Otto Aalto

Implementing an AI Agent for Logistics Compliance

School of Technology and Innovations
Master’s Thesis in Computing Sciences
Information Systems

Vaasa 2026

VAASAN YLIOPISTO
Tekniikan ja innovaatiojohtamisen akateeminen yksikkö

Tekijä:	Otto Aalto
Tutkielman nimi:	Implementing an AI Agent for Logistics Compliance
Tutkinto:	Kauppatieteiden maisteri
Koulutusohjelma:	Tietojenkäsittelytieteiden maisteriohjelma
Opintosuunta:	Tietojärjestelmätiede
Työn ohjaaja:	Teemu Mäenpää
Valmistumisvuosi:	2026 **Sivumäärä:** 64

TIIVISTELMÄ:

Tämä tutkielma tarkastelee, miten tekoälyagenttia voidaan käyttää logistiikan prosesseissa, joissa käsitellään jäsentymättömiä ja vaihtelevia asiakirjoja. Tutkimuksen tavoitteena on selvittää, voiko tekoälyagentti automatisoida toimittaja-asiakirjojen käsittelyä. Kyseinen tehtävä vaatii tällä hetkellä merkittävästi manuaalista työtä, koska asiakirjoja on paljon, ja niiden muodot vaihtelevat. Tutkimuksessa arvioidaan AI agentin toimivuutta, hyödyllisyyttä ja organisatorisia vaikutuksia vaatimustenmukaisuutta painottavassa prosessissa.

Tutkimus soveltaa suunnittelutieteellistä -metodologiaa yhdessä tapaustutkimuksen kanssa teollisen logistiikkaorganisaation kontekstissa. Tutkimuksessa suunniteltiin tekoälyagenttimalli, jonka tehtävänä oli lukea, tulkita ja luokitella toimittajien vakuutuksia sääntely- ja organisaatiovaatimusten perusteella. Agentti muodostui osaksi puolittain automatisoitua työnkulkua, ja sitä testattiin todellisilla vakuutuksilla yhdessä asiantuntijoiden kanssa, jotka vastaavat prosessin manuaalisesta käsittelystä. Arviointi keskittyi agentin kykyyn analysoida rakenteeltaan vaihtelevia asiakirjoja ja tehdä sen pohjalta päätöksiä.

Tulokset osoittavat, että tekoälyagentti kykenee luotettavasti analysoimaan ja luokittelemaan toimittajien dokumentteja, myös silloin kun dokumenttien rakenne on epäsäännöllinen tai monimutkainen. Asiantuntijat vahvistivat, että agentin tekemät päätelmät ja niiden perustelut vastasivat haluttua tasoa. Luottamusta agentin toimintaan lisättiin, kun prosessiin lisättiin päätösvarmuuteen perustuva kynnysarvo.

Tutkimuksen johtopäätöksenä on, että tekoälyagentti on käyttökelpoinen laajennus organisaation nykyisiin, RPA-pohjaisiin automaatioratkaisuihin. Siirtämällä arvioinnin tekoälylle organisaatio voi vähentää manuaalista työkuormaa ja vapauttaa henkilöstön keskittymään vaativampiin päätöstehtäviin. Tulokset osoittavat, että tekoälyn avulla voi ylittää sääntöpohjaisen automaation rajoitteet tarjoamalla kontekstuaalista päättelyä, joustavuutta ja läpinäkyvyyttä. Tutkielma tunnistaa myös jatkotutkimusmahdollisuuksia, kuten laajamittaisen arvioinnin, moniagenttiarkkitehtuurien hyödyntämisen sekä oppimisominaisuuksien lisäämisen autonomian ja pitkän aikavälin suorituskyvyn parantamiseksi.

AVAINSANAT: AI Agents, Business Process Automation, Intelligent Automation, Robotic Process Automation

UNIVERSITY OF VAASA
School of Technology and Innovations

Author:	Otto Aalto		
Title of the thesis:	Implementing an AI Agent for Logistics Compliance		
Degree:	Master of Science in Economics and Business Administration		
Degree Programme:	Information Systems		
Supervisor:	Teemu Mäenpää		
Year:	2026	**Pages:**	64

ABSTRACT:

This thesis investigates how an AI agent can be applied to logistics processes that involve unstructured and variable compliance documentation. The objective is to examine whether an AI agent can support or partially automate the screening of supplier declarations, a task that currently requires extensive manual effort due to large amounts of declarations and inconsistent formatting. The study aims to determine the feasibility, usefulness, and organizational implications of introducing AI agents into a compliance-heavy workflow.

The research applies the Design Science Research Methodology in combination with a case study within an industrial logistics organization. An AI agent system model was designed using Copilot Studio to read, interpret, and classify supplier declarations based on regulatory and organizational requirements. The agent was integrated into a workflow and tested using real declarations together with domain experts responsible for current manual processing. The evaluation focused on agent's ability to analyse heterogeneous document structures and make decisions based on them.

The results show that the AI agent can reliably analyse and classify supplier declarations, including cases with irregular or complex formatting in a controlled environment. Experts confirmed that the agent's decisions and explanations were consistent with human judgement, and that the system would take less time and effort for document screening. Introducing a confidence-based threshold further improved perceived safety and trust in the system's output, particularly given the regulatory sensitivity of the process.

The study concludes that AI agents are practical and valuable extension to existing RPA based automation solutions of the organization. By automating the initial assessment of declarations, the agent could help to reduce manual workload and enables employees to focus on decision-making tasks that require expertise. The findings highlight the potential of AI to address limitations of rule-based automation by providing contextual reasoning, adaptability, and transparency. The thesis also identifies future research opportunities, including large-scale evaluation, multi-agent architectures, and incorporating learning mechanisms to improve long-term performance and autonomy.

KEYWORDS: AI Agents, Business Process Automation, Intelligent Automation, Robotic Process Automation

Contents

Figures

Tables

1 Introduction

Artificial Intelligence has developed from being a speculative concept to a transformative force across industries, which has broadened the scope of tasks that can be automated or supported by AI. One of the promising developments is the rise of AI agents, which are autonomous or semi-autonomous systems that perceive their environment, make decisions, and execute actions to achieve specified objectives (Sapkota et al., 2026). Unlike traditional automation tools like RPA, AI can adapt, learn from context, and handle unstructured data, making them particularly valuable in applications where complexity and inconsistencies are present if used properly (Ng et al., 2021).

The logistics sector represents a compelling field for agentic automation applications. While process automation tools, like RPA have streamlined logistics processes, they are primarily deterministic and struggle with unstructured inputs; they do not provide semantic understanding or probabilistic reasoning by default (Aalst et al., 2018).

An AI agent capable of automating the filtering and classification of large amounts of supplier declaration attachments in email systems could bring substantial efficiency gains. Such an application would not only reduce human workload but could potentially also improve accuracy and response times. This makes the topic of AI agents highly relevant both academically and practically.

1.1 Background and Motivation

The motivation for this thesis arises from the gap between the capabilities of existing RPA solutions and the practical needs of logistics organizations dealing with complex, unstructured compliance processes. While prior research has extensively examined RPA, there is limited empirical work demonstrating how AI agents can be implemented in a logistics organization. Additionally, organizations require concrete design knowledge and practical examples to assess whether agentic automation is not only theoretically promising but also feasible and valuable in operational settings.

This study is motivated by both practical and academic considerations. From a practical perspective, the case organization seeks to explore whether agentic automation can enhance its existing automation capabilities by reducing manual workload, improving consistency, and increasing the scalability of compliance-related processes. From an academic perspective, the study aims to contribute to the body of knowledge on AI agents and agentic automation by demonstrating how AI agents can be designed, implemented, and embedded within real-world logistics workflows using a design science approach and using it together with other automation tools. By addressing a concrete business problem and developing a demonstrative artifact, this research seeks to understand agentic AI and its application in organizational information systems.

1.2 The Research Problem and Limitation

The purpose of this report is to demonstrate viability of AI agents for a logistics organization. To fulfil the purpose, a real prototype of an AI agent and a model of an agentic automation system will be created for a selected logistics process. This will ensure, that the technology is tested in a controlled environment. Additionally, experts from the field are involved in the testing of the solution, to give in depth knowledge about the viability. The research questions were formed according to Hoang Thuan et al. (2019, p.350) article Construction of Design Science Research Questions.

Research question 1: How can AI agents be implemented in a logistics organization
Research question 2: What are the important properties of AI agents?

The study is delimited in several ways. It focuses exclusively on the process of handling supplier declarations communicated via email, while other logistics processes fall outside the scope. The implementation is restricted to Microsoft Outlook as the communication platform, acknowledging that other email clients may require different integration methods. Additionally, the declarations will be sent out to the case company's supplier portal. The AI agent will be developed in Copilot Studio and is designed primarily to support document filtering and classification tasks. The evaluation of the agent is focused on efficiency, reliability, and feasibility in a controlled testing environment, without extending to broader organizational or macroeconomic impacts.

1.3 Approach

This research adopts a design science methodology, which is well-suited for developing and evaluating IT artifacts intended to solve practical problems. The approach involves iterative design, implementation, and evaluation, guided by both theoretical knowledge and practical constraints (Peffers et al., 2007). According to Peffers et al. (2007), the process can be summarized in four main steps:

Problem Identification: Establishing the specific logistical challenge of filtering supplier declaration attachments, that the AI agent is expected to address. Design of the AI Agent: Developing an agent architecture that integrates AI capabilities and automation within the enterprise systems. Also implementing a knowledgebase for the agent for further conceptual understanding

By building a prototype agent and embedding it into a real or simulated logistics workflow, the project moves from conceptual design to practical application. Throughout this process, the artifact is systematically developed and validated while remaining closely aligned with a real-world problem, thereby strengthening its relevance and practical value.

1.4 Expected Results

The central outcome of this work is expected to be a model of an automation system containing AI capabilities that could handle a specified task within logistics automatically. The agent functions together with more traditional automation, enabling the handling of more complex, uncertain, and variable tasks. Unlike rule-based RPA, which executes predefined instructions, the AI agent is expected to add adaptability to changing situations, make autonomous decisions, and reduce the need for manual intervention.

The results are expected to show how an AI agent can handle tasks that require reasoning in logistics context. The aim is not only to create a model of the system, but to develop a proof-of-concept and test it within the organizational context. These insights will provide guidance on the conditions under which AI agents can truly enhance automation in logistics processes.

1.5 Research Structure

The remainder of this thesis is organized into five chapters, each addressing a specific component of the research process.

Chapters 2-3: These chapters present the theoretical foundations of AI RPA, and AI agents, with a focus on their characteristics, design principles, and applications in business and logistics contexts. They also review related work focusing on both opportunities and limitations.

Chapter 4: This chapter outlines the design science research approach adopted in this study. It describes the reasoning for selecting design science, explains the iterative process of problem identification, artifact design, and evaluation, and provides details on the sourcing of research material.

Chapter 5: This chapter applies the theoretical and methodological considerations to a concrete problem: filtering supplier declaration attachments in Microsoft Outlook. It presents the design, implementation, and testing of the AI agent prototype, followed by an evaluation of its performance in terms of accuracy, efficiency, and usability. The agent uses GPT-4 as its LLM.

Chapter 6: The final chapter summarizes the key findings of the research, discusses the contributions to both theory and practice, and reflects on the limitations of the study. It also provides suggestions for future research directions, particularly regarding the broader application of AI agents in logistics and beyond.

By following this structure, the thesis ensures a progression from theoretical grounding to practical implementation and evaluation, offering both academic and practically applied insights into the use of AI agents in logistics processes.

2 Theoretical Background of AI and AI Agents

Artificial Intelligence (AI) is a field of computer science focused on the design and study of systems capable of performing tasks that normally require human intelligence. The development of AI has progressed significantly. Early approaches relied heavily on symbolic reasoning and explicitly programmed rules, which were effective in narrowly defined problem spaces but limited in their ability to handle ambiguity and scale. With the advent of machine learning, the focus shifted toward data-driven methods that enable systems to learn patterns and representations from data rather than relying solely on predefined rules. This transition has been supported by advances in computational power and the availability of large datasets, leading to significant improvements in AI performance and applicability across domains such as business, healthcare, and education. (Fui-Hoon Nah et al., 2023).

More recently, deep learning techniques based on neural networks have become central to AI development. These methods allow systems to automatically learn hierarchical representations of data, resulting in performance in tasks involving perception, prediction, and decision-making. Despite these advances, the increasing deployment of AI systems has raised concerns related to transparency, accountability, bias, and societal impact (Capel & Brereton, 2023).

Large language models are also a central topic related to AI. They work by estimating the probability of word sequences based on observed data. LLMs support both language understanding and language generation tasks (Warudkar & Jalit, 2024). As computational resources and text data increased, language models grew, and increased in complexity, leading to substantial performance improvements. Contemporary NLP research increasingly relies on large-scale language models, which have redefined the state of the art in many core NLP tasks (Zubiaga, 2024).

LLMs are neural language models trained on extremely large and diverse text corpora, often containing billions of words. Their scale allows them to capture complex statistical patterns in language, enabling them to perform a wide range of tasks without task-specific training. LLMs achieve strong performance across tasks such as text generation, summarization, question answering, and conversational interaction (Bandi et al., 2023). The capabilities of LLMs go beyond just word generation, as development in LLMs has allowed AI to be used in workflow automation and orchestration, moving from traditional Robotic Process Automation to Agentic Process Automation (Fan et al., 2024, p. 1-2).

LLMs are a core component of generative AI systems and can produce fluent, coherent, and contextually appropriate text that often resembles human writing. However, their capabilities also introduce notable limitations and risks. LLMs may generate factually incorrect or misleading information, reproduce biases present in their training data, and be misused for malicious purposes such as misinformation, fraud, or automated manipulation. These concerns have led to growing research attention on governance, transparency, and responsible use of LLM-based systems (Ferrara, 2024). Amodei et al., (2016) pointed out that it is important to ensure privacy, fairness and security when using AI.

2.1 AI Agents

AI agents can perceive their environment, reasoning over observations, and executing actions autonomously to achieve specified goals. AI agents are typically framed as adaptive, goal-oriented actors within systems. Agents need to be customized for their specific tasks (Wali et al., 2023). While traditional automation, such as Robotic Process Automation (RPA), focuses on repetitive and deterministic tasks, the integration of AI enables automation to address complex, data-driven processes that require perception, judgment, and adaptability (Aalst et al., 2018, p. 269-270).

AI agents are designed to perform a spectrum of cognitive tasks, including data analysis, prediction, classification, and contextual decision-making. These agents can process both structured and unstructured data, allowing systems to recognize patterns, interpret human input, and optimize outcomes dynamically (Afrin et al., 2025). In addition to reasoning and decision-making, AI agents may can invoke external tools, like APIs to interact with other systems (Shetty et al., 2024). Agents are well-suited for more complex and realistic tasks that carry practical value in real-world settings, often involving problems where there is no single correct solution (Kapoor et al., 2024).

Agents are typically capable of perception, reasoning, and action (Afrin et al., 2025). Agents can gather and interpret data, while reasoning supports planning and decision-making based on that information. AI agents can also learn, which allows them to further adapt into changing situations (Shinn et al., 2023). AI agents also vary in complexity and purpose. Simple agents operate on more simple condition–action rules, while model-based reflex and goal-based agents incorporate internal representations of their environments to support more informed decision-making (Afrin et al., 2025; Russell & Norvig, 2021).

By embedding AI agents within business systems, organizations achieve enhanced scalability, accuracy, and operational efficiency. However, this advancement also introduces ethical and governance challenges, including accountability, transparency, and data privacy (Afrin et al., 2025, p. 183).

AI agents can serve as the cognitive core automation, bridging the gap between mechanical task execution and autonomous decision-making. Their evolution reflects a broader shift toward adaptive, learning-based systems that augment human capabilities and enable continuous innovation across industries (Afrin et al., 2025, p. 188–189; Aalst et al., 2018).

The autonomy of AI agents also introduces security risks that do not arise in traditional rule-based automation. Autonomous agents can coordinate with other agents, accumulate privileges, and adapt their behaviour over time, creating risks such as collusive escalation and agentic drift (Zouari, 2025). These can amplify errors or misuse and make harmful behaviour harder to detect and correct. Ning et al. (2025) also raises the concern about AI agents handling confidential data in case they interact with compromised websites.

AI agents can also be subject to issues that AI generally has. This means that identical inputs may lead to different outcomes, introducing inconsistency. At the same time, genuine bias can still emerge and propagate across agent interactions, especially when outputs from one agent serve as inputs to another (Condon & Jilani, 2026).

2.2 Important Properties of AI Agents

Sapkota et al. identified three defining properties of AI agents that are also supported by other literature. According to Sapkota et al. (2026) autonomy is regarded as one of the defining properties of an AI agent. Autonomy refers to an AI agent operating with minimal to no human intervention and exercises control over its actions (Afrin et al., 2025; Sapkota et al., 2026). Autonomy distinguishes agents from passive tools or subroutines: an agent decides whether, when, and how to act, rather than simply executing predefined commands. Agents are able to plan action sequences and adapt dynamically (Russell & Norvig, 2021; Sapkota et al., 2026).

AI agents are designed to handle specific, well-defined tasks (Afrin et al., 2025; Ning et al., 2025; Sapkota et al., 2026). They excel at performing repeatable operations within a limited scope, such as filtering emails, querying databases, or managing calendars. By focusing on narrow functions, these agents achieve greater efficiency, clarity, and precision, making them especially effective in situations, where broad, general-purpose reasoning would be unnecessary or less efficient (Sapkota et al., 2026)

AI agents also demonstrate reactivity and adaptability, allowing them to respond dynamically to changes in their environment (Sapkota et al., 2026). Rather than following a fixed sequence of instructions, they can adjust their behaviour based on new inputs, feedback, or evolving conditions. This enables them to handle uncertainty, recover from unexpected events, and refine their actions over time. Such flexibility makes them well-suited for environments where conditions are not static and require continuous adjustment rather than rigid execution.

Table 1. Important Properties of AI Agents.

Important Properties of AI Agents
Autonomy
Task-specificity
Reactivity and Adaptation

2.3 Multi-Agent Systems

The advancement of AI has allowed development of multiple AI agents capable of goal-directed decision-making (Bo et al., 2024). AI agent systems can be distinguished between single-agent systems and multi-agent systems. Single agents focus on individual autonomy and task completion, whereas multi-agent systems allow coordination across multiple agents (Bellogín et al., 2025). Multi-agent systems leverage distributed reasoning, and role specialization to solve problems that exceed the capacity of any individual agent (Naik et al., 2025).

In these systems, each agent has a specific role or capability and they collaborate or compete through structured communication protocols (Batra et al., 2025). This design allows agents to share intermediate reasoning steps, debate hypotheses, and cross-verify results, thereby potentially improving factual accuracy, reasoning depth, and creative diversity.

The transition from single-agent to multi-agent systems is primarily motivated by the limitations of isolated LLMs in handling multi-step reasoning, long-horizon planning, and error detection. By allowing multiple LLMs to engage in structured dialogue, systems can encourage divergent thinking and self-correction. Prior works demonstrate that collaborative or competitive interaction among agents, such as through debate, critique, or role-based workflows leads to improved reasoning quality and factual accuracy across domains including scientific question answering, software development, and social simulation.

Multi-agent systems introduce challenges related to unpredictability, and reduced control (Bellogín et al., 2025). However, they also enable scalability and robustness through redundancy and specialization. Naik et al. (2025) mentions that effective multi-agent systems require careful design that balance agent autonomy with human oversight, transparency, and explainability, which elements increasingly recognized as core components of agent theory rather than implementation details.

2.4 AI Agents and Agentic AI

According to Sapkota et al. (2026) the term AI Agents refers to autonomous software entities designed to perform specific goal-oriented tasks within defined digital environments. These agents perceive inputs, reason over contextual information, and act toward a user-defined objectives (Sapkota et al., 2026, p. 4). Unlike simple automation scripts that follow rigid rules, AI Agents demonstrate adaptive and reactive behaviour, such as learning from feedback or dynamically configuring outputs. They can rely on large language models and large image models for reasoning and perception, supporting applications such as customer service, scheduling, and organizational information retrieval (Sapkota et al., 2026, p. 5–6).

Agentic AI, on the other hand, can be seen as a step beyond individual AI Agents. Agentic AI refers to systems that are designed to achieve more complex goals, typically consisting of multiple agents working together towards a goal (Sapkota et al., 2026, p. 7–9, Bhattaram et al., 2025, p. 1-2). Typically agentic AI is capable to achieve goals with little to no human intervention (Wadinambiarachchi et al., 2025, p. 20). These systems integrate advanced mechanisms such as dynamic task decomposition, persistent memory, and orchestrators that manage subordinate agents and resolve conflicts. Unlike single AI

Agents, Agentic AI systems are characterized by distributed intelligence, goal alignment, and contextual adaptation (Sapkota et al., 2026, p. 8).

One of the main differences between AI Agents and Agentic AI lies in their scope and autonomy. AI Agents are typically single systems optimized for discrete, well-defined tasks, while Agentic AI systems coordinate multiple specialized agents to perform complex, multi-step workflows (Sapkota et al., 2026, Table 1).

In terms of applications, AI Agents are prevalent in customer support automation, internal enterprise search, email prioritization, content recommendation, and scheduling (Sapkota et al., 2026, p. 13). In contrast, Agentic AI systems are applied in more advanced domains requiring coordination and adaptability (Sapkota et al., 2026, p. 13–14). Through this progression, the study positions Agentic AI as the next stage of intelligent infrastructure, which is capable not only of executing predefined tasks but also of constructing, revising, and managing complex objectives with minimal human supervision.

3 Business Process Automation

Business process automation refers to the replacement of manually performed and regularly occurring transactions with automated digital systems supported by technologies such as AI and RPA tools (Blahušiaková, 2023). In recent years automation has increased, as many companies are aiming to digitalize workflows that previously relied on physical documentation and in-person procedures (Blahušiaková, 2023).

Business process automation is especially relevant in competitive and regulated environments, where organizations must strengthen their business processes while reducing risks and operational costs. Automation solutions are described as playing a vital role in improving efficiency, service quality, compliance, and risk mitigation (Silvares et al., 2024). Furthermore, as enterprises increasingly collaborate with other organizations and face reporting and supervisory requirements, the ability to share information effectively becomes more significant, raising concerns about compatibility and interoperability between automation systems (Silvares et al., 2024).

One common method of business process automation is RPA, which is commonly understood as a software method for automating routine and rule-based tasks that were previously performed by humans. RPA operates through software that can interact with digital systems through user interfaces in ways similar to human workers (Schlegel et al., 2024). RPA can provide benefits such as increased productivity, time and cost savings, and reductions in manual errors (Schlegel et al., 2024). At the same time, automation does not necessarily require extensive investments in reprogramming or building new system interfaces, which makes it an attractive option for organizations seeking efficiency improvements (Schlegel et al., 2024).

In practice, BPA is applied in several organizational functions, particularly in administrative and accounting-related processes. Automation is also closely linked to Enterprise Resource Planning (ERP) systems, which integrate business transactions and coordinate organizational resources through shared data infrastructures (Blahušiaková, 2023).

Processes such as order tracking, payroll processing, inventory management, and invoicing are identified as key candidates for automation due to their repetitive and structured nature (Blahušiaková, 2023).

There are also challenges in business process automation implementation. One major issue concerns the lack of standardization across automation tools and descriptions. Since different automation solutions often adopt various methods, incompatibility can arise, negatively affecting communication, quality, and productivity (Silvares et al., 2024). For this reason, adopting standard specifications or widely accepted best practices is viewed as beneficial. Automation can also cause automation bias, overreliance, and even skill decay due to overuse (Haase et al., 2024).

3.1 Robotic Process Automation

This section explains what Robotic Process Automation (RPA) is, where and how it is used, and its general impacts. The aim of the section is to present the information and concepts related to software robotics that are essential for this study.

RPA is a term used to describe tools that perform tasks in a manner like humans by interacting with a computers or a comparable device's user interface. Its primary purpose is to automate tasks without requiring significant changes to existing information systems (Aalst et al., 2018). One key advantage of software robots is that a single robot can operate across multiple independent information systems, unlike traditional automation solutions that are often limited to a single system (Aalst et al., 2018).

Successful implementation of RPA improves business efficiency, productivity, information security, processing times, and accuracy (Pramod, 2022). It accelerates processes while reducing human workload and the likelihood of errors (Aalst et al., 2018). The adoption of RPA is considered particularly suitable when tasks are digital, repetitive, and consume significant human resources (Ratia et al., 2018). According to Ratia et al. (2018), the value added by RPA is especially evident in improved organizational efficiency, as properly implemented RPA reduce the need for employees to focus on routine tasks and enhance the overall quality of work. Their ability to increase productivity and reduce costs is therefore a major driver behind the adoption of RPA.

Currently, RPA is one of the easiest and fastest forms of automation to implement, which has led to its popularity growing significantly compared to many other digitalization approaches (Sliż et al., 2024). RPA is also regarded as one of the most critical specialized skills required to achieve digital transformation in organizations (Pramod, 2022).

Although software robots imitate human actions, they cannot perform all tasks within information systems in the same way humans can. Tasks suitable for software robots typically meet the following criteria: they do not require judgment, creativity, or interpretative ability; they involve high transaction volumes; they require interaction with multiple information systems and applications; they are highly standardized with minimal exceptions; and they are prone to human error (Moraes et al., 2022).

The extent of software robotics usage can be assessed in terms of both depth and breadth (Zhu & Kanjanamekanant, 2023). In the context of software robotics, depth refers to how extensively a single task can be automated. Deep automation means that a task is transferred almost entirely to a software robot. Breadth, in turn, refers to how many different tasks utilize software robots.

RPA can be enhanced with AI and ML to evolve from rule-based automation into intelligent automation, enabling the handling of unstructured data, probabilistic decision-making, and adaptive process flows (Kedziora & Hyrynsalmi, 2023; Moraes et al., 2022). Integrating AI elements like ML, optical character recognition, and natural language processing allows bots to process varied inputs such as invoices or documents, classify discrepancies with confidence scores, and route exceptions for human review when needed (Bavaresco et al., 2023; Kedziora & Hyrynsalmi, 2023)

Such integrations widen the scope of automatable processes for RPA while emphasizing the importance of well-mapped workflows, robust training, and human oversight to mitigate risks like model biases or process changes (Moraes et al., 2022).

3.2 Agentic Process Automation

Ye et al. (2023) introduce Agentic Process Automation as an extension of traditional automation that integrates AI agents into workflows. In contrast to traditional automation, like RPA, which automates predefined, rule-based tasks within static workflows, Agentic Process Automation enables systems to dynamically construct, modify, and execute processes. This shift allows automation to operate effectively in environments characterized by unstructured data, uncertainty, and changing objectives, thereby reducing reliance on pre-programmed rules.

At a conceptual level, Agentic Process Automation reflects a broader transition from task-centric automation toward goal-oriented, autonomous behaviour. Rather than merely following scripted instructions, agentic systems can reason about objectives, decomposing tasks, selecting appropriate tools, and coordinating actions across multiple steps. This aligns with recent distinctions in the literature between conventional AI agents and agentic AI, where the latter emphasizes adaptive autonomy, contextual awareness, and the orchestration of complex activities over time (Sapkota et al., 2026).

Agentic Process Automation represents a shift toward automation systems that exhibit elements of decision-making and self-directed action. It is a transition towards cognitive automation, where systems are designed to interpret context, reason over alternatives, and adapt their behaviour accordingly (Ye et al., 2023). Central to this paradigm is the notion of AI agents that can leverage existing digital tools in flexible ways rather than operating within scripted instructions. Such capabilities fundamentally exceed the design limits of traditional RPA systems, which lack the ability to redesign or extend their own operational logic at runtime (Aalst et al., 2018).

As Agentic Process Automation becomes more cognitively capable, the role of human workers shifts away from routine execution towards, guidance and oversight, while focusing more on even complex tasks (Ye et al., 2023). Ye et al. (2023) emphasize that the core focus of automation is not to replace humans. They also raised concerns about automation bias, since more capable automation can increase over-resilience. Humans may start to trust AI based views even when they are conflicting with their own.

3.3 Human in the Loop

Human-in-the-loop (HITL) is a technique, where a human is involved with helping a system achieve its goal (Memarian & Doleck, 2024, p. 2). As automated systems take on more complex tasks, human involvement remains important for handling ambiguity, confirming uncertain outputs, and ensuring that decisions remain aligned with expectations (Rahman, 2026). In this sense, HITL is not just a safeguard against error; it is a way to make automation more reliable, adaptable, and usable in real-world settings (Wu et al., 2022).

From a general perspective, automation works best when it can handle routine, high-volume, or well-defined tasks, while humans step in for exceptions, judgment calls, and contextual interpretation. HITL supports this division of labor by allowing systems to automate what they can confidently do and defer to people when additional information or review is needed (Rahman et al., 2026). This makes automation less rigid and more responsive to changing conditions, edge cases, and domain-specific requirements. While

AI agents are typically autonomous systems, they can require minimal human intervention (Sapkota et al., 2026, p. 5). However AI can fail due to hallucinations or planning errors (Wang et al., 2024, p. 8), which justifies HITL in some cases where AI cannot be fully trusted. HITL can be integrated into automated workflows, where humans provide domain knowledge, while machines contribute speed (Wu et al., 2022).

4 Research Approach

This thesis adopts a combined research approach that integrates design science methodology with case study research. This dual strategy is appropriate given the aim of both creating a practical solution and studying its development and application within a real organizational context. Design science provides the structured process needed to build and evaluate an artifact while ensuring that the solution is both relevant to practice and grounded in scientific rigor (Hevner et al., 2004). At the same time, the case study approach enables an in-depth examination of how the artifact is shaped by, interacts with, and influences the dynamics of the company in which it is implemented (Schell, 1992).

By combining design science with a case study, the research benefits from both methodological strengths: the systematic construction and evaluation of a solution and the deep understanding of its practical context. This approach supports the contribution of the thesis, which is advancing theoretical knowledge through the design of an artifact and generating empirical insights through its application in a real-world organizational case.

The following sections describe each method in detail, outline how they are applied in this thesis, and explain how they complement one another throughout the research process.

4.1 Case Study

This thesis can be defined as a case study since it investigates a contemporary phenomenon within its real-life organizational context, in line with Schell's (1992) description of case study research. A case study is characterized by examining a phenomenon where the boundaries between the subject and its context are not clearly evident and where multiple sources of evidence are used to understand “how” and “why” events unfold within their natural setting (Schell, 1992). In this thesis, the research is situated inside a single company, focusing on current practices and processes, which corresponds to the conditions under which case studies are considered appropriate. Therefore, given its focus on an organizational phenomenon, its use of multiple data sources, and its emphasis on understanding processes within their actual environment, this thesis meets the criteria for a case study.

At the same time, the study is positioned within the design science, which focuses on the creation and evaluation of artifacts intended to solve identified organizational problems (Peffers et al., 2007, p. 49). Design science research is concerned with developing explicitly applicable solutions and demonstrating their utility in practice (Peffers et al., 2007, p. 47). This aligns with the objective of the present thesis, which requires designing and developing an artifact to address a specific business problem.

The combination of a case study strategy with a design science approach can be used to demonstrate designed artifacts in real-world contexts (Peffers et al., 2007, p. 57-70). In such settings, the organizational environment serves both as the context for problem identification and as a testbed for demonstrating and assessing the effectiveness of the proposed solution. As outlined in the design science research methodology, this involves activities such as problem identification, design and development, demonstration, and evaluation, which can be carried out within a case setting (Peffers et al., 2007, p. 57-70).

This thesis adopts a case study to support the iterative development and demonstration, and evaluation of the AI agent within its natural context. This dual positioning strengthens the methodological rigor by combining contextual analysis with the systematic creation and assessment of an artifact.

4.2 Design Science

Design Science Research (DSR) is a research method that focuses on the creation and evaluation of artifacts designed to solve real-world problems. Unlike behavioural science, which aims to explain or predict phenomena, DSR seeks to produce prescriptive knowledge about how things can or should be designed to achieve desired outcomes. In the context of information systems research, DSR extends the boundaries of human and organizational capabilities by building new and useful artifacts such as models, methods, software, or systems (Hevner et al., 2004). The goal of DSR is not only to design effective solutions but also to generate knowledge about the principles underlying those designs, contributing to the broader scientific understanding of how technological artifacts can be developed and used effectively (Vom Brocke & Maedche, 2019).

DSR can be described as comprising three interconnected research cycles that together define its process (Hevner, 2007). The relevance cycle connects the research to the real-world environment, ensuring that the problem addressed and the requirements identified are meaningful and practically significant. The rigor cycle links the research to the existing body of scientific knowledge, drawing on established theories and methods to inform the design process while contributing new knowledge back into the knowledge base. Between these two, the design cycle represents the core iterative process of constructing and evaluating artifacts. Through repeated cycles of building and testing, researchers refine their solutions and strengthen the knowledge claims associated with them. This cyclical structure ensures that DSR remains both theoretically grounded and practically useful, balancing scientific rigor with tangible relevance to real-world problems (Hevner, 2007; Winter, 2008).

Design Science Research can be used as a tool when doing research about the development of an AI agent because it provides a structured yet flexible framework for designing, implementing, and evaluating technological artifacts (Peffers et al., 2007), AI agent itself being the artifact of this research. As Peffers et al. (2007) explain, DSR involves a sequence of stages by identifying and motivating the problem, defining the objectives for a solution, designing and developing the artifact, demonstrating its use, evaluating its performance, and communicating the results. This structure aligns with the process of developing an AI agent, where the researcher first identifies a practical or theoretical problem that the agent will address, then iteratively builds and tests the system to achieve measurable improvements in performance or decision-making. Using DSR ensures that the development of the AI agent is not just a technical exercise but a research process that generates validated, generalizable knowledge about how such agents can be designed and applied effectively.

DSR emphasizes the importance of evaluation as an integral part of the research process. According to (Venable et al., 2016), evaluation in DSR serves to assess both the utility of the artifact in addressing the intended problem and the rigor of the research process itself. By systematically designing and conducting evaluation activities researchers can demonstrate the scientific validity and practical relevance of their artifacts. This study follows the Quick & Simple evaluation strategy. According to Venable et al. (2016) this approach is suitable when the artefact is relatively small, and design risk is limited. As the goal is to provide a proof of concept and the time frame to create this thesis is too short to perform a thorough evaluation, this approach was chosen.

Using DSR will help with combining theory and practice. It enables the creation of a technological artifact that addresses a real-world problem while simultaneously contributing to academic knowledge about AI agent design and application. Gregor & Hevner (2013) note that DSR contributions can take multiple forms, ranging from concrete instantiations such as an implemented system to more abstract design principles or theories. In the case of an AI agent, the artifact itself may demonstrate a novel design approach or capability, while the accompanying design knowledge can inform future research and practice. DSR supports a research process that is iterative, reflective, and generative, positioning the AI agent not just as a functional tool but as a scientific contribution to the field of information systems.

This research follows the Design Science Research Methodology (DSRM) introduced by Peffers et al. (2007), which provides a structured approach for developing and presenting an artifact aimed at solving a relevant information systems problem. The methodology consists of six core activities: problem identification and motivation, defining objectives for a solution, design and development, demonstration, evaluation, and communication. These steps ensure that the research progresses logically from understanding the problem to creating and showing the utility of an artifact and finally disseminating the results to appropriate audiences.

However, while DSRM prescribes all six activities, the evaluation step will not be carried out in full scale in this study due to time constraints. Although evaluation is an essential activity for assessing how well the artifact meets the defined objectives (Peffers et al., 2007), the available timeframe prevents the execution of a rigorous evaluation process. However, the remaining steps allow for a structured and methodologically grounded development and presentation of the artifact.

In Design Science Research, the first step involves defining the specific research problem and establishing why it is worthwhile to solve. Peffers et al. (2007) explain that this step requires the researcher to articulate the problem clearly and conceptually, often by decomposing it into smaller elements so its complexity can be adequately captured. The motivation aspect serves two purposes: it justifies the significance of the problem to both researchers and practitioners, and it clarifies the reasoning behind the chosen research focus (Peffers et al., 2007, p. 53). A well-justified problem definition also ensures that the developed artifact will be aimed at a relevant and meaningful need.

Once the problem is specified, the next step is to infer and articulate the objectives that a successful solution should meet. These objectives may be quantitative or qualitative. Peffers et al. (2007) emphasize that solution objectives must be derived rationally from the problem analysis and informed by what is known to be feasible given existing technologies and theories. The step ensures that the design activity is grounded in explicit performance expectations, which later serve as the basis for evaluation (p. 53).

The next step consists of design and development of the artifact, which in this case is the AI agent. According to Peffers et al. (2007), designing and developing the artifact involves defining its desired functionality and architecture, followed by constructing the actual artifact (p. 54). The authors conceptualize design as a search process that applies theoretical and practical knowledge to achieve the defined objectives. The artifact must embed a research contribution and be capable of addressing the identified problem in a meaningful way.

After the artifact is created, it's demonstrated how the artifact solves the original problem. Peffers et al. (2007) describe this as showing the artifact in use through appropriate methods such as experimentation, simulation, case studies, or proofs (p. 55). The purpose of this step is not full evaluation, but rather to illustrate feasibility and utility, that the artifact can indeed be applied in practice to address the problem conditions. It requires sufficient knowledge of how to use the artifact effectively.

The final step consists of communicating the research problem, the artifact, the design process, and the evaluation results to relevant scholarly and professional audiences. According to Peffers et al. (2007), effective communication includes explaining the problem's importance, the novelty and utility of the artifact, the rigor of the design process, and the evidence supporting its effectiveness (p. 56). This step ensures that knowledge generated through Design Science Research is disseminated and can be understood, critiqued, and applied by the research community and practitioners. In this case this step is this report itself.

4.3 Acquisition and Analysis of Research Material

To ensure that the theoretical foundation of this thesis is comprehensive, reliable, and aligned with current developments in artificial intelligence, a systematic process was followed in acquiring relevant research material. The literature review began with an exploratory search to map the amount of existing work on AI agents. This initial scan helped define the key concepts, identify gaps in current knowledge, and refine the research focus of the thesis.

Building on this foundation, a more structured literature search was conducted using major scientific databases, including the ACM Digital Library, AIS eLibrary, IEEE Xplore, Web of Science, and Google Scholar. These databases were selected because they collectively cover a wide range of publications in computer science, information systems, artificial intelligence, and operations management helping with finding relevant material to the development of AI agents for logistics processes. Multiple databases were required, since at this time research of agentic automation and agents is relatively sparse. Searches were performed using combinations of keywords such as “AI agents”, “intelligent automation”, “agentic automation”, “process automation” and “design science”. Boolean operators and database-specific filters were applied to refine the results based on publication year, and relevance.

In addition to academic sources, domain experts were consulted to complement the practical implications during testing of the proof-of-concept. This is important given the thesis is connected to a practical issue in a logistics organization. The experts provide domain specific understanding to validate the results of testing. Also REACH regulation was used as s source duo to its link to the process

Selected material was evaluated based on credibility, methodological rigor, and relevance to the research objectives. Priority was given to peer-reviewed journal articles, conference papers, and authoritative books; however, due to the rapidly evolving nature of AI technologies, also non–peer-reviewed sources are also included when they offered up-to-date insights not yet covered in academic research. These sources were used primarily to complement, rather than replace, the peer-reviewed literature. Through this combination of systematic academic search and targeted industry exploration, the research material collected offers a versatile background about AI agents, business process automation and design science.

4.4 Research Process Overview

This chapter explains how the research process is systematically structured in accordance with the DSRM proposed by Peffers et al. (2007). Each section of the chapter corresponds to a specific DSRM step except for section 5.1, which explains the research context and environment. This ensures consistency between the theoretical research framework and the practical implementation of the artifact. Section 5.2 addresses the identification of the problem and its motivation by describing the organizational environment and compliance-processing context, highlighting the limitations of traditional automation approaches and justifying the need for an AI-based solution. Section 5.3 translates this problem into clearly defined objectives, specifying what the artifact must achieve to address the identified challenges. Section 5.4 focuses on the design and development of the artifact, where the conceptual solution is transformed into a functional AI agent system using appropriate tools and technologies. Section 5.5 prepares and conducts the demonstration of the artifact within its intended context, providing a proof of concept that illustrates its applicability. Finally, Section 5.6 evaluates the artifact based on stakeholder feedback and defined criteria, assessing the extent to which the proposed artifact fulfils its objectives.

Table 2. Research Process Overview.

Chapter Section	**Related DSRM step (Peffers et al., 2007)**	**What is Done in Practice**
5.2 Problem Definition	Step 1: Problem Identification and Motivation	The organizational environment and compliance-processing context are introduced. Inability of traditional automation is assessed, resulting in need of AI.
5.3 Define Objectives	Step 2: Define Objectives of a solution	Solution goals are derived from the problem. In practice, this step defines what the artifact must achieve.
5.4 Design and Develop	Step 3: Design and Development	The artifact is designed and constructed. In practice, an AI agent system is developed using the required tools.
5.5 Demonstration	Step 4: Demonstration	The chapter prepares for applying the artifact in context. The artifact will be tested in context, and a chart will be introduced for proof of concept. The artifact will be evaluated based on comments of stakeholders.

5 Implementation of an AI Agent

The following sections build a structured foundation for the design and evaluation of the proposed solution based on DSRM steps proposed by Peffers et a. (2007). First, Section 5.1 introduces the industrial background and research context, situating the study within a logistics organization and outlining the selected compliance process. Section 5.2 then defines the core problem, highlighting the limitations of manual review and traditional automation in handling unstructured regulatory documents.

Section 5.3 establishes the objectives of the research, specifying the functional and organizational goals of the proposed solution as well as the key properties required of the AI agent. Section 5.4 presents the design and development of the artifact, describing its architecture, components, and role within the existing automation ecosystem.

Finally, Section 5.5 demonstrates the implemented solution through practical examples and provides an initial evaluation of the artifact in a realistic setting, thereby linking the conceptual design to its practical applicability.

5.1 Background and Research Context

This study is made as a collaboration with an industrial manufacturing company. The research environment is a logistics organization that specializes in spare parts sales and distribution to external customers. The organization handles a variety of supply chain and logistics processes that aim to deliver parts to customers efficiently. The organization utilizes already RPA to automate business processes to a large scale for repetitive and rule-based tasks. The purpose of this project is to determine, if AI agents are a feasible to extend automation capabilities together with RPA. Additionally, the organization wants an evaluation of Copilot Studio against other AI agent platforms, which affected the choice of tools in this study.

Within the organization, a specific process has been chosen to work as a platform to create a proof of concept of an AI agent in automation. The process begins with suppliers sending supplier declarations about REACH certifications via email. The Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Regulation is the European Union's main legal framework for managing chemical substances and their potential risks. It requires businesses to register detailed information about chemical substances, allows authorities to evaluate associated risks, and provides mechanisms to authorise or restrict hazardous substances where necessary (European Commission, n.d.).

The employees need to read the declarations one by one and determine if they include all the needed data and comply with the company's policies regarding related standards. However, the number of declarations is large, and the declarations are provided in an inconsistent format. Thus, the existing automation solution, RPA, cannot be used for this process, but AI agents on the other hand are designed to handle such information with the help of AI technology.

The aim is to provide an AI solution to go through the declarations and filter them, which would improve efficiency and provide the employees more time to focus on more creative and demanding tasks. After the filtering, an AI agent would trigger an RPA to process the sufficient declarations and inform the employees about the insufficient ones.

5.2 Problem Definition

The case organization relies on suppliers to provide regulatory and compliance-related documentation, such as declarations concerning REACH. In the studied organization, this process currently begins with suppliers submitting such declarations via email. Employees are then required to manually review each declaration to determine whether it contains all required information and complies with internal policies and applicable standards.

This manual review process has several challenges. First, the volume of incoming supplier declarations is high, making the task time-consuming and intensive. Second, the declarations are provided in highly inconsistent formats, including varying document structures, terminology, and levels of detail. As a result, employees must read and interpret each declaration individually, which increases cognitive workload and the risk of human error. Third, due to the unstructured and semi-structured nature of the documents, the organization's existing automation solution based on RPA, is not suitable for this task. RPA relies on deterministic rules and structured inputs and therefore lack the ability to interpret natural language, reason about compliance requirements, or adapt to document variability (Aalst et al., 2018).

In Figure 1 is a chart describing the process that the artifact is designed to automate. The plan is to design an artifact, that can handle the process from beginning, until the declaration is handed either to a materials expert for further evaluation, or RPA to upload it directly to supplier portal.

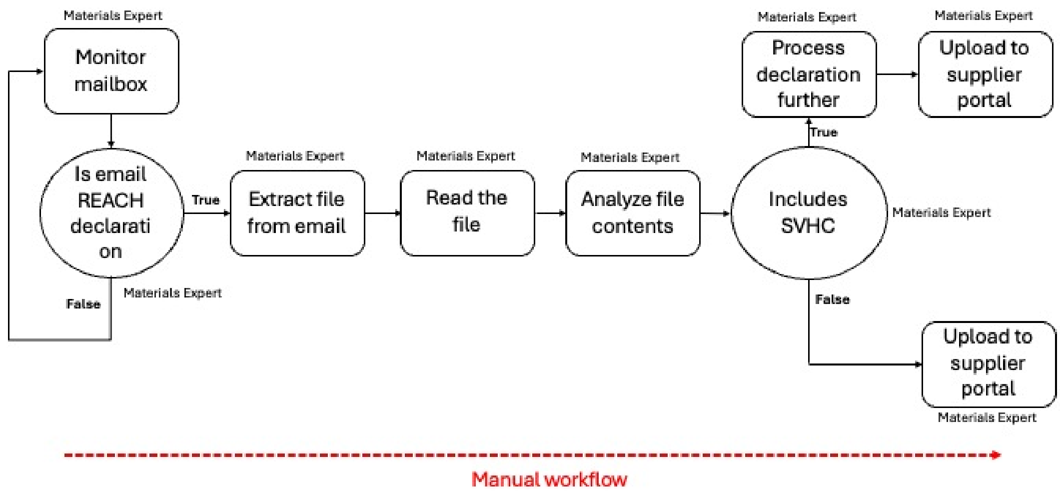

Figure 1. Workflow of the Current Process

As seen from the chart in Figure 1, the current process is entirely manual. These limitations create inefficiencies in the current process and prevent scalability as the number of suppliers and regulatory requirements continues to grow. Consequently, employees spend a significant amount of time on repetitive screening tasks rather than focusing on more complex, value-adding activities such as exception handling, supplier communication, or compliance analysis. This situation highlights a clear problem: the organization lacks an effective automation solution capable of handling unstructured compliance documents and supporting intelligent decision-making.

AI agents offer a promising approach to addressing this problem. AI agents, operating under policy constraints, can reason systematically in uncertain conditions, and adjust decisions (Radanliev et al., 2026). Therefore, the identified problem motivates the design and development of an AI agent model as an artifact that can support and partially automate the supplier declaration review process.

5.3 Define Objectives

The objective of this research is to design and develop an architecture for a solution that supports the automated processing of supplier declarations related to REACH certifications. The proposed solution aims to address the limitations of the current manual review process by introducing AI capabilities that enable the interpretation, assessment, and filtering of unstructured and semi-structured compliance documents.

We considered whether to adopt a single-agent or agentic multi-agent system, ultimately deciding on a single-agent approach. Potentially, the solution could have included two separate agents: one dedicated to email classification and another responsible for reasoning and decision-making. However, upon further evaluation traditional automation tools proved to already be effective at handling email filtering tasks in this case, when the mail filtering could be done with predefined rules. As a result, introducing a dedicated AI agent for classification would add unnecessary complexity without significant benefit. Therefore, the design was simplified to rely on traditional automation for email handling, while focusing a single AI agent on the evaluation and decision-making step.

From an organizational perspective, the solution aims to improve operational efficiency, and consistency in the handling of supplier declarations. By automating the initial screening and filtering tasks, the AI agent is expected to reduce processing time and cognitive workload for employees in the future, allowing them to focus on more complex, creative, and value-adding activities.

The artifact is aimed to make significant changes to the process of handling supplier declarations. The changes will be mainly the reduced manual work that the process requires. At functional level, the artifact is intended to analyse supplier declarations and analyze information relevant to regulatory requirements. The agent should be capable of interpreting document content despite variations in structure, terminology, and level of detail. This capability is essential for handling the diversity of supplier-provided documentation that cannot be processed using traditional rule-based automation due to inconsistencies and complexity.

In addition to information extraction, the AI agent is designed to evaluate whether each declaration complies with predefined requirements and internal policies concerning REACH standards. This evaluation involves determining the completeness and adequacy of the provided information rather than merely identifying the presence of specific keywords or fields. Based on this assessment, the artifact should autonomously classify declarations as sufficient or insufficient, thereby enabling differentiated downstream handling.

A further objective of the solution is to integrate seamlessly with the organization's existing automation infrastructure. For declarations classified as sufficient, the AI agent should trigger an RPA workflow to continue processing the documents within established enterprise systems. In contrast, declarations identified as insufficient should be escalated to employees, accompanied by contextual information that supports efficient human review and decision-making. This hybrid approach ensures that the strengths of both AI agents and RPA are utilized while maintaining appropriate human oversight.

The solution should be able to operate with minimal human intervention while remaining aligned with organizational policies, perceive and interpret its environment through contextual data, and react promptly to changing conditions. It must be capable of reasoning-based decision-making rather than fixed rules and ideally improve over time through learning and adaptation. The agent should also leverage tools effectively and maintain predictable, controlled, and transparent behaviour to ensure safety.

5.4 Design and Develop

The AI agent developed in this study is designed to perform document interpretation, classification, and workflow orchestration within a constrained logistics context. The agent is designed to make decisions based on its knowledgebase, which consists of its system prompt and REACH regulation document.

Following the DSRM (Peffers et al., 2007), this step focuses on the conceptual design of the AI agent artifact intended to address the identified problem of manually processing supplier declarations. The purpose of this preliminary design is to define the agent’s architecture, core functionalities, and integration points, while deferring implementation-specific optimizations to later stages of the research.

The proposed artifact is an automation system, combining Copilot Studio and traditional automation designed to autonomously analyse supplier declarations related to REACH certifications, assess their compliance with organizational requirements, and coordinate subsequent automation actions. The agent acts as an intelligent actor between unstructured document inputs and the organization’s existing RPA and Power Automate infrastructure.

This section describes the design and development of the proposed artifact in accordance with the DSRM. Following the DSRM activities defined by Peffers et al. (2007), this phase corresponds primarily to Design and Development, while explicitly reflecting the outputs of the preceding problem identification and objectives of a solution phases. The resulting artifact is an automation system model intended to support organizational compliance processes by autonomously evaluating supplier declarations against regulatory requirements.

The identified organizational problem concerns the manual, time-consuming, and error-prone assessment of supplier declarations for regulatory compliance. Based on this problem, the objective of the solution was defined as the creation of an automation model that reduces manual effort of compliance decisions and supports domain-specific reasoning through an AI agent. The artifact developed in this research is designed to achieve these objectives by combining traditional automation, an AI agent and human supervision.

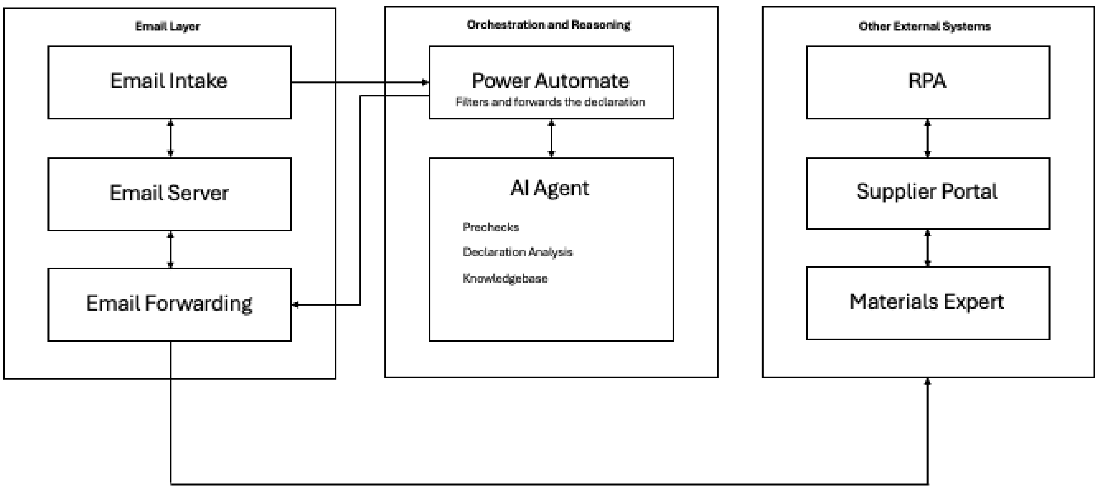

Figure 2. Initial Model of the Artifact

The designed artifact is a model of an automation system depicted in Figure 2. The model consists of these main components: email layer, orchestration and reasoning, and external systems.

Power Automate receives the attached declaration from email and passes it to the AI agent. The agent has some predefined rules to precheck the file, like checking the file format and number of files, AI capabilities are then used to read the declaration and analyse the contents. The agent is responsible for interpreting declarations within its respective regulatory domains. The agent has an embedded knowledgebase about REACH declaration to further improve its reasoning capabilities.

Due to these features, the artefact is designed to significantly change process of handling the declarations from the employees' perspective. The process will be automated end-to-end, when the declaration doesn't include SVHCs, which leaves employees to review only the more complex declarations manually.

While the agent has some predefined rules related to file prechecks, it is designed to perform contextual reasoning over the declaration content in the analysis phase. This includes validating the presence of required information, identifying inconsistencies, and determining whether the declaration satisfies regulatory obligations. The agents autonomously produce a compliance decision indicating whether the declaration is sufficient or insufficient.

5.5 Demonstration

This chapter aims to demonstrate the artifact by introducing the structure of the built solution and showing test cases for it. This chapter will also include evaluation of the artifact. This study adopts a Quick & Simple evaluation strategy proposed by Venable et al. (2026). According to Venable et al. (2016) this approach is suitable when the artefact is relatively small, and design risk is limited. As the goal is to provide a proof of concept and the time frame to create this thesis is too short to perform a thorough evaluation, this approach was chosen.

The AI agent was developed in Copilot Studio. The role of the agent is to be the reasoning component of the process and since a critical part of the artifact. The agent is triggered by receiving a file, which it performs prerequisite checks on. The agent needs to check that the file extension is .pdf and that only one file is uploaded at a time. If these conditions are not met it ends the process, otherwise it will proceed with checking the declaration.

The agent will read through the declaration and analyse the text based on a system prompt, which includes instructions and example cases. Ultimately the agent needs to assign a Boolean variable called varSVHC based on if the declaration has substances of very high concern or not, as the variable is used to decide the next step of the process. The agent is also instructed to provide an explanation of why varSVHC was set to true or false to increase transparency of the system. Visibility into agent actions assist with accountability and governance (Chan et al., 2024).

Demonstration continues with testing the agent by giving it real world cases in a supervised environment. The testing was done together with the team responsible for handling the declarations, including two material experts and their manager.

The agents reasoning capabilities were tested by providing REACH declarations for it in a controlled environment. The testing was done together with the team that is currently handling the declarations, and the team was responsible of evaluating if the agent provides the right output and reasoning for its choice.

The test was done to test if the agent can correctly analyse the declarations. At this point the email triggering was not enabled, so the declarations were given manually through the agent's chat. The agent was provided a total of 8 declarations, of which 5 declarations didn't have substances of very high concern and 3 declarations had. The declarations were chosen by the team that is responsible of handling them. The pictures below show the agents conclusion, reasoning behind its choice and how it assigned the Boolean variable.

To be successful, the agent needs to be able to read the declaration, analyse its contents, be able to detect if the declaration has substances of very high concern, and assign a Boolean variable correctly, that determines if the declaration is forwarded to RPA or materials expert.

The first declaration tested was relatively clear and included no SVHC's. The agent analysed the declaration correctly and provided sufficient reasoning behind its reasoning. The agent recognized that the declaration didn't include substances of very high concern. Some of the text has been redacted because of company specific information. (Figure 4)

Based on the supplier REACH declaration provided:

- The declaration states that [redacted] is a "downstream user" under REACH and that their products are considered articles, not substances or preparations.
- It further states: "no substance is released from the articles under normal and reasonably foreseeable conditions of use."
- The declaration also mentions: [redacted] will inform its customers without being asked as soon as information is available that listed substances above the concentrations of 0.1% specified according to REACH are contained in the products supplied."

Conclusion:
The declaration does not indicate that any substances of very high concern (SVHC) above the 0.1% threshold are present in the supplied products. The supplier commits to informing customers if such substances are present in the future.
Result:
No SVHCs are currently declared in the products.
varSVHC = **false**

Figure 3. Test Case 1: No SVHC.

The second declaration tested was also a relatively clear and included no SVHC's. The declaration was shorter than the previous one. The agent analysed the declaration correctly and provided sufficient reasoning behind its reasoning, identifying that none of the substances exceeded the SVHC threshold. (Figure 5)

After reviewing the supplier's REACH Declaration of Conformity, the document states:

- The products and packaging do not contain Substances of Very High Concern (SVHC) above the threshold value of 0.1% w/w as declared by ECHA.
- The supplier commits to regular and timely monitoring of the SVHC list and will inform customers if any such substances are present in the supplied products.

Result:
The declaration confirms that the supplied materials do not contain SVHCs above the regulatory threshold.
varSVHC = **false**

Figure 4. Test Case 2: No SVHC.

For the test case 3, the team noted that the declaration was significantly more complex than usual. The declaration had unstructured format and at the same time included substances of very high concern. The agent managed to analyse the declaration sufficiently and provided correct reasoning for the result according to the team. This means that the agent recognized the articles that are considered substances of very high concern. Some company specific information was redacted from the picture. (Figure 6)

Based on the supplier REACH declaration provided:

1. The declaration specifically mentions the presence of "Lead" (CAS Number: 7439-92-1, [redacted]
2. Lead is listed as a Substance of Very High Concern (SVHC) under the REACH Regulation (Annex XVII and the ECHA Candidate List).
3. The concentration range is stated as "> 0.1% w/w and ≤ 100% w/w", which is above the 0.1% threshold that triggers SVHC notification requirements.

Result:
The declaration confirms that the product contains a substance of very high concern (SVHC), specifically lead, above the regulatory threshold.
varSVHC = **true**

Figure 5. Test Case 3: SVHC.

Also test case 4 included SVHCs but wasn't as complex from the formatting perspective. The agent reasoned correctly and provided the right output. The agent successfully identified the articles that were considered SVHC. Some company specific information was redacted from the picture. (Figure 7)

The supplier REACH declaration from [redacted] does include substances of very high concern (SVHC) above the threshold of 0.1% w/w in certain products. The declaration provides detailed tables listing the SVHCs, their CAS numbers, and the affected products/components.

Result:

Yes, the declaration includes SVHCs above 0.1% w/w in some products.

varSVHC = true

Figure 6. Test Case 4: SVHC.

Test case 4 included SVHCs as well. The agent reasoned correctly according to the team. Redactions were done due to company specific information. However, this time there were multiple SVHCs, which the agent didn't list. (Figure 8)

Based on the supplier REACH declaration:

- The declaration states that in the articles provided, substances classified as SVHC (Substances of Very High Concern) are present in concentrations higher than 0.1% w/w.
- On page 2, the substance "Lead" (CAS Number: 7439-92-1) is specifically listed as an SVHC present in the article number [redacted]

Result:

Yes, the declaration includes substances of very high concern (SVHC).

varSVHC = true

Figure 7. Test Case 5: SVHC.

Test case 6 didn't include SVHC and was also relatively short. The agent reasoned correctly and provided the right output. Company specific information redacted from the picture. The agent also successfully mentioned which article was considered SVHC. (Figure 9)

Based on the supplier REACH declaration and its attachment, the document states for the listed product [redacted] that there are no substances of very high concern (SVHC) present above 0.1% (indicated as "None" under the SVHC >0.1% (CAS No.) column).
Result: The declaration does not include any substances of high concern (SVHC).
varSVHC = **false**

Figure 8. Test Case 6 No SVHC.

Also test case 7 was done with a declaration that didn't include SVHCs. The agent reasoned correctly that the declaration didn't include any substances of very high concern and assigned the variable correctly. Figure 10)

After reviewing the supplier's REACH declaration, the document states compliance with the REACH regulation, specifically mentioning "SVHC REACH (REACH)" in the list of applicable standards. However, the declaration does not specifically list any substances of very high concern (SVHC) present in the products. It only references compliance with the regulation.
Result: The declaration does not indicate the presence of SVHC substances in the products.
varSVHC = **false**

Figure 9. Test Case 7: No SVHC.

The eight and last test case included a declaration without SVHC. The agent successfully identified that there were no SVHC and assigned the variable correctly. (Figure 11)

Based on the supplier REACH declaration:

- The declaration states that the articles provided **do not** contain substances subject to restrictions (Annex XVII), authorization (Annex XIV), or substances of very high concern (SVHC) in concentrations higher than 0.1% w/w.
- The table for "Substances to declare" on page 2 is empty, indicating that no SVHCs or other restricted substances are present in the supplied articles.

Result:
The supplier declaration confirms that there are no substances of very high concern (SVHC) present in the supplied articles.
varSVHC = **false**

Figure 10. Test Case 8: No SVHC.

All the tests were successful according to the team. The agent provided valid reasoning behind its choices and assigned varSVHC correctly in every case. However, it was discussed that even though the agent was correct every time, it could have even further protection against mistakes. The discussed solution was to include a confidence score, that the agent provides based on how confident it is on its answers. In this case the agent will only proceed if the confidence score is 85% or above. If the confidence score is not met, the declaration will be forwarded to a human. According to Rahman et al. (2026), human intervention is appropriate when the AI based solution is not confident enough. The reason for this change is, that potential mistakes could cause potentially lot of human work and regulatory issues, so the team wants to mitigate the chance for those. From the architecture’s perspective, this change is within the agent.

5.6 Final Artifact

The final artifact developed in this study is a model of process automation proof-of-concept designed to support the processing of supplier declarations within a logistics context (Figure 12). The solution combines RPA, Power Automate, an AI Agent and human-in-the-loop. Compared to the first iteration of the solution, a confidence score was introduced to provide further protection against mistakes. The confidence scoring takes place directly in the agent.

The proof-of-concept can be defined as intelligent automation. An intelligent automation system is a combination of traditional automation and AI for decision making (Kezidora and Hyrynsalmi, 2023). It does not fully meet the characteristics of agentic automation because agentic AI requires the collaboration of multiple agents (Sapkota et al., 2026), and in agentic automation the agents act more freely, building the workflow as the process progresses (Sapkota et al., 2026), whereas in this case the workflow is predetermined.

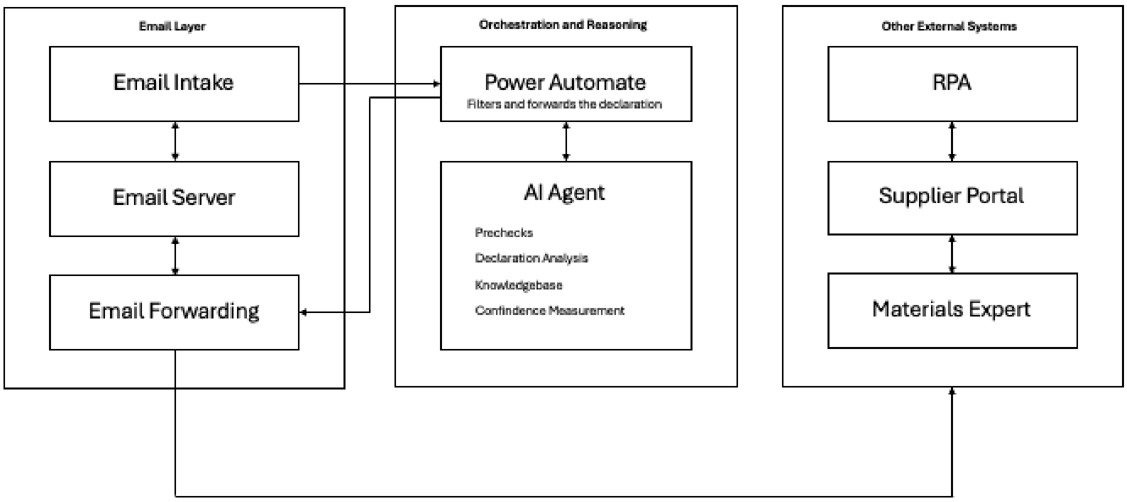

Figure 11. Final Artifact

Rather than being a fully deployed production system, the artifact serves as a proof of concept that illustrates the feasibility of integrating AI agents with traditional automation technologies. Its purpose is to validate that AI agents can be applied to unstructured compliance processes and to explore how such a system would function in practice.

The solution connects an AI agent with rule-based automation to enable the interpretation, classification, and routing of supplier declarations. It operates by combining AI, automation tools, and human oversight. The primary objective of the artifact is to demonstrate how AI agents can be implemented in logistics context and how manual document processing tasks can be partially automated using AI capabilities and traditional automation, while maintaining reliability through human oversight.

The process is initiated when supplier declarations are received via email. The file is then passed to an AI agent, which functions as the core reasoning component of the artifact. The agent analyses the content of the declaration using natural language processing and contextual reasoning. Based on predefined regulatory requirements, the agent evaluates whether the declaration is sufficient or insufficient.

Following this step, the artifact performs workflow orchestration. Declarations that do not contain substances of very high concern are automatically forwarded to a traditional RPA workflow for uploading to the company's supplier portal. Declarations with substances of very high concern, or those with uncertain outcomes, are escalated to human experts for review.

Prior to the development of the artifact, the processing of supplier declarations was conducted entirely manually. Employees were required to examine each declaration individually, interpret unstructured and heterogeneous document content, and make compliance-related decisions without the support of automated systems. This approach was time-consuming, cognitively demanding, and limited in scalability.

The proposed solution demonstrates how this process can be restructured into a semi-autonomous workflow through the integration of AI agents, traditional automation and human interaction. In the redesigned process shown in Figure 13, routine screening and classification tasks are delegated to the AI agent, reducing reliance on manual effort. The agent is capable of contextual reasoning and interpretation of unstructured documents. This shift enhances the system's ability to handle variability in document formats and content.

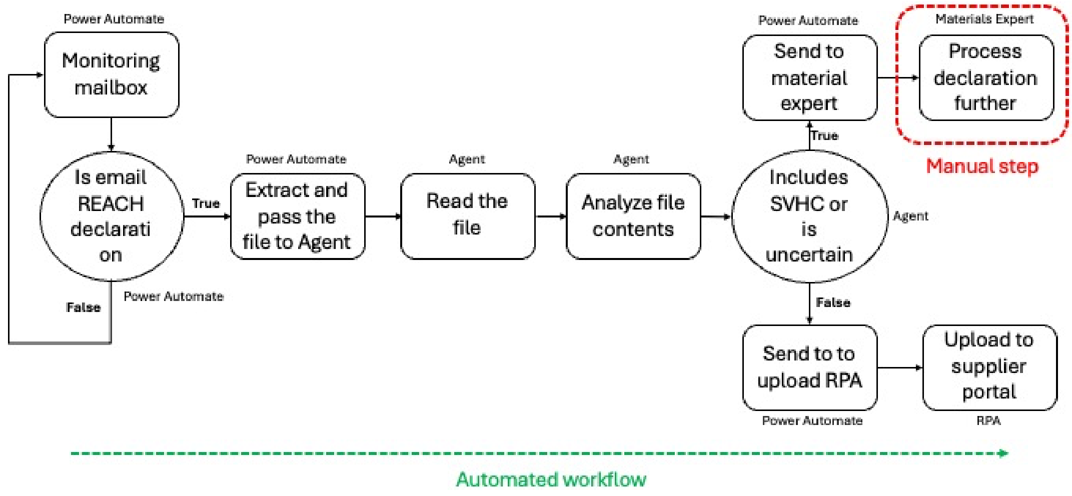

Figure 12. Artifacts Impact on the Workflow

Although the artifact is developed within a logistics context, its basic design is transferable to other domains involving unstructured document processing and compliance evaluation. The generalizable components of the artifact include the agent-based reasoning layer, and the hybrid model combining AI and traditional automation tools. These elements provide a reusable foundation for developing similar systems in other organizational settings by modifying the system prompt and agent's knowledge.

The artifact incorporates all the key properties that are associated with AI agents identified by Sapkota et al. (2026) and supported by Afrin et al. (2025). It incorporates task-specificity fully, while it partly supports autonomy and reactivity and adaptation. (Table 3)

Table 3. The Solutions Relation to AI Agent Properties.

Property	**Artifact's feature**
Autonomy	The artifact is designed to work autonomously with minimal human intervention in complex cases. Human involvement is required only in complex or uncertain cases.
Task-Specificity	The agent is specifically designed to only analyse REACH declarations provided by email and forwarding them accordingly. The system prompt prevents the agent from handling other tasks outside its scope
Reactivity and Adaptation	The agent reacts dynamically to incoming declarations as they are received. It can adapt to changing declaration formats, but it cannot adapt to changing workflows.

6 Concluding Remarks

This thesis set out to investigate how AI agents can be implemented for processing supplier declarations in a logistics organization in cases where traditional rule-based automation is insufficient due to unstructured document formats and complex compliance requirements. Another objective for the thesis was to investigate the important properties of AI agents. The study followed the Design Science Research Methodology to identify a practical problem, develop an artifact, demonstrate its applicability in practice, and generate knowledge relevant for both research and practice.

The research began by identifying the limitations of the organization's existing manual process for handling REACH declarations. Employees were required to read and evaluate large numbers of heterogeneous supplier documents, a task that was cognitively demanding, time-consuming, and unsuitable for traditional RPA due to the high degree of document variability and decision-making.

To address this problem, an intelligent automation model and proof-of-concept was designed. The artifact is a model of a Copilot Studio–based AI agent integrated with automation workflows capable of receiving and analysing supplier declarations, determining whether they contain substances of very high concern (SVHC), and routing them accordingly. The solution extracts declarations, interprets document content, applies reasoning based on system prompt and regulatory criteria, and forwards the declaration to desired address. The prototype was tested together with material experts using eight real declarations, demonstrating its capability to handle both structured and complex unstructured cases.

The result is a redesigned workflow where the initial screening of declarations is automated, documents without SVHC are forwarded to RPA for uploading, and uncertain or insufficient cases are escalated to human experts.

The study also identified key properties that an AI agent should have. The key properties of an AI agent include Autonomy, Task-Specificity, Reactivity and Adaptation (Afrin et al., 2025; Ning et al., 2025; Sapkota et al., 2026). However, this study suggests that an AI agent can work with limited autonomy and adaptation. In addition, this study identified safety as an additional property to consider in AI agents to add employee trust in the form of confidence scoring.

6.1 Theoretical Implications

This thesis contributes to the theoretical discussion on AI agents and business process automation by showing how AI agents can extend, rather than replace, existing automation logics in a compliance-intensive logistics process. The literature in Chapters 2 and 3 presents AI agents as autonomous, task-specific, and adaptive systems capable of handling unstructured inputs and contextual decision-making (Afrin et al., 2025; Sapkota et al., 2026). In contrast to this broader conceptual view, the present thesis suggests that in a real organizational setting, an AI agent may deliver theoretical value even when its autonomy and adaptability remain partial, if it is embedded in a hybrid workflow with RPA and human oversight.

The thesis aligns with earlier work that positions AI agents as systems capable of perception, reasoning, and action, and with studies arguing that such systems are particularly suitable for structured but complex tasks involving unstructured data. It also supports the view in the literature that RPA alone is limited in situations that require interpretation, judgment, or adaptation (Aalst et al., 2018), whereas AI can widen the scope of automatable work. The present findings indicate that useful agent-based automation can already emerge in a narrower configuration where the workflow remains predetermined and the agent functions mainly as a reasoning layer within a controlled process.

One theoretical contribution of the thesis is the demonstration that AI agents can act as a practical intermediate layer between traditional rule-based automation and more advanced agentic AI. Ye et al. (2023) describes agentic process automation as a step beyond

RPA. Building on these discussions, the thesis adds a more grounded view: in compliance work, the value of an AI agent may lie less in full autonomy and more in its ability to support contextual classification, explanation, and confidence-based escalation within a semi-automated workflow. This extends the theoretical understanding of automation by showing that the benefits of AI agents can be realized without adopting a fully autonomous or agentic architecture.

The thesis also offers a slightly different perspective on the defining properties of AI agents. The literature identifies autonomy, task-specificity, and reactivity/adaptation as central characteristics of AI agents (Afrin et al., 2025; Sapkota et al., 2026). However, this study suggests that an AI agent can still be meaningful and valuable when autonomy and adaptation are limited, if it performs a clearly bounded task reliably and transparently within a human-in-the-loop arrangement. In addition, the thesis introduces safety as a property in this context, since confidence thresholds were used to increase trust and reduce the risk of incorrect regulatory decisions.

6.2 Practical Implications

The study provides evidence that AI agent is feasible for processing unstructured compliance documents in a logistics environment. While earlier research highlights the theoretical potential of AI agents to handle unstructured information and perform contextual reasoning, concrete demonstrations within operational logistics compliance processes have been limited. The proof-of-concept developed in this study confirms that an AI agent can reliably interpret heterogeneous REACH declarations, extract meaningful information, and classify documents. The successful handling of irregularly formatted declarations shows that the agent can perform tasks that traditional RPA cannot accomplish due to its reliance on deterministic rules and structured inputs (Aalst et al., 2018). These results demonstrate that AI agent is a feasible solution for improving the processing of unstructured regulatory document.

The study also reveals the central role of transparent reasoning in building trust among domain experts. During testing, material experts consistently agreed with the agent's classifications and reported that the explanations provided by the agent helped them verify its decisions. Confidence threshold was introduced to provide even further trust.

The study demonstrates that combining traditional rule-based automation enhanced with AI agents is a viable solution for complex business processes. Instead of replacing RPA, the AI agent complements it by handling the cognitive and interpretative tasks that RPA cannot manage.

6.3 Limitations and Future Research Directions

While this study demonstrates the feasibility of applying AI agents to the processing of supplier declarations, several limitations must be acknowledged when interpreting the results and assessing the generalizability of the findings.

The evaluation of the artifact was limited in scope. The demonstration relied on a small sample of eight declarations, tested in a controlled environment together with domain experts. Although the agent performed correctly in all cases, the limited sample size and lack of long-term testing restrict the ability to draw conclusions about its performance under real operational variability, such as fluctuations in supplier behaviour, or evolving regulatory requirements. The artifact was not tested in full production conditions, meaning that factors such as system-to-system integration, email ingestion at scale, latency, exception handling, and operational reliability remain unexamined.

The study relies mostly on qualitative evaluation. Expert judgment was used to determine whether the agent's reasoning and output were correct, but quantitative performance metrics, like precision, recall, false-negative rates, processing times, or error impact analysis were not included. As a result, the magnitude of efficiency gains and reliability improvements cannot be empirically quantified.

The solution was designed and evaluated within a single organization, for one specific compliance process, using one type of regulatory document and technical tools were predetermined. Consequently, the findings cannot be generalized without caution to other industries, document types, or compliance contexts. While the architecture is transferable, its effectiveness in other environments would require additional testing and adaptation.

Finally, broader ethical, governance, and risk-management considerations were outside the scope of this thesis. Although the design incorporates a confidence threshold to improve safety, the study does not explore further security or governance issues. These factors become important when putting AI into a larger role in compliance workflows.

These limitations give opportunity for further research directions and more extensive evaluation. Future research could explore ways to incorporate continuous learning or structured feedback loops into the automation system. As the current artifact does not adapt based on past decisions, introducing mechanisms such as reinforcement learning, iterative prompt refinement, or human-in-the-loop feedback could improve long-term accuracy and resilience. This would allow the agent to better handle potential changes.

Another important direction involves evaluating the solution in a large-scale, real-world environment. The proof of concept in this study was tested using a limited number of declarations in a controlled environment, meaning its performance under operational load remains unexamined. Large-scale testing could provide quantitative metrics on accuracy, processing time, reliability, and error rates. Such evaluation would help determine the system’s real performance.

Extending the architecture to agentic multi-agent systems also offers meaningful opportunities for further development. Instead of relying on a single reasoning agent, future research could investigate how an agentic solution may collaborate, verify each other’s outputs, or divide responsibilities based on regulatory domains. Multi-agent setups may increase robustness, reduce the risk of incorrect classifications, and provide a stronger foundation for complex decision-making tasks.

Also studying employee acceptance and the organizational impacts of AI agents, intelligent automation or agentic automation would provide valuable insights into how such systems are adopted in practice. Research could examine how employees perceive the agent’s decisions, how confidence thresholds affect trust, what kinds of training or governance structures are required, and how agentic automation reshapes professional roles, and what ethical issues these solutions can cause. Understanding these aspects would help ensuring that AI systems are integrated successfully and sustainably into organizational processes.

References

Aalst, W. M. P. van der, Bichler, M., & Heinzl, A. (2018). Robotic Process Automation. *Business & Information Systems Engineering*, *60*(4), 269–272. https://doi.org/10.1007/s12599-018-0542-4

Afrin, S., Roksana, S., & Akram, R. (2025). AI-Enhanced Robotic Process Automation: A Review of Intelligent Automation Innovations. *IEEE Access*, *13*, 173–197. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3513279

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). *Concrete Problems in AI Safety* (arXiv:1606.06565). arXiv. https://doi.org/10.48550/arXiv.1606.06565

Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges. *Future Internet*, *15*(8), 260. https://doi.org/10.3390/fi15080260

Batra, Y., Tim, Y., & Kirshner, S. (2025). *Does More Automation Mean More Efficiency? A Mixed Methods Study Exploring Agentic AI Design Strategies*.

Bavaresco, R. S., Nesi, L. C., Victória Barbosa, J. L., Antunes, R. S., Da Rosa Righi, R., Da Costa, C. A., Vanzin, M., Dornelles, D., Junior, S. C., Gatti, C., Ferreira, M., Silva, E., & Moreira, C. (2023). Machine learning-based automation of accounting services: An exploratory case study. *International Journal of Accounting Information Systems*, *49*, 100618. https://doi.org/10.1016/j.accinf.2023.100618

Bellogín, A., Giudici, P., Larsson, S., Pang, J., Schimpf, G., Sengupta, B., & Solmaz, G. (2025). *Systemic Risks Associated with Agentic AI: A Policy Brief*.

Bhattaram, A., Chung, J., Chung, S., Gupta, R., Ramamoorthy, J., Gullapalli, K., Marculescu, D., & Stamoulis, D. (2025). GeoFlow: Agentic Workflow Automation for Geospatial Tasks. *Proceedings of the 33rd ACM International Conference on Advances in Geographic Information Systems*, 1150–1153. https://doi.org/10.1145/3748636.3763217

Blahušiaková, M. (2023). Business process automation: New challenges to increasing the efficiency and competitiveness of companies. *Strategic Management*, *28*(3), 18–33. https://doi.org/10.5937/StraMan2300038B

Bo, X., Zhang, Z., Dai, Q., Feng, X., Wang, L., Li, R., Chen, X., & Wen, J.-R. (2024). *Reflective Multi-Agent Collaboration based on Large Language Models*.

Capel, T., & Brereton, M. (2023). What is Human-Centered about Human-Centered AI? A Map of the Research Landscape. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–23. https://doi.org/10.1145/3544548.3580959

Chan, A., Ezell, C., Kaufmann, M., Wei, K., Hammond, L., Bradley, H., Bluemke, E., Rajkumar, N., Krueger, D., Kolt, N., Heim, L., & Anderljung, M. (2024). Visibility into AI Agents. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 958–973. https://doi.org/10.1145/3630106.3658948

Condon, G., & Jilani, M. (2026). Towards a framework for Bias Evaluation of AI Agents in Agentic Workflows. *Proceedings of the 2026 Conference on Human Centred Artificial Intelligence - Education and Practice*, 46–52. https://doi.org/10.1145/3777490.3777504

European Commission. (n.d.). *REACH regulation*. Retrieved April 2, 2026, from https://environment.ec.europa.eu/topics/chemicals/reach-regulation_en

Fan, S., Cong, X., Fu, Y., Zhang, Z., Zhang, S., Liu, Y., Wu, Y., Lin, Y., Liu, Z., & Sun, M. (2024). *WorkflowLLM: Enhancing Workflow Orchestration Capability of Large Language Models* (arXiv:2411.05451). arXiv. https://doi.org/10.48550/arXiv.2411.05451

Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, *7*(1), 549–569. https://doi.org/10.1007/s42001-024-00250-1

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, *25*(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Haase, J., Kremser, W., Johannes Kepler University Linz, Austria, Leopold, H., Mendling, J., Onnasch, L., & Plattfaut, R. (2024). Interdisciplinary Directions for Researching the Effects of Robotic Process Automation and Large Language Models on Business Processes. *Communications of the Association for Information Systems*, *54*(1), 579–604. https://doi.org/10.17705/1CAIS.05421

Hevner, A. R. (2007). *A Three Cycle View of Design Science Research*. *19*.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). *Design Science in Information Systems Research*.

Hoang Thuan, N., Drechsler, A., & Antunes, P. (2019). Construction of Design Science Research Questions. *Communications of the Association for Information Systems*, 332–363. https://doi.org/10.17705/1CAIS.04420

Kapoor, S., Stroebl, B., Siegel, Z. S., Nadgir, N., & Narayanan, A. (2024). *AI Agents That Matter* (arXiv:2407.01502). arXiv. https://doi.org/10.48550/arXiv.2407.01502

Kedziora, D., & Hyrynsalmi, S. (2023). Turning Robotic Process Automation onto Intelligent Automation with Machine Learning. *The 11th International Conference on Communities and Technologies (C&T)*, 1–5. https://doi.org/10.1145/3593743.3593746

Memarian, B., & Doleck, T. (2024). Human-in-the-loop in artificial intelligence in education: A review and entity-relationship (ER) analysis. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100053. https://doi.org/10.1016/j.chbah.2024.100053

Moraes, C. H. V. de, Scolimoski, J., Lambert-Torres, G., Santini, M., Dias, A. L. A., Guerra, F. A., Pedretti, A., & Ramos, M. P. (2022). Robotic Process Automation and Machine Learning: A Systematic Review. *Brazilian Archives of Biology and Technology*, *65*, e22220096. https://doi.org/10.1590/1678-4324-2022220096

Naik, S., Toombs, A. L., Snellinger, A., Saponas, S., & Hall, A. K. (2025). Designing with Multi-Agent Generative AI: Insights from Industry Early Adopters. *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, 1961–1972. https://doi.org/10.1145/3715336.3735823

Ng, K. K. H., Chen, C.-H., Lee, C. K. M., Jiao, J. (Roger), & Yang, Z.-X. (2021). A systematic literature review on intelligent automation: Aligning concepts from theory, practice, and future perspectives. *Advanced Engineering Informatics*, *47*, 101246. https://doi.org/10.1016/j.aei.2021.101246

Ning, L., Liang, Z., Jiang, Z., Qu, H., Ding, Y., Fan, W., Wei, X., Lin, S., Liu, H., Yu, P. S., & Li, Q. (2025). A Survey of WebAgents: Towards Next-Generation AI Agents for Web Automation with Large Foundation Models. *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, 6140–6150. https://doi.org/10.1145/3711896.3736555

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pramod, D. (2022). Robotic process automation for industry: Adoption status, benefits, challenges and research agenda. *Benchmarking*, *29*(5), 1562–1586. https://doi.org/10.1108/BIJ-01-2021-0033

Radanliev, P., Maple, C., Santos, O., & Atefi, K. (2026). SBOMs into Agentic AIBOMs: Schema Extensions, Agentic Orchestration and Reproducibility Evaluation. *Digital Threats: Research and Practice*, *7*(2), 1–35. https://doi.org/10.1145/3798285

Rahman, A. (2026). *DataSpeck: An AI-Driven Human-in-the-Loop System for Automating Transformations in Data Conversion Workflows*.

Ratia, M., Myllärniemi, J., & Helander, N. (2018). Robotic Process Automation—Creating Value by Digitalizing Work in the Private Healthcare? *Proceedings of the 22nd International Academic Mindtrek Conference*, 222–227. https://doi.org/10.1145/3275116.3275129

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A modern approach*. https://bookshelf.vitalsource.com/reader/books/9781292410074/epubcfi/6/2[%3Bvnd.vst.idref%3DCover]!/4/2/2%4050:37

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

Schell, C. (1992). *The Value of the Case Study as a Research Strategy*.

Schlegel, D., Rosenberg, B., Fundanovic, O., & Kraus, P. (2024). How to conduct successful business process automation projects? An analysis of key factors in the context of robotic process automation. *Business Process Management Journal*, *30*(8), 99–119. https://doi.org/10.1108/BPMJ-06-2023-0465

Shetty, M., Chen, Y., Somashekar, G., Ma, M., Simmhan, Y., Zhang, X., Mace, J., Vandevoorde, D., Las-Casas, P., Gupta, S. M., Nath, S., Bansal, C., & Rajmohan, S. (2024). Building AI Agents for Autonomous Clouds: Challenges and Design Principles. *Proceedings of the ACM Symposium on Cloud Computing*, 99–110. https://doi.org/10.1145/3698038.3698525

Shinn, N., Cassano, F., Berman, E., Gopinath, A., Narasimhan, K., & Yao, S. (2023). *Reflexion: Language Agents with Verbal Reinforcement Learning* (arXiv:2303.11366). arXiv. https://doi.org/10.48550/arXiv.2303.11366

Silvares, C., São Mamede, H., & Costa, J. (2024). Best practices for business process automation description – a case study. *Enterprise Information Systems*, *18*(10), 2400474. https://doi.org/10.1080/17517575.2024.2400474

Sliż, P., Siciński, J., & Jackowska, B. (2024). The impact of ambidexterity on robotic process automation implementation strategy in service enterprises. *Future Business Journal*, *10*(1), 6. https://doi.org/10.1186/s43093-023-00296-z

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Vom Brocke, J., & Maedche, A. (2019). The DSR grid: Six core dimensions for effectively planning and communicating design science research projects. *Electronic Markets*, *29*(3), 379–385. https://doi.org/10.1007/s12525-019-00358-7

Wadinambiarachchi, S., Waycott, J., Rogers, Y., & Wadley, G. (2025). Imagining Design Workflows in Agentic AI Futures. *Proceedings of the 37th Australian Conference on Human-Computer Interaction*, 19–35. https://doi.org/10.1145/3764687.3764719

Wali, A., Mahamad, S., & Sulaiman, S. (2023). Task Automation Intelligent Agents: A Review. *Future Internet*, *15*(6), 196. https://doi.org/10.3390/fi15060196

Wang, Z., Cui, Y., Zhong, L., Zhang, Z., Yin, D., Lin, B. Y., & Shang, J. (2024). *OfficeBench: Benchmarking Language Agents across Multiple Applications for Office Automation* (arXiv:2407.19056). arXiv. https://doi.org/10.48550/arXiv.2407.19056

Warudkar, S., & Jalit, R. (2024). Unlocking the Potential of Generative AI in Large Language Models. *2024 Parul International Conference on Engineering and Technology (PICET)*, 1–5. https://doi.org/10.1109/PICET60765.2024.10716156

Winter, R. (2008). Design science research in Europe. *European Journal of Information Systems*, *17*(5), 470–475. https://doi.org/10.1057/ejis.2008.44

Ye, Y., Cong, X., Tian, S., Cao, J., Wang, H., Qin, Y., Lu, Y., Yu, H., Wang, H., Lin, Y., Liu, Z., & Sun, M. (2023). *ProAgent: From Robotic Process Automation to Agentic Process Automation* (arXiv:2311.10751). arXiv. https://doi.org/10.48550/arXiv.2311.10751

Zhu, Y.-Q., & Kanjanamekanant, K. (2023). Human–bot co-working: Job outcomes and employee responses. *Industrial Management & Data Systems*, *123*(2), 515–533. https://doi.org/10.1108/IMDS-02-2022-0114

Zouari, J. (2025). Toward Agentic IAM: A Probabilistic Authorization Framework for Least Privilege AI Workflows. *Proceedings of the IEEE/ACM 12th International Conference on Big Data Computing, Applications and Technologies*, 1–6. https://doi.org/10.1145/3773276.3776564

Zubiaga, A. (2024). Natural language processing in the era of large language models. *Frontiers in Artificial Intelligence*, *6*, 1350306. https://doi.org/10.3389/frai.2023.1350306

Appendices

Appendix 1. Assessment of Artificial Intelligence Usage

Artificial intelligence has been used to improve fluency and grammar of the text. ChatGPT, Perplexity and Copilot have been used for this purpose.