International Journal of Information Systems and Project Management (IJISPM)
2025, 13(4), e130401, DOI: 10.12821/ijispm130401

RESEARCH ARTICLE

Ethical considerations in the AI lifecycle for design, developing and adopting AI in public sector – the case of Finland

Ari Alamäki
Haaga-Helia University of Applied Sciences, Ratapihantie 13, 00520 Helsinki, Finland, ari.alamaki@haaga-helia.fi

Umair Ali Khan
Haaga-Helia University of Applied Sciences, Ratapihantie 13, 00520 Helsinki, Finland, umairali.khan@haaga-helia.fi

Altti Lagstedt
Haaga-Helia University of Applied Sciences, Ratapihantie 13, 00520 Helsinki, Finland, altti.lagstedt@haaga-helia.fi

Abstract
This study explores the role of ethics in all phases of AI projects. Through qualitative interviews with the public sector actors in Finland, the study identifies key ethical concerns related to the lifecycle of AI. The findings highlight the need for embedding ethical requirements throughout the AI system lifecycle and emphasize the role of human-centered AI systems. By utilizing empirical data from multiple public sector case organizations, this study provided both theoretical insights and practical guidelines for developing ethically aligned AI systems. The findings emphasize the need for a comprehensive, lifecycle-oriented approach to ethical AI design, development, adoption, and use. The AI lifecycle spans various phases that collectively shape the ethical impact of AI applications. This research provides empirical insights into how ethical considerations can be practically integrated from the design to adoption phases of AI. By embedding ethical practices throughout the lifecycle, organizations can anticipate and mitigate risks more effectively.

Keywords
artificial intelligence; software development; ethics of AI; AI lifecycle; project management.

Received: 25 November 2024 | Accepted: 2 April 2025

1. Introduction

The public sector is in the early stages of adopting artificial intelligence (AI) (Selten & Klievink, 2024; Straub et al., 2023), although it will be an important technology for public management processes (Misra et al., 2023; Neumann et al., 2024). AI remarkably enhances the data collection and analysis possibilities, enabling automatic learning, decision making, and prediction. AI is an essential part of the Industry 4.0 revolution and the change of work (Frey et al., 2016; Jan et al., 2023; Leesakul et al., 2022) which also applies to public sector actors. Since AI will have a significant impact on the development of work and organizations, the related ethical discussion is of paramount importance (Ashok et al., 2022a; Huang et al., 2023). It is typical for AI solutions that they are updated with new teaching data even during use. Consequently, their operation and applicability may change over time. In addition, the ethical questions related to decommissioning the AI solution should also be considered. Therefore, ethical considerations should not be seen as just one step in the planning phase of AI solutions, but the entire lifecycle of an AI solution, including operations, maintenance, and retirement, should be taken into account (Huang et al., 2023).

The use of AI in organizations can be roughly divided into two different use case categories: i) the organization's own data is used, and their own AI solution is built, or ii) pre-trained generative AI is used. Especially the rapidly evolving field of generative AI is bringing forth new ethical and trustworthiness challenges. Generative AI holds significant potential to revolutionize the public sector, yet it simultaneously poses distinctive challenges. Especially, the increasing integration of large language models (LLMs) into critical areas such as education, healthcare, and business necessitates the development of robust methods to verify the authenticity, ownership, accuracy, and ethical use of the content they generate. Moreover, issues such as the lack of explainability, transparency, and accountability are becoming increasingly prominent in generative AI solutions. Recent evaluations (Bommasani et al., 2023) indicate that major LLMs, including both proprietary and open source, do not adequately meet the requirements of the EU AI Act (European Parliament, 2023). Although the providers of these LLMs excel in enhancing efficiency through the integration of extensive training data and adding multimodal capabilities, the growing focus on transparency and the categorization of AI applications' risks, as outlined by the EU AI Act, is crucial. This emphasis is essential for ensuring safety and upholding the fundamental rights of individuals and businesses (European Commission, 2024). Thus, it is crucial to understand the ethical implications of the integration of AI applications built on the LLMs into the public sector. Additionally, public organizations face many other challenges regarding AI, such as perceived financial costs, organizational innovativeness, governmental pressure, government incentives, data privacy concerns, lack of skilled personnel, and resistance to change (Misra et al., 2023; Neumann et al., 2024; Sun & Medaglia, 2019). Thus, more research is still needed in this field (Andrews, 2019).

The ethical discussion is often overshadowed by the technological and economic exploitation of AI's potential. Therefore, it is crucial to emphasize ethical questions more prominently in the discourse. There has already been some discussion about taking ethics into account during the AI life cycle (see e.g. Huang et al., 2023). Therefore, there has been relatively little thought about, for example, the use of AI in the public sector. There is clearly a need for more versatile studies on the matter. To address this, we formulate the following research questions:

1. What ethical issues are emphasized as a part of AI system design, development, and adoption in case organizations?
2. How should ethical consideration be integrated into AI system life cycle phases in the public sector's organizations?

To answer these research questions, we first analyzed the existing ethical frameworks at different stages of the AI lifecycle. Subsequently, we analyzed the ethical AI issues and AI frameworks at the AI system lifecycle level. In addition, we also analyzed this issue from the lens of information system (IS) development and suggested the main goals that should be considered in the AI development process. We found that most relevant studies assessing ethical considerations in the AI system lifecycle are conceptual and lack empirical evidence. Our case study, including interviews with the AI adopters and potential AI adopters in the public sector and the subsequent qualitative analysis of the interview data, highlighted the practical ethical challenges and concerns.

The rest of the paper is structured as follows. Section 2 presents the related work, providing an overview of existing ethical frameworks and studies on AI development, especially in the public sector. Section 3 outlines the research methods, detailing the qualitative case study approach, participant selection, data collection, and analysis techniques. Section 4 presents the findings from the interviews, categorized into key AI application areas, ethical challenges, and design considerations in public sector AI adoption. Section 5 offers a discussion of the results, connecting them with the existing literature and ethical AI frameworks. Section 6 highlights the limitations of the study and outlines a future research agenda for exploring ethical AI development further. Section 7 concludes the paper.

2. Background

Ethical considerations are increasingly crucial in the development and deployment of AI systems within their rapidly changing landscape. While AI presents substantial socio-economic advantages, it is important to acknowledge that the same mechanisms driving these benefits may also pose significant risks. For example, AI has the potential to cause both tangible harms, such as jeopardizing individual safety and health, including potential loss of life and property damage, and intangible harms, including breaches of privacy, constraints on freedom of expression, threats to human dignity, and discrimination (Nikolinakos, 2023). As AI applications continue to expand, establishing guidelines for ethical and responsible behavior in AI system development is becoming critically important. In recent years, several AI frameworks have been proposed, such as the EU's ethical guidelines for trustworthy AI (High-Level Expert Group on AI (AIHLEG), 2018) focusing on ensuring that AI systems are transparent, fair, and uphold privacy and dignity. AI Act (European Parliament, 2023) categorizes AI systems by risk and sets compliance standards for high-risk applications. OECD AI Principles (Floridi & Cowls, 2019) focus on human-centric values, transparency, and accountability across sectors, among others.

Although these frameworks serve as reference guides for developing responsible AI systems, they are typically high-level and do not provide concrete guidance on how to develop such systems (Lu et al., 2022, 2023; Sanderson et al., 2022). It remains largely unclear how these principles and values be converted into requirements for responsible AI systems, and what developers and the organizations developing these systems should do (Vakkuri et al., 2021). Due to this limitation, the developers often lack explicit knowledge on how to apply these guidelines during the system development process (Ronanki, 2023).

To tackle this issue, some studies adopted an empirical approach involving interviews with scientists, engineers, and developers to understand the practitioners' views on AI ethics principles and their implementation. Prior research (Kamila & Jasrotia, 2023; Lu et al., 2022; Sanderson et al., 2022) find privacy and security, bias and fairness, trust and reliability, transparency, and human-AI interactions as major ethical concerns. Lu et al. (2022) find that ethical risk assessments in AI are often one-time actions, ethical requirements are vaguely defined, system-level design considerations are overlooked, and there is insufficient support for continuous ethical monitoring post-deployment. Sanderson et al. (2022) adopted a software engineering approach covering four aspects: i) high-level view highlighting the do-once-and-forget approach in practice, maintaining trust from data providers, attaching ethics credentials to AI components and products, significance of system-level approach to AI development, ii) ethics requirements highlighting privacy and security as the prerequisites, normative, descriptive, and temporal aspects of responsibility, iii) design and implementation emphasizing overriding AI decisions by human's, reliability vs. fairness tradeoff, preferred use of trained AI models and related components, and explainability of decisions, and iv) deployment and operation highlighting monitoring and validation to ensure the adherence of AI systems post-development, and tracking the use of AI systems.

Some studies employed desktop research to review the existing academic literature on ethical concerns, requirements, limitations, and risks in AI system development (Huriye, 2023; Kamila & Jasrotia, 2023). Huriye (2023) identified bias, privacy, accountability, and transparency as key ethical issues, emphasizing the need for stakeholder collaboration and a human-centered approach that values local community needs. Kamila and Jasrotia (2023) also highlighted similar concerns, including privacy and security, bias and fairness, trust and reliability, transparency, and human-AI interaction. Ashok et al. (2022b) identified 14 ethical implications across seven technology archetypes and emphasized key principles such as accountability, fairness, and privacy. Similarly, Huang et al. (2023) addressed ethical risks such as privacy leakage, discrimination, and security concerns arising from AI applications.

Prior studies addressed this issue by proposing comprehensive methods to develop ethical AI systems. For instance, Vakkuri et al. (2021) proposed a sprint-by-sprint process that takes on the form of a deck of 21 cards, covering 8 AI ethics themes, which also covered the AIHLEG's Ethics Guidelines (High-Level Expert Group on AI (AIHLEG), 2018) including (1) human agency and oversight; (2) technical robustness and safety; (3) privacy and data governance; (4) transparency; (5) diversity, non-discrimination and fairness; (6) environmental and societal well-being; and (7) accountability. Using this framework, one can select the cards that are relevant to their work and then evaluate the situation again after each sprint. This approach results in a paper trail of choices and trade-offs that document the ethical considerations conducted during development. However, this method does not suggest how the cards should be selected for a specific development phase.

2.1. Ethical Issues at the AI System Lifecycle Level:

Several studies emphasize the application of ethical principles to the entire life cycle of AI systems (Lo Piano, 2020; Taddeo et al., 2024). The OECD Framework for the Classification of AI Systems (OECD, 2022) also highlights the need for ethical considerations throughout the entire life cycle of AI systems. Similarly, the NIST AI Risk Management Framework (AI NIST, 2023), which closely follows the OECD's definition of the AI lifecycle, also emphasizes the importance of ethical considerations at each development stage.

The AI system lifecycle encompasses several stages that guide a project from concept to operation and monitoring. Ethical considerations must be addressed from the early phases (Lahiri & Saltz, 2024), as several critical decisions are required, such as data collection, risk evaluation, infrastructure selection (including hardware, software, and platforms like Microsoft Azure), and AI model selection. Understanding the type of information management department, we are dealing with is crucial. Additionally, we need to assess how agile the organization is and how straightforward it is to make these decisions.

The granularity of the AI lifecycle varies from study to study. From an ethical perspective, the AI system lifecycle is more complex than the traditional system lifecycle. An insufficiently differentiated lifecycle can lead to blind spots and the creation of ethical risks. Conversely, identifying too many stages (with related tasks) can make the iterative application of ethical principles cumbersome, rendering the ethical guidelines unwieldy (Taddeo et al., 2024).

Different studies outline varying phases of the AI system lifecycle. Some studies, such as the OECD digital economy papers (OECD, 2019) and the NIST AI Risk Management Framework (AI NIST, 2023), define six phases: plan and design, collect and process data, build and use model, verify and validate, deploy and use, and operate and monitor. Similarly, other studies describe three (De Silva & Alahakoon, 2022), four (Amugongo et al., 2023; Floridi et al., 2022), and five (Huang et al., 2023) phases of the AI system lifecycle. Based on various studies, we identify five basic stages in the AI system lifecycle: design, develop, deploy, operate, and retire which encompass all the actions across all phases defined by the existing studies.

Each phase includes three dimensions: actions, actors, and ethical/risk considerations. Additionally, each phase encompasses specific AI principles that define the ethical aspects within that phase as overarching principles (OECD, 2019). Therefore, the AI system lifecycle can be viewed as a five-dimensional model.

Figure 1 illustrates the AI system lifecycle and its five dimensions. Each phase involves specific actions that represent the activities within that phase. Based on these activities, each phase requires certain expertise or roles, making the AI system lifecycle a multidisciplinary approach that demands more competence than the traditional system lifecycle.

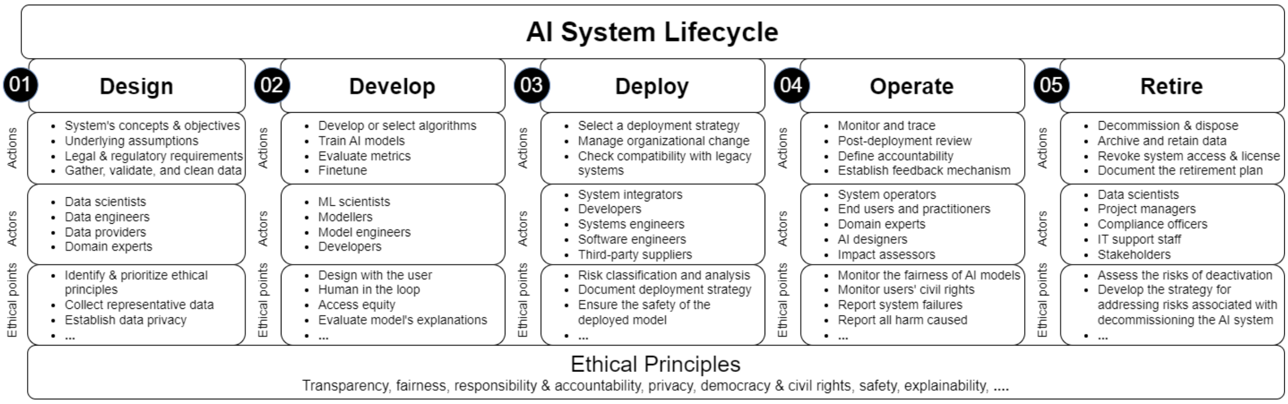

Fig. 1: Synthesis of existing studies: AI system lifecycle showing the dimensions of stages, actors, and the ethical/risk considerations

The design phase, in particular, involves a broader range of expertise, which signifies its importance. The data scientist responsible for the design phase typically holds a senior position with several years of experience. They must be able to formulate the problem and conceptualize a solution by drawing on existing literature and their past experiences with diverse AI projects. Additionally, they should be adept at identifying representative, required, and available data by collaborating with other phases of the lifecycle (De Silva & Alahakoon, 2022). Since the design phase sets the foundation for the entire AI project, ethical considerations at this stage are most important, as they have long-lasting impacts (Brey & Dainow, 2024).

Certain ethical principles govern ethical behavior within each phase of the AI system lifecycle. These principles, derived or inspired by those set by (High-Level Expert Group on AI (AIHLEG, 2019), are applied differently across various studies. For example, the OECD digital economy papers (OECD, 2019) outline five ethical principles: i) benefiting people and the planet, ii) human-centered values and fairness, iii) transparency and explainability, iv) robustness, security, and safety, and v) accountability. This study also explains how each AI principle can be assessed at specific phases of the AI lifecycle. Hence, for each phase, there exists a set of ethical considerations that correspond to specific AI principles.

Huang et al. (2023) assess ethical considerations in a five-phase AI system lifecycle using ethical principles such as transparency, fairness, responsibility and accountability, democracy and civil rights, sustainability, privacy, and safety. Similarly, Amugongo et al. (2023) focus on fairness, transparency, ethics by design, trust, precision, and safeguarding humanity in a four-phase lifecycle. Floridi et al. (2022) propose a five-phase model with principles like societal welfare, accountability, governance framework, responsibility, performance, environmental impact, fairness, accuracy, traceability, privacy, explainability, and robustness, providing guidelines for each phase. Taddeo et al. (2024) adopt Floridi's model, emphasizing transparency, responsibility, traceability, robustness, privacy, and accountability.

It is worth mentioning that privacy in the AI lifecycle is a two-fold approach that encompasses "Privacy by Design" and "Privacy by Default" to ensure both regulatory compliance and the protection of individual rights (Navaie, 2024). "Privacy by Design" involves integrating privacy measures into the design and architecture of information systems from the outset, rather than as an afterthought. This proactive approach ensures that privacy is an integral component of the system's functionality. "Privacy by Default" complements this by ensuring that personal data is automatically protected in any given system or business practice, meaning that the default settings are configured to the most privacy-friendly options. These concepts are emphasized in regulations such as the General Data Protection Regulation (GDPR), which mandates that data protection measures be implemented by design and by default. By incorporating these principles, organizations can ensure that personal data is handled with the highest standards of privacy protection throughout its lifecycle.

In summary, various studies have employed different levels of granularity in AI system lifecycle models and applied diverse sets of ethical principles across lifecycle phases. Although the number of ethical principles used in each study varies, they generally align with those established by (High-Level Expert Group on AI (AIHLEG), 2019) and its older version.

In our study, we primarily address ethical considerations in the AI lifecycle from a conceptual standpoint, by providing a grounded, empirical perspective. The case study in this paper captures and validates the views of various stakeholders directly involved in different stages of the AI lifecycle. This approach bridges the gap between theoretical ethical frameworks and practical, real-world challenges, aligning the insights from practitioners with ethical guidelines previously established in literature.

2.2. AI Ethics in Information System Development

The first three phases of the AI life cycle (Figure 1) are easily overlooked as they are not so visible to the end users (De Silva & Alahakoon, 2022). The developers (typically from an external company) have a key role in these phases. Commonly, development is outsourced, and it is tempting to think that the responsibility has been transferred to "someone else" (Hedlund & Persson, 2025). However, these three phases play a very significant role in the implementation of an ethically sustainable AI solution, and therefore, we will go through the first half of the AI life cycle in a little more detail next.

Typically, an information systems (IS) development methodology encompasses the first three phases of the previously introduced AI system lifecycle: design, development, and deployment (see Figure 1). There are several commonly known methods for developing information systems, all of which have their own strengths and weaknesses (Dahlberg & Lagstedt, 2018). On a general level, these methods can be divided into two groups: plan-driven and change-driven methods (Lagstedt, 2019; Moe et al., 2012).

The plan-driven methods (such as the waterfall method) are divided into successive steps, such as system requirements, software requirements, analysis, program design, coding, testing, operations (implementation) (Royce, 1970). The following step proceeds only when the previous one has been completed. The use of plan-driven methods requires that the goals of the development and requirements of the IS can be identified with sufficient accuracy in advance. If there are large uncertainties in the goals, or the environment is volatile, there is a high risk of unnecessary and expensive work in the plan-driven method, when one has to return to the planning phase from the testing phase (Hansen & Lyytinen, 2010; Sommerville, 2011).

If the goal can be identified accurately enough in advance, the plan-driven method is a very straightforward and efficient method to achieve the desired result. In cases like these, a significant advantage of the plan-driven method is that it enables professionals in ethical deliberation to comprehensively assess the impacts and considerations as part of the design phase (Figure 1) well in advance of implementing any solutions. Additionally, the waterfall model typically features more extensive documentation, which aids in the traceability and evaluation of ethical decisions. However, if there is uncertainty about the goals, or if all stakeholders are not taken into account, there is a significant risk that the implemented system will not meet all needs, and ethics will not be taken into account sufficiently. The rigidity of the waterfall model poses challenges in addressing ethical concerns that arise in the post-design phase. Similar to other design errors, ethical issues are more difficult to rectify if identified at later stages.

The change-driven methods (such as agile) are iterative and incremental in nature, meaning that planning, development, and implementation are done in small steps (sprints). Between these steps, results and objectives are re-evaluated, and necessary changes are made to the objectives. As such, change-driven methods are especially suitable for situations where there are big uncertainties related to both business execution (the existing business processes and practices) and development objectives (Dahlberg & Lagstedt, 2021). However, change-driven development poses a high risk of unnecessary steps and extra work, and, in addition, it seems to be prone to technical issues as well (Behutiye et al., 2017; Holvitie et al., 2018).

When significant uncertainties are present in the development process, it is crucial to acknowledge that not all ethical considerations can be fully addressed during the design phase. Agile development facilitates the continuous review and updating of ethical issues in each sprint or iteration, allowing for prompt responses to emerging concerns. This incremental approach simplifies the evaluation and management of ethical impacts, thereby mitigating associated risks. Moreover, the ongoing involvement of stakeholders ensures the seamless integration of ethical perspectives. However, this integration is not automatic. Continuous development discussions, primarily and typically focused on desired functionalities, must ensure that ethical considerations are included and documented in all changes and new objectives.

In plan-driven development, the main responsibility of ethical considerations lies with domain experts (such as requirement engineers). Whereas, in change-driven development, the responsibility easily rests with end users, and the product owner must ensure that the ethical discussion remains on the agenda. The ethical considerations in AI development must encompass the entire lifecycle, transitioning seamlessly from development to usage. This transition from development to operations is a recognized challenge in information system development, where changes in responsible personnel can lead to information loss, and maintenance teams may struggle to understand development decisions, hindering essential updates. This issue is particularly critical in AI, where solutions are dynamic, often requiring regular dataset updates. Ensuring data reliability and ethical use during maintenance is crucial. DevOps is a collaborative approach that integrates software development (Dev) and IT operations (Ops) to automate and streamline the software delivery process, ensuring faster, more reliable releases through shared tools, practices (e.g., CI/CD), and continuous feedback (Ebert et al., 2016). This approach, which could be called AIDevOps, should also be applied to ethical considerations in AI development such that the discussions in the development phase must not be interrupted or lost when the responsibility is transferred to the maintenance of artificial intelligence systems.

3. Methods

Given the limited existing research on AI adoption in the public sector, we opted for a qualitative multiple-case study as our research approach. This design is particularly suitable for addressing research questions related to under-explored topics, where new insights are essential, and multiple cases improve the diversity of research data (Stewart, 2012). Multiple case studies apply this requirement as it aims to understand phenomena across organizational boundaries with several in-depth case descriptions (Gustafsson, 2017; Stewart, 2012). Thus, it provides a richer sample representing different types of organizations that the interviewees represent. To comprehensively explore this under-researched phenomenon, we conducted qualitative interviews with managers and experts working for the public sector processes from 19 organizations. By delving into their experiences, expectations, and perspectives, we aimed to gain a deeper understanding. Our study employed a qualitative interview research design, focusing on the design and development processes of AI systems - an area that is still in its early stages of exploration (c.f., Gummesson, 2000). The public sector plays a significant role in the Information and Communication Technology (ICT) market, and many private AI and ICT companies work for public sector organizations (Ghezzi & Mikkonen, 2023; Hickok, 2024). Thus, we selected the public sector as the target group of our multiple case study.

3.1. Participants

The interviewees comprised AI developers, digitalization experts, and managers from the public sector's digital services domain. They represent a crucial expert group actively involved in digitization processes and AI adoption. Thus, their participation significantly enhances the study's validity. The interviews were conducted with 20 representatives from 19 organizations in Finland. These organizations included municipalities, government agencies, educational institutions, and private companies that provide services to the public sector (see Table 1). Our selection of case companies followed principles aimed at ensuring rich samples of real-world phenomena (Eisenhardt & Graebner, 2007). Specifically, we chose organizations that were recognized as leading AI adopters in the public sector or were recommended as such. Additionally, we focused on organizations dealing with large volumes of tasks, as the value potential for AI is particularly significant in such cases. Thus, the selection criterion was based more on AI adoption experience or data volume rather than the size of the organization.

Table 1. The interviewees, their affiliations, and organization types

	Position of Interviewee	Organization Type	Size	Case No
1	Expert	Educational Organization	Medium	16
2	Expert	Educational Organization	Small	14
3	Middle Management	City Organization	Large	2, 5, 6
4	Middle Management	City Organization	Medium	4, 7
5	Middle Management	Government Organization	Large	9, 11
6	Middle Management	Government Organization	Small	8, 10
7	Middle Management	Ministry	Large	12
8	Middle Management	Private Organization Solving Public Needs	Small	19
9	Top Management	City Organization	Large	3
10	Top Management	City Organization	Medium	1
11	Top Management	Educational Organization	Small	15, 17A, 17B
12	Top Management	Ministry	Medium	13
13	Top Management	Private Organization Solving Public Needs	Large	18

3.2. Data collection

The research data was collected from interviews with 19 case organizations that are working for public or government services in Finland. We selected individuals for interviews who were experts in the adoption of AI or in digital transformation within the public sector and thus could offer significant insights into this phenomenon. Hence, the interviewees represented ministries, cities, governmental organizations, educational institutions, and private organizations working for the public sector.

The researcher contacted the potential interviewees via email, outlining the study's objectives and seeking their willingness to participate. After obtaining informed consent, the researcher scheduled interview sessions and conducted interviews via Microsoft Teams, with an average duration of 45 minutes. Each session was video-recorded and automatically transcribed by Microsoft Teams. Additionally, one interview was conducted in a face-to-face setting.

The methodological approach employed semi-structured interviews (Appendix 1), wherein the researcher predetermined the main themes based on the research questions. The prior literature on AI adoption guided us to determine the themes of the interview script. The semi-structured interview methods allow for versatile and flexible data collection practices (Kallio et al., 2016). Thus, the flow of discussion within each interview session remained flexible, allowing exploration of those topics. The data analysis for this study focused on descriptions and comments related to ethical considerations in the design, deployment, and use of AI within public sector processes and tasks, or those closely associated with it. In addition, we were interested in the interviewees' views on research and development methods, processes, obstacles, opportunities, and best practices of AI adoption.

3.3. Data Analysis

The data analysis followed the principles of qualitative content analysis (Schreier, 2012). More specifically, the inductive grounded theory approach (Strauss & Corbin, 1998) was applied in data analysis and in forming new theoretical understanding. We initiated the analysis of transcribed interviews by identifying relevant excerpts based on our research themes. We applied three coding phases: open, axial, and selective (Matavire & Brown, 2013; Strauss & Corbin, 1998). During the initial round, we applied initial codes related to the AI development process and trustworthiness and ethics that were so-called sensitizing devices for us (cf. (Matavire & Brown, 2013). We annotated all comments that mentioned or explained these initial codes and their associated themes. All other material was excluded from the research data of this study. During the first coding phase, we transferred the initial relevant excerpts to a Word document, where we further organized them into smaller sub-categories based on their content. In this first coding round, we followed the principles of data-driven open coding and re-coded excerpts without predefined categories (Strauss & Corbin, 1998). In this coding phase, the role and significance of ethical issues as part of the AI development and adoption process began to crystallize.

In the second axial coding round, we created continuous improvements and categories, and drew inspiration during the data analysis, continuously moving back and forth between our data and emerging theoretical understanding (Mattarelli et al., 2013). Our data included experiences, thoughts, and insights from all phases of AI design, development, and adoption. Some of the interviewees held leadership positions in their organization's AI development, while others worked with the implementation of AI.

In the third coding phase, the results were organized into groups based on their content, from which six categories eventually emerged. In this selective analysis phase, selective coding was applied, and it resulted in six following categories: i) ethical competence becomes a key competence, ii) the ethical competence belongs to the design of AI, iii) ethical considerations begin from the requirement specification, iv) designing of human-aligned AI, v) AI solutions require careful monitoring for trustworthiness, and vi) ethical requirements vary depending on the use case. We noticed that these categories have interrelationships and dependencies with each other and follow the lifecycle and development stages of AI. We conceptualized the findings in Figure 2.

We are aware that qualitative analysis may have limitations in interpreting interview transcripts (Eisenhardt & Graebner, 2007). From the data analysis perspective, another researcher checked that the interview excerpts extracted from the interview transcripts were correctly classified into the six emerging categories and were aligned with the interview data and prior literature. We also used the ethics of AI guideline documents for strengthening evidence (see Chapter 4.1). To enhance the trustworthiness of our data analysis, we also connected our findings to existing literature. We used prior literature as a sensitizing device aligned with the evolved grounded theory approach (Matavire & Brown, 2013). Specifically, we continuously reviewed and critically reflected on our theoretical understanding of empirical findings by examining prior literature on AI adoption and development (Gummesson, 2000). In creating a new theoretical understanding from the results, we employed an abductive and iterative approach (Dubois & Gadde, 2002) which allowed us to establish explanations regarding the ethical development methods of AI. By iteratively comparing prior research with new insights from our results, we were able to make a novel theoretical contribution to the existing field. Ultimately, this abductive and iterative approach deepened our understanding of the data while simultaneously advancing the theoretical understanding of ethical AI development (Gummesson, 2000).

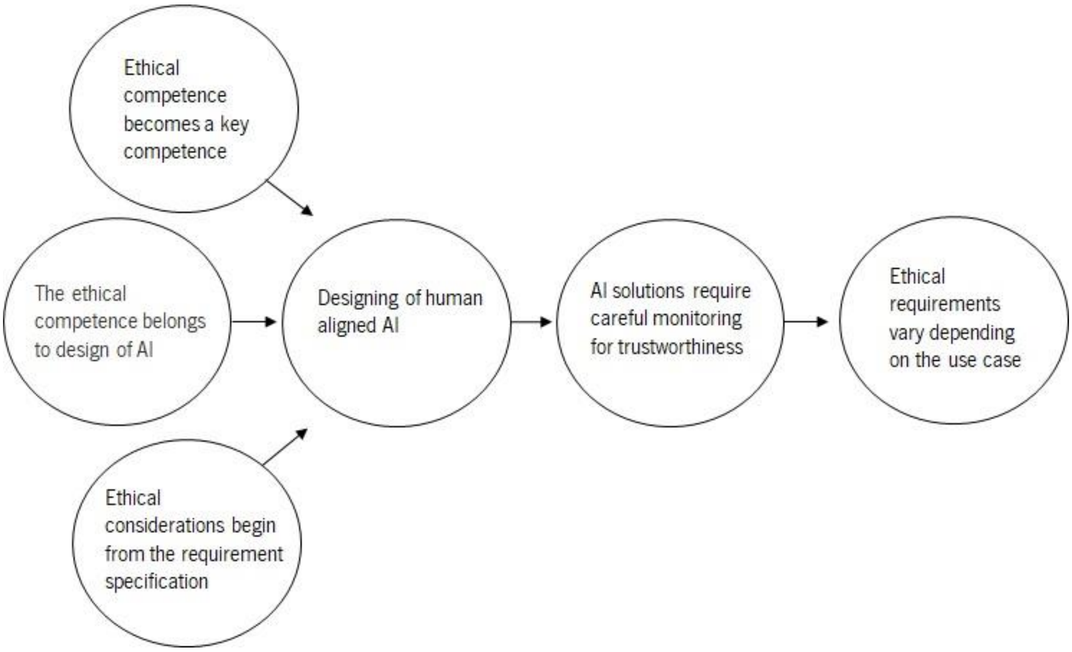

Fig. 2. The conceptual model of the key findings

4. Findings

4.1. Ethical competence becomes a key competence

The findings show that ethics plays a crucial role in designing, developing, and adopting AI. The designers, developers, and adopters should be able to manage those risks. For example, AI applications must be designed to prevent delving too deeply into personal data. They should also avoid revealing excessive information about individuals. These considerations are essential when planning and defining AI applications.

> *As we delve deeper into an individual's data, we start encountering quite personal information about their health and other aspects (Case 15).*

> *There can be ethical questions, but, well, they are not obstacles (Case 5).*

In addition to privacy, reliability is a crucial requirement when defining and designing AI applications. Reliability also impacts privacy because a dependable application is inherently secure. One interviewee pointed out that AI can start analyzing aspects beyond its original design or initial need. This represents a key risk in AI, as reliable and secure solutions are specifically tailored to address specific needs and are used accordingly. Thus, designers, developers, and adopters should have the competence to manage their AI-related actions.

> *If we speak about the adoption of technology to social and healthcare services, there are pretty much such attitudes, like ethical questions. (Case 6).*
>
> *Now, one thing is precisely this: one should understand what one is doing, because otherwise we have artificial intelligence that models something entirely different. AI solves a different problem based on its intended use (Case 16).*
>
> *Who is responsible if an artificial intelligence makes a mistake? Is it the person using the AI, or is it the developers who customized the AI, there can be risks realized, and many may hesitate to offer AI development if the legal boundaries are unclear (Case 18).*

Some public organizations (see e.g. Finnish Tax Administration, Ministry of Finance, Suomi.fi) have also published their ethical guidelines for AI, where they emphasize its importance in developing, adopting, and using AI.

4.2. AI systems should support users' ethical evaluation

The users and developers are learning to manage human-AI interaction from an ethical viewpoint. The informants stated that ethical questions or at least some doubts about the reliability of results always persist.

> *All models are inherently biased to some extent, so perhaps we should also have some form of tolerance for that. Often, people imagine that when we have an AI solution, it must always be one hundred percent right in all situations or surpass human capabilities for there to be some benefit from it (Case 9).*

Improving transparency of AI functionality and results through explainable AI was emphasized. This addresses ethical considerations where users can see the rationale of AI's decisions, leading to enhanced trustworthiness. This implies that evaluating the ethics of AI applications should be a shared responsibility, involving not only designers and developers but also the users themselves. Transparency and explainability of AI applications support this need.

> *Certainly, we should be able to evaluate precisely why artificial intelligence has arrived at a particular solution. In this one point, you might also consider the flip side: why it hasn't arrived at a different solution. Transparency is undoubtedly crucial in this regard (Case 9).*
>
> *Our big dream is to get AI for analyzing result data, so that we can see beyond the numbers (Case 15).*

4.3. The ethical competence belongs to the design of AI

Ethical competence should be part of employees' competencies. It is challenging to introduce ethical competence externally during the development process; thus, it should be built within the organization's capabilities, and some kind of ethical AIDevOps practices should be applied (see section 2.2). When designing and developing AI, ethics cannot be disregarded, as it is an essential aspect of modern AI phenomena. Hence, the ethics of AI is not only the requirement of responsible AI usage, but it also belongs to the design and development of AI systems (see Section 2.1).

> *Regarding these new AI solutions, I would perhaps add that we need individuals with expertise in AI ethics, because, currently, it's challenging to separate discussions about AI from ethical considerations (Case 9).*
>
> *Like, an assessment should be done already during the development preparation. And of course, also before implementation. If something has changed, we have an AI ethics group that kind of oversees things (Case 9).*

Evaluating ethical consequences or requirements is not an easy task, as it might be context dependent. Sometimes there are "gray areas" which require careful ethical considerations. In addition, developers, as well as end users, are concerned about accountability in cases when AI produces the wrong output. Therefore, accountability is an important pillar of the ethical AI framework and has been addressed by many ethical AI frameworks mentioned in Section 2.

> *In order for that student, for example, to be supported and guided, and for their studies to progress and everything else there. There are also many gray areas there, regarding what is meant or not to meant by fairly vague instructions, regulations, or laws, is a somewhat case-specific matter (Case 14).*

4.4. Ethical considerations begin from the requirement specification of the AI application

Ethical considerations are not solely related to the use and adoption of AI. When designing and developing AI solutions, we must assess how the combination of specific datasets might reveal information about individuals. For instance, a city organization could potentially gather substantial information about individuals if the data collected across different departments and areas were to be interconnected. This relates to the responsibility of AI developers and administrators of how they are using a large set of data with the offered AI systems. Thus, this consideration begins in the early phases of AI design, namely in the requirement specification.

> *But then, in this context of artificial intelligence, cities easily face significant challenge, and it is that basically like cities know much about individuals, and we arrive at the issue of where is the ethics of AI going on, and where we are using it, and there comes unruly situations (Case 3).*

One way to mitigate ethical risks in public sector AI projects is to select use cases that do not directly impact individuals but rather focus on the built environment, such as infrastructure and the physical environment.

> *Would it be the safest approach for cities regarding the utilization of artificial intelligence to initially apply it in areas where the actions do not directly impact people but rather relate to physical environments, transportation networks, building infrastructure, and natural areas and beyond (Case 3).*
>
> *Somewhat, I believe that when it comes to numerical data, the more unambiguous the initial information collected, the more reliable the results that artificial intelligence can produce (Case 19).*

4.5. Designing of human-aligned AI

Designing human-aligned AI applications requires the involvement of end-users in the design and development phases. This way, the designers and developers can iteratively test new AI applications and improve them based on the feedback. This user-centered design approach increases the ethical capabilities of the AI systems.

> *We tested it with a few users, and it soon became clear that they doubted where this 'guess box' was getting its 'good or bad' or 'accept or reject' decisions from. We figured it out and implemented explainable AI, which actually helped (Case 8).*

Design capability was emphasized. An informant of an advanced AI adopter organization pointed out that we should be able to design human-centric AI applications. This does not refer to a service design process, but to AI applications that are context-aware and human-aligned AI systems. Organizations should manage those design capabilities at a strategic level, as human centricity is a non-functional requirement that should be raised to a requirement in the early phases of design processes.

> *We increasingly understand the environment and humans, particularly human orientation. And we integrate these more and more, and the human orientation connected to the capabilities created by technology and its transformation. And our aim is to continually strengthen that at a very strategic level (Case 11).*

This is a management challenge of AI development, as the developers should be able to review design goals from a broader perspective than technological opportunities. Putting humans and usage context at the center of the design processes requires strategic changes in the management of AI adoption and development.

4.6. AI solutions require careful monitoring for trustworthiness

Unlike conventional software, the functioning of AI solutions might change. AI solutions or their algorithms require training with data before they are ready for use. This is in stark contrast to conventional software, which is considered 'ready-made' once it has undergone testing and is put into production use. Conventional software operates consistently, regardless of the volume of data it processes, adhering to its pre-programmed logic. However, it does not adapt its output based on new information flowing through it. In contrast, AI algorithms dynamically analyze data, adjusting their output based on the information in the data. Unfortunately, this adaptability can lead to gradual biases or discriminatory behavior in AI solutions, even if initial tests align with expert perspectives. This makes significant changes to the development and maintenance processes of AI solutions compared to conventional software maintenance in organizations. In addition, organizational restructuring might also create privacy risks if the access rights of AI systems are changed accordingly. The robustness of AI models is also an important property that refers to maintaining a decent performance for varying inputs and scenarios. This is particularly important for sensitive applications where a wrong prediction could have drastic consequences. At the same time, this also comes up as one of the major concerns or requirements of the end users.

> *If we had an AI model that, for instance, made decisions, we would need to monitor the behavior of this AI model to ensure that it doesn't exhibit any drift, so that it should consistently make decisions in a similar manner over time, or it does not learn discriminatory tendencies or other undesirable traits we need to actively monitor regarding to these models, especially when considering ethical aspects or similar ones. Perhaps this falls within the domain of expertise for ML engineers, addressing areas that traditional development processes often overlook (Case 9).*

In addition to technological challenges, organizational structures might also affect the ethical use of AI or ensure its trustworthiness.

> *This caused, once again, an administrative boundary where suddenly they couldn't access anything, almost like they were in the dark, unable to see anything, when previously, they had been able to get so with certain restricted rights (Case 14).*

The target of AI development moves constantly. A challenge in developing AI solutions is that AI technologies are developing at an enormous speed. Thus, the requirement specification may be outdated when a slow implementation process starts, as new models and methods may have been released during the definition and design phases. Hence, an ethical AI system today is not necessarily ethical tomorrow.

> *If we look at history, it's true that symbolic AI is now in an AI winter, while computational AI, you know. Deep learning networks are now mainstream (Case 9).*

4.7. AI-enhanced application categories – different contexts for ethics of AI

Organizations use AI for a number of applications including financial management systems, processing applications, predicting customer demand, anticipating the progress of studies, customer service, or monitoring visitor flows. Some have already integrated generative AI into the intranet, but it is largely embedded in office applications. Table 2 lists the AI applications that the interviewees mention to have or are going to have. Based on the lists and how they produce value for the organization, we clustered the applications into eight categories.

Table 2 shows that AI applications are associated with different processes, user groups, and usage contexts. In designing, developing, and deploying AI applications, we should review the special needs of processes, users, and the environment where AI applications will be used.

Table 2. AI-based information system categories found in the case organizations

AI-based application examples that the interviewees mentioned	AI-based information system category	Description
Predicting dropouts, predicting customer demand	Management AI	Assist in predicting needs or risk factors. Increased efficiency and productivity.
Financial process automation, proposal evaluation	Administration AI	Automation and optimization of administrative processes. Reduced costs.
Study planning, competence profiling, thesis planning support	Educational AI	Management support for educators and administration. Learning support for students.
Border control, waste container indicator, cleaning robot	Operative AI	Automation of the monitoring of routine tasks. Reduced costs.
Large language models such as ChatGPT for assisting knowledge work, automatic transcript, and generative AI embedded in the intranet	Office AI	Knowledge support for experts. Assist in routine tasks by saving time and resources.
Chatbot, customer feedback	Communication AI	24/7 services for customers. Assist in improving customer experience.
Software robotic in financial processes, email filtering, application classifier, a predictor of visitor flow, scanned document reader	Transactional AI	Automation and optimization of experts' workload. Reduced costs.
Traffic management, Internet of Things in a city infrastructure	Infrastructure AI	Automation and optimization of monitoring processes. Improved decision-making.

5. Discussion

This study identified six interrelated phenomena of AI ethics during the AI lifecycle (Table 3). Some of them are related to the organization's AI design and development capabilities and practices, while others are related to the adoption and use of AI. AI ethics is often thought to be related to the adoption of AI, and many ethical guidelines for AI focus on the adoption and usage phases rather than AI design and development. However, this study shows that AI ethics is also part of the organization's design and development capabilities. Thus, it also relates to the design and development processes, in addition to activities during adoption and use.

This study has several theoretical contributions. First, we expanded the ethical discussion of AI to include the design and development process of AI systems. Prior research on the ethics of AI has mainly focused on AI adoption, whereas the perspectives of design and development processes have remained scarcer. Also, we should focus on ethical considerations throughout the entire AI system lifecycle, not only from the viewpoint of AI adoption and usage. Our study is grounded in empirical evidence where participants validate the findings. Ethical considerations are particularly important from the initial phases as they lay the groundwork for the whole AI adoption and usage processes. This study further indicates that ethical considerations are relevant to every role involved in the AI system lifecycle. All actors within the lifecycle of AI applications should be engaged with ethical considerations. Moreover, different use cases set various requirements for ethical considerations since organizations and usage environments may raise special requirements. Consequently, this study contributes to the debate on ethical AI system design from the life cycle perspective of AI applications.

Our second contribution is to the design, development, and adoption of AI in the public sector, as our results particularly reflect the views of AI adopters in this sector. Existing studies (Misra et al., 2023; Neumann et al., 2024) demonstrate that AI is a significant technological enabler for developing public management practices and processes. For instance, city and government organizations could potentially gather substantial information about individuals if the data collected across different service sectors were to be interconnected. Public sector organizations often process large data flows and have significantly large data sets that provide an excellent starting point for automation. Public sector organizations have large amounts of data that do not directly impact individuals but rather focus on the built environment, such as infrastructure and the physical environment. In those cases, the ethical risk is lower.

The third contribution shows that ethical consideration is not only a single task in AI adoption, but it is an integral requirement from the design to the implementation and use of AI. Unlike conventional software systems, AI adapts to the situational environment and continuously uses data to enhance its performance. Hence, ethical risks and the trustworthiness of its results require continuous monitoring by AI adopters. Advancements in AI technologies and data management practices boost the development of more advanced, ethical, and trustworthy AI solutions. The frameworks for the ethical design of AI systems (Vakkuri & others, 2021) also lower design and development barriers to designing new ethical AI solutions. Prior studies (e.g. (Kaniadakis & Linturn, 2022)) emphasize the crucial role of interdisciplinary teams in technology adoption, who search for, define, and solve development challenges and opportunities. Our study extends this remark to all phases of the AI lifecycle, where interdisciplinary competencies play an essential role

The fourth contribution of this study (see Table 3) shows that ethical considerations regarding AI design and deployment should begin from the resourcing of the design team with persons with ethical competence, as several foundational decisions are made before the beginning of the development processes. This is consistent with the study by Smith (2019) which highlights the need for diverse teams, including members with ethical competence, to guide the development of trustworthy AI systems. Additionally, the findings reveal that ethical competence should be a part of AI-related design competencies and thus, an integral part of the design process. This is in line with IBM's AI design ethics overview which emphasizes that an ethical, human-centric AI must be developed in a manner consistent with the values and ethical standards of the affected community. Furthermore, the ethical design of AI relates to the designing of human-aligned AI. Human-aligned AI is both ethical and human-centric AI supporting human autonomy and trustworthy decision-making. This finding aligns with the study of He et al. (2021), who advocate for a human-centric AI framework that integrates high levels of human control with advanced automation. Our findings underscored the importance of design competence in AI, especially as human-aligned AI becomes increasingly central to its adoption. It is consistent with the study by Tjondronegoro et al. (2022) which highlights the importance of integrating ethical considerations throughout AI development, and by Goyal et al. (2024) which highlights the need for understanding human expectations in designing AI agents, emphasizing that aligning AI behaviour with user preferences is crucial for effective human-AI collaboration.

Unlike conventional software systems, AI solutions require careful monitoring of trustworthiness as they are continually learning and evolving systems. As the final construct, AI systems should support users' ethical evaluation while using AI applications. The demand for explainability as a part of AI systems supports this requirement. This finding is supported by Shin's (2021) empirical work that demonstrates that explainability promotes trust and empowers users to ethically engage with AI systems, supporting the need for user-centric explanatory interfaces.

The fifth contribution of this study relates to AI applications in the public sector. Based on the findings (Table 2), we conceptualized eight AI-based information system categories. The AI-based system categories have unique use cases and value propositions. Transactional AI and infrastructure AI create the basis and operating platforms for other information systems, whereas office AI, communication AI, educational AI, and management AI focus more on end-user value co-creation. Administration AI typically supports administrative processes in internal or cross-institutional reporting and monitoring needs. From the ethical viewpoint, all eight categories have their specific ethical risks depending on the data they are analyzing and managing, and their role in decision-making. Transactional AI, infrastructure AI, and office AI do not directly generate content and results that are directly associated with ethical issues, unlike educational AI and management AI, which are more related to human decisions and recommendations. However, all those categories should meet the requirements of trustworthy AI, which is a part of ethical principles.

Table 3. The role of ethics in designing and developing AI applications

Key findings	Explanation	Actions	Phase of lifecycle
Ethical competence becomes a key competence	Competence to define ethical requirements is a critical skill in AI projects	Ethical competence is a resource that should be included in the project plan	Design
Ethical competence belongs to the design of AI	Ethical consideration is a crucial part of the design phase	The design team should have a person who has the knowledge and skills to define ethical requirements	Design Develop
Ethical considerations begin with the requirement specification	Ethical consideration should begin in the early phases of an AI project when requirements for an AI application are defined	Ethics of AI is a non-functional requirement that affects other functional requirements in the specification of the AI application	Design Develop
Designing of human-aligned AI	Designing human-aligned AI applications includes the implementation of ethical requirements	The fulfillment of ethical requirements should be tested to ensure the AI application meets the principles of human-aligned AI application	Design Develop
AI solutions require careful monitoring for trustworthiness	The ability of AI applications to function ethically is associated with their continuous trustworthiness	The functions of explainability and human-in-the-loop ensure that AI applications generate ethically sustainable results	Operate
Ethical requirements vary depending on the use case	Ethical requirements are associated with the usage context of AI and thus they are context-dependent	The use case and usage context of an AI application affect the sensitivity of data that the AI application analyses, generates, and manages.	Deploy Operate Retire

The study found what ethical issues are emphasized as part of AI system development and adoption in case organizations. We summarized the findings in Table 3, which emphasizes the integration of ethical issues in all phases of the AI design and development process. This study emphasizes the role of ethical issues in the requirement specification where the most important ethical selections are made. Additionally, ethical competence should be seen as a key competence in AI design and development teams, as it is an integral part of AI models, data, and systems. More specifically, all ethical considerations are not associated with the usage and implementation phases of AI, but many fundamental decisions regarding ethics are made in the design and development phases of AI systems. In addition, it is important that the ethical discussion is not interrupted when moving from the development phase to the maintenance phase. To support this continuity, there is a need for ethical AIDevOps practices that integrate ethics into the entire AI lifecycle. What kind of AIDevOps practices should be in place to best support an ethical continuum throughout the entire lifecycle is an important subject for further research.

This study examined how ethical considerations should be integrated into AI development processes. Based on the findings, we further analyzed their implications for organizations, as well as for AI design and development practices, and ethical considerations. This study points out that rapidly developing AI technologies, system components, data sets, and AI models challenge the design and development processes from technical and ethical perspectives. AI design and development is a more complex process than designing conventional software systems. Consequently, it requires more ethical design competencies and multidisciplinary collaboration from various levels of an organization.

6. Limitations and future research agenda

The scope of the study was limited to public sector organizations in Finland, which may affect the generalizability of the findings. While the study provides an in-depth analysis of AI ethical challenges within this specific context, the ethical considerations and development practices may vary across different regions, cultural settings, and public sector environments.

The study relied on qualitative interviews as the primary data collection method. Although this approach allowed us to capture rich, detailed insights into the ethical challenges faced by public sector organizations, it also introduces potential biases. The subjective nature of interviews means that findings are shaped by the perspectives of the interviewees, and there is a risk that some ethical concerns may not have been fully disclosed or emphasized during the interviews. Especially, considering the AI lifecycle shown in Figure 1, it is evident that ethical considerations encompass the whole AI lifecycle which is much deeper than commonly understood. In contrast, the interviewees often see only a part of the AI lifecycle; therefore, their point of view may be limited. That said, it is still important to analyze their opinions as they are the major stakeholder of this system.

The dynamic nature of AI technologies presents a challenge for capturing the rapidly evolving ethical issues associated with AI adoption. AI technologies particularly generative AI and large language models are advancing at an unprecedented rate, and ethical concerns may shift as these technologies become more sophisticated and embedded into public sector systems. As a result, some ethical challenges discussed in this study may become outdated or less relevant as new AI models and frameworks emerge.

Future studies could expand the scope of this study to include private and public sector organizations from different regions or countries, allowing for a comparative analysis of ethical AI challenges and best practices across diverse socio-political contexts. Additionally, quantitative research could be conducted to measure the impact of ethical AI adoption on organizational performance and trustworthiness from both employees' and citizens' perspectives. Investigating the long-term effects of ethical AI systems on public services would also provide deeper insights into how continuous ethical monitoring can sustain trust in AI over time. Moreover, continuous research is needed to address the evolving challenges and ensure that ethical guidelines remain applicable and effective.

Future research should adopt a more comprehensive, lifecycle-oriented approach to examining AI ethics. A holistic perspective across the entire AI lifecycle, including design, development, implementation, operation, and retirement phases, can offer deeper insights into the cumulative ethical impacts. Such an approach could reveal critical interactions between phases that influence ethical outcomes, such as the long-term effects of design choices on model robustness and accountability in later stages.

7. Conclusions

This study contributes to the growing body of knowledge on ethical AI development by exploring the challenges faced by public sector organizations. Our findings highlight the importance of integrating ethical considerations into every stage of the AI system lifecycle, from the initial design and development phases to deployment, monitoring, and retirement. By utilizing empirical data from multiple public sector case organizations in Finland, we have provided both theoretical insights and practical guidelines for developing ethically aligned AI systems.

We emphasize the need for a comprehensive, lifecycle-oriented approach to ethical AI development. The AI lifecycle spans various phases that collectively shape the ethical impact of AI applications. Addressing ethical concerns at each stage ensures that foundational principles like transparency, accountability, and fairness are not just design concepts but integral components that guide the system's behavior over time. By embedding ethical practices throughout the lifecycle, organizations can anticipate and mitigate risks more effectively, which will lead to greater trust and reliability in AI solutions. This holistic perspective is especially relevant to the public sector, where AI systems directly impact citizen well-being and must align with public values. Therefore, prioritizing lifecycle-wide ethical considerations to support sustainable, responsible, and human-centered AI applications is very important.

From a scientific literature perspective, this study adds to the limited empirical research on ethical AI development in the public sector. While much of the existing literature has focused on high-level principles and frameworks, this research offers concrete, real-world insights into what ethical considerations should be integrated into AI development processes. The study highlights the importance of ethical competence as a core skill within AI design teams, emphasizing that ethics cannot be treated as an afterthought but must be embedded throughout the AI lifecycle. This contribution is particularly significant for public sector organizations, where trust, transparency, and fairness are critical due to the direct impact of AI systems on citizens and public services.

For practitioners, our research provides actionable recommendations on how to incorporate ethical guidelines into AI projects, addressing issues such as human alignment, accountability, transparency, and the continuous monitoring of AI systems. These recommendations can guide AI developers and managers in making more informed decisions about ethical risks and responsibilities in AI deployment.

Acknowledgments

Ministry of Education (OKM/108/523/2021) and Helia Foundation.

References

AI NIST. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Amugongo, L. M., & others. (2023). Operationalising AI ethics through the agile software development lifecycle: a case study of AI-enabled mobile health applications. *AI and Ethics*, 1–18.

Andrews, L. (2019). Public administration, public leadership and the construction of public value in the age of the algorithm and "big data." *Public Administration*, *97* (2), 296–310.

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022a). Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management*, *62* (September 2021), 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022b). Ethical Framework for Artificial Intelligence and Digital Technologies. *International Journal of Information Management*, *62*, 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433

Behutiye, W. N., Rodríguez, P., Oivo, M., & Tosun, A. (2017). Analyzing the concept of technical debt in the context of agile software development: A systematic literature review. *Information and Software Technology*, *82*, 139–158. https://doi.org/10.1016/j.infsof.2016.10.004

Bommasani, R., Klyman, K., Zhang, D., & Liang, P. (2023). *Do Foundation Model Providers Comply with the EU AI Act?* https://crfm.stanford.edu/2023/06/15/eu-ai-act.html

Brey, P., & Dainow, B. (2024). Ethics by design for artificial intelligence. *AI and Ethics*, *4* (4), 1265-1277.

Dahlberg, T., & Lagstedt, A. (2018). There Is Still No "Fit for All" IS Development Method: Business Development Context and IS Development Characteristics Need to Match. *Proceedings of the 51st Hawaii International Conference on System Sciences*, *9*. https://doi.org/10.24251/hicss.2018.604

Dahlberg, T., & Lagstedt, A. (2021). Fit to Context Matters – Selecting and Using Information Systems Development Methods to Develop Business in Digitalization Contexts. *Proceedings of the 54th Hawaii International Conference on System Sciences 2021*, 6902–6911. https://doi.org/10.24251/HICSS.2021.829

De Silva, D., & Alahakoon, D. (2022). An artificial intelligence life cycle: From conception to production. *Patterns*, *3* (6), 1-19.

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

Ebert, C., Gallardo, G., Hernantes, J., & Serrano, N. (2016). DevOps. *IEEE Software*, *May/June*, 94–100.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

European Commission. (2024). *AI Act | Shaping Europe's digital future*. https://ec.europa.eu/digital-single-market/en/ai-act

European Parliament. (2023). *EU AI Act: first regulation on artificial intelligence*. https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., & others. (2022). *CapAI-A procedure for conducting conformity assessment of AI systems in line with the EU artificial intelligence act*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4064091

Frey, C. B., Garlick, R., Friedlander, G., Mcdonald, G., Wilkie, M., & Lai, A. (2016). *TECHNOLOGY AT WORK v2.0* (Issue January). CityGroup and University of Oxford.

Ghezzi, R., & Mikkonen, T. (2023). On Public Procurement of ICT Systems: Stakeholder Views and Emerging Tensions. *International Conference on Software Business*, 61–76.

Goyal, N., Chang, M., & Terry, M. (2024). *Designing for Human-Agent Alignment: Understanding what humans want from their agents*.

Gummesson, E. (2000). *Qualitative Methods in Management Research* (2nd ed.). Sage Publications.

Gustafsson, J. (2017). *Single case studies vs. multiple case studies: A comparative study*. Halmstad University, School of Business, Engineering and Science.

Hansen, S., & Lyytinen, K. (2010). Challenges in contemporary requirements practice. *Proceedings of the Annual Hawaii International Conference on System Sciences*. https://doi.org/10.1109/HICSS.2010.98

He, H., Gray, J., Cangelosi, A., Meng, Q., McGinnity, T. M., & Mehnen, J. (2021). The Challenges and Opportunities of Human-Centered AI for Trustworthy Robots and Autonomous Systems. *Proceedings of the IEEE*, 109(4), 1–17. https://doi.org/10.1109/JPROC.2021.3068223

Hedlund, M., & Persson, E. (2025). Distribution of responsibility for AI development: expert views. *AI & Society*. https://doi.org/10.1007/s00146-024-02167-9

Hickok, M. (2024). Public procurement of artificial intelligence systems: new risks and future proofing. *AI & Society*, *39*(3), 1213–1227.

High-Level Expert Group on AI (AIHLEG). (2018). *Ethics guidelines for trustworthy AI | Shaping Europe's digital future*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

High-Level Expert Group on AI (AIHLEG). (2019). *Ethics Guidelines for Trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Holvitie, J., Licorish, S. A., Spínola, R. O., Hyrynsalmi, S., MacDonell, S. G., Mendes, T. S., Buchan, J., & Leppänen, V. (2018). Technical debt and agile software development practices and processes: An industry practitioner survey. *Information and Software Technology*, *96*(Nov. 2017), 141–160. https://doi.org/10.1016/j.infsof.2017.11.015

Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An Overview of Artificial Intelligence Ethics. *IEEE Transactions on Artificial Intelligence*, *4*(4), 799–819. https://doi.org/10.1109/TAI.2022.3194503

Huriye, A. Z. (2023). The ethics of artificial intelligence: examining the ethical considerations surrounding the development and use of AI. *American Journal of Technology*, *2*(1), 37–44.

IBM. (n.d.). *AI design ethics overview*. Retrieved April 7, 2025, from https://www.ibm.com/design/ai/ethics/

Jan, Z., Ahamed, F., Mayer, W., Patel, N., Grossmann, G., Stumptner, M., & Kuusk, A. (2023). Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities. *Expert Systems with Applications*, *216*, 119456. https://doi.org/10.1016/j.eswa.2022.119456

Kallio, H., Pietilä, A.-M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, *72*(12), 2954–2965. https://doi.org/10.1111/jan.13031

Kamila, M. K., & Jasrotia, S. S. (2023). Ethical issues in the development of artificial intelligence: recognizing the risks. *International Journal of Ethics and Systems*.

Kaniadakis, A., & Linturn, L. (2022). Organizational adoption of Robotic Process Automation: managing the performativity of hype. *International Journal of Information Systems and Project Management*, *10* (4), 20–36.

Lagstedt, A. (2019). *Selecting the Right Method for the Right Project*. Turku University, Turku School of Economics.

Lahiri, S., & Saltz, J. (2024). The need for a risk management framework for data science projects: a systematic literature review. *International Journal of Information Systems and Project Management*, *12* (4), 41–57.

Leesakul, N., Oostveen, A. M., Eimontaite, I., Wilson, M. L., & Hyde, R. (2022). Workplace 4.0: Exploring the Implications of Technology Adoption in Digital Manufacturing on a Sustainable Workforce. *Sustainability (Switzerland)*, *14* (6). https://doi.org/10.3390/su14063311

Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, *7* (1), 1–7. https://doi.org/10.1057/s41599-020-0501-9

Lu, Q., & others. (2022). Software engineering for responsible AI: An empirical study and operationalised patterns. *Proceedings of the 44th International Conference on Software Engineering: Software Engineering in Practice (pp. 241-242).*

Lu, Q., Zhu, L., Xu, X., & Whittle, J. (2023). Responsible-AI-by-design: A pattern collection for designing responsible AI systems. *IEEE Software*. *40* (3), 63-71.

Matavire, R., & Brown, I. (2013). Profiling grounded theory approaches in information systems research. *European Journal of Information Systems*, *22*(1), 119–129. https://doi.org/10.1057/ejis.2011.35

Mattarelli, E., Bertolotti, F., & Macrì, D. M. (2013). The use of ethnography and grounded theory in the development of a management information system. *European Journal of Information Systems*, *22*(1), 26–44. https://doi.org/10.1057/ejis.2011.34

Misra, S. ~K., Sharma, S. ~K., Gupta, S., & Das, S. (2023). A framework to overcome challenges to the adoption of artificial intelligence in Indian Government Organizations. *Technological Forecasting and Social Change*, *194*, 122721.

Moe, N. B., Aurum, A., & Dybå, T. (2012). Challenges of shared decision-making: A multiple case study of agile software development. *Information and Software Technology*, *54*(8), 853–865. https://doi.org/10.1016/j.infsof.2011.11.006

Navaie, K. (2024). Personal Data Protection in AI-Native 6G Systems. *ArXiv Preprint ArXiv:2411.03368*. https://arxiv.org/abs/2411.03368

Neumann, O., Guirguis, K., & Steiner, R. (2024). Exploring artificial intelligence adoption in public organizations: a comparative case study. *Public Management Review*, *26*(1), 114–141.

Nikolinakos, N. T. (2023). Realising Trustworthy AI: Implementation of the Key Ethical Requirements. In *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies - The AI Act* (Vol. 53). Springer, Cham. https://doi.org/10.1007/978-3-031-27953-9_4

OECD. (2019). *Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)*. OECD Digital Economy Papers, No. 291. Retrieved April 7, 2025, from https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/11/scoping-the-oecd-ai-principles_71e1b6dc/d62f618a-en.pdf

OECD. (2022). *OECD Framework for the Classification of AI systems* (Issue 323). OECD Publishing. https://doi.org/10.1787/cb6d9eca-en

Ronanki, K. (2023). Towards an AI-centric Requirements Engineering Framework for Trustworthy AI. *2023 IEEE/ACM 45th International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*.

Royce, Dr. W. W. (1970). Managing the Development of large Software Systems. *IEEE Wescon*, *August*, 1–9.

Sanderson, C., & others. (2022). Towards implementing responsible AI. *2022 IEEE International Conference on Big Data (Big Data)*.

Schreier, M. (2012). *Qualitative Content Analysis in Practice*. SAGE Publications.

Selten, F., & Klievink, B. (2024). Organizing public sector AI adoption: Navigating between separation and integration. *Government Information Quarterly*, *41* (1), 101885.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551.

Smith, C. J. (2019). *Designing Trustworthy AI: A Human-Machine Teaming Framework to Guide Development*. *arXiv preprint arXiv:1910.03515*

Sommerville, I. (2011). *Software Engineering* (9th ed.). Addison-Wesley. https://doi.org/10.1111/j.1365-2362.2005.01463.x

Stewart, J. (2012). Multiple-case study methods in governance-related research. *Public Management Review*, *14* (1), 67–82. https://doi.org/10.1080/14719037.2011.589618

Straub, V.J., Morgan, D., Bright, J., & Margetts, H. (2023). Artificial intelligence in government: Concepts, standards, and a unified framework. *Government Information Quarterly*, *40* (4), 101881.

Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (2nd ed.). Sage Publications.

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare, *Government Information Quarterly*, 36 (2), 368-383.

Taddeo, M., Blanchard, A., & Thomas, C. (2024). From AI ethics principles to practices: a teleological methodology to apply AI ethics principles in the defence domain. *Philosophy & Technology*, *37* (1), 42.

Tjondronegoro, D., Yuwono, E., Richards, B., Green, D., & Hatakka, S. (2022). *Responsible AI Implementation: A Human-centered Framework for Accelerating the Innovation Process*. *arXiv preprint arXiv:2209.07076*.

Vakkuri, V., & others. (2021). ECCOLA—A method for implementing ethically aligned AI systems. *Journal of Systems and Software*, *182*, 111067. https://doi.org/10.1016/j.jss.2021.111067

Appendix 1

The interview script (semi-structured interview)

1. What is the current state of AI application?
2. What opportunities do you see? Which processes could benefit from it?
3. What obstacles are there to AI adoption?
4. What would be the minimum level of expertise required in the organization at different levels when implementing AI solutions?
5. What collaboration should take place between different departments when implementing AI?
6. What kind of task volumes do you have and how are they managed? Where are they located, who is responsible, and could they be optimized with AI?
7. What ethical or reliability issues related to AI have you observed or should be considered?

Biographical notes

Ari Alamäki is a Principal Lecturer at Haaga-Helia University of Applied Sciences, Helsinki, Finland and Adjunct Professor at University of Turku, Finland. His current research focuses on the applications of artificial intelligence in education and business services. He received his PhD in technology education in 1999 from the University of Turku, Finland. He was a member of European Commissions' expert group on artificial intelligence and data in education and training in 2021-2022, and he has coordinated AI Innovation Hub of Ulysseus European University. He has also worked in management positions in ICT industry from 2000 to 2011.

Umair Ali Khan is a Senior Researcher in Digitalization & Artificial Intelligence group at Haaga-Helia University of Applied Sciences in Helsinki, Finland, where his research focuses on applied AI, machine learning, and generative AI. In particular, his areas of interest include knowledge extraction and processing using large language models, AI agents, advanced retrieval generation methods, and knowledge graph integration with LLMs. Umair also serves as an AI Advisor for the Finnish AI Region (FAIR) project - one of the initiatives of European Digital Innovation Hubs (EDIHs) under the Digital Europe Program - where he assists SMEs on their AI adoption journey.

Altti Lagstedt is a principal lecturer at Haaga-Helia University of Applied Sciences, Helsinki, Finland. He holds a PhD in Information Science and an MSc in Technology. His expertise includes information systems development, business process digitalization, and artificial intelligence, for which he has written several academic and practical publications. In several projects, he has been studying digital transformation and artificial intelligence from different perspectives and in various industries, with a focus on business processes. He is also leading Haaga-Helia's portion of the EUSAiR project (https://eusair-project.eu/), which, during the years 2024-2026, will define AI sandbox solutions in accordance with the AI regulation for the entire EU.