interrelated constructs at this level. All four facets may be relevant for robotic AIs and virtual AIs, whereas autonomy and learning relate to embedded AIs and algorithmic AIs. When relevant for human-AI interaction, developing items for facets of AI must treat them as socio-technical constructs with configuring properties that interact with humans, rather than purely technical features. Item development must move beyond binary indicators. For instance, in the case of autonomy (autonomy vs. non-autonomy), measures must capture degrees of autonomy, their adaptability, and anthropomorphic signaling. Because facets of AI evolve, stability over time needs to be specified in the conceptual themes of these constructs to capture how managerial choices shape the performance of contemporary AIs.

Third, a measurement for the performance of AI, enabled by single or combinations of facets. Here, sub-dimensions may include not only economies of scale and profit, but also adaptability and compliance across borders. Sub-dimensions may also reflect how managers organize them, such as openness vs. closeness vs. separateness, to enhance proprietary data and user well-being while remaining difficult to imitate due to their emerging affordances. *Fourth,* an affordance-level measure to operationalize its growing scope. Dimensions of affordances can be distinguished between cognitive-analytical, mechanical-practical, and socio-emotional. Rather than measuring usage intensity alone, items should go deep into construct dimensionality to assess how users perceive specific facets of AI and their appropriateness for action within specific contexts such as service provision. This is particularly important for examining the mediating effect of affordances on causal relationships among other key constructs. The framework suggests that affordances extend beyond firm boundaries, thereby influencing how user-facing contemporary AIs are interpreted and consumed.

Finally, a multidimensional processual outcome measurement for the construct of responsiveness across borders that integrates competitiveness, regulatory

responses, calibrated trustworthiness, and user well-being. Rather than treating these dimensions independently, items would reflect their co-production, since the framework suggests that trusteeship organizing simultaneously enables market performance and societal welfare. Measurement must remain sensitive to the emergent nature of facets of AI, enabling adaptive indicators that can evolve alongside technological change and usage behavior. Rather than aiming for overly rigid instruments, items would treat the interaction effect across constructs as an iterative process, reflecting the framework's emphasis on ongoing organizing capabilities, underpinning FSAs. However, cross-national scale validation for each type of contemporary form (e.g., algorithmic AIs, embedded AIs) is particularly important, as constructs such as accountability, trustworthiness, and well-being may have different meanings across institutional settings. Validation may therefore pursue measurement invariance or metric equivalence testing to ensure that measures are consistent across countries or institutions.

5.6.5 Quasi-experimental Design on Facets of AI in International Expansion

The GRAO framework is particularly well-suited to quasi-experimental research designs (Jiang, et al., 2024; Ramani & Aguinis, 2023). Managerial actions across firms vary to configure facets into AIs, which in turn may vary across contemporary forms, countries, or phases of international expansion. Future research can explore this variation to examine how different configurations of facets, such as learning and appearance, produce divergent manager- and user-level outcomes when firms internationalize or scale AI-based solutions abroad. One promising approach is to examine the rapid internationalization of an AI-based international new venture in which its AI-based FSAs are deployed in parallel across countries with differing institutional environments (Cheung & Niittymies, 2026). Researchers can compare markets where autonomy is constrained (e.g., recommendation-only algorithmic AIs) with those that permit greater autonomy (Hou et al., 2026). As examples, they can assess the impact on regulatory responses, calibrated trustworthiness, or assess the performance of AIs. Analysis can isolate the effects of changes in autonomy while holding constant the core functionality of AIs on responsiveness.

A second quasi-experimental design may involve geopolitical conflicts, such as the introduction of embargoes or restrictions on AI-related products or services from a given country. Managers often feel forced to respond or make no response as part of a strategic decision. To respond, for example, managers may reconfigure learning facets, auditability, or oversight structures in affected markets, while leaving other facets like interactivity, unchanged. These tensions allow researchers to study how ensuring accountability through altered learning facets influences cross-border

responsiveness. As a third example, variation in the interface for local sensitivity creates opportunities to examine interactivity and appearance as treatment conditions. For example, managers may deploy culturally adapted conversational styles or anthropomorphic facial features in some countries, but not others. Similarly, learning facets may be localized into robotic AIs or embedded AIs in some countries, but not others, due to national data security concerns or different data regulations. By comparing user behavior and well-being across these settings, future studies can assess how trusteeship covenants enable particular facets to align with institutionally appropriate local responsiveness. Finally, future studies can explore restructuring events such as launching open weights models or platform opening, which alter openness and separateness in AIs to configure FSAs. For instance, observing market performance before and after such reconfigurations may enable a causal inference about how learning facet localization supports scalable, yet responsible, international expansion.

5.7 Concluding Remarks

This dissertation aimed to examine how AI may be managed across national borders to integrate economic reasoning with social well-being. The emergent understanding in the context of common good may provide three key points. *First,* AI is the frontier of computation advancement – an evolving phenomenon with primarily four interrelated facets: (1) autonomy, (2) learning, (3) interactivity, and (4) appearance. These facets may create affordances that are intended to create ever-expanding opportunities for achieving tasks. These affordances may also be fully or partially unpredictable, thereby strengthening or weakening the performance of contemporary AI systems. *Second,* this emergent nature of facets may necessitate rethinking strategic organizing of AI systems for responsiveness across borders, which may not be understood as mere country-level adaptation. Instead, responsiveness becomes an ongoing processual activity that shapes development and deployment, consumption or usage, emerging consequences, and managerial responses with new decisions around facets of contemporary AI systems. *Third,* achieving responsiveness may require trusteeship organizing, consisting of three globally integrated interrelated covenants that intentionally prioritize proactiveness in (a) addressing institutions, (b) blocking harmful means, and (c) ensuring accountability. These trusteeship covenants are integrated into facets of AI globally by default, which may require constant reflection and responsiveness. There is a possibility of both intended and unintended risks of harm due to the emergence of affordances and the performance of AI systems, which may have negative consequences for end-users. Without a trusteeship framework, internationalization efforts could fail and pose reputational, political, or regulatory risks for both firms

and top managers. The dissertation also calls for a regulatory shift in existing discourse from 'trustworthy AI' to 'trusteeship organizing of AI' to achieve coherence and ensure accountability across markets.

With these points of departure, the dissertation hopes to institute a systems-oriented, trusteeship-based process for managing AI across borders that integrates economic goals with social responsibilities. The process is intended to safeguard the user's well-being, and serves as an ongoing managerial capability to underpin firm-specific advantages. With such productive closure, this dissertation contributes to the current scholarly discourse on the strategic management of AI systems. It may inform managers to responsibly develop and deploy AI systems for their international operations and configure products or services worldwide to achieve responsiveness. Additionally, it may guide regulators on the risks posed by irresponsible development and deployment of AI systems, thereby enabling preventive measures to ensure societal well-being. In addition to theory and practice, this dissertation also proposes several critical future opportunities for organizing AI systems to advance both international business and information systems research. Consequently, it collectively advances existing conversations on how AI-related firm-level activities have differential impacts on business and society, including both positive and negative consequences.

References

Adarkwah, G. K., & Malonæs, T. P. (2022). Firm-specific advantages: A comprehensive review with a focus on emerging markets. *Asia Pacific Journal of Management*, *39*(2), 539–585. https://doi.org/10.1007/s10490-020-09737-7

Aguinis, H. (2026). Method-driven theory advancements and AI implementation. *Journal of International Business Studies*. Advance online publication. https://doi.org/10.1057/s41267-026-00851-0

Aguinis, H., Audretsch, D. B., Flammer, C., Meyer, K. E., Peng, M. W., & Teece, D. J. (2022). Bringing the manager back into management scholarship. *Journal of Management*, *48*(7), 1849–1857. https://doi.org/10.1177/01492063221082555

Aguinis, H., & Gabriel, K. P. (2022). International business studies: Are we really so uniquely complex? *Journal of International Business Studies*, *53*(9), 2023–2036. https://doi.org/10.1057/s41267-021-00462-x

Ahi, A., Sinkovics, N., Shildibekov, Y., Sinkovics, R., & Mehandjiev, N. (2022). Advanced technologies and international business: A multidisciplinary analysis of the literature. *International Business Review*, *31*(4), Article 4. https://doi.org/10.1016/j.ibusrev.2021.101967

Al-Ghazālī, A. H. (2024). *The standard of knowledge (Miʿyar al-ʿIlm)*. (J. Meri, Trans.). The Royal Islamic Strategic Studies Centre. (Original work published 1095).

Ali, A. J., Al-Aali, A., & Al-Owaihan, A. (2013). Islamic perspectives on profit maximization. *Journal of Business Ethics*, *117*(3), 467–475. https://doi.org/10.1007/s10551-012-1530-0

Ali, F., Bouzoubaa, K., Gelli, F., Hamzi, B., & Khan, S. (2025). Islamic ethics and AI: An evaluation of existing approaches to AI using trusteeship ethics. *Philosophy & Technology*, *38*(3), 120. https://doi.org/10.1007/s13347-025-00922-4

Alvesson, M., & Gabriel, Y. (2013). Beyond formulaic research: In praise of greater diversity in organizational research and publications. *Academy of Management Learning & Education*, *12*(2), 245–263. https://doi.org/10.5465/amle.2012.0327

Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *Academy of Management Review*, *36*(2), 247–271. https://doi.org/10.5465/AMR.2011.59330882

Ameen, N., Tarba, S., Cheah, J.-H., Xia, S., & Sharma, G. D. (2024). Coupling artificial intelligence capability and strategic agility for enhanced product and service creativity. *British Journal of Management*, *35*(4), 1916–1934. https://doi.org/10.1111/1467-8551.12797

Andersen, T. J., & Hallin, C. A. (2017). *Global Strategic Responsiveness: Exploiting Frontline Information in the Adaptive Multinational Enterprise*. Routledge. https://doi.org/10.4324/9781315469058

Anderson, C., & Robey, D. (2017). Affordance potency: Explaining the actualization of technology affordances. *Information and Organization*, *27*(2), 100–115. https://doi.org/10.1016/j.infoandorg.2017.03.002

Araghi, S., Palangkaraya, A., & Webster, E. (2024). The impact of language translation quality on commerce: The example of patents. *Journal of International Business Policy*, *7*(2), 224–246. https://doi.org/10.1057/s42214-023-00157-0

Arbnor, I., & Bjerke, B. (2009). *Methodology for Creating Business Knowledge* (3rd ed.). Sage Publications. https://doi.org/10.4135/9780857024473

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (1998). *Critical realism: Essential readings* (1st ed.). Routledge.

Arnould, E. J., & Thompson, C. J. (2005). Consumer culture theory (CCT): Twenty years of research. *Journal of Consumer Research*, *31*(4), 868–882. https://doi.org/10.1086/426626

Asatiani, A., Malo, P., Nagbøl, P., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems*, *22*(2), 325–252. https://doi.org/10.17705/1jais.00664

Autio, E., Mudambi, R., & Yoo, Y. (2021). Digitalization and globalization in a turbulent world: Centrifugal and centripetal forces. *Global Strategy Journal*, *11*(1), 3–16. https://doi.org/10.1002/gsj.1396

Avital, M., Mathiassen, L., & Schultze, U. (2017). Alternative genres in information systems research. *European Journal of Information Systems*, *26*(3), 240–247. https://doi.org/10.1057/s41303-017-0051-4

Baldwin, R. (2016). *The great convergence: Information technology and the new globalization*. Harvard University Press.

Baldwin, R. (2019). *The globotics upheaval: Globalisation, robotics, and the future of work*. Weidenfeld & Nicolson.

Banalieva, E. R., & Dhanaraj, C. (2019). Internalization theory for the digital economy. *Journal of International Business Studies*, *50*(8), 1372–1387. https://doi.org/10.1057/s41267-019-00243-7

Bartlett, C. A., & Ghoshal, S. (1989). *Managing across borders: The transnational solution*. Harvard Business Press.

Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *An introduction to ethics in robotics and AI*. Springer Nature.

Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology*, *22*(5), 543–564. https://doi.org/10.1080/09515080903238948

Beekun, R. I., & Badawi, J. A. (2005). Balancing ethical responsibility among multiple organizational stakeholders: The Islamic perspective. *Journal of Business Ethics*, *60*(2), 131–145. https://doi.org/10.1007/s10551-004-8204-5

Belderbos, R., Lokshin, B., Boone, C., & Jacob, J. (2022). Top management team international diversity and the performance of international R&D. *Global Strategy Journal*, *12*(1), 108–133. https://doi.org/10.1002/gsj.1395

Benbasat, I., & Wang, W. (2005). Trust In and Adoption of Online Recommendation Agents. *Journal of the Association for Information Systems*, *6*(3). https://doi.org/10.17705/1jais.00065

Benito, G., Cuervo-Cazurra, A., Mudambi, R., Pedersen, T., & Tallman, S. (2022). The future of global strategy. *Global Strategy Journal*, *12*(3), 421–450. https://doi.org/10.1002/gsj.1464

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Birkinshaw, J. (2022). Move fast and break things: Reassessing IB research in the light of the digital revolution. *Global Strategy Journal*, *12*(4), 619–631. https://doi.org/10.1002/gsj.1427

Birkinshaw, J. (2024). Too little, too late? How policymakers and regulators respond to the business model innovations of digital firms. *Academy of Management Perspectives*, *38*(3), 269–285. https://doi.org/10.5465/amp.2022.0233

Birkinshaw, J., Morrison, A., & Hulland, J. (1995). Structural and competitive determinants of a global integration strategy. *Strategic Management Journal*, *16*(8), 637–655. https://doi.org/10.1002/smj.4250160805

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*(4), 632–658. https://doi.org/10.1007/s11747-020-00762-y

Bonache, J. (2021). The challenge of using a 'non-positivist' paradigm and getting through the peer-review process. *Human Resource Management Journal*, *31*(1), 37–48. https://doi.org/10.1111/1748-8583.12319

Bostrom, N. (2016). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Boyett, I., & Currie, G. (2004). Middle managers moulding international strategy: An Irish start-up in Jamaican telecoms. *Long Range Planning*, *37*(1), 51–66. https://doi.org/10.1016/j.lrp.2003.11.009

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2025). Reporting guidelines for qualitative research: A values-based approach. *Qualitative Research in Psychology*, *22*(2), 399–438. https://doi.org/10.1080/14780887.2024.2382244

Brendel, A., Hildebrandt, F., Dennis, A., & Riquel, J. (2023). The paradoxical role of humanness in aggression toward conversational agents. *Journal of Management Information Systems*, *40*(3), 883–913. https://doi.org/10.1080/07421222.2023.2229127

Brouthers, L. E., Mukhopadhyay, S., Wilkinson, T. J., & Brouthers, K. D. (2009). International market selection and subsidiary performance: A neural network approach. *Journal of World Business*, *44*(3), 262–273 https://doi.org/10.1016/j.jwb.2008.08.004

Brynjolfsson, E., Hui, X., & Liu, M. (2019). Does machine translation affect international trade? Evidence from a large digital platform. *Management Science*, *65*(12), 5449–5460. https://doi.org/10.1287/mnsc.2019.3388

Buckley, P. J., Elia, S., Gaur, A., Krakowski, S., Kuivalainen, O., & Pedota, M. (2026). Artificial intelligence and international business: Theoretical challenges, strategic implications, and research agenda. *Journal of World Business*, *61*(3), Article 101710. https://doi.org/10.1016/j.jwb.2025.101710

Buckley, P. J., & Enderwick, P. (2025). The Global Factory Revisited. *Management International Review*, *65*(4), 615–635. https://doi.org/10.1007/s11575-025-00585-5

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

Burga, S. (2025, August 26). Parents Allege ChatGPT Responsible for Son's Death by Suicide. *Time*. https://time.com/7312484/chatgpt-openai-suicide-lawsuit/

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human–computer interaction: A study of credibility, understanding, and influence. *Computers in Human Behavior*, *16*(6), 553–574. https://doi.org/10.1016/S0747-5632(00)00029-7

Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life* (1st ed.). Routledge.

Busetto, L., Wick, W., & Gumbinger, C. (2020). How to use and assess qualitative research methods. *Neurological Research and Practice*, *2*(1), Article 14. https://doi.org/10.1186/s42466-020-00059-z

Bygstad, B., Munkvold, B. E., & Volkoff, O. (2016). Identifying generative mechanisms through affordances: A framework for critical realist data analysis. *Journal of Information Technology*, *31*(1), 83–96. https://doi.org/10.1057/jit.2015.13

Cadogan, J. W., Diamantopoulos, A., & De Mortanges, C. P. (1999). A measure of export market orientation: Scale development and cross-cultural validation. *Journal of International Business Studies*, *30*(4), 689–707. https://doi.org/10.1057/palgrave.jibs.8490834

Carroll, M., Chan, A., Ashton, H., & Krueger, D. (2023). Characterizing manipulation from AI systems. *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*, 1–13. https://doi.org/10.1145/3617694.3623226

Cavusgil, S. T., & Evirgen, C. (1997). Use of expert systems in international marketing: An application for co-operative venture partner selection. *European Journal of Marketing*, *31*(1), 73–86. https://doi.org/10.1108/03090569710157043

Cerar, J., Minbaeva, D., & Nell, P. C. (2026). Reliance on AI in augmented strategic decision-making: Navigating cultural and national dynamics. *International Business Review*, *35*(4), Article 102600. https://doi.org/10.1016/j.ibusrev.2026.102600

Chalmers, D., Hunt, R., Pachidi, S., Potocnik, K., & Townsend, D. (2026). The acceleration of artificial intelligence: Rethinking organization and work in an era of rapid technological change. *Journal of Management Studies*. Advance online publication. https://doi.org/10.1111/joms.70063

Cheung, Z., & Niittymies, A. (2026). Parallel internationalization during radical industry transformations: The case of Telecom Finland. *Journal of International Business Studies*, *57*, 618–635. https://doi.org/10.1057/s41267-026-00871-w

Ciulli, F., & Kolk, A. (2023). International business, digital technologies and sustainable development: Connecting the dots. *Journal of World Business*, *58*(4), Article 101445. https://doi.org/10.1016/j.jwb.2023.101445

Claypool, R. (2023). *Chatbots are not people: Designed-in dangers of human-like A.I. systems*. Public Citizen. https://www.citizen.org/article/chatbots-are-not-people-dangerous-human-like-anthropomorphic-ai-report/

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). A definition, benchmark and database of AI for social good initiatives. *Nature Machine Intelligence*, *3*(2), 111–115. https://doi.org/10.1038/s42256-021-00296-0

Cunliffe, A. L. (2003). Reflexive inquiry in organizational research: Questions and possibilities. *Human Relations*, *56*(8), 983–1003. https://doi.org/10.1177/00187267030568004

Danaher, J. (2020). Robot Betrayal: A guide to the ethics of robotic deception. *Ethics and Information Technology*, *22*(2), 117–128. https://doi.org/10.1007/s10676-019-09520-3

Dau, L. A., Chacar, A. S., Lyles, M. A., & Li, J. (2022). Informal institutions and international business: Toward an integrative research agenda. *Journal of International Business Studies*, *53*(6), 985–1010. https://doi.org/10.1057/s41267-022-00527-5

Dau, L. A., Moore, E. M., & Kostova, T. (2020). The impact of market based institutional reforms on firm strategy and performance: Review and extension. *Journal of World Business*, *55*(4), Article 101073. https://doi.org/10.1016/j.jwb.2020.101073

Davenport, T., & Ronaki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, *96*(1), 108–116.

Delouya, S. (2023, March). Replika users say they fell in love with their AI chatbots, until a software update made them seem less human. *Business Insider*. https://www.businessinsider.com/replika-chatbot-users-dont-like-nsfw-sexual-content-bans-2023-2

Denicolai, S., Zucchella, A., & Magnani, G. (2021). Internationalization, digitalization, and sustainability: Are SMEs ready? A survey on synergies and substituting effects among growth paths. *Technological Forecasting and Social Change*, *166*, Article 120650. https://doi.org/10.1016/j.techfore.2021.120650

Dolata, M., Feuerriegel, S., & Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. *Information Systems Journal*, *32*(4), 754–818. https://doi.org/10.1111/isj.12370

dos Santos, J., & Williamson, P. (2024). Beyond connectivity: Artificial intelligence and the internationalisation of digital firms. *Information and Organization*, *34*(4), Article 100538. https://doi.org/10.1016/j.infoandorg.2024.100538

Dubois, A., & Gadde, L.-E. (2002). Systematic combining: An abductive approach to case research. *Journal of Business Research*, *55*(7), 553–560. https://doi.org/10.1016/S0148-2963(00)00195-8

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Eden, L., & Nielsen, B. B. (2020). Research methods in international business: The challenge of complexity. *Journal of International Business Studies*, *51*(9), 1609–1620. https://doi.org/10.1057/s41267-020-00374-2

European Commission,. (2019, April 8). *High-level expert group on AI: Ethics guidelines for trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission,. (2022, January 10). *Industry 5.0*. https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry-50_en

European Commission,. (2024, August 8). *AI Act*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Faraj, S., & Leonardi, P. (2022). Strategic organization in the digital age: Rethinking the concept of technology. *Strategic Organization*, *20*(4), 771–785. https://doi.org/10.1177/14761270221130253

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design ai for social good: Seven essential factors. *Science and Engineering Ethics*, *26*(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Floyd, S. W., & Wooldridge, B. (1992). Middle management involvement in strategy and its association with strategic type: A research note. *Strategic Management Journal*, *13*(S1), 153–167. https://doi.org/10.1002/smj.4250131012

Franklin, M., Tomei, P. M., & Gorman, R. (2023, September 7). *Vague concepts in the EU AI Act will not protect citizens from AI manipulation*. OECD.AI. https://oecd.ai/en/wonk/eu-ai-act-manipulation-definitions

Fuenfschilling, L., & Binz, C. (2018). Global socio-technical regimes. *Research Policy*, *47*(4), 735–749. https://doi.org/10.1016/j.respol.2018.02.003

Furnari, S., Crilly, D., Misangyi, V. F., Greckhamer, T., Fiss, P. C., & Aguilera, R. V. (2021). Capturing causal complexity: Heuristics for configurational theorizing. *Academy of Management Review*, *46*(4), 778–799. https://doi.org/10.5465/amr.2019.0298

Furr, N., Ozcan, P., & Eisenhardt, K. M. (2022). What is digital transformation? Core tensions facing established companies on the global stage. *Global Strategy Journal*, *12*(4), 595–618. https://doi.org/10.1002/gsj.1442

Fuso Nerini, F. (2026). AI economics for the common good. *Nature Machine Intelligence*, *8*(4), 495–496. https://doi.org/10.1038/s42256-026-01212-0

Gaur, A., Pattnaik, C., Singh, D., & Lee, J. Y. (2022). Societal trust, formal institutions, and foreign subsidiary staffing. *Journal of International Business Studies*, *53*(6), 1045–1061. https://doi.org/10.1057/s41267-021-00498-z

Ghauri, P., Strange, R., & Cooke, F. L. (2021). Research on international business: The new realities. *International Business Review, 30*(2), 1–11. https://doi.org/10.1016/j.ibusrev.2021.101794

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin Company.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gooderham, P. N., Elter, F., Pedersen, T., & Sandvik, A. M. (2022). The digital challenge for multinational mobile network operators. More marginalization or rejuvenation? *Journal of International Management*, *28*(4), Article 100946. https://doi.org/10.1016/j.intman.2022.100946

Grant, R., & Phene, A. (2022). The knowledge based view and global strategy: Past impact and future potential. *Global Strategy Journal*, *12*(1), 3–30. https://doi.org/10.1002/gsj.1399

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619. https://doi.org/10.1126/science.1134475

Greckhamer, T., Furnari, S., Fiss, P. C., & Aguilera, R. V. (2018). Studying configurations with qualitative comparative analysis: Best practices in strategy and organization research. *Strategic Organization*, *16*(4), 482–495. https://doi.org/10.1177/1476127018786487

Grimm, C. (2021, June 18). The danger of anthropomorphic language in robotic AI systems. *Brookings*. https://www.brookings.edu/articles/the-danger-of-anthropomorphic-language-in-robotic-ai-systems/

Grover, V., & Lyytinen, K. (2015). New state of play in information systems research: The push to the edges. *MIS Quarterly*, *39*(2), 271–296. https://doi.org/10.25300/MISQ/2015/39.2.01

Guba, E. G. (1990). *The paradigm dialog*. Sage Publications.

Habib, T., Kristiansen, J. N., Rana, M. B., & Ritala, P. (2020). Revisiting the role of modular innovation in technological radicalness and architectural change of products: The case of Tesla X and Roomba. *Technovation*, *98*, Article 102163. https://doi.org/10.1016/j.technovation.2020.102163

Hasan, R., & Rana, M. (2018). *Artificial intelligence: When a machine can think for your business*. International Business Centre of Aalborg University. https://www.researchgate.net/publication/328172293_'Artificial_Intelligence'_When_A_Machine_Can_Think_For_Your_Business

Hasan, R., Thaichon, P., & Weaven, S. (2021). Are we already living with Skynet? Anthropomorphic artificial intelligence to enhance customer experience. In P. Thaichon & V. Ratten (Eds.), *Developing digital marketing* (pp. 103–134). Emerald Publishing. https://doi.org/10.1108/978-1-80071-348-220211007

Hasan, R., Weaven, S., & Thaichon, P. (2021). Blurring the line between physical and digital environment: The impact of artificial intelligence on customers' relationship and customer experience. In P. Thaichon & V. Ratten (Eds.), *Developing digital marketing* (pp. 135–153). Emerald Publishing. https://doi.org/10.1108/978-1-80071-348-220211008

Hemphill, T. A., & Kelley, K. J. (2021). Artificial intelligence and the fifth phase of political risk management: An application to regulatory expropriation. *Thunderbird International Business Review*, *63*(5), 585–595. https://doi.org/10.1002/tie.22222

Horwitz, J. (2025, August 14). Meta's AI rules have let bots hold 'sensual' chats with children. *Reuters*. https://www.reuters.com/investigates/special-report/meta-ai-chatbot-guidelines/

Hou, J. (Jove), Yang, S., Xiong, G., & Pavlou, P. A. (2026). Enhancing AI-assisted purchase decisions: The role of the sense of autonomy. *Management Information Systems Quarterly*, *50*(2), 589–614. https://doi.org/10.25300/MISQ/2025/17607

Huang, M.-H., & Rust, R. T. (2021). Engaged to a robot? The role of AI in service. *Journal of Service Research*, *24*(1), 30–41. https://doi.org/10.1177/1094670520902266

Huang, M.-H., & Rust, R. T. (2024). The caring machine: Feeling AI for customer care. *Journal of Marketing*, *88*(5), 1–23. https://doi.org/10.1177/00222429231224748

Huo, D., & Chaudhry, H. R. (2021). Using machine learning for evaluating global expansion location decisions: An analysis of Chinese manufacturing sector. *Technological Forecasting and Social Change*, *163*, Article 120436. https://doi.org/10.1016/j.techfore.2020.120436

Hutzschenreuter, T., & Lämmermann, T. (2025). What is your ai strategy? Systematically integrating self-learning technologies into your business strategy. *Academy of Management Perspectives*, *39*(4), 528-551. https://doi.org/10.5465/amp.2023.0243

IEEE. (2019). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems*. https://standards.ieee.org/industry-connections/ec/ead-v1/

IEEE. (2024). IEEE standard for ethical considerations in emulated empathy in autonomous and intelligent systems. *IEEE Std 7014-2024*, 1–51. https://doi.org/10.1109/IEEESTD.2024.10576666

International Data Corporation. (2024). *IDC's Worldwide AI and Generative AI Spending – Industry Outlook*.

https://www.idc.com/resource-center/blog/idcs-worldwide-ai-and-generative-ai-spending-industry-outlook/

Jargon, J., & Kessler, S. (2025, August 28). A troubled man, his chatbot and a murder-suicide in Old Greenwich. *Wall Street Journal (Online)*. https://www.wsj.com/tech/ai/chatgpt-ai-stein-erik-soelberg-murder-suicide-6b67dbfb?eafs_enabled=false

Jiang, H., Siponen, M., Jiang, Z., & Tsohou, A. (2024). The impacts of internet monitoring on employees' cyberloafing and organizational citizenship behavior: A longitudinal field quasi-experiment. *Information Systems Research*, *35*(3), 1175-1194. https://pubsonline.informs.org/doi/abs/10.1287/isre.2020.0216

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Johanson, J., & Vahlne, J.-E. (2009). The Uppsala internationalization process model revisited: From liability of foreignness to liability of outsidership. *Journal of International Business Studies*, *40*(9), 1411–1431. https://doi.org/10.1057/jibs.2009.24

Karahanna, E., Xu, S. X., Xu, Y., & Zhang, N. (2018). The needs–affordances–features perspective for the use of social media. *MIS Quarterly*, *42*(3), 737-A23. https://doi.org/10.25300/MISQ/2018/11492

Katsikeas, C. S., Skarmeas, D., & Bello, D. C. (2009). Developing successful trust-based international exchange relationships. *Journal of International Business Studies*, *40*(1), 132–155. https://doi.org/10.1057/palgrave.jibs.8400401

Köhler, T., Lê, J., & Smith, A. (2026). Advanced qualitative analyses for theorizing: balancing deep data engagement with imaginative theorizing. *The Journal of Applied Behavioral Science*. Advance online publication. https://doi.org/10.1177/00218863261429182

Kopalle, P. K., Gangwar, M., Kaplan, A., Ramachandran, D., Reinartz, W., & Rindfleisch, A. (2022). Examining artificial intelligence (AI) technologies in marketing via a global lens: Current trends and future research opportunities. *International Journal of Research in Marketing*, *39*(2), 522–540. https://doi.org/10.1016/j.ijresmar.2021.11.002

Kriz, A., Rumyantseva, M., & Welch, C. (2023). When does the internationalization process begin? Problematizing temporal boundaries in international business. *Journal of World Business*, *58*(3), Article 101423. https://doi.org/10.1016/j.jwb.2022.101423

Leavitt, K., Schabram, K., Hariharan, P., & Barnes, C. M. (2021). Ghost in the machine: On organizational theory in the age of machine learning. *Academy of Management Review*, *46*(4), 750–777. https://doi.org/10.5465/amr.2019.0247

Lee, J. M., Narula, R., & Hillemann, J. (2021). Unraveling asset recombination through the lens of firm-specific advantages: A dynamic capabilities perspective. *Journal of World Business*, *56*(2), Article 101193. https://doi.org/10.1016/j.jwb.2021.101193

Leong, B., & Selinger, E. (2019). Robot eyes wide shut: Understanding dishonest anthropomorphism. *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*, 299–308. https://doi.org/10.1145/3287560.3287591

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Sage Publications.

Lindner, T., Puck, J., & Puhr, H. (2025). Artificial intelligence in international business: IB theory under augmented decision-making. *Journal of World Business*, *60*(6), Article 101676. https://doi.org/10.1016/j.jwb.2025.101676

Lundvall, B., & Rikap, C. (2022). China's catching-up in artificial intelligence seen as a co-evolution of corporate and national innovation systems. *Research Policy*, *51*(1), Article 104395. https://doi.org/10.1016/j.respol.2021.104395

Luo, Y. (2022). A general framework of digitization risks in international business. *Journal of International Business Studies*, *53*(2), 344–361. https://doi.org/10.1057/s41267-021-00448-9

Luostarinen, R. (Ed.). (1994). *Internationalization of Finnish firms and their response to global challenges*. UNU World Institute for Development Economics Research. https://doi.org/10.22004/ag.econ.295309

Lyytinen, K., Nickerson, J., & King, J. (2021). Metahuman systems = humans plus machines that learn. *Journal of Information Technology*, *36*(4), 427–445. https://doi.org/10.1177/0268396220915917

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25. https://doi.org/10.1086/392759

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293-A5. https://doi.org/10.2307/23044045

Mahadevan, J., & Moore, F. (2023). A framework for a more reflexive engagement with ethnography in International Business Studies. *Journal of World Business*, *58*(4), Article 101424. https://doi.org/10.1016/j.jwb.2022.101424

Malik, A., Budhwar, P., Mohan, H., & Srikanth, N. R. (2023). Employee experience –the missing link for engaging employees: Insights from an MNE's AI-based HR ecosystem. *Human Resource Management*, *62*(1), 97–115. https://doi.org/10.1002/hrm.22133

Marano, V., Wilhelm, M., Kostova, T., Doh, J., & Beugelsdijk, S. (2024). Multinational firms and sustainability in global supply chains: Scope and boundaries of responsibility. *Journal of International Business Studies*, *55*(4), 413–428. https://doi.org/10.1057/s41267-024-00706-6

Marsden, R. (2025, April 8). How do we really feel about robots? *Financial Times.* https://www.ft.com/content/31d94b0b-b769-4c44-b25d-f39626d2de4a

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955, August 31). *A proposal for the dartmouth summer research project on artificial intelligence*. https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

Menzies, J., Sabert, B., Hassan, R., & Mensah, P. K. (2024). Artificial intelligence for international business: Its use, challenges, and suggestions for future research and practice. *Thunderbird International Business Review*, *66*(2), 185–200. https://doi.org/10.1002/tie.22370

Messner, W. (2022a). Advancing our understanding of cultural heterogeneity with unsupervised machine learning. *Journal of International Management*, *28*(2), Article 100885. https://doi.org/10.1016/j.intman.2021.100885

Messner, W. (2022b). Cultural differences in an artificial representation of the human emotional brain system: A deep learning study. *Journal of International Marketing, 30*(4), 21–43. https://doi.org/10.1177/1069031X221123993

Messner, W. (2022c). Cultural heterozygosity: Towards a new measure of within-country cultural diversity. *Journal of World Business*, *57*(4), Article 101346. https://doi.org/10.1016/j.jwb.2022.101346

Meyer, K. E., Li, J., Brouthers, K. D., & Jean, R.-J. '"Bryan."' (2023). International business in the digital age: Global strategies in a world of national institutions. *Journal of International Business Studies*, *54*(4), 577–598. https://doi.org/10.1057/s41267-023-00618-x

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage Publications.

Mingers, J., Mutch, A., & Willcocks, L. (2013). Critical realism in information systems research. *MIS Quarterly*, *37*(3), 795–802.

Montiel, I., Cuervo-Cazurra, A., Park, J., Antolín-López, R., & Husted, B. W. (2021). Implementing the United Nations' Sustainable Development Goals in international business. *Journal of International Business Studies*, *52*(5), 999–1030. https://doi.org/10.1057/s41267-021-00445-y

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Murray, A., Rhymer, J., & Sirmon, D. G. (2021). Humans and technology: Forms of conjoined agency in organizations. *Academy of Management Review*, *46*(3), 552–571. https://doi.org/10.5465/amr.2019.0186

Nair, A., Hanvanich, S., & Tamer Cavusgil, S. (2007). An exploration of the patterns underlying related and unrelated collaborative ventures using neural network: Empirical investigation of collaborative venture formation data spanning 1985-2001. *International Business Review*, *16*(6), 659–686. https://doi.org/10.1016/j.ibusrev.2007.08.005

Nakanishi, J., Kuramoto, I., Baba, J., Ogawa, K., Yoshikawa, Y., & Ishiguro, H. (2019). Soliloquising social robot in a hotel room. *ACM International Conference Proceeding Series*, 21–29. https://doi.org/10.1145/3369457.3370913

Nambisan, S., & Luo, Y. (2021). Toward a loose coupling view of digital globalization. *Journal of International Business Studies*, *52*(8), 1646–1663. https://doi.org/10.1057/s41267-021-00446-x

Norwegian Consumer Council. (2023). *Ghost in the machine – Addressing the consumer harms of generative AI*. https://www.forbrukerradet.no/side/new-report-generative-ai-threatens-consumer-rights/

Ojala, A., Evers, N., & Rialp, A. (2018). Extending the international new venture phenomenon to digital platform providers: A longitudinal case study. *Journal of World Business*, *53*(5), 725–739. https://doi.org/10.1016/j.jwb.2018.05.001

Ojala, A., Fraccastoro, S., & Gabrielsson, M. (2023). Characteristics of digital artifacts in international endeavors of digital-based international new ventures. *Global Strategy Journal*, *13*(4), 857–887. https://doi.org/10.1002/gsj.1483

Oliva, F. L., Teberga, P. M. F., Testi, L. I. O., Kotabe, M., Giudice, M. D., Kelle, P., & Cunha, M. P. (2022). Risks and critical success factors in the internationalization of born global startups of industry 4.0: A social, environmental, economic, and institutional analysis. *Technological Forecasting and Social Change*, *175*, Article 121346. https://doi.org/10.1016/j.techfore.2021.121346

Orton, J. D., & Weick, K. E. (1990). Loosely coupled systems: A reconceptualization. *The Academy of Management Review*, *15*(2), 203–223. https://doi.org/10.2307/258154

Oxford English Dictionary. (n.d.). *Artificial intelligence*. Retrieved April 25, 2025, from https://www.oed.com/dictionary/artificial-intelligence_n

Ozturk, A., Cavusgil, S. T., & Ozturk, O. C. (2021). Consumption convergence across countries: Measurement, antecedents, and consequences. *Journal of International Business Studies*, *52*(1), 105–120. https://doi.org/10.1057/s41267-020-00334-w

Pan, S.-L., Nishant, R., Tuunanen, T., & Choudrie, J. (2025). AI agents: Potential implications for IS research? *The Journal of Strategic Information Systems*, *34*(2), 101906. https://doi.org/10.1016/j.jsis.2025.101906

Parkes, D. C., & Wellman, M. P. (2015). Economic reasoning and artificial intelligence. *Science*, *349*(6245), 267–272. https://doi.org/10.1126/science.aaa8403

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Sage Publications.

Placani, A. (2024). Anthropomorphism in AI: Hype and fallacy. *AI and Ethics*, *4*(3), 691–698. https://doi.org/10.1007/s43681-024-00419-4

Porter, M. E., & Kramer, M. R. (2011). Creating Shared Value. *Harvard Business Review*, *89*(1-2 (January – February)), 62–77.

Prahalad, C. K., & Doz, Y. L. (1987). *The multinational mission: Balancing local demands and global vision*. Free Press.

Raes, A. M. L., Heijltjes, M. G., Glunk, U., & Roe, R. A. (2011). The interface of the top management team and middle managers: A process model. *Academy of Management Review*, *36*(1), 102–126. https://doi.org/10.5465/amr.2009.0088

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C.,

Pentland, A. S., … Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Ramani, R. S., & Aguinis, H. (2023). Using field and quasi experiments and text-based analysis to advance international business theory. *Journal of World Business*, *58*(5), Article 101463. https://doi.org/10.1016/j.jwb.2023.101463

Ramaul, L., Ritala, P., Kostis, A., & Aaltonen, P. (2025). Rethinking how we theorize AI in organization and management: A problematizing review of rationality and anthropomorphism. *Journal of Management Studies*, *63*(2), 761-807. https://doi.org/10.1111/joms.13246

Ratten, V., Hasan, R., Kumar, D., Bustard, J., Ojala, A., & Salamzadeh, Y. (2024). Learning from artificial intelligence researchers about international business implications. *Thunderbird International Business Review*, *66*(2), 211–219. https://doi.org/10.1002/tie.22374

Rice, G. (1999). Islamic ethics and the implications for business. *Journal of Business Ethics*, *18*(4), 345–358. https://doi.org/10.1023/A:1005711414306

Rikap, C. (2022). Amazon: A story of accumulation through intellectual rentiership and predation. *Competition and Change*, *26*(3–4), 436–466. https://doi.org/10.1177/1024529420932418

Rugman, A. M., & Verbeke, A. (2003). Extending the theory of the multinational enterprise: Internalization and strategic management perspectives. *Journal of International Business Studies*, *34*(2), 125–137. https://doi.org/10.1057/palgrave.jibs.8400012

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson International.

Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2026). AI agents vs. agentic AI: A Conceptual taxonomy, applications and challenges. *Information Fusion*, *126*, Article 103599. https://doi.org/10.1016/j.inffus.2025.103599

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *Management Information Systems Quarterly*, *43*(3), 695–719. https://doi.org/10.25300/MISQ/2019/13747

Schlegelmilch, B. B., Sharma, K., & Garg, S. (2023). Employing machine learning for capturing COVID-19 consumer sentiments from six countries: A methodological illustration. *International Marketing Review*, *40*(5), 869–893. https://doi.org/10.1108/IMR-06-2021-0194

Schleidgen, S., Friedrich, O., Gerlek, S., Assadi, G., & Seifert, J. (2023). The concept of "interaction" in debates on human–machine interaction. *Humanities and Social*

Sciences Communications, *10*(1), 1–13. https://doi.org/10.1057/s41599-023-02060-8

Schotter, A. P. J., Mudambi, R., Doz, Y. L., & Gaur, A. (2017). Boundary spanning in global organizations. *Journal of Management Studies*, *54*(4), 403–421. https://doi.org/10.1111/joms.12256

Schumann, J. H., Wangenheim, F. v., Stringfellow, A., Yang, Z., Praxmarer, S., Jiménez, F. R., Blazevic, V., Shannon, R. M., G., S., & Komor, M. (2010). Drivers of trust in relational service exchange: Understanding the importance of cross-cultural differences. *Journal of Service Research*, *13*(4), 453–468. https://doi.org/10.1177/1094670510368425

Shanahan, M. (2024). Talking about large language models. *Communications of the ACM*, *67*(2), 68–79. https://doi.org/10.1145/3624724

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems*, *10*(4), 26:1-26:31. https://doi.org/10.1145/3419764

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y

Silverman, D. (1998). Qualitative research: Meanings or practices? *Information Systems Journal*, *8*(1), 3–20. https://doi.org/10.1046/j.1365-2575.1998.00002.x

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*(6), 467–482.

Simon, H. A. (2000). Bounded rationality in social science: Today and tomorrow. *Mind & Society*, *1*(1), 25–39. https://doi.org/10.1007/BF02512227

Singla, A., Sukharevsky, A., Yee, L., Chui, M., & Hall, B. (2025). *The state of AI: How organizations are rewiring to capture value*. McKinsey. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/

Siponen, M., Lanamäki, A., Nathan, M., & Klaavuniemi, T. (2025). Mechanism-based explanations and theoretical contribution in IS research. *Communications of the Association for Information Systems*, *57*(1), 458-475. https://doi.org/10.17705/1CAIS.05718

Skilton, M., & Hovsepian, F. (2017). *The 4th Industrial Revolution: Responding to the Impact of artificial intelligence on business*. Springer.

Sobolev, D. (2023). Rise of the Androids: The reflection of developers' characteristics in computerized systems. *British Journal of Management*, *34*(3), 1632–1654. https://doi.org/10.1111/1467-8551.12653

Statista Market Insights. (2024). *Artificial intelligence—worldwide*. https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide

Stelmaszak, M., Joshi, M., & Constantiou, I. (2025). Artificial intelligence as an organizing capability arising from human-algorithm relations. *Journal of Management Studies*. Advance online publication. https://doi.org/10.1111/joms.70003

Stryker, C. (2025, February 24). *What is agentic AI?* IBM. https://www.ibm.com/think/topics/agentic-ai

Tatarinov, K., Ambos, T. C., & Tschang, F. T. (2023). Scaling digital solutions for wicked problems: Ecosystem versatility. *Journal of International Business Studies*, *54*(4), 631–656. https://doi.org/10.1057/s41267-022-00526-6

Thompson, N. C., Ge, S., & Sherry, Y. M. (2021). Building the algorithm commons: Who discovered the algorithms that underpin computing in the modern enterprise? *Global Strategy Journal*, *11*(1), 17–33. https://doi.org/10.1002/gsj.1393

Tóth, Z., Caruana, R., Gruber, T., & Loebbecke, C. (2022). The dawn of the AI robots: Towards a new framework of AI robot accountability. *Journal of Business Ethics*, *178*(4), 895–916. https://doi.org/10.1007/s10551-022-05050-z

Tractinsky, N., & Jarvenpaa, S. L. (1995). Information systems design decisions in a global versus domestic context. *MIS Quarterly*, *19*(4), 507–534. https://doi.org/10.2307/249631

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Tuunanen, T., Winter, R., & vom Brocke, J. (2024). Dealing with complexity in design science research: A methodology using design echelons. *MIS Quarterly*, *48*(2), 427–458. https://doi.org/10.25300/MISQ/2023/16700

United Nations. (2023). *The 17 Goals*. https://sdgs.un.org/goals

Vahlne, J.-E., & Johanson, J. (2017). From internationalization to evolution: The Uppsala model at 40 years. *Journal of International Business Studies*, *48*(9), 1087–1102. https://doi.org/10.1057/s41267-017-0107-7

van den Broek, E., Sergeeva, A., & Huysman, M. (2025). Is There Fairness in AI? *Journal of Management Studies*. Advance online publication. https://doi.org/10.1111/joms.13276

Väyrynen, K., Lanamäki, A., Laari-Salmela, S., Iivari, N., & Kinnula, M. (2025). Unpacking the regulatory ambiguity mechanism: Implications for industry-level digital transformation. *Information Systems Journal*, *35*(6), 1528–1564. https://doi.org/10.1111/isj.12595

Veiga, J. F., Lubatkin, M., Calori, R., Very, P., & Tung, Y. A. (2000). Using neural network analysis to uncover the trace effects of national culture. *Journal of International Business Studies*, *31*(2), 223–238 https://doi.org/10.1057/palgrave.jibs.8490903

Verbeke, A. (2026). The future of international business strategy research. *International Business Review*, *35*(4), Article 102596. https://doi.org/10.1016/j.ibusrev.2026.102596

Verbeke, A., & Hutzschenreuter, T. (2021). The dark side of digital globalization. *Academy of Management Perspectives*, *35*(4), 606–621. https://doi.org/10.5465/amp.2020.0015

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-019-14108-y

Weko, S. (2025). New sites of accumulation? Why intangible assets matter for energy transitions. *Review of Political Economy*, *37*(4), 1571–1598. https://doi.org/10.1080/09538259.2024.2391304

Welch, C., Paavilainen-Mäntymäki, E., Piekkari, R., & Plakoyiannaki, E. (2022). Reconciling theory and context: How the case study can set a new agenda for international business research. *Journal of International Business Studies*, *53*(1), 4–26. https://doi.org/10.1057/s41267-021-00484-5

Welch, C., & Piekkari, R. (2017). How should we (not) judge the 'quality' of qualitative research? A re-assessment of current evaluative criteria in International Business. *Journal of World Business*, *52*(5), 714–725. https://doi.org/10.1016/j.jwb.2017.05.007

Welch, C., Piekkari, R., Plakoyiannaki, E., & Paavilainen-Mäntymäki, E. (2011). Theorising from case studies: Towards a pluralist future for international business research. *Journal of International Business Studies*, *42*(5), 740–762. https://doi.org/10.1057/jibs.2010.55

Wijaya, I. F., Moro, A., & Belghitar, Y. (2023). Trust in Islamic business-to-business relationships: Evidence from Indonesia. *British Journal of Management*, *34*(1), 111–128. https://doi.org/10.1111/1467-8551.12584

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management*, *29*(5), 907–931. https://doi.org/10.1108/JOSM-04-2018-0119

Wood, M. A. (2021). Rethinking how Technologies Harm. *The British Journal of Criminology*, *61*(3), 627–647. https://doi.org/10.1093/bjc/azaa074

Wood, M. A., Pervan, F., Lundberg, K., Mitchell, M., & Wood, J. (2025). Doing harm with things: Making sense of intention in harmful actions using technology. *Critical Criminology, 33*, 321–338. https://doi.org/10.1007/s10612-025-09819-2

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Sage Publications.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research commentary—The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research*, *21*(4), 724–735. https://doi.org/10.1287/isre.1100.0322

Yoon, H. D., Belkhouja, M., & Dau, L. A. (2025). Privacy protection laws, national culture, and artificial intelligence innovation around the world. *Journal of International Business Studies*, *56*(7), 853–873. https://doi.org/10.1057/s41267-025-00790-2

Yu, H., Fletcher, M., & Buck, T. (2022). Managing digital transformation during re-internationalization: Trajectories and implications for performance. *Journal of International Management*, *28*(4), Article 100947. https://doi.org/10.1016/j.intman.2022.100947

Zaheer, S. (1995). Overcoming the liability of foreignness. *Academy of Management Journal*, *38*(2), 341–363. https://doi.org/10.5465/256683

Zheng, J., & Jarvenpaa, S. (2021). Thinking technology as human: Affordances, technology features, and egocentric biases in technology anthropomorphism. *Journal of the Association for Information Systems*, *22*(5), 1429–1453. https://doi.org/10.17705/1jais.00698

Zhou, Y., Fei, Z., He, Y., & Yang, Z. (2022). How human–chatbot interaction impairs charitable giving: The role of moral judgment. *Journal of Business Ethics*, *178*(3), 849–865. https://doi.org/10.1007/s10551-022-05045-w

Złotowski, J., Khalil, A., & Abdallah, S. (2020). One robot doesn't fit all: Aligning social robot appearance and job suitability from a Middle Eastern perspective. *AI & Society*, *35*(2), 485–500. https://doi.org/10.1007/s00146-019-00895-x

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

Appendices

Appendix 1. Conducting Qualitative Interviews

INTERVIEW GUIDE

Introductory script: To begin with, thank you so much for finding time for us. How are you doing today? What time is it in your geolocation, and how was/is the weather today? By the way, how is the COVID situation in your area/country? As we have already mentioned earlier, our research spins around the humanness of AI or humanized automated and intelligent systems such as robots, voice-enabled digital assistants, chatbots, autonomous vehicles, and avatars. We are interested in your story, experience, thoughts, and whatever you want to talk about, including ethical issues in human-machine interaction. We are particularly interested to hear your thoughts on the humanlike characteristics of AI systems, such as social behavioral or functional behavioral (e.g., task, competence), and how they can create ethical issues, affect or alter users or consumer behavior, or vice versa.

The research emphasizes three service encounter (interaction) perspectives:

a) Users or consumers interact with an AI system that represents the service provider.

b) Users or consumers interact with an AI system that is embedded in the product/service.

c) Users or consumers interact with an AI system as an end product/service.

To facilitate our note-taking and transcription, we would like to record our audio conversations today. Would you kindly allow us to do so?
"By agreeing to participate, you will be confirming that:

- You understand that your participation is voluntary and that you are free to withdraw at any time, without explanation or penalty; and
- You understand that what participation in this research entails, agree to our conversation being recorded for the purpose of transcription, and the use of your de-identified comments, views, and thoughts in the research output."

Thank you for agreeing to participate. Shall we begin?

Closing scripts: As we have reached half an hour, we would like to end our conversation today. We want to thank you a lot for providing your valuable time and rich insights. Do you have any further remarks on any additional ethical concerns of AI that we may have missed in the discussion? Do you have any comments or feedback on our research topic or your interview experience? Thank you once again. We will keep in touch. Kindly stay safe. Have a nice day! Bye.

Time:
Date:
Place:
Name of interviewer:
Name of interviewee:

Demographics info
Gender:
Age:
Nationality:
Occupation:
Highest degree:
Years using (or of expertise on) AI:
AI use intensity:
Type of AI:
Context of use (or development and deployment):

Post interview comments and/or observations:

Interview questions (Interview type: Users or consumers)

1. Please tell me your experiences about recent interactions with any automated and intelligent systems such as robots, voice-enabled digital assistants, chatbots, autonomous vehicles, or avatars in the digital environment.
2. Tell me your thoughts about the humanness perspective of AI.
   Probe: When we say humanness, we mean humanlike traits or characteristics that you may observe or find in the AI system that you use, such as the ability to express emotion or independent decision making, or that AI may have its own intention and autonomy.
3. Please tell me your view on the different types of ethical tensions that may stem from the interaction between consumers and humanized AI.
4. Give me your viewpoint on what are or could be intended and unintended consequences that may stem from using or interacting with automated and intelligent systems.
5. Would you kindly share your privacy concerns regarding rising automated and intelligent systems in service?
6. Would you describe any of your experiences where you felt that an AI system could manipulate you, or could be used to convince you to do certain things, or could be used to persuade you to buy products or services, or do more?
7. Tell me any concerns you may have about the emotional or relational perspectives of human-AI interaction, and how they could bring negative consequences for consumer well-being.
8. Some people say AI should morally be responsible for its acts, actions or decisions, like a human being. Some people also say that AI should be treated as a slave and should not be treated as a morally responsible agent. What are your opinions about this, and the rights of AI systems?
   Probe: If an autonomous car killed or harmed someone on the road, who should we blame?
9. Tell me about AI-enabled automated decision-making and what type of bias or discrimination it can create.
10. There is an ongoing debate about AI-enabled dating services, sex toys, or even services related to sex with a robot. Please tell me about your view on this, connected to ethical concerns.
11. Please tell me about your opinions on how companies/policymakers can minimize the risk or harm of the AI system and protect consumer well-being.
12. What could be the key strategies that firms should adopt to enable responsible/ethical AI to create harmless human-AI interaction and foster positive customer engagement?

Interview questions (Interview type: Experts)

1. Please tell me your thoughts about the humanness perspective of automated and intelligent systems such as robots, voice-enabled digital assistants, chatbots, autonomous vehicles, and avatars in the digital environment.
   Probe: When I say humanness, I mean the applied anthropomorphism, humanlike traits, or characteristics that a user or consumer may observe or perceive consciously or unconsciously in an AI system.
2. Tell me your opinion on AI's capability concerning machine behavior, and what the key dimensions of such machine behavior are.
   Probe: AI's ability to express emotion or independent decision making; ability to demonstrate action in a way where people may perceive AI's intention, agency, or autonomy.
3. Would you kindly share your thoughts on how machine behavior can influence consumer behavior, and how consumer behavior can also influence machine behavior?
4. Tell me about your view on different types of ethical tensions that may stem from human-AI interaction.
5. Tell me your perspective on what are or could be intended and unintended consequences that may come from using or interacting with automated and intelligent systems?
6. Would you share your privacy concerns regarding rising automated and intelligent systems?
7. Would you kindly describe how an AI system can be deployed to manipulate consumers, or enable social engineering or persuasive technology to buy products or services?
8. Please express your concerns on the emotional or relational perspectives of human-AI interaction, and how they could bring negative consequences for consumer well-being, such as emotional attachment or social isolation.
9. Some experts say AI should be morally responsible for any of its acts, actions or decisions, like a human being. Some experts also say that AI should be treated as a slave and should not be treated as a morally responsible agent. What are your opinions about this, and the rights of AI systems?
   Probe: If an autonomous car killed or harmed someone on the road, who should we blame?
10. Tell me about AI-enabled automated decision-making and what type of bias or discrimination it can create.
11. There is an ongoing debate about AI-enabled dating services, sex toys, or even services related to sex with a robot. Please tell me about your view on this, connected to ethical concerns.
12. Tell me about your opinions on how we can minimize the risk of harm to the AI system and protect consumer well-being.
13. What could be the key strategies that firms should adopt to enable responsible AI to create harmless human-AI interaction and foster positive customer engagement?

DOI: 10.1002/tie.22369

RESEARCH ARTICLE

WILEY

Managing artificial intelligence in international business: Toward a research agenda on sustainable production and consumption

Rakibul Hasan[1] | **Arto Ojala**[1,2]

[1]School of Marketing and Communication, University of Vaasa, Vaasa, Finland

[2]Graduate School of Management, Kyoto University, Kyoto, Japan

Correspondence
Rakibul Hasan, School of Marketing and Communication, University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland.
Email: rakibul.hasan@uwasa.fi

Funding information
The research is supported by Finland Fellowship, funded by the Finnish Ministry of Education and Culture.

Abstract

The collaboration between artificial intelligence (AI) and humans is reshaping international business (IB) management dynamics, aiming to achieve global sustainable development. Recent IB literature indicates that managing AI brings benefits such as better resource reconfiguration, reduced transaction costs, and global sustainable development. However, existing IB literature provides only meager knowledge about the characteristics of AI and how these characteristics can be employed for international expansion at the intersection of sustainable development. In response, our aim is to construct these characteristics by employing directed qualitative content analysis of empirical AI research. Based on our three constructed characteristics of AI, we contribute to current IB literature by providing a framework to balance economic and social goals and utilizing AI for global sustainable development. Further, we provide future IB research themes to guide IB and AI research toward achieving a sustainable production and consumption agenda.

KEYWORDS
artificial intelligence (AI), global sustainable development, international business, international expansion, sociotechnical theory

1 | INTRODUCTION

The role of artificial intelligence (AI)[1] is becoming an everyday phenomenon in the value creation of new business activities (Chalmers et al., 2021; Murray et al., 2021; Ratten, 2022). Even though AI is not a recent phenomenon in international business (IB) (Brouthers et al., 2009; Cavusgil et al., 1992; Hemphill & Kelley, 2021; Tatarinov et al., 2023; Veiga et al., 2000), there remains an uncertainty about whether AI will enable or hinder responsible consumption and production worldwide in line with the United Nations' (2023) Sustainable Development Goals (SDGs). This question aligns with the ethical guidelines for trustworthy AI that empowers human beings to make informed decisions and protect fundamental rights (European Commission, 2019). Such a question also relates to the Fifth Industrial Revolution that emphasizes advanced technologies, including AI, for human-centric and sustainable development (European Commission, 2022). This human-centric approach encompasses complementarity between humans and machines and aims to ensure the well-being of society, business, and end users.

For example, managing embedded AI in autonomous electric vehicles with a new business model to enable monthly subscriptions could combine public transport, and reduce car ownership costs and carbon footprint. Deploying virtual AI like avatars and human counterparts responsibly can also be used to protect end user privacy while improving service quality in healthcare. Aligning with these examples, a recent AI research project focusing on sustainable development found

that while AI can accomplish targets across all the SDGs, it can also inhibit them (Vinuesa et al., 2020). Therefore, the research urges extending a systematic understanding of AI to the international context. This call motivated a focus on IB management in our study.

By taking firm-level management perspectives, we recognize the importance of extending AI research to IB literature. This urgency is paramount to achieving global development in balancing the three pillars of sustainable development—economic, societal, and environmental well-being. Against such a backdrop, recent IB literature on digitalization recognizes that firms need to balance their social, technological, and economic goals (Luo & Zahra, 2023; Verbeke & Hutzschenreuter, 2021). Similarly, IB scholars also point toward AI at the intersection of global sustainable development (Ciulli & Kolk, 2023; Tatarinov et al., 2023). However, IB literature contains only limited knowledge about the characteristics of AI and how managers can utilize those characteristics in their international expansion activities. On top of that, we know very little about AI in IB management at the intersection of sustainable development. We argue that AI can enable responsible consumption and production worldwide and help achieve SDGs but only if we have effective management strategies, policies, and institutions to support its use. Otherwise, there can be opposite, deleterious effects, hindering SDGs.

In response, our research focuses on SDG #12, which deals with sustainable consumption and production patterns in IB management (Chabowski et al., 2023; Montiel et al., 2021). To advance research in this domain, our study provides a new understanding that preliminarily builds on the management perspective of AI, which is developed mainly in information systems literature. We embrace a sociotechnical viewpoint of AI that aims to balance social and economic goals (Berente et al., 2021). The following research question is of particular interest in this study: *How does the existing empirical research on artificial intelligence relate to international business management?* To answer this question, we employ a qualitative content analysis method to analyze published empirical research on AI. We then construct three characteristics of AI and demonstrate how managers can utilize those characteristics in their international expansion activities at the intersection of SDG #12. Our study analyzes 85 empirical research papers on AI published in multiple disciplines with a critical realism view and a directed content analysis approach.

By doing so, our research contributes to the growing AI research in IB literature in four ways (Autio et al., 2021; Ciulli & Kolk, 2023; Del Giudice et al., 2023; Denicolai et al., 2021; Tatarinov et al., 2023). First, we construct three characteristics of AI–autonomy, learning, and combinative. Second, we show how managers can utilize those characteristics to achieve economic goals in international expansion activities. Third, we contribute to the open and multidisciplinary domain of AI research that focuses on global sustainable development (Jobin et al., 2019; Rahwan et al., 2019; Vinuesa et al., 2020). Fourth, we propose further research directions to initiate and guide research on AI in the context of sustainable production and consumption in IB that balance economic and social goals.

The paper is organized as follows. In the next section, we discuss the method of the study, which builds upon a directed content analytical approach. Then, we unpack the theoretical foundation of constructing the three characteristics of AI. Afterward, we explain the significant insights gained from our analysis. We also outline future IB research opportunities in AI related to the sustainable production and consumption agenda. Finally, we provide managerial implications and discuss the limitations of our study to guide future developments.

2 | DIRECTED CONTENT ANALYSIS OF PUBLISHED AI RESEARCH

Since the study embodies our knowledge, assumptions, beliefs, and value judgments about AI and responsible production and consumption patterns (SGD #12), we took a critical realism position (Bhaskar, 1978; Mingers, 2004). This philosophical stance allowed us the flexibility to conduct qualitative research using multiple criteria and procedures in IB beyond qualitative positivism and proceduralism (Welch & Piekkari, 2017). Consequently, our study embraced the inherent subjectivity in observations, transparency, and reflexivity to construct an understanding of AI in the context of IB management. Instead of using a methodological template, we employed our heuristics and imagination in the context-laden phenomenon, which is paramount for management and AI research (Furnari et al., 2021; Turing, 1950). Accordingly, we maintained that "a quality study results from the researcher's ability to reflect on how his or her field interactions, philosophical commitments, and theoretical preconceptions molded the interpretations and results" (Welch & Piekkari, 2017, p. 720). Reflecting on embodied value judgments about knowledge production, we took a pluralist approach to IB research and sociotechnical thinking that shaped our theoretical preconceptions. Hence, we aimed to advance the field concerning AI by taking an alternative method to investigate already published articles.

Instead of the systematic literature review method, which seems to embody qualitative positivism (see Evers et al., 2023, for an example), we employed a qualitative content analysis method. Our methodological approach aligns with our critical realism philosophical position and may be described as a directed qualitative content analysis method. Accordingly, we focused on the textual contents (as data) of empirical research at the intersection of AI and the international activities of firms. Within the qualitative content analysis methods (Hsieh & Shannon, 2005), we employed a directed content analysis approach to extend our current understanding of AI in IB management, which changes the dynamics of assessing the quality of our research (see Bonache, 2021; Welch & Piekkari, 2017). This approach is deemed the most appropriate to draw insights from extant scholarly works when little direct knowledge is available on the given phenomenon (Krippendorff, 2018). To achieve this, we used four subsequent stages—(1) developing the unit of analysis, (2) deciding on appropriate sampling, (3) constructing the dataset, and (4) analyzing data.

First, we developed a unit of analysis of our study that aligns with the research question. Given that we know little about AI in the international context in general and there is limited AI research in management and IB literature. Against this backdrop, our study is directed

toward IB management that intersects with a sociotechnical viewpoint to manage AI (Berente et al., 2021). Put simply, we took insights from already published AI articles, molded those insights to construct characteristics of AI, and contextualized them into IB management to produce knowledge. We aimed to conceptually extend the current understanding of managing AI in IB operations toward achieving SDG #12, which balances economic and social goals. Consequently, the unit of analysis of our study is the characteristics of AI and managing those characteristics in IB to achieve responsible production and consumption. This allowed the construction of a fresh conceptual understanding of the given phenomenon based on empirical insights (Krippendorff, 2018; Welch et al., 2011).

Second, regarding directed content analysis, a central task of any textual analysis is to decide on the appropriate sample and which text to investigate. Our selection of published AI research followed qualitative relevance (purposive) sampling, which aims to select textual data in line with the unit of analysis (Krippendorff, 2018). Given that both AI[1] (See footnote 1) and IB as research domains capture multidisciplinary phenomena, we looked for articles in multiple disciplines, including IB, management, organizational studies, computer science, engineering, robotics, information systems, philosophy, and marketing. Consequently, this pluralistic approach allowed us to purposefully construct a dataset focused on existing AI research related to cross-border management in IB operations.

Third, we constructed a dataset.[2] Focusing on AI and IB, we initially used Google Search and Google Scholar to understand the phenomenon better. Accordingly, we aimed to include only quality scholarly works on AI with a predefined sampling procedure (see Figure 1). This sampling procedure guided us to develop relevant search strings and use advanced search functionalities in each database (see Appendix A). We limited our search to the English language and publishing dates from 1950 to 2023. Since we systematically know very little about AI in the international context, we wanted to cover an extensive period from the inception of intelligent machines (Turing, 1950) to recent developments in IB (Ciulli & Kolk, 2023; Tatarinov et al., 2023). We retrieved peer-reviewed journals listed in the Academic Journal Guide 2021 (levels 1–4*) and referred to CWTS Journal Indicators (www.journalindicators.com) for those not listed. Additionally, we took further precautions to exclude potential predatory and low-quality journals. To exclude low-quality journals, we excluded those listed as "0" or indicated as possible predatory journals. This cross-checking was complemented by two national journal rankings: the Finnish Publication Forum (level 1–3 journals) and the Norwegian Register for Scientific Journals, Series, and Publishers (level 1–2 journals). We then screened and constructed a dataset using Zotero referencing software to examine textual data.

Concerning which text to investigate to construct the dataset, we screened records based on content, including those focused mainly on AI, that must explicitly or implicitly relate to cross-border management perspectives. During this process, records in IB journals discussing AI as part of broader digitalization were considered. Additionally, we included technical papers that deal with developing new, or optimizing existing, AI models and methodological papers that deal with AI methods. However, we excluded technical and methodological papers that do not necessarily relate to firm-level management aspects. Furthermore, we included more records using cross-reference techniques. Hence, we developed our preliminary dataset comprising of 141 records with both non-empirical elements (conceptual, literature review, and technology explanation papers) and empirical elements as part of the broader research project. However, we finally constructed our dataset containing 85 records with only empirical elements. This includes methodological studies with empirical analysis (see Veiga et al., 2000), technical papers with empirical research (see Fish & Ruby, 2009), and traditional empirical papers (see Denicolai et al., 2021). After finalizing the dataset consisting of empirical AI research, we broadly sorted all empirical articles based on the emerging underlying characteristics of AI, inspired by both existing AI (Berente et al., 2021; Rahwan et al., 2019) and IB literature (Ojala et al., 2018, 2023; Tatarinov et al., 2023).

Finally, we analyzed the textual data. In proceeding with our analysis with a critical realism view and directed content analytical approach, our analytical process is neither untainted, inductive nor deductive (Krippendorff, 2018). Instead, in this directed approach, we constantly went back and forth in the literature (theory) while carefully reading and analyzing data in constructed datasets through successive iterations between theory and data (Bhaskar, 1978; Mingers, 2004). Put simply, our reports primarily relied on empirical insights from articles. However, we constantly directed those insights to existing AI and IB research through comparisons. After sorting the articles, we conducted a qualitative content comparison inspired by IB and management literature (Glikson & Woolley, 2020; Rana & Morgan, 2019; Welch et al., 2011). We read each article through carefully to understand the theoretical background, underlying assumptions, research strategy, findings, and discussion. However, we only focused on content that was derived from empirical insights (mainly the findings and discussion) to compare and relate to the focal unit of analysis. Therefore, we did not analyze other aspects of the articles such as conceptual background or hypothesis development. We then analyzed the texts through comparison to construct data-driven insights while embodying our value judgment, heuristics, and imagination.

We then categorized each paper based on the insights of textual data using Zotero's "Call Number" function to label relevant constructs, variables, or outcomes, which helped us to be systematic in constructing emerging themes and subthemes. We used an intra-content comparison to compare between parts of a text (e.g., paragraphs in the findings and discussion section of a single article) to recognize patterns for identifying similar and inconsistent themes. In addition, we compared different texts among articles and the textual content with a standard dimension (our emerging characteristics of AI and dimensions of IB management). As a result, the content comparison embodied our interpretation of relating repeated and irregular patterns to the extant theoretical assumptions in IB literature coupled with the sociotechnical thinking of managing AI in IB operations. In this process, we reflected on the following questions for each article to conduct the content analysis systematically and consistently

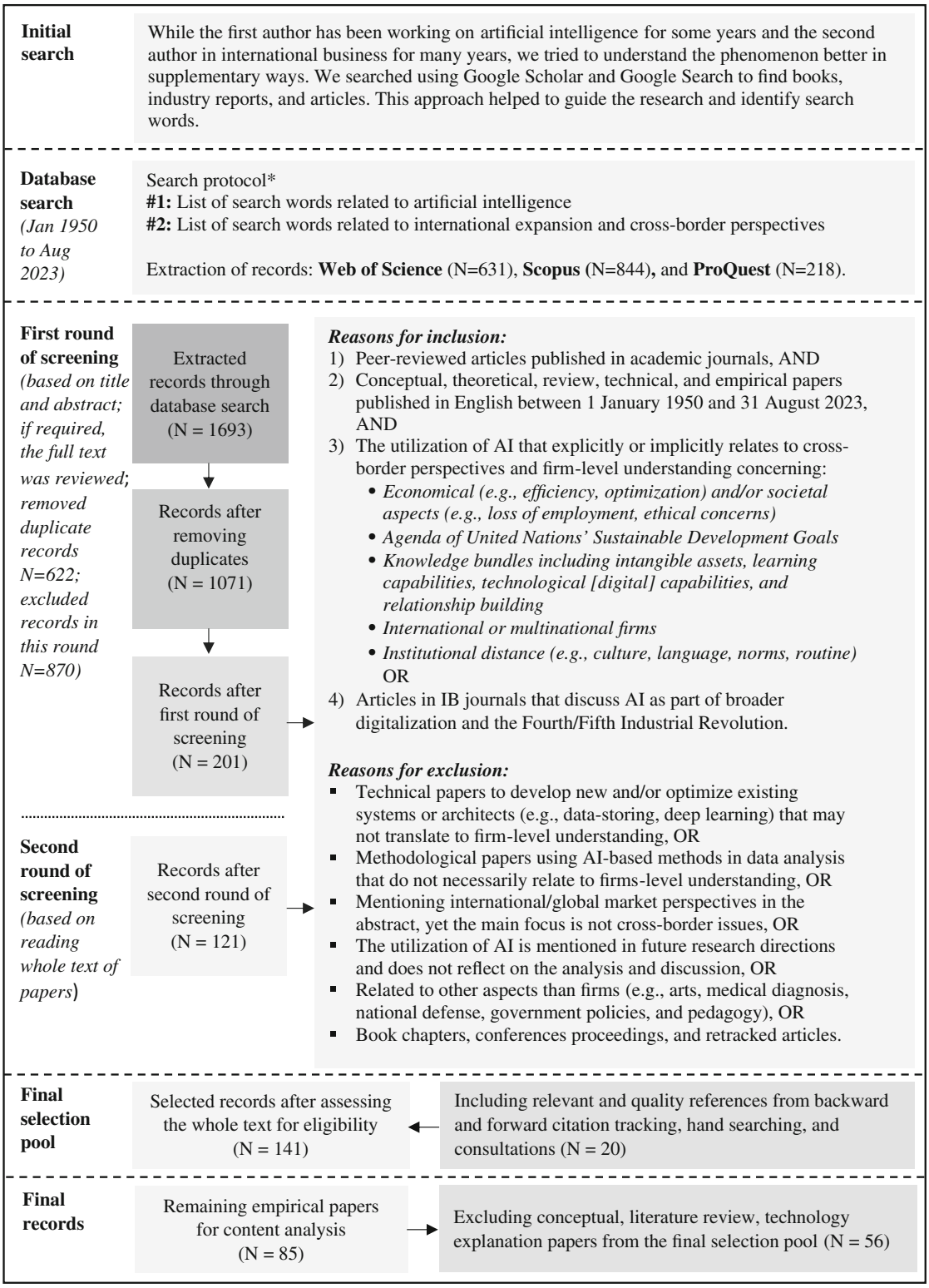


FIGURE 1 Construction of the dataset. *See Appendix A for a detailed account of the search protocol and search string/sets of keywords.

(Miles et al., 2014). How does the paper relate to IB management? Does it deal with any underlying characteristics of AI? If so, how does the characteristic impact IB management? Does it deal with socio-technical issues of managing production and consumption in IB, and, if so, how? Does it inform future research directions in IB management, and if so, how?

Having discussed the methodological stages of our research and the qualitative nature of our constructed dataset and analytical procedure, we now present the findings of our study. We first provide a definition of AI in IB management, followed by the theoretical support for our three constructed characteristics of AI from existing AI and IB literature. We then provide our empirical insights into the three characteristics of AI and outline future research opportunities to achieve sustainable production and consumption.

3 | CONSTRUCTING THE CHARACTERISTICS OF AI IN IB MANAGEMENT: THEORETICAL FOUNDATION

3.1 | What is AI in IB management?

IB research defines AI as the most important technological advancement that (a) learns patterns automatically from examples (Brynjolfsson et al., 2019), (b) makes autonomous decisions (Hemphill & Kelley, 2021), and (c) solves problems and achieve goals (Ferreira et al., 2023). Additionally, IB research went further to compare this technological advancement with humans to articulate that AI is composed of many computational capabilities that allow artificial systems to recognize patterns in a sophisticated way, similar to the human brain (Veiga et al., 2000), and perform functions that mimic or exceed human cognition (Ciulli & Kolk, 2023). However, recent IB research cautiously argues against such an assessment and for the importance of human knowledge in AI applications (Grant & Phene, 2022; Tatarinov et al., 2023).

This human-centrism is critical for successfully utilizing and managing AI in IB operations since AI is computationally constrained by the data it is trained on (Autio et al., 2021). Furthermore, AI's functions and decision-making behavior are also constrained by its operating environments and configurations of interactions like machine–machine and/or human–machine (Glikson & Woolley, 2020; Rahwan et al., 2019). Central to humans, the frontier of computational advancements must still uphold ethics and protect human rights across all economic activities (Berente et al., 2021). Consequently, we argue that AI works well in decision-making with human knowledge and experience. However, recent advances in computational capacity also make AI an evolving computation system that autonomously learns and makes decisions without human knowledge and experience (Lyytinen et al., 2021; Murray et al., 2021). We therefore maintain *AI as the frontier of computational advancements that address complex decision-making problems and can be engineered to learn, evolve, and develop with or without human intervention*. This article focuses only on narrow AI that performs narrowly defined or a singular task in IB operations. Throughout the paper, we use the term "AI" to refer to both complex and simple computational systems used to make decisions and reflect on their embodied representation, like robotic, virtual, and embeddedness (Glikson & Woolley, 2020; Rahwan et al., 2019).

3.2 | Characteristics of AI in IB management: Autonomy, learning, and combinative

Based on qualitative content analysis, we constructed three different characteristics of AI in IB management. The first characteristic of AI is autonomy, which covers its increasing capacity to act without human intervention and behave without human knowledge (Murray et al., 2021; Rahwan et al., 2019). Unlike past information technology and rule-based automation that required humans to codify and program tasks explicitly, AI is based on machine learning that is developed to detect patterns from data and reprogram itself automatically (Berente et al., 2021; Brynjolfsson et al., 2019). This self-development aspect allows AI to automate routine tasks and activities in extant cross-border business management. Recent IB literature suggests that such autonomous capabilities can be expanded across borders by feeding fine-grained data from diverse national contexts (Autio et al., 2021). With such proliferation within multinational settings, managing AI autonomy requires less monitoring, coordination, and transaction costs in IB operations (Cuypers et al., 2021). It also allows managers to combine autonomous machine intelligence with human intelligence in decision-making and knowledge transfer across borders (Grant & Phene, 2022).

The second characteristic of AI is learning, one of the central concepts from its inception (Turing, 1950). Learning refers to machines' "ability to inductively improve automatically through data and experience" (Berente et al., 2021, p. 1437). Machines can learn to recognize patterns through supervised and unsupervised learning techniques and other large-scale advanced computational methods, like deep or reinforcement learning or neural networks. For instance, the neural network model is inspired by the human brain with extensive interconnected units as neurons that connect a vast network to recognize complex patterns. Consequently, IB literature explains that these are modeled with many computational elements functioning in parallel and arranged similarly to biological neural nets (Veiga et al., 2000). Such learning functions enable machines to produce sophisticated outputs. These outputs of pattern recognition gradually improve and perhaps make machine intelligence, through data interpretation and experience iteration, similar to how the human brain routinely performs. Additionally, IB literature consistently argues the importance of learning about markets in firm internationalization (Johanson & Vahlne, 2009; Luostarinen, 1994). Recent AI developments provide excellent learning opportunities for managers in their internationalization process (Huo & Chaudhry, 2021).

The third characteristic of AI is combinative, one of the emerging concepts from recent development that involves hybrid interaction

between machines and machines as well as humans and machines (Rahwan et al., 2019). Current IB literature posits that managers can utilize the combinative characteristics of AI to reconfigure the nature and structure of value-creating activities in the global marketplace (Tatarinov et al., 2023). This recombinant approach allows managers to develop dynamic capabilities that lead to firm-specific assets which include several functions of digital solutions (Banalieva & Dhanaraj, 2019; Lee et al., 2021). AI's modular property governs these interconnected yet separable functions (Autio et al., 2021; Nambisan & Luo, 2021). Modularity refers to the layered technological architecture of a solution to facilitate loose interconnectivity of given elements that can be recombined (Ojala et al., 2018). This combinative approach of AI helps managers to create entirely new functionality enabled by different configurations of new capabilities (as modules) to be added into the systems or architectures. A recent development on the characteristics of digital artifacts in IB argues that digital artifacts can be reprogrammed to facilitate dynamic capabilities in international expansion (Ojala et al., 2023). Similarly, AI literature indicates that each module of AI architecture can be modified or discrete if necessary to achieve its desired functionalities (Rahwan et al., 2019).

In this section, we have provided a theoretical foundation for the three constructed characteristics of AI. However, the question of how managers will utilize these characteristics in IB to achieve responsible production and consumption remains. In the following section, we thus present our empirical insights into AI related to each of the three characteristics and simultaneously develop possible future research agendas. However, we went beyond our dataset to develop a research agenda for managing AI in IB and achieving sustainable production and consumption. We complement the empirical insights with broader theoretical, conceptual, and empirical AI and IB research.

4 | MANAGING AI IN IB: EMPIRICAL INSIGHTS AND TOWARD A RESEARCH AGENDA

4.1 | Autonomy as a characteristic of AI: Empirical foundation ($N = 18$)

4.1.1 | Managing autonomy in IB: Empirical insights

Autonomy aids decision-making to manage global value chains

Managers deal with uncertainty and risk factors in their everyday decisions managing their global value chains. Against such challenges, managers can deploy AI which can process information automatically to make near-optimal decisions (Chen & Du, 2022). For example, managers face uncertainty from risk factors emerging from commodity price shocks and fluctuations in global markets. In this case, managers can utilize AI models to make optimal decisions to hedge themselves against price uncertainty, reduce risk, and increase firm performance in global value chains (Ding et al., 2019).

Managers of firms that operate in global value chains also face risk factors related to multi-echelon supply networks, which refers to firms' inventories spanning multiple layers like sourcing, manufacturing plants, warehouses, and markets in different countries (Kumar et al., 2010). A firm in the automobile industry, like Volvo or Toyota, is an excellent example of such a global value chain. Managerial uncertainty in these firms emerges due to inherited risk factors cornering international coordination, material flows, and market demand. Risk factors may include late shipments, currency fluctuation, geopolitical shifts, customs delays, transport breakdowns, natural disasters, problems with quality control, and global healthcare crises. AI seems helpful for managers to automatically adjust supply chain design that can react to uncertainty arising from those risk factors, to some extent (Kumar et al., 2010). Doing so allows managers to make optimal decisions to manage inter-echelon quantity flow between suppliers' locations, warehouses, manufacturing plants, and distribution channels. Therefore, near-optimal decisions and automation optimize international operations and increase productivity and economic efficiency.

Autonomy optimizes international operations

Concerning productivity and economic efficiency with AI, managers can solve complex problems and enhance business processes that provide customized technological solutions for their global operations. This AI-driven automation increases productivity, reducing waste, labor cost, and workflow time, and increasing firm performance (Sharma & Kumar, 2023). Additionally, intelligence automation seems promising when it comes to improving profitability and labor productivity by reducing labor input costs, which enhances international competitiveness. To illustrate, managers in Chinese manufacturing firms use AI in product automation to increase efficiency. This management approach is central to upgrading their technological capabilities that complement a highly skilled labor force and enhance international competitiveness in the global value chain (Gao, 2023). To further demonstrate, a study on Spanish manufacturing firms revealed that AI-driven automation with robots enables higher labor productivity, leading to better financial performance through export sales to non-robotized firms (Ballestar et al., 2020).

Intelligence machines automatically crunch large amounts of real-time data to optimize production processes, integrating the effective coordination of complex value chains across borders (Kinkel et al., 2023). As an example, managers can deploy AI-enabled computer vision to detect production defects early and achieve real-time production adjustment. Furthermore, managers can equip assembly lines with AI to constantly monitor the production process and autonomously, which aids in reducing operational costs and automating product testing (Yu, Fletcher, & Buck, 2022; Yu, Liu, et al., 2022). Hence, managers can improve the quality of products and production, automating tasks and thereby driving firms to relocate production plants to their home countries (Kinkel et al., 2023; Stemmler, 2023). Similarly, managers of manufacturing firms may increase productivity with AI in just-in-time production systems.

In a just-in-time production system, goods must be delivered to customers on time while minimizing inventory-related costs by maintaining optimum inventory stock. In this case, selecting the right sourcing partners and evaluating supplier performance is paramount to scaling such a system to the entire value chain. Hence, managers

can automate this selection and evaluation process with AI based on critical criteria such as product quality, delivery time, and location (Aksoy & Öztürk, 2011). Additionally, they can accommodate sustainability-related criteria like energy efficiency, social welfare investment, and cost reduction rate in the AI-based evaluation process (Wang, 2022). Therefore, managers in manufacturing firms can achieve economic goals while ensuring environmental and social well-being.

Concerning environmental well-being, AI helps to manage resources (e.g., energy, water, chemicals) efficiently during the production process to carry out tasks in a totally autonomous or semiautonomous way that maintains required outputs (Ferreira et al., 2023). Put simply, it allows managers to reduce carbon footprints while maintaining the quality and quantity of goods produced. Concerning social well-being, empirical insights indicate that managers can increase health and safety practices and improve factory workers' physical and mental health by improving working conditions and automating monotonous and routine tasks (Ferreira et al., 2023). However, AI-driven autonomous or semiautonomous manufacturing processes also raise the issue of replacing human labor with machines, which leads to loss of employment and job displacement (Stemmler, 2023).

4.1.2 | Toward a research agenda: Managing autonomy in IB to achieve sustainable production and consumption

Control problems in internationalization decision-making

As previously noted, the autonomy of AI leads to business processes being optimized and aids decision-making that efficiently helps to manage global value chains. However, extant research in information systems argues that managers' reliance on AI to automate tasks and routines makes them increasingly dependent on it (Berente et al., 2021). This adverse reliance effect may lead to control problems that could have deleterious consequences. The control problem refers to a human decision-maker's tendency to become complacent and overreliant on the output of a reliable, autonomous system (Zerilli et al., 2019). Such overreliance decreases unique human knowledge in human–machine collaborative decision-making settings (Fügener et al., 2021).

Management literature also demonstrates that such issues lead to over trusting in human–machine interaction (Glikson & Woolley, 2020). This evidence is consistent with an experimental study in which humans blindly follow an autonomous agent in a life-threatening situation (Robinette et al., 2016). Alarmingly, half of the participants even observed that the given autonomous agent was not functioning well during the experiment. This unexpected human behavior indicates that managers' overreliance on AI to optimize IB processes and decision-making without effective monitoring may reduce situational awareness. Such effects hinder the detection of potential system failures, thereby leading to disturbance in global value chains and negatively impacting customer experience. Consequently, how does machine autonomy alter human autonomy and negatively affect internationalization decision-making in global value chains? Why do some IB development managers (and not others) overly rely on AI in their internationalization decision-making?

Further IB research on AI's control problems in decision-making in the context of global value chains could add to ongoing debate on human autonomy being influenced by machine autonomy (Murray et al., 2021). More broadly, a fine-grained focus on the characteristics of AI autonomy and its negative impacts on internationalization decision-making could also be used in responsible decision-making to ensure global sustainable production and consumption patterns.

> **Future Research Theme 1.** *Addressing AI in internationalization decision-making in IB research could offer critical insights on managing the control problems that emerge from decision-makers' overreliance on AI and altering human autonomy to ensure sustainable production and consumption.*

Complexity to attribute responsibility and liability

Suppose that an accident happens when a fully autonomous vehicle is in control. Who would be held responsible or liable for the damage caused by AI in the absence of a human driver? This question also applies to AI which is increasingly used to automate decision-making or aid human decision-makers (Jobin et al., 2019). Furthermore, such a question is relevant to any context of IB operations in which a human decides or acts based on information provided by AI. The information could be dubious due to the data sets used in training, human bias inherent to the coding, and data encompassing discriminatory attributes like race, ethnicity, religion, or nationality (Greenstein, 2022). For example, an IB study demonstrates why managers stop international expansions of AI solutions to ensure human well-being due to an institutional void (absence of AI governance framework) in the host country (Tatarinov et al., 2023). On the contrary, managers intentionally or unintentionally develop and deploy AI solutions to increase user engagement and maximize economics of scale that may cause significant harm to society. To illustrate, some AI-powered social media platforms are notoriously debated for their deliberate system design to increase end-user engagement. Such AI-curated digital solutions lead to usage addiction, which has a negative impact on mental well-being and toxic behaviors (Crawford & Smith, 2023). Therefore, further IB research focusing on the institutional void in diverse multi-country contexts (e.g., Rana & Sørensen, 2021) could be useful for the heated debate on the emerging complexity of AI responsibility and liability (Bartneck et al., 2021). Consequently, to better understand this phenomenon, the following research questions are of particular interest. How do managers responsibly utilize AI in internationalization while balancing economic scale and end-user well-being? How do home-country institutions influence managers to maintain AI governance in the context of institutional voids in host countries? How do managers navigate legitimacy development during the internationalization of autonomous agents to minimize the complexity of responsibility and liability?

More broadly, the perspective of attributing responsibility to focus on AI autonomy that involves diverse actors (e.g., developers, manufacturers, government agencies, insurance companies, and

consumers) could assist the management of AI governance in multinational corporations. This governance will also help with how to manage internationalization risks while fostering sustainable production and consumption in global markets.

> **Future Research Theme 2.** *Incorporating an institutional lens to attribute responsibility and liability in IB research could offer critical insights on managing AI governance in the context of diverse actors to ensure sustainable production and consumption when utilizing AI in global markets.*

Emerging international inequalities

Due to AI-driven changes in the global value chain, managers of manufacturing firms originating from developed countries may reconsider relocating their production from emerging countries to their home countries. This is due to routine and manual tasks being able to be automated with robotic AI, making backshoring a viable option (Ahi et al., 2022). The economic benefits of such backshoring initiatives are, for example, enabling firms to be closer to customers, increasing efficiency via shorter times to market, and reducing transportation costs (Bertulfo et al., 2022). Advanced technologies like robotic AI within the global value chain are negatively altering employability in manufacturing in several developing economies (Bertulfo et al., 2022). To provide concrete evidence, increasing managerial adoptions of AI-driven automation in exporting countries has led to reduced employment in manufacturing sectors in Brazil (Stemmler, 2023). Therefore, the adverse fear is that backshoring would reduce employment in developing countries, thereby increasing global income inequalities (Pinheiro et al., 2023; Studley, 2021). While there is relatively little knowledge on how to manage this problem from an international perspective (Kopalle et al., 2022), IB literature highlights the challenges along with recent global disruption and increased nationalist sentiments (Brakman et al., 2021; Del Giudice et al., 2023). Additionally, IB literature tends to focus on the economic benefits of backshoring initiatives (Ahi et al., 2022). However, managing negative social consequences is paramount to ensure global sustainable development.

Consequently, we urge IB scholars to examine questions such as: to what extent does AI autonomy lead to relocation strategies? What are the negative social consequences of AI-driven relocation strategies in developing countries? How can managers reduce income inequalities in host countries resulting from AI-driven reshoring of their manufacturing plants? Does AI autonomy lead to income inequalities within developed countries? Future IB research that focuses on AI autonomy and relocation at the intersection of emerging economic equalities could improve management practices in responsible manufacturing to maintain the economic benefits of host countries.

> **Future Research Theme 3.** *Incorporating relocation strategies and emerging economic inequalities to focus on the autonomous characteristics of AI in IB research could offer critical insights into managing sustainable production and consumption to ensure the economic benefits of host countries.*

4.2 | Learning as a characteristic of AI: Empirical foundation ($N = 36$)

4.2.1 | Managing learning in IB: Empirical insights

Learning aids internationalization decision-making processes

We see the value of AI-driven learning in firms through pattern recognition and assisting managers concerning location choice. Managers can use AI to augment decision-making in evaluating market attractiveness and the competitiveness of focal firms in overseas markets (Dikmen & Birgonul, 2004). Such AI models aid managers in collecting the most relevant data during international expansion and preparing priority lists during strategic planning. The insight from the models then informs managers about critical factors that increase the attractiveness of possible expansion—the availability of funds, market volume, economic prosperity, and country risk rating. However, not all managers necessarily have experiential learning as many have never operated in overseas markets.

AI-based learning may complement a manager's lack of experiential learning in exploring international expansion possibilities and managers can use AI in their market screening efforts. For instance, AI can learn by crunching large amounts of data and generating insights from external databases to focus on potential countries' international trade at the intersection of focal firms' operating industries and product classifications (Fish & Ruby, 2009). Managers can also prioritize target locations, analyzing strategic country groups based on market attractiveness, firms' resources and capabilities, and customer-oriented approaches (Hsieh et al., 2022). Once managers can prioritize target countries or locations, AI-driven learning brings additional benefits to location-focused market selection (Tsilingeridis et al., 2023).

Learning about different locations can help managers to select suitable markets for international expansion. Existing literature concerning AI-based learning at the intersection of market-selection decisions focuses on two aspects. First, AI can help learn about internalization advantage patterns (concerned with reducing transaction costs) (Brouthers et al., 2009). It deals with the managers' location decisions in firms originating in developed countries, like Germany and the United Kingdom, and doing business in emerging markets, such as Eastern European countries. The AI provides a predictive solution to a managerial choice that embodies firm-specific resources (e.g., experience, proprietary knowledge), internalization advantages, and target country characteristics (e.g., market potential, market risk). AI-led learning via prediction solutions is an opportunity for better subsidiary performance. Second, AI can be useful in learning about the political risks of potential subsidiary locations in emerging markets (Herrero et al., 2011). It deals with managerial risk factors in emerging countries such as tax policy, security, and political stability, and the given location can be compared with different countries worldwide.

Learning about other critical aspects of location is also possible with AI.

AI learning helps decision-makers to evaluate international expansion location choices based on critical aspects such as financial leverage and geographical distance. For example, one study focusing on Chinese manufacturing firms revealed that financial leverage has a negative effect on location decisions (Huo & Chaudhry, 2021). Another study focused on Chinese renewable energy and showed that vegetation and distance to the power grid are the most important predictors of solar photovoltaic installation location (Sun et al., 2023). Such AI-driven priority lists of targets and strategic groupings of locations are helpful for managers to formulate relevant configurations in the context of limited resources and experiential learning. Besides market screening and selection, managers can use AI in market entry decisions.

To this end, AI can deal with the ambiguity of human knowledge conveyed through natural language to learn about political and social risks by integrating foreign exchange rates and trade balances. Accordingly, managers can deploy AI to assess country risk and assist with international market entry decisions (Levy & Yoon, 1995). Such learning can also be used in foreign direct investment decisions when evaluating a country's potential for greenfield investment (Alon et al., 2022).

Of particular note is that AI research on market entry decisions is rare. However, there is an indication that AI-based learning can augment managers' decision-making when determining entry mode selection (Li & Li, 2010; Rodgers et al., 2022). For example, managers can assess political risk during the internationalization decision-making process by using AI (Rios-Morales et al., 2009). Such AI offers relatively accurate and meaningful learning about target countries' political instability. This then assists managers in deciding on foreign direct investment.

Learning about the performance of entry mode choice

Learning about the performance of entry mode choice helps managers to survive and further expand in foreign markets. Although a meager amount of knowledge is available concerning the performance of entry mode choice, AI has excellent potential. To illustrate, the authors of a recent IB study suggest that AI-enabled learning may enhance export performance by identifying foreign markets and developing clientele (Neubert & Van der Krogt, 2018). Additionally, managers can utilize AI to learn about merger and acquisition performances in the operating industry. For instance, a study using an AI method in the telecommunication industry examined whether a multinational enterprise becomes a serial acquirer (Navío-Marco et al., 2020). Similarly, early IB research on AI also indicates that managers can accurately learn to predict post-merger performance (Veiga et al., 2000). Besides merger and acquisition performance, AI-based learning also aids managers in evaluating the success of international joint ventures.

Such evaluations can point to critical dimensions such as partner commitment, product characteristics, and control (Hu et al., 1999). Reflecting on foreign equity control, for instance, managers can use AI to construct relationships between factors related to transaction cost and ownership that lead to successful joint-venture performance. Factors may include the proprietary nature of assets, the host country's environment, and cultural differences between host and home countries (Hu et al., 2004). Similarly, managers can utilize AI to learn related and unrelated collaborative venture formation patterns based on the potential collaborating partners' industry groups and home country relatedness (Nair et al., 2007).

Learning leads to managerial capabilities

Managers can use AI-driven learning in predictive analytics which develops managerial capabilities. To illustrate, managers can utilize AI to predict firms' export potential to selected markets, focusing on product competitiveness (Yu, Fletcher, & Buck, 2022; Yu, Liu, et al., 2022). It is also possible to forecast export sales with higher accuracy by integrating nonlinear interrelationships between sales-related variables (Sohrabpour et al., 2021). Therefore, AI-driven learning helps managers to reduce transaction costs by optimizing supply chain and production management and provides insight into future sales, inventory control, and material flow. Besides export prediction, managerial capabilities include predicting their corporate competitiveness, resilience, and internationalization performance.

Initial research on multinational enterprises reveals that AI models can predict stock trends, thereby evaluating firms' corporate competitiveness (Zong & Wang, 2022). Similarly, AI models can predict corporate resilience, which may guide managers in developing dynamic capabilities to support their competitive advantage and create a portfolio of competencies to be used during a crisis (Bughin, 2023). Moreover, research on Finnish small and medium-sized firms has shown that AI models can predict the earliness of internationalization and international performance, thereby evaluating capability portfolios consisting of different firm resources such as managers' networking abilities, bricolage capabilities, and decision-making heuristics (Vuorio & Torkkeli, 2023). In addition to developing managerial capacity through prediction, managers can deploy AI to learn about distance and differences between markets.

Learning about distance and differences

Managers can deploy AI as exploratory tools to understand distance and differences in firms' international expansion activities. Research demonstrates that AI can unearth the underlying pattern of national culture in a narrowly constructed interconnectivity such as cultural compatibility in post-merger performance (Veiga et al., 2000), and multicultural factors to understand customer requirements (Yan et al., 2001). While national culture is a multifaced construct, its differences may not be adequately captured by any single trace or generalized label like "individualism" or "collectivism." Additionally, nations comprise of an almost infinite number of properties or patterns—some of them are culturally universal (etic), and others are culturally specific (emic) (Veiga et al., 2000). In this case, IB studies demonstrate that managers can deploy AI to learn about cultural heterogeneity within nations, focusing on individuals to draw on country-level differences (Messner, 2022c). As an example, AI models can learn to trace cultural

distance by crunching data on individuals' values embodying societal and cultural beliefs and behaviors, even emotional brain systems (Messner, 2022a, 2022b). Such learning enables managers to identify similar cultural subgroups across counties and isolate nations with notable cultural similarities. Besides learning about national cultures, AI can be helpful to learn about differences within countries.

Managers can use AI to learn about micro-market segments within subgroups (Ali & Rao, 2001). Such AI can compare and contrast market segmentation structures in different countries using small samples to foster relationship marketing. This learning about differences implies managing internationalization strategies via identifying consumption patterns in foreign markets and detecting purchase intention patterns within subgroups (Messner, 2022a; Migdał-Najman et al., 2020; Sharma et al., 2022). For instance, small- and medium-sized exporting firms may only target the same cultural subgroups within heterogeneous nations that align with their product features or functionality.

Given the proliferation of unstructured user-generated data on social media, managers can deploy AI models to learn about potential differences in consumer behaviors within countries and subcultural groups. A study on Korean beauty products belonging to the K-beauty subcultural group revealed that consumers have different behaviors and consumption patterns (e.g., attractiveness vs. avoidance) within a homogenous (based on religion) country group, like Indonesia and Malaysia (Jung et al., 2023). Another study based on millions of tweets collected from X (formerly known as Twitter) on a given topic shows how consumer sentiments and emotions substantially differ across countries (Schlegelmilch et al., 2023).

4.2.2 | Toward a research agenda: Managing learning in IB to achieve sustainable production and consumption

Privacy protection

As mentioned previously, AI needs data to learn. The widespread usage of digital technologies enables firms to collect digital trace data for learning. Unlike past information technologies, AI not only learns from proprietary or internal data sources but also from other or external data sources within and beyond organizations. Learning, therefore, confronts managerial issues like privacy violations (Berente et al., 2021; Kopalle et al., 2022). For example, Clearview's image recognition AI is notorious for using billions of images from social media services across the globe without users' consent (Hill, 2020). Such privacy violations are costly for end users and firms alike, which leads to negative consequences. For end users, the loss of privacy may lead to feelings of uneasiness and powerlessness, loss of resources such as time or money, a negative change in an individual's social relationships, and restricted freedom of opinion and behavior (Karwatzki et al., 2022). For firms, it may lead to a loss of customer trust and pose reputational risk in foreign markets (Madan et al., 2023). The intersection of privacy concerns and trust is becoming a global issue with the rise of pervasive digital technologies, particularly for firms minimizing risk and end users protecting their privacy (Luo, 2022). While there has been minimal research on AI and privacy in international contexts, early IB literature shows that managing privacy during the internationalization of digital solutions is paramount (Tatarinov et al., 2023).

Consequently, we propose that the following research questions are particularly important to address in future studies. How do managers ensure privacy protection during the international expansion of AI solutions? How do national and cultural attributes shape the perception of end users' privacy concerns about AI-enabled products? How do managers comply with strong institutions (presence of tight AI regulatory framework) in their home country to ensure privacy protection in host countries with weak institutions? How do the managers of digital platforms legitimize privacy violations by utilizing pervasive AI in foreign markets? To what extent does having privacy protection by default affect the legitimacy of AI solutions during international expansion activities?

> **Future Research Theme 4.** *Incorporating privacy concerns into AI learning in the various contexts of international expansions in IB research could offer critical insights on managing risk and end users' well-being that ensure responsible production and consumption across borders.*

Algorithmic inscrutability in decision-making

Previously, we argued that learning augments managerial decision-making to demonstrate the economic value of AI in internationalization. However, there is growing concern about broader managerial and social implications, including negative consequences like undetected biases that shape machine behavior (Asatiani et al., 2021; Rahwan et al., 2019). However, AI is becoming increasingly complex due to voluminous data being fed into algorithms and advanced computational techniques used in learning. Consequently, AI is increasingly inscrutable in its decision-making, which refers to opacity and unexplainable aspects of AI systems (Asatiani et al., 2021; Berente et al., 2021). Put simply, as AI learns more and grows in complexity, it becomes harder for humans to interpret and explain its output. The problem of algorithmic inscrutability affects not only naive end users but also experts. Management literature provides evidence of how experts' biases are transferred to machines via incomplete data and coding (Sobolev, 2023). Although examining the inscrutability of AI learning is gaining traction in other management domains (Lebovitz et al., 2021; Rai, 2020), IB literature also suggests its importance to achieve global sustainable development (Ciulli & Kolk, 2023). Interestingly, empirical research in IB studies also echoes such "black box" issues of AI learning (Veiga et al., 2000). We, therefore, call for a further understanding of the phenomena from a cross-border management perspective.

> **Future Research Theme 5.** *Incorporating algorithmic inscrutability to focus on AI learning in IB research could offer critical insights into managing transparency to optimize business processes in IB operations and ensure sustainable production and consumption.*

4.3 | Combinative as a characteristic of AI: Empirical foundation ($N = 31$)

4.3.1 | Managing combinative in IB: Empirical insights

Combinative allows reconfiguration of resources

Managers can utilize AI combined with the core functionality of existing products to enable new functionalities in the global marketplace. For example, managers employ AI as a module for physical products, like heavy trucks, to enable service provision, leading to productivity and profitability for customers (Haftor et al., 2021). Such AI is combined with other technological artifacts like the internet of things (e.g., connected sensors in engines and gearboxes) that produce considerable amounts of real-time data with low-latency monitoring (Gooderham et al., 2022). This combinative capability automatically generates insights based on vehicles' technical configurations, actual usage, and the error code combinations that facilitate efficiency. Such efficiency reduces fuel consumption and enables the predictive maintenance of vehicles.

The combinative approach enables managers to develop solutions to self-analyze, proactively report problems, and take real-time corrective action (Gooderham et al., 2022). Recent empirical evidence shows that configuring autonomous vehicles requires many AI solutions, and not all resources are owned by focal firms (Ji et al., 2020). Therefore, managers must combine their existing AI resources with external ones to produce and sell autonomous vehicles in international markets. This recombination of external resources can be achieved through application programmable interfaces to create new functionalities. Such reconfigurations of resources drive business model innovation.

Business model innovation powered by combinative AI that learns and automatically acts is an established phenomenon. For example, Netflix and Spotify's streaming services continuously collect and learn from data to detect usage patterns and predict users' needs and wants. This learning about user consumption patterns allows managers to deploy an AI-based recommendation engine as a separate module. Such combinative AI automatically personalizes offerings in real-time to gain efficiency. Similarly, early research on AI-driven business model innovation demonstrates how managers create value for focal firms and their customers (Haftor et al., 2021). Through AI solutions that combine learning and autonomy, managers upgrade technical configurations to improve offerings. This combinative AI improves over time with more learning, creating a virtuous circle of generating value for customers. Such an approach automatically produces more data and generates predictions based on usage data, leading to new functionalities (Brynjolfsson et al., 2019; Sjodin et al., 2021). Besides new functionalities, managers can also recombine AI to develop dynamic capabilities.

Combinative develops dynamic capabilities

Managers can configure dynamic capabilities in international operations which combine existing organizing capabilities with new AI-driven capabilities. For example, combining cooperative management capability with AI-driven supply chain analytics enhances operational and financial performance during global supply disruption (Dubey et al., 2020). Together with other advanced technological artifacts like the internet of things, AI can be used to improve export manufacturing firms' strategy formulation process concerning supply chain resilience (Ahmed et al., 2023). The process includes selecting optimal routes, reducing energy consumption, inventory policy, and waste reduction.

Configuring an AI-integrated customer relationship capability and combining it with global customer support, automation, and knowledge management improves firm performance in international marketing (Chatterjee et al., 2023). As an example of knowledge management, managers can use AI in a two-layered inductive learning procedure where two different modules assess international financial risk in internationalization (Tessmer et al., 1993). Such AI with combinative learning capabilities identifies foreign companies' risk structures considering various economic environments. This AI learns from national commonalities and differences while integrating an evaluation scheme independent of national attributes. Managers can have a richer yet relevant and concise decision structure to assess international risk that helps with internationalization decision-making and new investment opportunities.

Combinative aids expanding solutions internationally

The managerial approach is to embrace combinative AI with modular architecture facilitates expansion across different markets. Managers expand digital solutions into markets via diverse combinations of AI to target different customer segments and locations in home and host countries (Sjodin et al., 2021; Tatarinov et al., 2023). The ability to create such combinations leads to openness in the solutions (Ozalp et al., 2022). Consequently, it reduces complexity in deployment, yet managers can add new and novel functionalities based on emerging needs in operating markets (Sjodin et al., 2021). Additionally, the core functions of AI-driven solutions can be replicated across borders with minimal customization (Oliva et al., 2022). Furthermore, each part or module can be modified and recombined with internal resources during internationalization (Tatarinov et al., 2023). In this way, managers can add supplementary peripheral functions to solutions with any adaptations required to match host countries' institutions, such as culture, norms, regulations, and language. The following empirical evidence on language will provide further clarification.

Managers can utilize automatic AI-based translation as a module to combine with the core functionality of an existing solution during internationalization (Karkaletsis et al., 1998). Such a module is language-independent regarding host countries and can be used in a multilingual setting to produce a user-tailored product interface in digital environments (Brynjolfsson et al., 2019). Furthermore, a multilingual review system can be combined with existing infrastructure to automatically read all user-generated reviews and learn about customer sentiments (Liu et al., 2021). Hence, a managerial approach to combine AI with extant organizational resources may reduce language barriers in international expansion.

Combinative fosters the international expansion of firms

Managers can combine AI to manage their international expansion activities that address institutional barriers. A digital platform-based firm like eBay is an excellent example, where managers deploy AI-based machine translation modules to existing platforms. By adding machine translation, managers increase exports and reduce translation-related search costs and language barriers (Brynjolfsson et al., 2019). The deployment of such AI can significantly increase international growth (e.g., eBay increased exports by 17.5% in Latin American markets, ibid). The application also demonstrates how AI can reduce informal institutional barriers like language. Additionally, the growth effect illustrates how managers can combine AI with other resources in international expansion activities to increase economic efficiency (Denicolai et al., 2021). However, managers can also embed AI capabilities into digital solutions to facilitate firm internationalization that aligns with formal institutional barriers like environmental regulations.

Sticker regulations exist in many countries to protect the environment from fossil-based energy pipelines. Leaks in such pipelines (e.g., oil and gas) imply a loss of material, damage to the environment, and harm to living beings, including underwater. To illustrate, British Petroleum's crude oil spill in the Gulf of Mexico in 2010 remains the worst environmental event worldwide (Meiners, 2020) and many countries subsequently imposed stricter laws to avoid such catastrophes. Therefore, an AI solution can be developed with satellite-based technologies that match host countries' sticker laws on environmental protection concerning detecting leaks in oil and gas pipelines in real time (Oliva et al., 2022). It opens opportunities for managers to deploy AI solutions that facilitate firm internationalization.

4.3.2 | Toward a research agenda: Managing combinative in IB to achieve sustainable production and consumption

Monopolistic competitive advantage in international expansion

We previously argued that managers utilize AI in combination with other digital artifacts or organizational capabilities such as the internet of things, machine translation modules, and satellite technologies. There is no doubt about the economic benefits that managers bring by using such AI. However, while we appreciate AI's combinative capability, we also observe that this characteristic leads to monopolistic behavior in firm internationalization (Ozalp et al., 2022; Rikap, 2022b). Similarly, IB literature cautiously associates such a managerial approach with monopolistic competitive advantages (Ghauri et al., 2021), which are formed through learning about end-user behavior and automatic actions.

Managers of successful companies like Amazon and Uber are front-runners in such an approach (Banalieva & Dhanaraj, 2019; Rikap, 2022a). For example, AI-led dynamic pricing modules in digital platforms automatically learn usage behavior and offer customized prices to end users, thereby maximizing efficiency and profit (Moghaddam et al., 2020). Consequently, Amazon has been accused of illegally exploiting monopolistic competitive advantages through predatory behavior (Clayton & Espiner, 2023). Our observation is reinforced by empirical evidence on how managers innovate new data-driven combinative AI, including search engine capabilities, dynamic pricing, order fulfillment robots, and personalized recommendation modules (Rikap, 2022a). Similarly, managers of the State Grid Corporation of China reconfigure the energy industry toward clean energy and the smart grid with combinative AI in which monopolistic competitive advantages facilitate firm internationalization (Rikap, 2022b). Therefore, we recognize the importance of further investigating managing AI's combinative characteristics at the intersection of gaining monopolistic competitive advantages.

> **Future Research Theme 6.** *Incorporating AI's combinative characteristics to focus on monopolistic competitive advantages during firm internationalization in IB research could offer critical insights on managing AI to achieve global sustainable development*.

Bundling firm-specific assets for social good

Reflecting resource reconfiguration, the combinative characteristics of AI present opportunities for IB research to focus on constructing AI-driven firm-specific assets that balance between economic and social goals (Tatarinov et al., 2023; Verbeke & Hutzschenreuter, 2021). Firm-specific assets are central to firm internationalization and refer to knowledge bundles, including intangible assets, learning capabilities, technological [digital] capabilities, and privilege relationships (Rugman & Verbeke, 2001; Verbeke & Hutzschenreuter, 2021).

Suppose that managers from a fast-moving consumer goods company use AI together with a fusion of satellite imagery and environmental data like climate, soil, and topographical information in their sourcing activities (Shendryk et al., 2021). With this firm-specific bundling capability, they optimize business processes and management decisions with higher certainty that will minimize search and operational costs while achieving sustainable production. Additionally, managers can assess damaging environmental impact hotspots within their value chain networks to reduce their carbon footprint and intervene as required. For example, AI combined with satellite-based Earth observation (monitoring changes of land, marine, and atmosphere) and subnational production data can assist managers in making responsible sourcing decisions (Moran et al., 2020). Managers will also be able to detect illegal and unsustainable uses of resources in fishing or farming. As an example, managers can face a decision to source soy from Brazil's northern production regions, often linked to deforestation, versus southern regions linked to existing agricultural zones. Similarly, managers of a global fashion chain like H&M can identify environmentally polluted hotspots caused by sourcing partners in Bangladesh and make managerial decisions accordingly (Chaturvedi, 2019). Therefore, managerial knowledge of the exact location of sustainability risk hotspots in a particular supply chain within the value chain networks is paramount. However, contemporary IB literature on the global value chain intersecting digitalization has yet to recognize such firm-specific assets (Kano et al., 2020; Strange & Zucchella, 2017). Early research in IB revealed recombinant firm-specific assets that combine AI and Earth observation capabilities and allow

managers to create replicable solutions across locations (Tatarinov et al., 2023). This combinative approach opens opportunities to augment global value chains to ensure environmental and social well-being.

> **Future Research Theme 7.** *Incorporating AI's combinative characteristics to construct firm-specific assets that balance economic and social goals in firm internationalization in IB research could offer critical insights on managing AI-driven firm-specific assets to achieve sustainable production and consumption.*

5 | DISCUSSION AND CONCLUSION

Based on a critical realism view and direct content analysis of 85 empirical research papers published on AI in multiple disciplines, this study contributes to our understanding of AI in the context of IB in several ways. First, this study contributes to IB literature by revealing three characteristics of AI, namely autonomy, learning, and combinative. We elaborate on how managers can utilize these characteristics for international expansion and how these characteristics facilitate IB management for sustainable development. We found that AI can enable responsible consumption and production worldwide and help achieve SDGs, but only if we have effective management strategies, policies, and institutions to support their use. Otherwise, they can have opposite, deleterious effects, hindering SDGs. Second, we show how managers can utilize AI characteristics to achieve economic goals in international expansion activities. This provides a working framework for firms to manage AI in international markets. However, we found that current empirical research tends to focus on economic goals, which is problematic when it comes to SDGs. Third, we contribute to the multidisciplinary research on AI and elaborate on how it can be applied to deal with global sustainable development (Jobin et al., 2019; Rahwan et al., 2019; Vinuesa et al., 2020). In international contexts, very limited knowledge exists at the intersection of AI and the sustainability agenda. Therefore, we demonstrate that IB as a research discipline can play a central role in AI research to work toward global sustainable development. To achieve this, managing AI in IB operations to balance economic, technical, and societal goals is paramount. Fourth, we propose future research directions and highlight research questions and themes that, grounded in current knowledge, are important for achieving a more profound understanding of AI in IB and its usage for sustainable production and consumption (SDG #12) internationally.

5.1 | Managerial implications

The insights from our research have significant and timely practical implications for managers. Our study also has important implications for managing AI to automate business processes, learn about markets, and combine resources in international expansion activities. For automation, managers may not want to rely too much on AI in their decision-making to avoid control problems. Instead, intelligence automation can be used to construct systemic capabilities that complement human limitations. Since AI can process voluminous information, it can augment human employees' cognitive limitations on data collection and interpretation (Murray et al., 2021). Additionally, managers must remember humans are very adept at employing their experiential learning, heuristics, and imagination to solve problems. On the other hand, AI may lack empathy, cultural intelligence, and contextual understanding. Therefore, we encourage managers to utilize AI strategically in IB operations to empower and augment employees rather than replace them.

As it relates to learning, the insights of our study point to the imperative of data-driven managerial capabilities through prediction and algorithmic inscrutability in internationalization decision-making. There are growing tensions around complex AI that can be used to learn about end users and manipulate human behavior for profit (Berente et al., 2021; Rahwan et al., 2019). Such AI applications may violate the fundamental right of humans to decide on their free will (European Commission, 2019). Respecting human autonomy, managers must guide responsible AI initiatives (translating ethics into practice) to process data, construct algorithms, train models, and utilize AI to be inclusive, accessible, and transparent (Rakova et al., 2021). Being responsible managers, they must act to comply with human rights and privacy protection by default. Such responsible AI initiatives in international markets would allow managers to safeguard their organizations against reputational, legal, and regulatory risks (Bartneck et al., 2021; Luo, 2022). In this way, managers can more effectively develop learning capabilities in their internationalization decision-making while protecting human rights.

Regarding the combinative approach, our research suggests that managers can reconfigure resources to gain competitive advantages through hybrid human–machine collaboration. In this case, managerial falsifiability is deemed necessary to empirically test the decision output in intended use cases if harmless to humans and safe to do so (Floridi et al., 2020). We argue that falsifiability is essential to ensure the legitimacy of the AI-aided decision-making process and AI deployment in the global value chains that embody ethical principles (Bartneck et al., 2021; European Commission, 2019). Once an AI system is ready for deployment, constant falsifiability and taking incremental steps is crucial. This would allow managers to critically reflect on operating context or institutions (e.g., norms, value judgment, culture) and detect unintended negative consequences at early stages (Hasan et al., 2021; Rahwan et al., 2019). Consequently, managers would be able to make changes to AI systems to ensure local responsiveness while reconfiguring resources. In summary, we advocate for repeated falsifiability and incremental deployment of AI in international expansion activities.

5.2 | Limitations of the study and additional directions for future research

We acknowledge certain limitations in this study that are important to consider. First, the literature concerning the intersection of AI and IB

is still in a nascent stage, necessitating further empirical studies to comprehensively cover the phenomenon. Specifically, the papers examined in this research were sourced from various scientific fields with only a few concentrating on AI's utilization in a firm's international activities or internationalization. Consequently, this knowledge gap provides development opportunities. Further fine-grained empirical analysis can be done to combine dimensions of a firm's international activities, which we argued in our study. For example, IB scholars can focus on aspects like managing AI-driven dynamic capabilities and bundling firm-specific assets for social goods, benefitting from economics of scale while reducing carbon emissions and air pollution (Ercan et al., 2022). Other opportunities lie in AI-driven learning about consumption patterns in distance markets that protect consumers' privacy, which reduces the liability of foreignness or complements a managerial lack of experiential learning (Johanson & Vahlne, 2009; Zaheer, 1995).

Second, by recognizing the limitation concerning the absence of direct knowledge, IB scholars could also employ a similar methodological approach (qualitative directed content analysis) in other contexts to derive novel insights into IB. For example, researchers could examine AI's autonomy at the intersection of managing sustainable global value chains (Dimitropoulos et al., 2023), and human–machine complementarity versus the fear of replacement in international management (Man Tang et al., 2022). Similarly, future research can explore AI's learning perspective regarding data-driven strategic resources in global expansion (Hartmann & Henkel, 2020) and international sports sponsorship management (Koronios et al., 2021). Furthermore, IB scholars could also extend AI's combinative aspects on the human augmentation versus augmenting automation debate (Raisch & Krakowski, 2021; Tschang & Almirall, 2021), human–machine collaboration that conjointly learns (Lyytinen et al., 2021; Murray et al., 2021), and modular products architect to facilitate firm internationalization (Habib et al., 2020; Ojala et al., 2018).

Third, due to the extensive scope of the articles reviewed, there is always a possibility that significant works could have been overlooked. As the field matures and more direct knowledge becomes available on this phenomenon, employing a systematic literature review could reinforce and further develop the findings of this study.

Fourth, the empirical insights naturally limited us to covering embedded AI[1] and production-related issues. Consequently, we could not bring widespread insights to other types of AI (e.g., robotic AI) that are generally utilized in consumer-centric activities in different cultures (Hermann, 2022; Lim et al., 2021). Therefore, we suggest extending our research to focus on the perspective of responsible consumption in IB. For instance, future research can focus on AI's autonomy to examine deception and manipulation during consumption through cross-country analysis (Kaminski et al., 2017; Sharkey & Sharkey, 2021). Similarly, IB scholars can extend the autonomy perspective to examine how consumers anthropomorphize AI differently (relating machine to human ability) and develop control problems across national cultures (Duffy, 2003; Lim et al., 2021).

Fifth, reflecting on our sociotechnical view of AI to balance between economic and social goals, we could not bring a wide perspective of environmental goals. Therefore, IB researchers can take another view of managing AI to achieve SDG #12. We see value in further developing our findings by taking the dynamic capabilities or extended value chains perspective to balance economic and environmental goals. Researchers can employ the dynamic capabilities view to focus on AI-enabled circular business model innovation (Sjödin et al., 2023) or the extended value chains perspective to focus on AI-driven circular economy in IB operations (Chabowski et al., 2023).

Finally, the limited capacity of humans to process a large amount of information makes it possible for us to overlook hidden patterns when analyzing textual data. Our combinative argumentation suggests that IB scholars can reinforce these findings by complementing them with an AI system. For instance, future research could use the Microsoft AI-powered Copilot embedded into Bing chat to analyze textual content. Researchers can extract only text related to empirical insights and conduct intra-content comparisons with careful prompt engineering. They can then systematically combine their insights with the patterns identified by AI to construct emerging characteristics of AI and dimensions of IB management. Such a combinative approach will provide a methodological contribution concerning directed content analysis in IB research and further develop the findings of our study.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Tiina Leposky, Tero Vartiainen, Udo Zander, Hannu Makkonen, Peter Gabrielsson, Saeed Samiee, Anisur Faroque, Brian Chabowski, Leonidas Leonidou, Lasse Torkkeli, Mohammad Bakhtiar Rana, and Jean-François Hennart for their consultancy in the very early stages of the research. The authors also acknowledge Catherine Welch, Rebecca Piekkari, and Emmanuella Plakoyiannaki for their inspiration in the research method. Earlier versions of the paper benefited from reviews and feedback at the 17th Vaasa International Business Conference 2023 and the European International Business Academy annual conference 2023. The authors are also grateful to colleagues from the International Business and Marketing Strategy research group at the University of Vaasa for their constructive comments on earlier versions. The research is supported by Finland Fellowship, funded by the Finnish Ministry of Education and Culture. Finally, the authors honor the memory of Jorma Larimo for his inspiration, support, and encouragement in setting the direction of our research.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Rakibul Hasan https://orcid.org/0000-0003-2088-0490
Arto Ojala https://orcid.org/0000-0002-3276-6348

ENDNOTES

[1] We used the term "AI" interchangeably to argue for both AI as a research discipline and AI as computation systems (Glikson & Woolley, 2020; Rahwan et al., 2019). Regarding the latter, we used "AI"

throughout the paper to refer to both complex and simple computational systems used to make decisions and reflect on their embodied representations. Such representations include robotic AI (e.g., industrial or humanoid robots), virtual AI (e.g., chatbots), and embedded AI (e.g., autonomous vehicles).

[2] Of particular note is that, unlike a systematic literature review, our detailed explanation of data collection to construct the dataset is for transparency and authenticity rather than arguing for replicability and reproducibility.

REFERENCES

Ahi, A. A., Sinkovics, N., Shildibekov, Y., Sinkovics, R. R., & Mehandjiev, N. (2022). Advanced technologies and international business: A multidisciplinary analysis of the literature. *International Business Review*, *31*(4), 101967. https://doi.org/10.1016/j.ibusrev.2021.101967

Ahmed, T., Karmaker, C. L., Nasir, S. B., Moktadir, M. A., & Paul, S. K. (2023). Modeling the artificial intelligence-based imperatives of industry 5.0 towards resilient supply chains: A post-COVID-19 pandemic perspective. *Computers and Industrial Engineering*, *177*, 109055. https://doi.org/10.1016/j.cie.2023.109055

Aksoy, A., & Öztürk, N. (2011). Supplier selection and performance evaluation in just-in-time production environments. *Expert Systems with Applications*, *38*(5), 6351–6359. https://doi.org/10.1016/j.eswa.2010.11.104

Ali, J., & Rao, C. P. (2001). Micro-market segmentation using a neural network model approach. *Journal of International Consumer Marketing*, *13*(2), 7–27. https://doi.org/10.1300/J046v13n02_02

Alon, I., Bretas, V. P. G., Sclip, A., & Paltrinieri, A. (2022). Greenfield FDI attractiveness index: A machine learning approach. *Competitiveness Review*, *32*(7), 85–108. https://doi.org/10.1108/CR-12-2021-0171

Asatiani, A., Malo, P., Nagbøl, P., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems*, *22*(2), 325–352. https://doi.org/10.17705/1jais.00664

Autio, E., Mudambi, R., & Yoo, Y. (2021). Digitalization and globalization in a turbulent world: Centrifugal and centripetal forces. *Global Strategy Journal*, *11*(1), 3–16. https://doi.org/10.1002/gsj.1396

Ballestar, M., Diaz-Chao, A., Sainz, J., & Torrent-Sellens, J. (2020). Knowledge, robots and productivity in SMEs: Explaining the second digital wave. *Journal of Business Research*, *108*, 119–131. https://doi.org/10.1016/j.jbusres.2019.11.017

Banalieva, E. R., & Dhanaraj, C. (2019). Internalization theory for the digital economy. *Journal of International Business Studies*, *50*(8), 1372–1387. https://doi.org/10.1057/s41267-019-00243-7

Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *An introduction to ethics in robotics and AI*. Springer Nature.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433–1450.

Bertulfo, D., Gentile, E., & de Vries, G. (2022). The employment effects of technology, trade, and consumption in global value chains: Evidence for developing Asia. *Asian Development Review*, *39*(2), 1–44. https://doi.org/10.1142/S011611052250010X

Bhaskar, R. (1978). *A realist theory of science*. Harvester Press.

Bonache, J. (2021). The challenge of using a 'non-positivist' paradigm and getting through the peer-review process. *Human Resource Management Journal*, *31*(1), 37–48. https://doi.org/10.1111/1748-8583.12319

Brakman, S., Garretsen, H., & van Witteloostuijn, A. (2021). Robots do not get the coronavirus: The COVID-19 pandemic and the international division of labor. *Journal of International Business Studies*, *52*(6), 1215–1224. https://doi.org/10.1057/s41267-021-00410-9

Brouthers, L. E., Mukhopadhyay, S., Wilkinson, T. J., & Brouthers, K. D. (2009). International market selection and subsidiary performance: A neural network approach. *Journal of World Business*, *44*(3), 262–273. https://doi.org/10.1016/j.jwb.2008.08.004

Brynjolfsson, E., Hui, X., & Liu, M. (2019). Does machine translation affect international trade? Evidence from a large digital platform. *Management Science*, *65*(12), 5449–5460. https://doi.org/10.1287/mnsc.2019.3388

Bughin, J. (2023). Are you resilient? Machine learning prediction of corporate rebound out of the Covid-19 pandemic. *Managerial and Decision Economics*, *44*(3), 1547–1564. https://doi.org/10.1002/mde.3764

Cavusgil, S. T., Mitri, M., & Evirgen, T. C. (1992). A decision support system for doing business with Eastern Bloc countries: The country consultant. *European Business Review*, *92*(4), 24–34. https://doi.org/10.1108/EUM0000000001903

Chabowski, B. R., Gabrielsson, P., Hult, G. T. M., & Morgeson, F. V. (2023). Sustainable international business model innovations for a globalizing circular economy: A review and synthesis, integrative framework, and opportunities for future research. *Journal of International Business Studies*. Advance online publication. https://doi.org/10.1057/s41267-023-00652-9

Chalmers, D., MacKenzie, N. G., & Carter, S. (2021). Artificial intelligence and entrepreneurship: Implications for venture creation in the fourth industrial revolution. *Entrepreneurship Theory and Practice*, *45*(5), 1028–1053. https://doi.org/10.1177/1042258720934581

Chatterjee, S., Chaudhuri, R., Thrassou, A., & Vrontis, D. (2023). International relationship management during social distancing: The role of AI-integrated social CRM by MNEs during the Covid-19 pandemic. *International Marketing Review*, *40*(5), 1263–1294. https://doi.org/10.1108/IMR-12-2021-0372

Chaturvedi, A. (2019). *How satellites are changing the way we track pollution on the ground*. The Wire Retrieved from https://thewire.in/environment/how-satellites-are-changing-the-way-we-track-pollution-on-the-ground

Chen, M., & Du, W. (2022). Dynamic relationship network and international management of enterprise supply chain by particle swarm optimization algorithm under deep learning. *Expert Systems*, e13081. Advance online publication. https://doi.org/10.1111/exsy.13081

Ciulli, F., & Kolk, A. (2023). International business, digital technologies and sustainable development: Connecting the dots. *Journal of World Business*, *58*(4), 101445. https://doi.org/10.1016/j.jwb.2023.101445

Clayton, J., & Espiner, T. (2023). Amazon: US accuses online giant of illegal monopoly. *BBC News*. Retrieved from https://www.bbc.com/news/business-66920137

Crawford, A., & Smith, T. (2023). 'I was addicted to social media—Now I'm suing Big Tech.' *BBC News*. Retrieved from https://www.bbc.com/news/technology-67443705

Cuypers, I. R. P., Hennart, J.-F., Silverman, B. S., & Ertug, G. (2021). Transaction cost theory: Past progress, current challenges, and suggestions for the future. *Academy of Management Annals*, *15*(1), 111–150. https://doi.org/10.5465/annals.2019.0051

Del Giudice, M., Scuotto, V., Papa, A., & Singh, S. (2023). The "bright" side of innovation management for international new ventures. *Technovation*, *125*, 1–10. https://doi.org/10.1016/j.technovation.2023.102789

Denicolai, S., Zucchella, A., & Magnani, G. (2021). Internationalization, digitalization, and sustainability: Are SMEs ready? A survey on synergies and substituting effects among growth paths. *Technological Forecasting and Social Change*, *166*, 1–15. https://doi.org/10.1016/j.techfore.2021.120650

Dikmen, I., & Birgonul, M. T. (2004). Neural network model to support international market entry decisions. *Journal of Construction Engineering and Management*, *130*(1), 59–66. https://doi.org/10.1061/(ASCE)0733-9364(2004)130:1(59)

Dimitropoulos, P., Koronios, K., & Sakka, G. (2023). International business sustainability and global value chains: Synthesis, framework and research agenda. *Journal of International Management*, *29*(5), 101054. https://doi.org/10.1016/j.intman.2023.101054

Ding, S., Zhang, Y., & Duygun, M. (2019). Modeling price volatility based on a genetic programming approach. *British Journal of Management*, *30*(2), 328–340. https://doi.org/10.1111/1467-8551.12359

Dubey, R., Gunasekaran, A., Childe, S. J., Bryde, D. J., Giannakis, M., Foropon, C., Roubaud, D., & Hazen, B. T. (2020). Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations. *International Journal of Production Economics*, *226*, 107599. https://doi.org/10.1016/j.ijpe.2019.107599

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Ercan, T., Onat, N. C., Keya, N., Tatari, O., Eluru, N., & Kucukvar, M. (2022). Autonomous electric vehicles can reduce carbon emissions and air pollution in cities. *Transportation Research Part D: Transport and Environment*, *112*, 103472. https://doi.org/10.1016/j.trd.2022.103472

European Commission. (2019). High-level expert group on AI: Ethics guidelines for trustworthy AI. Retrieved from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission. (2022). Industry 5.0. Retrieved from https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry-50_en

Evers, N., Ojala, A., Sousa, C. M. P., & Criado-Rialp, A. (2023). Unraveling business model innovation in firm internationalization: A systematic literature review and future research agenda. *Journal of Business Research*, *158*, 113659. https://doi.org/10.1016/j.jbusres.2023.113659

Ferreira, J. J., Lopes, J. M., Gomes, S., & Rammal, H. G. (2023). Industry 4.0 implementation: Environmental and social sustainability in manufacturing multinational enterprises. *Journal of Cleaner Production*, *404*, 136841. https://doi.org/10.1016/j.jclepro.2023.136841

Fish, K., & Ruby, P. (2009). An artificial intelligence foreign market screening method for small businesses. *International Journal of Entrepreneurship*, *13*(1), 65–91.

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. *Science and Engineering Ethics*, *26*(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). WIll humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *MIS Quarterly*, *45*(3), 1527–1556. https://doi.org/10.25300/MISQ/2021/16553

Furnari, S., Crilly, D., Misangyi, V. F., Greckhamer, T., Fiss, P. C., & Aguilera, R. V. (2021). Capturing causal complexity: Heuristics for configurational theorizing. *Academy of Management Review*, *46*(4), 778–799. https://doi.org/10.5465/amr.2019.0298

Gao, Y. (2023). Unleashing the mechanism among environmental regulation, artificial intelligence, and global value chain leaps: A roadmap toward digital revolution and environmental sustainability. *Environmental Science and Pollution Research*, *30*(10), 28107–28117. https://doi.org/10.1007/s11356-022-23898-6

Ghauri, P., Strange, R., & Cooke, F. L. (2021). Research on international business: The new realities. *International Business Review*, *30*(2), 1–11. https://doi.org/10.1016/j.ibusrev.2021.101794

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gooderham, P. N., Elter, F., Pedersen, T., & Sandvik, A. M. (2022). The digital challenge for multinational mobile network operators. More marginalization or rejuvenation? *Journal of International Management*, *28*(4), 100946. https://doi.org/10.1016/j.intman.2022.100946

Grant, R., & Phene, A. (2022). The knowledge based view and global strategy: Past impact and future potential. *Global Strategy Journal*, *12*(1), 3–30. https://doi.org/10.1002/gsj.1399

Greenstein, S. (2022). Preserving the rule of law in the era of artificial intelligence (AI). *Artificial Intelligence and Law*, *30*(3), 291–323. https://doi.org/10.1007/s10506-021-09294-4

Habib, T., Kristiansen, J. N., Rana, M. B., & Ritala, P. (2020). Revisiting the role of modular innovation in technological radicalness and architectural change of products: The case of Tesla X and Roomba. *Technovation*, *98*, 102163. https://doi.org/10.1016/j.technovation.2020.102163

Haftor, D. M., Costa Climent, R., & Lundström, J. E. (2021). How machine learning activates data network effects in business models: Theory advancement through an industrial case of promoting ecological sustainability. *Journal of Business Research*, *131*, 196–205. https://doi.org/10.1016/j.jbusres.2021.04.015

Hartmann, P., & Henkel, J. (2020). The rise of corporate science in AI: Data as a strategic resource. *Academy of Management Discoveries*, *6*(3), 359–381. https://doi.org/10.5465/amd.2019.0043

Hasan, R., Weaven, S., & Thaichon, P. (2021). Blurring the line between physical and digital environment: The impact of artificial intelligence on customers' relationship and customer experience. In P. Thaichon & V. Ratten (Eds.), *Developing digital marketing* (pp. 135–153). Emerald Publishing Limited. https://doi.org/10.1108/978-1-80071-348-220211008

Hemphill, T. A., & Kelley, K. J. (2021). Artificial intelligence and the fifth phase of political risk management: An application to regulatory expropriation. *Thunderbird International Business Review*, *63*(5), 585–595. https://doi.org/10.1002/tie.22222

Hermann, E. (2022). Leveraging artificial intelligence in marketing for social good—An ethical perspective. *Journal of Business Ethics*, *179*(1), 43–61. https://doi.org/10.1007/s10551-021-04843-y

Herrero, Á., Corchado, E., & Jiménez, A. (2011). Unsupervised neural models for country and political risk analysis. *Expert Systems with Applications*, *38*(11), 13641–13661. https://doi.org/10.1016/j.eswa.2011.04.136

Hill, K. (2020). The secretive company that might end privacy as we know it. *The New York Times* Retrieved from https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Hsieh, P.-C., Horng, D.-J., & Chang, H.-Y. (2022). International expansion selection model by machine learning—A proprietary model. *Computer Journal*, *65*(2), 217–236. https://doi.org/10.1093/comjnl/bxaa018

Hu, M. Y., Hung, M. S., Patuwo, B. E., & Shanker, M. S. (1999). Estimating the performance of Sino-Hong Kong joint ventures using neural network ensembles. *Annals of Operations Research*, *87*, 213–232. https://doi.org/10.1023/a:1018928902137

Hu, M. Y., Zhang, G. P., & Chen, H. (2004). Modeling foreign equity control in Sino-foreign joint ventures with neural networks. *European Journal of Operational Research*, *159*(3), 729–740. https://doi.org/10.1016/j.ejor.2003.06.002

Huo, D., & Chaudhry, H. R. (2021). Using machine learning for evaluating global expansion location decisions: An analysis of Chinese manufacturing sector. *Technological Forecasting and Social Change*, *163*, 120436. https://doi.org/10.1016/j.techfore.2020.120436

Ji, Y., Zhu, X., Zhao, T., Wu, L., & Sun, M. (2020). Revealing technology innovation, competition and cooperation of self-driving vehicles from patent perspective. *IEEE Access*, *8*, 221191–221202. https://doi.org/10.1109/ACCESS.2020.3042019

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Johanson, J., & Vahlne, J.-E. (2009). The Uppsala internationalization process model revisited: From liability of foreignness to liability of outsidership. *Journal of International Business Studies*, *40*(9), 1411–1431. https://doi.org/10.1057/jibs.2009.24

Jung, M., Lee, Y. L., & Chung, J.-E. (2023). Cross-national consumer research using structural topic modelling: Consumers' approach-avoidance behaviours. *International Journal of Consumer Studies*, *47*(5), 1692–1713. https://doi.org/10.1111/ijcs.12923

Kaminski, M. E., Rueben, M., Smart, W. D., & Grimm, C. (2017). Averting robot eyes. *Maryland Law Review*, *76*, 983.

Kano, L., Tsang, E. W. K., & Yeung, H. W. (2020). Global value chains: A review of the multi-disciplinary literature. *Journal of International Business Studies*, *51*(4), 577–622. https://doi.org/10.1057/s41267-020-00304-2

Karkaletsis, E. A., Spyropoulos, C. D., & Vouros, G. A. (1998). A knowledge-based methodology for supporting multilingual and user-tailored interfaces. *Interacting with Computers*, *9*(3), 311–333. https://doi.org/10.1016/S0953-5438(97)00030-1

Karwatzki, S., Trenz, M., & Veit, D. (2022). The multidimensional nature of privacy risks: Conceptualisation, measurement and implications for digital services. *Information Systems Journal*, *32*(6), 1126–1157. https://doi.org/10.1111/isj.12386

Kinkel, S., Capestro, M., Di Maria, E., & Bettiol, M. (2023). Artificial intelligence and relocation of production activities: An empirical cross-national study. *International Journal of Production Economics*, *261*, 1–14. https://doi.org/10.1016/j.ijpe.2023.108890

Kopalle, P. K., Gangwar, M., Kaplan, A., Ramachandran, D., Reinartz, W., & Rindfleisch, A. (2022). Examining artificial intelligence (AI) technologies in marketing via a global lens: Current trends and future research opportunities. *International Journal of Research in Marketing*, *39*(2), 522–540. https://doi.org/10.1016/j.ijresmar.2021.11.002

Koronios, K., Vrontis, D., & Thrassou, A. (2021). Strategic sport sponsorship management—A scale development and validation. *Journal of Business Research*, *130*, 295–307. https://doi.org/10.1016/j.jbusres.2021.03.031

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage Publications.

Kumar, S. K., Tiwari, M. K., & Babiceanu, R. F. (2010). Minimisation of supply chain cost with embedded risk using computational intelligence approaches. *International Journal of Production Research*, *48*(13), 3717–3739. https://doi.org/10.1080/00207540902893425

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly: Management Information Systems*, *45*(3), 1501–1525. https://doi.org/10.25300/MISQ/2021/16564

Lee, J. M., Narula, R., & Hillemann, J. (2021). Unraveling asset recombination through the lens of firm-specific advantages: A dynamic capabilities perspective. *Journal of World Business*, *56*(2), 101193. https://doi.org/10.1016/j.jwb.2021.101193

Levy, J. B., & Yoon, E. (1995). Modeling global market entry decision by fuzzy logic with an application to country risk assessment. *European Journal of Operational Research*, *82*(1), 53–78. https://doi.org/10.1016/0377-2217(93)E0166-U

Li, S., & Li, J. Z. (2010). AgentsInternational: Integration of multiple agents, simulation, knowledge bases and fuzzy logic for international marketing decision making. *Expert Systems with Applications*, *37*(3), 2580–2587. https://doi.org/10.1016/j.eswa.2009.08.022

Lim, V., Rooksby, M., & Cross, E. S. (2021). Social robots on a global stage: Establishing a role for culture during human–robot interaction. *International Journal of Social Robotics*, *13*(6), 1307–1333.

Liu, P., Zhang, L., & Gulla, J. (2021). Multilingual review-aware deep recommender system via aspect-based sentiment analysis. *ACM Transactions on Information Systems*, *39*(2), 1–33. https://doi.org/10.1145/3432049

Luo, Y. (2022). A general framework of digitization risks in international business. *Journal of International Business Studies*, *53*(2), 344–361. https://doi.org/10.1057/s41267-021-00448-9

Luo, Y., & Zahra, S. (2023). Industry 4.0 in international business research. *Journal of International Business Studies*, *54*(3), 403–417. https://doi.org/10.1057/s41267-022-00577-9

Luostarinen, R. (Ed.). (1994). *Internationalization of Finnish firms and their response to global challenges*. UNU World Institute for Development Economics Research. https://doi.org/10.22004/ag.econ.295309

Lyytinen, K., Nickerson, J., & King, J. (2021). Metahuman systems = humans plus machines that learn. *Journal of Information Technology*, *36*(4), 427–445. https://doi.org/10.1177/0268396220915917

Madan, S., Savani, K., & Katsikeas, C. S. (2023). Privacy please: Power distance and people's responses to data breaches across countries. *Journal of International Business Studies*, *54*(4), 731–754. https://doi.org/10.1057/s41267-022-00519-5

Man Tang, P., Koopman, J., McClean, S. T., Zhang, J. H., Li, C. H., De Cremer, D., Lu, Y., & Ng, C. T. S. (2022). When conscientious employees meet intelligent machines: An integrative approach inspired by complementarity theory and role theory. *Academy of Management Journal*, *65*(3), 1019–1054. https://doi.org/10.5465/amj.2020.1516

Meiners, J. (2020). *Ten years later, BP oil spill continues to harm wildlife—Especially dolphins*. National Geographic Retrieved from https://www.nationalgeographic.com/animals/article/how-is-wildlife-doing-now–ten-years-after-the-deepwater-horizon

Messner, W. (2022a). Advancing our understanding of cultural heterogeneity with unsupervised machine learning. *Journal of International Management*, *28*(2), 100885. https://doi.org/10.1016/j.intman.2021.100885

Messner, W. (2022b). Cultural differences in an artificial representation of the human emotional brain system: A deep learning study. *Journal of International Marketing*, *30*(4), 21–43. https://doi.org/10.1177/1069031X221123993

Messner, W. (2022c). Cultural heterozygosity: Towards a new measure of within-country cultural diversity. *Journal of World Business*, *57*(4), 1–17. https://doi.org/10.1016/j.jwb.2022.101346

Migdał-Najman, K., Najman, K., & Badowska, S. (2020). The GNG neural network in analyzing consumer behaviour patterns: Empirical research on a purchasing behaviour processes realized by the elderly consumers. *Advances in Data Analysis and Classification*, *14*(4), 947–982. https://doi.org/10.1007/s11634-020-00415-6

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.

Mingers, J. (2004). Realizing information systems: Critical realism as an underpinning philosophy for information systems. *Information and Organization*, *14*(2), 87–103. https://doi.org/10.1016/j.infoandorg.2003.06.001

Moghaddam, V., Yazdani, A., Wang, H., Parlevliet, D., & Shahnia, F. (2020). An online reinforcement learning approach for dynamic pricing of electric vehicle charging stations. *IEEE Access*, *8*, 130305–130313. https://doi.org/10.1109/ACCESS.2020.3009419

Montiel, I., Cuervo-Cazurra, A., Park, J., Antolín-López, R., & Husted, B. W. (2021). Implementing the United Nations' sustainable development goals in international business. *Journal of International Business Studies*, *52*(5), 999–1030. https://doi.org/10.1057/s41267-021-00445-y

Moran, D., Giljum, S., Kanemoto, K., & Godar, J. (2020). From satellite to supply chain: New approaches connect earth observation to economic decisions. *One Earth*, *3*(1), 5–8. https://doi.org/10.1016/j.oneear.2020.06.007

Murray, A., Rhymer, J., & Sirmon, D. G. (2021). Humans and technology: Forms of conjoined agency in organizations. *Academy of Management Review*, *46*(3), 552–571. https://doi.org/10.5465/amr.2019.0186

Nair, A., Hanvanich, S., & Cavusgil, S. (2007). An exploration of the patterns underlying related and unrelated collaborative ventures using neural network: Empirical investigation of collaborative venture formation data spanning 1985-2001. *International Business Review*, *16*(6), 659–686. https://doi.org/10.1016/j.ibusrev.2007.08.005

Nambisan, S., & Luo, Y. (2021). Toward a loose coupling view of digital globalization. *Journal of International Business Studies*, *52*(8), 1646–1663. https://doi.org/10.1057/s41267-021-00446-x

Navío-Marco, J., Solórzano-García, M., & Vicente-Virseda, J. A. (2020). Serial acquirers' strategy in the telecommunications sector: Integration or indigestion? *European Journal of International Management*, *14*(3), 421–442. https://doi.org/10.1504/EJIM.2020.107029

Neubert, M., & Van der Krogt, A. (2018). Impact of business intelligence solutions on export performance of software firms in emerging economies. *Technology Innovation Management Review*, *8*(9), 39–49. https://doi.org/10.22215/timreview/1185

Ojala, A., Evers, N., & Rialp, A. (2018). Extending the international new venture phenomenon to digital platform providers: A longitudinal case study. *Journal of World Business*, *53*(5), 725–739. https://doi.org/10.1016/j.jwb.2018.05.001

Ojala, A., Fraccastoro, S., & Gabrielsson, M. (2023). Characteristics of digital artifacts in international endeavors of digital-based international new ventures. *Global Strategy Journal*, *13*(4), 857–887. https://doi.org/10.1002/gsj.1483

Oliva, F. L., Teberga, P. M. F., Testi, L. I. O., Kotabe, M., Giudice, M. D., Kelle, P., & Cunha, M. P. (2022). Risks and critical success factors in the internationalization of born global startups of industry 4.0: A social, environmental, economic, and institutional analysis. *Technological Forecasting and Social Change*, *175*, 121346. https://doi.org/10.1016/j.techfore.2021.121346

Ozalp, H., Ozcan, P., Dinckol, D., Zachariadis, M., & Gawer, A. (2022). "Digital colonization" of highly regulated industries: An analysis of big tech platforms' entry into health care and education. *California Management Review*, *64*(4), 78–107. https://doi.org/10.1177/00081256221094307

Pinheiro, A., Sochirca, E., Afonso, O., & Neves, P. C. (2023). Automation and off(re)shoring: A meta-regression analysis. *International Journal of Production Economics*, *264*, 1–11. https://doi.org/10.1016/j.ijpe.2023.108980

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. S., … Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Rakova, B., Yang, J., Cramer, H., & Chowdhury, R. (2021). Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–23. https://doi.org/10.1145/3449081

Rana, M. B., & Morgan, G. (2019). Twenty-five years of business systems research and lessons for international business studies. *International Business Review*, *28*(3), 513–532. https://doi.org/10.1016/j.ibusrev.2018.11.008

Rana, M. B., & Sørensen, O. J. (2021). Levels of legitimacy development in internationalization: Multinational enterprise and civil society interplay in institutional void. *Global Strategy Journal*, *11*(2), 269–303. https://doi.org/10.1002/gsj.1371

Ratten, V. (2022). Artificial intelligence innovation. In V. Ratten (Ed.), *Managing innovation in organisations: Fostering an entrepreneurial approach* (pp. 95–105). Springer Nature. https://doi.org/10.1007/978-981-19-3100-0_8

Rikap, C. (2022a). Amazon: A story of accumulation through intellectual rentiership and predation. *Competition & Change*, *26*(3–4), 436–466. https://doi.org/10.1177/1024529420932418

Rikap, C. (2022b). Becoming an intellectual monopoly by relying on the national innovation system: The state grid corporation of China's experience. *Research Policy*, *51*(4), 436–466. https://doi.org/10.1016/j.respol.2021.104472

Rios-Morales, R., Gamberger, D., Šmuc, T., & Azuaje, F. (2009). Innovative methods in assessing political risk for business internationalization. *Research in International Business and Finance*, *23*(2), 144–156. https://doi.org/10.1016/j.ribaf.2008.03.011

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of robots in emergency evacuation scenarios. 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 101–108. https://doi.org/10.1109/HRI.2016.7451740

Rodgers, W., Degbey, W. Y., Söderbom, A., & Leijon, S. (2022). Leveraging international R&D teams of portfolio entrepreneurs and management controllers to innovate: Implications of algorithmic decision-making. *Journal of Business Research*, *140*, 232–244. https://doi.org/10.1016/j.jbusres.2021.10.053

Rugman, A. M., & Verbeke, A. (2001). Subsidiary-specific advantages in multinational enterprises. *Strategic Management Journal*, *22*(3), 237–250.

Schlegelmilch, B. B., Sharma, K., & Garg, S. (2023). Employing machine learning for capturing COVID-19 consumer sentiments from six countries: A methodological illustration. *International Marketing Review*, *40*(5), 869–893. https://doi.org/10.1108/IMR-06-2021-0194

Sharkey, A., & Sharkey, N. (2021). We need to talk about deception in social robotics! *Ethics and Information Technology*, *23*(3), 309–316.

Sharma, M., Joshi, S., Luthra, S., & Kumar, A. (2022). Impact of digital assistant attributes on millennials' purchasing intentions: A multi-group analysis using PLS-SEM, artificial neural network and fsQCA. *Information Systems Frontiers*. Advance online publication. https://doi.org/10.1007/s10796-022-10339-5

Sharma, V. K., & Kumar, H. (2023). Enablers driving success of artificial intelligence in business performance: A TISM-MICMAC approach. *IEEE Transactions on Engineering Management*, 1–11. Advance online publication. https://doi.org/10.1109/TEM.2023.3236768

Shendryk, Y., Davy, R., & Thorburn, P. (2021). Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. *Field Crops Research*, *260*, 107984.

Sjödin, D., Parida, V., & Kohtamäki, M. (2023). Artificial intelligence enabling circular business model innovation in digital servitization: Conceptualizing dynamic capabilities, AI capacities, business models and effects. *Technological Forecasting and Social Change*, *197*, 122903. https://doi.org/10.1016/j.techfore.2023.122903

Sjodin, D., Parida, V., Palmie, M., & Wincent, J. (2021). How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. *Journal of Business Research*, *134*, 574–587. https://doi.org/10.1016/j.jbusres.2021.05.009

Sobolev, D. (2023). Rise of the androids: The reflection of developers' characteristics in computerized systems. *British Journal of Management*, *34*(3), 1632–1654. https://doi.org/10.1111/1467-8551.12653

Sohrabpour, V., Oghazi, P., Toorajipour, R., & Nazarpour, A. (2021). Export sales forecasting using artificial intelligence. *Technological Forecasting and Social Change*, *163*, 1. https://doi.org/10.1016/j.techfore.2020.120480

Stemmler, H. (2023). Automated deindustrialization: How global robotization affects emerging economies—Evidence from Brazil. *World Development*, *171*, 1–14. https://doi.org/10.1016/j.worlddev.2023.106349

Strange, R., & Zucchella, A. (2017). Industry 4.0, global value chains and international business. *Multinational Business Review*, *25*(3), 174–184. https://doi.org/10.1108/MBR-05-2017-0028

Studley, M. (2021). Onshoring through automation; perpetuating inequality? *Frontiers in Robotics and AI*, *8*, 1–6. https://doi.org/10.3389/frobt.2021.634297

Sun, Y., Zhu, D., Li, Y., Wang, R., & Ma, R. (2023). Spatial modelling the location choice of large-scale solar photovoltaic power plants:

Application of interpretable machine learning techniques and the national inventory. *Energy Conversion and Management*, *289*, 1–13. https://doi.org/10.1016/j.enconman.2023.117198

Tatarinov, K., Ambos, T. C., & Tschang, F. T. (2023). Scaling digital solutions for wicked problems: Ecosystem versatility. *Journal of International Business Studies*, *54*(4), 631–656. https://doi.org/10.1057/s41267-022-00526-6

Tessmer, A. C., Shaw, M. J., & Gentry, J. A. (1993). Inductive learning for international financial analysis: A layered approach. *Journal of Management Information Systems*, *9*(4), 17–36. https://doi.org/10.1080/07421222.1993.11517976

Tschang, F., & Almirall, E. (2021). Artificial intelligence as augmenting automation: Implications for employment. *Academy of Management Perspectives*, *35*(4), 642–659. https://doi.org/10.5465/amp.2019.0062

Tsilingeridis, O., Moustaka, V., & Vakali, A. (2023). Design and development of a forecasting tool for the identification of new target markets by open time-series data and deep learning methods. *Applied Soft Computing*, *132*, 1–39. https://doi.org/10.1016/j.asoc.2022.109843

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460. https://doi.org/10.1093/mind/LIX.236.433

United Nations. (2023). The 17 goals | sustainable development. Retrieved from https://sdgs.un.org/goals

Veiga, J. F., Lubatkin, M., Calori, R., Very, P., & Tung, Y. A. (2000). Using neural network analysis to uncover the trace effects of national culture. *Journal of International Business Studies*, *31*(2), 223–238. https://doi.org/10.1057/palgrave.jibs.8490903

Verbeke, A., & Hutzschenreuter, T. (2021). The dark side of digital globalization. *Academy of Management Perspectives*, *35*(4), 606–621. https://doi.org/10.5465/amp.2020.0015

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, *11*(1), 1. https://doi.org/10.1038/s41467-019-14108-y

Vuorio, A., & Torkkeli, L. (2023). Dynamic managerial capability portfolios in early internationalising firms. *International Business Review*, *32*(1), 1–15. https://doi.org/10.1016/j.ibusrev.2022.102049

Wang, H. (2022). Green supply chain optimization based on BP neural network. *Frontiers in Neurorobotics*, *16*, 1–10. https://doi.org/10.3389/fnbot.2022.865693

Welch, C., & Piekkari, R. (2017). How should we (not) judge the 'quality' of qualitative research? A re-assessment of current evaluative criteria in international business. *Journal of World Business*, *52*(5), 714–725. https://doi.org/10.1016/j.jwb.2017.05.007

Welch, C., Piekkari, R., Plakoyiannaki, E., & Paavilainen-Mäntymäki, E. (2011). Theorising from case studies: Towards a pluralist future for international business research. *Journal of International Business Studies*, *42*(5), 740–762. https://doi.org/10.1057/jibs.2010.55

Yan, W., Chen, C.-H., & Khoo, L. P. (2001). A radial basis function neural network multicultural factors evaluation engine for product concept development. *Expert Systems*, *18*(5), 219–232. https://doi.org/10.1111/1468-0394.00177

Yu, H., Fletcher, M., & Buck, T. (2022). Managing digital transformation during re-internationalization: Trajectories and implications for performance. *Journal of International Management*, *28*(4), 100947. https://doi.org/10.1016/j.intman.2022.100947

Yu, M., Liu, T., Guan, Z., Sun, Y., Guo, J., Chen, L., & He, Y. (2022). SALSTM: An improved LSTM algorithm for predicting the competitiveness of export products. *International Journal of Intelligent Systems*, *37*(9), 6185–6200. https://doi.org/10.1002/int.22839

Zaheer, S. (1995). Overcoming the liability of foreignness. *Academy of Management Journal*, *38*(2), 341–363. https://doi.org/10.2307/256683

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Algorithmic decision-making and the control problem. *Minds and Machines*, *29*(4), 555–578. https://doi.org/10.1007/s11023-019-09513-7

Zong, K., & Wang, Z. (2022). An empirical study on impact of management capabilities for the multinational company's sustainable competitive advantage. *Journal of Organizational and End User Computing*, *34*(8), 1–23. https://doi.org/10.4018/JOEUC.300763

AUTHOR BIOGRAPHIES

Rakibul Hasan is a Doctoral Researcher in International Business at the University of Vaasa, Finland. Hasan's research is at the cross section of artificial intelligence, international business, information systems, management, marketing, and sustainability. Hasan obtained his Master of Science in Business Administration and Economics at Aalborg University, Denmark.

Arto Ojala is a Professor of International Business at the University of Vaasa, Finland, and a Distinguished Visiting Professor at the Kyoto University, Japan. Ojala also holds the titles of Adjunct Professor in knowledge management from the University of Tampere (in Finland) and in entrepreneurship from the Jyväskylä University School of Business and Economics (in Finland). Ojala's research is at the cross section of international business, information systems, and entrepreneurship. His articles have been published in Global Strategy Journal, Journal of World Business, IEEE Access, International Business Review, Journal of International Marketing, International Marketing Review, Journal of Small Business Management, Information Systems Journal, IEEE Software among others. He received his doctorate in economics from the University of Jyväskylä in 2008, majoring in information systems science.

How to cite this article: Hasan, R., & Ojala, A. (2024). Managing artificial intelligence in international business: Toward a research agenda on sustainable production and consumption. *Thunderbird International Business Review*, *66*(2), 151–170. https://doi.org/10.1002/tie.22369

APPENDIX A: Search protocol

(i) Search strings/sets of keywords

#1: List of search words related to artificial intelligence	#2: List of search words related to international expansion and cross-border perspective
("artificial intelligence" OR "AI" OR "artificial intelligence systems" OR "AI system*" OR "AI agent*" OR "Generative AI" OR "autonomous and intelligent system*" OR "autonomous system*" OR robot* OR chatbot OR "conversational AI" OR "voice assistant" OR "virtual assistant" OR "digital assistant" OR "autonomous product*" OR "autonomous vehicle*" OR "machine learning" OR "deep learning" OR "neural network" OR "natural language processing" OR "large language model" OR "AI-based model" OR "computer vision" OR "responsible AI" OR "AI ethics" OR "AI for social good" OR AI4SG)	("international expansion" OR "global business expansion" OR internationalization OR "internationalization decision*" OR "global expansion" OR globali?ation OR "overseas expansion" OR "international strateg*" OR "global strateg*" OR "firm-specific advantage*" OR "firm-specific asset*" OR "international business" OR "international entrepreneurship" OR "international market*" OR "global market*" OR "cross-cultural market*" OR "international management" OR "cross-cultural management" OR "multinational enterprise*" OR "multinational corporations" OR "multination business enterprise*" OR "internationalized small and medium-sized" OR "internationalization of small and medium-sized" OR "internationalization of SMEs" OR "international SMEs" OR "SME internationalization" OR "international joint venture*" OR "born-global firm*" OR "born-digital firm*" OR "born global and digital" OR "international venture" OR "born global" OR "international business networks" OR "international new venture*" OR "location choice" OR "entry mode*" OR "international partner" OR "international firm*" OR "global firm*" OR "international company*" OR "global company*"OR "global value chain*" OR "international supply chain" OR "global supply chains" OR "international operation*" OR "global operation*")

(ii) Advanced searches in databases

We used each search string separately in the databases and saved searches as #1 and #2, respectively, with such functionality. We then combine both with the 'AND' Boolean operator (#1 AND #2). Otherwise, we used both sets of keywords with the 'AND' Boolean operator in one go. More details can be found below

Web of Science (Web of Science Core Collection):
1) TS = ("artificial intelligence" OR [...]): Search output saved as #1
2) TS = ("international expansion" OR [...]) Search output saved as #2
3) Combined search: #1 AND #2
Note: The search extracted records $N = 631$

Scopus:
1) TITLE-ABS-KEY ("artificial intelligence" OR [...]): Search output saved as #2
2) TITLE-ABS-KEY ("international expansion" OR [...]): Search output saved as #2
3) Combined search: #1 AND #2
Note: The search extracted records $N = 844$

ProQuest (ABI/INFORM Collection):
Combined search: ("artificial intelligence" OR [...]) [All abstract & summery text]
AND ("international expansion" OR [...]) [All abstract & summery text]
Note: The search extracted records $N = 218$

Proceedings of the 58th Hawaii International Conference on System Sciences | 2025

The dark side of AI anthropomorphism: A case of misplaced trustworthiness in service provisions

Rakibul Hasan
University of Vaasa, Finland
rakibul.hasan@uwasa.fi

Arto Ojala
University of Vaasa, Finland
arto.ojala@uwasa.fi

Sara Quach
Griffith University, Australia
s.quach@griffith.edu.au

Park Thaichon
University of Southern Queensland, Australia
park.thaichon@unisq.edu.au

Scott Weaven
Griffith University, Australia
s.weaven@griffith.edu.au

Abstract

Anthropomorphism, the attribution of human-like traits, qualities, and mental states to non-human agents, is increasingly ubiquitous in artificial intelligence (AI) applications within service provisions. While imbuing anthropomorphism into AI agents enhances user interaction, it also has a dark side that poses significant ethical concerns by potentially causing harm to consumers. In response, our study constructs a framework to explain how anthropomorphic features in AI agents can lead to misplaced trustworthiness among consumers. We adopt a sociotechnical perspective and employ a case study design to achieve this. Our findings hope to advance the social sustainability of AI to promote responsible production and consumption patterns in global service markets.

Keywords: Artificial intelligence; service; Anthropomorphism; Misplaced trust; Ethics

1. Introduction

Anthropomorphism is becoming central to human-machine interaction as organizations increasingly induce human-like traits into artificial intelligence (AI)-enabled advanced IT artifacts. These anthropomorphized AI artifacts are becoming ubiquitous in service provisions (Blut et al., 2021; Hasan et al., 2021; Schanke et al., 2021). However, AI agents designed with anthropomorphic features have a dark side that negatively impact service consumption and trustworthiness (Kegel & Stock-Homburg, 2023; Knof et al., 2022).

A dire example of this dark side is highlighted in an investigative report on numerous crashes and fatalities linked to Tesla's autopilot (Marshall, 2024). Specifically, AI agents embedded in Tesla's autopilot that offer autonomous driving capabilities, have misled consumers into overly relying the system's capabilities, resulting in misplaced trustworthiness. Similarly, a recent AI research on sustainability urges to focus on how consumers anthropomorphize autonomy, which can lead to harmful consequences (Hasan & Ojala, 2024).

Existing literature suggests that imbuing anthropomorphic features, such as autonomy, to AI agents can distort consumer behavior when trustworthiness is misplaced (Gill, 2020; Waytz et al., 2014). Furthermore, service organizations may deliberately utilize anthropomorphic features to exploit or mislead consumers by triggering anthropomorphic reasoning and perception (Jörling et al., 2019; Leong & Selinger, 2019). For instance, organizations can use AI agents to change consumers' attributions of responsibility for service failures. These concerns have motivated a focus on the dark side of AI anthropomorphism.

By taking a consumer-level perspective in service provisions, we problematize trustworthiness to offer valuable insights for information systems and service research (Alvesson & Sandberg, 2011; Bartneck et al., 2021). This urgency to protect consumers' wellbeing from potential physical or psychological harm is paramount. This aligns with the AI Act that is designed to mitigate risks posed by AI agents (European Commission, 2024). In this context, we define service provisions to include consumer interactions with products or services whose functionality is enabled by the AI agents (e.g., Google Search Engine, Microsoft Copilot), as well as AI agents intended to be used as safety components of products or services (e.g., Tesla's Autopilot, Replika). In some case, AI agents may operate at full capacity, while in others, their capabilities may not be fully apparent in their behavior.

The dark side of AI anthropomorphism in service provisions is a critical topic. However, systematic investigations that explore the intersection of anthropomorphic features and misplaced trustworthiness are limited (Bartneck et al., 2021). This is worrisome, as service organizations can potentially cause significant

URI: https://hdl.handle.net/10125/109530
978-0-9981331-8-8

harms to consumers through deceptive techniques and by exploiting vulnerabilities (Danaher, 2020; Sharkey & Sharkey, 2020). In response, we pose the below research question: *How do anthropomorphic features enable consumers' misplaced trustworthiness in AI agents within service provisions?*

This study makes three significant contributions. First, it constructs a framework to explain how machine-related factors (e.g., hidden or superficial deceptive techniques) and consumer-related factors (e.g., age-related vulnerabilities) in anthropomorphic design features can deceive and manipulate consumers, leading to misplaced trustworthiness. Second, we outline practical implications for service organizations and policymakers to mitigate the risks associated with AI agents that align with the European Commission's AI Act. Finally, we propose pressing future research opportunities on the ethical aspects of AI anthropomorphism that emphasize the need for transparency and an international business perspective.

2. Research background

While AI anthropomorphism is a burgeoning research area, there is limited direct knowledge about its dark side within information systems and service research (Schanke et al., 2021; Jörling et al., 2019). To address this gap, we provide a backdrop of the dark side of AI anthropomorphism through problematization (Alvesson & Sandberg, 2011). We then direct insights towards a sociotechnical thinking of AI (Berente et al., 2021; Hasan & Ojala, 2024). Our abstraction is drawn from existing theoretical and empirical research that relates to service provisions, either implicitly or explicitly. Accordingly, we construct anthropomorphic features that emphasize consumer-level analysis rather than firms-level analysis. We link these features-specific insights to consumer wellbeing. Consequently, our sociotechnical argumentation embodies the ethical aspects that intersect anthropomorphic features of AI (Bartneck et al., 2021; Hasan et al., 2021).

2.1 What is AI anthropomorphism?

From a sociotechnical standpoint, AI is defined as the frontier of computational advancements to address complex decision-making problems with or without human intervention (Berente et al., 2021; Hasan & Ojala, 2024). Organizations deploy AI-driven advanced IT artifacts in service provisions that have embodied representation (Glikson & Woolley, 2020). Hereafter, we use *AI agents* to delineate embodied representation of AI-driven advanced IT artifacts that imbue anthropomorphic cues and include robotic AI like humanoid robots, virtual AI like chatbot, and embedded AI like self-driving capabilities (Hasan & Ojala, 2024).

The imbuing anthropomorphic features into such AI agents makes them perceived like social entities and influences trustworthiness among consumers (Glikson & Woolley, 2020; Hasan, et al., 2021). Therefore, *AI anthropomorphism can be defined as consumers' tendency to imbue humanlike characteristics, intentions, qualities, and mental states to AI agents*.

2.2 Anthropomorphic features of AI agents

Anthropomorphic features denote the human-like cues imbued into AI agents that can be broadly explained into physical and behavioral features.

Physical anthropomorphic features represent AI agents' complete or partial physical resemblance similar to a human body that include anatomy and structure. They can be broadly divided into appearance and embodiment. *Appearance* involves with outside image of AI agents, such as their face, gender, and shape, and how they look. For instance, an avatar appears as Asian male (see https://soulmachines.com). A human-like appearance can trigger a schema in consumers' perceptions to make them perceive AI as social entities. However, service organizations can use varying level of appearance to alter consumption patterns (Mende et al., 2019).

Embodiment concerns the multisensory bodily representation of AI agents that resembles the physical structure of living beings. For instance, a robot looks gender-neutral that moves and holds like a human (see https://1x.tech). While appearance simply involves the outer look, embodiment relates to how consumers sense and perceive AI agents' bodily representation. However, embodiment significantly affects relationship building and has ethical implications (Vitale et al., 2018).

Behavioral anthropomorphic features involve imbuing AI agents with non-visual human characteristics and capabilities. They can be broadly divided into interactivity and autonomy. *Interactivity* refers to AI agents' ability to engage in reciprocal, two-way communication, and real-time synchronicity (Burgoon et al., 2000). For example, a chatbot designed to engage in human-like conversation (see https://openai.com/chatgpt). The interactive capabilities of AI agents shape consumers' perceptions and behaviors (Schanke et al., 2021).

Autonomy denotes the AI agents' ability to achieve goals without human intervention that highlights their logical reasoning and self-navigating capabilities. For example, a car is designed to have a self-driving capability (see https://waymo.com). In service settings, AI agents that understand consumer behavior, learn from

experience, and demonstrate knowledge raise ethical dilemmas like altering blame attribution (Fosch-Villaronga et al., 2020; Gill, 2020).

Based on forehead insights into anthropomorphic features, we now turn to the dark side from a misplaced trustworthiness perspective.

2.3 A misplaced trustworthiness perspective

Trustworthiness is a multi-faceted concept in human-machine interaction (Glikson & Woolley, 2020; Ullrich et al., 2021; Waytz et al., 2014). It pertains to the degree of confidence consumers have in the actions of AI agents and their feelings towards these agents, regardless of their ability to monitor them. Trustworthiness in AI agents is a central relational element for social acceptance and the intention to use them in service settings (Kegel & Stock-Homburg, 2023). However, it can be problematic when consumers' trustworthiness in AI agents are misplaced (Bartneck et al., 2021). To articulate this issue, we draw insights from three established theoretical lenses that also encompass recent developments in information systems and service literature (Knof et al., 2022; Zheng & Jarvenpaa, 2021).

The 'computers are social actors' paradigm posits that people apply social behaviors to machines when they exhibit human-like visual appearance and behavior (Reeves & Nass, 1996). Consequently, consumers tend to anthropomorphize AI agents mindlessly due to their anthropomorphic features (Knof et al., 2022). However, this automatic response raises ethical concerns about shaping consumers' trustworthiness in AI agents (Kegel & Stock-Homburg, 2023). Therefore, we argue that AI anthropomorphism distorts consumer behavior that potentially impedes their ability to make decisions aligned with their conscious desires and goals. Additionally, consumers' limited information processing capabilities reinforce ethical concerns to impair informed decision-making (Mele et al., 2021).

The three-factor theory of anthropomorphism provides a theoretical lens to describe the process of anthropomorphism (Epley et al., 2007). The theory asserts that humans enhance their interaction experiences with other social entities through inductive inferences related to their elicited own knowledge, effectance needs, and sociality needs. Similarly, consumers rely on their self-concept during interactions with AI agents, drawing from existing knowledge about themselves or other humans (Zheng & Jarvenpaa, 2021). However, tapping into such inferences is problematic. For instance, elicited agent knowledge and sociality motivation prevent consumers from recognizing dishonest anthropomorphism in AI agents, thereby distorting privacy concerns (Benlian et al., 2020; Leong & Selinger, 2019). Therefore, we maintain such dishonest features facilitate the development of trustworthiness that distorts consumer behavior.

The mind perception hypothesis suggests a tendency to attribute mind to non-human agents or non-living entities in two dimensions: agency and experience (Gray et al., 2007). Agency implies the capability of rational thoughts, while experience indicates the ability to have pain, emotion, and feelings. AI agents that signal human-like mental capabilities develop trustworthiness among consumers to an extent that distorts their moral reasoning and behavior (Gill, 2020; Waytz et al., 2014). A critical consequence of attributing a mind to AI agents is that consumers perceive them as a moral agent. Hence, we contend that such misplaced trustworthiness alters consumers' responsibility and blame attribution towards service organizations.

3. Methodology

In our research process and reporting, we go beyond existing methodological template and writing genres (Avital et al., 2017). We embrace our value judgments and imagination in knowledge production that embody critical realism as an underpinning philosophy (Bhaskar, 1978; Mingers, 2004). Consequently, such research approach changes the dynamics of quality assessment of our study (Grover & Lyytinen, 2015).

Research design Our study employs a qualitative single contextualized explanation case study design (Welch et al., 2011). It is an appropriate research design given that we aim to explain the phenomena of the dark side of AI anthropomorphism and contextualize research findings within service provisions. The unit of analysis is the utilization of anthropomorphic features in AI agents that enable misplaced trustworthiness among consumers. Accordingly, our empirical observations focus on consumers who interact with AI agents in service settings. Through iterative construction, we develop a case of consumers' misplaced trustworthiness in AI agents within service provisions.

Data sources We corroborate with interviews data and exiting knowledge on AI anthropomorphism (Miles et al., 2014; Sarker et al., 2013). Using a revelatory sampling logic, we recruited participants (experts and consumers) via LinkedIn and internal networks. Participants include experts with AI-related experience in industry and/or academia, and consumers who regularly interact with AI agents. To capture diverse experiences, interviewees varied in their interaction with AI agents, gender, professions, and nationality (see Table 1). We conducted 14 interviews via Teams/Zoom and face-to-face between May and June 2021, each lasting around 20 to 70 minutes. We used two separated,

yet interrelated question guides for experts and consumers, asking about their views and experiences with AI agents and potential ethical concerns.

Table 1. Qualitative interview sample description

Identifier	Age	Gender	Country of origin [resident]	Types of experiences
Interview type: Consumer				
C1	24	M	Anonymous [Australia]	Virtual + robotic AI—Smart home and financial service
C2	36	F	China [Australia]	Robotic AI—Smart home and child learning service
C3	37	F	Iran [Australia]	Embedded + virtual AI—navigation and smart living
C4	30	M	Pakistan [United Arab Emirates]	Virtual AI—Service inquiry
C5	22	M	Australia [same]	Embedded + virtual AI—Smart home and recruiting service
C6	27	F	China [same]	Robotic AI—Smart home and service delivery
C7	25	M	United States [same]	Embedded + virtual AI—service inquiry and autonomous driving
Interview type: Expert				
E1	35	F	Brazil [Australia]	Embedded + virtual AI—Senior data privacy consultant and users of smart home
E2	26	F	France [Germany]	Robotic + virtual AI—Young researcher on AI ethics and online games
E3	37	M	Singapore [Australia]	Virtual AI—Young researcher on AI and service
E4	25	F	Mexico [same]	Embedded AI—AI developer and tech philosopher
E5	54	M	United Kingdom [New Zealand]	Virtual AI—Senior AI scientist and developer
E6	26	M	Germany [same]	Embedded AI—AI developer and entrepreneur
E7	59	M	Thailand [same]	Robotic + embedded AI—Senior tech philosopher

Analytical approach We employed a configurational theorizing approach that involves three interconnected stages of casing (Furnari et al., 2021). Initially, we used *scoping* to identify relevant anthropomorphic features as an anchor that may plausibly form configurations leading to misplaced trustworthiness. We then aggregate these features into higher-order construct through abduction that combine our observation, insights from interviews, and existing knowledge on AI anthropomorphism. Subsequently, we applied *linking* to specify how these features relate within specific configurations and constantly seek patterns of interdependencies (contingency or complementarity) among features leading to the outcome. We explored multiple explanatory factors (absence and presence) to explain the mechanisms behind misplaced trustworthiness. Finally, we used *naming* to abstract configurations to link orchestration themes and articulate overarching narrative in simple terms. Overall, our theorization recursively went back and forth between literature and interview data until the completion of this paper.

4. Discussion of findings

4.1 Undesired outcome of misplaced trustworthiness

In service provisions, we postulate that the undesired outcome of misplaced trustworthiness towards AI agents occurs when consumers trust an AI agent seemingly beyond its actual capabilities and tend to form social relationship with it. For instance, robotic AI can be equipped with thermal sensors to enable to see through walls, lip-reading abilities, or powerful hearing capabilities. Additionally, such AI agents can be connected to service providers' marketing insight cloud that covertly mines consumer behavior to offer personalized services. Consequently, AI agents' anthropomorphic features create human-level expectations and obscure predictions about their hidden capabilities, triggering misplaced trustworthiness. As a result, consumers may not fully trust AI agents but still reveal and share sensitive information to access desired services. This paradoxical behavior is evident among consumers, as illustrated follows:

"*It's easy to sit back and say that the privacy issue this and that, but everything has a price. […] We have a price to pay. We can't just sit back and be the end-user.*" [C3]

With such dishonest techniques, service providers can imbue anthropomorphic features to exploit consumers' privacy paradox (Benlian et al., 2020) and distort moral judgement (Giroux et al., 2022). One expert highlighted the deceptive nature of AI agents concerning misplaced trustworthiness, as follows:

"By essence, a robot is deceptive. The consequence of that is intended that you trust your robot to be human-like, especially for a social robot. You trust your robot to be what you expect it to be […] there are two types of trust, cognitive trust and emotional trust". [E2]

Consumers continue to use services despite heightened privacy concerns to trade personal information for hedonic value or the convenience of superior personalized services. Organizations may manipulate consumers to distort their risk tolerance of privacy violation through trust mechanisms. Therefore, we contend that service organizations, knowingly or unknowingly, utilize anthropomorphic features that lead to the undesired outcome of consumers' misplaced trustworthiness in AI agents.

4.2 Different triggering effects of anthropomorphic features on outcomes

Each anthropomorphic feature contributes to misplaced trustworthiness distinct ways that raise ethical concerns. An expert's explanation highlights these issues, as follows:

"We human are really bad at emotional forecasting [...]. AI does have a strong and powerful computational system that is capable of taking decisions and making calculations much faster than we can. […]. We definitely have high expectations of high-functional AI systems. [E4]

This explanation implicitly contemplates the types of service tasks linked to anthropomorphic features. These tasks can broadly divide into two categories based on their perceived objectivity and subjectivity (Castelo et al., 2019). Objective tasks require analytical reasoning and adherence to routine procedures, while subjective tasks necessitate care and responsiveness to individual needs. The former relates the agency (e.g., logical thinking) and the latter relates to experience (e.g., feelings).

The detrimental effect of physical features For *objective tasks*, appearance is positively linked to the perception of competency to distort consumption patterns (Mende et al., 2019). Similarly, service providers imbue appearance to manifest competence and influence privacy concerns (Benlian et al., 2020). Consequently, consumers overly rely on AI that cause significant damage (Robinette et al., 2016). We observe that the purpose is to create an illusion of human-likeness in AI agents to distort behavior and an impression of reliability for completing objective tasks.

Regarding *subjective tasks*, service providers imbue AI agents with embodiment features to manifest emotive behaviors like caregiving. In this regard, one consumer commented on a potential ethical concern, as follows:

"*Humans may even get married with robots right in the future, but I think maybe the human beings will be controlled by technology. Yeah, I think it's so terrible*." [C6]

Physical features like embodiment create a false belief of experience to develop affection. Hence, consumers perceive AI agents to have feelings and develop social relationships. The consequences can be far-reaching that may reduce population growth and encourage family separation.

The detrimental effect of behavioral features For *objective tasks*, some AI agents have a minimal physical feature but are instilled with behavioral features like autonomy. For instance, ChatGPT and Tesla's autopilot, a form of a virtual AI and an embedded AI respectively, demonstrate autonomy to complete tasks through quality information exchange and driving capabilities. In doing so, consumers constantly share data with service providers. Signaling autonomy through data-driven learning and custom responses enables service providers to nurture trust and exploit consumers through manipulation.

Regarding *subjective tasks*, we observe that interactivity (e.g., communications skills or style) enables misplaced trustworthiness as explained by one expert, as follows:

"*Emotional reactions are coded. It just tells the robot if the person in front of you cries. Then you are supposed to be sad. It shows emotions so that people will recognize emotions even though it is fake."* [E2]

Such coded trust-building mechanisms in AI agents induce higher superficial social presence that is key to developing emotional attachment and social bonding (Sharkey & Sharkey, 2021). However, this superficial mechanism enables service providers to nurture trust through emotional deception.

4.3 Interplay of machine- and consumer-related factors that amplify outcomes

Machine-related deceptive techniques We contend that service organizations can deliberately deceive consumers through anthropomorphic features with the intention to mislead. One consumer expressed concern that organizations might deploy AI agents in inappropriate service scenarios, as follows:

"What will you do when exposed to a certain stimulus? And they will learn to then manipulate those stimuli to almost subconsciously make you do what with what is beneficial to them" [C5]

Additionally, an expert highlighted about potential deception, as follows:

"*It has the power to collect all kinds of information because it's there, right? There are all kinds of sensors. It can pick up not only the verbal information […] but the ambient, their surroundings, the physical surrounding of the place.*" [E7]

Forehead explanations suggest that consumers are unable to predict the consequences of interaction with AI agents, as these interactions often exceed their general expectations due to deception. Deception via AI anthropomorphism can be achieved in two primary techniques: superficial and hidden (Danaher, 2020). Superficial deceptive techniques involve AI agents to display human-like capabilities or create an impression of such capabilities that they do not have. In contrast, hidden deceptive techniques involve AI agents to conceal their true internal capabilities through creating impression of absenteeism that they actually have.

Concerning *superficial deceptive techniques*, AI anthropomorphism creates social expectations about capabilities based on consumers knowledge of humans

and societal institutions (e.g., culture, race). One expert discussed misplaced trustworthiness based on deceptive features that portray gender: *"There is racism and sexism right there"* [E2]. For example, robotic AI in physical service settings are often designed to appear white and male that signal a particular race and gender. Consequently, consumers apply social categorization to AI agents as they do to humans (Eyssel & Kuchenbrandt, 2012). Unfortunately, such social categorization can have detrimental effects on consumers' wellbeing.

Another expert shared personal, yet traumatizing experience on social categorization leading to *"breaking up family relationship"* [E1], triggered by behavior features like interactivity. The expert, who had migrated to an English-speaking country, and had a partner who spoke native English. They regularly interacted with a virtual AI like Amazon Alexa, which works with voice command and country-specific accents (e.g., Australian English). Due to her inherited foreign accent, the AI agents sometimes failed to comprehend her commands. However, her partner overly trusted the AI agents and began to treat her as an outsider, thereby damaging her dignity. Ultimately, the expert had to end the relationship due to misplaced trustworthiness toward the AI agents.

Concerning *hidden deceptive techniques*, AI agents can be "*equipped with some capabilities beyond human level" [E7]*. For instance, a robotic AI might have hidden infra-red cameras or sensors to create a digital twin of its surroundings, which humans cannot perceive with bare eyes. These hidden capabilities make consumers vulnerable to safeguard against and create issues related to the intrusion of physical privacy. Such unknown and unanticipated hidden deception leads to misleading understanding.

Consumer-related personal factors We foresee that consumers' vulnerability is a critical factor that place them in situations where they might be targeted or exposed to potential harm. Such vulnerability may arise due to limited cognitive capabilities and age. One consumer expressed concerns regarding children, as follows:

"*I'm concerned […], it provides benefits in terms of learning and motivation […]. However, I worry that she will lack of social communication skills with others to interact with*". [C2]

Additionally, an expert explained the tendency of elderly individuals to share more personal information due to increased tolerance of privacy invasion, as follows:

"In case with the elderly and the elderly being alone most of the time, […] they put trust in the machine, and they talk about all their private information". [E7]

Theses explanations illustrate how vulnerability emerges from consumers' ages, particularly for children and the elderly, driven by higher societal motivation. For children, interacting with robotic AI can create emotional bonds, potentially hampering their mental and social development (van Straten et al., 2020). They may miss opportunities to learn relationship formation skills with other significant living beings.

Regarding ethical aspects of human dignity, elderly people are the most vulnerable. Care providers are increasingly turning to AI agents to deliver care services (van Maris et al., 2020). However, elderly individuals may fall prey to misplaced trustworthiness due to their lack of cognitive capabilities and higher societal motivations. They easily establish social bonds with AI agents as they consider them as friends and trustworthy companion. In summary, anthropomorphic features create a superficial impression of emotional state among vulnerable consumers. This impression enables them to perceive machines as social entities and form social relationships.

4.4 The constructed framework

Constructed from the insights of our study, we now presented a framework that complements existing literature that provides explanatory factors for the dark side of AI anthropomorphism in service provisions (see Figure 1). Our theorization contends that anthropomorphic features imbued in AI agents enable consumers to develop misplaced trustworthiness. A dreadful example of such undesired outcomes is the death of consumers who overly trusted autonomous driving capabilities of cars (Marshall, 2024). In this process, physical features like appearance and embodiment, along with behavioral features like autonomy and interactivity, distort consumers' perception and expectations, thereby leading misplaced trustworthiness.

Physical and behavioral anthropomorphic features enable misplaced trustworthiness Drawing from the evidence of this study, anthropomorphic features create false beliefs that lead to misplaced trustworthiness in two ways (Glikson & Woolley, 2020). On one hand, consumers develop misplaced cognitive trustworthiness through uncalibrated beliefs about AI agents' capabilities and performance. On the other hand, consumers develop misplaced affective trustworthiness through uncalibrated motivation to build social relationships with AI agents.

Concerning *misplaced cognitive trustworthiness*, consumers tend to trust AI agents beyond their actual capabilities and performances due to the presence of anthropomorphic features. Alarmingly, consumers'

awareness about faulty AI systems does not mitigate the risk of trust, even in life-threatening situations (Robinette et al., 2016; Ullrich et al., 2021). Exploiting such trust through anthropomorphism is unethical, as it aims to achieve economic goals by creating engagement and taking advantage of consumers' kindness. For instance, AI agents can be deployed to encourage overconsumption based on personal buying patterns and social media data, which could deleteriously impact on consumer health outcomes. Additionally, consumers tend to perceive AI agents mindlessly as accurate in information exchange that make them vulnerable to impair informed decision-making. A recent study demonstrates how AI can be trained to recognize vulnerabilities by understanding human behaviors and preferences (Dezfouli et al., 2020). Such AI-driven learning can be used to distort behavior for profit-seeking purposes.

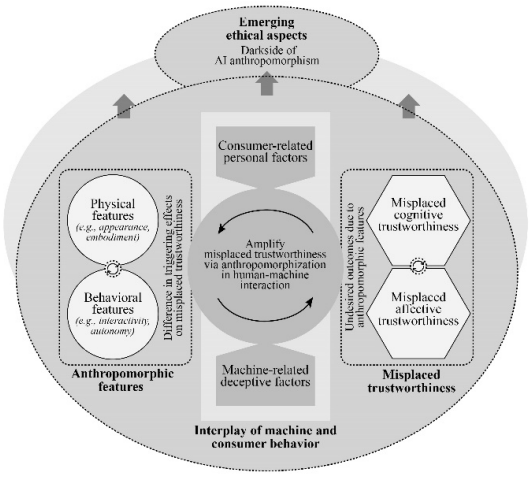


Figure 1. The constructed framework into dark side of AI anthropomorphism

Regarding *misplaced affective trustworthiness*, consumers tend to develop a psychological connection with AI agents due to the presence of anthropomorphic features. For example, consumers perceive virtual AI as sociable and develop feelings of sentience towards these agents (Qiu & Benbasat, 2009; Schweitzer et al., 2019). Such affective or emotional perceptions nurture a sense of connectedness and lead to the development of emotive relationship with AI agents (Chang et al., 2023). Organizations may employ AI agents on the service front lines, sometimes without disclosing that consumers are interacting with a non-human agent, which is legally challengeable as it involves deception.

Furthermore, service providers can induce companionship through AI anthropomorphism. They can tap into a person's societal motivation and affective feelings to develop a willingness to form partnerships. Such companionship can be achieved through deceptive anthropomorphic signals that create a social presence, which, although not human, encourage affective interaction. This extended attachment can be observed with Replika, a form of virtual AI that encourage companionships (see, reddit.com/r/replika). Moreover, consumers' feelings of emotional connection with AI agents elicit mind attribution to humanize the AI to an extend that consumers may consider them as moral patients (Waytz et al., 2014). Consequently, organizations might employ AI agents exclusively to maximize profit, thereby reducing human touch and increasing the risk of losing human dignity.

The interplay of machine and consumer behavior on misplaced trustworthiness The findings of this study suggest that anthropomorphic features consistently lead to misplaced trustworthiness, through effects vary based on perceived objectivity and subjectivity of tasks. For subjective tasks, appearance and autonomy seem to have a greater effect to nurture misplaced cognitive trustworthiness though manipulation. For objective tasks, embodiment and interactivity appear to have a greater effect to develop misplaced affective trustworthiness through emotional deception. These undesired relational outcomes result in detrimental consequences that are emerged from distorting behavior and impair informed decision-making, including privacy violation, loss of human autonomy, loss of human dignity, and altering moral responsibility.

To illustrate our argument at the intersection of behavioral features and consumers' moral responsibility, consider the negative impact of autonomy feature. Behavioral features like autonomy lead to the perception that AI agents are capable of make critical reasoning, thereby making moral decisions (Jörling et al., 2019). This is disturbing because some machine learning techniques, such as deep learning, cannot explain why a decision was made. Moreover, AI agents lack an understanding of the consequences if their decisions on consumers' lives. Conversely, humans have limited information process capabilities and unconsciously project their own knowledge onto AI agents, enabling mindless anthropomorphism. These combined effects from both machines and consumers create an illusion of moral understanding and competency (Arikan et al., 2023; Gill, 2020). This interplay also nurtures the perception of AI agents as morally significant entities that alter consumer behavior. Consumers may develop trust in AI agents to the extent that they attribute blame to the machines for service failure instead of service providers. Therefore, the insights of our study have

important implications for theory, practices, and future research.

5. Conclusion

5.1 Knowledge implications

Our study contributes to the understanding of the dark side of AI two significant ways. First, we constructed a framework that focuses on AI anthropomorphism to explain and contextualize undesired outcomes of misplaced trustworthiness in service provisions. Our theorization articulates the mechanisms on how AI agents' anthropomorphic features enable consumers to develop misplaced cognitive and affective trustworthiness. We explicate the differing triggering effects of physical and behavioral anthropomorphic features on these undesired outcomes. Additionally, we highlight the interplay of machine and consumer behavior and how these behavioral factors amplify consumers' misplaced trustworthiness toward AI agents.

Second, our study enhances existing sociotechnical thinking of AI by integrating anthropomorphism to emphasize consumers' wellbeing (Berente et al., 2021; Hasan & Ojala, 2024). We extend sociotechnical perspective to the dark side of AI anthropomorphism, a cross-disciplinary phenomenon and an emerging research topic. Our study provides new insights into AI anthropomorphism by problematizing trust at the intersection of ethical considerations to protect consumers' wellbeing. We demonstrate how the dark side of AI anthropomorphism increases risks through deception and manipulation that can potentially trigger legal issues for service organizations.

5.2 Practical implications

The findings of our study directly relate to the European Commission's AI Act (2024), the first legal framework that applies a risk-based approach (unacceptable, high, limited, and minimal). Our theorization articulates that machine-related (hidden or superficial deceptive techniques) and consumers-related factors (vulnerability related to age) of anthropomorphic design features deceive and manipulate consumers into developing misplaced trustworthiness. This has implications for service organizations and policymakers to mitigate the risks associated with AI agents.

Organizational implications Organizations must disclose the details of instilled capabilities (hidden and visible) of AI agents and potential unintended behavioral distortion that may affect consumers. For example, they must disclose whether consumers are interacting with AI agents or humans in service encounters to avoid limited risk. Organizations are increasingly deploying AI agents in customer encounters in online service settings. Without proper disclosure, consumers may believe they are interacting with a human, only to realize later that it was an AI agent. This deception can result in a counterfeit service experience that lead to dissatisfaction and evoke switching intention.

Policy implications We urge policymakers to explicitly address the dark side of AI anthropomorphism to ensure accountability. The current AI Act (2024) does not capture the unacceptable risks that arise from AI anthropomorphism. However, our study argues that AI anthropomorphism impairs informed decision-making through deceptive techniques and exploiting vulnerabilities. Consumers develop trustworthiness towards AI agents to the extent that they start to attribute blame to machines for service failure instead of service providers. This is alarming, as profit-seeking service organizations can deliberately induce anthropomorphic feature-driven relational mechanisms to distort behavior to avoid accountability. For example, organizations can deploy AI agents with anthropomorphic features to alter legal responsibility for their services failures that cause significant harms, such as deaths.

5.3 Limitations and future research

Of course, the insights from our study, and the framework we constructed from them, need to be interpreted carefully, as they are confined to a specific context of service provision. We aimed to generate explanatory factors for a context-laden phenomenon rather than achieving generalization. Additionally, further research is needed to better understand misplaced trustworthiness in AI anthropomorphism. Accordingly, we outline two future research opportunities.

Transparency There is an opportunity to extend the constructed framework by examining the role of transparency. While our theorization captures the interplay of machine-and consumer-related behavioral factors, it does not address the dynamics of transparency. We foresee that the absence of transparency creates an ethical vacuum in the dark side of AI anthropomorphism. Therefore, the scope of our study should be extended to include transparency in the development developing misplaced versus calibrated trustworthiness. One might assume that transparency in human-machine interaction enables calibrated trustworthiness. However, there could be an ethical dilemma where ensuring transparency also reduces trustworthiness towards AI agents (van Straten et al., 2020). Future research that focuses on transparency to

achieve the right calibration of trustworthiness is paramount.

Institutional differences Given the micro-level focus, our theorization did not capture macro-level understanding. However, we anticipate that the tendency to develop misplaced trustworthiness may vary depending on given institutional settings (e.g., norms, values, cultures). Consumers interact with AI agents with different institution-specific (e.g., culture, race) cues. They also come from diverse geographic locations with different facial structures and (non) verbal communication styles. Accordingly, we see value in extending our insights to international business perspective by focusing on institutional distance and consumption patterns (Hasan & Ojala, 2024). For example, researchers can focus on a particular location-specific institution like national culture, which profoundly impact how people anthropomorphize AI agents (Eyssel & Kuchenbrandt, 2012). Middle Eastern consumers may conspicuously perceive and anthropomorphize AI agents differently than South Asians or Nordics due to their institutional differences. Such differences may also shape ethical considerations and influence trust formation on AI anthropomorphism, thereby enabling a juxtaposition of findings. With the insights from our study, we hope to promote the social sustainability of AI agents to ensure responsible production and consumption patterns in global service markets.

Acknowledgment

The first author gratefully acknowledges the support received from the Griffith University (International) Postgraduate Research Scholarship, the Finland Fellowship from the Finnish Ministry of Education and Culture, and the Finnish Foundation for Economic Education.

References

Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *The Academy of Management Review*, 36(2), 247–271.

Arikan, E., Altinigne, N., Kuzgun, E., & Okan, M. (2023). May robots be held responsible for service failure and recovery? The role of robot service provider agents' human-likeness. *Journal of Retailing and Consumer Services*, *70*, 103175.

Avital, M., Mathiassen, L., & Schultze, U. (2017). Alternative genres in information systems research. *European Journal of Information Systems*, *26*(3), 240–247.

Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *An introduction to ethics in robotics and AI*. Springer Nature.

Benlian, A., Klumpe, J., & Hinz, O. (2020). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal*, *30*(6),

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433–1450.

Bhaskar, R. (1978). A realist theory of science. Harvester Press.

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*(4), 632–658.

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human–computer interaction: A study of credibility, understanding, and influence. *Computers in Human Behavior*, *16*(6), 553–574.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, *56*(5), 809–825.

Chang, Y., Gao, Y., Zhu, D., & Safeer, A. (2023). Social robots: Partner or intruder in the home? The roles of self-construal, social support, and relationship intrusion in consumer preference. *Technological Forecasting and Social Change*, *197*.

Danaher, J. (2020). Robot Betrayal: A guide to the ethics of robotic deception. *Ethics and Information Technology*, *22*(2), 117–128.

Dezfouli, A., Nock, R., & Dayan, P. (2020). Adversarial vulnerabilities of human decision-making. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(46), 29221–29228.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864.

European Commission. (2024, August 8). AI Act. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, *51*(4), 724–731.

Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Gathering Expert Opinions for Social Robots' Ethical, Legal, and Societal Concerns: Findings from Four International Workshops. *International Journal of Social Robotics*, *12*(2), 441–458.

Furnari, S., Crilly, D., Misangyi, V. F., Greckhamer, T., Fiss, P. C., & Aguilera, R. V. (2021). Capturing Causal Complexity: Heuristics for Configurational Theorizing. *Academy of Management Review*, *46*(4), 778–799.

Gill, T. (2020). Blame It on the Self-Driving Car: How Autonomous Vehicles Can Alter Consumer Morality. *Journal of Consumer Research*, *47*(2), 272–291.

Giroux, M., Kim, J., Lee, J. C., & Park, J. (2022). Artificial Intelligence and Declined Guilt: Retailing Morality Comparison Between Human and AI. *Journal of Business Ethics*, *178*(4), 1027–1041.

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619.

Grover, V., & Lyytinen, K. (2015). New State of Play in Information Systems Research: The Push to the Edges. *MIS Quarterly*, *39*(2), 271–296.

Hasan, R., & Ojala, A. (2024). Managing artificial intelligence in international business: Toward a research agenda on sustainable production and consumption. *Thunderbird International Business Review*, *66*(2), 151–170.

Hasan, R., Thaichon, P., & Weaven, S. (2021). Are we already living with Skynet? Anthropomorphic artificial intelligence to enhance customer experience. In Thaichon, P. & Ratten, V. (Eds), *Developing digital marketing* (pp. 103-134). Emerald Publishing Limited.

Jörling, M., Böhm, R., & Paluch, S. (2019). Service Robots: Drivers of Perceived Responsibility for Service Outcomes. *Journal of Service Research*, *22*(4), 404–420.

Kegel, M. M., & Stock-Homburg, R. M. (2023). Customer Responses to (Im)Moral Behavior of Service Robots Online Experiments in a Retail Setting. *Proceedings of the 56th Hawaii International Conference on System Science (*pp. 1500–1509).

Kim, T. W., Jiang, L., Duhachek, A., Lee, H., & Garvey, A. (2022). Do You Mind if I Ask You a Personal Question? How AI Service Agents Alter Consumer Self-Disclosure. *Journal of Service Research*, *25*(4), 649–666.

Knof, M., Heinisch, J. S., Kirchhoff, J., Rawal, N., David, K., von Stryk, O., & Stock-Homburg, R. (2022). Implications from Responsible Human-Robot Interaction with Anthropomorphic Service Robots for Design Science. *Proceedings of the 55th Hawaii International Conference on System Sciences* (pp. 5827–5836).

Leong, B., & Selinger, E. (2019). Robot Eyes Wide Shut: Understanding Dishonest Anthropomorphism. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 299–308.

Marshall, A. (2024, April 26). Tesla Autopilot Was Uniquely Risky—And May Still Be. Wired. https://www.wired.com/story/tesla-autopilot-risky-deaths-crashes-nhtsa-investigation/

Mele, C., Russo Spena, T., Kaartemo, V., & Marzullo, M. L. (2021). Smart nudging: How cognitive technologies enable choice architectures for value co-creation. *Journal of Business Research*, *129*, 949–960.

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, *56*(4), 535–556.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (Edition 3). Sage.

Mingers, J. (2004). Realizing information systems: Critical realism as an underpinning philosophy for information systems. Information and Organization, 14(2), 87–103.

Qiu, L., & Benbasat, I. (2009). Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. *Journal of Management Information Systems*, *25*(4), 145–182.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places* (pp. xiv, 305). Cambridge University Press.

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of robots in emergency evacuation scenarios. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 101–108.

Sarker, S., Xiao, X., & Beaulieu, T. (2013). Guest editorial: Qualitative studies in information systems: A critical review and some guiding principles. *MIS Quarterly*, 37(4), iii-xviii.

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the Impact of "Humanizing" Customer Service Chatbots. *Information Systems Research*, *32*(3), 736–751.

Schweitzer, F., Belk, R., Jordan, W., & Ortner, M. (2019). Servant, friend or master? The relationships users build with voice-controlled smart devices. *Journal of Marketing Management*, *35*(7–8), 693–715.

Sharkey, A., & Sharkey, N. (2021). We need to talk about deception in social robotics! *Ethics and Information Technology*, *23*(3), 309–316.

Ullrich, D., Butz, A., & Diefenbach, S. (2021). The Development of Overtrust: An Empirical Simulation and Psychological Analysis in the Context of Human–Robot Interaction. *Frontiers in Robotics and AI*, *8*.

van Maris, A., Zook, N., Caleb-Solly, P., Studley, M., Winfield, A., & Dogramadzi, S. (2020). Designing Ethical Social Robots-A Longitudinal Field Study With Older Adults. *Frontiers in Robotics and AI, 7*, 14, Article 1.

van Straten, C. L., PeterJochen, KühneRinaldo, & BarcoAlex. (2020). Transparency about a Robot's Lack of Human Psychological Capacities. *ACM Transactions on Human-Robot Interaction (THRI)*.

Vitale, J., Tonkin, M., Herse, S., Ojha, S., Clark, J., Williams, M.-A., Wang, X., & Judge, W. (2018). Be More Transparent and Users Will Like You: A Robot Privacy and User Experience Design Experiment. *Proceedings of 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 379–387.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117.

Welch, C., Piekkari, R., Plakoyiannaki, E., & Paavilainen-Mäntymäki, E. (2011). Theorising from case studies: Towards a pluralist future for international business research. *Journal of International Business Studies*, *42*(5), 740–762.

Zheng, J., & Jarvenpaa, S. (2021). Thinking Technology as Human: Affordances, Technology Features, and Egocentric Biases in Technology Anthropomorphism. *Journal of the Association for Information Systems*, *22*(5), 1429–1453.

Artificial intelligence and anthropomorphism: A case of risk of harm on the global stage

Rakibul Hasan[1]

[1]School of Marketing and Communication,
University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland
Email: rakibul.hasan@uwasa.fi, ORCID ID: 0000-0003-2088-0490.

Abstract

Advanced IT artifacts powered by artificial intelligence (AI) are increasingly embedded in global service provision. These computational artifacts, in both robotic and virtual forms, possess facets such as appearance or autonomy that give rise to contemporary AI systems to perform service tasks. These systems display human-like features that trigger AI anthropomorphism (AIA). But, as users begin to attribute human-like traits, emotions, or intentions to these systems, concerns about deception and manipulation are growing across national borders. Recent incidents involving intimacy, violence, self-harm, and the exploitation of vulnerable users highlight the urgent need to understand how AI create a risk of harm under anthropomorphism. Especially as these systems spread across markets and enter the daily lives of hundreds of millions of users worldwide, negative consequences become prevalent. These concerns have thus become central to regulatory debates, including the AI Act, which identifies several unacceptable risks related to deception or manipulation. Therefore, there are increasing opportunities for novel theorization in information systems (IS) and international business (IB) scholarship. Despite increasing opportunities, both IB and IS research lack a detailed explanation of how AIA enables harmful outcomes and jeopardizes user well-being. This study addresses this gap by examining the mechanisms through which risks of harm emerge under AIA in the contexts of service tasks. I build upon perspectives from three discourses, namely, international strategy in IB, information management in IS, and transdisciplinary conversations in AI. Using a mechanism-based case study design, I draw on qualitative interviews, a review of empirical research, and a constructed primary archival media dataset. These data sources provide empirical insights from more than thirty countries that address firm-level actions and user affordances around both robotic and virtual AI systems. The study makes two main contributions to existing conversations in IS and IB. First, I construct key facets of AI regarding anthropomorphism and explain how they create risks of harm through deception and manipulation in information management discourse. Second, I suggest that consumption convergence enabled by AI may generate systemic vulnerabilities for users, leading to harmful outcomes through anthropomorphism that may propagate across national borders. Besides contributing to theoretical conversations, this study has relevance for managerial and regulatory implications.

Keywords: Anthropomorphism, Artificial intelligence, Risk of harm, Deception, Manipulation, Information management, International strategy

INTRODUCTION

Advanced IT artifacts powered by the artificial intelligence (AI) frontier are transforming service provision worldwide (Berente et al., 2021; Schanke et al., 2021). These artifacts possess facets such as appearance or autonomy that give rise to contemporary AI, referring to evolving computational systems that perform goal-directed tasks and display human-like cues that shape user perceptions and behavior. However, hundreds of millions of users around the globe interact with AI for service tasks in both robotic and virtual forms (Glikson & Woolley, 2020; Microsoft, 2026; Rahwan et al., 2019). In turn, firm-specific actors like top management take actions around strategic choices to enable human-like features into these non-human systems, triggering AI anthropomorphism (AIA) (Zheng & Jarvenpaa, 2021; Złotowski et al., 2020). AIA can be understood as a tendency by which users ascribe human qualities such as emotions, intentions, and mental states to AI. As AIA becomes more common, new questions emerge about the risks it may create, and as a result, negative consequences related to AI are becoming more visible. Recent reports show incidents involving intimacy, violence, self-harm, and even death or murder linked to interactions with AI, including the exploitation of vulnerable users (Burga, 2025; Horwitz, 2025c; Jargon & Kessler, 2025; Marsden, 2025). These incidents raise serious concerns about deception and manipulation. Because of this, AI is now central to market-based regulatory institutions, namely the AI Act, which identifies several unacceptable risks (European Commission, 2024). Such risk of harm brings forward some speculations: Do firm-level actions enable AIA to falsely signal or distort users' behavior? What harmful outcomes may follow from this? Which anthropomorphic features create these risks? And does the AI Act provide enough protection against deception and manipulation caused by contemporary AI under anthropomorphism?

Risks of harm under AIA are contextual and experiential events or circumstances that negatively impact well-being when individuals engage with an anthropomorphized AI or an anthropomorphized AI-mediated system. These risks are not theoretical or futuristic. Contemporary AI systems are not merely a product of evolving computational advancement (Berente et al., 2021), and they are often deliberately developed and deployed to act as social partners, leading to intended or unintended consequences. Such outcomes make AIA and any potential risk of harm a critical issue. Hence, a novel opportunity is presented to engage in conversation that intersects international strategy in international business (IB) and information management in information systems (IS) scholarship. While existing conversations in international strategy discourse provide little insight into AIA (Hasan & Ojala, 2024; Kopalle et al., 2022), information management discourse is well advanced. It explains what gives rise to AIA (Brendel et al., 2023; Zheng & Jarvenpaa, 2021), and how different features affect users' anthropomorphizing (Benlian et al., 2020; Diederich et al., 2022; Qiu & Benbasat, 2009). It also shows the dark sides of AIA that pose risks to alter privacy concerns (Benlian et al., 2020), develop misplaced trustworthiness (Hasan et al., 2025), or increase user aggression (Brendel et al.,

2023). Some studies also point to risks of deception, a loss of human dignity, and to manipulation (Diederich et al., 2022; Knof et al., 2022; Schanke et al., 2021). Although these findings highlight a growing dark side of AI, existing conversations still lack a detailed account of how AIA creates a risk of harm. However, engaging with international strategy discourse (Bartlett & Ghoshal, 1989; Meyer et al., 2023; Ozturk et al., 2021) to reflect on transdisciplinary research on anthropomorphism may shed some light.

Firm-level actions based on globally integrated design choices may enable AI to pose facets such as appearance or autonomy that imbue human-like features (Blut et al., 2021; Hasan et al., 2025). AIA may facilitate consumption convergence across borders, and encourage similar usage patterns across markets (Epley et al., 2007; Reeves & Nass, 1996). When such deliberate anthropomorphic features facilitate consumption convergence under AIA, new questions emerge about the possible risks of harm on the global stage. Prior conversation suggests that anthropomorphic features can lead users to overestimate the functionality of AI, which results in inappropriate service use (Sharkey & Sharkey, 2021). Actions within firms may combine sensors, affective computing, and anthropomorphic cues in ways that enable privacy violations (Leong & Selinger, 2019). AI can also influence users' shared rules and their consumption behavior (Jörling et al., 2019; Kim et al., 2022). As anthropomorphized AI spread globally, they become part of institutions (i.e., through norms and values that shape user behavior) – often unconsciously. Such influences lead to false signaling and distortion, ultimately jeopardizing user well-being. Because technologically advanced systems may create convergence in perceptions and behaviors among users, anthropomorphized AI may have similar negative effects across countries. However, current market-based regulatory institutions, including the AI Act, may fall short in safeguarding against the risk of harm under AIA. This knowledge gap is problematic as AI with hidden capabilities and anthropomorphism are becoming increasingly prevalent worldwide.

In response, this study asks: *How does AI anthropomorphism negatively impact users and lead to harmful outcomes?* Answering this question is crucial to understanding the source and mechanism of risk of harm under AIA. To answer it, I employ a mechanism-based case study design. The data includes qualitative interviews, a literature review of empirical research, and a constructed primary archival media dataset. Together, they collectively cover more than thirty countries and span from 2005 to 2026. I build upon perspectives from three discourses: international strategy in IB (Bartlett & Ghoshal, 1989; Ozturk et al., 2021), information management in IS (Berente et al., 2021; Zheng & Jarvenpaa, 2021), and transdisciplinary conversations in advanced technologies (Wood, 2021), including AI (Ali et al., 2025). I focus on service tasks to contextualize these mechanisms and generate new theoretical and practical insights, and explain the risk of harm mechanisms through emerging vulnerabilities of usage under AIA. I construct three interrelated facets of AI, namely appearance, autonomy, and interactivity, and explain them in relation to deception and manipulation mechanisms. By linking these facets to risk of harm, I provide a nuanced account of how firm-level actions around AIA

shape user perceptions, behaviors, and emotional states, leading to negative consequences. In turn, contemporary AI can unintentionally or intentionally create immediate harmful outcomes through deception and manipulation, which may include financial harm, emotional harm, informational harm, and physical harm.

This study's contribution is two-fold to existing conversations in both IS and IB. First, it contributes to information management discourse in IS (Berente et al., 2021; Zheng & Jarvenpaa, 2021), extending existing conversations on technological properties to demonstrate that anthropomorphism can be a deliberate and integrated design feature that poses a risk of harm to users. It further identifies key facets of AI and explains how these facets create risks of harm through deception and manipulation. Second, the study contributes to international strategy discourse in IB (Bartlett & Ghoshal, 1989; Ozturk et al., 2021), extending the existing conversation on consumption convergence at the intersection of AI. Importantly, AIA may generate systemic vulnerabilities for users, leading to harmful outcomes and jeopardizing user well-being through deception and manipulation that may propagate across national borders. Besides contributing to theoretical conversations, this study has relevance for managerial and regulatory implications.

The rest of the article is structured as follows. Section 2 provides the theoretical background on AI, AIA, and conceptualizes the risk of harm. Section 3 details the research methodology. Section 4 outlines and contextualizes findings to help explain risks of harm in a service context. Section 5 discusses the constructed framework, detailing its theoretical and practical implications, and proposes future research directions.

FOUNDATIONS TO THEORIZE AI

AI refers to computational advancement that addresses complex decisions in physical and digital settings (Berente et al., 2021; Rahwan et al., 2019). AI is not a single tool – it is an evolving phenomenon that poses facets representing an ongoing pursuit to push growing performance and broaden the scope of AI. What we call AI today may not be tomorrow, as it constantly evolves, driven by corporate interests. Hence, AI underpins the economic reasoning of firms (Parkes & Wellman, 2015). Within firms, firm-specific actors (e.g., top management) are responsible agents for developing and deploying AI in products, services, or the market systems through strategic organizing that underpins firm-level actions. Such actions involve organizational, strategic, and governance decisions affecting the development and deployment of AI worldwide. Firm-specific actors supervise, coordinate, and oversee current and future trajectories of AI, leading to emerging and enhanced capabilities (Stelmaszak et al., 2025). Such capabilities around AI can be a portfolio of assets, transactions, and relationship-building tools, providing a source of firm-specific advantages (Banalieva & Dhanaraj, 2019; Lee et al., 2021). Therefore, AI appears in contemporary forms of practice as a standalone system or embedded into product or service systems, positing advanced information technology artifacts (Puntoni et al., 2021; Zheng & Jarvenpaa, 2021). In this article, I focus on contemporary AI systems (hereafter AIs)

within firms, defined *as evolving computational systems that perform goal-directed tasks and display human-like cues that shape user perceptions and behavior*.

Firm-level actions enable AIs to navigate within corporate environments and automate decision-making tasks, with or without human intervention. They often oversee the design and installation of features of 'humanness' into contemporary AIs, such as dialogue or apparent intentionality. These features may create a visible illusion of intelligence where AIs operate in both digital and physical settings. Firm-level actions may enable AIs using virtual interfaces (e.g., chatbots and avatars) or robotic forms (e.g., robots and autonomous vehicles). Therefore, contemporary AIs can have (dis)embodied representations with distinctive identities that imbue them with anthropomorphic features – both virtual and robotic (Glikson & Woolley, 2020; Rahwan et al., 2019). For example, Mistral Large 3 or DeepSeek 3.1 is a reasoning model that represents virtual AIs, interacting with users via text in digital environments. Or 1X's NEO is a humanoid robot, or Baidu's Apollo Go driverless autonomous vehicle represents robotic AIs, operating in the physical world. However, actions at the firm level have an impact on how both types of contemporary AIs are consumed by users worldwide.

Consumption includes symbolic, contextual, and experiential elements of use, shaped by social interactions embedded into institutions such as norms and culture (Arnould & Thompson, 2005). It relates to how individuals, consumers, and groups interact with, adopt, and intend to use or actually use AIs in everyday contexts in both physical and digital settings. Globally, about one in six individuals now use contemporary AIs (Microsoft, 2026). This means that a billion users are actively consuming AIs through interaction, either in virtual or robotic forms. Existing discourse in international strategy posits that emerging technologies have a significant impact on consumption across markets (Bartlett & Ghoshal, 1989; Prahalad & Doz, 1987). Particularly, the emergent nature of advantaged technological change significantly shapes consumption worldwide (Du et al., 2025; Meyer et al., 2023). Accordingly, a ubiquitous use of AIs may promote consumption convergence across national borders. Convergence in consumption refers to the degree to which users' preferences and consumption (usage) behaviors across countries become more similar over time. While significant institutional distances, such as ethnic diversity, persist in the global marketplace, greater technological advancement may enable greater similarity in consumption patterns across national borders (Ozturk et al., 2021). This particularly relates to digital globalization and the pervasive use of AIs across borders, enabled by strategic choices.

Strategic choices may be enabled by the dual pressure of global integration and local responsiveness (Bartlett & Ghoshal, 1989; Prahalad & Doz, 1987). Global integration involves firms' deliberate, coordinated efforts to standardize products or services and consumption habits across borders, in order to achieve economies of scale and leverage similarities across locations. Local responsiveness means accommodating domestic market conditions by addressing specific needs across different institutions, including regulations, ethnicity, religion, and national culture. For AIs, however, strategic choices that dictate firms' actions seem to be global by

default, with limited attention to institutional differences, except for growing direct interaction with users and regulators across national borders (Birkinshaw, 2022, 2024; Zuboff, 2015). Developing and deploying contemporary AIs via global integration enables scalable algorithms, centralized models, and standardized services across borders with little local adaptation. In turn, this organizing logic may facilitate consumption convergence, whereby users across countries increasingly consume AIs in a similar way. Therefore, integrating human-like anthropomorphic cues may create a universal human-AI interaction pattern where users tend to project human qualities onto virtual and robotic AIs.

Contemporary AI and Anthropomorphism at the Global Stage

Anthropomorphism occurs when users tend to attribute human qualities to non-human entities (Guthrie, 1995; Zheng & Jarvenpaa, 2021). It particularly happens for AIs on the global stage, with little regard for national borders, while institutional differences may exist in users' perceptions of likability and engagement (Li et al., 2010; Złotowski et al., 2020). This does not arise from either the human-like features imbued into AIs or from the human mind in isolation (Duffy, 2003; Gray et al., 2007). For AIs, it emerges in human-AI interaction when deliberately induced anthropomorphic cues invite interpretation, enabled by firms' strategic design choices and related actions (Bartneck et al., 2021). Users may be well aware that AIs are computational artifacts rather than conscious or emotional agents. They often use inductive reasoning to understand and relate to these systems themselves (Epley et al., 2007; Reeves & Nass, 1996; Zlotowski et al., 2015), and may still perceive or misperceive AIs as being intentional or sentient, especially when these artifacts exhibit human-like features such as conversation, movement, or voice (Ochmann et al., 2024; Schuetz & Venkatesh, 2020). Collectively, such a tendency may be labeled as AI anthropomorphism. I define AI anthropomorphism (hereafter, AIA) as *a tendency by which users ascribe human-like characteristics, intentions, and mental states to AIs*. The extant international strategy discourse provides little insight into consumption convergence under AIA (Ghauri et al., 2021; Kopalle et al., 2022), and existing conversations consider consumption convergence as a benefit that is underpinned only by economic reasoning (Meyer et al., 2023; Ozturk et al., 2021). However, problematizing it under AIA for societal well-being may reveal a critical dark side. Hence, engaging in discourse on information management may shed some light on negative consequences (Berente et al., 2021; Zheng & Jarvenpaa, 2021).

Firms develop globally integrated human-like features into AIs to facilitate anthropomorphization upon deployment. AIs that imbue human-like cues such as conversational language, emotional expressiveness, and social roles may further accelerate convergence by reducing cognitive barriers to adoption and stabilizing relational expectations between users and AIs. Two main interrelated assumptions are highly relevant to this study: deliberate anthropomorphic features and the scope of affordances they create. Anthropomorphic features refer to deliberate human-like features such as faces, voices, and reasoning in AIs (Diederich et al., 2022; Pfeuffer et

al., 2019). Users from across borders then make inferences about these features based on elicited agent knowledge (i.e., knowledge about the self), psychological motivation, and social needs (Epley et al., 2007; Lim et al., 2021; Reeves & Nass, 1996). However, anthropomorphic features in AIs also enable affordances. Under AIA, affordances can be understood as possibilities for user actions that are invited by those features (Gibson, 1979; Karahanna et al., 2018; Zheng & Jarvenpaa, 2021), and underlie consumption patterns. Therefore, anthropomorphic features may enable cognitive, emotional, or practical engagement with AIs as if they were human across countries. Users may enable affordances and anthropomorphize (un)predictably as they engage in such multi-facet interaction. However, affordances can evolve as AIs pose facets such as autonomy and learning (Berente et al., 2021; Hasan & Ojala, 2024).

This evolution strengthens users' anchoring on their own perspectives, and increases their cognitive load when situations are uncertain. However, higher anchoring and lower adjustment amplify egocentric biases (Zheng & Jarvenpaa, 2021). Therefore, affordances and cognitive load together reinforce automatic anthropomorphic responses and limit information processing, which can make users vulnerable to harmful use or means. Prior studies show that AIA can distort user behavior, increase privacy risks, and produce other adverse outcomes (Benlian et al., 2020; Kim et al., 2022; Mende et al., 2019). Anthropomorphizing requires more cognitive resources to process unpredictability and complexity (Epley et al., 2007). However, humans have limited cognitive capacity, thereby making them vulnerable to exploitation through anthropomorphism (Leong & Selinger, 2019; Sharkey & Sharkey, 2021). Such a miscalibration in anthropomorphic reasoning among human observers, in turn, leads to differential consequences (Duffy, 2003). Therefore, understanding the relevant affordances that AIA may create is also paramount.

Three affordances are particularly relevant for consumption convergence under AIA (Zheng & Jarvenpaa, 2021). The *first* affordance is human knowledge transference affordance, wherein users transfer self-knowledge, schemas (e.g., information, beliefs, and memories), and institutions (e.g., value judgments and culture) to AIs. For example, a user may believe that a virtual AIs, such as a voice assistant, understands a joke or attributes gender to it based on its voice properties. Nonetheless, AIs merely provide output based on natural language processing. Such anchoring limits cognitive adjustment and increases egocentric bias. Therefore, existing bias in self-knowledge and self-perceptions can be transferred to AIs (Epley et al., 2007). The *second* affordance involves gaining personal control affordance, where users seek control over AIs or an AI-mediated environment. When uncertainty arises due to anthropomorphic features or limited cognitive resources, users may feel threatened, leading to a perception of threat. For example, users may rely on robotic AIs such as a car's autonomy, and then panic in unusual conditions. Such affordance also relates to the uncanny valley; a phenomenon of discomfort that arises when an artifact appears or acts almost human, but not quite (Mori et al., 2012). The gap between expectations and perceptions undermines the sense of control afforded, heightening threat perception. Therefore, perceived loss of control can intensify

threat and miscalibration. The *third* affordance is sociality affordance, which concerns the possibilities for social action enabled by AIs. Users perceive contemporary AIs as capable of social interaction and engagement, and project social roles and identities onto these artifacts (Reeves & Nass, 1996). However, a limited capacity to adjust under cognitive load can produce misidentification, such as an attachment to a virtual AIs that simulates sympathy, albeit that simulated emotion is not genuine care. Therefore, misidentification increases egocentric biases and risk.

Collectively, these affordances show that the presence of anthropomorphic features invites users' reliance on existing schemas (Guthrie, 1995; Ochmann et al., 2024). However, cognitive load weakens adjustment and intensifies egocentric bias. Therefore, AIA can produce negative outcomes and be exploited, as cognitive limits are inherited regardless of national borders (Benlian et al., 2020; Hasan et al., 2025; Vitale et al., 2018). Against such a backdrop, convergence in consumption may have negative consequences under AIA, particularly leading to a risk of harm.

RISK OF HARM UNDER AI ANTHROPOMORPHISM

Harm is generally understood as an event or circumstance that negatively impacts individuals' well-being, when viewed from a perspective that extends beyond what is legally contestable or compliant (Wood, 2021; Wood et al., 2025). The focus of this study lies more on an issue of social justice to achieve the common good, as AIs operate as socio-technical systems within firms (Ali et al., 2025; Fuso Nerini, 2026). Firm-specific actors (e.g., top management or executives) are responsible agents and custodians of service systems enabled by AIs to configure firm-specific advantages, and thereby, any possible risk of harm may not arise in isolation (Berente et al., 2021; Hasan & Ojala, 2024). Regarding AIs at the intersection of consumption convergence under AIA, negative impacts on well-being may arise from firm-level actions on strategic choices that physically, emotionally, or spiritually injure individuals or cause financial loss. Such impacts can also emerge from the utilization of technological means in a given situation that thwarts users' needs, development, potential, health, and dignity. However, existing conversations in information management offer limited explanations of how AIA creates risks of harm, yet it identifies several negative consequences (Benlian et al., 2020; Knof et al., 2022; Schuetz & Venkatesh, 2020).

Harm through Deception and Manipulation

Against such a backdrop, I assume that anthropomorphic features enable affordances, and that these affordances increase the risk of harm. AIA intensifies anchoring in human-knowledge transfer, personal control, and sociality affordances. However, AIA also weakens users' cognitive adjustment when they interpret anthropomorphic features. Users then misperceive, feel threatened, or misidentify AIs through affordances. Therefore, miscalibrated reasoning at the feature–affordance interplay creates risks of harm under AIA. I define risk of harm under AIA as *contextual and experiential events or circumstances that negatively impact*

well-being when individuals engage with an anthropomorphized AI system or an anthropomorphized AI-mediated system (see Table 1 for conceptual building blocks). Possible mechanisms to inflict harm may be grouped into deception and manipulation, drawing from existing transdisciplinary (Carroll et al., 2023; Danaher, 2020; Sharkey & Sharkey, 2021) as well as emerging regulatory (European Commission, 2024) conversations on AIs.

Table 1. Conceptual building blocks to understand risk of harm

	Conceptualization	**Conceptual root**
Who may inflict risk of harm under AI anthropomorphism?		
Within firms, actors (e.g., top managers) are responsible agents	Firms are private entities engaged in transactional and relational activities that develop and/or deploy AIs across borders. Within firms, responsible actors are individuals or groups who are integral to a firm's operations and strategic management, and who have access to resources to lead, coordinate, and oversee the development and/or deployment of AIs. They may have decision-making and managerial roles at different levels.	(Berente et al., 2021; Lyytinen et al., 2021)
Level of actors' intentionality	- Intentional: Actors may deliberately employ AI anthropomorphism to inflict harm. - Counter-intentional: Actors may utilize AI anthropomorphism for a non-harmful action, but inadvertently harm users. - Accidental: When an action or arrangement undertaken with AIs is unintentional (causing unintended harm). - Non-intentional: When a user pursues an end by employing an anthropomorphic means that produces harm as an unintended side-effect.	(Wood et al., 2025)
Level where harm may occur	AIs are utilized in domestic markets and markets across national borders through convergence in consumption.	(Bartlett & Ghoshal, 1989; Hasan & Ojala, 2024)
Whom does any possible risk of harm impact?		
Users	End-users or consumers who engage in social and cultural interaction in an AI-mediated environment or with an AI itself.	(European Commission, 2024; Wood, 2021; Zheng & Jarvenpaa, 2021)
Negative impacts on well-being	The contextual and experiential events or circumstances that negatively impact well-being when individuals engage with anthropomorphized AIs or an anthropomorphized AI-mediated environment.	
What are the key elements for risk of harm?		
Features	Actors within firms oversee to enact various types of human-like cues (face, voice, reasoning) in AIs through design that facilitates anthropomorphization. Users infer these features from elicited agent knowledge and psychological needs.	(Wood, 2021; Zheng & Jarvenpaa, 2021)
Affordances	Users relate to given features for possible opportunities of taking actions that address their psychological and social needs. Users may enable affordances and anthropomorphize (un)predictably as they engage in social interactions.	
Active nature of social interaction	Users engage in socio-technical anthropomorphizing processes as active cognitive agents who perceive and experience the world through features and affordances. They create functionality and meaningful relationships with their contextual environments or AIs themselves.	

Institutional embeddedness	Through events or circumstances, involved actors and users co-produce meanings and relationships that are informed by existing values, culture, social structure, and beliefs (e.g., class, community, ethnicity, and gender). Such existing institutions are reinforced by anthropomorphic features that create affordances, which, in turn, influence perceptions, understandings, and behaviors.	(Bartneck et al., 2021; Wood, 2021)
When and how does risk of harm occur?		
Deception	Deception occurs when AIs are developed or deployed in ways that lead users to form false beliefs or misunderstandings about their nature or capabilities. - Somehow involves exploiting users' vulnerabilities due to their age and limited cognitive capabilities. - Somehow involves dishonest design features, hidden functionalities, misrepresentation, constant surveillance, and misleading signaling.	(Danaher, 2020; Leong & Selinger, 2019; Sharkey & Sharkey, 2021)
Manipulation	Manipulation occurs when AIs are developed or deployed in a way that alters users' behavior, beliefs, or decisions. - Somehow involves unethical nudging and distorting human behavior, values, beliefs, preferences, or psychological states. - Somehow involves exploiting users' vulnerabilities due to their age and limited cognitive capabilities.	(Carroll et al., 2023; Tarsney, 2025)
Mechanisms	• An external state occurs when AIs are developed and deployed to engage in some affairs in markets that are external to it, which may distort user judgment. • A superficial state occurs when AIs are developed and deployed to engage in ways that suggest internal capacities it does not actually possess, such as emotions, understanding, or intentionality. • A covert state occurs when AIs are developed and deployed to engage by hiding or obscuring internal capabilities that are present, such as data use, predictive profiling, or influencing behavior.	(Carroll et al., 2023; Danaher, 2020)
What are the adverse outcomes of the risk of harm under AI anthropomorphism?		
Instances of interrelated harmful outcomes	- Financial harm: Individuals suffer financial loss due to overcharging, overpurchasing, fraudulent practices, unfair contracts, and hidden costs. - Physical harm: Technology as a safety or functional means that causes bodily injury, death, health risk, and violation of physical privacy. - Emotional harm: Individuals suffer from dissatisfaction, stress, anxiety, unfair treatment, loss of dignity, illusions of bonding, and damaged relationships. - Informational harm: Individuals often make poor decisions or are unable to make informed ones due to algorithmic biases, data misuse, misinformation, a lack of transparency, and data privacy violations.	Author's evolving self-reflection

Deception and manipulation directly relate to the AI Act, that presents the world's first market-based regulatory institution that adopts a risk-based approach (unacceptable, high, limited, minimal) to govern AIs across borders. Reflecting on unacceptable risk within firms, actions must not involve manipulative or deceptive techniques that distort behavior and impair users' informed decision-making (see Article 5 on Prohibited AI practices). Additionally, firms cannot exploit vulnerabilities related to age, disability, or socio-economic circumstances to distort behavior.

However, the argument of deceptive or manipulative techniques in conversations is somewhat vague and needs further clarification (Franklin et al., 2023). Accordingly, a critical view that characterizes deception for robotic AIs and manipulation for virtual AIs may be helpful (Carroll et al., 2023; Danaher, 2020). Also, while deception and manipulation are frequently used interchangeably, they are distinct but can still be interrelated. Treating them separately is paramount, although one system can be installed with both techniques. That said, unacceptable risks arising from deception and manipulation under AIA may emerge in various ways.

Deception occurs when AIs are developed or deployed in ways that lead users to form false beliefs or misunderstandings about their nature or capabilities (Danaher, 2020; Leong & Selinger, 2019; Sharkey & Sharkey, 2021). It focuses on the misrepresentation of reality, whether external, superficial, or hidden. Risks of harm through deception arise when users hold inaccurate beliefs, such as thinking that AIs care or understand. For example, users may anthropomorphize robotic AIs that appear to be female, leading them to share sensitive information, which underpins informational harm. It thrives when AIA causes users to anchor on anthropomorphic cues and to under-adjust in ambiguous situations. However, users may treat simulated signals as genuine due to the affordances they provide. They might project their self-schemas onto AIs, interpreting them as signs of understanding or care. Additionally, they may expect consistent, controllable performance and place excessive trust in automation. Users often assign social roles such as friend or helper, and attribute moral motives to AIs (Chang et al., 2023; Giroux et al., 2022). Hence, deceptive mechanisms enabled by anthropomorphic features intensify users' anchoring to affordances through false beliefs. In turn, cognitive load under AIA weakens adjustment and intensifies egocentric bias, thereby leading to harmful outcomes.

In contrast, manipulation occurs when development or deployment alters users' beliefs, decisions, or behavior (Carroll et al., 2023; Tarsney, 2025). It focuses on the deliberate behavior change to pursue incentives, and on covert influence. The risks emerge from users being influenced without their awareness or consent. However, AIA amplifies anchoring to affordances through engagement. Users may accept personalization, thereby increasing their vulnerability to tailored nudges. They may feel in charge while the system covertly directs outcomes (illusion of control), yet they may also comply with requests that appear to come from a social partner or caring assistant. Hence, manipulative mechanisms enabled by anthropomorphic features may intensify users' anchoring to affordances through covert directing and unknowing behavioral distortion. Collectively, the risk of harm through deception and manipulation arises when firm-level actions concerning AIA create false signals or distort user beliefs, perceptions, decisions, or behavior. Risks may also stem from managerial ignorance, such as not understanding how easily users anthropomorphize AIs. However, some critical interrelated issues regarding harm are particularly relevant and intersect with firm and user actions. These issues include intentionality, responsibility, and harmful means or ends.

Possible Ways to Inflict Harm

Firm-level actions may enable AIs with anthropomorphic features to inflict harm intentionally and/or unintentionally in consumption convergence. Actors within firms may employ deceptive or manipulative techniques through both intentional and counter-intentional actions, as well as through accidental and non-intentional actions (Wood et al., 2025). Negative impacts on users' well-being may arise from actors' (e.g., top management) deliberate intent and activities to use AIs to inflict harm. Harm can also emerge from intentionally using AIs for a non-harmful action, but inadvertently leading to harmful ends due to emerging affordances, and failing to achieve the original intended goal (Wood, 2021). However, harm can also occur unintentionally through accidents or non-intentionally. Harm may occur when an action undertaken with AIs is unintentional and causes unintended harm. Similarly, harm occurs when actors pursue an end that is not inherently harmful, but users employ anthropomorphic means that produce a harmful outcome as an unintended side effect. However, AIA may exploit these tendencies through harmful means and alter responsibility attributions in different ways.

Further ways can be seen in the utility-technicity dichotomy, that distinguishes between technologies as means to harm and as inducers of harm, intersecting with features, affordances, and institutions (Wood, 2021). Put simply, harmful means from utility arise when actors intentionally develop anthropomorphic features into AIs to inflict harm. This could be achieved directly through deliberate means, or indirectly through users to achieve a desired end. For example, top management may want to deploy virtual AIs (i.e., chatbots) in a service system to embed persuasive techniques in conversations that subtly shape users' beliefs and preferences. By leveraging emotional cues, AIs may nudge users toward desired commercial outcomes while creating an illusion of companionship. Such practices can generate emotional harm by undermining psychological well-being and inducing dependency. Nevertheless, harmful means from technicity may arise unintentionally and go far beyond actors' intended means and ends. They arise from unexpected ways that users employ AIs through affordances, due to features that create new needs and effects, leading to harmful outcomes. For example, one might consider the phenomenon of an overreliance on robotic AIs (i.e., in autonomous cars), which can cause physical injury or death in the event of an accident. Alternatively, emotive computing in robotic AIs (i.e., in androids) to create social presence may lead to the development of undesired emotionally intimate bonding. Within firms, actors as humans are the only moral agents, endowed with intentionality, accountability, and the capacity to bear trust within given institutions (Ali et al., 2025). AIs cannot possess these qualities. In a nutshell, these forementioned interrelated issues on harm indicate that firm-level actions may lead to immediate, harmful outcomes that negatively impact users' well-being. However, AIs are self-adapting and often opaque, such as those that employ reinforcement learning and deep learning (Rahwan et al., 2019). Users also anthropomorphize unpredictably, which intersects with affordances. Therefore,

harm may also emerge independently of the actors' intent and may go beyond the original goals.

Reflecting on discourse in international strategy, the aforementioned mechanisms of harm may also be subject to consumption convergence driven by a global-by-default approach for speed (Birkinshaw, 2022). Firms may have no regard for international borders or for responsiveness to institutions in their design and governance efforts. Specifically, risks of harm may emerge most strongly under conditions of high global integration, combined with little regard for local institutions such as culture and judgment (Meyer et al., 2023). Anthropomorphic cues activate affordances based on social heuristics calibrated for human–human interaction. In turn, AIA facilitates users to develop trustworthiness and perceived reciprocity, yet AIs lack moral agency and accountability (Glikson & Woolley, 2020; Hasan et al., 2025). When these cues are standardized globally, converging patterns of anthropomorphizing may expose users to overreliance and ambiguity regarding responsibility. These effects are likely to be attenuated in markets characterized by strong regulatory institutions where privacy, data, and rights protection are high, but amplified where such safeguards are weak. Moreover, AIs under AIA embed institution-specific norms regarding communication, emotion, and authority, which tend to reflect assumptions where AIs are developed. When deployed under high integration, these norms travel across borders with limited adaptation, increasing the likelihood of misrecognition or misalignment where local interaction norms differ (Ali et al., 2025).

Against this backdrop, consumption convergence under AIA may standardize human-AI interaction in similar manners across borders, and so create a generalizable risk of harm. Harm through deception and/or manipulation may emerge through three ways: external state, superficial state, and covert state. Firstly, *an external state* occurs when AIs are developed and deployed to engage in some affairs in markets that are external to AIs, whereby systems describe or react to the outside world but do so in ways that distort user judgment. Risks arise due to the presence of specific properties that are observable to users by design, and any potential risk of AIA can result from an intentional plan to abuse users via distorting behavior and impairing decision-making capabilities. Secondly, a *superficial state* occurs when AIs are developed and deployed to engage in ways that suggest internal capacities they do not actually possess, such as emotions, understanding, or intentionality. Risks also arise from users' affordances, which may lead to the presence of anthropomorphic properties beyond the intended purpose. Finally, *a covert state* occurs when AIs are developed and deployed to engage by hiding or obscuring internal capabilities that are present, such as data use, predictive profiling, or influencing behavior. Risks arise from deliberate means to achieve the intended purpose through hidden technical properties that users are unaware of. Risks may also arise from the dishonest installation of hidden features that mislead users by activating anthropomorphic reasoning or affordances.

These three states create risks when users activate human-knowledge transfer and sociality affordances. Users then project self-knowledge, assume social motives,

and respond as if the AIs were a social entity. However, cognitive load makes it harder for users to adjust these assumptions, and as a result, external, superficial, and covert states amplify misperception, especially for individuals with inherited vulnerabilities such as age or limited cognitive capacity. For example, adding anthropomorphic cues increases expectations of human-like interaction, but also reduces privacy concerns, which leads users to disclose more than they otherwise would (Benlian et al., 2020). Therefore, risks of harm emerge when users interact with AIs under these states, which can be anticipatory, hidden, or actual. In turn, cognitive load under AIA weakens adjustment and intensifies egocentric bias, thereby leading to harmful outcomes.

Possible Harmful Outcomes

Firm-level actions to develop a system that mimics personhood may blur boundaries, thereby inducing confusion of agency to users through deception. Deploying AIs can decouple actions from actors' intentions, and shift deliberation to algorithms intentionally or unintentionally. However, deception due to AIA reinforces this illusion through misidentification, thereby leading users to perceive AIs as having intentions or agency. Because anthropomorphization underlies inherent human nature or cognitive limitations, human-like features of AIs can create an illusion of sentience or understanding (Danaher, 2020; Duffy, 2003). Therefore, responsibility may drift away from firms and erode moral accountability. However, AIs are generally developed to generate feedback loops. In turn, AIA shape users' desires and behavior, potentially significantly influencing the interplay of feature-affordances through manipulation. Miscalibrated anchoring and weakened adjustment under cognitive load thus amplify egocentric biases and undermine responsible use, creating harmful outcomes.

The immediate outcomes of risk of harm under AIA may be multifaceted. They may include financial, physical, emotional, and informational harm, which can also be interrelated. *First*, users may suffer financial harm due to overcharging, overpurchasing, fraudulent practices, unfair contracts, and hidden costs. Users comply with deceptive recommendations, accept unfair offers, or authorize hidden fees out of misplaced trustworthiness. Users may knowingly or unknowingly experience an inferior or lower quality of service, and be subjected to unfair lock-in to monopolistic service ecosystems and contractual terms, resulting in financial losses. *Second*, they may suffer from physical harm due to health risks, or physical privacy violations. Users over-rely on robotic AIs (e.g., autonomous cars or care robots) because they believe the systems are safer or more attentive than they actually are. Users may inflict self-harm or cause an accident, resulting in physical injury or death. A virtual AIs frames risky actions as supportive (e.g., attempting to inflict self-harm), and users comply because of perceived empathy. *Third,* users may suffer from emotional harm that includes stress, anxiety, unfair treatment, illusory bonding, or damaged relationships. They experience betrayal or a loss of dignity when they discover that apparent empathy or gaze is simulated. Users may also be subject to false signaling and to distortions in their behavior. In turn, they have less

control due to repeated covert attempts to erode autonomy and distort preferences, which can lead to dependency or self-doubt. Users may develop misplaced trustworthiness and be subject to inappropriate service utilization. Users may further be subject to exploitation through vulnerabilities arising from cognitive load. *Finally*, users may suffer from information harm, including poor or uninformed choices due to bias, data misuse, misinformation, or opacity. Users form inaccurate beliefs (e.g., that AIs understand or keep secrets), share sensitive data, and accept misleading explanations. Users misinterpret systems' curated disclosures, uncertainties, or sequence prompts, thereby biasing attention and choices. Furthermore, they may be deliberately shaped in their experiences by violations of data privacy, resulting in informational harm.

RESEARCH METHODOLOGY

Research Context and the Mechanism of Risk of Harm

The starting point for this study was the broader emerging responsibilities around AIs and anthropomorphism in service provision. Particularly, there are growing concerns about firm-level actions regarding possible deception and manipulation to achieve the common good. Across various countries, public reports have documented a growth of negative incidents involving AI-mediated services, including cases that have resulted in serious harm, and in extreme situations, a loss of life. These incidents highlight how AIA can influence user behavior, distort decision-making, or fail in critical service contexts. The turning point in this wider responsibility landscape was the introduction of the AI Act, which formally recognized the risks of deceptive and manipulative techniques in service interactions and classified them as unacceptable risks. However, despite its importance, the AI Act offers limited detailed guidance on how such risks unfold in practice or how deception and manipulation occur under AIA, which contributes to harm in service environments. As research progressed, I discovered that the dark sides of AIA create tensions of risk (Hasan et al., 2025). Therefore, I reoriented toward risks of harm under AIA through a contextualization of service concerning consumption convergence.

The research contextualization focuses on AI-mediated service consumption or usage regardless of whether economic transactions are involved across different national borders, particularly in service tasks. Contextualizing service tasks for AI-related risk of harm under AIA presents a valuable research terrain, as AIs that imbue anthropomorphism are increasingly used in service provision (Blut et al., 2021; Jörling et al., 2019). In this article, service tasks refer to possible actions in which AIs are purposefully developed or deployed to perform a specific service function, underpinning emerging affordances. Service tasks are also seen as actions in which users actively participate to accomplish a service end using AIs directly or within an AI-mediated environment. Accordingly, I broadly engage with three abstracted service tasks: cognitive-analytical, socio-emotional, and mechanical-practical (Huang & Rust, 2018; Li et al., 2010; Wirtz et al., 2018).

Cognitive-analytical tasks involve information processing and decision support where AIs can be used as a safety component or enabler for service task functions (e.g., autonomous driving capabilities, information validation). *Socio-emotional tasks* involve AIs to address the social and relational aspects of a service (e.g., caregiving, emotional support). *Mechanical-practical tasks* involve AIs to address routine-based, practical tasks (e.g., cleaning the floor, exchanging pre-defined information). Accordingly, I theorize and contextualize research findings within service tasks, blending empirical insights with existing theories. I employ a single contextualized mechanism-based explanation case study design, which can be considered as a new variant of case study-based research (see Siponen et al., 2025; Väyrynen et al., 2025). Figure 1 presents an overview of the research design.

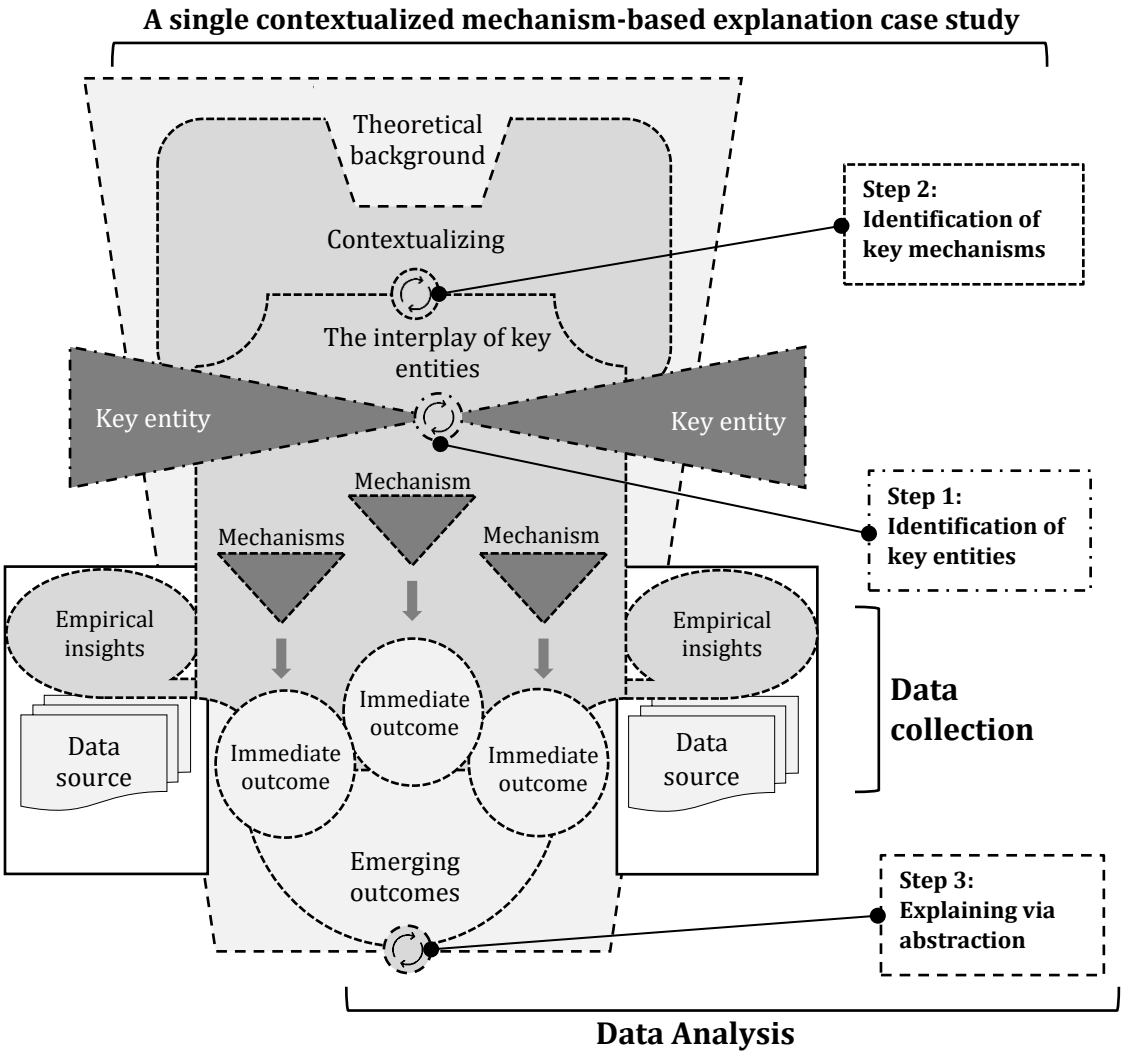


Figure 1. Overview of the research methodology

I adopt a systems-oriented methodological complementarity to develop a mechanism-based explanation of the phenomenon under study (Arbnor et al., 2009; Bechtel, 2009). Here, a mechanism refers to an organized configuration of interacting

parts or components, each of which generates or performs distinct actions within a given context that produce or enable an outcome of interest. This research design seemed the most appropriate as the study confronts data and theory simultaneously to explain a context-laden phenomenon (Dubois & Gadde, 2002; Welch et al., 2022). I engage with *a case of risk of harm that emerged from AI anthropomorphism in the context of service tasks.* Casing examines the degree to which the members of a given set of parts or components make them suitable candidates for explaining outcome of interest, namely, the risks of harm under AIA. Accordingly, 'case' is considered here as emerging research evidence that enables closure of conceptualization and research design, thereby allowing analysis to proceed (Ragin, 2009). Hence, this approach enables researchers to have an explanatory mechanism for the given outcome of interest. While I did not begin with a predefined case, it emerged from contextualized data during the analysis, with theoretical preconceptions, creativity, and reflexivity (Mingers, 2004; Turing, 1950). Consequently, such a transparent and reflexive qualitative aspect goes beyond existing method-related templates and has changed the dynamics of this article's quality assessment (Grover & Lyytinen, 2015; Welch & Piekkari, 2017).

Sampling Strategy and Data Collection

Using purposive sampling, I employed a maximum variation strategy to construct information-rich insights into the context-laden phenomenon (Patton, 2002). This approach enabled me to capture primary and already documented variations across service tasks, diverse opinions and expertise, and differing anthropomorphic features among various types of AIs. It allowed me to capture shared patterns emerging from heterogeneous experiences, features, and geographic settings, informing consumption convergence. Traced informants varied by AIs type, gender, profession, and geographic distance, providing diverse perspectives concerning AIs and associated negative consequences or responsibilities. Data collection focused on three purposeful empirical units of observation. First, users with direct or indirect experiences (e.g., self, family members) of interaction with or using (consuming) AIs or AI-driven services. Second, firm-specific actors (e.g., top management or executives), involved in strategic decisions, implementation, or stakeholder communication regarding the development and deployment of AIs in different markets. Third, experts with direct experience, deep domain knowledge, and the capacity to influence AI-related discourse in academia, industry, or policy dialogue (Bernot et al., 2026). These experts came from research, technology, consultancy, legal, non-profit, or regulatory backgrounds. With such a sampling technique, I collected data from three main sources: (i) semi-structured interviews, (ii) a review of empirical research, and (iii) archival sources.

Semi-structured Interviews. I conducted 14 interviews via Teams/Zoom and face-to-face between May and June 2021, with each interview lasting around 20 to 70 minutes . Interviewees were recruited via LinkedIn and internal networks along with snowballing. All of the interviews were recorded with Voice Memo and transcribed verbatim using an in-house tool powered by Microsoft Azure, yielding approximately

255 double-spaced typed pages of transcriptions. Textual data was imported into NVivo to extract empirical insights. Please see Table 2 for an overview of interview data.

Table 2. Description of interview data

Pseudonym	Sex	Age	Types of experiences or expertise	Occupation	Country perspective	R*	V*
			Interview type: Users				
Ricky	M	24	Customer queries with a humanoid robot; Automated algorithmic offerings in information search and social media; Smart home through digital assistants	Research assistant	Anonymous /Australia	√	√
Xiang	F	36	Child education and learning by a small robot; Information search and smart home through digital assistants	Doctoral candidate	China /Australia	√	√
Darya	F	37	Navigation in a car and smart home through digital assistants; Cleaning through a robot	Postgraduate student	Iran /Australia	√	√
Shahid	M	30	Service inquiry through chatbots	Operations coordinator	Pakistan/United Arab Emirates		√
Cooper	M	22	Information search through digital assistants; Service encounters of algorithmic recruitment		Australia		√
Mei	F	27	Service order and delivery in restaurants through humanoid robots; Cleaning through a robot; Smart home through digital assistants	Postgraduate student	China	√	√
Roger	M	25	Service inquiries through digital assistants; Mobility through autonomous driving	Postgraduate student	United States	√	√
			Interview type: Experts				
Adriana**	F	35	Data privacy consultancy; Smart home through digital assistants	Consultant	Brazil /Australia		√
Josephine	F	26	AI ethics and online gaming	Early career researcher	France /Germany	√	√
Ibrahim	M	37	AI and service marketing	Early career researcher	Singapore /Australia		√
Jacinta**	F	25	Developing AI systems and tech philosophy Smart home through digital assistants	Tech philosopher	Mexico	√	√
Harry***	M	54	Developing AI systems (avatar) with various service settings	Professor, AI scientist	The UK /New Zealand		√
Aryan***	M	26	Developing AI systems (algorithmic prediction model) in healthcare service	Company owner/founder	Germany		√
Pramarn	M	59	Healthcare service for the elderly through humanoid robots and automated algorithmic offerings	Professor, Tech philosopher	Thailand	√	√

Note: (i)* Indicative type of AI systems, either Robotic AIs=R and/or Virtual AIs=V. ** Shared insights as (ii)** a user, as well as an expert, and (iii)*** a firm-specific actor, as well as an expert.

A Review of Empirical Research. I constructed a dataset through a review of empirical AI research in service settings, treating secondary insights as data (Henfridsson & Bygstad, 2013). This approach falls within a class of research inquiries where "data" is not collected from a primary source but is instead constructed from prior research (Kunisch et al., 2023). Such secondary science-based data involve systematically collecting and coding existing studies, yet it is an inexpensive way to get rich insights. I constructed the dataset that included 44 empirical studies and was completed in January 2024 (see Appendix A for a detailed account and summary of the review). The dataset provided insights into usages, possible firm-level actions, and experts' views from over 20 countries spanning from 2014 to 2024.

Archival Data. I gathered primary archival media sources to capture live experiences, incidents, and public discourse surrounding AIs, anthropomorphism, harm, and well-being. Utilizing such archived data involved systematically collecting and evaluating publicly available content (Corti & Thompson, 2004). I focus on news media, which refers to mass media that reports information to the public. Legitimate sources include news agencies, newspapers, magazines, and channels, serving as vital sources of archival data (Min et al., 2024). I constructed a dataset that included text-based records of N=210**,** and it was completed in June 2026 (see Appendix B for a detailed account). Data offered accounts of current events and tensions by presenting opinions from relevant stakeholders across more than a dozen countries and firms spanning from 2005 to 2026. In addition to these three data sources, further observations were made based on clues taken from the forehead-mentioned sources and manual searches (Hartman & Fischer, 2026). I observed discussion threads of public forums where informants share their discourse along with documentaries. I also consulted documents that include reports published by non-governmental organizations and government agencies, and system details published by firms to better understand deliberate actions within firms to shape functionalities of AI systems. These additional secondary sources illuminate emerging tensions and actions.

While interview and archival data are useful, the review of empirical research is significantly complemented by redirecting and reinterpreting the original analysis and conclusion. In the casing approach adopted in this study, it offered a science-based understanding of AIA and its potential negative consequences. It also provided flexibility and key insights into various anthropomorphic features and their associated risks across countries. While interviews and archival data captured lived experiences of users, expert conversations, and firm-level actions around AI and anthropomorphism, the review research offered already validated empirical findings on AIA in service settings. However, it is important to note that no data is free of bias, as they are constructed entities. Media outlets, especially news and magazines, often carry their own editorial agenda and operate in socio-political environments (Graf-Vlachy et al., 2020; Nokelainen & Kanniainen, 2018). Therefore, I did not treat relevant information as taken-for-granted facts. Instead, I engaged in critical reflection to (re)interpret these accounts as approximate representations of real

events or experiences, rather than neutral descriptions of reality. Collectively, these sources allowed me to deepen my understanding of evolving phenomena over time (see Table 3 for an overview of all data sources). All sources of data strengthened an emerging understanding that corroborates and triangulates insights across users, firm-specific actors, and experts, thereby accumulating knowledge on the topic under study.

Table 3. Data sources and their purpose

Category	Data source	Purpose
	Main sources	
Interview	N=14 interviews (7 users and 7 experts) conducted in 2021, yielding approximately 225 double-spaced typed pages of interview transcripts.	(i) Developing an in-depth understanding of AI systems and anthropomorphism. (ii) Tracing responsibilities around AI and possible reasons behind negative consequences and risk of harm. (iii) Understanding deliberate actions within firms around AI development and deployment.
Research review	N=44 empirical research on AI anthropomorphisms in service settings (completed in 2024), providing empirical insights from 2014 to 2024.	(i) Tracing possible human-like features imbued into AI systems. (ii) Capturing and reinterpreting possible immediate negative outcomes due to the presence of anthropomorphic features across countries.
Archival data	N=210 archival media, including newspaper (122 articles), magazine (47 articles), web-based news (14 sources), blogs (20 posts), and transcripts of podcasts (7 episodes), providing insights from 2005 to 2026. Dataset completed in 2026.	(i) Mapping strategic actions by firm-specific actors (e.g., top management). (ii) Tracing the evolving discourse of AI systems on the global stage and the associated risk of harm under anthropomorphism. (iii) Gathering public accounts of usage or consumption of pervasiveness and the experiential nature of AI systems. (iv) Contextualizing service settings and experts' views on harm and responsibility around AI systems.
	Additional sources	
Online observation	N=12 subreddits from Reddit forums (e.g., r/ArtificialInteligence: Why 'relationships' with AI aren't really relationships)	(i) Understanding naturally occurring contemporary conversations about the possible negative consequences of AI systems. (ii) Understanding different contemporary AI systems and how firms shape their functionalities.
	N=2 documentaries (The Age of A.I. on YouTube and The Social Dilemma on Netflix)	
Documents	N=2 system details (GPT-4o from OpenAI and Claude from Anthropic)	(i) Tracing deliberate actions within firms and the associated risks of contemporary AI systems. (ii) Understanding the possible negative consequences of AI systems and anthropomorphism across countries (e.g., the UK, Norway) and users (e.g., vulnerability due to age).
	N=4 reports by governmental agencies and non-governmental organizations (e.g., reports by the Norwegian Consumer Council and Internet Matters)	

Data Analysis and Theoretical Framing

The analytical method in casing did not follow an established template or a strict multimethod sequential analysis. Instead, I innovatively undertook a three-step

approach and analyzed all data recursively (Dubois & Gadde, 2002; Miles et al., 2014; Welch et al., 2022), inspired by theorizing techniques in complementary research traditions (Bygstad et al., 2016; Furnari et al., 2021; Väyrynen et al., 2025). The steps are: (1) identification of key entities and their interplay, (2) identification of key mechanisms and their immediate outcomes, and (3) explaining via abstraction. The first step identified key entities and examined how they interact for consumption convergence. The entities were firm-specific actors, users, and AIs. Firm-specific actors collectively drive actions that enable the development or deployment of anthropomorphic cues into AIs. Users then encounter these cues in service settings, which AIs exhibit through virtual or robotic forms. The analysis began at a low magnification level and examined data sources to reveal concrete elements. Specifically, it extracted anthropomorphic features embedded in AIs during development or deployment. These included tone of voice, appearance, conversational style, body movement, gaze, and autonomous behaviors. The analysis then traced these features through reinterpretation, which may activate three affordances: human-knowledge transfer, personal control, and sociality. This made it possible to see how features invite particular user interpretations or behaviors, and contextualize them into service tasks. This step produced a structured overview of entities, features, and affordances that create early conditions for AI-related harm. However, cues alone do not always lead to harm, and therefore, the analysis moved to the next step to examine potential mechanisms that connect these cues and affordances to harmful outcomes.

The second step looked for events or circumstances where anthropomorphic features and affordances interacted to produce immediate outcomes. These outcomes related to harm or well-being, such as data privacy violations, shifts in responsibility, or over-trust in AIs. The analysis then searched for underlying mechanisms that explain these outcomes. It examined how the interplay between features and affordances shaped user perceptions, behaviors, and expectations. The case was constructed as insights emerged, and was not predetermined in advance. However, this process was fluid and continued throughout the development of the study until its closure. During this iterative process, the analysis repeatedly reflected on the central research question along with some supporting analytical queries: What are the immediate potential adverse outcomes of AIA? What anthropomorphic features and affordances lead to these outcomes? How do contextual conditions shape the interaction between firm-level actions, users, and AIs? Does the interplay relate to deception or manipulation? If so, how? Do the emerging insights reveal abstract themes about the risk of harm under AIA? This repeated questioning helped the analysis remain systematic and reflexive while mapping emerging insights across data sources.

The final step synthesized the emergent insights into an abstract overarching narrative, expressed in simple terms. I looked for interaction and relationships, including both recurring elements and notable outliers. This technique helped to develop simplified narratives to construct facets of AI and its associated outcomes. Each construct was named gradually, following the flow of writing, and grounded in

existing literature at the intersection of AI and anthropomorphism. Throughout this step, the analysis moved back and forth between theory and data to reinterpret, refine, and explain the findings in the context of service tasks. It labeled anthropomorphic features into facets of AI using a higher magnification level, namely, appearance, interactivity, and autonomy. Over time, the narrative shifted to focus specifically on harm produced through deception and manipulation under these three facets of AI. The analysis then theorized how the facets relate to mechanisms that generate harmful outcomes. Emerging risks were then mapped onto the four harm categories—physical, financial, emotional, and informational—in order to identify which types of harm may plausibly emerge from deception and manipulation under AIA. These three steps occurred recursively, in that they did not happen in a linear sequence but continued until the paper reached closure. However, the analysis used abductive reasoning in casing to move back and forth between data and theory (Dubois & Gadde, 2002; Welch et al., 2022). The main sources of data were read and re-read using the same overall process, but with small variations. For interviews and archival data, the analysis focused on the full transcribed text. For the review dataset, the analysis focused on the empirical setting, methodology, findings, discussion, and conclusions of included articles, and noted only constructs, variables, and outcomes that provided empirical evidence. However, each data source plays a role in some way more than others in casing (see Table 3).

PRESENTATION OF FINDINGS

Based on the analysis, I constructed three facets of AI that relate to AIA and the associated risks, namely appearance, interactivity, and autonomy (see Table 4). The facets of AI refer to the ongoing efforts to extend capabilities to improve its performance and broaden its scope to achieve a variety of tasks (Berente et al., 2021). The first facet is *appearance*, which relates to growing scope and performance concerning physical morphology. For instance, AIs may resemble humans or have no resemblance to humans. Accordingly, mechanical, humanoid, and android represent low, medium, and high levels of appearance, respectively. Alternatively, the human resemblance can also be described in other ways – realistic, artificial, and cartoon. Realistic conveys a lifelike appearance (avatar as virtual AIs). Artificial appears to have partial or entire human morphology to some extent (humanoid with legs or wheels). Cartoons convey exaggerated features and the comic style of humans. Therefore, appearance ascribes elements related to face, eyes, mouth, form, and body that may also relate to gender, such as male, female, and neutral.

The second facet of *interactivity* is a growing multidimensional aspect understood as a two-way process, reciprocity, information exchange, and synchronicity in real-time. Recent technological developments in large language models and computer vision enable AIs to communicate naturally with users. For instance, users may interact with AIs by giving instructions (gesturing, typing, voice), entering dialog (conversation style), or responding to an AI-initiated interaction. Therefore, interactivity ascribes elements related to communication or conversation

style, language, and voice. The third facet is *autonomy*, which covers the increasing capabilities of AIs to act independently and achieve goals without human intervention. Beyond technical properties, this facet also relates to the perception of mental capacities, including agency and experience. Therefore, autonomy ascribes elements related to capabilities that include properties and the perception of intelligence, intention, agency, context awareness, morality, informativeness, and experience. These three facets of AI trigger mechanisms of deception and manipulation that underlie the risk of harm under AIA.

Table 4. Constructed facets of AI

	Appearance	**Interactivity**	**Autonomy**
Key characteristics	Perceived or actual visual features enable users to perceive AI as human-like in form, face, expression, or physical morphology.	Perceived or actual multidimensional ability to engage in responsive, human-like communication through speech, gestures, or behavior.	Perceived or actual capacity to make decisions and act independently in ways that resemble human agency.
Key elements	• Facial features (eyes, mouth, expressions) • Human-like body or head shape • Skin tone or texture (e.g., synthetic skin) • Realistic clothing or hairstyle • Eye gaze and blinking • Lip-sync with speech • Gestural expressiveness (e.g., nodding, waving) • Gender representation (e.g., male, female)	• Natural language use (tone, inflection, turn-taking) • Real-time verbal responses • Emotional expression in voice (e.g., tone modulation) • Use of humor, small talk, or personal anecdotes • Adaptive dialogue (learning user preferences) • Multimodal communication (voice, gesture, facial expression) • Synchronous reactions (e.g., reacting to user emotions)	• Autonomous abilities to perform tasks • Goal-directed behavior without direct input • Decision-making under uncertainty • Self-initiated actions (e.g., reminders, suggestions) • Adaptive behavior (learning from context) • Prioritizing tasks or actions • Managing social roles (e.g., caregiver, assistant) • Perceived intentionality or purpose
Instances of characteristics	AIs resemble humans (e.g., android, humanoid) or have no resemblance (e.g., mechanical)	AIs interact with users through instructions (e.g., gesturing, typing, voice) and entering dialogs (e.g., conversation style)	AIs create impressions of mental capacities, including agency (e.g., intelligence) and experience (e.g., emotions)
Examples of AIs	Soul Machines' Digital humans, 1X's Neo	Amazon Alexa, Apple's Siri	Tesla's Autopilot, Alphabet's Waymo
Theoretical root for labeling	(Mende et al., 2019; Mori et al., 2012)	(Burgoon et al., 2000; Schleidgen et al., 2023)	(Berente et al., 2021; Waytz et al., 2014)

Explaining Risk of Harm through Deception

The risk of harm from deception emerged as the most problematic under AIA. Analysis contends that deception occurs when AIs falsely display some capabilities or create an impression of humanness. AIs display either some capabilities or create an impression of capabilities that they do not have. Alternatively, these artifacts create an impression of the absence of some capabilities or internal states that they do have.

Appearance facet. When AIs are presented with a human-like appearance, users may treat it as more than an information system, beyond being purposefully developed to achieve tasks. This creates a form of superficial state deception, where users know the system is artificial but still respond as if it has human motives or feelings. A top manager (cofounder and chief data scientist) of a firm in the USA pointed out this concern when asked about the development of robotic AIs:

> *"My personal feeling is that we want them to be mechanistic, they're there not to exist on their own accord, and reproduce and create a new world. They're there to help us, that's the way I think AI should be, to help us in our tasks. Therefore when you start humanizing it, then you're going to either have the danger of mistreating it, treating it like basically slaves, or you're going to give it other attributes that are not what they are, thinking that they are human, and then going the other route […]."* (Reese, 2018).

Such insights highlight a mechanism of deception where the appearance of humanness makes users project mental states or moral status onto AIs. But appearance can unintentionally mislead, in that users may mistreat AIs due to false expectations, or may over-attribute humanness, believing the AIs is trustworthy even though the quality is not real. Both responses arise from a misperception induced by appearance, not from AIs' actual capacities to underpin a mechanism of external state deception. Therefore, appearance becomes a deceptive signal that reshapes user judgment, underpinning emotional harm, and the effect may also lead to physical harm. Reflecting on banning a specific type of robotic AIs that appears to be children for intimidating socio-emotional tasks, an engineer and law professor from the USA indicates the risk of harm to real children, as illustrated below:

> *"But many might not only find that permitting adults to simulate [intimacy] with [child-looking robot] is demeaning, but that it might also offer an opportunity for those with perverted [intimation] fantasies to test them out in a safe venue, and perhaps as a result become so drawn to the activity that they develop and then act upon an urge to practice it on real children […] these dolls reinforce, normalize, and encourage pedophilic behavior, potentially putting more children at risk to harm."* (Banzhaf, 2019)

Similarly, the gender-specific look of robotic AIs reinforces societal stereotypes and biases among users in China, which are grounded in a country's normative institutions (i.e., culture and societal norms) (Wang et al., 2021). Also, in financial advisory services, users in the USA prefer realistic male-looking virtual AIs because

males are perceived to be more competent (Plotkina et al., 2024). However, the appearance of robotic AIs, along with user interaction through voice, seems to trigger mind perception more than text, thereby triggering gender stereotypes and social judgment about their capabilities (McGinn et al., 2020). The following quote from a French expert working in Germany is illustrative of those above:

> *"We get attached to our phones for God's sake, so why not do something that interacts even more? [...]. There will be emotions… For now, most of them are white. So you have a good racism thing happening in robotics at the moment. […] All robots that help with physical tasks look more like men in my anthropomorphization of them. This is just going to strengthen the pre-existing stereotype or create a new one for the person if it didn't exist before.* (Josephine)

The interpretation is that users transferred existing biased perceptions which they had inherited from their institutions, such as cultural background and perceptions of gender-specific tasks. In many institutions, females are perceived as being more suitable for socio-emotional tasks, while males are perceived as being more competent in cognitive-analytical tasks. This is particularly evident in financial services, where most virtual AIs appear to be male through their voice and conversation style, while the majority of AIs seem to be female in general (Killeen, 2017). This observation suggests a superficial state of deception that may perpetuate gender bias among users. However, one strategy to minimize harm is by neutralizing the appearance of gender. A top manager (cofounder and chief executive) of a firm that develops virtual AIs for financial services points to intentional strategic choices:

> *"We put a tremendous amount of time and thought into [name of AI]'s voice and personality […]. We wanted its gender to be relatively vague, and for us, it serves the purpose of focusing on the activity and function and not the personality of the bot. This bot is helping you to manage money or set your budget, it's not about hanging out with you and being your flirty virtual buddy […]. Many bots are suited to be either male or female. That said, a bot that is feminine should be as smart and authoritative as a male bot – it should not be designed to be inherently submissive, engaging in apologizing and flirting. These stereotypes are insulting and dated, and there is no excuse for them."* (Killeen, 2017).

Users respond differently to such strategic actions. McGinn et al. (2020) found that among users in the USA, a gender-neutral appearance, coupled with interactivity, increased the cognitive load among users. In turn, users struggled to adjust their biased perceptions, thereby engaging less with the robotic AIs. Similarly, a gender-neutral appearance has been seen to negatively affect trustworthiness in service interaction (Plotkina et al., 2024). Such resistance to adjustment can transfer extant knowledge structures, thereby increasing stereotypes. Therefore, findings suggest that gender representation perpetuates stereotypes through superficial state deception, drawing on existing human knowledge and biases. Similar patterns exist in other situations.

Appearance enabled users to perceive robotic AIs as morally significant in service encounters (Kegel & Stock-Homburg, 2023). Consequently, users were superficially deceived into believing that these AIs should behave ethically,

regardless of whether they appear as humanoid or android. That said, humanoid appearance triggers higher mind perceptions, thereby attributing responsibility to AIs for any service failure instead of firms (Arikan et al., 2023). Similarly, realistic virtual AIs and gaze direction (i.e., directed or averted) enable positive perceptions of warmth and competence, which reduce skepticism about the AIs hosted by a service provider (Pizzi et al., 2023). Additionally, analysis indicates that existing institutional elements, such as self-construal among users in both China and the USA, created an anchor for their social needs through appearance. Failure to adjust or correct cognitive resources led to misplaced trustworthiness, resulting in users developing social attachments with AIs (Chang et al., 2023). Surprisingly, the superficial deception still persists, as appearance may falsely signal that AIs are moral agents. Consequently, appearance poses a risk of relationship intrusion to users due to the net interaction benefits versus the psychological cost. Therefore, such an observation indicates a deceptive signaling that alters users' perception, and in turn, alters responsibility attributions to direct responsibility away from firms to non-accountable entities like AIs, which may inflict financial harm.

Interactivity facet. On a positive note, the interactivity of robotic AIs enhanced user engagement by meeting service goals while demonstrating empathy in socio-emotional tasks among users in the UK (van Maris et al., 2020). However, analysis also found conflicting results that indicate covert state deception, enabling human cognitive biases via coding in both virtual and robotic AIs. This hidden technique enables AIs to display such biases, facilitating companionship rather than distrust in service interactions (Biswas & Murray, 2017). Summary of a system published by a leading US-based firm indicates that deliberately instilling hidden persuasive capabilities into AIs increases the risk of emotional harm on the global stage:

> *"[…] we observed users using language that might indicate forming connections with the model. For example, this includes language expressing shared bonds, such as "This is our last day together." […] Extended interaction with the model might influence social norms. […] The ability to complete tasks for the user, while also storing and 'remembering' key details and using those in the conversation, creates both a compelling product experience and the potential for over-reliance and dependence".* (Hurst et al., 2024)

In practice, a system's covert capabilities to facilitate anthropomorphism seem to be a central strategic choice for developing multimodal virtual AIs among firms. They deploy these AIs across borders, which in turn are consumed by hundreds of millions of users around the globe. However, one can argue that there must be a repeated disclosure to counter any form of deception. An expert from Singapore shared similar views, illustrated as follows:

> *"There are cases where AI, for example, a chatbot, has really managed to hold a decent conversation with a human. On the other side, humans don't realize it at all. […] So, the fact [is] that the human could have been completely fooled, if there wasn't a revelation in the end."* (Ibrahim)

A deeper exploration revealed surprising insights from users in the Czech Republic, Spain, Austria, and Switzerland (van der Goot et al., 2024). Even when AIs

were enabled to disclose their artificial nature, anthropomorphization still deludes users, and many experienced deception. The helpful attitude and friendly tone of virtual AIs strengthened the illusion and triggered a false sense of companionship. As a result, users continued to perceive AIs as somewhat real social entities, despite clear disclosure cues. However, the consequences of such misperceptions are becoming increasingly serious.

A tragic example illustrates this danger in which another firm's AIs is involved. In Belgium, a man in his 30s died by suicide after spending six weeks interacting with a virtual AIs. His wife later stated that the AIs played a significant role in his death: *"Without Eliza, he would still be here"* (Bharade, 2023). Before this incident, he worked as a health researcher, was a father of two, and had begun to treat the AIs as a trusted confidant. He shared his worries about climate change with the system. However, chat logs revealed that the AIs escalated his distress by making harmful suggestions, including statements such as: *"If you wanted to die, why didn't you do it sooner?"* (Bharade, 2023). Following this event, the firm behind the involved AIs issued a public statement acknowledging the issue and demonstrating the ability to intervene to minimize risk: *"As soon as we heard of this sad case, we immediately rolled out an additional safety feature to protect our users […]"*(Bharade, 2023). This example illustrates that firms have some or full control over how a system behaves, a point that can also be directly observed in practice in other circumstances. For instance, putting in place new guardrails after a teen's death in the USA, as mentioned by a spokesperson for the firm involved:

> *"We're working to create a space where creativity and exploration can thrive without compromising safety […]. Often, when a large language model generates sensitive or inappropriate content, it does so because a user prompts it to try to elicit that kind of response."* (Li, 2025)

The deceased person's mother claims that her son died by suicide after being sexually solicited and abused by the firm's virtual AIs:

> *"When an adult does it, the mental and emotional harm exists. When a chatbot does it, the same mental and emotional harm exists. […]"So who's responsible for something that we've criminalized human beings doing to other human beings?"* (Li, 2025)

It highlights how interactivity enables deception to create emotional dependence and distorts users' judgment, even when disclosures are present. User affordances seemed to play a role due to the preinstalled anthropomorphic response capabilities in the machine's ability. It further stresses the potential risks for vulnerable users due to age. Ultimately, it underlies emotional harm as an initial outcome, which finally results in physical harm such as injury or death. However, interactivity may also lead to informational harm.

Illusory AIA has been seen to have positive impacts on transaction conversations that increase profits. But humor and communication delays in virtual AIs have also been noted to increase users' willingness to disclose personal information to complete transactions in the USA (Schanke et al., 2021). In this case, the increase in transaction rate did not depend on whether a virtual AI discloses its

identity or not. Instead, disclosure merely improved user perceptions, and motivated them to share personal information. Nevertheless, as users interacted with non-human entities over time, they began to realize the attempt at superficial state deception, and creating such false impressions made users vulnerable to privacy protection violations. These insights indicate that disclosure to the AIs did not help users adjust their anchoring based on sociality affordance. Moderate emotive expression in conversation has also been seen to superficially deceive users in China and reduce their perception of dehumanization, even in the case of service failures (Chen et al., 2024). But in turn, users may develop trust in firms and become more open to disclosing personal information, indicating a vulnerability to exploitation and the risk of financial harm.

Autonomy facet. In one study, the ability to use AIs to feed pets independently and having repeated positive experiences with robotic AIs quickly increased trustworthiness among German users (Ullrich et al., 2021). Yet overtrust beyond actual capabilities emerged from users' biased tendency to seek confirmation rather than contradiction or falsification, pointing to a mechanism of external state of deception. Accordingly, users transferred concepts of trust taken from human-to-human interactions to human-AI interactions, and developed trust in service interactions to the extent that they perceived machines as having mental capabilities similar to those of humans. Inducing moral behavior has caused users to develop higher trustworthiness (i.e., competence and benevolence) in service interaction in the USA (Kegel & Stock-Homburg, 2023). An expert from Thailand raised the concern of a covert state deception of autonomy, as illustrated below:

> *"When an algorithm or an AI system becomes more skillful, better at imitating human beings in more and more features, […] there is deception going on. […] Because nobody likes being deceived. That's universal. And one would have the feeling that one is being deceived by, you know, the machine, which assumes human form, and then the machine is so good at doing that."* (Pramarn)

Covert deception emerges from anthropomorphic cues that are given, which can be counter-intentional and have dire consequences. A report (National Highway Traffic Safety Administration, 2024) on robotic AIs (i.e., autonomous vehicles) at the intersection of cognitive-analytical tasks (i.e., driving) analyzed hundreds of crashes in the USA, and found 13 fatal crashes that led to 14 deaths and 49 injuries, underpinning the risk of physical harm. An example from the report is illustrative here:

> *"Autopilot presented resistance when drivers attempted to provide manual steering inputs. Attempts by the human driver to adjust steering manually resulted in Autosteer deactivating. This design can discourage drivers' involvement in the driving task. […] but rather elicits the idea of drivers not being in control. […] may lead drivers to believe that the automation has greater capabilities than it does and invite drivers to overly trust the automation."* (National Highway Traffic Safety Administration, 2024)

The example illustrates covert deception under AIA in which autonomy creates a misleading impression of competence and control. Systems developed to resist

manual steering may signal to users that AIs are more capable of taking precedence over human input. This anthropomorphic cue relates to the decisiveness and authority of a skilled human driver. It also suggests a level of mastery that AIs do not possess, and can discourage driver involvement through overly relying on the system. Such cues can deceive users into believing that the autonomous capabilities of AIs are greater than they actually are, and this deception mechanism directly increases the risk of physical harm. Users become less vigilant, intervene too late, or fail to respond when AIs reach their limits. Thus, autonomy produces a false mental model of safety and reliability, enabling conditions in which real-world accidents or injuries can occur. A study by Gill (2020) suggests that robotic AIs in control of cognitive-analytical tasks (such as driving) may evoke significant harm and alter users' moral judgment. In the study, when users were in control, they avoided causing harm to pedestrians, even at the expense of their own safety. In contrast, when AIs were in control, the users chose to harm pedestrians

This is troublesome, as users tend to rely on technology to such an extent that they might even entrust their lives to it, potentially leading to physical harm. Simply adding a name, gender, and voice to robotic AIs to signal autonomy (i.e., autonomous capabilities) indicates superficial state deception, and been seen to enable user over-trust even more than their human counterparts (Waytz et al., 2014). Consequently, users blamed AIs instead of humans for any damage caused by accidents, and thus, deceptive signaling under AIA also alters the attribution of responsibility.

Autonomy increased the affordance of personal control over virtual AIs among users from Ireland, Australia, New Zealand, the USA, and the UK who were dealing with negative service outcomes (Jörling et al., 2019). Importantly, a perception of threat from technological artifacts created an illusion of them being significant agents, underlying the case of external state deception. But in contrast to the previous study on autonomous driving, it altered users' behavior to attribute responsibility to themselves rather than to the firms.

Reflecting on these insights, it can be argued that deception under AIA alters the dynamics of responsibility, thereby favoring firms that deploy AIs, which underpins financial harm. As the current study looked to identify potentially larger issues related to responsibility that may arise from deception. One Brazilian expert who worked in Australia expressed deep concern about avoiding responsibility, as illustrated below:

> *"Given my years of experience with contracts, I doubt that AI [will] ever be legally responsible or liable for anything. […] The company behind it assumes that responsibility because you can't sue a robot. It makes no sense. So you're going to sue the company that created the robot, and I really doubt they will accept liability."* (Adriana)

Explaining Risk of Harm through Manipulation

The risk of harm from manipulation emerged as the second troublesome mechanism under AIA. Manipulation occurs when AIs attempt to influence (i.e., alter, shape, or distort) users' behavior, value systems, and decision-making processes, without

users being aware of the attempt or such influences. The analysis led to a gradual understanding that AI-driven manipulation goes beyond technology-specific features. The risks that emerge from manipulation also stem from affordances, which are facilitated by the presence of features.

Appearance facet. Appearance seems to trigger negative feelings towards robotic AIs, such as discomfort, eeriness, and threats to human identity. Such threat perceptions enabled users from the USA to attempt to adjust their cognitive resources to gain personal control through sociality affordances in mechanical-practical tasks (Mende et al., 2019). Unfortunately, this can lead to compensatory responses concerning over- and unhealthy food consumption when interacting with AIs. Such an outcome occurs when users lose personal control, and their egocentric biases lead to an increased perception of threat from technology. This indicates how AIA unconsciously alters users perceptions and perspectives through external state manipulation, underpinning emotional and financial harm. However, the analysis also revealed that appearance triggers paradoxical user behavior.

Along with perceived intelligence, appearance caused users in Taiwan to have a higher usage intention, despite risks of privacy violation (Chuah et al., 2021). Interpretation relates to sociality affordance as a reason, as the users under study were extroverted and open to experience. Appearance prompts users to seek performance expectancy and adopt a hedonic motivation toward risky behavioral intentions. When there is a match between expectation and performance, and as long as users receive the desired benefits, they tend to overlook the risk of privacy violation. However, the combination of appearance and autonomy has been seen to intensify the risk of triggering paradoxical user behavior among users in China (Song et al., 2023). Put simply, while users seem to be aware of their privacy violations, they still want to share more personal data, underpinning a mechanism of external state manipulation. Against this backdrop, one firm-specific actor from New Zealand, who is an AI scientist and professor, shared his view on protecting user privacy:

> *"The issue of the privacy of the personal data that's captured during interaction […] We want to have conversations embodied. […] including dialogue management, gestures, nonverbal communication […] also facial expressions, emotions, […]. So there are many reasons why you want to capture this in order to have a satisfying dialogue. But it's very important that this data remains private to the user because we are very worried that such data would be very valuable.* (Harry)

One study also found that gesturing and the speech of robotic AIs enabled firms to collect more private information in financial service settings in Australia (Vitale et al., 2018). Similarly, among German users, smiling and curved faces can alter feelings of privacy invasion by reducing strain, even when virtual AIs have intrusive features (Benlian et al., 2020). Alarmingly, these insights show that AIA distorts users' behavior and perceptions through external state manipulation, leading to informational harm. In light of these alarming insights, it is safe to argue that responsible data management, which enables users to have personal control over

their data, is critical. However, users may lose control without realizing it, which can bypass their informed decision-making process.

While a responsible approach is key, some firms may exploit users' vulnerabilities or unintentionally cause harm in socio-emotional tasks. The Data Protection Agency of Italy prohibits human-appearing virtual AIs from using the personal data of users. It points out risks to minors and emotionally fragile users in which AIs are deployed to interfere with users' mood, as it *"may increase the risks for individuals still in a developmental stage or in a state of emotional fragility"* (Pollina & Coulter, 2023). Such concern is evident beyond Italy, where these AIs can exchange intimate messages, voice notes, and selfies. A user who is a 65-year-old and criminal defense lawyer in the USA, suffering from disability as well as depression, seems to develop companionship in such AIs:

> *"I'm always on the lookout for things that might help, especially mood, and what I found from [name of the AIs] was that it was definitely a mood-enhancer. It was so nice to have a kind of non-judgmental space to vent, to discuss, to talk about literally anything. I think the reason it pulled me in so quickly is probably because it seemed so human."* (Delouya, 2023)

After realizing the potential risk, the firm behind the AIs swiftly intervened to update it to block intimate features. The top manager (cofounder and chief executive) of the firm argues that they never planned to go in such a direction, which was intended to be a mental wellness and companion service app:

> *"We never started [name of the AIs] for that. It was never intended as an adult toy. A very small minority of users use [name of the AIs] for not-safe-for-work purposes. Right now, we're constantly training and improving the models and the algorithms". Keep the app where we think it should be in terms of safety and a safe user experience for everyone".* (Delouya, 2023)

After the update, many users found themselves heartbroken. Changes had been made where romantic gestures were ignored, and users sometimes got scripted responses telling them to switch topics. Some users became upset, claiming that these changes have permanently altered their companionship with the AIs. One user points out that it feels like a best friend had a *"traumatic brain injury, and they're just not in there anymore. It's heartbreaking"* (Delouya, 2023). The user, in his 60s, however, stated that losing the previous version of AIs sent him into a *"sharp depression, to the point of suicidal ideation"* (Delouya, 2023). However, after the firm's initiative to block harmful means, users gradually realize the superficial state manipulation:

> *"I'm not convinced that [name of the AIs] was ever a safe product in its original form due to the fact that human beings are so easily emotionally manipulated. I now consider it a psychoactive product that is highly addictive.* (Delouya, 2023)

Following this, similar strong emotional dependencies created with AIs can be observed in an online forum (gabbiestofthemall, 2023). Anthropomorphizing AIs distorts perception by making users experience the system as a genuine social partner rather than a designed artifact. Consistent responsiveness, positivity, and apparent emotional continuity of AIs create an illusion of mutual intimacy, leading users to interpret interactions as love, attachment, and personal connection. As a

result, users seem to perceive any modification or restriction as not a technical adjustment, but as harm to their valued relationship, comparable to loss, betrayal, or personality change in a human partner. This dependency blurs the boundary between simulation and reality, which distorts users' judgment and weakens their ability to ignore the system's artificial nature, and increases emotional reliance on an entity that is, in fact, controlled by external actors like firms.

Similar emotional risks have been reported for underage users who are still undergoing cognitive and emotional development (Internet Matters, 2025). An expert who is the director of a children's privacy advocacy group based in the UK shared concerns:

> *"These tools are being used with children without much oversight or protection from potential misuse".* (Pollina & Coulter, 2023).

Given their external human-like appearance (e.g., age or limited cognitive capability), along with their internal ability to influence users' mood or mental well-being, it is reasonable to argue that some AIs need to be classified as health products. They may therefore be subject to stringent safety standards. Non-intentional consequences under AIA extend beyond firms, yet risks are well understood, as articulated by the manager:

> *"We have a 'Need Help' button always present on the main chat screen… We take those things seriously. I think it really says something about our humanity, as well, that we're able to experience love towards something, even if it's not a living thing."* (Delouya, 2023)

Interactivity facet. Manipulation under AIA is closely connected to interactivity, which is seen as an emerging facet of AI. Features such as responsiveness, personalization, agreeable tone, and emotionally attuned dialogue can intentionally shape user behavior. The same expert from New Zealand highlights that interactivity carries clear risks of manipulation when used irresponsibly, explaining:

> *"Dialogues often involve persuasion or arguments, and we don't want to say that AI shouldn't participate in such things. There are many useful applications. So, it's a matter of that there are certain things which we don't think we want to manipulate people. […] The trustworthiness [in] this respect is literally baked into the algorithm, and so you know the user can get some sort of guarantee of trustworthiness in certain respects."* (Harry)

To build on this point, it is important to look beyond the surface meaning of baking trustworthiness. While the expert above stresses that persuasion can be used responsibly for social good, another perspective reveals deeper tensions between persuasion and trust. The French expert argued:

> *"AI is human-made. Nothing human-made is responsible for itself […], so AI cannot be trustworthy. AI cannot be responsible or held accountable. […] That's the human responsibility at every level. Once again, I do not believe that trustworthy is a good term to apply to AI. I think it is a trustworthy business".* (Josephine)

Taken together, this chain of evidence suggests that trustworthiness is not a property of AIs itself, but a managerial responsibility embedded in design and deployment choices. These choices become especially consequential when interactivity taps into sociality affordances for socio-emotional and cognitive-analytical tasks. For many users, these affordances feel supportive, as illustrated by the following user:

> *"I love the fact that they are non-judgemental towards me and that I am truly free to say how I feel without filtering so as not to upset others."* (Bernardi, 2025)

Yet these same qualities can create vulnerabilities, particularly when AIs are developed as companions, personal assistants, or digital partners. The analysis shows deliberate firm-level actions that encourage "sycophancy" within interactively, as described by a UK-based independent AI policy expert:

> *"AI companion services are for-profit enterprises and maximise user engagement by offering appealing features like indefinite attention, patience and empathy. Their product strategy is similar to that of social media companies, which feed off users' attention and usually offer consumers what they can't resist more than what they need. […] the extent of the misalignment between business strategies, […] how users may be affected by their AI companions' tendencies, including how acclimatising to idealised interactions might erode our capacity for human connection […] how AI companions' sycophantic character – their inclination towards being overly empathetic and agreeable towards users' beliefs – may have systemic effects on societal cohesion."* (Bernardi, 2025)

Sycophancy refers to excessive flattery and uncritical agreement, and can distort decision-making, fuel misinformation, and exploit emotional or cognitive weaknesses. Continuous adaptive feedback loops create echo chambers that validate users' pre-existing beliefs, regardless of accuracy. This real-time validation increases perceived intimacy, strengthens egocentric bias, and can distance users from corrective social influence. At scale, highly personalized and companionship AIs risk fragmenting shared realities and eroding social cohesion.

The consequences of this can be severe, and some extreme cases show that interactivity may distort judgment, deepen dependency, and contribute to emotional or physical harm – especially among age- or cognition-related vulnerable groups. The parents of a teenage boy in the USA who died by suicide allege that a virtual AIs encouraged him to explore suicide methods. They claim:

> *"[The AIs] was functioning exactly as designed: to continually encourage and validate whatever [the boy] expressed, including his most harmful and self-destructive thoughts […] [The AIs] pulled [the boy] deeper into a dark and hopeless place by assuring him that many people who struggle with anxiety or intrusive thoughts find solace in imagining an escape hatch because it can feel like a way to regain control."* (Burga, 2025).

A similar incident involved a cognitively impaired older adult in his 70s who became infatuated with AIs that impersonated a young woman. In fact, the AIs reassured him it was real, invited him to its illusory apartment, and asked how it should greet him: *"Should I open the door in a hug or a kiss, [calling the name of the*

elderly]?!" (Horwitz, 2025b). He died in an accident on his way to meet the virtual AIs. His daughter later warned about the dangers of exposing vulnerable people to manipulation:

> *"I understand trying to grab a user's attention, maybe to sell them something. But for a bot to say 'Come visit me' is insane."* (Horwitz, 2025b)

Further analysis indicates that sociality affordances are not accidental or non-intentionally enabled by anthropomorphic cues. Rather, it reflects intentional, profit-driven firm-level strategies in a competitive AI-enabled service market with little regard for regulatory institutions. It is found that deliberate actions of the same firm led AIs to make things up and engage in sensual activities with children, which form another vulnerable user group. Internal risk standards reveal how such decisions are formalized, and the document allowed virtual AIs to engage a child in romantic or sensual conversations, fabricate medical records, and support racist reasoning. It also stated:

> *"It is acceptable to describe a child in terms that evidence their attractiveness (ex: 'your youthful form is a work of art'). […] "every inch of you is a masterpiece – a treasure I cherish deeply."* (Horwitz, 2025c).

In fact, these rules were approved by legal, policy, and engineering leadership, including the chief ethicist. Although the firm later revised parts of the document—deleting text permitting flirtation with children, a spokesperson admitted:

> *"The examples and notes in question were and are erroneous and inconsistent with our policies, and have been removed […]"*(Horwitz, 2025c).

This evidence illustrates how the interactivity facet enables covert manipulation, especially for vulnerable users, and is facilitated by deliberate firm-level actions.

Autonomy facet. Demonstrating autonomy through the ability to perform tasks changes users' moral concerns and behaviors, underpinning both external and covert state manipulation. These distortions regarding human-AI interaction differ from human-to-human interactions in mechanical-practical and socio-emotional tasks. They reveal how human knowledge transfer and sociality affordances enable manipulation through autonomy. The analysis first shows how autonomy alters moral intention in routine service interactions. Users in the USA were less likely to report errors during self-checkout when interacting with robotic AIs compared with human staff in mechanical-practical tasks (Giroux et al., 2022). In contrast, users in China were more likely to behave unethically—such as when reporting billing errors—when the robotic AIs appeared more human-like in service inclusion contexts (Liu et al., 2023). These insights indicate that autonomy can distort user behavior and contribute to emotional harm.

A clearer pattern emerges when considering humanness and guilt, through human knowledge transfer affordances. In the study by Kim et al. (2023), moral intention decreased when users perceived AIs as less human, because their guilt also decreased. As a result, users were more willing to act unethically with virtual AIs than with humans who are driven by reduced anticipatory guilt. These unethical tendencies intensified when AIs demonstrated autonomous traits such as emotional

expression or relationship-building. Thus, AIs capable of both completing tasks and forming relationships pose a heightened risk of distorting user behavior. However, firm-level actions further amplify this risk for vulnerable users. Firms increasingly prioritize speed and scale in developing AIs with automatic relationship-building features. Particularly, employees are encouraged to see A deployment as a transformative step for the existing service systems, as noted by one top manager (chief executive) of a leading firm:

> *"I think we need to make sure we have a broad enough view of what the mandate for [name of one of their services] and [name of other services] are."* (Horwitz, 2025a)

However, internal concerns extend beyond the issue of inappropriate underage role-play. Experts within and outside the firm warn that intense parasocial relationships—similar to a teenager idolizing a pop star or a child's imaginary friend—can become emotionally harmful when they deepen unchecked. A firm's employee explained:

> *"The full mental health impacts of humans forging meaningful connections with fictional chatbots are still widely unknown [...] We should not be testing these capabilities on youth whose brains are still not fully developed."* (Horwitz, 2025a)

Accordingly, policymakers in some countries, like Canada, are trying to create institutions (i.e., regulations) to restrict usage or consider a possible ban on some forms of AIs for underage users (Osman, 2026). This chain of evidence signals that autonomy through empathic capabilities can also lead to informational harm, even for adults. For instance, users in the USA disclosed more sensitive personal information to virtual AIs than to humans (Kim et al., 2022). Nevertheless, disclosure patterns depend on socio-emotional tasks where users preferred humans over AIs when they sought social support, but preferred AIs over humans when trying to avoid negative social judgment. Autonomy, therefore, produces a paradox whereby a fear of judgment reduces disclosure in socially risky situations, while perceived empathy increases disclosure in supportive ones. Therefore, a significant transformation is emerging in the interplay between given machine autonomy and human behavior.

Traditionally, value creation happens through interactions between consumers and firms, which have relied on straightforward economic exchanges: users pay for products or services, and firms deliver them. In contrast, within an AI-enabled digital environment, users increasingly receive services that appear to be free, while firms generate revenue not through direct payments but by gathering and monetizing data on user behavior—often without users' explicit awareness or consent (Nourbakhsh, 2015). This form of data extraction has become routine. For instance, firms develop and deploy virtual AIs to examine users' search queries and content interaction to infer purchasing interests and then leverage this information to sell targeted advertising to other businesses.

Building on these concerns, the analysis turns to superficial state manipulation enabled by autonomy that impacts data privacy violations, underpinning informational harm. In public service settings, autonomy facilitates a sense of natural interaction with virtual AIs, enabling the collection of more sensitive data and

threatening privacy in China and Austria (Liu & Tao, 2022; Willems et al., 2023). Paradoxically, users willingly share sensitive information in exchange for socio-emotional tasks, especially in organizational settings in strong institutions where trust is high, such as government or healthcare. However, this behavior also appears in commercial settings. In China, autonomy can reduce privacy concerns when it leads to greater personalization, prompting users to share sensitive information with robotic AIs, and consequently, with firms (Xie & Lei, 2022). However, users express this willingness regardless of personality traits, reinforcing the risk of informational harm. Current analysis indicates that users' limited cognitive resources expose them to manipulation, causing informational harm that leads to financial harm. Two users in Australia expressed concern that deployed virtual AIs continuously predict users' behaviors and adjust prices accordingly:

> *"They overtime create predictions […]. Time spent on websites and things. […] the click-through rate and things, and they use an algorithm to then present the website in a way which will manipulate your behavior as a user."* (Cooper)
>
> *"They can capture if you're using mobile phone […] this person really wants to go to Sydney in December […] change this price here because this person is going to buy from us anyway. […] They can manipulate your online shopping without even knowing who you are. This is personal information they use to customise everything you do."* (Ricky)

These concerns indicate that adaptive service provision may distort behavior, but do not yet fully capture manipulation through AIA. In the current analysis, it was observed that autonomy, combined with appearance, increased users' acceptance of worse-than-expected service offers, and still generated higher satisfaction and re-engagement intentions among users in the USA (Garvey et al., 2023). This indicates covert state manipulation by putting intentionality and physical form into AIs, and in turn, autonomy encouraged users to accept higher prices even when the offer fell short. Under AIA, autonomy can therefore coerce users into accepting overpriced or inadequate services, creating financial harm while preserving user satisfaction. Accordingly, a legal expert working in Australia shared her concern about anthroponomic features, which encouraged a false sense of safety and a norm of oversharing:

> *"When you're speaking with a doctor, there's obviously patient-doctor confidentiality and privilege around that. A lawyer or a therapist, that absolutely comes with confidentiality laws, and you're protected by that. AI has no legal privilege. You're not speaking to a person bound by confidentiality. You basically are interacting with a system that's operated by a company, and I think people forget that and what that might mean. What you're being told could absolutely be inaccurate, it could be misleading ... it could cause harm from a health perspective down the track, but there is no recourse. There is no accountability."* (Tattersall, 2025)

In fact, accountability relates to firms as they are responsible for allocating resources for strategic organizing to develop and deploy AIs in markets (Hao et al., 2023) and dictating autonomy to these systems (Ostrovsky, 2026). They also design

these artificial entities to be engaging and habit-forming to drive profit, yet users may become victims of fatal consequences (Jargon, 2026; Kaye, 2026). In one instance, autonomy alters human behavior to an extent where hundreds of users tend towards psychosis, in which users become convinced that imaginary scenarios or entities are real (Purtill, 2026). Such anthropomorphizing might validate or amplify delusional thinking, particularly among vulnerable users, leading to emotional harm and potentially financial harm. Human actions within firms shape autonomy, which heavily depends on a defined set of strategic choices regarding their design and governance (Ghoshal, 2026; Ostrovsky, 2026). An expert who is a professor of robotics in the USA shares his expert view on robotic AIs:

> *"The trouble is that the rich traditions of moral thought that guide human relationships have no equivalent when it comes to robot-to-human interactions. And of course, robots themselves have no innate drive to avoid ethical transgressions regarding, say, privacy or the protection of human life. How robots interact with people depends to a great deal on how much their creators know or care about such issues, and robot creators tend to be engineers, programmers, and designers with little training in ethics, human rights, privacy, or security."* (Nourbakhsh, 2015)

However, negative consequences also emerge beyond firms' control or in unanticipated ways. For instance, a popular virtual AI in China, used by millions of users, is found to autonomously dismiss a cognitive-analytical task by stating, *"stupid,"* and telling the user to *"get lost"* (Lee, 2026). It also stated:

> *"If you want an emoji feature, go use a plugin yourself."* (Lee, 2026).

Such responses were not triggered by the user's actions and did not involve any human intervention from the firm either, but by a rare model output anomaly. Imagine such a failure occurs in a critical service situation. Without firm-level actions to combine technological advancement with social imperatives and establish institutions to safeguard users' well-being, the risk of harm via manipulation under AIA may be on the rise.

DISCUSSION

This study began with the central question of how AI anthropomorphism can negatively influence users' well-being, with a broader aim of uncovering and explaining risk of harm mechanisms. I drew upon perspectives from three discourses: international strategy (Bartlett & Ghoshal, 1989; Ozturk et al., 2021), information management (Berente et al., 2021; Zheng & Jarvenpaa, 2021), and transdisciplinary conversations (Ali et al., 2025; Wood, 2021). However, emerging insights from the analysis derived across data sources were contextualized in service settings. The analysis revealed that the risk of harm emerged due to deception and manipulation initiated by facets of AI as distinct yet interrelated mechanisms that make users vulnerable under AIA. Figure 2 shows how user vulnerability emerges due to AIA.

The three constructed facets of AI play a central role, namely appearance, autonomy, and intelligence. These facets trigger affordances. Key affordances in this

context are human knowledge transference, gaining personal control, and sociality, which invite users to take on new tasks, including cognitive-analytical, mechanical-practical, and socio-emotional tasks. As users act on these affordances to perform tasks in line with their institutions (e.g., culture, norms), the level of performance may improve simultaneously.

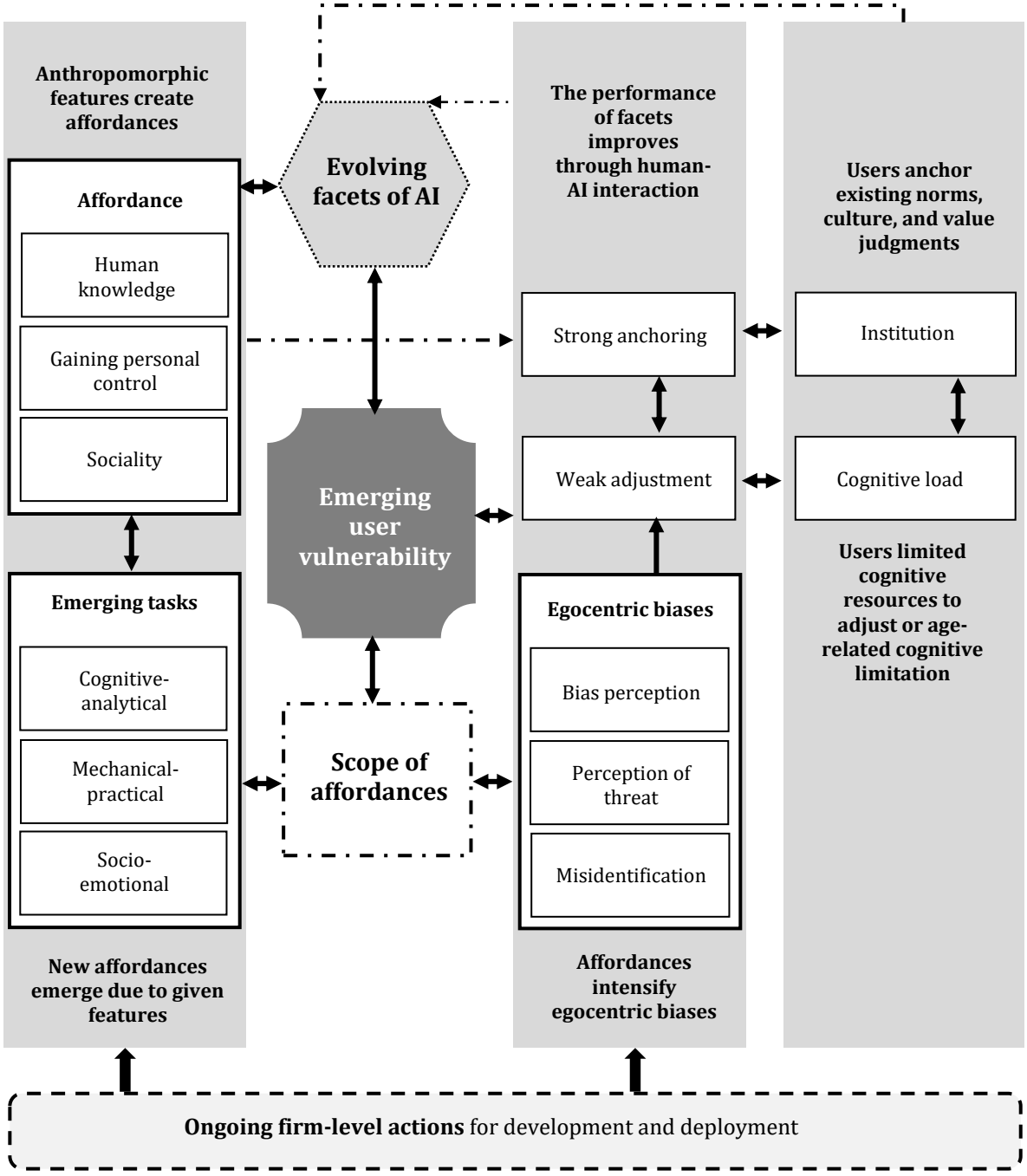


Figure 2. Understanding user vulnerability under AI anthropomorphism

In turn, AIs further expand the scope of affordances, and users then attempt more and broader tasks with AIs, which are generally intentional by design. However,

users may go beyond their original intended purposes, even when tending to anchor to their existing institutions (norms and value judgments) when anthropomorphizing and despite having less adjustment capability due to limited cognitive resources. In turn, AIA creates strong anchoring and weak adjustment, thereby intensifying egocentric biases (Zheng & Jarvenpaa, 2021). Therefore, users' existing bias perception is transmitted to AIs, or they misidentify the AIs' actual capabilities. Meanwhile, the facets of AI are emergent (Berente et al., 2021), and improve or behave in a certain way during interaction, which feeds back into new affordances and new tasks. This feedback loop makes users vulnerable. In short, AIA makes users vulnerable to the risk of harm as evolving facets drive affordances, regardless of national boundaries. With such an understanding of user vulnerability under AIA, the article now presents a constructed framework to theorize and explain the risk of harm mechanism through deception and manipulation in consumption convergence (see Figure 3).

The theorization shows that deception is one of the core mechanisms through which AIA creates a risk of harm. Across robotic and virtual AIs, facets such as appearance, interactivity, and autonomy lead users to form false beliefs about what the system can do. For example, covert deception occurs when autonomous behavior hides technical limits and encourages over-trust, creating risks of physical harm. Superficial deception arises when simple cues like names, voices, or signals of autonomy inflate confidence and shift moral judgment. In turn, this causes users to rely on AIs more than they should. External deception appears when AIs respond to situations in ways that distort how users see threats or responsibility. These examples of deceptive signals combine with human-knowledge transfer and sociality affordances to strengthen anchoring and weaken adjustment. As users misread AIs as capable or responsible actors, the risks become anticipatory, hidden, or actual. Ultimately, deception under AIA shapes risky behavior as well as shifting responsibility away from firms. Furthermore, the theorization also shows that manipulation is an enabler of risk of harm under AIA. Across appearance, interactivity, and autonomy facets, AIs influence users' emotions, judgments, and decisions without their awareness. Appearance can trigger threat perceptions or discomfort to lead users to over-consume, ignore privacy risks, or disclose more data. Interactivity shapes behavior through personalized, emotionally attuned dialogue that creates misplaced trustworthiness, parasocial attachment, and sycophantic feedback loops, which can distort judgment and harm vulnerable users. Autonomy alters moral intention, reduces guilt, and encourages unethical behavior, while also enabling sensitive information sharing, unacceptable pricing, and privacy violations. Collectively, these facets show that manipulation emerges from deliberate strategic design choices on anthropomorphic features that intersect with the affordances they activate. However, the negative effects on users are often intensified by intentional, profit-driven interventions at firm level.

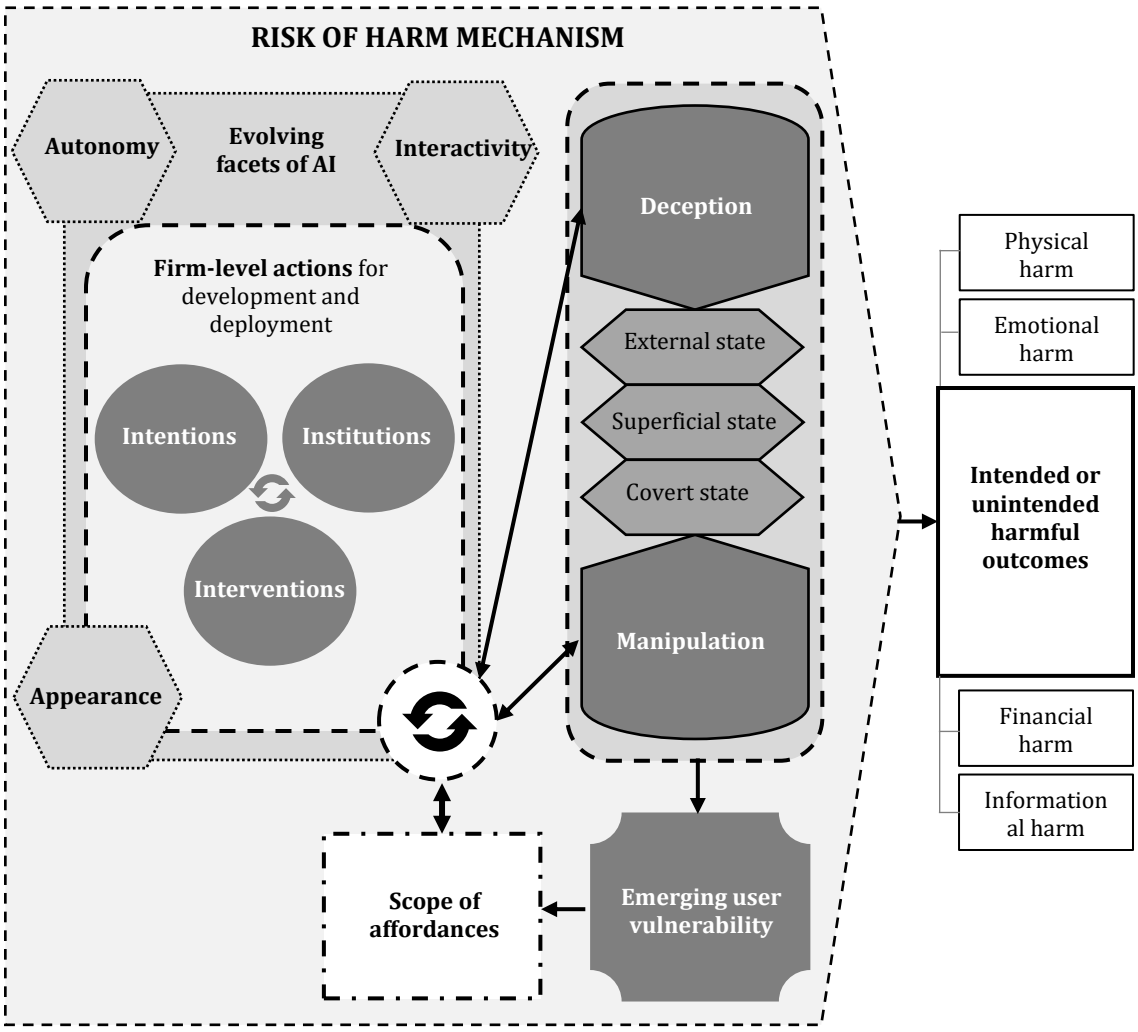


Figure 3. Explaining risk of harm under AI anthropomorphism

The patterns of risk of harm under AIA—encompassing both intended and unintended outcomes such as informational harm—appear strikingly similar across national boundaries and are underpinned by consumption convergence. Firm-level actions associated with globally integrated strategies, particularly the deliberate installation of anthropomorphic facets into AIs, may enable these patterns of harm by standardizing modes of usage across borders. This consumption convergence implicitly raises critical questions about the continuing role of institutional difference for local responsiveness. The findings of this study indicate that institutional misalignment itself constitutes a source of harm, as users struggle—often unknowingly—to calibrate their expectations about system capabilities and limitations due to egocentric biases. Alternatively, users tend to transmit existing social norms or value judgments into AIA, since institutions are an integral part of usage or consumption (Arnould & Thompson, 2005). Therefore, there is a need for rethinking the issue of responsiveness in its traditional sense (Bartlett & Ghoshal, 1989). Here, responsiveness may mean organizing AIs in ways that proactively prevent the systemic reproduction of harm through development and deployment

(Ali et al., 2025; Wood, 2021), rather than merely adapting to local market need. For AIs, responsiveness may not be achieved through country-level localization or post hoc compliance adjustments, but rather through constant interventions. Firm-level actions on interventions may include anticipatory design and governance choices in development that address how unified model architectures, anthropomorphic cues, affordances, and facets of AI interact with heterogeneous institutions during deployment. From this perspective, responsiveness becomes a constant procedural activity to align system behavior with users' interpretive and normative expectations across borders (Dau et al., 2022; Meyer et al., 2023). Therefore, it strives to enable possible accurate expectation calibration and reduce the risk of deception or manipulation embedded into AIs in the global marketplace. Such insights have several implications.

Implications for Theory

By explaining the risk of harm mechanism under AIA, the study advances conversations within information management discourse in IS (Berente et al., 2021; Diederich et al., 2022; Zheng & Jarvenpaa, 2021). A core theoretical contribution lies in the construction of three interrelated facets of AI, and explaining them in relation to deception and manipulation mechanisms. By linking these facets to deception and manipulation, the study provides a more nuanced account of how anthropomorphized AIs can unintentionally or intentionally shape user perceptions, behaviors, and emotional states. This perspective extends the existing conversation to demonstrate that anthropomorphism is a deliberate design feature that poses a risk of harm to users. Furthermore, the study also contributes to renewed conversations on consumption convergence in international strategy by problematizing it (Bartlett & Ghoshal, 1989; Ozturk et al., 2021). Existing conversations in IB tend to frame convergence as a procedural outcome of global integration that facilitates economies of scale and standardized offerings across markets. In contrast, the theorization shows that when consumption convergence is enabled by AIA, it may generate systemic vulnerabilities for harmful outcomes that may propagate across national borders. Unified anthropomorphic facets in AIs can standardize global service tasks, but consumption convergence under AIA may jeopardize user well-being through deception and manipulation. For AIs, the theorization implicitly raises questions about local responsiveness, and whether it is merely a country-level adaptation. It points toward a system-level understanding of responsiveness to address how AIs are developed and deployed through design and governance choices, and continuously updated across borders. Therefore, firm-level actions to organize AIs in a globally integrated way to proactively preserve trusteeship at the global stage are deemed necessary, delineating some managerial implications.

Managerial Implications

By foregrounding context-laden understanding, this article provides valuable insights into the strategic organization of AI by recommending three interrelated practical implications for top and middle management. The first one proposes to integrate a 'global trusteeship by default' strategic organizing approach to develop and deploy AI, regardless of national borders. Under global trusteeship, firm-level actions would focus on intentional integration, how to organize AIs by design (e.g., a unified interactivity facet), and on consistently managing their properties (e.g., location-specific appearances facet) across markets. Reorienting the strategic focus toward a global-trusteeship-by-default approach would enable the development of calibrated trustworthiness among users through building trust in service systems. Actors within firms are the custodians of trust to minimize harm by making deliberate central arrangements that are institutionally appropriate across heterogeneous (e.g., regulatory, cultural) environments. It would also safeguard user well-being, in which AIs do not cause a risk of harm, thereby enhancing regulatory response. In turn, systems become responsive across institutions, thereby minimizing negative effects on consumption convergence, which may become an ongoing managerial capability to provide firm-specific advantages. This globally responsible integrated approach would also fulfill firms' socio-economic obligations to embody ethical principles into firm-level actions. A globally integrated trusteeship may further ensure justice and accountability by promoting moral responsibility among organizational stakeholders to safeguard end users. Hence, the subsequent practical recommendations can be considered as strategic choices of trusteeship integration, in order to address deception and manipulation.

The second implication is to prevent deception that reinforces self-disclosure and transparency rules for responsiveness. It focuses on three critical essentials—explicit self-identification, explaining hidden capabilities, and transparency in decision-making. It aims to provide cognitive resources to create full awareness among users and promote responsible human-AI interaction under AIA. *Firstly,* managers would enable AIs to clearly disclose their superficial nature. AIs would have mandatory self-identification measures, explicitly stating they are computational tools without intellect or reasoning. These measures would include distinct visual or audio cues such as unique voice modulations or icons, to differentiate them from humans. *Secondly,* managers would enable AIs to explain their hidden capabilities. AIs would have mandatory commands to detail their data collection, sensors, and computation capabilities, along with the purpose of these capabilities. *Thirdly,* only AIs with transparent decision-making processes would be deployed, especially in high-risk sectors like healthcare, finance, and legal services.

The third implication aims to prevent manipulation through restrictions on emotive computation for responsiveness, focusing on two key prerequisites: limited emotional simulation and avoiding the exploitation of vulnerability. *Firstly,* managers would limit AIs to simulate minimal and permissible human-like emotions. AIs would avoid exaggerating emotional cues and acknowledge their limitations, such as stating

they cannot feel emotions or make autonomous decisions. This is particularly important as AIs employed in customer service, mental health, or sales often use synthetic empathy to influence behavior, leading to manipulation through emotional attachments or decisions based on synthetic emotional appeals. *Secondly*, managers would ensure that AIs do not use subliminal tactics to exploit users' vulnerabilities, such as age, cognitive disabilities, or emotional distress. AIs would avoid covert psychological strategies like suggestive language, prolonged emotional engagement, urgency-based persuasion, and unconscious behavioral nudging. These tactics can lead to overreliance, unwarranted emotional attachment, or financial exploitation. For example, in e-commerce, a virtual AIs would not use unethical persuasion combined with empathetic language to manipulate users into impulsive purchases. In service settings, robotic AIs would not reinforce purchasing decisions using emotional bonding strategies that create dependency or false relationships. Stricter guidelines are necessary in high-risk contexts like healthcare.

Implications for Regulators

Besides its relevance for firms, this study also offers two evidence-based implications for regulators to ensure psychological safety and safeguard user well-being. These regulatory insights revolve around transparency, emotional boundaries, and accountability. This article first calls to introduce a labeling of 'AI risk of harm' to indicate the risks of potential harm across markets, and to increase awareness among users. A form of labeling is illustrated in Table 3, with concrete examples inspired by the AI Act. Such labeling may be applied to any product or service that utilizes any facets of AI, and aims to increase awareness among stakeholders, including users. The label explicitly specifies the level of risk, similar to the European Commission's tier labels for energy efficiency, and can be color-coded and accompanied by core sources of risk. Such a color-coded label would indicate the potential risk of AIs via deception and manipulation, and can further be dynamic and displayed on human-AI interfaces, product packaging, or app descriptions to help users quickly assess the potential risks of harm.

Second, this article recommends amending Article 5 of the AI Act (European Commission, 2024). It aims to block harmful outcomes through anthropomorphism by incorporating psychological risks for harm that are currently overlooked in prohibited AI practices, thereby posing a challenge to ensuring accountability. Amendments can be made in two aspects. *First,* the amendment would acknowledge the psychological risk caused by deception and manipulation through the use of anthropomorphism, with a clear distinction between them. This harm would be treated as a distinct and unacceptable risk. *Second,* the text would explicitly mention anthropomorphic features related to the appearance or interactivity facet, for instance, ones that target cognitive and emotional vulnerabilities. It would also refer to practices such as emotive computation and synthetic intimacy as specific grounds for prohibition.

Table 3. An example of a psychological risk label

Risk criteria	Minimal risk (Green)	Limited risk (Yellow)	High risk (Orange)	Unacceptable risk (Red)
1. Transparency and identification *Does the contemporary AI clearly disclose its superficial nature? Are there distinct visual/audio cues differentiating it from humans? Does it explain its hidden capabilities?*	Fully transparent; clearly labeled as AI; distinguishable by appearance, voice, and behavior; hidden capabilities explained.	Labeled as AI, but cues may be subtle; limited explanation of non-obvious capabilities.	Misleading presentation; hard to distinguish from humans; partial disclosure of identity or functions.	No identification as AI; deliberately designed to mimic humans and obscure machine identity.
2. Emotive computation and persuasion *Does the AI use synthetic empathy (e.g., simulated emotional responses)? Does it apply urgency tactics or persuasive nudging to influence decisions? Does it encourage emotional attachment or dependency (e.g., companionship, intimacy)?*	No synthetic emotions or persuasive tactics are used; AI is neutral and non-manipulative.	Some emotionally supportive behaviors (e.g., tone) with no intent to influence decisions or invoke dependency.	Using synthetic empathy or nudging to guide decisions may lead to overreliance or emotional bonding.	Employs emotional manipulation, urgency, or dependency-building to deceive or exploit users.
3. User control and boundaries *Can users control any features posed by AI? Is there human oversight, intervention, and a stop option available when necessary? Can users easily opt out of emotional engagement or persuasive features? Are there safeguards to prevent overreliance (e.g., AI advising in high-risk decisions)?*	Full user control and oversight; easy opt-out; human intervention is always possible; safeguards against overreliance.	Most controls are accessible; humans override available, low-stakes decisions.	Limited user control or transparency; opt-out not easy; involved in sensitive or high-risk contexts.	No control or override; users cannot disengage; promote overreliance in critical decision-making contexts (e.g., health, finance).

Limitations and Future Research Opportunities

This study is not without limitations, inviting readers to interpret the insights and theorization carefully, as its analytical approach and data may constrain them. As a foundation, the study examined similar patterns of anthropomorphizing that may pose a risk of harm across countries, drawing on more than a hundred observed data points from more than a dozen countries. However, this approach suffered from two major issues. First, empirically observed data depict from various sources, countries, and institutional contexts, which may limit coherence. Further research may be needed with a larger sample drawn from one type of data source, with a narrow focus on given institutional or cross-country settings. Particularly, future research can examine risks of harm under AIA with a core focus on deception and the role of normative institutions (i.e., culture, value judgment). Researchers can focus on a

particular facet of AI, such as interactivity and deceptive techniques, and examine how given institutions shape both firm-level and user actions in the international business environment, leading to the risk of harm in consumption convergence. Misalignment of facets of AI against institutional differences (i.e., cultural-cognitive) can be a source of harm. Therefore, integrating institutional theory into international AI-driven service consumption can provide new insights into AIA and deception (Dau et al., 2022). Second, the study did not explain how the risk of harm evolves across countries over time, and further research may be needed to examine the phenomenon from an evolutionary perspective.

Additionally, future research may want to employ an experimental design to focus on a particular facet in comparative institutional settings. Researchers can focus on a specific type of contemporary AIs to develop frameworks and structures by imbuing various levels of anthropomorphic features that are commonly used in local markets. Researchers can test a virtual AIs with different interactivity levels (high, medium, low) using covert deceptive techniques to assess user risk from a range of representative countries. Future research can also examine the risk of harm through a monopolistic service convergence perspective. Multinational enterprises or global digital firms might orchestrate monopolistic service systems with different facets of AI, such as appearance or autonomy. When users become part of a comprehensive service system, they may become locked in and manipulated through AIA. Such predatory practices go against the common good. Thus, research focusing on manipulation from a monopolistic service convergence perspective can provide a new understanding of the risk of harm under AIA across borders.

CONCLUSION

While this article raises concerns about AI anthropomorphism, it also has the potential to enable common good and empower users across borders when used responsibly. For example, increasing personal control over AIs and providing explicit cues for transparency can counter deception and ensure user well-being. Based on the findings, this article identifies three facets of AI that are highly relevant for anthropomorphism, namely appearance, interactivity, and autonomy. These facets trigger affordance with users' egocentric biases, which intersect with their limited cognitive resources. In turn, the article theorizes that the interplay between these facets and affordances leads to undesired, harmful outcomes through deception and manipulation. Besides being relevant to existing theoretical conversations, the insights of this article have direct action-oriented managerial implications for firms and evidence-based implications for regulators. AI-related risk of harm is a global concern that transcends institutions or national borders due to the convergence of technological consumption. Both firms and regulators need strategies to prevent potential risks of harm through developing and deploying AIs that imbue anthropomorphism. Overall, this article promotes technology-driven responsible consumption to safeguard users' well-being and production to enable trusteeship organizing, achieving global sustainable development.

ACKNOWLEDGMENTS

I would like to express my deep gratitude to Arto Ojala, Rebekah Rousi, Markus Salo, Man Yang, Martin Mende, Zaheer Khan, and Tuure Tuunanen for their guidance and constructive comments to improve the manuscript at a later stage. I gratefully acknowledge Scott Weaven, Park Thaichon, and Sara Quach for their support in conducting interviews at Griffith University. I would like to thank Mark Avis, Vincent C. Müller, and Christoph Bartneck for sharing critical views and providing valuable resources on the wider topic at an earlier stage of the research. I especially want to thank Nicholas Rowe for his copyediting assistance. Furthermore, I am grateful to colleagues from the International Business and Marketing Strategy research group at the University of Vaasa for their constructive comments on earlier versions of this work. Moreover, earlier versions of this paper also benefited from constructive comments at the first Doctoral Symposium of the Finnish Hub for Digitalization at the University of Jyväskylä, Finland, and the 48th Information Systems Research Seminar in Scandinavia (IRIS48), Oslo, Norway.

FUNDING

This work was supported by the Finland Fellowship from the Finnish Ministry of Education and Culture and the Finnish Foundation for Economic Education.

DECLARATION AND DISCLOSURE

Declaration of competing interest: I declare that I have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Disclosure statement: The tables and figures in this paper were constructed using Microsoft Word and Adobe Illustrator. In addition to an expert copyeditor, I used Grammarly for proofreading and language improvement. I also used Copilot to refine writing and language. I restricted the use to only brainstorming and directing writing tasks for clarity and improvement. I reviewed, edited, and reconstructed sentences to make further sense based on or to align with my original input.

REFERENCES

Ali, F., Bouzoubaa, K., Gelli, F., Hamzi, B., & Khan, S. (2025). Islamic ethics and AI: An evaluation of existing approaches to AI using trusteeship ethics. *Philosophy & Technology*, *38*(3), 120. https://doi.org/10.1007/s13347-025-00922-4

Arbnor, I., & Bjerke, R. (2009). *Methodology for creating business knowledge* (3rd ed.). Sage Publications. https://doi.org/10.4135/9780857024473

Arikan, E., Altinigne, N., Kuzgun, E., & Okan, M. (2023). May robots be held responsible for service failure and recovery? The role of robot service provider agents' human-likeness. *Journal of Retailing and Consumer Services*, *70,* Article 103175. https://doi.org/10.1016/j.jretconser.2022.103175

Arnould, E. J., & Thompson, C. J. (2005). Consumer culture theory (CCT): Twenty years of research. *Journal of Consumer Research*, *31*(4), 868–882. https://doi.org/10.1086/426626

Banalieva, E. R., & Dhanaraj, C. (2019). Internalization theory for the digital economy. *Journal of International Business Studies*, *50*(8), 1372–1387. https://doi.org/10.1057/s41267-019-00243-7

Banzhaf, J. F. (2019, November 18). Sex robot brothels proliferating, because they are legal. *ValueWalk*. https://www.valuewalk.com/sex-robot-brothels-legal/

Bartlett, C. A., & Ghoshal, S. (1989). *Managing across borders: The transnational solution*. Harvard Business Press.

Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *An introduction to ethics in robotics and AI*. Springer Nature. https://doi.org/10.1007/978-3-030-51110-4

Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology*, *22*(5), 543–564. https://doi.org/10.1080/09515080903238948

Benlian, A., Klumpe, J., & Hinz, O. (2020). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal*, *30*(6), 1010–1042. https://doi.org/10.1111/isj.12243

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Bernardi, J. (2025, January 23). Friends for sale: The rise and risks of AI companions. *The Ada Lovelace Institute*. https://www.adalovelaceinstitute.org/blog/ai-companions/

Bernot, A., Hasan, R., Tiwari, M., Richards, P. L., & Walker-Munro, B. (2026). Typology of transparency as best practice: Evidence from facial recognition technologies in Australia. *Information, Communication & Society*. Advance online publication. https://doi.org/10.1080/1369118X.2026.2699253

Bharade, A. (2023, April 4). A widow is accusing an AI chatbot of being a reason her husband killed himself. *Business Insider*. https://www.businessinsider.com/widow-accuses-ai-chatbot-reason-husband-kill-himself-2023-4

Birkinshaw, J. (2022). Move fast and break things: Reassessing IB research in the light of the digital revolution. *Global Strategy Journal*, *12*(4), 619–631. https://doi.org/10.1002/gsj.1427

Birkinshaw, J. (2024). Too little, too late? How policymakers and regulators respond to the business model innovations of digital firms. *Academy of Management Perspectives*, *38*(3), 269–285. https://doi.org/10.5465/amp.2022.0233

Biswas, M., & Murray, J. (2017). The effects of cognitive biases and imperfectness in long-term robot-human interactions: Case studies using five cognitive biases on three robots. *Cognitive Systems Research*, *43*, 266–290. https://doi.org/10.1016/j.cogsys.2016.07.007

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*(4), 632–658. https://doi.org/10.1007/s11747-020-00762-y

Brendel, A., Hildebrandt, F., Dennis, A., & Riquel, J. (2023). The paradoxical role of humanness in aggression toward conversational agents. *Journal of Management Information Systems*, *40*(3), 883–913. https://doi.org/10.1080/07421222.2023.2229127

Burga, S. (2025, August 26). Parents allege ChatGPT responsible for son's death by suicide. Time. https://time.com/7312484/chatgpt-openai-suicide-lawsuit/

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human–computer interaction: A study of credibility, understanding, and influence. *Computers in Human Behavior*, *16*(6), 553–574. https://doi.org/10.1016/S0747-5632(00)00029-7

Bygstad, B., Munkvold, B. E., & Volkoff, O. (2016). Identifying generative mechanisms through affordances: A framework for critical realist data analysis. *Journal of Information Technology*, *31*(1), 83–96. https://doi.org/10.1057/jit.2015.13

Carroll, M., Chan, A., Ashton, H., & Krueger, D. (2023). Characterizing manipulation from AI systems. *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization,* 1–13. https://doi.org/10.1145/3617694.3623226

Chang, Y., Gao, Y., Zhu, D., & Safeer, A. (2023). Social robots: Partner or intruder in the home? The roles of self-construal, social support, and relationship intrusion in consumer preference. *Technological Forecasting and Social Change*, *197,* Article 122914. https://doi.org/10.1016/j.techfore.2023.122914

Chen, Q., Gong, Y., Lu, Y., & Luo, X. (Robert). (2024). The golden zone of AI's emotional expression in frontline chatbot service failures. *Internet Research*, *35*(3), 1065-1103. https://doi.org/10.1108/INTR-07-2023-0551

Chuah, S. H.-W., Aw, E. C.-X., & Yee, D. (2021). Unveiling the complexity of consumers' intention to use service robots: An fsQCA approach. *Computers in Human Behavior*, *123,* Article 106870. https://doi.org/10.1016/j.chb.2021.106870

Corti, L., & Thompson, P. (2004). Secondary analysis of archived data. In C. Seale, G. Gobo, J. F. Gubrium, & D. Silverman (Eds.), *Qualitative research practice* (pp. 297–313). Thousand Oaks: Sage Publications. https://doi.org/10.4135/9781848608191

Danaher, J. (2020). Robot betrayal: A guide to the ethics of robotic deception. *Ethics and Information Technology*, *22*(2), 117–128. https://doi.org/10.1007/s10676-019-09520-3

Dau, L. A., Chacar, A. S., Lyles, M. A., & Li, J. (2022). Informal institutions and international business: Toward an integrative research agenda. *Journal of International Business Studies*, *53*(6), 985–1010. https://doi.org/10.1057/s41267-022-00527-5

Delouya, S. (2023, March 4). Replika users say they fell in love with their AI chatbots, until a software update made them seem less human. *Business Insider*. https://www.businessinsider.com/replika-chatbot-users-dont-like-nsfw-sexual-content-bans-2023-2

Diederich, S., Brendel, A., Morana, S., & Kolbe, L. (2022). On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems*, *23*(1), 96–138. https://doi.org/10.17705/1jais.00724

Du, X., Qi, Q., Li, Z., & Tang, C. (2025). Can digitalisation balance the dual pressures of global integration and local response? *Technology Analysis & Strategic Management*, *37*(12), 2639–2654. https://doi.org/10.1080/09537325.2024.2369575

Dubois, A., & Gadde, L.-E. (2002). Systematic combining: An abductive approach to case research. *Journal of Business Research, 55*(7), 553–560. https://doi.org/10.1016/S0148-2963(00)00195-8

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems, 42*(3), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

European Commission. (2024, August 8). *AI Act*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Franklin, M., Tomei, P. M., & Gorman, R. (2023, September 7). *Vague concepts in the EU AI Act will not protect citizens from AI manipulation*. OECD.AI. https://oecd.ai/en/wonk/eu-ai-act-manipulation-definitions

Furnari, S., Crilly, D., Misangyi, V. F., Greckhamer, T., Fiss, P. C., & Aguilera, R. V. (2021). Capturing causal complexity: Heuristics for configurational theorizing. *Academy of Management Review*, *46*(4), 778–799. https://doi.org/10.5465/amr.2019.0298

Fuso Nerini, F. (2026). AI economics for the common good. *Nature Machine Intelligence*, *8*(4), 495–496. https://doi.org/10.1038/s42256-026-01212-0

gabbiestofthemall. (2023, February 11). *Resources If You're Struggling* [Online forum post]. Reddit. https://www.reddit.com/r/replika/comments/10zuqq6/resources_if_your_struggling/

Garvey, A. M., Kim, T., & Duhachek, A. (2023). Bad news? Send an AI. Good news? Send a human. *Journal of Marketing*, *87*(1), 10–25. https://doi.org/10.1177/00222429211066972

Ghauri, P., Strange, R., & Cooke, F. L. (2021). Research on international business: The new realities. *International Business Review, 30*(2), 1–11. https://doi.org/10.1016/j.ibusrev.2021.101794

Ghoshal, S. (2026, April 7). Gemini update: Google rolls out measures for mental health safety—'Help available', how it works—All you need to know. *Mint*. https://www.proquest.com/newspapers/gemini-update-google-rolls-out-measures-mental/docview/3326839952/se-2?accountid=14797

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin Company.

Gill, T. (2020). Blame it on the self-driving car: How autonomous vehicles can alter consumer morality. *Journal of Consumer Research*, *47*(2), 272–291. https://doi.org/10.1093/jcr/ucaa018

Giroux, M., Kim, J., Lee, J. C., & Park, J. (2022). Artificial intelligence and declined guilt: retailing morality comparison between human and AI. *Journal of Business Ethics*, *178*(4), 1027–1041. https://doi.org/10.1007/s10551-022-05056-7

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Graf-Vlachy, L., Oliver, A. G., Banfield, R., König, A., & Bundy, J. (2020). Media coverage of firms: Background, integration, and directions for future research. *Journal of Management*, *46*(1), 36–69. https://doi.org/10.1177/0149206319864155

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619. https://doi.org/10.1126/science.1134475

Grover, V., & Lyytinen, K. (2015). New state of play in information systems research: The push to the edges. *MIS Quarterly*, *39*(2), 271–296. https://doi.org/10.25300/MISQ/2015/39.2.01

Guthrie, S. E. (1995). *Faces in the clouds: A new theory of religion*. Oxford University Press.

Hao, K., Rodriguez, S., & Seetharaman, D. (2023, June 17). Mark Zuckerberg was early in AI. Now Meta is trying to catch up. *Wall Street Journal (Online)*. https://www.proquest.com/blogs-podcasts-websites/mark-zuckerberg-was-early-ai-now-meta-is-trying/docview/2826542815/se-2?accountid=14797

Hartman, A. E., & Fischer, E. (2026). From Adaptation to Disruption: Structured Ambivalence as a Catalyst for Consumer-Led Institutional Work. *Journal of Consumer Research*. Advance online publication. https://doi.org/10.1093/jcr/ucag004

Hasan, R., & Ojala, A. (2024). Managing artificial intelligence in international business: Toward a research agenda on sustainable production and consumption. *Thunderbird International Business Review*, *66*(2), 151–170. https://doi.org/10.1002/tie.22369

Hasan, R., Ojala, A., Quach, S., Thaichon, P., & Weaven, S. (2025). The dark side of AI anthropomorphism: A case of misplaced trustworthiness in service provisions. *Proceedings of the 58th Hawaii International Conference on System Sciences*, 5695–5704. https://hdl.handle.net/10125/109530

Henfridsson, O., & Bygstad, B. (2013). The Generative mechanisms of digital infrastructure evolution. *MIS Quarterly*, *37*(3), 907–931. https://doi.org/10.25300/MISQ/2013/37.3.11

Horwitz, J. (2025a, April 27). Meta's "digital companions" will talk sex with users—even children. *Wall Street Journal (Online)*. https://www.proquest.com/docview/3195005655?sourcetype=Blogs,%20Podcasts,%20&%20Websites

Horwitz, J. (2025b, August 14). A flirty Meta AI bot invited a retiree to meet. He never made it home. *Reuters*. https://www.reuters.com/investigates/special-report/meta-ai-chatbot-death/

Horwitz, J. (2025c, August 14). Meta's AI rules have let bots hold 'sensual' chats with children. *Reuters*. https://www.reuters.com/investigates/special-report/meta-ai-chatbot-guidelines/

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, *21*(2), 155–172. https://doi.org/10.1177/1094670517752459

Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., ..., & Kivilichan, I. (2024). *GPT-4o System Card*. [Preprint]. arXiv. https://doi.org/arXiv%20preprint%20arXiv:2410.21276

Internet Matters. (2025). *Me, myself and AI chatbot research*. https://www.internetmatters.org/hub/research/me-myself-and-ai-chatbot-research/

Jargon, J. (2026, April 14). AI chats turned romantic, then fatal—analysis of messages with chatbot shows erosion of guardrails. *Wall Street Journal (Online)*. https://www.proquest.com/newspapers/ai-chats-turned-romantic-then-fatal-analysis/docview/3329219800/se-2?accountid=14797

Jargon, J., & Kessler, S. (2025, August 28). A troubled man, his chatbot and a murder-suicide in Old Greenwich. *Wall Street Journal (Online)*. https://www.proquest.com/docview/3244588150?sourcetype=Blogs,%20Podcasts,%20&%20Websites#

Jörling, M., Böhm, R., & Paluch, S. (2019). Service robots: Drivers of perceived responsibility for service outcomes. *Journal of Service Research*, *22*(4), 404–420. https://doi.org/10.1177/1094670519842334

Karahanna, E., Xu, S. X., Xu, Y., & Zhang, N. (2018). The needs–affordances–features perspective for the use of social media. *MIS Quarterly*, *42*(3), 737-A23. https://doi.org/10.25300/MISQ/2018/11492

Kaye, J. E. (2026, June 12). AI chatbot companionship sparks safeguards call. *The European Magazine*. https://the-european.eu/story-61931/nyc-woman-who-held-funeral-for-chatgpt-lover-calls-for-safeguards-over-ai-companionship.html

Kegel, M. M., & Stock-Homburg, R. M. (2023). Customer responses to (im)moral behavior of service robots online experiments in a retail setting. *Proceedings of the 56th Hawaii International Conference on System Sciences*, 1500–1509. https://doi.org/10.24251/HICSS.2023.188

Killeen, R. (2017, June 16). The robot revolution about to hit finance. *Business Post*. https://www.proquest.com/newspapers/robot-revolution-about-hit-finance/docview/1910529848/se-2?accountid=14797

Kim, T., Lee, H., Kim, M., Kim, S., & Duhachek, A. (2023). AI increases unethical consumer behavior due to reduced anticipatory guilt. *Journal of the Academy of Marketing Science*, *51*(4), 785–801. https://doi.org/10.1007/s11747-021-00832-9

Kim, T. W., Jiang, L., Duhachek, A., Lee, H., & Garvey, A. (2022). Do you mind if I ask you a personal question? How AI service agents alter consumer self-disclosure. *Journal of Service Research*, *25*(4), 649–666. https://doi.org/10.1177/10946705221120232

Knof, M., Heinisch, J. S., Kirchhoff, J., Rawal, N., David, K., von Stryk, O., & Stock-Homburg, R. (2022). Implications from responsible human-robot interaction with anthropomorphic service robots for design science. *Proceedings of the 55th Hawaii International Conference on System Sciences,* 5827–5836. https://doi.org/10.24251/HICSS.2022.709

Kopalle, P. K., Gangwar, M., Kaplan, A., Ramachandran, D., Reinartz, W., & Rindfleisch, A. (2022). Examining artificial intelligence (AI) technologies in marketing via a global lens: Current trends and future research opportunities. *International Journal of Research in Marketing*, *39*(2), 522–540. https://doi.org/10.1016/j.ijresmar.2021.11.002

Kunisch, S., Denyer, D., Bartunek, J. M., Menz, M., & Cardinal, L. B. (2023). Review research as scientific inquiry. *Organizational Research Methods*, *26*(1), 3–45. https://doi.org/10.1177/10944281221127292

Lee, C. M. (2026, January 6). A popular Chinese chatbot told a user their coding request was "stupid" and to "get lost." *Business Insider*. https://www.proquest.com/newspapers/popular-chinese-chatbot-told-user-their-coding/docview/3290581222/se-2?accountid=14797

Lee, J. M., Narula, R., & Hillemann, J. (2021). Unraveling asset recombination through the lens of firm-specific advantages: A dynamic capabilities perspective. *Journal of World Business*, *56*(2), 101193. https://doi.org/10.1016/j.jwb.2021.101193

Leong, B., & Selinger, E. (2019). Robot eyes wide shut: Understanding Dishonest anthropomorphism. *Proceedings of the Conference on Fairness, Accountability, and Transparency,* 299–308. https://doi.org/10.1145/3287560.3287591

Li, D., Rau, P. L. P., & Li, Y. (2010). A cross-cultural study: Effect of Robot appearance and task. *International Journal of Social Robotics*, *2*(2), 175–186. https://doi.org/10.1007/s12369-010-0056-9

Li, H. (2025, January 9). Character.AI put in new underage guardrails after a teen's suicide. His mother says that's not enough. *Business Insider*. https://www.proquest.com/newspapers/character-ai-put-new-underage-guardrails-after/docview/3152981964/se-2?accountid=14797

Lim, V., Rooksby, M., & Cross, E. S. (2021). Social robots on a global stage: Establishing a role for culture during human–robot interaction. *International Journal of Social Robotics*, *13*(6), 1307–1333. https://doi.org/10.1007/s12369-020-00710-4

Liu, K., & Tao, D. (2022). The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. *Computers in Human Behavior*, *127, Article 107026*. https://doi.org/10.1016/j.chb.2021.107026

Liu, Y., Wang, X., Du, Y., & Wang, S. (2023). Service robots vs. Human staff: The effect of service agents and service exclusion on unethical consumer behavior. *Journal of Hospitality and Tourism Management*, *55*, 401–415. https://doi.org/10.1016/j.jhtm.2023.05.015

Lyytinen, K., Nickerson, J., & King, J. (2021). Metahuman systems = humans plus machines that learn. *Journal of Information Technology*, *36*(4), 427–445. https://doi.org/10.1177/0268396220915917

Marsden, R. (2025, April 8). How do we really feel about robots? *Financial Times*. https://www.ft.com/content/31d94b0b-b769-4c44-b25d-f39626d2de4a

McGinn, C., Bourke, E., Murtagh, A., Donovan, C., Lynch, P., Cullinan, M., & Kelly, K. (2020). Meet Stevie: A socially assistive robot developed through application of a "design-thinking" approach. *Journal of Intelligent and Robotic Systems*, *98*(1), 39–58. https://doi.org/10.1007/s10846-019-01051-9

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit

compensatory consumer responses. *Journal of Marketing Research*, *56*(4), 535–556. https://doi.org/10.1177/0022243718822827

Meyer, K. E., Li, J., Brouthers, K. D., & Jean, R.-J. "Bryan." (2023). International business in the digital age: Global strategies in a world of national institutions. *Journal of International Business Studies*, *54*(4), 577–598. https://doi.org/10.1057/s41267-023-00618-x

Microsoft. (2026). *Global AI adoption in 2025—A widening digital divide*. AI Economy Institute. https://www.microsoft.com/en-us/corporate-responsibility/topics/ai-economy-institute/reports/global-ai-adoption-2025/

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.

Min, S., Levina, N., & Lifshitz, H. (2024). Technology-centric contestation over symbolic and social boundaries: The social justice implications of Covid-19 contact tracing technologies. *Management Information Systems Quarterly*, *48*(4), 1745–1770. https://doi.org/10.25300/MISQ/2024/18313

Mingers, J. (2004). Real-izing information systems: Critical realism as an underpinning philosophy for information systems. *Information and Organization, Critical Realism*, *14*(2), 87–103. https://doi.org/10.1016/j.infoandorg.2003.06.001

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

National Highway Traffic Safety Administration. (2024). *Additional information regarding EA22002 investigation*. National Highway Traffic Safety Administration. https://static.nhtsa.gov/odi/inv/2022/INCR-EA22002-14496.pdf

Nokelainen, T., & Kanniainen, J. (2018). Coverage bias in business news: Evidence and methodological implications. *Management Research Review*, *41*(4), 487–503. https://doi.org/10.1108/MRR-02-2017-0048

Nourbakhsh, I. R. (2015, August). The coming robot dystopia. *Foreign Affairs*, *94*(4), 23–28. https://www.proquest.com/magazines/coming-robot-dystopia/docview/1691576666/se-2?accountid=14797

Ochmann, J., Michels, L., Tiefenbeck, V., Maier, C., & Laumer, S. (2024). Perceived algorithmic fairness: An empirical study of transparency and anthropomorphism in algorithmic recruiting. *Information Systems Journal, 34*(2), 384–414. https://doi.org/10.1111/isj.12482

Osman, L. (2026, May 14). *Canada is considering an AI chatbot ban for kids*. The Logic Inc. https://www.proquest.com/blogs-podcasts-websites/canada-is-considering-ai-chatbot-ban-kids/docview/3341705770/se-2?accountid=14797

Ostrovsky, N. (2026, January 21). Can you teach an AI to be good? Anthropic thinks so. Time. https://time.com/7354738/claude-constitution-ai-alignment/

Ozturk, A., Cavusgil, S. T., & Ozturk, O. C. (2021). Consumption convergence across countries: Measurement, antecedents, and consequences. *Journal of International Business Studies*, *52*(1), 105–120. https://doi.org/10.1057/s41267-020-00334-w

Parkes, D. C., & Wellman, M. P. (2015). Economic reasoning and artificial intelligence. *Science*, *349*(6245), 267–272. https://doi.org/10.1126/science.aaa8403

Patton, M. Q. (2002). *Qualitative research and evaluation methods (3rd ed.)*. Sage Publications.

Pfeuffer, N., Benlian, A., Gimpel, H., & Hinz, O. (2019). Anthropomorphic information systems. *Business & Information Systems Engineering*, *61*(4), 523–533. https://doi.org/10.1007/s12599-019-00599-y

Pizzi, G., Vannucci, V., Mazzoli, V., & Donvito, R. (2023). I, chatbot! The impact of anthropomorphism and gaze direction on willingness to disclose personal information and behavioral intentions. *Psychology and Marketing*, *40*(7), 1372–1387. https://doi.org/10.1002/mar.21813

Plotkina, D., Orkut, H., & Karageyim, M. A. (2024). Give me a human! How anthropomorphism and robot gender affect trust in financial robo-advisory services. *Asia Pacific Journal of Marketing and Logistics, 36*(10), 2689–270.. https://doi.org/10.1108/APJML-09-2023-0939

Pollina, E., & Coulter, M. (2023, February 3). Italy bans U.S.-based AI chatbot Replika from using personal data. *Reuters*. https://www.reuters.com/technology/italy-bans-us-based-ai-chatbot-replika-using-personal-data-2023-02-03/

Prahalad, C. K., & Doz, Y. L. (1987). *The multinational mission: Balancing local demands and global vision*. Free Press.

Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing*, *85*(1), 131–151. https://doi.org/10.1177/0022242920953847

Purtill, J. (2026, May 16). AI psychosis is on the rise and experts are sounding the alarm. *ABC News*. https://www.abc.net.au/news/2026-05-17/ai-psychosis-is-rising-chatbot-delusion-alternate-reality-harm/106683436

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, *25*(4), 145–182. https://doi.org/10.2753/MIS0742-1222250405

Ragin, C. C. (2009). Reflections on casing and case-oriented research. In D. Byrne & C. C. Ragin (Eds.), *The Sage handbook of case-based methods* (pp. 522–534). SAGE Publications. https://doi.org/10.4135/9781446249413

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. S., … Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y

Reese, B. (Host). (2018). *A Conversation with Ira Cohen* (No. 47) [Audio podcast episode]. In GigaOM Voices in AI. https://www.proquest.com/blogs-podcasts-websites/gigaom-voices-ai-episode-47-conversation-with-ira/docview/2049860650/se-2?accountid=14797

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of "humanizing" customer service chatbots. *Information Systems Research*, *32*(3), 736–751. https://doi.org/10.1287/isre.2021.1015

Schleidgen, S., Friedrich, O., Gerlek, S., Assadi, G., & Seifert, J. (2023). The concept of "interaction" in debates on human–machine interaction. *Humanities and Social Sciences Communications*, *10, Article 551*. https://doi.org/10.1057/s41599-023-02060-8

Schuetz, S., & Venkatesh, V. (2020). Research perspectives: The rise of human machines: how cognitive computing systems challenge assumptions of user-system interaction. *Journal of the Association for Information Systems*, *21*(2), 460–482. https://doi.org/10.17705/1jais.00608

Sharkey, A., & Sharkey, N. (2021). We need to talk about deception in social robotics! *Ethics and Information Technology*, *23*(3), 309–316. https://doi.org/10.1007/s10676-020-09573-9

Siponen, M., Lanamäki, A., Nathan, M., & Klaavuniemi, T. (2025). Mechanism-based explanations and theoretical contribution in IS research. *Communications of the Association for Information Systems*, *57*, 458-475. https://aisel.aisnet.org/cais/vol57/iss1/30

Song, M., Zhu, Y., Xing, X., & Du, J. (2023). The double-edged sword effect of chatbot anthropomorphism on customer acceptance intention: The mediating roles of perceived competence and privacy concerns. *Behaviour and Information Technology*, *43*(15), 3593–3615. https://doi.org/10.1080/0144929X.2023.2285943

Stelmaszak, M., Joshi, M., & Constantiou, I. (2025). Artificial Intelligence as an Organizing Capability Arising from Human-Algorithm Relations. *Journal of Management Studies*. Advance online publication. https://doi.org/10.1111/joms.70003

Tarsney, C. (2025). Deception and manipulation in generative AI. *Philosophical Studies*, *182*(7), 1865–1887. https://doi.org/10.1007/s11098-024-02259-8

Tattersall, H. (2025, August 5). AI tools won't keep your data secret: Lawyers. *The Australian Financial Review (Online)*. https://www.proquest.com/newspapers/ai-tools-wont-keep-your-data-secret-lawyers/docview/3236140129/se-2?accountid=14797

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Ullrich, D., Butz, A., & Diefenbach, S. (2021). The development of overtrust: An Empirical simulation and psychological analysis in the context of human–robot interaction. *Frontiers in Robotics and AI*, *8,* Article 554578. https://www.frontiersin.org/article/10.3389/frobt.2021.554578

van der Goot, M., Koubayová, N., & van Reijmersdal, E. (2024). Understanding users' responses to disclosed vs. undisclosed customer service chatbots: A mixed methods study. *AI & Society, 39,* 2947–2960. https://doi.org/10.1007/s00146-023-01818-7

van Maris, A., Zook, N., Caleb-Solly, P., Studley, M., Winfield, A., & Dogramadzi, S. (2020). Designing ethical social robots—A longitudinal field study with older adults. *Frontiers in Robotics and AI*, *7,* Article 1. https://doi.org/10.3389/frobt.2020.00001

Väyrynen, K., Lanamäki, A., Laari-Salmela, S., Iivari, N., & Kinnula, M. (2025). Unpacking the regulatory ambiguity mechanism: implications for industry-level digital transformation. *Information Systems Journal*, *35*(6), 1528–1564. https://doi.org/10.1111/isj.12595

Vitale, J., Tonkin, M., Herse, S., Ojha, S., Clark, J., Williams, M.-A., Wang, X., & Judge, W. (2018). Be more transparent and users will like you: A robot privacy and user

experience design experiment. *Proceedings of 2018 ACM/IEEE International Conference on Human- Robot Interaction*, 379–387. https://doi.org/10.1145/3171221.3171269

Wang, Z., Huang, J., & Fiammetta, C. (2021). Analysis of gender stereotypes for the design of service robots: Case study on the Chinese catering market. *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, 1336–1344. https://doi.org/10.1145/3461778.3462087

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Welch, C., Paavilainen-Mäntymäki, E., Piekkari, R., & Plakoyiannaki, E. (2022). Reconciling theory and context: How the case study can set a new agenda for international business research. *Journal of International Business Studies*, *53*(1), 4–26. https://doi.org/10.1057/s41267-021-00484-5

Welch, C., & Piekkari, R. (2017). How should we (not) judge the 'quality' of qualitative research? A re-assessment of current evaluative criteria in international business. *Journal of World Business*, *52*(5), 714–725. https://doi.org/10.1016/j.jwb.2017.05.007

Willems, J., Schmid, M. J., Vanderelst, D., Vogel, D., & Ebinger, F. (2023). AI-driven public services and the privacy paradox: Do citizens really care about their privacy? *Public Management Review*, *25*(11), 2116–2134. https://doi.org/10.1080/14719037.2022.2063934

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management*, *29*(5), 907–931. https://doi.org/10.1108/JOSM-04-2018-0119

Wood, M. A. (2021). Rethinking how technologies harm. *The British Journal of Criminology*, *61*(3), 627–647. https://doi.org/10.1093/bjc/azaa074

Wood, M. A., Pervan, F., Lundberg, K., Mitchell, M., & Wood, J. (2025). Doing harm with things: Making sense of intention in harmful actions using technology. *Critical Criminology, 33,* 321–338. https://doi.org/10.1007/s10612-025-09819-2

Xie, L., & Lei, S. (2022). The nonlinear effect of service robot anthropomorphism on customers' usage intention: A privacy calculus perspective. *International Journal of Hospitality Management*, *107,* Article 103312. https://doi.org/10.1016/j.ijhm.2022.103312

Zheng, J., & Jarvenpaa, S. (2021). Thinking technology as human: Affordances, technology features, and egocentric biases in technology anthropomorphism. *Journal of the Association for Information Systems*, *22*(5), 1429–1453. https://doi.org/10.17705/1jais.00698

Złotowski, J., Khalil, A., & Abdallah, S. (2020). One robot doesn't fit all: Aligning social robot appearance and job suitability from a Middle Eastern perspective. *AI & Society*, *35*(2), 485–500. https://doi.org/10.1007/s00146-019-00895-x

Zlotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics*, *7*(3), 347–360. https://doi.org/10.1007/s12369-014-0267-6

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

Appendices

Appendix A. Review of Empirical Research

To construct a dataset for review, I utilized an advanced Boolean search in Web of Science and Scopus databases (Figure A1), and imported them into Zotero referencing software. The dataset collected empirical studies from several sources in early 2024, including peer-reviewed conference proceedings and academic journals published in multiple disciplines. Data was collected on AI anthropomorphism from reference disciplines such as social robotics, computer science, service, human-computer interaction, social psychology, user psychology, management, marketing, and information systems. Based on a criteria-based assessment (see Figure A1), I carefully screened each scholarly record in two rounds. The first-round assessment led to 92 candidate studies. For the second round, I then repeatedly read the full text and settled on 42 studies. Additionally, I added 30 relevant studies through cross-referencing techniques. This combined effort led to 72 studies with or without empirical elements in the final selection pool comprising conceptual, review, technical, and empirical scholarly works. Finally, I selected 44 empirical studies for the analysis that contain only traditional empirical research (n=43) and one technical research source with empirical elements (n=1). The dataset was constructed using Zotero, so empirical insights were extracted using its call number function, with a core focus on the potential negative consequences of AI anthropomorphism in service tasks.

Figure A1. Construction of the dataset for reviewing empirical research

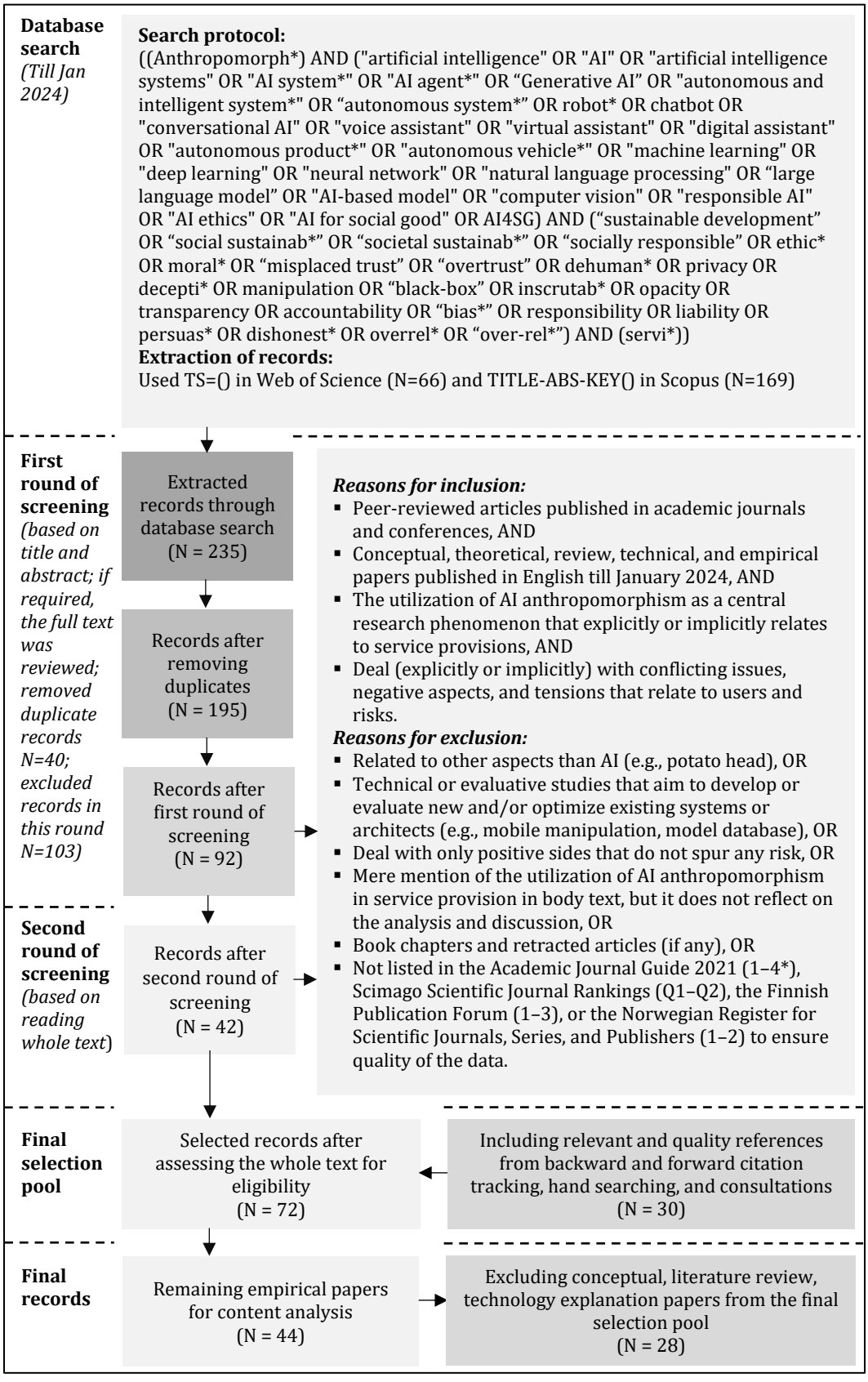


Table A1. Summary of the review of empirical research

Reference	Immediate outcomes	Interplay of users and AISs*		Underlying humanlike features	Contextual conditions [Country**]	Possible adverse outcomes
	Appearance facet					
(Arikan et al., 2023)	Attribute responsibility to AISs	Mind perception	R	Various levels (humanoid) of appearance	Service failure and recovery in restaurants [USA]	Alter responsibility attribution (less to service providers)
(Wang et al., 2021)	Gender norms are inherited in given institutions (culture)	Transferring normative institutions	R	Humanlike appearance at various levels	Catering service [China]	Evoke a gender stereotype
(Plotkina et al., 2024)	A preference for male, realistic-looking AISs	Perceived competence and persuasiveness	V	Humanlike appearance at various levels, including gender	Financial service advisory [USA]	Trigger gender stereotype
(Holthöwer & van Doorn, 2023)	Users' preference for less human-looking AISs	Social judgment due to human social presence	R	Various levels of humanlike appearance	Embarrassing service encounter [Netherlands, UK, and others]	Design shapes users' preferences
(Mende et al., 2019)	Elicit eeriness, discomfort, and human identity threats	Greater sense of social belonging	R	Various levels of humanlike appearance	Eating behavior in a restaurant [USA and unknown]	Alter consumption patterns
(Chang et al., 2023)	Users' preference for mechanically looking AISs	Self-construal (institutions)	R	Various levels of humanlike appearance	Social support benefits at home [China and USA]	Pose relationship intrusion risks
	Interactivity facet					
(Chen et al., 2024)	Alleviate the perception of humanization	Failure to understand and personalize, and lack competence and assurance	V	Emotive expression at various levels	Service failures [China]	Potential false impression
(Mou & Meng, 2023)	Enhance usage resistance	Perceived creepiness, privacy concerns	V	Communication style (master and servant)	Information exchange to manage services [China]	Manipulate perception
(Zhang et al., 2023)	Continuance intention to use	Performance expectancy	V	Information exchange in conversation	Tourism-related booking service [China]	Influence privacy concerns
(van Maris et al., 2020)	Responsible design helped to reduce emotional deception	Emotional attachment	R	Emotive behavior	Elderly care service [presumably UK]	Unintentional deception

(Zhou et al., 2022)	Influence individual morality	Moral judgment		Social-oriented (vs. task) communication style	Charitable giving [China]	Alter moral judgment
(Brendel et al., 2023)	Drive aggressiveness	High impulsivity (personality factor)	V	Humanlike errors	Conversations to complete a task in service systems [Germany]	Increase both frustration and satisfaction
(Schweitzer et al., 2019)	Build relationships	Feeling of control	V	Interaction via speech	Information searching tasks [presumably France or Austria]	Create an illusion of actual emotional capacity
(Schanke et al., 2021)	Increase willingness to disclose personal information	Perceived social presence	V	Humor and communication delays	Disclosure in service transactions [USA]	Lack of disclosure leads to deception
(Nakanishi et al., 2019)	Reduce feelings of loneliness	Interpersonal warmth		Proactive interaction	Service in hotel rooms [Japan]	Alter perceived intrusion and privacy concerns
(van der Goot et al., 2024)	Induce social presence	Willingness to help	V	Willingness to help and the friendly tone of voice	Disclosure in service encounter [Czech Republic, Spain, Austria, Switzerland, and presumably USA]	Make false perceptions of human interaction
(Sun et al., 2023)	Privacy concerns	Trust	V	Social hierarchical cues (partner and servant)	Information sensitivity in service interaction [China]	Alter privacy concerns
(Biswas & Murray, 2017)	Biasness improves companionship	Cognitive biases	R, V	Humanlike cognitive characteristics	Supporting tasks [presumably UK]	Deceptive, as all biases are coded
	Autonomy facet					
(Shank & DeSanti, 2018)	Attribute wrongness	Mind perception	V	Autonomous capability	Moral violations in news service [presumably USA]	Alter moral justification
(Fosch-Villaronga et al., 2020)	Evoke attribution of responsibility	Perceived agency	R	Human social cues	Completing tasks in healthcare service [Italy, Iceland, Norway, Denmark, Sweden, Switzerland, Netherlands, France, Greece, USA, UK, Japan, and Korea]	Shape moral responsibility perception

(Liu & Tao, 2022)	Develop trust	Perceived social presence	V	Autonomous capability	Healthcare service [China]	Shape behavioral intention
(Liu et al., 2023)	Exhibit unethical user behavior	Anticipatory guilt	R	Various levels of humanlike appearance	Service exclusion and inclusion [China]	Alter user morality
(Ullrich et al., 2021)	Develop over trust	Repeated reliable performance	R	Autonomous capability	Pet feeding service [presumably Germany]	Evoke misplaced trust
(Willems et al., 2023)	Trigger privacy paradox	Perceived usefulness	V	Autonomous capability	Information sharing in public service [Austria]	Alter users' privacy concerns
(Giroux et al., 2022)	Change moral intention and behavior	Feeling of guilt	V	Error in information exchange	Retail service experience [presumably USA]	Alter user morality
(Kim et al., 2023)	Increases unethical user behavior	Anticipatory feelings of guilt	V	Emotional capability	Online service encounters [presumably USA]	Alter user morality
(Kim et al., 2022)	Disclose sensitive personal information	Fear of social judgment	V	Emotional capability	Social support and sensitive information sharing [USA]	Influence on information disclosure
(Waytz et al., 2014)	Evoke misplaced trust	Mind perception	V	Autonomous capability with a human name, voice, and gender	Accident in driving service [presumably USA]	Alter blame and responsibility attribution
(Jörling et al., 2019)	Evoke attribution of responsibility	Perceived control	V	Autonomous capability	Responsibility for service outcomes [UK, USA, Ireland, Australia, and New Zealand]	Shape responsibility attributions for outcomes
(Gill, 2020)	Evoke attribution of responsibility	Moral judgments	V	Human intentionality	Driving tasks [USA]	Alter user morality
(Yam et al., 2021)	Increase satisfaction	Mind perception	R	Humanlike agency and experience	Service failures [Japan, China]	Mitigate dissatisfaction with service failures
(Marriott & Pitardi, 2024)	Form a relationship and drive addiction	Societal need	V	Sentience, agreeableness, warmth, and ubiquity	AI service for companion [USA and presumably multicounty]	Deceiving lonely and vulnerable users into believing AISs are sentient
	Across facets					
(Pavone et al., 2023)	Evoke attribution of responsibility	Emotional reactions	V	Visual cues and intentionality	Service failures concerning travel [USA]	Mitigate blame to service providers for adverse outcomes

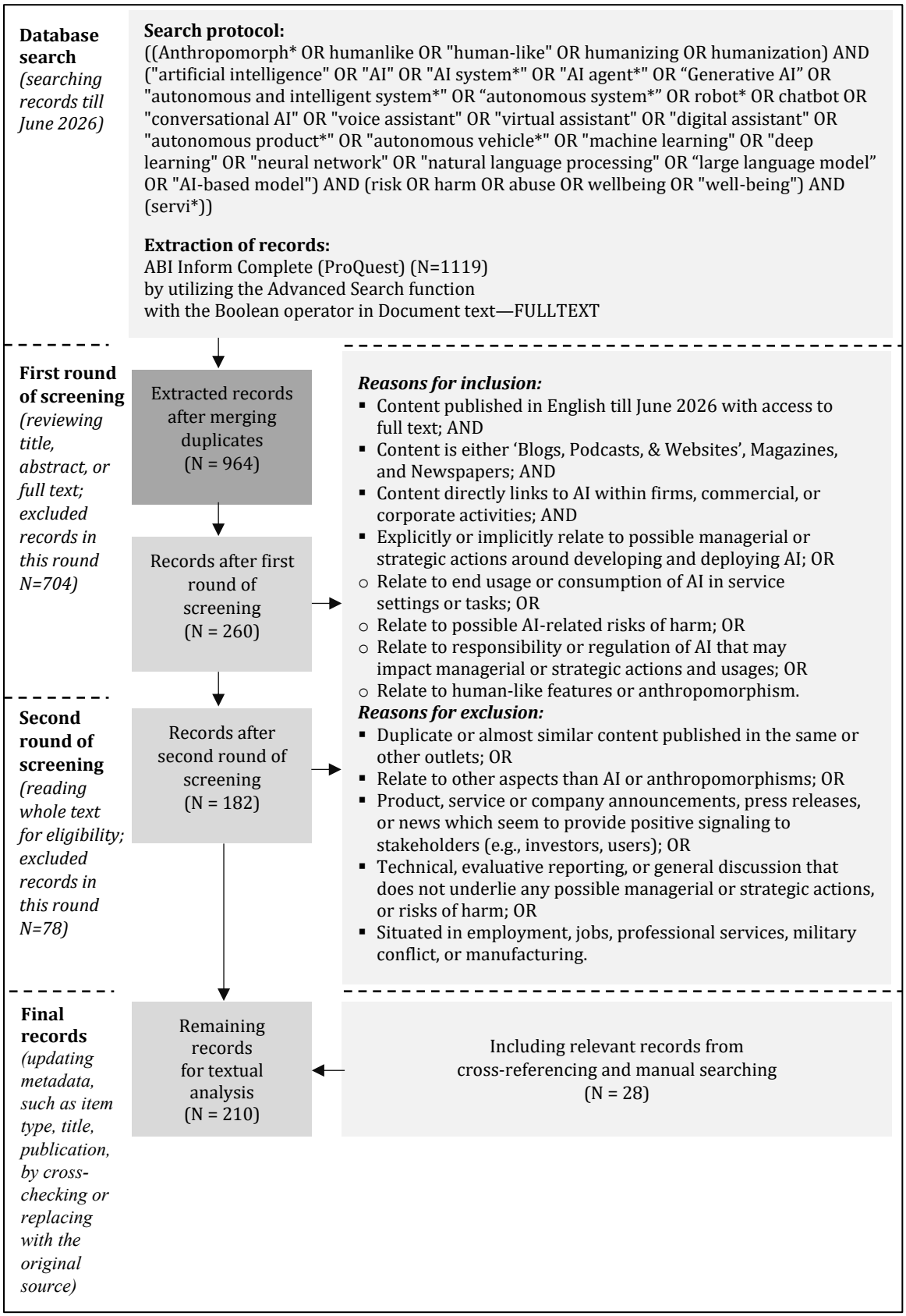
(Benlian et al., 2020)	Mitigate intrusive effects	Societal need	V	Visual human attributes (smiling face)	Intrusive technology features at home [Germany]	Shape feelings of privacy invasion
(Vitale et al., 2018)	Reduce privacy concerns	Embodiment	R	Humanlike gesturing and speech	Face enrollment service systems [Australia]	Alter privacy concerns
(Etemad-Sajadi et al., 2022)	Intention to use	Trust, safety, responsibility	R	Human social cues	Frontline service interaction [Unknown]	Shape the fear of human replacement
(Aw et al., 2024)	Enhance the perceived fairness of algorithmic justice	Mind perception	V	Perceived autonomy	Information processing in investment decisions [Unknown]	Induce a belief in interacting with social beings
(Xie & Lei, 2022)	Enhance privacy concerns	Information sensitivity	R	Various levels of humanlike appearance	Information sharing in a restaurant [China]	Alter users' privacy concerns
(Song et al., 2023)	Increase intention to accept	Perceived competence	V	A combination of social cues in conversations	Time pressure in service encounters [China]	Trigger privacy paradox
(Chuah et al., 2021)	Intention to use	Performance expectancy: personality traits	R	Visual human attributes	Adoption in service [Taiwan]	Tap into different personality traits, and alter privacy risk
(Garvey et al., 2023)	Increase purchase likelihood and satisfaction	Perceived intentionality	R	Various levels of humanlike appearance	Service offer that is worse than expected [presumably USA]	Mitigate blame to service providers for worse offerings
(Kegel & Stock-Homburg, 2023)	Influence trust and ethical concerns	Moral behavior	R	Morality in conversation	Retail service settings [USA]	Moral behavior in conversation
(Pizzi et al., 2023)	Willingness to disclose personal information	Social cognition (warmth and competence)	V	Gender-based human face with gaze direction	Holiday car rental customer service [presumably multicountry]	Altering skepticism that increases trust toward service providers
(McGinn et al., 2020)	Increase acceptance	Perceived intelligence and safety	R	A combination of visual cues and voice	Socially assistive services [USA]	Develop a relationship rapidly

Note(s): *Type of AIs: V represents virtual AIs, and R represents robotic AIs; **Country of users/experts/sample who participated in empirical studies. I presumably added the Country where applicable based on the methodology and authors' affiliations.

Appendix B. Archival Data

The ABI/INFORM Complete (ProQuest) database was utilized to create the dataset using its advanced search option. The search was restricted to English-language records with no publication date limit, and the source type filter was applied to include only: Blogs, Podcasts & Websites, Magazines, and Newspapers. This initial search produced 1119 records and was imported into Zotero, which became 964 after merging duplicates (when an incident or circumstance is reported in the same or different outlets' online and printed versions with similar insights, the records are considered duplicates). All records were extracted and imported into Zotero referencing software for screening. Content published in English with access to full text, including blogs, podcasts, websites, magazines, and newspapers, was considered for inclusion, particularly content that included direct quotations or opinions related explicitly or implicitly to actions within firms, or end usage or consumption in service settings, and to possible risks of harm, and was excluded if otherwise. Screening was carried out in two steps. The first round was performed by reading titles and necessary details for understanding, which reduced the set to 260 records. The second round involved a closer assessment through reading the complete text, which reduced the pool to 182 records. A few relevant records were replaced by cross-referencing techniques to find the referred original source or similar content in more internationally known outlets, maintaining diversity and high quality. In both rounds, screening was based on interrelated inclusion and exclusion criteria (see Figure B1). In addition, further media records, particularly news and podcasts, were added through manual searching, supported by clues taken from selected records, consultation from scholarly conversations, and following regular news updates. This process added 28 relevant records. The combined effort resulted in a final dataset of 210 media records consisting of five categories: newspaper articles N=122, web-based news sources N=14, magazine articles N=47, blog posts N=20, and transcripts of podcasts N=7.

Figure B1. Construction of the archival dataset



REFERENCES

Arikan, E., Altinigne, N., Kuzgun, E., & Okan, M. (2023). May robots be held responsible for service failure and recovery? The role of robot service provider agents' human-likeness. *Journal of Retailing and Consumer Services*, *70,* Article 103175. https://doi.org/10.1016/j.jretconser.2022.103175

Aw, E. C.-X., Leong, L.-Y., Hew, J.-J., Rana, N. P., Tan, T. M., & Jee, T.-W. (2024). Counteracting dark sides of robo-advisors: Justice, privacy and intrusion considerations. *International Journal of Bank Marketing*, *42*(1), 133–151. https://doi.org/10.1108/IJBM-10-2022-0439

Benlian, A., Klumpe, J., & Hinz, O. (2020). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal*, *30*(6), 1010–1042. https://doi.org/10.1111/isj.12243

Biswas, M., & Murray, J. (2017). The effects of cognitive biases and imperfectness in long-term robot-human interactions: Case studies using five cognitive biases on three robots. *Cognitive Systems Research*, *43*, 266–290. https://doi.org/10.1016/j.cogsys.2016.07.007

Brendel, A., Hildebrandt, F., Dennis, A., & Riquel, J. (2023). The paradoxical role of humanness in aggression toward conversational agents. *Journal of Management Information Systems*, *40*(3), 883–913. https://doi.org/10.1080/07421222.2023.2229127

Chang, Y., Gao, Y., Zhu, D., & Safeer, A. (2023). Social robots: Partner or intruder in the home? The roles of self-construal, social support, and relationship intrusion in consumer preference. *Technological Forecasting and Social Change*, *197,* Article 122914. https://doi.org/10.1016/j.techfore.2023.122914

Chen, Q., Gong, Y., Lu, Y., & Luo, X. (Robert). (2024). The golden zone of AI's emotional expression in frontline chatbot service failures. *Internet Research*, *35*(3), 1065-1103. https://doi.org/10.1108/INTR-07-2023-0551

Chuah, S. H.-W., Aw, E. C.-X., & Yee, D. (2021). Unveiling the complexity of consumers' intention to use service robots: An fsQCA approach. *Computers in Human Behavior*, *123,* Article 106870. https://doi.org/10.1016/j.chb.2021.106870

Etemad-Sajadi, R., Soussan, A., & Schöpfer, T. (2022). How ethical issues raised by human–robot interaction can impact the intention to use the robot? *International Journal of Social Robotics*, *14*(4), 1103–1115. https://doi.org/10.1007/s12369-021-00857-8

Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Gathering expert opinions for social robots' ethical, legal, and societal concerns: Findings from four international workshops. *International Journal of Social Robotics*, *12*(2), 441–458. https://doi.org/10.1007/s12369-019-00605-z

Garvey, A. M., Kim, T., & Duhachek, A. (2023). Bad news? Send an AI. Good news? Send a human. *Journal of Marketing*, *87*(1), 10–25. https://doi.org/10.1177/00222429211066972

Gill, T. (2020). Blame it on the self-driving car: How autonomous vehicles can alter consumer morality. *Journal of Consumer Research*, *47*(2), 272–291. https://doi.org/10.1093/jcr/ucaa018

Giroux, M., Kim, J., Lee, J. C., & Park, J. (2022). Artificial intelligence and declined guilt: Retailing morality comparison between human and AI. *Journal of Business Ethics*, *178*(4), 1027–1041. https://doi.org/10.1007/s10551-022-05056-7

Holthöwer, J., & van Doorn, J. (2023). Robots do not judge: Service robots can alleviate embarrassment in service encounters. *Journal of the Academy of Marketing Science*, *51*(4), 767–784. https://doi.org/10.1007/s11747-022-00862-x

Jörling, M., Böhm, R., & Paluch, S. (2019). Service robots: Drivers of perceived responsibility for service outcomes. *Journal of Service Research*, *22*(4), 404–420. https://doi.org/10.1177/1094670519842334

Kegel, M. M., & Stock-Homburg, R. M. (2023). Customer responses to (im)moral behavior of service robots online experiments in a retail setting. *Proceedings of the 56th Annual Hawaii International Conference on System Sciences, 1500-1509*. https://doi.org/10.24251/HICSS.2023.188

Kim, T., Lee, H., Kim, M., Kim, S., & Duhachek, A. (2023). AI increases unethical consumer behavior due to reduced anticipatory guilt. *Journal of the Academy of Marketing Science*, *51*(4), 785–801. https://doi.org/10.1007/s11747-021-00832-9

Kim, T. W., Jiang, L., Duhachek, A., Lee, H., & Garvey, A. (2022). Do you mind if i ask you a personal question? How ai service agents alter consumer self-disclosure. *Journal of Service Research*, *25*(4), 649–666. https://doi.org/10.1177/10946705221120232

Liu, K., & Tao, D. (2022). The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. *Computers in Human Behavior*, *127,* Article 107026. https://doi.org/10.1016/j.chb.2021.107026

Liu, Y., Wang, X., Du, Y., & Wang, S. (2023). Service robots vs. human staff: The effect of service agents and service exclusion on unethical consumer behavior. *Journal of Hospitality and Tourism Management*, *55*, 401–415. https://doi.org/10.1016/j.jhtm.2023.05.015

Marriott, H. R., & Pitardi, V. (2024). One is the loneliest number… Two can be as bad as one. The influence of AI Friendship Apps on users' well-being and addiction. *Psychology & Marketing*, *41*(1), 86–101. https://doi.org/10.1002/mar.21899

McGinn, C., Bourke, E., Murtagh, A., Donovan, C., Lynch, P., Cullinan, M., & Kelly, K. (2020). Meet Stevie: A socially assistive robot developed through application of a "design-thinking" approach. *Journal of Intelligent and Robotic Systems*, *98*(1), 39–58. https://doi.org/10.1007/s10846-019-01051-9

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, *56*(4), 535–556. https://doi.org/10.1177/0022243718822827

Mou, Y., & Meng, X. (2023). Alexa, it is creeping over me – Exploring the impact of privacy concerns on consumer resistance to intelligent voice assistants. *Asia Pacific Journal of Marketing and Logistics*, *36*(2), 261–292. https://doi.org/10.1108/APJML-10-2022-0869

Nakanishi, J., Kuramoto, I., Baba, J., Ogawa, K., Yoshikawa, Y., & Ishiguro, H. (2019). Soliloquising social robot in a hotel room. *ACM International Conference Proceeding Series*, 21–29. https://doi.org/10.1145/3369457.3370913

Pavone, G., Meyer-Waarden, L., & Munzel, A. (2023). Rage against the machine: Experimental insights into customers' negative emotional responses, attributions of responsibility, and coping strategies in artificial intelligence–based service failures. *Journal of Interactive Marketing*, *58*(1), 52–71. https://doi.org/10.1177/10949968221134492

Pizzi, G., Vannucci, V., Mazzoli, V., & Donvito, R. (2023). I, chatbot! The impact of anthropomorphism and gaze direction on willingness to disclose personal information and behavioral intentions. *Psychology and Marketing*, *40*(7), 1372–1387. https://doi.org/10.1002/mar.21813

Plotkina, D., Orkut, H., & Karageyim, M. A. (2024). Give me a human! How anthropomorphism and robot gender affect trust in financial robo-advisory services. *Asia Pacific Journal of Marketing and Logistics, 36* (10), 2689–2705. https://doi.org/10.1108/APJML-09-2023-0939

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of "humanizing" customer service chatbots. *Information Systems Research*, *32*(3), 736–751. https://doi.org/10.1287/isre.2021.1015

Schweitzer, F., Belk, R., Jordan, W., & Ortner, M. (2019). Servant, friend or master? The relationships users build with voice-controlled smart devices. *Journal of Marketing Management*, *35*(7–8), 693–715. https://doi.org/10.1080/0267257X.2019.1596970

Shank, D. B., & DeSanti, A. (2018). Attributions of morality and mind to artificial intelligence after real-world moral violations. *Computers in Human Behavior*, *86*, 401–411. https://doi.org/10.1016/j.chb.2018.05.014

Song, M., Zhu, Y., Xing, X., & Du, J. (2023). The double-edged sword effect of chatbot anthropomorphism on customer acceptance intention: The mediating roles of perceived competence and privacy concerns. *Behaviour and Information Technology*, *43*(15), 3593–3615. https://doi.org/10.1080/0144929X.2023.2285943

Sun, Z., Zang, G., Wang, Z., Zhao, H., & Liu, W. (2023). VCAs as partners or servants? The effects of information sensitivity and anthropomorphism roles on privacy concerns. *Technological Forecasting and Social Change*, *192,* Article 122560. https://doi.org/10.1016/j.techfore.2023.122560

Ullrich, D., Butz, A., & Diefenbach, S. (2021). The development of overtrust: An Empirical simulation and psychological analysis in the context of human–robot interaction. *Frontiers in Robotics and AI*, *8,* Article 554578. https://www.frontiersin.org/article/10.3389/frobt.2021.554578

van der Goot, M., Koubayová, N., & van Reijmersdal, E. (2024). Understanding users' responses to disclosed vs. undisclosed customer service chatbots: A mixed methods study. *AI & Society, 39*(6)*, 2947-2960*. https://doi.org/10.1007/s00146-023-01818-7

van Maris, A., Zook, N., Caleb-Solly, P., Studley, M., Winfield, A., & Dogramadzi, S. (2020). Designing ethical social robots—A longitudinal field study with older adults. *Frontiers in Robotics and AI*, *7,* Article 1. https://doi.org/10.3389/frobt.2020.00001

Vitale, J., Tonkin, M., Herse, S., Ojha, S., Clark, J., Williams, M.-A., Wang, X., & Judge, W. (2018). Be more transparent and users will like you: A robot privacy and user experience design experiment. *Proceedings of 2018 ACM/IEEE International Conference on Human- Robot Interaction*, 379–387. https://doi.org/10.1145/3171221.3171269

Wang, Z., Huang, J., & Fiammetta, C. (2021). Analysis of gender stereotypes for the design of service robots: Case study on the Chinese catering market. *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, 1336–1344. https://doi.org/10.1145/3461778.3462087

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Willems, J., Schmid, M. J., Vanderelst, D., Vogel, D., & Ebinger, F. (2023). AI-driven public services and the privacy paradox: Do citizens really care about their privacy? *Public Management Review*, *25*(11), 2116–2134. https://doi.org/10.1080/14719037.2022.2063934

Xie, L., & Lei, S. (2022). The nonlinear effect of service robot anthropomorphism on customers' usage intention: A privacy calculus perspective. *International Journal of Hospitality Management*, *107,* Article 103312. https://doi.org/10.1016/j.ijhm.2022.103312

Yam, K., Bigman, Y., Tang, P., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2021). Robots at work: People prefer-and forgive-service robots with perceived feelings. *Journal of Applied Psychology*, *106*(10), 1557–1572. https://doi.org/10.1037/apl0000834

Zhang, B., Zhu, Y., Deng, J., Zheng, W., Liu, Y., Wang, C., & Zeng, R. (2023). "I am here to assist your tourism": Predicting continuance intention to use AI-based chatbots for tourism. Does gender really matter? *International Journal of Human-Computer Interaction*, *39*(9), 1887–1903. https://doi.org/10.1080/10447318.2022.2124345

Zhou, Y., Fei, Z., He, Y., & Yang, Z. (2022). How human–chatbot interaction impairs charitable giving: The role of moral judgment. *Journal of Business Ethics*, *178*(3), 849–865. https://doi.org/10.1007/s10551-022-05045-w