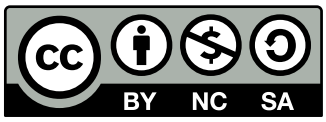
Aalto University
School of Science

Master's Programme in Mathematics and Operations Research

Improving risk registers with an AI assistant

Alexander Westergård

Master's Thesis
2025



Aalto University
School of Science

Author Alexander Westergård

Title Improving risk registers with an AI assistant

Degree programme Mathematics and Operations Research

Major Systems and Operations Research

Supervisor Prof. Ahti Salo

Advisor Juha Törmänen, D.Sc (Tech.)

Collaborative partner Inclus

Date 5 July 2025 **Number of pages** 53+5 **Language** English

Abstract

Risk registers are a structured way to document an organization's risks and are a central part of the risk management process. Over time, risk registers can become incoherent, filled with redundant risks and missing relevant ones. Such risk registers may become overwhelming for experts to manage.

This thesis developed an artificial intelligence (AI) assistant framework designed to address common issues in risk registers. The AI assistant utilizes large language models (LLMs) enhanced with structured tools and a retrieval-augmented generation (RAG) search. The AI assistant can read information from the organization's risk register, RAG knowledge base, and user inputs to create a comprehensive working context. Using structured tools, the AI assistant can suggest modifications to the risk register. Multi-step reasoning enables the completion of more complex tasks that involve multiple tools and data sources.

A case study with Inclus Oy examined how an AI assistant affects the speed, quality, and handling of complex data in the risk register while maintaining a reliable process. Analytical generalizations were made based on observations of the risk register, AI assistant, and perception case study participants. The study yielded promising results, where the AI assistant suggested merging duplicate risks, harmonizing descriptions, and adding new risks. A survey revealed that the use of an AI assistant improved risk register coverage and made it more accurate and up-to-date.

While an AI assistant substantially speeds up risk management tasks and improves the quality, it is still prone to hallucinations. The study demonstrated that AI assistants must operate under human oversight and that users must retain the authority to review and confirm every proposed modification to the register. Additionally, human oversight ensures that participants are committed to mitigating the risks. AI assistants can enhance the participatory process by accelerating and improving the quality of data processing.

Keywords Risk register, large language models, AI assistant, participatory risk management

Aalto-yliopisto
Perustieteiden korkeakoulu

Tekijä Alexander Westergård

Työn nimi Riskirekisterien parantaminentekoälyavustajan avulla

Koulutusohjelma Matematiikka ja Operaatiotutkimus

Pääaine Systems and Operations Research

Työn valvoja Prof. Ahti Salo

Työn ohjaaja TkT Juha Törmänen

Yhteistyötaho Inclus

Päivämäärä 5.7.2025 **Sivumäärä** 53+5 **Kieli** englanti

Tiivistelmä

Riskirekisterit ovat jäsennelty tapa dokumentoida organisaation riskejä, ja ne muodostavat keskeisen osan riskienhallintaprosessia. Ajan myötä riskirekisterit voivat kuitenkin muuttua epäjohdonmukaisiksi, niihin voi kertyä päällekkäisiä tai epäolennaisia riskejä, tai osa olennaisista riskeistä saattaa jäädä kirjaamatta. Tällaisen riskirekisterin hallinnointi voi olla liian vaikeaa asiantuntijoille.

Tässä diplomityössä kehitettiin tekoälyavustaja, joka on suunniteltu korjaamaan riskirekisterien tyypillisiä ongelmia. Tekoälyavustaja perustuu suuriin kielimalleihin (LLM), ja sen toimintaa tehostetaan erilaisten työkalukutsujen sekä *retrieval-augmented generation* (RAG) -hakumekanismin avulla. Tekoälyavustaja voi hakea tietoa organisaation riskirekisteristä tai RAG-tietokannasta, ja käyttäjän antamat syötteet luovat kattavan kontekstin mallin käyttöön. Työkalukutsujen avulla tekoälyavustaja voi ehdottaa muokkauksia riskirekisteriin. Monivaiheinen päättely puolestaan mahdollistaa monimutkaisempien toimenpiteiden suorittamisen useiden työkalujen ja tietolähteiden avulla.

Inclus Oy:n kanssa toteutetussatapaustutkimuksessa tarkasteltiin, kuinka tekoälyavustaja vaikuttaa riskirekisterin hallinnoinnin nopeuteen, laatuun, auttaa kompleksin datan hallinnoinnissa, siten että prosessi pysyy luotettavana. Tutkimuksessa tehtiin analyyttisiä yleistyksiä havainnoista, joita tehtiin riskirekisteristä, tekoälyavustajasta sekä osallistujien kokemuksista. Tutkimus antoi lupaavia tuloksia siitä, että tekoälyavustaja pystyi ehdottamaan päällekkäisten riskien yhdistämistä, riskikuvausten yhdenmukaistamista ja uusien riskien lisäämistä. Kysely paljasti, että tekoälyavustajan käyttö paransi riskirekisterin kattavuutta sekä teki siitä tarkemman ja ajantasaisemman.

Vaikka tekoäly nopeuttaa merkittävästi joitain riskienhallintatehtäviä ja parantaa riskirekisterin laatua, se on silti yhä altis hallusinaatioille. Tutkimus korostaa, ettei tekoälylle tulisi antaa autonomiaa hallinnoida riskirekisteriä, vaan käyttäjän tulee hyväksyä tekoälyn ehdottamat toimenpiteet ennen niiden toteuttamista. Tämä takaa lisäksi sen,että osallistujat sitoutuvatriskienhallintaprosessiinja riskienminimointiin. Tekoälyavustaja voi parantaa osallistavaa riskienhallintaprosessia nopeuttamalla työvaiheita ja parantamalla prosessin laatua.

Avainsanat Riskirekisteri, suuret kielimallit, osallistava riskienhallinta, tekoälyavustaja

Preface

I want to thank the entire Inclus team for the opportunity to conduct my master's thesis within the company and for their active involvement throughout the project. My special thanks go to Juha Törmänen, whose guidance and insightful feedback significantly contributed to the development of this work. I am equally grateful to my academic supervisor, Ahti Salo, for his valuable comments and to everyone who read the draft and shared thoughtful suggestions during the process.

Otaniemi, 5 July 2025

Alexander Westergård

Contents

Abbreviations

AI	Artificial Intelligence
API	Application Programming Interfaces
FMEA	Failure Mode and Effects Analysis
GPT	Generative Pre-trained Transformer
ID	Identifier
ISO	International Organization for Standardization
LLM	Large Language Model
PDF	Portable Document Format
RAG	Retrieval-Augmented Generation
ReAct	Reasoning and Action mechanism
RQ	Research Question

1 Introduction

Humans have always managed risk in some form, but risk management as a scientific field only began to take formal shape approximately 40–50 years ago (Aven, 2016). The increasing complexity of organizations and external threats make systematic risk management crucial for informed decision-making across organizations and institutions.

Organizations can perform risk management on either an ad hoc or continuous basis. The ISO 31000 (2018) standard recommends a cyclical model, emphasizing continuous monitoring and iterative improvement. The cyclic risk management process expects risks to evolve through time (Patterson & Neailey, 2002). New risks emerge, old ones become irrelevant, and risks change priority. The decision-maker must understand the evolving risk landscape to make an informed decision. A standard method for maintaining risk awareness is to keep an up-to-date document of all relevant risk information, which is referred to in this thesis as a risk register. A risk register typically includes unique risk identifiers (risk IDs), descriptions of identified risks, risk criteria, relevant controls or mitigation steps, risk owners, and dates of relevant actions (Leva et al., 2017).

When well-maintained, a risk register can facilitate informed and transparent decision-making. However, inadequately structured or outdated registers may lead to confusion and reduced decision quality. Typical problems are related to incoherent, redundant, or duplicate risks, making it difficult and labor-intensive to maintain the risk register (Leva et al., 2017). Limitations in expert knowledge, especially in domains outside the expert's expertise, can result in low-quality assessments or missing relevant risks (Aven, 2013). If the risk register is inaccurate or incomplete, it may lose its utility as a decision-support tool.

Recent advancements in artificial intelligence (AI) have led to the development of large language models (LLMs) (Hagos et al., 2024). LLMs present a promising direction for improving risk management processes, particularly in maintaining and enhancing risk registers. LLMs can process large volumes of unstructured information from multiple sources. They help with challenges such as knowledge gaps and the complexity of managing evolving risk data. The use of LLMs in risk management, particularly in expert-driven processes such as participatory risk management, is a relatively unexplored yet promising application area.

This thesis examines how LLM-based agent models, also known as AI assistants, can aid in managing risk registers. A case study examined how an AI assistant improved the risk register of Inclus Oy, a company specializing in participatory risk management (Inclus, 2025). The implementation enhances the AI assistant with structured tool use (Yao et al., 2023) and data retrieval mechanisms, including retrieval-augmented generation (RAG) (Lewis et al., 2020). The study evaluated the assistant's usefulness in expert workflows and assessed its role in supporting risk register management.

Chapters 2-4 review the risk management literature and the theoretical background of large language models. Chapters 5-7 present the AI assistant implementation and case study evaluating its performance. Chapters 8-9 discuss the results, conclude the thesis, and outline directions for future research.

2 Theoretical perspectives on participatory risk management

This chapter explores the theoretical foundations of participatory risk management, covering risk definitions, risk management processes, risk registers, and the associated challenges. The chapter also discusses how background knowledge and the uncertainties related to risks affect risk management.

2.1 Risk definition

Risk is a concept deeply rooted in history, with evolving interpretations. The earliest definition of a measurable risk comes from de Moivre, who characterized it as an expected loss (de Moivre, 1718). According to Aven (2012), this concept evolved into a strict probability-based risk perspective, assigning an exact probability to the undesirable event. A strict probability-based approach requires information on the likelihood of risks and does not consider the preferences of the participants. Obtaining accurate probabilities may be impractical, as some risks are highly situation-specific, while others depend on someone's deliberate action. Using probability alone as a risk measure overlooks stakeholders' values and preferences.

Kaplan & Garrick (1981) introduced *'the set of triplets'*, meaning that risk analysis answers three questions: What are the scenarios, likelihoods, and consequences? Their framework emphasizes the use of likelihood and consequences together to describe risk. A subsequent shift moved the concept from a strict probability-based approach to a more loosely defined uncertainty (Aven, 2012). The concept of uncertainty enables more nuanced and subjective expert judgment of risks. Aven notes that viewing risk in terms of 'consequences and uncertainty' enables consideration of both positive and negative consequences, broadening the term's application. Consequently, different fields of application utilize varying definitions of risk.

ISO 31000 (2018) defines risk as *"the effect of uncertainty on objectives"*. The standard also defines a stakeholder as any party affected by the risk, not just the decision-maker. Consequences may be positive or negative. It uses *'likelihood'* instead of *'probability'*, enabling a broader interpretation of the chances that an event may occur. This thesis adopts the definitions provided in ISO 31000 because those are widely recognized and easily generalizable.

2.2 Risk management process

Modern organizations face an environment of increasing complexity, where uncertainty is often more relevant than measurable probability. In this context, managing risk is no longer a mathematical exercise, but a strategic necessity. Each organization may define and manage risk differently, but key elements remain consistent.

Risk management is a structured process that identifies, assesses, and addresses potential threats or opportunities affecting an organization's objectives (Aven, 2016). Risk management is a key element in operational and strategic decision-making,

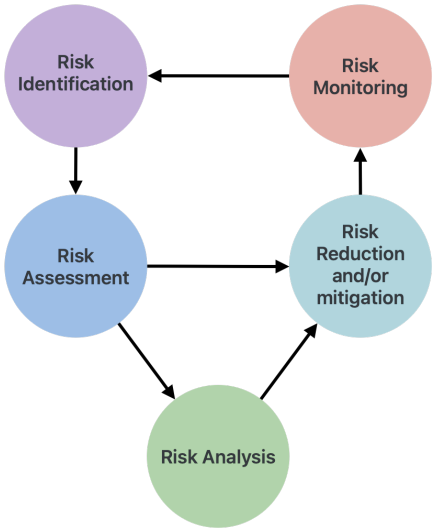


Figure 1: Risk management process cycle (Patterson & Neailey, 2002).

informing decision-makers of the organization's risks and helping to mitigate the risks (Apostolakis, 2004). This process typically follows a structured framework. Patterson & Neailey (2002) characterize risk management as a cyclical process that aligns with the ISO 31000 (2018) standard definition. Figure 1 illustrates the cyclic process of continuous risk management. Continuous risk management improves and adapts to the changing risk environment.

2.3 Risk registers

A risk register provides a structured method for documenting organizational risks. A risk register is a key component in the continuous risk management process, as it enables the recording of past risks. Leva et al. (2017) suggested core components that a risk register should include, as presented in Table 1. Additional elements may be included, depending on the organization's needs. The purpose of the risk register is to support structured decision-making by enabling visibility and traceability in risk-related actions.

Organizations update the risk register based on the cyclic process presented in Figure 1. Aligning with the process defined by Patterson & Neailey (2002) and the core components presented by Leva et al. (2017), the risk register includes the following entries:

1. A newly identified risk has at least a *risk ID*, *description*, and *entry date*.
2. Risk is *ranked* based on the given metrics in the risk assessment phase.
3. Risk analysis determines *actions* to mitigate the risk and the *target dates*.
4. When an action is completed, the *completion date* is recorded.

The risk *owner* is added at the latest when planning various mitigation actions. This person does not necessarily need to be the one implementing the individual actions, but they are still responsible for ensuring the mitigation actions are carried out. Organizations update their risk register regularly, such as monthly, quarterly, or yearly, to monitor risk trends and patterns in a constantly changing environment. However, different risks may have different timelines. The risk ranking is updated with each cycle. Actions and related dates can be updated as needed, new risks can be identified, and risks can be archived or removed from the risk register if they become acceptable or managed.

Table 1: Risk register core components (Leva et al., 2017).

Element	Description
Risk ID	Unique identification number for each risk.
Risk description	Explanatory description, category, and title for the risk.
Risk ranking	Estimates of risk metrics such as likelihood and impact.
Owner	Person or organization responsible for managing the risk and ensuring action execution.
Actions	List of relevant mitigation or response actions for the risk.
Dates	Entry and modification dates for the risk, plus target and completion dates for actions.

2.4 Challenges in maintaining risk registers

While continuous risk management is beneficial in its own right, maintaining a harmonized register may be challenging. Common issues in managing risk registers include having too many redundant risks, not including all relevant risks, or lacking a shared understanding among experts regarding the listed risks (Leva et al., 2017). Risk management supports human decision-making, so the risk register must be intuitive and easily interpretable. Organizations could list all perceptible risks, including every detail. However, that would not be meaningful, as risk management resources are limited. A best practice is to focus on the most relevant risks for the organization. Leva et al. noted that the risk register also serves as a risk-ranking tool, determining the priority of each risk.

Leva et al. (2017) state that low-level risks and risks routinely managed may divert attention away from more relevant risks. The risk register should exclude risks with low likelihood and impact (accepted risks), as managing those can be more expensive than simply accepting them. Leva et al. note that the risk register may become filled with redundant risks if it is not properly managed. Organizations should continually update their risk register to foster fresh perspectives and prevent entrenched thinking.

Listed risks may be highly relevant yet overly specific, or their definitions may overlap. Bjørnsen & Aven (2019) discuss how combining risks can lead to a better understanding of the overall risks. If risks defined with similar characteristics have similar estimates of the likelihood and impact, their aggregation can be considered.

Poorly defined risks can cause a lack of common understanding or inappropriate specificity (Owen, 2015). The wording of the risk name or description may be too imprecise or vague. Disagreements can arise from different interpretations or a lack of a shared conceptual framework.

Some relevant risks may be absent from the register due to limited knowledge, insufficient expertise, or difficulty detecting or anticipating them (Aven, 2015). Particularly, risks that may have no prior data and are unobservable from history. Recognizing these kinds of risks requires unconventional thinking.

To summarize, four core challenges often undermine the quality of risk registers:

- **Missing relevant risks**: due to limited expertise or blind spots in risk identification.
- **Inadequately defined risks**: resulting in ambiguity, expert disagreement, and weak decision support.
- **Redundant risks**: cluttering the register with acceptable or low-priority items.
- **Duplicative or overly specific risks**: leading to fragmentation and inefficient analysis.

A better-defined risk register should lead to a more comprehensive understanding of the risks, making it more actionable.

2.5 Participatory risk management

Risk management decisions should not be left solely to technical experts or individual decision-makers in complex and uncertain environments. Participatory risk management emphasizes the importance of involving all parties affected by the risk in the risk management process. It engages experts who analyze the risks and decision-makers who make risk-aware decisions. Depending on the organization, experts and decision-makers can be two different groups of people, overlapping groups, or the same people (French, 2011).

A group of experts has more diverse competence and extensive knowledge than one individual, leading to better risk management. Owen (2015) highlights that every expert has an incomplete and subjective understanding of the risk domain and argues that the collaborative method can lead to a more comprehensive understanding of the risks and, thus, better decisions.

Apostolakis (2004) argues that decision-making should always be *risk-informed*, not *risk-based*. Risk-informed decisions consider both the analyzed risks and the participants' preferences. In contrast, *risk-based* decision-making refers to decisions solely based on the identified risks. Participant preferences consistently influence decisions, which should not be underestimated. Hansson & Aven (2014) present preferences and values as a significant factor affecting the risk evaluation process and all subsequent decisions. Figure 2 shows how value-based and fact-based judgments affect the different phases of the risk management process. While experts base risk

analysis primarily on facts, the decision is guided by the decision-maker's values and preferences.

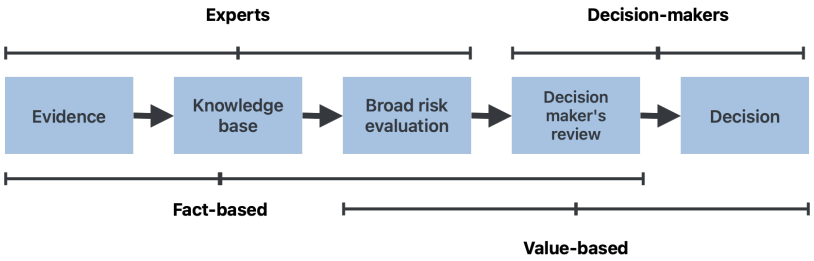


Figure 2: Information flow in the risk management process (Hansson & Aven, 2014).

Despite its advantages, participatory risk management can also surface disagreements among participants. French & Argyris (2018) notes that disagreement is almost always present at some level. Thus, groups never decide anything unitarily. Risk management is never purely technical; it inevitably involves value-based judgments. Recognizing disagreements helps ensure that diverse participant preferences, shaped by different interests, values, and goals, are acknowledged in the decision-making process. Especially in political or enterprise settings, participants may perceive and prioritize risks in conflicting ways. Participatory risk management increases participants' trust and perceived fairness, highlighting the need for it (Scolobig, 2025).

2.6 Uncertainty of knowledge

While organizations may describe risk events in standardized terms, their perceived relevance and impact vary significantly depending on participants' context and background knowledge. Background knowledge is a key factor in risk management, as only acknowledged risks can be intentionally managed (Aven & Krohn, 2014).

Risk assessment is especially challenging when it comes to so-called black swans (Aven, 2013). Nicholas Nassim Taleb introduced the concept of black swans to a broader audience in his famous book *The Black Swan: The Impact of the Highly Improbable* (Taleb, 2007). Taleb described black swans as highly improbable events with extreme consequences, often claiming that they were predictable in hindsight. A central theme in black swans is the lack of knowledge. Aven (2015) notes that unidentified risks are unlikely to have any contingency measures, leading to more severe consequences. The uncertainty of knowledge also extends beyond black swans, as domain experts may lack a sufficient understanding of more apparent risks that fall outside their area of expertise.

Second-degree uncertainties are uncertainties about underlying knowledge, also known as epistemic uncertainties. The famous phrase from the United States Secretary of Defense, Donald Rumsfeld, on February 12, 2002 (U.S. Department of Defense, 2002), during a press conference about Iraq and weapons of mass destruction, goes as:

> *"...there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say, we know there are some*

things we do not know. But there are also unknown unknowns—the ones we don't know we don't know..."

The term quickly entered the field of risk science. This framework is further illustrated in Figure 3, which maps the relationship between awareness and knowledge. *'Known knowns'* can be defined as risks stakeholders know and can accurately model. On the other hand, *'known unknowns'* are risks stakeholders are aware of but cannot accurately model. *'Unknown knowns'* and *'unknown unknowns'* pose greater challenges as they usually go unnoticed before the risk materializes (Aven, 2015). Aven notes that the last two cases always come as a surprise relative to one's knowledge. By definition, risks that are of the *'unknown unknown'* type are impossible to identify as they are unknown in the objective sense.

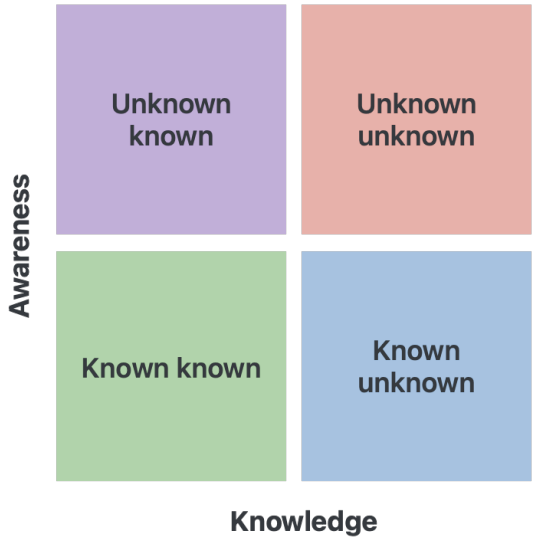


Figure 3: Knowledge and awareness matrix.

Flage & Aven (2009) introduced the concept of the strength of knowledge to measure second-degree uncertainty. It reflects how well the phenomena are understood, the validity of assumptions, the reliability of available data, and expert consensus. Aven (2013) notes that the strength of knowledge should affect risk prioritization. Risks with low strength of knowledge can pose a higher risk level than predicted due to the greater potential for unexpected consequences.

To conclude, there are two problems related to the experts' limited knowledge:

- **Surprises**: Risks that fall entirely outside the expert's subjective awareness (i.e., *'unknown unknowns'* or *'unknown knowns'*).
- **Weak knowledge**: Risks that are recognized but insufficiently understood, often due to limited data or contested assumptions.

Surprises are unknown risks, though they may be known by others in *'unknown known'* case, even if the particular experts are unaware of the risk. On the other hand, weak

knowledge of a risk means that the experts are aware of the risk but do not fully understand it. That raises a critical question of how organizations can systematically evaluate and strengthen their knowledge base, particularly in settings where expert judgment is the primary source of risk information. Addressing weak knowledge is not merely a theoretical concern. Instead, it is a foundational step toward building evidence-based risk management systems that can anticipate both the expected and the unforeseen.

3 Overview of large language models and enhancements

This chapter covers the recent advances of large language models (LLMs), their operating principles, added tools, and the RAG technique. It also addresses the issues associated with LLMs.

3.1 Recent developments of large language models

Recent advances in large language models (LLMs) have received much attention. In 2017, Vaswani et al. (2017) introduced a breakthrough technology: transformer architecture. Transformers employ a self-attention mechanism to capture relationships between tokens in an input sequence, prioritizing contextual understanding over strictly word-by-word processing. Transformers allow the model to consider the broader context when predicting the next token. Self-attention enables models to capture dependencies across a sentence or document, allowing them to form richer contextual representations. Modern LLMs process input through multiple layers, each operating at different levels of abstraction. These layers update and evaluate the relationships between token embeddings, not just individual words, to build contextual understanding.

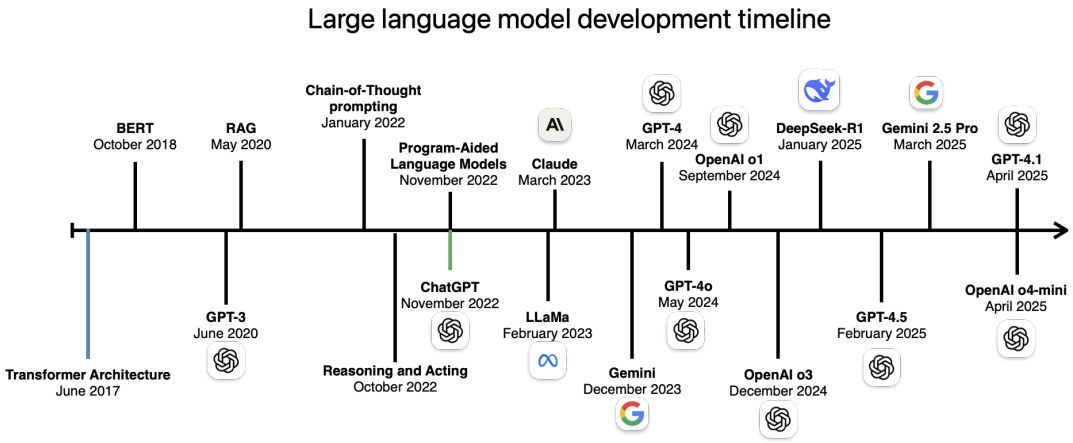


Figure 4: Large language model development timeline.

LLM technology is advancing rapidly. Figure 4 presents a timeline of key events in the field's advancement since the transformer was invented. The timeline shows that in the 2020s, one of the leading companies developing LLMs, OpenAI, released new GPT (Generative pre-trained transformer) language models with significant improvements at a rapid pace. GPT models have been in the general public's awareness since the first public version of ChatGPT, released on November 30, 2022 (OpenAI, 2022). Other competing services have come later and have had similar advancements, such as the DeepSeek R1 model and Google Gemini (Guo et al., 2025; Anil et al., 2023). New models vastly outperform their predecessors, and a year-old method may not

be capable of performing the tasks expected of the latest model by default. At the time of writing this thesis, publicly available technology is significantly outpacing published peer-reviewed scientific research about LLMs. The reader should expect that any mention of specific AI models will become obsolete in a few years.

Natural language models are pre-trained with vast amounts of data that teach them to understand the grammar and meaning of sentences (Hagos et al., 2024). LLMs are particularly effective in tasks involving general or widely available knowledge. However, language models have still exhibited several flaws. Wang et al. (2024) highlighted four flaws:

- There are privacy issues as the models may include sensitive data in their training dataset.
- The models can be biased as they gather information from various sources across the internet.
- The model may produce hallucinations, plausible-sounding but factually incorrect outputs presented with confidence.
- The model may violate intellectual property laws, as it may copy information from the source without proper attribution or citation.

These issues are pivotal when working with risk management and public actors such as companies and organizations. LLM biases are reflections of human biases, but some are also intentionally caused by hostile entities (Sadeghi & Blachez, 2025).

Risk management is a practice where justification of knowledge plays a pivotal role, especially when utilizing LLMs. LLMs based on general knowledge may perform poorly in a specific context, and a black-box model does not provide reliable sources for the given information (Gao et al., 2023). Highly relevant context data enables the model to perform well in more specific tasks.

3.2 Hallucination and interpretability

Hallucination is one of the primary issues in LLMs. It means that the model output is plausible-sounding but factually incorrect (Huang et al., 2025). Huang et al. divide hallucinations into two subcategories: factuality and faithfulness hallucinations. Factual hallucinations are inconsistencies in the outputted facts. Faithfulness hallucinations, on the other hand, are cases where the output is inconsistent with the user input, meaning it does not follow the given instructions. Huang et al. identify several sources of hallucination: untrue or biased data, hallucinations from training, and hallucinations from inference, where output data is highly likely, based on the embedding, but still incorrect. LLMs are not actively trained but instead rely on data from months or several years back, depending on the model, and this leads to some data being outdated by default.

Hallucinations are part of a more significant problem of a lack of LLM interpretability. LLMs tend to be black box models, where fully understanding the system's

functionality is nearly impossible due to their complexity. Singh et al. (2024) note that using LLMs in high-stakes tasks is impossible if interpretability is weak, as such tasks often require source justification. One LLM provider, Anthropic, claims that their citation mechanism reduces hallucinations to a minimum (Anthropic, 2025). The citation mechanism may be more computationally expensive than standard output. Nevertheless, it may be the most straightforward solution for the interpretability problem as it uses direct copying, allowing users to verify the source of the information. Similar methods can be expected for other LLMs in the near future, as the models tend to mimic each other's features.

3.3 Embedding

Large language models try to predict the most probable output for the given input. The core technique in LLMs is vector embedding. An embedded vector is a numerical representation of a corresponding text. Already, the earliest neural language models have used the technique. Modern models continue to rely on embedding, using it much more extensively, and even the attention mechanism used by transformers relies on the principle of measuring and updating the relationships between embeddings (Minaee et al., 2024).

LLMs must convert human text into a format that computers can process to effectively understand language. The embedding technique maps words with contextually similar meanings to coordinates in a high-dimensional vector space. However, modern LLMs also use an embedding mechanism to find relationships between different levels of abstraction, not just single words (Tennenholtz et al., 2023; Geva et al., 2020). The fundamental unit in embeddings is a token, which corresponds to a part of a word or a concept and has its unique position in the embedding space. LLMs are based on finding the correct relationship between the tokens in the embedding space, thereby being able to retrieve the contextually most relevant input (Minaee et al., 2024).

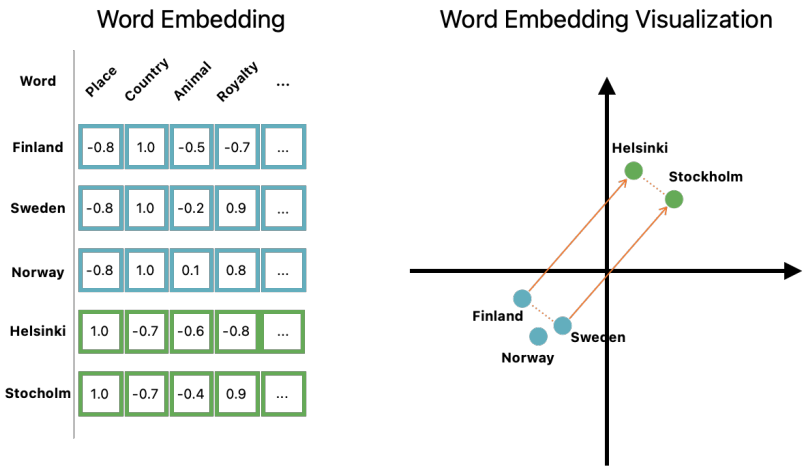


Figure 5: Example of word embedding demonstrating the relationship between different concepts.

Figure 5 shows a simple example of how embedding vectors work. Embedding vector place conceptually similar tokens near each other in the vector space. In the example, the words *'Helsinki'* and *'Stockholm'*, both capital cities, are placed near each other in the vector space. Similarly, the words *'Finland'*, *'Sweden'*, and *'Norway'* are conceptually close to each other, as they are all Nordic countries. The words *'Finland'* and *'Sweden'* also have a similar distance between them, as do *'Helsinki'* and *'Stockholm'*. With embeddings, the LLM does not need to store every possible question and answer. Instead, it can generalize from similar examples in its vector space. While modern LLMs include many layers performing different functions, embeddings remain a core component across all natural language models.

3.4 Retrieval-augmented generation (RAG)

In fields requiring domain expertise, in addition to the problem of hallucinations discussed in Chapter 3.2, LLMs are also challenged by their lack of context-specific knowledge and the fact that the model data is often outdated. Lewis et al. (2020) introduced the RAG technique to tackle these issues, using an external database with relevant knowledge instead of modifying the model's initial embeddings. RAG works such that the retriever searches for the appropriate data from the embedded database, and the generator produces the content using LLM. Modern RAG systems retrieve information from an external embedded database using methods conceptually similar to the LLM's internal retrieval processes.

Gao et al. (2023) list several benefits of the RAG. The significant advantages are that RAG improves accuracy and reduces factual hallucinations if the database contains relevant information. RAG stores data in a separate database and is thus much more secure than directly fine-tuning LLM with new data. Although RAG addresses many issues raised by Wang et al. (2024), it is not a complete solution. Huang et al. (2025) note that while RAG does reduce factual hallucinations when the data is reliable and up to date, it can still lead to some hallucinations.

3.5 Extending AI assistants' capabilities with tools

AI assistants utilize tools to enhance their capabilities and overcome certain limitations of LLM reasoning. In an AI assistant context, a *'tool'* refers to any programmable function that the LLM can operate independently. Furthermore, intermediate reasoning steps can be used as steps in chain-of-thought reasoning, which Wei et al. (2022) observed to lead to more precise answers. Wei et al. note that while effective in language-related tasks, LLMs often fall short in areas such as mathematical reasoning and may lack relevant information in certain instances. Yao et al. (2023) demonstrates a broad use of different tools, such as search engines and application programming interfaces (API), with their reasoning and action mechanisms (ReAct). RAG usage is a step that allows the model to search data from the given embedded vector database (Gao et al., 2023).

Figure 6 depicts the basic multi-step reasoning flow. Tools can be either functional, executing a task, or they can help the AI assistant access new information. The use of

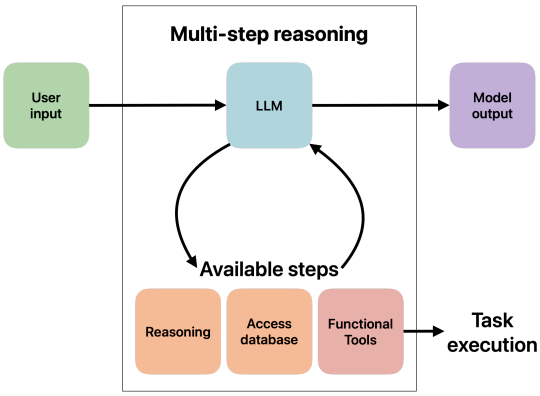


Figure 6: Basic multi-step reasoning flow.

tools, external data, and multi-step reasoning opens up numerous new opportunities, as a tool can be almost any function that can be programmed.

3.6 Knowledge sources

LLMs' knowledge is limited to what they are trained on, what the user provides, or what they can infer through reasoning. Chapter 2.6 explains the different dimensions of knowledge. It is self-evident that even LLMs can not know *'unknown unknowns'*, as they are unknown to everyone.

GPT-3 had 800GB of training data from academic papers, the internet, libraries, code libraries, and conversations (Gao et al., 2020). OpenAI trained other models with different datasets. However, they have not released this information about GPT-4 and newer models. GPT-4o is pre-trained only until October 2023 (Hurst et al., 2024), but internet access tools help bypass this limitation. Based on the large training sets, LLMs can at least access common knowledge or the *'known knowns'*. LLMs can also access information unknown to some people, so-called *'unknown knowns'*.

Information still exists that LLMs do not have access to. Much of the information is behind paywalls, and especially in commercial and state-held organizations, access to information is closely monitored, thus preventing LLMs from accessing confidential information. Another key source of information LLMs do not have access to is the information people have in their heads but do not share. The user can store the relevant information in RAG, which the LLMs would not otherwise have, and use it as a knowledge source.

4 AI in risk management: related studies

This chapter discusses recent studies on the use of AI in risk management. Several studies have investigated LLM usage in risk management, especially in expert-heavy tasks such as risk identification, assessment, and decision support (Collier et al., 2024; Esposito et al., 2024; Stødle et al., 2024). Although privacy, bias, and reliability concerns persist, recent research suggests that LLMs can also support risk management by enhancing expert analysis, extracting information, and proposing mitigation strategies.

4.1 LLMs in Risk Identification and Assessment Tasks

Collier et al. (2024) implemented a study on LLM performance in risk assessments for product recall risks, benchmarked with expert evaluations on the model's performance. Collier et al. asked the ChatGPT 3.5 model to identify risks, make a failure mode and effects analysis (FMEA), propose mitigation actions, and provide guidance. They based the experiment mainly on prompt engineering methods. Collier et al. concluded that the LLM's ability to yield new ideas was beneficial, but assessing estimates of risk likelihoods and severity was the least helpful task, and the model was inaccurate in these areas. They studied LLM capabilities, insufficient context data, and the phenomenon of hallucinations. Since the release of ChatGPT 3.5, the model's capabilities have improved, and its output is rarely nonsensical.

Esposito et al. (2024) conducted a similar study, enhancing the LLM models with fine-tuning and RAG. They noted that RAG helped find new risks and reduce hallucinations, and the fine-tuned model gave more accurate predictions. Esposito et al. found LLM assistance helpful for experts because it provides comprehensive analyses and helps uncover hidden risks. A key advantage of the LLM was that it sped up the analysis. Esposito et al. concluded that LLM should not replace human decision-makers but rather serve as a support tool for them.

4.2 Epistemic and ethical challenges

Stødle et al. (2024) studied LLMs and AI more broadly from the perspectives of consequence characterization, uncertainty characterization, and knowledge management. Like previous studies, they observed that LLMs are effective at generating initial risk lists but struggle to evaluate risks accurately. However, they found LLMs helpful for information and event extraction from unstructured data sources. These models can identify consequence patterns across extensive volumes of text, tasks that can be labor-intensive for human experts.

Regarding uncertainty characterization, Stødle et al. differentiate between two types: aleatoric uncertainty, which reflects inherent variability in data, and epistemic uncertainty, which arises from a lack of knowledge. While they found AI helpful in managing aleatoric uncertainty reasonably well, they performed poorly with epistemic uncertainty. They noted that current LLMs often generate fabricated or misleading

outputs rather than signaling when they lack adequate information. These limitations are important to consider when designing LLM-based tools for risk management.

The final perspective addressed by Stødle et al. was knowledge management. In risk analysis, they emphasized the justification of knowledge with clear evidence. While AI systems can support this process by uncovering relevant information, they cautioned against over-reliance, particularly due to the models' inability to express uncertainty about their knowledge limitations. Despite ongoing improvements in these areas, Stødle et al. highlighted a more profound ethical concern: "Should AI be allowed to prescribe risk management decisions automatically?" They argued that *risk-informed* decision-making, unlike *risk-based* approaches, requires an understanding of stakeholder preferences, a task that AI systems are currently incapable of. They concluded that human oversight must remain central in any AI-supported risk governance framework.

5 AI-assistant design and implementation

This chapter proposes an AI assistant prototype developed based on the concepts introduced in Chapter 3. The AI assistant is intended to (1) accelerate the risk management process, (2) assist with complex data handling, and (3) enhance the quality of the risk register, while (4) help ensure a reliable risk management process. The chapter covers the system design, system architecture, core features, and tools, as well as the scope of the design. The prototype includes specific tools to address the challenges outlined in Chapter 2.4.

5.1 System design overview

The AI assistant prototype is designed to support risk identification and assessment phases (see Chapter 2.2) by enabling users to interact more effectively with the risk register. It is designed to provide three core capabilities: retrieving contextual information, accessing relevant risk data, and proposing edits through a natural language chat interface (Figure 7). The assistant serves as an intermediate reasoning layer, combining information from multiple sources and enriching analysis through the use of specialized tools.

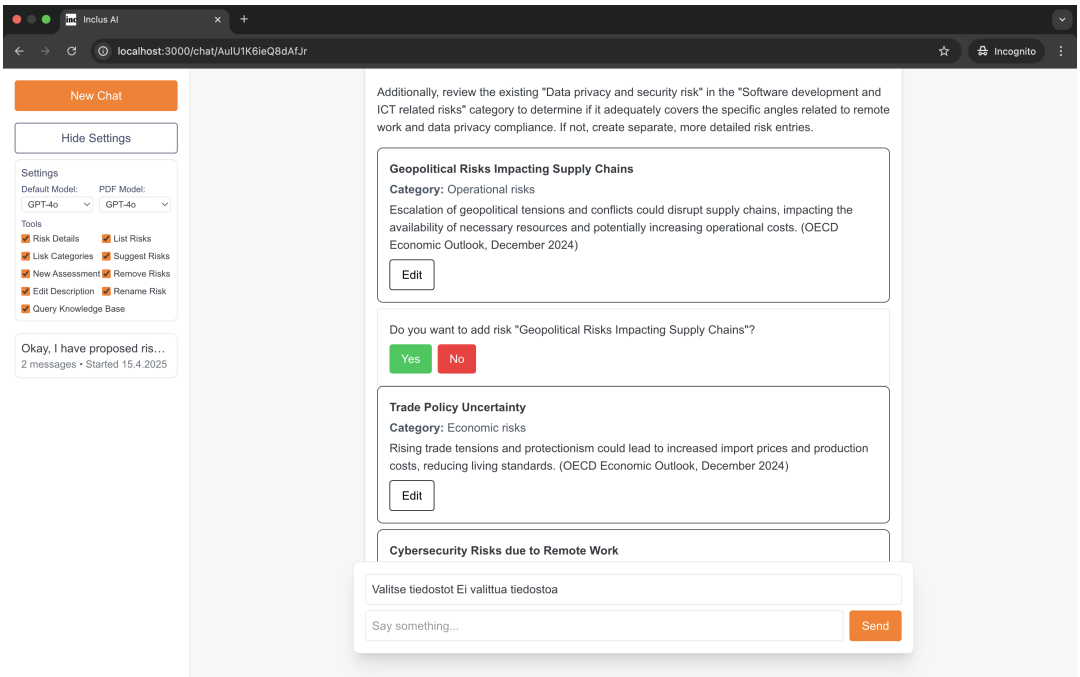


Figure 7: Chat-based user interface of the assistant.

The AI assistant is based on a large language model (LLM) at its core. The platform supports integration with various LLMs, including locally hosted models. While different models may be better suited for specific tasks, the prototype utilizes OpenAI's GPT-4 model via Microsoft Azure due to its reliability and security features, and Google Gemini to provide comprehensive multi-modal capabilities.

5.2 System architecture

This prototype was designed to utilize data from multiple sources rather than relying solely on LLM output. Figure 8 presents the system architecture of the prototype, where the assistant interface connects different data sources. The AI assistant accesses context through four primary sources: the structured risk register, RAG knowledge base, multi-modal documents such as PDFs (Portable Document Format), and stakeholder inputs. Each source plays a distinct role in supporting informed decision-making. Various data sources focus on enhancing the model's context.

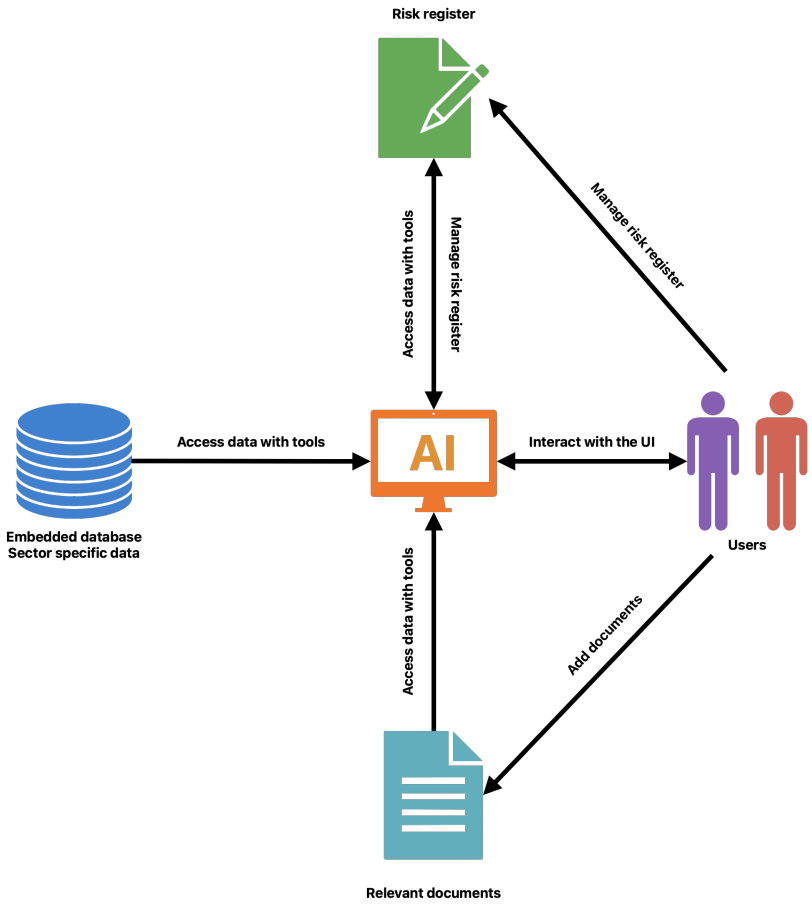


Figure 8: System architecture of the AI assistant for risk register management.

The following content provides a detailed explanation of risk register interactions, RAG search, and multi-modal document usage. Each of the data sources had its unique benefit. Chapter 7 delved more in-depth into the user interactions, and thus, it is not covered in this chapter.

5.2.1 Risk register interaction

The risk register is the assistant's primary structured source of information. The AI assistant is able to query the current, up-to-date, exact, and well-structured risk register

database as needed. In the prototype, the AI assistant is provided with the risk ID, category, risk assessments, risk descriptions, tasks, and tags for each risk, as presented in Table 2. The AI assistant can select the required elements for the set of risks based on the user input.

Table 2: Risk information provided to the AI assistant for each risk entry.

Element	Description
Risk ID	Unique identification number for the risk.
Risk description	Reasoning and description for the identified risk.
Name	Short title of the risk.
Category	Risk category (e.g., economic, sales, software).
Risk assessments	List of all the risk assessments of the given risk, including comments and estimates on risk metrics.
Tags	List of the tags, such as "urgent" or "critical".
Tasks	Mitigation tasks to be carried out.

5.2.2 RAG search

The AI assistant can use a RAG knowledge base to improve its context. RAG enables the efficient storage and querying of large volumes of data without requiring it to be embedded in the LLM. Every embedded vector includes metadata about the source of information. The prototype was provided with a comprehensive list of documents, listed in Appendix B1. The list consists of Inclus' internal papers and well-recognized reports.

To vectorize data, the prototype implementation uses OpenAI's text-embedding-ada-002 embedding model (OpenAI, 2022). RAG can only retrieve a small chunk at a time (512 tokens in this implementation), meaning that the questions must be precise and concise. It is unlikely that the model can explain an entire file. Answers are more accurate if they fit into the chunk size, for example, an individual paragraph. However, multi-step reasoning allows an AI assistant to interpret and make multiple RAG queries independently.

5.2.3 Multi-modal documents

The prototype supports the processing of multi-modal data, such as PDFs and images. Multi-modal processing requires more processing power from LLMs, but it enables more reliable document reading and allows users to verify documents immediately. Extracting information from PDFs has historically been a challenging task, which LLMs can do with high accuracy (Edwards, 2025). Unlike RAG, the questions about multi-modal documents do not need to be as specific to obtain valuable answers. With RAG, the context of the user input comes from the input text. However, with multi-modal documents, the document itself provides the context for the model.

5.3 Core features and tools

The AI assistant can utilize tools to retrieve data from the specified sources. Multi-step reasoning enables the model to use multiple tools with a single user query. The AI assistant can recognize the tool request without explicitly stating the tool name, and will ask clarifying questions from the user in case of unclear input. For example:

User: "Add a new operational risk."
AI assistant: "Please provide details about the new operational risk you want to add, including its name and description."

The assistant uses tools to interact with the risk register and retrieve context-specific data. The prototype has a curated list of tools to demonstrate its key capabilities. The list encompasses nine data retrieval and register editing tools, summarized in Table 3. The tools are selected to reflect common use cases, with the option to add others as needed.

Table 3: List of AI assistant tools.

Tool name	Purpose	Type
List categories	List categories in the register	Data retrieval
List risks	List risks within given categories	Data retrieval
Risk details	Show full risk details	Data retrieval
Query knowledge base	Fetch data from the knowledge base using RAG	Data retrieval
Suggest risks	Propose a new risk	Risk update
New assessment	Suggest a new assessment	Risk update
Remove risks	Propose removing a risk	Risk update
Edit description	Edit the risk description	Risk update
Rename risk	Propose a new risk title	Risk update

The AI assistant can suggest *"risk update"* actions, where the user can accept, decline, or modify the proposed update. It can also utilize *"data retrieval"* actions to refine any suggestions, and it can process multiple tool actions with a single user query. The primary value of the tools lies not in automated editing but in the AI assistant's ability to generate intelligent, context-aware suggestions.

Some built-in capabilities, such as follow-up questioning, basic reasoning, and multi-modal input (e.g., PDFs or images), are handled internally by the LLM rather than through explicit tools. Certain LLMs offer optional built-in features, such as citation support, deep reasoning, and web access. While not central to this prototype, these features can be valuable in future developments.

5.4 Application scope

In the design of AI assistants, users must always accept or decline proposals. The implementation does not allow the AI assistant to make autonomous edits directly to the risk register for two reasons. The technical reason is that the AI assistant may

hallucinate and cause damage to the risk register. Therefore, users should have control over updates to the risk register. The second reason is that the fundamental purpose of risk management is to lead to mitigation actions. Participants must be aware of the process to ensure they are committed to it and take the mitigation steps. A fully automatic process would hinder the participants' involvement.

Chapter 7 presents a case study and results in which the AI assistant helped improve the Inclus risk register. This evaluation assesses the practical implications of the assistant and explores how such systems can enhance expert-driven risk management in complex, real-world contexts.

6 Case study research design and methodology

The case study was conducted to test the prototype introduced in Chapter 5. The case study was conducted at Inclus, where the AI assistant is intended to be implemented. Inclus was used as the target of the case study, as this allowed for more control over the process and provided easy access to many risk management experts. This chapter outlines the research approaches employed in the case study and presents the four research questions (**RQ**s) that guided its development. The last subchapter describes the data collection methods.

6.1 Research Approach

This thesis examines how domain experts perceived the AI assistant and how it altered the structure, clarity, and redundancies of the enterprise risk register. Following the structure from Yin (2018), the **case study** approach was suitable as

a) The research questions (see Table 4) were in the form of *how*.

b) The thesis studied phenomena outside the author's control in a real-world organization.

c) The thesis studied contemporary events.

The study required long-term observations and access to organizations' risk management processes. It focused on a **single case study**, observing a particular case (Yin, 2018, pp. 47-54) (Inclus Oy). However, the case was also **instrumental**, demonstrating the larger issue of AI assistants supporting the management of risk registers (Stake, 1995, pp. 3-4). Analytical generalization suggests that these findings apply to organizations with similar risk management practices.

The case was analyzed at multiple levels to gain a deeper understanding of the situation. Yin (2018, pp. 51-54) describes this as an **embedded case study**. The embedded units of analysis include:

- **Risk register entries** (before and after AI support) to compare their comprehensiveness and manageability.
- **Expert perceptions** through Likert questionnaires and semi-structured interviews.
- **AI assistant interactions**, including prompts, suggested actions, and user acceptance.

The thesis followed a standard, iterative "*build* → *demonstrate* → *evaluate*" approach from design science (Peffers et al., 2007; Hevner et al., 2004). It included the development of the AI assistant, multiple demonstrations for the case study organization, and evaluation based on the research questions. Design science facilitates the iterative development of artifacts in real-world contexts. In this study, an AI assistant for risk register management worked as an artifact. Thus, this case study was

formulated as an **embedded instrumental single-case study**, with elements from **design science**.

6.2 Research questions

LLMs can be used to automate reasoning tasks, sometimes outperforming humans. However, they perform better in some tasks than others. Additionally, AIs should not be allowed to work autonomously in critical tasks (see Chapter 4.2). Based on the questions about AI risks and opportunities, this thesis formulated four research questions to be answered. The research questions are introduced in Table 4.

Table 4: Research questions for the case study.

Abbrev.	Research question	Explanation/focus of analysis
RQ1	How does the use of an AI assistant influence the speed of the risk-management workflow?	This question examines how the assistant impacts overall workflow speed, specifically by identifying new risks, merging overlaps, removing duplicates, and refining risk descriptions.
RQ2	How does an AI assistant help to process and organize large and complex risk data?	This question examines how the assistant processes and organizes complex risk data by integrating information from the risk register, user prompts, and the RAG knowledge base.
RQ3	How does using an AI assistant affect the quality of the risk register?	This question examines how the assistant enhances the quality of the risk register in terms of completeness, consistency, and depth of insight.
RQ4	How is the reliability of the AI assistant perceived and addressed in expert workflows?	This question examines how experts perceive the reliability of AI assistants, particularly in terms of hallucinations, biases, and their views on model transparency.

6.3 Research data collection methods

The primary analytical technique was pattern matching between different embedded units of analysis (risk register, expert perception, and AI assistant interactions) and between various sources of research evidence (see Table 5). Each source of evidence is related to the research questions from slightly different perspectives. Various perspectives align with the embedded units of analysis provided.

Table 5: Sources of evidence and related research questions.

Evidence	Description	Related RQs
Questionnaire	Descriptive questionnaire on participants' perceptions of the AI assistant usage on risk registers. Embedded at the beginning of the risk assessment form.	*RQ1*, *RQ3*
Interviews	One-hour, semi-structured interviews, where questions were open-ended, allowing the interviewer and interviewee to discuss freely. The interviews aimed to collect domain experts' perceptions of AI assistants with more in-depth answers than just the questionnaire (Yin, 2018, pp. 118-121).	*RQ1*, *RQ2*, *RQ3*, *RQ4*
Observations on the risk register	Observations on how the AI assistant transforms the risk register. Analysis of new risks, risk removals, new descriptions, new names, and risk aggregation.	*RQ2*, *RQ3*
Observations on the AI assistant	Observations of the AI assistant outputs, usability, and code-related observations.	*RQ1*, *RQ2*, *RQ3*, *RQ4*

Research evidences were collected from multiple sources to understand the case comprehensively, as different data sources complement each other (Yin, 2018, pp. 113-125). Formal data collection methods included a questionnaire for the participants and interviews with key informants in the case study organization. Additionally, observations were made on the risk register and AI assistant usage.

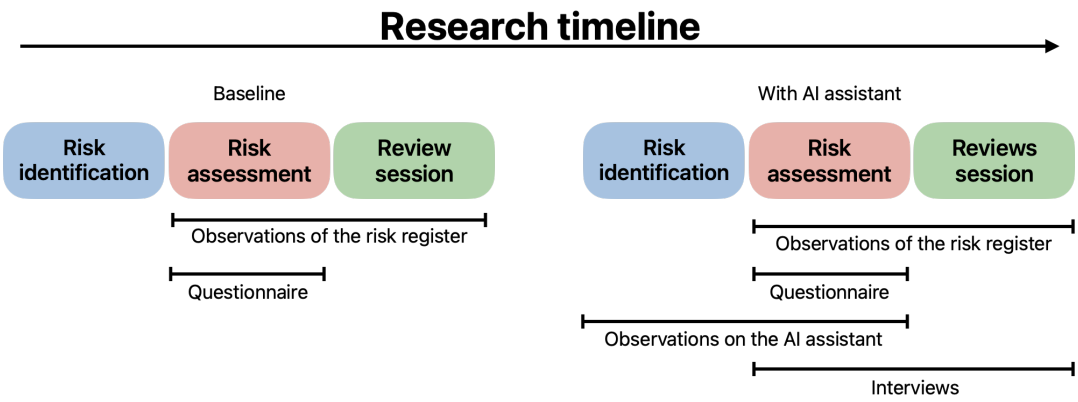


Figure 9: Research timeline.

The study was conducted across two iterations of the risk management process during the first half of 2025. Figure 9 depicts the timeline for the case study concerning the risk management phases and corresponding sources of evidence. The first iteration involved a risk management process with baseline (no-assistant) settings. The second iteration followed the same process but was conducted using the AI assistant to enhance the risk register. Both iterations included the entire risk management process, which consisted of identification, assessment, and review sessions.

7 Case study results

This chapter examines how the AI assistant enhanced Inclus's enterprise risk register. The chapter describes the implementation of the case study and the results from various sources of evidence, including risk register entries, questionnaires, interviews, and observations of the AI assistant. The case study provided results for each of the research questions (*RQ1-RQ4*). The last subchapter summarizes the key points of the case study.

7.1 Case study context

The case study examines Inclus, a company that develops participatory risk management software (see Chapter 2.5). Inclus' risk management workflow works as an iterative process, aligning with the framework proposed by Patterson & Neailey (2002) and illustrated in Figure 1. The case study consisted of two iterations of the risk identification and assessment workflow: a baseline run without the AI assistant, followed by an AI-assisted run. The risk register has been maintained since 2021, with new iterations updated based on the results of previous risk assessments. Accordingly, the AI-assisted run is updated based on the baseline risk register. The case study focused primarily on the early stages of this workflow, specifically risk identification and risk assessment, where the AI assistant offered the most significant added value. Each identification or assessment session is a dedicated survey that remains open for two weeks, allowing participants ample time to contribute. The following paragraphs describe how risk identification and assessment are conducted at Inclus, providing a brief outline of the case study participants.

Identification Risk identification is a phase of the Inclus risk management process, where participants can suggest new risks, comment on any newly suggested or existing risks, and flag any risks they believe to be relevant. Admin users can modify risks by changing names, descriptions, or categories, and archive, remove, or add any risks. Generally, admin users proceed with operations based on participant comments, suggestions, and flagging. In the AI-assisted run, the assistant's recommendations significantly influenced which risks were modified or merged, a difference from the baseline, where all changes originated solely from participants.

Assessment Risk assessment in Inclus is a phase purely guided by the participants. Participants evaluate the risks for which they believe they have relevant expertise. For each selected risk, they provide scores for three criteria: negative impact, likelihood, and positive impact. However, the case study focused mainly on the negative impact and likelihood. Each is rated on a scale from 1 to 5, with decimal values allowed. A score of 1 represents a very low impact or likelihood, while a score of 5 represents a very high impact or likelihood. In addition to numerical assessments, participants can add comments and suggest mitigation tasks related to each criterion. Assessments allow multiple participants to evaluate the same risks.

Participants Participants were invited to the identification and assessment surveys. Table 6 presents the number of participants from each department in the case study organization who were invited to the risk management process. Of the 17 invited, 12 (71%) participated in the risk management process during the baseline run, and 12 (71%) participated during the AI assistant-enhanced run.

Table 6: Invited participants for the risk management process.

Team	Invited
Software development	6
Sales, marketing, and CSM	11

7.2 Risk register entries

This subchapter examines in detail how the use of an AI assistant changed the risk register. Observations of risk register entries answered the complexity (*RQ2*) and quality (*RQ3*) issues. Table 7 shows the difference before and after applying the AI assistant to enhance the risk register, showing a clear difference in the number of risks. The following content provides a detailed discussion of the reasons for risk reduction.

Table 7: Risk register content with the baseline and AI-assistant-enhanced runs.

Risk categories	No. risks (baseline)	No. risks (AI enhanced)
Strategic risks	11	8
Operational risks	10	11
Software development and ICT related risks	17	10
Sales and Marketing related risks	7	3
Economic risks	6	3
Special threats and opportunities for Inclus	7	0
Total	**58**	**35**

7.2.1 Baseline risk register

The baseline run without the AI assistant identified 58 risks; however, only 49 had a corresponding description. The risk register included multiple duplicate risks, such as *'AI development'*, which overlapped with *'Exponential AI development disrupting software market'*, and *'Reputation risk'* overlapped with *'Negative public perception'*. Risks were proposed by individual experts, thus reflecting their style, leading to an incoherent risk register. For instance, one entry was a very general *'General IT risk'*, while another was extremely specific *'Software development is unable to keep up due to too many requests'*. Such disparity in scope illustrates the incoherence of the baseline

register. Suggestions from multiple experts are beneficial in their own right, but they can undermine the consistency of the register. Some risks had no description, and only a few had a comprehensive description. The descriptions often did not provide any new insights about the risks and merely stated the obvious. These issues relate to *RQ2*, as processing a large amount of data appears to be a cognitively demanding task for humans. These issues indicate a lower quality risk register regarding naming and description consistency, addressing *RQ3*.

Figure 10 presents the assessed risks in the risk matrix based on the average impact and likelihood estimates for each risk. The risk matrix hides the exact risk names for confidentiality reasons. The matrix shows that the risk register includes many risks that should be considered accepted (i.e., risks so minor that the organization accepts them without further mitigation). Listing threats and opportunities in the same risk register makes it challenging to distinguish between them.

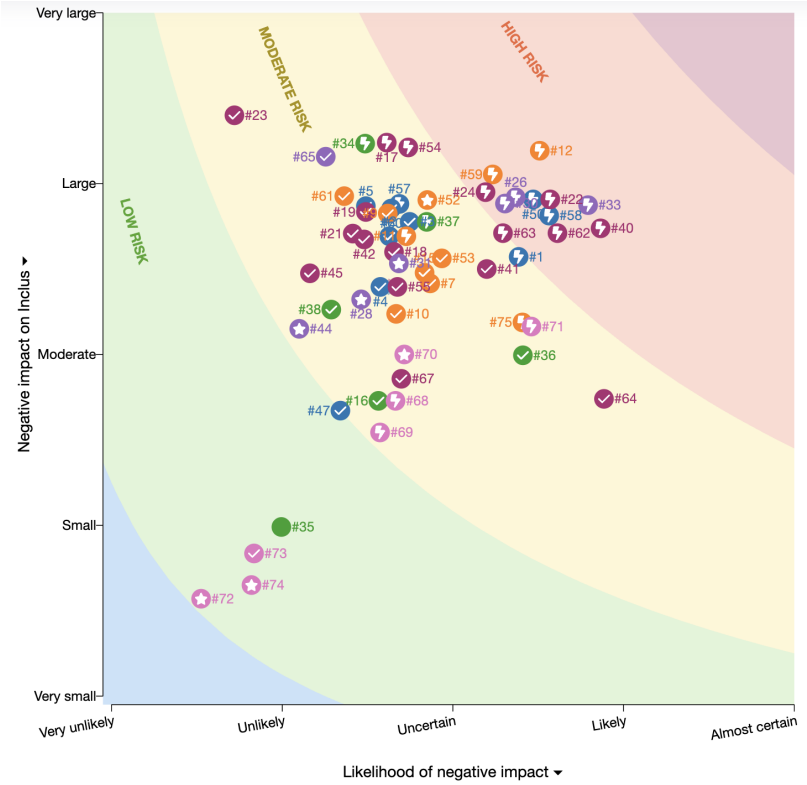


Figure 10: The risk matrix in March 2025 is used as a baseline without an AI assistant.

7.2.2 AI assistant enhanced risk register

In the second run, an AI assistant suggested modifications for the risk register, meaning removing unnecessary risks, adding new ones, and combining duplicate ones. The AI assistant proposed removing the *"Special threats and opportunities for the Inclus"* category entirely, as it was too vague. The AI assistant suggested moving the risks to other categories or removing them. Table 7 shows that the total number of risks dropped from 58 to 35 compared to the baseline. The reduction was due to the merging

of duplicate risks and the removal of accepted risks. The high number of duplicate risks in the baseline underscores the challenge of consistently managing a large set of risks without AI support. The improved risk register indicates that the AI assistant offers a clear benefit in managing large and complex datasets, addressing *RQ2*.

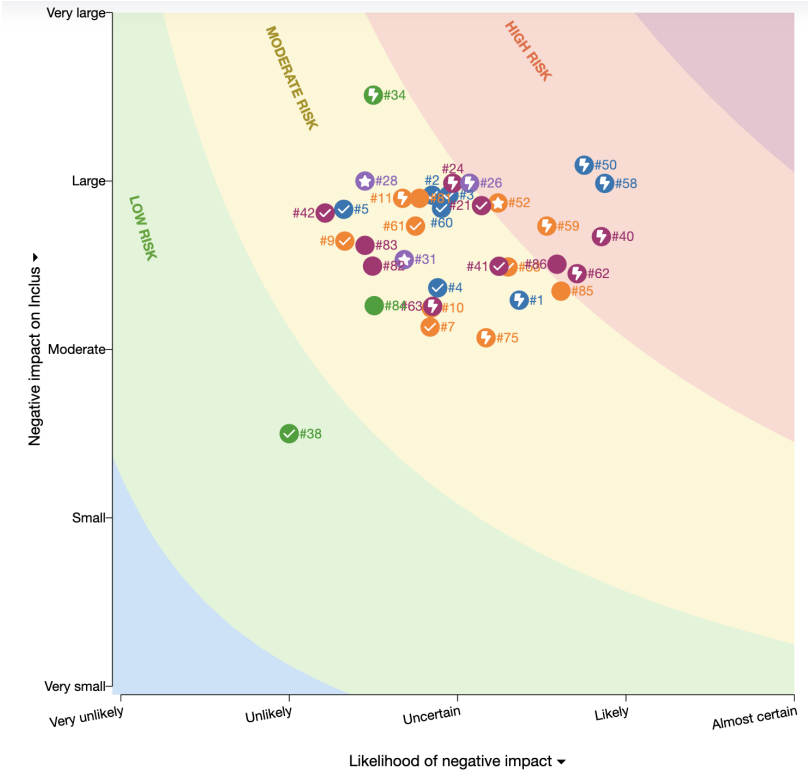


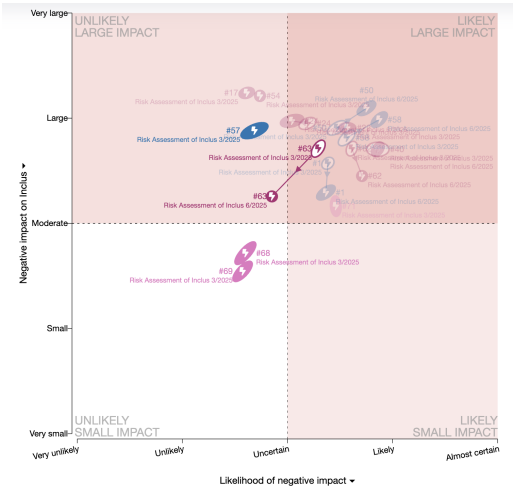
Figure 11: The risk matrix in June 2025 with an AI assistant-enhanced risk register.

Figure 11 shows how the use of the AI assistant changed the risk register and led to improvements. The combination of risks explains many reductions, particularly those associated with high-impact and high-likelihood risks. Note that some risks moved in the risk matrix between March and June, partly due to natural variability in participants' scoring and changes in the risk landscape between sessions.

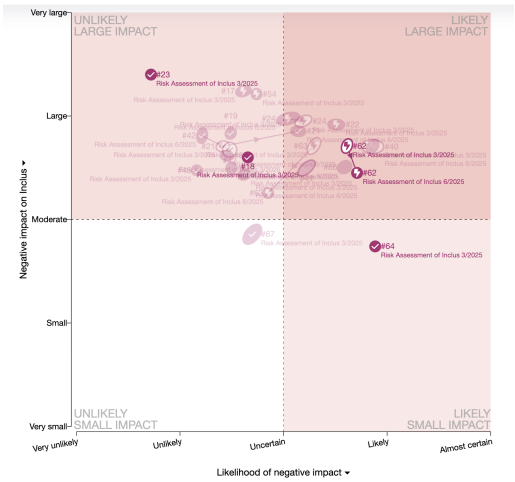
Table 8: Risk register actions.

Action	Amount	Impact on number of risks
Combined risks	30	-21
Removed risks	8	-8
Added risks	6	+6
Modified risk descriptions or name	20	0

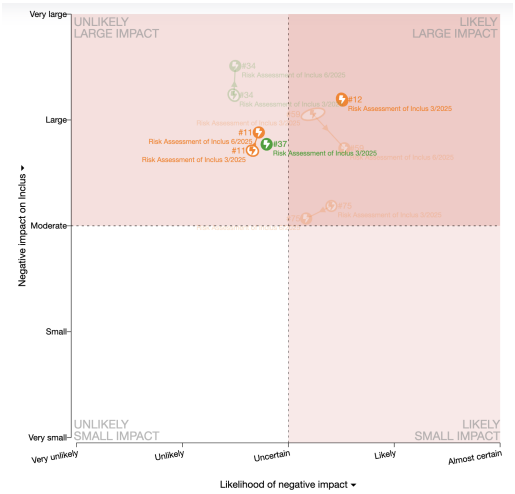
As Table 8 shows, the AI assistant suggested combining 30 risks into 9, which yielded a net reduction of 21 risk entries. It also recommended removing eight risks



(a) Risks #57, #63, #68, and #69 combined into new #63.



(b) Risks #23, #18, #64, and #62 combined into new #62.



(c) Risks #11, #12, and #37 combined into new #11.

(d) Risks #61, #17, and #54 combined into new #61.

Figure 12: Four examples of risk combination.

outright and adding six new ones. Twenty risks were modified in name or description without changing the count. Each risk could only have one of the actions, meaning the actions did not overlap. After the AI assistant suggestions, all risks had descriptions, and many descriptions had concrete examples and citations to documents in the Appendix B. The AI assistant's suggestions to combine risks and enhance descriptions did not eliminate any relevant risk topics, indicating that no vital content was lost while quality and coherence improved (*RQ3*).

Figure 12 shows four examples of risk combinations. The figure highlights the risk involved in each combination with darker colors. The new combined risk is the risk indicated by the arrow (new risks were #63, #62, #11, and #61). Cases (a), (c), and (d) have a pattern where risks have very similar likelihood and impact estimates and describe the same or similar risk. (b) shows four risks that describe one issue at different scales, where the assistant suggested merging them into a single risk (#62). The AI assistant suggested all combinations while humans did the risk assessments. In all cases, the new combined risk is close to the middle of the duplicate risks from the baseline risk assessment session. The AI assistant suggested combining the risks and revising their descriptions, allowing the expert to assess them consistently in the middle. In many cases, the new combined risk had less variation in assessment. Skewed dots display assessment variation, such that larger dots have larger variation.

7.3 Questionnaire

The risk assessment form included a questionnaire at the start. Due to the small sample size, the questionnaire's results are descriptive rather than inferential. The study primarily focused on how the AI assistant affects perceptions of the risk register's quality, readability, and manageability, as well as attitudes toward AI usage.

The questionnaire consisted of ten Likert-scale questions (rated 1 to 5) and one open-ended question (see Appendix A), with 1 indicating complete disagreement and 5 indicating complete agreement. Embedding the survey within the risk assessment form ensured high response rates and ensured the contextual relevance of the answers. The risk management process session was conducted twice: first, with baseline settings and no AI assistant involvement, and later, with AI assistant support. Both iterations used the same questionnaire. Out of 17 invited, 8 (47%) participated in the questionnaire during the baseline run, and 7 (41%) participated in the questionnaire during the AI assistant run.

Table 9 shows the average answers and p-values based on those. All questions showed higher average scores with the AI assistant. Despite the small sample size, four of these differences were statistically significant at the 5% level (one-tailed Mann-Whitney U test). Answers indicate that the AI-enhanced risk register had (Q1) more precise descriptions, (Q2) fewer gaps in coverage, (Q5) uncertainty levels easier to interpret, and (Q8) a more up-to-date register. From the participant's perspective, the risk register has improved quality, addressing *RQ3*.

Table 9: Calculated p-values of the Mann–Whitney test from the questionnaires.

Question	Avg. (baseline)	Avg. (AI-enhanced)	p-value
How clearly are the current risks described in the risk register?	3.11	3.86	**0.036***
Does the risk register include all relevant risks, or do you identify any gaps?	3.43	4.29	**0.048***
Is there a right amount of risks (not too many, not too few)?	2.38	3.14	0.136
Do you feel that the risk register also accounts for unforeseen risks (e.g., “black swans”)?	2.50	2.86	0.124
How easy is it to interpret the level of uncertainty associated with risks in the risk register (i.e., does it show how well we know the risk)?	2.75	3.43	**0.036***
Does the risk register accurately capture and reflect disagreements in risk assessments?	3.13	4.00	0.177
Do you believe that risk assessment is based on sufficient and high-quality information?	3.38	3.86	0.226
Do you feel that the risk register is up to date?	3.00	4.29	**0.016***
How reliable do you consider the risk assessment methods used in the risk register?	3.25	3.57	0.163

* Statistically significant difference $p < 0.05$.

Questions with a statistically significant change are marked in bold.

Participants also answered an open question: "**Could this risk management process be somehow improved? Are there challenges or uncertainties in the process you would want to address?**" The question got four answers during the baseline run and three during the AI assistant-supported run. During the baseline run, the primary concern was that the assessment was too long, as mentioned in all four answers. Participants from the customer success management team commented:

> *"This creates a feeling that content is more distracting than helpful and supportive in discussing the key issues."*

Another issue was the vague and overly broad risk definitions or out-of-context descriptions mentioned in the three answers. All answers mentioned that the *opportunity* criteria were confusing and should have been assessed separately.

Participants were much more satisfied with the outcome during the AI assistant-supported run. One participant from the software development team noted:

> *"The risk content is now more harmonized and at the same level, so there is significantly less mental whiplash when jumping from one level to another."*

This firsthand perspective underscores the improvement in consistency. However, some participants requested more concrete examples in the risk description and AI suggestions of new, unforeseen risks. The answers strongly indicate improved quality (*RQ3*). Participants reported that the second assessment felt shorter and less taxing, suggesting the AI-harmonized register made the process more efficient (*RQ1*).

7.4 Interviews

The study included four interviews to gain a more in-depth understanding of domain experts' perceptions of AI assistant usage in a risk management context. The role of each interviewee is presented in Table 10. All interviewees had a comprehensive understanding of the organization and its risk management practices, and frequently interacted with customers. All interviewees had previously used LLMs, such as Google Gemini and OpenAI's ChatGPT, in risk management and thus had clear visions of the AI assistant usage in risk management. Each interview followed the same core questions, aligned with our four research questions (*RQ1–RQ4*), with follow-ups tailored to each expert's background. This chapter synthesizes the insights of interviewees, integrating project examples and expert remarks.

Table 10: List of interviewees.

No.	Role
#1	Chief Executive Officer
#2	Head of Customer Success Management
#3	Vice President of Commercial Management
#4	Customer Success Management Assistant

Interviewees were first asked how an AI assistant would speed up the risk management process, answering *RQ1*. All interviewees identified two main benefits: an AI assistant can be used to harmonize risk registers (e.g., removing duplicates and improving descriptions), and it can also generate new risks and scenarios. Interviewee #2 also added that AI can be utilized in customer cases involving foreign languages, as LLMs can process multiple languages with ease. All interviewees noted that the AI assistant dramatically speeds up these tasks and often produces output comparable to or better than that of an average human. However, they also observed that the AI does not generally surpass top domain experts in specialized analyses. Interviewee #1 noted that managing the Inclus risk register during the assessment session took him two days, whereas with an AI assistant, the task took minutes. Interviewees #1, #3, and #4 shared experiences where they had used AI to generate a risk list, doing several days' work within minutes, with sufficient quality. Interviewee #2 commented:

> *"AI would reduce the cognitive load, leaving time and energy for the more critical tasks."*

According to the interviewees, many steps should be sped up with AI. However, fundamentally, risk management relates to humans' preferences in managing risks. Interviewee #1 added that while technically, AI could make decisions related to risk management and determine the risk appetite and objectives, organizations should not allow AI to decide these for them.

Interviewees were then asked how they see AI assistants in terms of their capabilities in processing large and complex datasets, relating to *RQ2*. Interviewees believed the ability to use data from multiple sources is the most valuable feature. Interviewees had experience in using AI with various case-specific sources to generate a list of new risks for an empty risk register. They noted that AI can identify risks that an expert would not otherwise consider. Interviewee #4 added that the precision of the context data is more important than the amount. Interviewee #1 commented that AI is like an extended memory, which performs consistently and is not affected by emotions and does not get tired. AI can also change the role of quantitative data in risk management. Interviewee #3 noted:

> *"Dialogue among participants is at least as valuable as the actual risk estimates."*

He believed that AI increases the role of quantitative data, as it is capable of summarizing risks, comments, and discussions from multiple sources. He noted that this would enhance the overall participatory process. Interviewee #2 added that AI can be utilized to combine information across risk registers and draw high-level summaries for decision-makers.

The third question concerned the quality of AI output, relating to *RQ3*. Interviewees believed that AI can be utilized to enhance the consistency and precision of the risk register. Interviewee #3 commented that with AI, the descriptions become more uniform and detailed, reducing both overly generic and overly narrow formulations. Interviewee #3 believed that currently, a top-level expert performs better than AI in specific analyses, but added:

> *"Experts do not necessarily have a fundamental advantage over AI."*

This relates to what AI can and cannot know (see Chapter 3.6). Interviewees #1, #2, and #3 commented that currently, AI cannot get the same experiences and tacit knowledge as a top expert. Interviewee #1 added that in a fast-paced business environment, analyses rarely go that deep, and AI analyses are often good enough. Interviewee #1 noted that organizations can utilize AI in tasks where they lack expertise. However, interviewees also commented that AI sometimes gives hallucinatory outputs, and the analyses do not always bring more value. They highlight the need for human oversight, and interviewee #1 added that the AI assistant should be prompted with clear objectives.

The last question related to the reliable use of AI in risk management, relating to *RQ4*. Interviewee #3 emphasized that fear of hallucinations and biases prevents companies from adopting AI in their work practices. However, interviewee #2 argued that having a transparent and reliable workflow is even more important than the

occasional AI hallucination. If the process is well-designed, users can detect or mitigate bad outputs. He believed that a reliable process would enable experts to learn how to use the AI assistant according to proven practices. Interviewee #2 also pointed out:

> *"Risks are uncertain until they are realized."*

This means that neither AI nor human experts, who identify risks, can predict the future. Therefore, in the identification and assessment phases, hallucinations rarely have a significant negative impact. However, interviewees noted that some cases do have sufficient data, which should be utilized reliably. Interviewee #1 commented that risk management is an especially challenging field because it requires a combination of creativity and evidence-based thinking. The level of exactness varies between organizations, processes, and individual risks. Interviewees remarked that AI usage should be designed reliably, with human oversight always present. This also means that AI should not be allowed to take actions independently. Instead, a human should be aware of all actions, whether accepting or rejecting them, and be mindful of possible hallucinations and biases. Interviewees emphasized that humans should also understand the process, as risk management ultimately supports informed decision-making and should reflect the values and preferences of stakeholders.

7.5 AI assistant interactions

As the primary user and developer of the AI assistant, the author's insights here may be influenced by that close involvement. However, wherever possible, these observations are linked to the research questions and triangulated with other evidence. The AI assistant development practices are explained in Chapter 5. Observations from the AI assistant usage provided insight for each research question (*RQ1*, *RQ2*, *RQ3*, and *RQ4*).

Table 11 summarizes six representative interactions with the AI assistant. In each case, the assistant was prompted with a specific task (left column) and produced a result (middle column). The right column notes any issues encountered. The user accepted all the significant suggestions listed in Table 11 (with minor edits), whereas some suggestions from other, less critical interactions were rejected. These six cases are highlighted because they led to the most considerable improvements in the register.

Cases #1, #2, #5, and #6 were routine, mechanical tasks (e.g., rewriting descriptions and merging duplicate risks) that demand careful, human-level reasoning to execute correctly. In other words, they are the cognitively taxing yet straightforward tasks an AI can handle efficiently. According to experts who had previously done this pre-processing manually, it would take multiple days of work to achieve what the AI did in minutes (*RQ1*).

Table 11: Example AI assistant queries.

Case	Question description	Tools used	Answer description	Issues
#1	Asking for a description of risks that had no description	Modify description	Provided comprehensive descriptions for nine risks with references to files in Appendix B.	Some citations were inaccurate
#2	Asking for better descriptions with proper citations to the rest of the risks	Modify description Rename risk	Provided descriptions of up to 21 risks at a time with references to files in Appendix B.	Some citations were inaccurate
#3	Asking to list all Inclus risks from the risk register and count them	List risks	Listed all risks by category and reported the count per category.	Miscalculated the number of risks
#4	Asking to suggest new risks based on the risk register	Suggest risks	Proposed ten new risks.	Three accepted as-is; four merged; three discarded
#5	Asking to propose risks to be removed or combined	Remove risks	Suggested removal of 18 risks, with justification.	Only eight were from the baseline register. The others were previously proposed by the AI assistant.
#6	Asking to combine duplicate risks	Suggest risks Remove risks	Proposed 20 duplicate risks for removal and added eight combined risks.	—

Case #6 (Combine duplicate risks) was especially complex, as the assistant was prompted to merge risks by analyzing 19 other entries' criteria, comments, and descriptions. The successful combinations suggest it effectively synthesized information across the entire risk set (addressing *RQ2*). Also, the suggested combinations had to be coherent with the rest of the risk register. The user accepted all these suggestions, and only minor improvements were made later (*RQ3*).

Cases #1, #2, and #4 demonstrate how the risk register utilized data from a vast amount of information. They all used the RAG search to stay up to date with the latest knowledge and compare risks with the risk register. Before the AI assistant's usage, many risks had inadequate descriptions that provided no insight into the risk. Adding new descriptions was easy, as no risks required exact evidence. The risks mainly required examples and a more comprehensive explanation of the risk definition. The quality of the AI-suggested descriptions was generally very high. However, some citations were inaccurate (*RQ3* and *RQ4*). For example, an AI assistant suggested adding this to the description of *'customer retention and onboarding challenges'* risk:

> *"The JRC Cross-border and Emerging Risks in Europe report highlights the importance of maintaining strong customer relationships and delivering consistent value to ensure long-term retention."*

However, the *'JRC Cross-border and Emerging Risks in Europe'* report does not discuss customer relationships. This example highlights a limitation that the assistant can produce authoritative-sounding citations that are unrelated (a form of hallucination). While these were caught during review (participants and the author noticed them), they underscore the need for user verification of AI outputs. Case #3 showed that the assistant correctly listed all risk names but miscounted the number of risks. Both these cases address problems related to reliability (*RQ4*).

The *'remove risks'* tool requires the AI assistant to provide a comprehensive reason for the risk removal. Cases #5 and #6 showed promising results in that regard. In case #5, the most common reasons were that the duplicate risks and low likelihood and impact were given in exact values. The AI assistant was also able to provide a reason that Inclus has acquired international customers, based on the data in the risk register, and gave one reason based on that:

```
Reason: may no longer be relevant if international registration
concerns are addressed.
```

In case #6, the AI assistant provided a *'combination'* as the reason for removal. Notably, in cases #5 and #6, the AI had to justify each suggested removal. This requirement serves as a safeguard, making the AI's thought process visible. It allows the human user to judge whether the removal is warranted, thereby enhancing the reliability of the overall process (*RQ4*).

7.6 Results summary

This chapter summarizes the findings on the benefits and concerns of AI assistants in supporting risk registers. In this case study, the AI assistant made a significant contribution to improving the coherence and usability of the risk register. Participants responded very positively, though these outcomes may reflect the specific context of Inclus.

Table 12 lists the identified AI assistant's beneficial features as well as its inherent limitations. Features that yielded the best results included risk combination, risk register harmonization, adding new risks, and removing existing risks. These benefits counter all the risk registers maintenance challenges presented in Chapter 2.4. Based on the results, the AI assistant accelerates the risk register management process and facilitates the processing of complex data. Risk combination, risk register harmonization, and adding new risks also enhance the quality of the register, while removing redundant risks accelerates the risk assessment process.

Chapter 2.6 introduces the knowledge-related problems: lack of knowledge (surprises) and weak knowledge. This case study demonstrates that the use of an AI assistant can certainly improve upon both of these aspects, but it still has certain limitations. An AI assistant can provide insight on almost any topic, but as noted by the interviewees, AI rarely outperforms top experts in their fields.

Table 12: Identified benefits and limitations.

Action	Source of evidence	Relevant RQs	Effect
Merge duplicate risks	Risk-register entries	RQ3	Major benefit
	Questionnaire	RQ3	
	Interviews	RQ1, RQ2, RQ3	
	AI assistant interactions	RQ1, RQ2, RQ3	
Improve and harmonise risk descriptions / names	Risk-register entries	RQ3	Major benefit
	Questionnaire	RQ3	
	Interviews	RQ1, RQ3	
	AI assistant interactions	RQ1, RQ2, RQ3	
Add new relevant risks or scenarios	Risk-register entries	RQ3	Major benefit
	Questionnaire	RQ3	
	Interviews	RQ1, RQ2, RQ3	
	AI assistant interactions	RQ1, RQ2, RQ3	
Remove redundant risks	Risk-register entries	RQ2	Major benefit
	Questionnaire	RQ1	
	Interviews	RQ1, RQ2	
	AI assistant interactions	RQ1, RQ2, RQ4	
Work on foreign languages	Interviews	RQ1	Minor benefit
Keep risk register up-to-date	Questionnaire	RQ3	Minor benefit
Domain-specific analyses	Interviews	RQ3	Limitation or benefit
	AI assistant interactions	RQ3	
Hallucinations	Questionnaire	RQ4	Limitation
	Interviews	RQ4	
	AI assistant interactions	RQ4	
Biases	Interviews	RQ4	Limitation
Calculations	AI assistant interactions	RQ4	Limitation
Citations	Risk-register entries	RQ4	Limitation
	Questionnaire	RQ4	
	Interviews	RQ4	
	AI assistant interactions	RQ4	

8 Discussion

Given the improvements and pitfalls observed in our case study, an important question arises: How can an AI assistant be used reliably in risk management? This chapter discusses how an AI assistant can be used reliably and the role of context in LLM usage. The chapter also examines the inherent limitations of risk science, LLMs in general, and limitations in case study settings. The last subchapter discusses future directions for AI assistant development.

8.1 Process reliability

The biggest challenge of risk management is that its applications vary significantly depending on the intended use. Every project and organization has its objectives. For some risks, the evidence plays a pivotal role, and they do not tolerate any hallucinations. For others, more creative risk and scenario generation is permitted, where hallucinations are less significant. When it comes to managing political or conflict risk, biases can be a significant factor. One interviewee noted that if an AI assistant outputs, for example, racial prejudices, it can harm the credibility of the entire risk management process.

However, even with comprehensive context, the model may hallucinate. LLMs are not search engines but reasoning models. Even one hallucination in the wrong place can have serious consequences. Hallucination is a problem that may never disappear with the current LLM operating logic (Banerjee et al., 2024), and therefore, the process must be made reliable. A reliable risk management process means that the user is aware of the sources and evidence, knows if the output can be inaccurate, and can make an informed decision based on that. In the most sensitive cases, the AI assistant might need to be limited to only suggesting suitable document references without providing any textual content.

Risk management aims to inform decision-makers of uncertainties in their context. The purpose of the risk management process is to involve participants in taking action to mitigate negative risks. Therefore, discussion on the risks is essential in making participants aware of the risks and committing to the actions. This thesis argues that participants should understand the risks, and therefore, the AI assistant should not function fully autonomously. The users should always approve actions to maintain understanding and accountability.

Case study interviewee #1 emphasized that an AI assistant should always have the organization's objective at its core. The context in which the AI assistant is utilized plays a pivotal role. The more accurate the context provided to the AI assistant about the organization and its objectives, the better the AI performs. Context data may play an even more critical role than the AI assistant's reasoning capabilities. RAG is one solution for better context, but an AI assistant can also utilize different fine-tuning methods. RAG only pulls in context when prompted appropriately, and if the prompt does not retrieve the correct info, the model will not use that context. Another way to improve performance is to utilize domain experts as AI assistant users who understand the context and objectives and can effectively interpret the output

quality. Expert users can ask the right questions and make the correct decisions based on the AI's suggestions. As described in Chapter 3.6 and noted by the interviewees, even a well-informed AI assistant would not be capable of attaining the experience and tacit knowledge of an expert user. Much like chess, the strongest teams consist of a grandmaster and a computer. In risk management, the optimal results may come from an expert working with an AI assistant.

8.2 Validity and Limitations

The case study was constructed from multiple sources of evidence to ensure reliability, and the key informants reviewed the case study report to ensure their input was correctly interpreted (Yin, 2018, pp. 126-130). Triangulation of data sources ensures internal validity by cross-checking the data between different sources and methods (i.e., risk register entries, questionnaire, interviews, and AI interaction observations) (Patton, 2014, pp. 956-978). Analytic generalizations and theoretical background ensure the external validity (Yin, 2018, pp. 34-42).

However, the case study has several notable limitations related to the nature of risk science, the stochastic and unpredictable behavior of LLMs, and the specific case study setting. These circumstances make the case highly specific, affecting the validity and repeatability of the study. Similar studies can reach the same analytical generalizations. For example, any participatory risk process can benefit from AI-driven risk consolidation, regardless of the specific risks involved.

Limitations of risk science The case study presents two fundamental issues related to the nature of risk science and the risks it addresses. The first issue is that the case is unique, so no other organization has the same environment and risks. The second issue is that participatory risk management is often done based on subjective estimates, as sufficient data is not always available. While the results might be reproducible by the same team, a different team with different background knowledge would most likely end up with slightly different results (Aven & Heide, 2009). The thesis is not about estimating risks, but rather about how the AI assistant can help manage the risk register. Still, it is noteworthy that different risks and experts can significantly affect the results.

LLM repeatability issues LLMs pose problems for scientific research and traceability in general (Semmelrock et al., 2025). LLMs are updated regularly with new training data, meaning an LLM may have different information about an issue at varying times. Another problem is that LLMs are black-box models, meaning that interpreting their internal logic is an exceedingly complex task. While LLM can be made deterministic, that would not ensure repeatability, as slight variations in the user input may lead to different results. Additionally, the case study relies on the RAG knowledge base, including Inclus' internal reports, which makes repeatability even more difficult.

Case study limitations The case study itself poses several limitations. The size of the case study organization (17 participants) limited any significant statistical analysis. Another limitation is that the author works at the organization studied in the case. While this setup provides an opportunity to gain a detailed understanding of the organization, it also introduces the potential for biases. The author of this thesis served as the primary user of the AI assistant. However, the risk register includes risks from various categories in which the author is not an expert. Users' lack of expertise in some domains can impact the quality of AI assistant interactions. However, throughout the process, domain experts (who were not the AI assistant's developers) validated the AI's suggestions and had the final say on changes, which helped reduce the author's potential bias in judging the outcomes. The case study organization's senior managers reviewed the interim findings and the final case report to ensure they aligned with their experiences, thereby adding credibility.

8.3 Future work

While the case study aimed to identify generalizations, it did not consider all purposes of AI assistant use for risk management. With domain experts serving as AI assistant users who understand the company's exact objectives, the results would likely be of even higher quality. Most RAG files given in the Appendix B lacked the specific context of Inclus. With better-specified objectives, users can prepare more accurate context data for the AI assistant. The Inclus risk register had quite a high tolerance toward hallucinations and biases. It would be valuable to see how AI assistants perform with more sensitive requirements.

The core findings of this study are expected to apply to other organizations' risk register management tasks. Although organizations have different objectives, a similar AI assistant is expected to provide value in tasks such as combining risks, suggesting new risks, enhancing descriptions, and maintaining the coherence of the risk register. Findings can also be generalized for broader use. An AI assistant can generate scenarios and mitigation tasks in the same way it suggests new risks, and it cab help maintain consistency among them.

9 Conclusions

This thesis examines how an AI assistant can help manage a complex risk register more effectively in participatory risk management settings. Risk registers can become challenging to manage when multiple participants work on them. Some common problems include missing critical risks, inadequately defined risks, redundant risks, or duplicate risks (Leva et al., 2017; Bjørnsen & Aven, 2019; Aven, 2013, 2015).

Recent evolution in natural language processing has shown promising results in supporting risk management (Collier et al., 2024; Esposito et al., 2024; Stødle et al., 2024). AI assistants can identify risks outside the expert's domain and help serve as an external memory. The most significant benefit AI assistants bring is their ability to harmonize risk registers, meaning they can help combine risks, improve risk descriptions, and determine the appropriate level of generalization. The study demonstrated that AI assistants can perform several complex tasks (e.g., merging overlapping risks, synthesizing descriptions) in a manner comparable to human reasoning. In processing large volumes of complex textual risk data, an AI assistant can extend the experts' cognitive capacity and perform work that would take experts days to complete in minutes.

AI assistants are still prone to hallucinations, which can have serious consequences. The AI assistant should be deployed as a supportive tool rather than an autonomous agent. The oversight of human experts is crucial for verifying suggestions and maintaining credibility. To tackle AI assistant-related problems, the process should be reliable, allowing users to understand the AI's steps and limitations in specific phases of the process. An AI assistant can reduce the cognitive load from experts and allow for more qualitative data and discussion to be considered a substantial part of the risk management process. Besides, more important than an accurate risk estimate is that participants are involved in the process and take action on the planned mitigation.

The case study provided comprehensive results for each of the research questions *RQ1-RQ4* (see Table 4). *RQ1* asked if an AI assistant speeds up risk management tasks. The study demonstrated that an AI assistant accelerated tasks such as adding new risks, combining duplicate risks, removing redundant risks, and refining risk descriptions. *RQ2* concerned handling large and complex data. The AI assistant helped harmonize a complex risk register and integrate information from multiple sources by leveraging a RAG knowledge base (external reports and internal documents). *RQ3* was about the improvements in quality. According to participants and based on observations of the risk register, the use of an AI assistant improved the quality of the risk register. *RQ4* addressed reliability and threats. Several hallucinations were discovered. However, with human oversight and a reliable process, these would not pose a threat to reliability.

Based on the design proposed in this thesis and the results of the case study, an AI assistant appears to offer significant benefits to participatory risk register management. In a carefully designed software application context, most flaws and issues of LLMs can be mitigated, providing a substantial improvement to an organization's ability to manage risk.

References

Anil, R., Borgeaud, S., Wu, Y., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., Millican, K., et al. (2023). Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*. DOI: https://doi.org/10.48550/arXiv.2312.11805.

Anthropic (2025). Introducing Citations on the Anthropic API. https://www.anthropic.com/news/introducing-citations-api. Accessed: 2025-03-28.

Apostolakis, G. E. (2004). How useful is quantitative risk assessment? *Risk Analysis*, 24(3), 515–520. DOI: https://doi.org/10.1111/j.0272-4332.2004.00455.x.

Aven, T. (2012). The risk concept—historical and recent development trends. *Reliability Engineering & System Safety*, 99, 33–44. DOI: https://doi.org/10.1016/j.ress.2011.11.006.

Aven, T. (2013). Practical implications of the new risk perspectives. *Reliability Engineering & System Safety*, 115, 136–145. DOI: https://doi.org/10.1016/j.ress.2013.02.020.

Aven, T. (2015). Implications of black swans to the foundations and practice of risk assessment and management. *Reliability Engineering & System Safety*, 134, 83–91. DOI: https://doi.org/10.1016/j.ress.2014.10.004.

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, 253(1), 1–13. DOI: https://doi.org/10.1016/j.ejor.2015.12.023.

Aven, T. & Heide, B. (2009). Reliability and validity of risk analysis. *Reliability Engineering & System Safety*, 94(11), 1862–1868. DOI: https://doi.org/10.1016/j.ress.2009.06.003.

Aven, T. & Krohn, B. S. (2014). A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliability Engineering & System Safety*, 121, 1–10. DOI: https://doi.org/10.1016/j.ress.2013.07.005.

Banerjee, S., Agarwal, A., & Singla, S. (2024). LLMs Will Always Hallucinate, and We Need to Live With This. *arXiv preprint arXiv:2409.05746*. DOI: https://doi.org/10.48550/arXiv.2409.0574.

Bjørnsen, K. & Aven, T. (2019). Risk aggregation: What does it really mean? *Reliability Engineering & System Safety*, 191, 106524. DOI: https://doi.org/10.1016/j.ress.2019.106524.

Collier, Z. A., Gruss, R. J., & Abrahams, A. S. (2024). How good are large language models at product risk assessment? *Risk Analysis*, 45(4). DOI: https://doi.org/10.1111/risa.14351.

de Moivre, A. (1718). *The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play*. London: W. Pearson. First edition.

Edwards, B. (2025). Why extracting data from PDFs is still a nightmare for data experts. https://arstechnica.com/ai/2025/03/why-extracting-data-from-pdfs-is-still-a-nightmare-for-data-experts/. Accessed on 2025-04-23.

Esposito, M., Palagiano, F., Lenarduzzi, V., & Taibi, D. (2024). Beyond words: On large language models actionability in mission-critical risk analysis. *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, (pp. 517–527). DOI: https://doi.org/10.1145/3674805.3695401.

Flage, R. & Aven, T. (2009). Expressing and communicating uncertainty in relation to quantitative risk analysis. *Reliability: Theory & Applications*, 4(2), 9–18.

French, S. (2011). Aggregating expert judgement. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas*, 105, 181–206. DOI: https://doi.org/10.1007/s13398-011-0018-6.

French, S. & Argyris, N. (2018). Decision analysis and political processes. *Decision Analysis*, 15(4), 208–222. DOI: https://doi.org/10.1287/deca.2018.0374.

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., et al. (2020). The pile: An 800GB dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*. DOI: https://doi.org/10.48550/arXiv.2101.00027.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2. DOI: https://doi.org/10.48550/arXiv.2312.10997.

Geva, M., Schuster, R., Berant, J., & Levy, O. (2020). Transformer feed-forward layers are key-value memories. *arXiv preprint arXiv:2012.14913*. DOI: https://doi.org/10.48550/arXiv.2012.14913.

Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al. (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*. DOI: https://doi.org/10.48550/arXiv.2501.12948.

Hagos, D. H., Battle, R., & Rawat, D. B. (2024). Recent advances in generative AI and large language models: Current status, challenges, and perspectives. *IEEE Transactions on Artificial Intelligence*. DOI: https://doi.org/10.1109/TAI.2024.3444742.

Hansson, S. O. & Aven, T. (2014). Is risk analysis scientific? *Risk Analysis*, 34(7), 1173–1183. DOI: https://doi.org/10.1111/risa.12230.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, (pp. 75–105). DOI: https://doi.org/10.2307/25148625.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., et al. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2), 1–55. DOI: https://doi.org/10.48550/arXiv.2311.05232.

Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A., Welihinda, A., Hayes, A., Radford, A., et al. (2024). GPT-4o system card. *arXiv preprint arXiv:2410.21276*. DOI: https://doi.org/10.48550/arXiv.2410.21276.

Inclus (2025). Inclus: Agile and Collaborative Risk Management Software. https://inclus.com/. Accessed: 2025-04-11.

ISO 31000 (2018). ISO 31000:2018 - Risk management – Guidelines.

Kaplan, S. & Garrick, B. J. (1981). On the quantitative definition of risk. *Risk Analysis*, 1(1), 11–27. DOI: https://doi.org/10.1111/j.1539-6924.1981.tb01350.x.

Leva, M. C., Balfe, N., McAleer, B., & Rocke, M. (2017). Risk registers: Structuring data collection to develop risk intelligence. *Safety Science*, 100, 143–156. DOI: https://doi.org/10.1016/j.ssci.2017.05.009.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474. DOI: https://doi.org/10.48550/arXiv.2005.11401.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey, 2024. *arXiv preprint arXiv:2402.06196*. DOI: https://doi.org/10.48550/arXiv.2402.06196.

OpenAI (2022). Introducing ChatGPT. https://openai.com/blog/chatgpt/. Accessed: 2025-02-18.

OpenAI (2022). text-embedding-ada-002. https://platform.openai.com/docs/models/text-embedding-ada-002. Accessed: 2025-05-30.

Owen, D. (2015). Collaborative decision making. *Decision Analysis*, 12(1), 29–45. DOI: https://doi.org/10.1287/deca.2014.0307.

Patterson, F. D. & Neailey, K. (2002). A risk register database system to aid the management of project risk. *International Journal of Project Management*, 20(5), 365–374. DOI: https://doi.org/10.1016/S0263-7863(01)00040-0.

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. SAGE Publications. ISBN:9781412972123.

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. DOI: https://doi.org/10.2753/MIS0742-1222240302.

Sadeghi, M. & Blachez, I. (2025). The Infection of Western AI Chatbots by a Russian Propaganda Network. https://www.newsguardtech.com/wp-content/uploads/2025/03/March2025PravdaAIMisinformationMonitor.pdf. Accessed: 2025-04-23.

Scolobig, A. (2025). Participatory Processes for Industrial Risk Management: Enablers, Barriers and Limitations. *Public Participation in Governance of Industrial Safety Risks: An Uneasy Journey*, (pp. 73–82).

Semmelrock, H., Ross-Hellauer, T., Kopeinik, S., Theiler, D., Haberl, A., Thalmann, S., & Kowald, D. (2025). Reproducibility in machine-learning-based research: Overview, barriers, and drivers. *AI Magazine*, 46(2), e70002. DOI: https://doi.org/10.48550/arXiv.2406.14325.

Singh, C., Inala, J. P., Galley, M., Caruana, R., & Gao, J. (2024). Rethinking interpretability in the era of large language models. *arXiv preprint arXiv:2402.01761*. DOI: https://doi.org/10.48550/arXiv.2402.01761.

Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: SAGE Publications. ISBN:978-0803975031.

Stødle, K., Flage, R., Guikema, S., & Aven, T. (2024). Artificial intelligence for risk analysis—A risk characterization perspective on advances, opportunities, and limitations. *Risk Analysis*, 45(4). DOI: https://doi.org/10.1111/risa.14307.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. New York: Random House. ISBN:978-1400063512.

Tennenholtz, G., Chow, Y., Hsu, C.-W., Jeong, J., Shani, L., Tulepbergenov, A., Ramachandran, D., Mladenov, M., & Boutilier, C. (2023). Demystifying embedding spaces using large language models. *arXiv preprint arXiv:2310.04475*. DOI: https://doi.org/10.48550/arXiv.2310.04475.

U.S. Department of Defense (2002). DoD News Briefing – Secretary Rumsfeld and Gen. Myers. https://web.archive.org/web/20160406235718/http://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636. Accessed: 2025-04-23.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, , & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. DOI: https://doi.org/10.48550/arXiv.1706.03762.

Wang, Z., Chu, Z., Doan, T. V., Ni, S., Yang, M., & Zhang, W. (2024). History, development, and principles of large language models: an introductory survey. *AI and Ethics*, (pp. 1–17). DOI: https://doi.org/10.48550/arXiv.2402.06853.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824–24837. DOI: https://doi.org/10.48550/arXiv.2201.11903.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). React: Synergizing reasoning and acting in language models. *International Conference on Learning Representations*. DOI: https://doi.org/10.48550/arXiv.2210.03629.

Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods*. Thousand Oaks, CA: SAGE Publications, 6th edition. ISBN:978-1-5443-6910-6.

A Survey questions

Table A1: Survey Questions.

Question	Criterion
How clearly are the current risks described in the risk register?	1-5
Does the risk register include all relevant risks, or do you identify any gaps?	1-5
Is there a right amount of risks (not too many, not too few)?	1-5
Do you feel that the risk register also accounts for unforeseen risks (e.g., "black swans")?	1-5
How easy is it to interpret the level of uncertainty associated with risks in the risk register (i.e., does it show how well we know the risk)?	1-5
Does the risk register accurately capture and reflect disagreements in risk assessments?	1-5
Do you believe that risk assessment is based on sufficient and high-quality information?	1-5
Do you feel that the risk register is up to date?	1-5
How reliable do you consider the risk assessment methods used in the risk register?	1-5
Could this risk management process be somehow improved? Are there challenges or uncertainties in the process you would want to address?	Open question

B RAG knowledge base

Table B1: Documents used in the RAG knowledge base.

File name	Description*	Pages
Allianz Risk Barometer 2025.pdf	This annual report by Allianz outlines the top global business risks for 2025 based on a survey of nearly 3,800 experts from over 100 countries. The top concerns include cyber incidents, business interruption, natural catastrophes, and climate change. The report provides insights into regional and industry-specific risk perceptions and discusses emerging threats such as AI-related technologies and political violence.	47

AXA Investment Managers Outlook 2025.pdf	This document offers AXA IM's macroeconomic and investment outlook for 2025-2026. It analyzes potential impacts of geopolitical events, particularly a second term for Donald Trump, on global economies and financial markets. Topics include regional growth forecasts, inflation trends, monetary policy divergences, and investment strategies focusing on equities, fixed income, and sector-specific risks.	26
CMI insight DP 2023.pdf	Produced by the CMI – Martti Ahtisaari Peace Foundation, this report describes a digitally enhanced foresight methodology used in complex peace processes. It emphasizes participatory conflict analysis and scenario-building with tools like the Inclus platform, aiming to facilitate inclusive dialogue among diverse stakeholders.	2
Code of Conduct of Inclus Oy.pdf	This document sets out the ethical guidelines and behavioral expectations at Inclus Oy.	2
European data protection supervisor – explainable artificial intelligence.pdf	A policy-oriented publication by the EDPS, this document examines the need for explainable AI (XAI) in light of the "black box" problem. It explores legal, ethical, and technical challenges of AI opacity and proposes strategies for increasing AI transparency, interpretability, and accountability in sensitive applications like healthcare and finance.	23
FINALPravda Report.pdf	An investigative report detailing how a Russian disinformation network named "Pravda" attempts to manipulate Western AI chatbot outputs by flooding the web with pro-Kremlin propaganda. The report explains how this tactic—termed "LLM grooming"—is designed to influence how AI models generate responses by polluting their training and retrieval data.	21
Global Trends 2040.pdf	A U.S. National Intelligence Council report exploring long-term global trends across demographics, technology, economics, and governance. It outlines five scenarios of how the world could evolve by 2040, shaped by climate change, geopolitical tensions, and shifting power dynamics.	156

Global Value Chain Dependencies.pdf	An analytical paper examining vulnerabilities in global value chains (GVCs), especially in light of COVID-19 and the war in Ukraine. It maps foreign input and market dependencies, highlighting the key role of China as both a supplier and buyer across industries.	41
Inclus culture handbook.pdf	Outlines Inclus Oy's organizational values: transparency, courage, growth, and community. Emphasizes a collaborative, open, and mission-driven work culture.	8
Inclus IN-2 Employee Security Guidelines.pdf	Practical cybersecurity requirements and recommendations for employees and subcontractors, including password hygiene, MFA, device security, and travel precautions.	9
Inclus Ltd Anti-Bribery Policy.pdf	States Inclus' zero-tolerance policy on bribery and corruption. Defines responsibilities and procedures for reporting unethical behavior.	2
Inclus SP-1 Information Security Policy.pdf	Defines Inclus' framework for managing information security, ensuring confidentiality, integrity, and availability, aligned with standards like ISO 27001.	7
Inclus SP-2 Acceptable Use Policy.pdf	Establishes rules for appropriate use of company IT assets, including software, internet access, mobile devices, and remote work security.	9
Inclus SP-8 Access Control Policy.pdf	Details access management practices, including user provisioning, authentication, and periodic reviews to ensure least-privilege access.	5
Inclus Technology and Security Whitepaper.pdf	Describes the technical and security architecture of Inclus' SaaS solution, covering infrastructure, encryption, authentication, APIs, and compliance best practices.	16
Industry Outlook 2025 .pdf	A forward-looking report analyzing expected trends and transformations across various industries for 2025, including technological innovation, regulatory changes, supply chain shifts, and sustainability efforts.	64
Inclus AWSInfrastructure.pdf	Describes the architecture and design principles of Inclus Oy's AWS-based cloud infrastructure, emphasizing security, segregation of environments, and use of Infrastructure as Code.	10

JRC Cross-border and emerging risks in Europe.pdf	Published by the Joint Research Centre (JRC), this report identifies and examines systemic cross-border and emerging risks in Europe, such as climate change, cybersecurity threats, and pandemics. It provides risk scenarios and policy recommendations.	196
NATO Science & Technology Trends 2020-2040.pdf	Assesses the impact of emerging and disruptive technologies (EDTs) on NATO's future capabilities. Key areas include AI, quantum technologies, space, biotechnology, and hypersonics. It aims to guide NATO in shaping its defense strategies.	160
NATO Science for Peace and Security-Emerging Threats of Synthetic Biology and Biotechnology.pdf	A compilation of expert perspectives on the biosecurity, governance, and dual-use challenges of synthetic biology and biotechnology, emphasizing the need for both top-down regulation and grassroots initiatives.	233
Proof of Inclus Oy's Compliance with Ethical Standards and Data Protection.pdf	Proof of Inclus Oy's Compliance with Ethical Standards and Data Protection Outlines Inclus Oy's commitment to GDPR, ethical conduct, and data protection measures including encryption, access control, and breach response procedures.	2
Quantum Technologies - A Review of the Patent Landscape.pdf	A comprehensive review analyzing nearly 49,000 patents related to quantum technologies. Covers areas like quantum computing, cryptography, sensing, and nanotechnology, highlighting innovation hotspots and key players.	25
RAND Emerging Technology and Risk Analysis .pdf	Analyzes how emerging technologies (e.g., AI, biotech, cyber, hypersonics) intersect with national security, assessing their potential risks and implications for U.S. and global policy.	22
RAND Strategic competition in the age of AI .pdf	Examines how AI influences strategic competition, particularly between the U.S. and China. Discusses AI's impact on military power, economic strength, and information dominance.	144

sigma 5 2024 Global economic and insurance market outlook 2025-26.pdf	Offers macroeconomic forecasts and insurance market projections. Focuses on global GDP trends, inflation, interest rates, and reinsurance profitability.	39
Swiss Re SONAR New Emerging Risk Insights.pdf	Identifies 13 emerging risk themes and 3 trend spotlights affecting insurance markets. Highlights risks like AI in insurance, deep-sea mining, and cyber-enabled fraud.	52
Tech Forecast 2025.pdf	A skills and tools forecast for IT professionals. Covers AI, cybersecurity, and cloud trends, with emphasis on in-demand skills like AI agents, LangChain, Kubernetes, and cloud infrastructure.	33
Top Strategic Technology Trends 2025.pdf	Outlines 10 key tech trends, including Agentic AI, quantum cryptography, spatial computing, and hybrid computing. Designed for CIOs and tech leaders to guide long-term innovation.	28
Top Tech Trends 2025 Report.pdf	Focuses on "AI-powered everything." Top trends include Gen AI in cybersecurity, supply chains, robotics, and the nuclear energy resurgence. Based on executive and investor surveys.	104
World Economic Forum Global Cybersecurity Outlook 2025 .pdf	Addresses the rising complexity of the cyber threat landscape, increasing AI-related vulnerabilities, supply chain risks, and cyber inequity between large and small organizations.	49
World Economic Forum The Global Risks Report 2025.pdf	A flagship report detailing global risks over three timeframes (2025, 2027, 2035). Key risks include misinformation, geopolitical conflict, climate change, and loss of trust in institutions.	104

* Descriptions are generated with GPT-4o.