e-Informatica Software Engineering Journal, Volume 17, Issue 1, 2023, pages: 230101, DOI: 10.37190/e-Inf230101

Governance in Ethical and Trustworthy AI Systems: Extension of the ECCOLA Method for AI Ethics Governance Using GARP

Mamia Agbese*, Hanna-Kaisa Alanen*, Jani Antikainen*, Erika Halme*, Hannakaisa Isomäki*, Marianna Jantunen*, Kai-Kristian Kemell*, Rebekah Rousi*, Heidi Vainio-Pekka*, Ville Vakkuri*

*Faculty of Information Technology, The University of Jvaskyla, Finland

mamia.o.agbese@jyu.fi, hanna-kaisa.h-k.alanen@student.jyu.fi, jani.p.antikainen@student.jyu.fi, erika.a.halme@jyu.fi, hannakaisa.isomaki@jyu, marianna.s.p.jantunen@jyu.fi, kai-kristian.o.kemell@jyu.fi, rebekah.rousi@uwasa.fi, heidi.vainiopekka@gmail.com, ville.vakkuri@jyu.fi

Abstract

Background: The continuous development of artificial intelligence (AI) and increasing rate of adoption by software startups calls for governance measures to be implemented at the design and development stages to help mitigate AI governance concerns. Most AI ethical design and development tools mainly rely on AI ethics principles as the primary governance and regulatory instrument for developing ethical AI that inform AI governance. However, AI ethics principles have been identified as insufficient for AI governance due to lack of information robustness, requiring the need for additional governance measures. Adaptive governance has been proposed to combine established governance practices with AI ethics principles for improved information and subsequent AI governance. Our study explores adaptive governance as a means to improve information robustness of AI ethical design and development tools. We combine information governance practices with AI ethics principles using ECCOLA, a tool for ethical AI software development at the early developmental stages.

Aim: How can ECCOLA improve its robustness by adapting it with GARP® IG practices?

Methods: We use ECCOLA as a case study and critically analyze its AI ethics principles with information governance practices of the Generally Accepted Recordkeeping principles (GARP®).

Results: We found that ECCOLA's robustness can be improved by adapting it with Information governance practices of retention and disposal.

Conclusions: We propose an extension of ECCOLA by a new governance theme and card, # 21.

Keywords: AI, AI Ethics, Trustworthy AI, AI Governance, Adaptive Governance, ECCOLA

Submitted: 30 Mar. 2022; Revised: 28 Jul. 2022; Accepted: 18 Sep. 2022; Available online: 28 Sep. 2022

1. Introduction

The continuous progress of artificial intelligence (AI) necessitates that AI ethical development tools and method models used in implementing ethics in the engineering of AI improve their robustness in informing AI governance practices [1]. As AI becomes one of the preferred emerging technologies for software startups [2], which utilize its application in diverse critical sectors such as transport, health, retail, and recently in warfare [3, 4], increased calls for effective AI governance practices are on the rise [5]. Current governance practices implemented in AI engineering lack robustness and often lead to inefficiency in fostering AI governance or failed AI governance practices [5]. The incident involving an autonomous Uber vehicle resulting in a pedestrian's loss of life and the associated failure in identifying the source of accountability [6] due to insufficient governance information represents a growing AI governance failure [7]. Inefficiencies of governance measures in the design and development of the AI posed a legal challenge in clearly delineating responsibility or governance between the malfunctioned AI autonomous driving system and the distracted driver [6]. Consequently, raising questions about the efficacy of current AI governance measures [5].

Governance issues in AI or AI governance concerns often arise when humans interact with AI during usage with no clear delineation of responsibilities associated with roles and the associated impact on humanity and society [8]. The current predominant approach to AI governance is the guideline approach based on AI ethical principles [1]. The approach involves AI principles such as the European Union (EU) Ethics Guidelines for Trustworthy AI [9] used as "soft laws" in the design, development, and deployment of AI to facilitate AI governance [1]. It is also the foundational approach used by AI ethical design and development tools such as method cards, model cards, and ethics canvas to implement AI ethics and subsequently AI governance practices in the development of ethical AI [10, 11]. However, Eitel-Porter [12] explains that the governance practices or ethical principles identified in these principles are insufficient for creating ethical AI that can foster effective AI governance. The guideline approach generally provides a foundation for AI governance but is inadequate due to a lack of information robustness in AI ethics principles [1, 7, 12]. Hamon et al. and Taeihagh [1, 13] corroborate by explaining that information robustness in AI ethical development tools for governance requires a solid information base due to constant changes in the AI terrain. Private information technology organizations that experiment more with AI are often ahead of guidelines and principles in terms of information [1]. Suggesting that insufficient information robustness in ethical AI ethical development tools and method models can lead to duplicated efforts with little practical benefits slowing the pace of research [13]. Overall, Taeihagh [1] stresses the urgency for reassessing current traditional approaches to AI governance to determine their potency as the constant speed of change in information threatens to outpace current AI governance measures [1]. Therefore, more robust governance measures are necessary to manage processes and create associated audits for principles enforcement [1, 13] to help mitigate the increasing inefficiency of AI ethics principles [1, 7].

The adaptive governance and hybrid approach (Adaptive governance) is one of the emerging approaches being explored to improve the robustness of AI governance practices [1, 14, 15]. The adaptive governance approach recommends that successful governance practices be emulated and adapted to current AI governance initiatives to improve AI governance overall [1]. This paper explains that successfully utilized governance practices or frameworks used in regulating previous technologies can be adapted to govern or complement

existing principles or governance approaches for new and emerging technologies such as AI [1]. In this way, lessons learned and successful practices can inform or be incorporated into existing practices to help build robust structures that increase governance capacity [16]. This approach also extends to AI ethical design and development tools to aid the implementation of robust governance practices in the developmental phase that can help mitigate AI governance risks [1, 17]. Currently, most approaches to adaptive governance for ethical AI are explored at policy and regulatory levels [1, 14, 16, 18], with scarce empirical assessment of the approach for AI ethical developmental tools and at the design and development stages [19]. Consequently, virtually no practices have been identified in literature.

Hamon [13] explains that AI ethical development tools can attain a robust or resilient information base by being subjected to rigorous evaluations to benchmark areas that have not been considered or fully exploited. This serves as a motivation for us to explore adaptive governance in AI ethical development tools. In a previous study [20], we analyzed an AI ethical developmental card tool, ECCOLA [10], to identify areas of governance vulnerabilities. ECCOLA is a tool used in the ethically aligned design of trustworthy AI [10] and developed using AI ethics principles (guideline approach). One of the findings from the study revealed a vulnerability in the form of Information governance (IG) practices. We, therefore, leverage the study and extend our research to examine how to improve ECCOLA's robustness by improving its IG vulnerability. We aim to critically analyze ECCOLA's ethical practices with IG practices of the Generally Acceptable Recordkeeping principles® (GARP®) to identify key areas where its information robustness can be improved. By so doing, we can extend the tool for improved governance practices and AI governance by adapting it with IG practices from the GARP® principles. We frame our research question as:

RQ: How can ECCOLA improve its robustness by adapting it with GARP® IG practices?

Our work can help similar AI ethical development tools leverage and optimize the approach to improve their information robustness and add to the scarce body of research on adaptive governance in AI ethical developmental tools.

The rest of the paper is structured as follows: Section 2 focuses on the background and related work on the concepts of AI and its associated technologies, AI ethics, AI governance and the governance frameworks employed in the study. Section 3 describes the methodology employed in this study. Section 4 provides the results, and Section 5 elaborates the findings and discusses them. In Section 6, we provide a conclusion for the research and avenues for further works.

2. Background and related work

2.1. Artificial Intelligence

There is no general or standard definition for AI as it constantly evolves. Collins et al. [21] explains that this is not a challenge as most scientific concepts are truly defined upon maturity. They identify a prevalent definition for AI as systems that mimic reasoning functions associated with human characteristics like learning speech and problem-solving [21]. AI is considered relatively new even though its origin can be traced back to the 1950s with Alan Turing, the Turing machine and Turing's research on making computers

intelligent and capable of replicating the human brain [22]. Paper [23] describes it as various technologies that produce computation associated with human intelligence. AI can be represented as software and possibly hardware systems that act in digital or physical dimensions to perceive their environment, collect data, process it into information, and learn to decide the best course of action to complex goals [24]. AI has evolved over the years to include expert systems (ES) which mirror human intelligence by using knowledge-based applications and inference procedures [25]. ML systems use data to learn from experience with respect to some class of tasks and performance measures [26]. ES includes computer programs or models capable of solving problems in a specific area of knowledge with as much expertise as a human expert [27]. They also automate tasks carried out by human specialists in a particular problem domain. Over time, the growth of the internet alongside computing technology gave rise to the influx of big data, which has evolved AI in a different dimension requiring a ML approach for handling or processing Big data.

ML are computer programs that learn from experience with respect to some class of tasks and performance measures [26]. They employ algorithms in generating numeric models to compute data decisions and are more effective than traditional quantitative methods. ML technology existed separately to AI but has fast become the central paradigm and a sub-field of AI [26]. ML can process structured and unstructured data in real-time to provide accurate predictions and efficiently model predictive data analytic applications and computing tasks where algorithm design is difficult or nearly impossible [26]. ML has evolved over two crucial phases, shallow learning and deep learning (DL) [28]. The shallow learning phase established during the 1990s is characterized by shallow models featuring single or no hidden layers. This phase has seen success in many applications such as web search sorting systems, spam partial filtering systems, and recommendation systems [28]. The DL phase emerged in the 2000s and stems from the study of Artificial Neural Networks (ANN) and how it imitates the human brain's neural structure [28]. The approach follows how the human brain processes information by establishing a simple model, forming different multi-layer learning models with multiple hidden layers and an extensive training data set or sets. Bengio in [29] explains that automatically learning features at multiple extraction levels enables DL systems to learn complex functions and improves their capability to map input functions to output directly from data instead of depending entirely on crafted human features. DL has become well-suited for complex unconstrained problems such as speech and image recognition, and its versatility is harnessed in autonomous and semi-autonomous AI systems [30].

AI is classified into three categories, narrow, general and super intelligence [31]. The narrow stage or phase which is the current state of AI is classified as narrow intelligence. Narrow intelligence is characterised by human-level intelligence (text, speech, and sound = data) to produce outputs such as voice and text recognition capabilities [31]. The other two stages of AI, generalised intelligence and super intelligence which is outside the scope of this study suggests stages where AI systems denote strong human intelligence and above human intelligence respectively. This however, is yet to be achieved [31]. Due to progress made so far in AI, it is increasingly implemented in application areas such as automation of workflow processes, improved service quality, and faster information processing. However, larger data sets used in DL tend to make the network topology complex and challenging to interpret in the design of AI and AI enabled systems. This is one of the leading ethical concerns of AI [30, 32].

2.2. AI ethics

Ethics to begin with is an extremely broad field, crossing diverse disciplines, while possessing roots in normative ethics, which examines what makes actions right or wrong [33]. It involves guidelines or sets of rules and principles to help determine what is good or right and the moral obligations and responsibilities of the entities involved [33]. These entities can range from humans to artificial agents. As artificial agents such as AI become more advanced and their interactions with human agents less defined, concerns have arisen regarding their design and development and the ethical impact of their actions. Some include ethical or moral judgment and decision-making of AI systems, cybersecurity, threats from AI, and goal alignment between humans and AI [8] fuelling the need for ethics to be applied to AI. AI ethics can be described as the field associated with studying ethical issues in AI [33]. The research on AI ethics continues to grow. Michael et al. [34] discuss using ML to design utility systems such as biometrics and how associated biases can be alleviated. Article [35] analyze AI using a contrarian approach on how its dark side, i.e., aspects of AI discussed as negatives, can help influence the development of ethical AI. Their work helps to draw insight into the challenging areas of AI and how these negative aspects can help inform the work on ethical or responsible AI. Trocin et al. [36] helps provide insight into responsible or ethical AI development. They analyze core AI and ML issues from the literature mainly concerned with establishing ethical principles and human values to reduce biases and promote fairness [36]. While their work primarily addresses the healthcare sector, the implications of their findings are relevant to mainstream AI ethics literature.

AI ethics also covers the ethical design and development of AI and the ethical issues that result from AI's interaction with humans [33]. While some have attributed the root of ethical AI issues to design and development in the form of the black-box nature of AI systems [37] others attribute it to their interaction with humans [38]. With all these arguments, the general consensus is for appropriate ethical regulatory measures that can enhance governance practices to be implemented in AI to help address these concerns. One of the main responses to these demands has come in the form of principles or guidelines to act as "soft laws" in regulating AI [1]. Currently, over 80 AI ethics guidelines exist. These stem from different governments, international bodies, private sectors, and the research community [39]. The guidelines or principles center on implementing ethical practices in AI to help mitigate the associated risks. While these principles have helped to identify what is needed to develop ethical AI, they are yet to provide practical solutions on how this will be achieved [40]. As a result, most of the work on AI ethics principles as guidelines focus on frameworks and checklists and not enough on how the guidelines can be implemented as groundwork for governance and innovation [41]. Most sets of principles possess strong similarities to one another, and thus converge in many ways. However, the inconsistencies in how they are interpreted or defined have made it challenging to assign accountability to actors and created room for "digital ethical shopping" [41]. Increased incidence of AI-related accidents continues to occur without appropriate accountability, necessitating that tangible action is needed to move AI ethics from high-level abstractions and arguments to the creation of practical accountability mechanisms [41].

2.2.1. AI governance

In AI ethics, governance has emerged as instrumental in addressing accountability issues where an AI mishap could have monumental consequences [5]. Governance designates how

actions are designed, maintained, regulated and usually represents how accountability is assigned [42, 43]. Governance is a broad concept and rooted in different issues that imply diverse meanings, but generally include processes that facilitate regulation [42, 43]. AI governance or incorporating governance in AI is considered complex and requires several approaches [32]. AI incorporates various technologies such as ML, which is considered to be unpredictable, complex, and random; also, various actors are involved in AI, leading to several conceptualizations of AI governance without an overarching definition [32]. Ashok et al. [44] explain AI governance or the principle of governance in AI as creating and implementing policies procedures and standards for the appropriate development, use, and management of the infosphere, implying that the governance of AI pertains to both the cyberspace and the analog world.

For the implementation of ethics in AI, Sondergaard in [42] explains that AI governance should be explainable, transparent, and ethical, although the meaning of each term is contextual. For example, the term transparency used within technical or legal contexts implies different meanings. Transparency in technology might indicate software or code transparency and in legal context may imply transparency of policies [42]. Therefore, AI governance involves considering several factors and components to aid the governing process. It should also encompass how humanity attempts to navigate the transition of AI as it evolves across different touchpoints [45, 46]. AI governance is considered in research as a layered structure overall which embraces a flexible approach to accommodate the various layers involved [45]. This paper explains that each layer deals with the different ethical and socio-ethical issues associated with AI. For example, the technical layer of AI governance can deal with technical components like data, information, and information security; the political layer can deal with political elements such as regulations, standardization, and the responsible actors [45]. The field of AI governance is still in its infancy [32]. Its current stage is argued as being disorganized, involving different stakeholders vying for their own best interests [47]. Most of the research is centered on different frameworks to aid AI governance at different layers [47], with each tailored to suit a particular context without a clear consensus or framework of how on implementation [5, 45, 46]. This general lack of agreement could be attributed to the tension involved in unifying the different components of AI governance.

One of the dominant and popular approaches to AI governance involves using the principles approach or ethical principles as guidelines for regulation [48–51]. The reasoning is that ethical guidelines can serve as a foundation for regulating AI, encompassing its development, deployment, and usage (its life cycle) [52]. The principles approach encourages the use of guidelines to serve as an ethical risk assessment to help reduce the impact of ethical exposure [46]. However, the principles approach is criticized as ineffective [7, 53]. One contention is that the principles approach lacks methods and practices to translate principles to practice with robust legal and accountability mechanisms to actualize the principles approach lacking [53]. In addition, guidelines may not always be adhered to as they may lack uniformity in the enforcement or be influenced by stakeholders [47]. Further research lends credence to this school of thought by arguing that the principles approach is expensive in terms of costs and processes required in implementation [7].

As a result, there have been arguments for different approaches to AI governance. Some of these approaches include the independent audit governance approach, governance of AI systems in their design, and Adaptive governance [1, 7, 54, 55]. Governance by independent audit calls for an AAA audit style approach to governance [7]. The AAA AI governance approach involves a prospective risk **A**ssessment which entails assessing AI systems before

implementation, and **A**udit trail for failure analysis and accountability; and a system **A**dherence to jurisdictional requirements [7]. Governance of AI systems in their design advocates for governance measures to be applied to AI systems in their design phases to aid accountability and audit in the systems as they evolve [56]. However, the issue of translating guidelines to codes in design is proving to be one of the challenges of enforcing this approach, as ethical guidelines are challenging to translate into codes [1].

The Adaptive Governance approach is the most advocated approach to emerge from the discourse for better AI governance [1]. The adaptive governance approach advocates the co-regulation of governing practices by relevant stakeholders [1]. X. Cao et al. in [57] explains that the concept of adaptive governance goes far back to the 1960s to Arnold Kaufman, who proposed the idea of "participatory democracy." Participatory in the sense that various stakeholders are involved in governance or management processes. Adaptive governance is analyzed as a flexible approach to AI governance [45], where new information is gathered from reiterative adjustment, and guidelines enhancement from successful frameworks [1]. Adaptive governance is also referred to or likened to co-regulation, or hybrid governance approach [15]. With the similarity explained as when AI governance practices are adapted with successful existing practices, a new form of governance emerges, a hybrid of the two [58]. A hybrid approach of governance can help provide flexibility in positioning governance guardrails that proactively identify foreseeable risks emerging from AI as it evolves [1]. Ayling and Chapman [59] corroborates this by explaining that long established techniques exist that help lay down governance practices and having such practices incorporated in the governance of AI can aid AI governance.

Research carried out by [57] identified the adaptive governance approach in a survey as having a significant impact on responsible innovation. Also, S.Y. Tan et al. [16] proposed an adaptive governance approach as a possible solution for policy regulation of autonomous vehicles in Singapore. However, clear implementation of this approach is yet to emerge in research, particularly at the development and design level of ethical AI systems. Consequently, virtually no practices for adaptive governance as it pertains to AI ethical development tools like ECCOLA have been identified in literature.

2.3. Frameworks

The frameworks used in this study are ECCOLA and the GARP®. ECCOLA represents the principles approach to AI governance and GARP® represents a successful governance frame work.

2.3.1. ECCOLA

The ECCOLA method is an actionable tool to aid the design and development of trustworthy and ethical AI systems presented in a previous study [10]. ECCOLA is considered a low-threshold framework that assists practitioners in their ethically aligned design of AI systems and forms part of the software development process. It also serves as an actionable tool that facilitates AI ethics in a method agnostic manner for AI developers. The method is card-based and comprises 21 cards split into eight ethical themes. The themes are built on ethical principles incorporated from major ethical guidelines, including the IEEE EAD and the EU Trustworthy AI guidelines [10]. The eight themes are analyze, Transparency, Data, Agency and Oversight, Safety and Security, Fairness, Well-being, and Accountability. Each theme comprises one to six cards, and each card provides a detailed approach to the

theme it represents. ECCOLA works by asking AI developers questions to consider and weigh the various ethical issues present in the development of ethical AI systems. It is also method agnostic and can aid software developers, product managers, and consultants with practices and tools for implementing ethically aligned designs in developing AI systems. The questions raised in the cards are based on ethical principles. These are pertinent from the conception stage where ideas are conceived, the mental stage where thinking tools, practices, and principles are analyzed, and the operational stage of the development process. By doing so, ECCOLA embodies the principles approach as the ethical principles on which the method is founded [5], also serves as a governance base in the development process. Table 1 gives a breakdown of the ECCOLA card themes and how they are broken down.

Table 1. ECCOLA card themes

Card themes (8)	Card number (0–20)	Card amount (total 21)
Analze	0	1
Transparency	#1–6	6
Safety and Security	#7–9	3
Fairness	#10–11	2
Data	#12–13	2
Agency and Oversight	#14–15	2
Wellbeing	#16–17	2
Accountability	#18–20	3

2.3.2. GARP® governance framework

Governance frameworks can be described as governance structures that mirror interconnected relationships, factors, and influences within an institution [60]. They typically comprise a conceptual layout with sets of rules on managing and controlling the asset they represent to perform at an efficient level [60]. While many governance frameworks such as information governance (IG), data governance, and corporate governance exist, the focus of our study is on IG to address the gap identified in ECCOLA.

Information governance (IG) is analyzed as a framework that contains mechanisms for guiding the creation, collection, storage, analysis, use distribution, and deletion of information relevant to the business to achieve value creation [61]. IG has its roots in the 20th century when the need arose to develop comprehensive and effective management of increasing volumes of data and information [62]. Previous efforts at governing information focused on basically archiving and retrieving information systematically [62]. However, as the volume of information and particularly electronically stored information increased alongside the speedy development of complex and interlinked systems, it gave rise to policies and rules to govern information and safeguard against loss and distortion [62]. At the time, only large and governmental organizations were inclined to invest in any IG due to the level of complexity, and pricing [62]. With the growth and development of computers and the internet, IG has become a concept available for all institutions, from government and large organizations to small firms and start-ups. In recent times, as the nature of data and information is evolving to include big data and other forms of unstructured data, there is a need to standardize IG and its infrastructure with generally suitable methods and measures [62].

Several IG frameworks exist as IG employs frameworks to enable it to carry out its governance practices [63]. The Unified Governance model also called the Information Governance Reference model IGRM from the eDiscovery community, EDRM [64] has been developed to address governance issues pertinent to eDiscovery issues. IBM developed an IG model, Information Lifecycle Management which initially focused on data but has since expanded to include other areas of record and information management [64]. The Generally Accepted Recordkeeping Principles (GARP®) or "The Principles" by The Association of Record Managers and Administrators (ARMA) to help address governance of information and records management [65] and even the Control Objectives for Information and Related Technology (COBIT) which focuses on IT governance has been discovered to share some commonalities with GARP® on some non-IT governance requirements such as protection [64]. However, the GARP® framework by ARMA has been recognized as having a widely leveraged global standard which identifies critical hallmarks of IG good practices at a high-level [65]. GARP® is also versatile and agnostic in its approach and can be used within different context [65]. For example [66] used the GARP® in their analysis of IG practices in block chain technology and the General data protection regulation (GDPR), implying that it can be applied to AI development tools such as the ECCOLA and ideal for our study.

GARP® approach provides guidance on information management, and the governance of record creation, organization, security maintenance, and other activities used to support record-keeping [65] efficiently. Records refer to content that could be data or information contextually created, recorded, or received in the initiation, conduct, and completion of an organizational or individual activity [67]. It also encompasses how it relates to other records, the organization or entity that created it, and the metadata likely to define its context. GARP® comprises eight principles – **Accountability**, **Transparency**, **Integrity**, **Protection**, **Compliance**, **Availability**, **Retention**, **Disposition**, and a maturity model made up of five milestones (sub-standard, in development, essential, proactive, and transformational) [65]. However, the scope of this study requires the analysis of GARP® and not the maturity model. The principles are explained further.

1. Accountability: Requires that a senior executive oversees IG programs and delegates responsibilities accordingly for information management, policies, and procedure adoption that guide personnel and ensure auditability.
2. Transparency: This deals with how documentation of activities and processes in an open and verifiable manner is made available to appropriate personnel and interested parties.
3. Integrity: Deals with how IG programs are constructed to reflect the authenticity and reliability of information assets generated or managed.
4. Protection: Deals with IG programs constructed with the appropriate level of protection for information assets on privacy, confidentiality, privilege, secrecy, classified documents, and how they relate to the continuity of the business and protection.
5. Compliance: Deals with how IG programs should be constructed to comply with applicable laws, binding authorities, and organization policies.
6. Availability: Focuses on how IG is exercised in information assets in a timely, efficient, and accurate retrieval manner.
7. Retention: Concerns with how IG is exercised consistently with an organization maintaining its information assets for an appropriate period considering its legal, regulatory, fiscal, operational, and historical requirements.

8. Disposition: Deals with how IG is carried out to provide secure and appropriate disposal of information assets that are no longer required to be maintained in Compliance with organizational laws and policies ([65]).

3. Methodology

Due to the novelty of our study and the limited resources identified in literature, we use ECCOLA as a case study. Case studies help to improve understanding of data derived within a specific context. Case studies within a research is often criticized as being difficult to conduct as they can lead to increased documentation [68]. However, a case study can allow for in-depth interpretation and evaluation of data by aiding the development of conceptual categories to help us add our judgment to the phenomenon found in the data [68]. In addition, since the study involves extending ECCOLA, we also incorporate aspects of design science in conceptualizing the extension of ECCOLA.

3.1. ECCOLA as a case study

Yin [68] explains that a case study can be used to describe or illustrate certain topics within an evaluation in a descriptive manner to provide an in-depth understanding pertinent to the phenomenon under study [68]. Using a case study also provides new or unexpected insights into the subject that can help propose practical courses of action to resolve identified issues and open up possible new directions for future research [68]. Case study is described as a linear but iterative process covering six main steps of planning, designing, preparing, collecting, analyzing and sharing [68]. We explain these steps in the context of ECCOLA.

The planning step aids the researchers in deciding on the use of a case study. Yin [68] explains that case studies are preferred when a "how" question needs to be answered in a research dealing with contemporary issues within real life contexts where control is not the focus. He also explains that case studies are unique in their ability to deal with a full variety of evidence such as documents, artifacts and observations [68]. We explore ECCOLA as a case study based on our research goal hinging on "how" AI ethical tools like ECCOLA can improve their robustness, AI ethical issues transcending technical boundaries to become socio-technical issues within contemporary real-life contexts, and the need for a representative AI ethical developmental tool like ECCOLA to help us deal with the variety of evidence encountered during the course of the study. Yin [68] also explains that a case study approach is for analytical generalization and not for statistical representation which is the case for our study.

For the design process, Yin [68] recommends that case studies are planned in a way that the evidence addresses the research question. As our study focuses on improving information robustness of AI ethical development tools, using such a tool in the form of ECCOLA as our unit of analysis can help ensure that we match the findings to the research question. Also Yin [68] explains that a case can be an event or an entity, of which ECCOLA qualifies as one. As such, we use ECCOLA as a single case study because a single case study can help provide credible test similar to critical experiments [68].

For preparation, Yin [68] explains the need for in-depth preparation to precede the case study. We carried out a comprehensive and rigorous analysis of ECCOLA in our previous study [20] to enable us determine its validity which enabled it for this study. In addition,

The GARP® IG framework was carefully sourced from literature as a pertinent source of information practices for the study.

For data collection, [68] explains collection of data or sources of evidence from documentation, archival records, interviews, direct observation, participant observation, illustrative materials and physical artifacts. ECCOLA served as the form of data collection in the form of documentation [68]. Documentation is described as communicable materials that are used in describing, explaining or instructing regarding attributes of procedures, objects or systems provided using different mediums either digital or analogue [67]. Also, ECCOLA is developed from AI ethics principles from which the method was created to ensure uniformity or convergence. This is because AI ethics principles usually form the basis for most AI ethical development tools like ECCOLA. As such ECCOLA as a data collection for the study also served as the unit of analysis.

Yin [68] explains that the analysis process usually begins with the data collection in case studies. He explains that establishing the data analysis early with well defined analytical tools can help to accomplish many other important aspects of the study [68]. We applied this in our research, as explained earlier, our analysis started with ECCOLA as part of the data collection. The analysis continued using the rigour of content analysis for a more critical analysis with the practices in the GARP® IG framework.

For sharing or communicating the results, Yin [68] recommends that awareness of the audience is important in disseminating the results. Indicating that care must be taken to ensure that the findings are appropriately communicated to the target audience. This is what we aim to do in this study.

3.2. Design science for extending ECCOLA

For the extension of ECCOLA to conceptualize and artifact, we incorporate aspects of the Design Science approach. Hevner et al. [69] explains that design science is naturally a problem solving process which requires knowledge and understanding of the design problem for its solution domain. They outline design processes build and evaluate, where purposeful solutions are built to address hitherto unsolved problems [69]. Our study focuses on the build phase which forms the first guideline out of the seven outlined by [69] for Design Science. They explain that while seven guidelines exists, researchers are not expected to follow them in a mandatory or mechanical fashion but use their creative skills to determine where, when and how to apply the guidelines in each specific projects.

We therefore follow the first guideline, *Design as an artifact* where the research must produce an artifact [69]. However, they explain that artifacts at this stage are rarely full grown information systems used in practice but are ideas and practices through which analysis, design, implementation and use of information systems can be effectively carried out [69]. Arguments exist that building artifacts using design science can be challenging due to the complexities of creative advances in fields with limited theories [69]. However, Design science provides a pragmatic approach that helps extend human and organizational boundaries to create new and innovative artefacts [69], making it suitable for the study. In addition, the pragmatism of the DSRM helps to provide a better process for answering "how" research questions and enables liberalism in exploring the answers [70].

We incorporate aspects from the Design Science research methodology (DSRM) outlined by [71]. Peffers et al. [71] describe the Design science approach as a rigorous process that aims for the design of an artifact to help solve identified problems, make research contributions, evaluate designs and communicate results to appropriate audience [71]. Six

process elements or guidelines that can be used to actualize an output is outlined. They include problem identification and motivation, defining the objectives for a solution, Design and Development, Demonstration, Evaluation and communication [71]. However, they explain that while the DSRM processes can follow a nominally sequential order, researchers are not expected to follow the steps rigidly and can start at any step in the process and proceed accordingly [71].

Peffers et al. explains that a problem centered approach can be the basis for the nominal sequence if the research idea results from an observed problem or from suggested future research from a prior project [71]. Our first study [20] provides this entry point. The paper [20] helped us to identify AI governance research gap in the ECCOLA method in the form of IG with the findings communicated in the paper. We therefore continue the process in this paper by following the next two steps as identified by [71]. The processes are defining the objectives of a solution and designing and developing an artifact where we communicate our findings in this paper [71].

3.2.1. Defining the objectives of the solution

For the second stage, the objective of a solution, i.e., what a better artefact can accomplish, we use a literature review to outline benefits of IG practices and how it affects ethical AI systems and development tools such as ECCOLA. The findings are provided to enable a broader understanding of the subject matter and create knowledge on some of the key concepts of IG practices as it pertains to AI governance and AI ethical development tools like ECCOLA [72].

3.2.2. Importance of information governance (IG)

Peffers et al. [71] explains that this stage involves inferring the importance of a solution from the identification of a problem and knowledge of what is possible and feasible. We outline the possibility and feasibility on what is possible with IG practices in AI ethical development tools like ECCOLA.

IG deals with the activities and technologies employed to maximize the value of information and minimize associated risks, and costs [73]. It comprises IG programs and measures that ensure information is appropriately controlled and accessible without compromise [67]. IG is often confused with data governance; however, data governance is narrowly focused on a specific information resource, data which is information in its raw, unstructured and unprocessed form. Data governance mainly deals with how data is governed by implementing appropriate measures and systems for producing and maintaining high-quality data [67]. IG covers information management which deals with managing information assets (IA). Overall, IG strategically provides a framework for managing information legally and ethically and helps balance risks associated with information and the value the information provides [63].

Software developers may find that through utilizing actionable ethical development tools such as ECCOLA, risks associated with the lack of IG governance practices are reduced [63]. Some examples of these risks include storage, disposition, and pre-processing [63]. Improper storage without the guidance of IG can lead to inaccessibility of information and inability to retrieve information. This results in a decrease in value and financial loss for developers. With the current surge in data and Big data used and generated by AI systems during their development life-cycle, having IG guidance on how to store information can help

developers avoid this loss. Information appropriately stored in line with IG practices can lead to effective retrieval, and reduced losses from e-discovery and other legal issues [63].

Disposition of information poses another risk for developers of AI items. Disposition risk involves storing too much retrievable data and the risks that may emanate from such practice [63]. Over storing data and information can lead to increased costs and risk using outdated or irrelevant, misleading or ineffective data [63]. Lack of disposition can also lead to control risk associated with storing data over a prolonged period [63]. When AI systems store data beyond its relevance period, it can pose a risk if hackers gain control of information that should have been long disposed of, putting users' personal information at risk. In addition, if users discover that their personal information supplied to a system at some point is retained after a prolonged period, it can also lead to a lack of trust and confidence in the system. IG advocates for timely disposition of data and information assets so they are not accessed and used maliciously and reduce redundancy costs. For pre-processing risks, IG can aid developers in managing data so that inconsistencies and missing aspects of data that can arise when gathering data are mitigated timely and do not lead to redundancies in the AI systems [63].

3.2.3. Design and development

The third and final step in the nominal process used in this study involves designing and creating an artifact [71]. Peffers et al. [71] explains that artifacts could be potentially models, methods, constructs, instantiations or new properties which could be technical, social or informational resources. They clarify further that a design research artifact could conceptually be any designed object where the results or contributions of a research are embedded in [71]. In the next sections, we present the analysis and findings used in the creation of artifact.

4. Results

This section presents the analysis and presents the findings from the analysis. For the analysis of the study, we employ the use of content analysis. Content analysis enables replicable and valid inferences from texts to the context of their use [74]. It is also useful in evaluating work to compare communication content against previously documented objectives [74]. In addition, content analysis is effective in analyzing text or data such as principles, interviews, field research notes, journals, books, guidelines and reports [75] making it the most suitable option for our study. Using content analysis allowed us to evaluate the languages used within our data, search for bias and make inferences [76]. We also found it useful in reducing our data to concepts to describe the research study by creating categories. Arguments exist that content analysis can be subjective and reductive, but they are also transparent, provide flexibility, and are replicable [76].

4.1. Analysis of ECCOLA with GARP®

We use interpretive content analysis which is qualitative in nature requiring no statistical inference but rather focuses on summarizing and describing meanings in an interpretive and narrative manner [74]. The interpretive approach is described by [74] as a procedure that enables researchers make inference about source and receiver communication from

evidence in the messages they exchange. The approach also allows for both manifest and latent content to be considered and analysed [74]. Latent content refers to meaning that is not overt or obvious but is implicit or implied in the communication while manifest content refers to the more obvious meaning within the communication [74].

4.1.1. Process

The content for analysis are the ECCOLA cards. The GARP® framework is used to create an index for the analysis. The index of analysis is created by first defining the units and categories of analysis. The unit of meaning to be coded are identified as activities or processes indicative of the GARP® principles, and the set of categories used for the coding are identified in the table of index, Table 2. A set of rules to determine coding of the ECCOLA cards against the index table are identified as Exist, partially exist and does not exist. The existence of all the practices identified in the index against the card is coded as **Exist**, the existence of one or more but not all is coded as **Partially exist** and the complete absence of activities is coded as **Does not exist**. Each ECCOLA card was manually coded by identifying the units of meaning into the conceptually defined categories in the table of index and the codes documented accordingly [74].

Table 2. Table of index

Unit of meaning	Set of categories
Accountability: activities or processes	Accountability structure Documentation Guiding (policies, procedures, decisions) Audit
Transparency: activities or processes	Open documentation of IA Available documentation of IA Verifiable documentation of IA Accessibility of IA
Integrity: activities or processes	Authentic management of IA, Reliable management of IA
Protection: activities or processes	Protection of private IA Protection of confidential IA Protection of privileged IA Protection of secret IA Protection of essential IA Categorization of IA (private, confidential, privileged, secret, classified)
Compliance: activities or processes	Compliance with applicable laws Compliance with binding authorities Compliance with organizational policies Compliance in (documentation, storage)
Availability: activities or processes	Maintenance of IA for timely retrieval Maintenance of IA for efficient retrieval Maintenance of IA for accurate retrieval Documentation of IA for accessibility
Retention: activities or processes	Maintenance period of IA for legal requirements Maintenance period of IA for regulatory requirements Maintenance period of IA for fiscal requirements Maintenance period of IA for operational requirements Maintenance period of IA for historical requirements Documentation of IA Retention period of IA Storage of IA
Disposition: activities/processes	Secure disposition of irrelevant IA by laws Secure disposition of irrelevant IA by policies Appropriate disposition of IA by laws Appropriate disposition of IA by policies Disposition documentation of IA

4.2. Findings

The result of the analysis is presented in the heat map in Figure 1 and the findings highlighted as Primary Empirical Contributions (PECs).

1. Accountability: The analysis of the ECCOLA cards against the GARP® index of Accountability reveals that four cards (4, 9, 18, and 20) have an "exist" status. Activities and practices within the four cards indicate GARP® IG practices such as transparent **documentation, audit**, and activities that demonstrate a **reporting structure** (who makes decisions) or allude to an **accountability structure** aligned with **regulatory bodies** and **policies** [65]. The remaining 17 cards have a status of partially exist. Actions and practices in the cards indicate one or more but not all the IG practices in line with the GARP® index, thus having a partial representation, Indicating that these cards can be improved with GARP® accountability practices. This leads us to our first PEC.
   PEC1: 17 of the ECCOLA cards can be adapted fully with GARP® IG practices of Accountability by referencing activities and practices such as documentation, accountability, and auditing of information assets.
2. *Transparency*: The transparency analysis revealed five cards (4, 5, 6, 9, and 19) with an "exist" status showing clear indications of practices and activities that facilitate **documentation** and **accessibility** of IA in a clear and verifiable manner **accessible** by the appropriate personnel. The remaining 16 cards have a "partially exist" status. There are indications of either of the practices in these cards, but not all of them, For example, the practices and activities in card #8 (Data quality) point towards proper accessibility of data but makes no reference to documentation or mode of documentation of IA that can result from these practices forming PEC2.
   PEC2: ECCOLA can be adapted fully with The GARP® IG Transparency practices with references such as documentation, mode of documentation, and accessibility in 16 cards.
3. *Integrity*: The integrity analysis indicates that all the 21 ECCOLA cards reflect activities and practices that align with Integrity practices of the GARP®. Activities and practices within the cards indicate **authenticity** and **reliability** in the handling of IA affiliated with the GARP® of Integrity. In Card #10 (Human Agency), the activities and practices are geared at sensitizing developers on the need for authentic practices that lead to reliability for the AI users. While some cards do not state the practices expressly, there is an allusion to them leading us to the third PEC.
   PEC3: ECCOLA has exist status and does not require further adaptation with the GARP® principle of Integrity.
4. *Protection*: The Protection analysis reveals eight cards (6, 7, 8, 9, 12, 13, 18, and 20) have an "exist" status. GARP® IG practices such as **protection mechanisms** for **designated** or **categorized** IA exist in these cards. One of the cards #7 (Privacy and Data) sensitizes developers on the need for protection of data by asking questions on encryption, anonymization of data, and accessibility of protected IA. In the remaining 13 cards, either one of these practices was identified, resulting in a "partially exist" status. This forms the fourth PEC.
   PEC4: The ECCOLA method can be adapted fully with GARP® practices of Protection with references such as protection/protection mechanisms and categorization or designation of IA in 13 cards.

5. *Compliance*: The compliance analysis reveals that 13 cards, (1, 6, 7, 8, 9, 13, 14, 15, 16, 17, 18, 19, and 20) have an "exist" status by displaying IG practices that align with the GARP® index. Implying that these cards create awareness for AI developers of compliance practices such as **documentation, storage, applicable laws, organizational policies**, and **authorities** for IA. Only one or two of the practices could be identified in the remaining eight cards forming our PEC5.
   PEC5: ECCOLA can be adapted fully with Compliance practices of the GARP® principle in eight cards by making references such as documentation, storage, applicable laws, organizational policies, and authorities for IA.
6. *Availability*: The Availability analysis shows that only one card has an "exist" status. Activities and processes within the lone card #9 ask AI developers questions such as who has access to users' data and the circumstances they are granted access—ensuring that IA **documentation**, **accessibility**, and **retrieval** align with the Availability GARP® IG practices in the index. The remaining 20 cards have a "partially exist" status by indicating one or two of these practices but not all of them.
   PEC6: ECCOLA can be adapted fully with the GARP® practices of Availability in 20 of its cards by referencing practices of documentation, accessibility, and retrieval of IA.
7. *Retention*: The Retention analysis reveals that nine of the cards (0, 4, 5, 6, 7, 8, 9, 18, and 20) have a "partially exist" status. The Cards asks AI developers questions on GARP® Retention IG practices such as **documentation, storage**, **maintenance** and **retention/period** in one or two contexts or allude to them in activities for IA. However, the remaining 12 cards show no indication or allusion to these practices leading us to our PEC7.
   PEC7: ECCOLA has partially exist status of GARP® principle of Retention in nine cards and does not exist status in 12 cards. ECCOLA can be extended to incorporate the GARP® IG practices of Retention such as documentation, storage, maintenance and retention.
8. *Disposition*: The Disposition analysis reveals that nine of the cards have a "partially exist" status. The cards (0, 4, 5, 6, 7, 8, 9, 18, and 20) have one or two GARP® Disposition practices and activities such as **documentation, transfer**, and **Disposition** of IA that can facilitate IG. In comparison, these practices could not be identified in the remaining 12 cards. Card #3 (Communication) asks AI developers pertinent questions on the need for transparent practices in developing AI systems; however, there are no questions or practices on how generated IA are disposed of, leading us to PEC8.
   PEC8: ECCOLA has partially exist status of GARP® principle of Disposition in nine cards and does not exist in 12 cards. ECCOLA can be extended to incorporate GARP® IG practices of Disposition such as documentation,transfer, and Disposition of IA.

5. Discussion

We discuss the outcome of the analysis which yielded eight PECs in Table 3. The PECs and their implications are discussed in this section.

PEC1 is based on adapting accountability practices of documentation, accountability, and auditing of information assets to ethical AI development tools like ECCOLA to help facilitate audit practices. The absence of some of its practices can translate to crucial accountability practices being omitted while using the ECCOLA cards. Where these practices are not present, software developers may cultivate practices and activities that

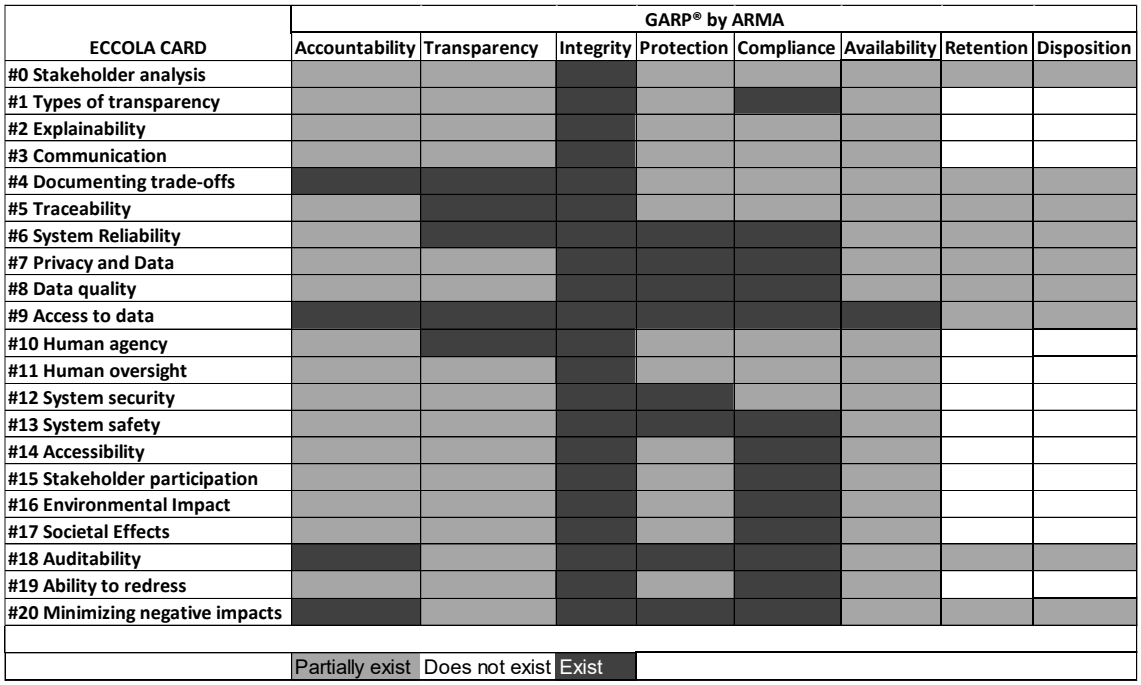

Figure 1. ECCOLA analysis with GARP®

overlook accountability structure or non-documentation of crucial information that does not comply with policies. In addition, adapting GARP® practices with already existing ethical principles can help ECCOLA mitigate accountability risks that may arise in the development of AI systems [1].

Reddy et al. [5] analyses accountability as a challenge as regards implementation in terms of governance. They explain that appropriate stages are needed for effective accountability practices and stress the need for an approval structure by governing bodies or regulating authorities that oversee and preview processes to ensure proper documentation of IA that can aid audits in governance [5]. This article further explains that accountability in governance requires regulatory oversight and needs to be in place in the development stage of AI. Guiding actions by an accountability structure and explanations in the form of IA documentation can facilitate internal or external audits. It also makes developers of AI accountable in the documentation of their IA and may help reduce the opacity of AI governance [5]. The principle of transparency (PEC2) highlights the need for transparent practices such as documentation, mode of documentation of IA to make them more accessible in IG for AI developers. M.S. Caron [77] argues that transparent processes in AI systems help to improve auditability. She explains that open and verifiable practices that generate IA must be transparent in the development process. Adapting transparency governance practices can aid understanding when IA is accessed by appropriate personnel, which is vital for auditability. Also, in developing AI, different cognitive biases and heuristics exist [77] warranting the need for transparent obligations to be imposed. These obligations can be in the form of auditable governance measures to help mitigate these practices so that when these IA are accessed, there is a clear understanding that helps audit in governance [77]. Kiener in [37] agrees with this argument and discusses the need for open and verifiable processes in the development of AI in sensitive fields like medicine. He explains that transparent processes and activities of AI can aid human oversight, risks, and audits in governance frameworks like IG.

PEC3 highlights the principle of integrity, which explains the need for reliability and authenticity in the processes, activities, and practices that generate IA in AI development. Adapted practices of integrity help facilitate proper audits in governance frameworks. Jansen et al. [78] explains that a governance framework in the development of AI can enable authentic and reliable IA. They analyze that a governance structure (such as IG) can enable IA of integrity by managing the quality, validity, security, and associated risks in practice to preserve the integrity.

PEC4 is based on the principle of protection. It emphasizes the need for protection practices such as protection/protection mechanisms and categorization or designation of IA for designated IA (private, confidential, privileged, secret, classified). Protection mechanisms and practices can provide safeguards to mitigate risks from incidental access or disclosures. Culnan in [79] supports this and explains that the IG protection practices in the development of AI can help ensure they are secure and properly categorized to help avert security and privacy breaches. While introducing regulatory mechanisms such as the GDPR and the California Consumer Privacy Act (CCPA) exists to help protect AI users, they can be insufficient in protecting IA that are not adequately categorized or labeled. To ensure that a suitable form of protection is provided [79], incorporating these practices can aid the audit processes involved in governance for AI developers [79].

PEC5 findings relate to how compliance practices like documentation, storage of IA in a manner that complies with applicable laws, organizational policies, and other binding authorities adapted in the development stage of AI can help aid AI governance overall. J. In et al. [80] argues that IG practices of maintaining IA in a manner that conforms to compliance (internal and external) can help mitigate risk and increase efficiency. He explains that when developers or organizations familiarise themselves with compliance practices and streamline the maintenance of their IA in line with governance frameworks like IG, such practices become routine and make it easy to produce systems compliant with applicable laws and binding authorities. Furthermore, in legal matters and regulation, these practices can help audit processes in AI governance [80].

PEC6 is based on the principle of availability. It explores how adapting practices such as documentation, accessibility, and retrieval practices can facilitate the maintenance of IA to ensure timely efficiency. Hind et al. [81] explains that AI developers are usually faced with the challenge of documentation of IA as there are no clear guidelines on how much to document to provide enough clarity. Therefore, governance practices geared towards appropriate documentation, accessibility, and retrieval of IA to ensure that IA is effectively and adequately documented and, upon retrieval, provide holistic information may aid clarity. Documentation of IA in line with a governance framework like IG also provides confidence that information made available is wholesome and suitable for all interested parties. In addition, these practices also aid the governance frameworks in audit processes to ensure unity [81].

PEC7 is based on the principle of retention and how incorporating governance practices such as effective maintenance, documentation, storage, and retention (period of retention) of IA can improve AI governance in development methods like ECCOLA. Kroll in [82] recommends the minimization of retention of collected records or the disposal of aggregate records where possible to enhance efficient governance of records. He explains that retention of IA should be appropriately maintained, documented and subject to a governance structure to reduce the risks of retaining them beyond their retention period. Retention of IA combined with the principles approach can help minimize the risk from legitimate requests from law enforcement [82]. When Information Assets are retained beyond their

life cycle, they can pose a risk if authorities request them and utilize them beyond the scope for which they were acquired [82].

PEC8 works on how adapting disposition practices like secure and appropriate transfer, disposition, and documentation of records or information (IA) in compliance with applicable laws and policies can improve AI governance in development methods like ECCOLA. Kroll [82] explains that regular disposal of aggregate and redundant records or reducing them to the lowest level of sensitivity can reduce privacy risks and increase the efficiency of AI governance. When Information Assets are maintained for a period, a need for them must be further established to enable them not to pose a risk of redundancy which may hamper efficiency. A clear need exists for records or information to be retained for a period and disposed of accordingly, as it can help make development methods more trustworthy when AI audits are carried out.

Table 3. Primary Empirical Contributions (PECs)

Primary Empirical Contributions (PECs)
PEC1 – 17 of the ECCOLA cards can be adapted fully with GARP® IG practices of Accountability by referencing activities and practices such as documentation, accountability, and auditing of information assets.
PEC2 – ECCOLA can be adapted fully with GARP® IG Transparency practices with references such as documentation, mode of documentation, and accessibility in 16 cards.
PEC3 – ECCOLA has exist status and does not require further adaptation with the GARP® principle of Integrity.
PEC4 – The ECCOLA method can be adapted fully with GARP® practices of Protection with references such as protection/protection mechanisms and categorization or designation of IA in 13 cards.
PEC5 – ECCOLA can be adapted fully with Compliance practices of the GARP® principle in eight cards by making references such as documentation, storage, applicable laws, organizational policies, and authorities for IA.
PEC6 – ECCOLA can be adapted fully with the GARP® practices of Availability in 20 of its cards by referencing practices of documentation, accessibility, and retrieval of IA.
PEC7 – ECCOLA has partially exist status of GARP® principle of Retention in nine cards and does not exist status in 12 cards. ECCOLA can be extended to incorporate the GARP® IG practices of Retention such as documentation, storage, and retention.
PEC8 – ECCOLA has partially exist status of GARP® principle of Disposition in nine cards and does not exist in 12 cards. ECCOLA can be extended to incorporate GARP® IG practices of Disposition such as documentation, transfer, and Disposition of IA.

5.1. Extension of ECCOLA

From the discussion, it has been identified that ECCOLA can be improved and extended to incorporate IG practices to improve its information robustness. While six of the principles can be improved in the cards by making references to pertinent practices, a need exist for proper representation of the unidentified principles of retention and disposition and governance practices as a whole. The discussion has also outlined how the lack of governance principles identified and particularly those of retention and disposition can pose risks for developers of AI systems who use tools like ECCOLA. As such, it is important to highlight these key governance practices in the form of a new theme and card. As such, we propose a new governance theme in ECCOLA and a new card which incorporates the deficient IG

Figure 2. Card 21

practices of retention and disposition. Therefore, we propose a new card, #21 under a new theme, governance illustrated in Figure 2.

Motivation Letting users know the period of retention of their records by the AI system creates Transparency and encourages trust that their records will not be stored indefinitely.

What to do: Ask yourself:

– Are records maintained in line with policies?
– Are users informed of the retention period of their records?
– Are users informed of the need to retain their records beyond the retention period?
– Are obsolete and redundant records destroyed or disposed in line with regulatory requirement?
– Are records securely disposed and not retained beyond their retention period?

Practical example: Establishing clear time limits for collected records and explaining the need to retain records beyond retention period makes the system trustworthy. AI systems that provide recruitment recommendations will find their services ineffective if users' information is not updated regularly to reflect their skill set.

When AI developers create a culture of openly communicating to users and end-users, the period of retention and disposition of their records or information, it can enable them to trust such AI or AI-enabled systems. The trust can be due to increased confidence that their information or records will not be stored and used indefinitely [82]. Also, frequently updating users' records and disposing of redundant records can enable developers of AI systems ensure that they have access to current and better quality records or information [82].

5.1.1. Significance of a new theme and card

To a large extent the discourse on governing information assets as it relates to AI ethics has largely focused on data governance. This may explain the major focus of proposed ethical model tools such as google model cards, data ethics canvas and even the current version of ECCOLA on data governance. These tools help draw attention of AI developers and relevant stakeholders to data ethics governance challenges which are vital in implementing AI ethics. The current version of ECCOLA has visible representation of data governance practices in the data cards(#7, #8 and #9) but there is no clearly delineated representation for IG practices for records or information that can be generated from data.

Model cards as AI ethical development tools help to provide *information* on trained machine learning models [83] or AI models to improve ethical practices such as transparency. As such it is pertinent that data governance practices are in place for data collected or sourced at the basic data level. But the records generated from the data used in training these models which are actually represented as *information* require that IG practices are in place for these sets of information assets as well. As Glasser et al. [45] explains that AI governance is a layered structure requiring governance practices at each layer to achieve the goal of governance. Therefore, incorporating IG practices such as retention and disposal and highlighting them in a governance theme can improve improve information robustness and help create a culture of visibility and informed communication for both developers and users [45].

Therefore, in comparison to other proposed ethical AI development tools that promote AI governance practices like the current version of ECCOLA, the proposed theme and card can help provide visibility and improve the information base of these tools towards AI governance. The current ECCOLA has no clear governance representation as most of them are embedded in the AI ethics practices. However the growing emphasis on the need for visible governance practices to be implemented at the developmental stages [1, 5, 10, 39, 45] necessitates that AI ethical developmental tools like ECCOLA have visible measures that aim to improve governance information and to tackle governance challenges that can occur in development. Visibility of Governance measures at this stage can help bring to the forefront pertinent AI governance challenges and provide the necessary information needed to address them [5, 45].

5.2. Validity threats

The reliability of the content analysis is a potential validity threat to the research. While interpretive content analysis is reproducible [74], the study could be subject to our own interpretation due to the nature of subjectivity of the documents such as the AI ethical principles in the ECCOLA cards and the researchers in determining the index terms from the GARP® documents used in the actual analysis.

The use of a single case study also poses a validity threat to the study. As highlighted by [68], the use of a single case study can pose a challenge for generalization of results. However, [68] still explains that a single case study can be used for analytic generalization and not for statistical generalization which is the case for our study.

A third and possibly one of the biggest threats to the validity of our study is the incomplete Design science approach used. We understand that most Design science usually require a valid artifact as output for the study [69]. Following the build and and evaluate approach [69], our study at this stage focuses on the build stage or the design an artifact

stage. We are aware that this can serve as a limitation to the study, however, we argue that following the recommendations of [69], we are not following all the methodology steps in a rigid or mechanical fashion and will use the outcome of each study to enable us leverage the next study until we have gone through all the DSRM steps to produce a viable artifact [69].

6. Conclusions and further works

This study explored how ethical practices in AI ethical developmental tools like ECCOLA can be adapted with IG practices for improved information robustness in tackling AI governance issues using the adaptive or hybrid governance highlighted by [1]. ECCOLA was critically analysed with the GARP® IG framework using content analysis to highlight areas where the practices in the cards could be improved with IG practices for a more information robust base. The results reveal 8 PECs, which indicate that most of the IG practices in the GARP®: Accountability, Transparency, Integrity, Protection, Compliance, and Availability are represented in ECCOLA either partially or fully to varying degrees. The findings further reveal that all 21 cards in ECCOLA comply with IG practices of Integrity. However, the principles of Retention and Disposition were shown to be the least represented and lacking IG practices in ECCOLA as they could not be identified in 12 of the cards. Implying that ECCOLA can improve its information robustness in all 21 cards and also be extended with a governance theme to improve its governance practices.

Regarding the implication of the findings to AI governance, it may imply that AI ethical practices in ECCOLA may be insufficient in addressing some governance issues that may arise in the development process of AI systems. AI governance issues that deal with Integrity may be fully addressed and partially addressed for Accountability, Transparency, Protection, Compliance, and Availability practices. However, issues that may arise from Retention and Disposition AI governance challenges may be partially or not addressed, indicating that the method's ethical practices can be improved in terms of its information robustness to aid AI governance. As a solution, the index terms from the analysis are suggested as potential modifiers for the cards with partial practices. However, for a better representation of governance practices in the tool, a governance theme is proposed. The theme will help address pertinent governance issues for developers as well as improve governance information robustness of ECCOLA. The first of the cards in this theme is proposed as #21 which addresses concerns from governance issues as it pertains to retention and disposition. The card provides motivation, suggested activities, and a practical example of how the card can effectively tackle ethical AI governance challenges that may arise at the developmental stage. Therefore, the analysis offers an answer and a possible solution to our research question.

In addition to analyzing ECCOLA, this study also explored an adaptive AI governance approach initiative at the development stage. The ethical practices were modified or adapted with IG practices to create a more practical approach to governance issues that developers of AI systems can encounter at the development stage.

The findings or outcome from this study will form the basis for the next phase of our study which will involve evaluation of the artifact to continue the DSRM process. Paper [71] explains that the DSRM is not a linear process to be followed rigidly but allows for the incorporation of different elements of the methodology as the research progresses. Therefore,

we aim to continue the remaining parts of the research in subsequent studies which are outside the scope of this one.

Information for funding/support or the lack of it

This research is partially funded by Business Finland (business-finland.fi)research projects. The are Sea4value and Stroke Data and ITEA4. The authors are grateful to the funders for their support.

References

[1] A. Taeihagh, "Governance of artificial intelligence," *Policy and Society*, 2021, pp. 1–21.
[2] M. Schulte-Althoff, D. Fürstenau, and G.M. Lee, "A scaling perspective on AI startups," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2021, p. 6515.
[3] C. Giardino, S.S. Bajwa, X. Wang, and P. Abrahamsson, "Key challenges in early-stage software startups," in *International conference on agile software development*. Springer, 2015, pp. 52–63.
[4] C. Newton, J. Singleton, C. Copland, S. Kitchen, and J. Hudack, "Scalability in modeling and simulation systems for multi-agent, AI, and machine learning applications," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III*, Vol. 11746. International Society for Optics and Photonics, 2021, p. 1174626.
[5] S. Reddy, S. Allan, S. Coghlan, and P. Cooper, "A governance model for the application of AI in health care," *Journal of the American Medical Informatics Association: JAMIA*, Vol. 27, No. 3, 2020, pp. 491–497.
[6] P. Liu, M. Du, and T. Li, "Psychological consequences of legal responsibility misattribution associated with automated vehicles," *Ethics and Information Technology*, Vol. 23, No. 4, 2021, pp. 763–776.
[7] G. Falco, B. Shneiderman, J. Badger, R. Carrier, A. Dahbura et al., "Governing AI safety through independent audits," *Nature Machine Intelligence*, Vol. 3, No. 7, 2021, pp. 566–571.
[8] S. Du and C. Xie, "Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities," *Journal of Business Research*, Vol. 129, 2021, pp. 961–974.
[9] "Ethics guidelines for trustworthy AI," European Commission, High-Level Expert Group on AI, Tech. Rep., 2019. [Online]. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[10] V. Vakkuri, K.K. Kemell, M. Jantunen, E. Halme, and P. Abrahamsson, "ECCOLA – A method for implementing ethically aligned AI systems," *Journal of Systems and Software*, Vol. 182, 2021, p. 111067.
[11] D. Lewis, W. Reijers, H. Pandit, and W. Reijers, *Ethics canvas manual*, ADAPT Centre and Trinity College Dublin and Dublin City University, 2017. [Online]. https://www.ethicscanvas.org/download/handbook.pdf
[12] R. Eitel-Porter, "Beyond the promise: Implementing ethical AI," *AI and Ethics*, Vol. 1, No. 1, 2021, pp. 73–80.
[13] R. Hamon, H. Junklewitz, and I. Sanchez, "Robustness and explainability of artificial intelligence," 2020.
[14] I. Linkov, B.D. Trump, K. Poinsatte-Jones, and M.V. Florin, "Governance strategies for a sustainable digital world," *Sustainability*, Vol. 10, No. 2, 2018, p. 440.
[15] U. Pagallo, P. Aurucci, P. Casanovas, R. Chatila, P. Chazerand et al., "On good AI governance: 14 priority actions, a SMART model of governance, and a regulatory toolbox," 2019.
[16] S.Y. Tan and A. Taeihagh, "Adaptive governance of autonomous vehicles: Accelerating the adoption of disruptive technologies in Singapore," *Government Information Quarterly*, Vol. 38, No. 2, 2021, p. 101546.
[17] S. Jain, M. Luthra, S. Sharma, and M. Fatima, "Trustworthiness of artificial intelligence," in *6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2020, pp. 907–912.

[18] I. Brass and J.H. Sowell, "Adaptive governance for the Internet of Things: Coping with emerging security risks," *Regulation and Governance*, Vol. 15, No. 4, 2021, pp. 1092–1110.
[19] A.B. Whitford and D. Anderson, "Governance landscapes for emerging technologies: The case of cryptocurrencies," *Regulation and Governance*, Vol. 15, No. 4, 2021.
[20] M. Agbese, H.K. Alanen, J. Antikainen, E. Halme, H. Isomäki et al., "Governance of ethical and trustworthy AI systems: Research gaps in the ECCOLA method," in *29th International Requirements Engineering Conference Workshops (REW)*. IEEE, 2021, pp. 224–229.
[21] C. Collins, D. Dennehy, K. Conboy, and P. Mikalef, "Artificial intelligence in information systems research: A systematic literature review and research agenda," *International Journal of Information Management*, Vol. 60, 2021, p. 102383.
[22] A.B. Simmons and S.G. Chappell, "Artificial intelligence-definition and practice," *IEEE Journal of Oceanic Engineering*, Vol. 13, No. 2, 1988, pp. 14–42.
[23] S. Leijnen, H. Aldewereld, R. van Belkom, R. Bijvank, and R. Ossewaarde, "An agile framework for trustworthy AI," in *Proceedings of the First International Workshop on New Foundations for Human-Centered AI*, 2020, pp. 75–78.
[24] "A definition of AI: Main capabilities and scientific disciplines," European Commission, High-Level Expert Group on AI, Tech. Rep., 2019. [Online]. https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines
[25] C.F. Tan, L. Wahidin, S. Khalil, N. Tamaldin, J. Hu et al., "The application of expert system: A review of research and applications," *ARPN Journal of Engineering and Applied Sciences*, Vol. 11, No. 4, 2016, pp. 2448–2453.
[26] L. Ma and B. Sun, "Machine learning and AI in marketing–Connecting computing power to human insights," *International Journal of Research in Marketing*, Vol. 37, No. 3, 2020, pp. 481–504.
[27] E.J. Rykiel Jr., "Artificial intelligence and expert systems in ecology and natural resource management," *Ecological Modelling*, Vol. 46, No. 1–2, 1989, pp. 3–8.
[28] P. Hu, Y. Lu et al., "Dual humanness and trust in conversational AI: A person-centered approach," *Computers in Human Behavior*, Vol. 119, 2021, p. 106727.
[29] Y. Bengio, *Learning deep architectures for AI*. Now Publishers, Inc., 2009.
[30] R.R. Kumar, M.B. Reddy, and P. Praveen, "Text classification performance analysis on machine learning," *International Journal of Advanced Science and Technology*, Vol. 28, No. 20, 2019, pp. 691–697.
[31] K. Oosthuizen, E. Botha, J. Robertson, and M. Montecchi, "Artificial intelligence in retail: The AI-enabled value chain," *Australasian Marketing Journal*, No. 3, 2020.
[32] B.W. Wirtz, J.C. Weyerer, and B.J. Sturm, "The dark sides of artificial intelligence: An integrated AI governance framework for public administration," *International Journal of Public Administration*, Vol. 43, No. 9, 2020, pp. 818–829.
[33] K. Siau and W. Wang, "Artificial intelligence (AI) ethics: Ethics of AI and ethical AI," *Journal of Database Management (JDM)*, Vol. 31, No. 2, 2020, pp. 74–87.
[34] K. Michael, R. Abbas, P. Jayashree, R.J. Bandara, and A. Aloudat, "Biometrics and AI bias," *IEEE Transactions on Technology and Society*, Vol. 3, No. 1, 2022, pp. 2–8.
[35] P. Mikalef, K. Conboy, J.E. Lundström, and A. Popovič, "Thinking responsibly about responsible AI and 'the dark side' of AI," *European Journal of Information Systems*, Vol. 31, No. 3, 2022, pp. 257–268.
[36] C. Trocin, P. Mikalef, Z. Papamitsiou, and K. Conboy, "Responsible AI for digital health: A synthesis and a research agenda," *Information Systems Frontiers*, 2021, pp. 1–19.
[37] M. Kiener, "Artificial intelligence in medicine and the disclosure of risks," *AI and Society*, Vol. 36, No. 3, 2021, pp. 705–713.
[38] V.C. Müller, "Ethics of artificial intelligence and robotics," in *The Stanford Encyclopedia of Philosophy*, E.N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2021. [Online]. https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/
[39] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence*, Vol. 1, No. 9, 2019, pp. 389–399. [Online]. http://www.nature.com/articles/s42256-019-0088-2

[40] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, "From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices," in *Ethics, Governance, and Policies in Artificial Intelligence*. Springer, 2021, pp. 153–183.
[41] M. Hickok, "Lessons learned from AI ethics principles for future actions," *AI and Ethics*, Vol. 1, No. 1, 2021, pp. 41–47.
[42] P. Sondergaard, *AI governance – what are the KPIs? And who is accountable?*, 2021.AI, (2019, Nov). [Online]. https://2021.ai/ai-governance-kpi [Accessed: 16Jun. 2021].
[43] R.I. Rotberg, "Good governance means performance and results," *Governance*, Vol. 27, No. 3, 2014, pp. 511–518.
[44] M. Ashok, R. Madan, A. Joha, and U. Sivarajah, "Ethical framework for artificial intelligence and digital technologies," *International Journal of Information Management*, Vol. 62, 2022, p. 102433.
[45] U. Gasser and V.A. Almeida, "A layered model for AI governance," *IEEE Internet Computing*, Vol. 21, No. 6, 2017, pp. 58–62.
[46] A.F. Winfield, K. Michael, J. Pitt, and V. Evers, "Machine ethics: The design and governance of ethical AI and autonomous systems," *Proceedings of the IEEE*, Vol. 107, No. 3, 2019, pp. 509–517.
[47] J. Butcher and I. Beridze, "What is the state of artificial intelligence governance globally?" *The RUSI Journal*, Vol. 164, No. 5-6, 2019, pp. 88–96.
[48] H. Yu, Z. Shen, C. Miao, C. Leung, V.R. Lesser et al., "Building ethics into artificial intelligence," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI' 18. AAAI Press, 2018, p. 5527–5533.
[49] A. Daly, T. Hagendorff, H. Li, M. Mann, V. Marda et al., "AI, governance and ethics: global perspectives," The Chinese University of Hong Kong, Faculty of Law, Research Paper 2020/05, 2020. [Online]. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3684406
[50] B. Perry and R. Uuk, "AI governance and the policymaking process: Key considerations for reducing AI risk," *Big data and cognitive computing*, Vol. 3, No. 2, 2019, p. 26.
[51] W. Wu, T. Huang, and K. Gong, "Ethical principles and governance technology development of AI in China," *Engineering*, Vol. 6, No. 3, 2020, pp. 302–309.
[52] P. Cihon, "Standards for AI governance: International standards to enable global coordination in AI research and development," Future of Humanity Institute, University of Oxford, Technical Report, 2019. [Online]. https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf
[53] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nature Machine Intelligence*, Vol. 1, No. 11, 2019, pp. 501–507.
[54] R. Leenes and F. Lucivero, "Laws on robots, laws by robots, laws in robots: Regulating robot behaviour by design," *Law, Innovation and Technology*, Vol. 6, No. 2, 2014, pp. 193–220.
[55] M. Firlej and A. Taeihagh, "Regulating human control over autonomous systems," *Regulation and Governance*, Vol. 15, No. 4, 2021, pp. 1071–1091.
[56] K. Yeung, A. Howes, and G. Pogrebna, "AI governance by human rights – centered design, deliberation, and oversight," in *The Oxford Handbook of Ethics of AI*. Oxford University Press, 2020, p. 77.
[57] X. Cao, D. Lv, L. Zhang, and Z. Xing, "Adaptive governance, loose coupling, forward-looking strategies and responsible innovation," *IEEE Access*, Vol. 8, 2020, pp. 228 163–228 177.
[58] R. Radu, "Steering the governance of artificial intelligence: National strategies in perspective," *Policy and Society*, 2021, pp. 1–16.
[59] J. Ayling and A. Chapman, "Putting AI ethics to work: Are the tools fit for purpose?" *AI and Ethics*, Vol. 2, 2021, pp. 405–429.
[60] O.E. Williamson, "The economics of governance: Framework and implications," *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics*, 1984, pp. 195–223.
[61] H. Borgman, H. Heier, B. Bahli, and T. Boekamp, "Dotting the I and crossing (out) the T in IT governance: New challenges for information governance," in *49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2016, pp. 4901–4909.

[62] H. Dogiparthi, "History of information governance," University of the Cumberlands, Dept. of Information Technology, Research Paper, 2019. [Online]. https://www.researchgate.net/publication/330844911_History_of_Information_Governance
[63] E.M. Coyne, J.G. Coyne, and K.B. Walker, "Big data information governance by accountants," *International Journal of Accounting and Information Management*, 2018.
[64] J. Hagmann, "Information governance–beyond the buzz," *Records Management Journal*, 2013.
[65] *Generally Accepted Recordkeeping Principles®*, ARMA International, 2017. [Online]. https://www.arma.org/page/principles
[66] D. Hofman, V.L. Lemieux, A. Joo, and D.A. Batista, ""The margin between the edge of the world and infinite possibility": Blockchain, GDPR and information governance," *Records Management Journal*, 2019, pp. 240–257.
[67] E. Lomas, "Information governance: information security and access within a uk context," *Records Management Journal*, No. 2, 2010.
[68] R.K. Yin, *Case Study Research: Design and Methods*. Sage Publications, 1994.
[69] A.R. Hevner, S.T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, No. 1, 2004, pp. 75–105.
[70] J.W. Creswell, W.E. Hanson, V.L. Clark Plano, and A. Morales, "Qualitative research designs: Selection and implementation," *The Counseling Psychologist*, Vol. 35, No. 2, 2007, pp. 236–264.
[71] K. Peffers, T. Tuunanen, M.A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, Vol. 24, No. 3, 2007, pp. 45–77.
[72] R.E. Stake, *The art of case study research*. Sage Publications, 1995.
[73] S. Bennett, "What is information governance and how does it differ from data governance?" *Governance Directions*, Vol. 69, No. 8, 2017, pp. 462–467.
[74] J.W. Drisko and T. Maschi, *Content analysis*. Pocket Guide to Social Work Re, 2016.
[75] R.P. Weber, *Basic content analysis*, Quantitative Applications in the Social Sciences. Sage, 1990, No. 49.
[76] S. Elo, M. Kääriäinen, O. Kanste, T. Pölkki, K. Utriainen et al., "Qualitative content analysis: A focus on trustworthiness," *SAGE Open*, Vol. 4, No. 1, 2014.
[77] M.S. Caron, "The transformative effect of AI on the banking industry," *Banking and Finance Law Review*, Vol. 34, No. 2, 2019, pp. 169–214.
[78] M. Janssen, P. Brous, E. Estevez, L.S. Barbosa, and T. Janowski, "Data governance: Organizing data for trustworthy artificial intelligence," *Government Information Quarterly*, Vol. 37, No. 3, 2020, p. 101493.
[79] M.J. Culnan, "Policy to avoid a privacy disaster," *Journal of the Association for Information Systems*, Vol. 20, No. 6, 2019, p. 1.
[80] J. In, R. Bradley, B.C. Bichescu, and C.W. Autry, "Supply chain information governance: Toward a conceptual framework," *The International Journal of Logistics Management*, 2018.
[81] M. Hind, S. Houde, J. Martino, A. Mojsilovic, D. Piorkowski et al., "Experiences with improving the transparency of AI models and services," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–8.
[82] J.A. Kroll, "Data science data governance [ai ethics]," *IEEE Security and Privacy*, Vol. 16, No. 6, 2018, pp. 61–70.
[83] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman et al., "Model cards for model reporting," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 220–229.