TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Sofia Paes 211979IVGM

From Ethics to Action: A Study of Human-Centric AI Implementation in Public Services, Comparing the Estonian Approach with Approaches Used in Other Countries

Master's thesis

Supervisor: Innar Liiv
PhD

Tallinn 2023

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Sofia Paes 211979IVGM

Eetikast praktiliste tegevusteni: uuring inimkeskse tehisintellekti rakendamisest avalikes teenustes, Eesti lähenemisviisi võrdlus teiste riikidega

magistritöö

Juhendaja: Innar Liiv
PhD

Tallinn 2023

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Sofia Paes

18.12.2023

Abstract

Approximately 120 AI solutions have been implemented in the Estonian public sector as of today, and this number remains growing. However, widespread adoption of AI technology is not only a source of pride, but also a significant responsibility for the state, which seeks to automate its work. Implemented AI solutions should be human-centric, generating public value while protecting individuals' fundamental rights and upholding democracy and the rule of law. Today, numerous ethical guidelines attempt to establish a framework for progressing towards more human-centric AI, but there remains a gap in understanding how to effectively implement existing ethical principles. This qualitative study based on semi-structured interviews explores how countries move from ethics to action when developing, deploying, and using AI tools in public services, with a final research objective of making Estonia's human-centric AI approach more complete.

Research reveals a significant level of interest in the human-centric AI approach across countries, caused by the rapid advancement of AI technology and its broader integration in both the private and public sectors. For addressing the need in the real-life application of AI ethics countries developed and applied number of legal regulations, practical tools, and supportive measures that advance the human-centric AI approach. In comparison to other countries, Estonia stands out for its highly centralised approach to AI in the public sector, which allows for both innovation and control over agencies implementing AI solutions. Meanwhile, there is still some unrealized potential in the ethical assessment and transparency enhancement of AI solutions used in the public sector, as well as room to grow in knowledge promotion techniques for achieving a more comprehensive human-centric AI approach.

Keywords: artificial intelligence, human-centric artificial intelligence, public sector, ethics

This thesis is written in English and is 75 pages long, including 6 chapters, and 4 tables.

Annotatsioon

Eetikast praktiliste tegevusteni: uuring inimkeskse tehisintellekti rakendamisest avalikes teenustes, Eesti lähenemisviisi võrdlus teiste riikidega

Tänapäeval Eestis avalikus sektoris on rakendatud ligikaudu 120 tehisintellekti lahendust ning nende arv jätkuvalt kasvab. Siiski tehisintellekti ulatuslik levik avalikus sektoris ei ole ainult põhjus uhkuseks digitaalselt arenenud riigi üle, vaid tehisintellekti kasutamisega kaasneb ka suur vastutus. Peab tegema kindlaks, et rakendatud tehisintellekti lahendused oleks inimkesksed ehk tooks ühiskonnale kasu, oleks inimeste põhiõigustega kooskõlas ning tugineks demokraatlike väärtustele. Vaatamata sellele, et praeguseajal ilmub palju juhiseid, mis püüavad luua raamistikku inimkeskse tehisintellekti loomiseks ja kasutuselevõtuks, üleminek eetilistest põhimõtetest päris tegevusteni sama kiirelt ei toimu ning tihti jääb lünklikuks. Antud kvalitatiivse uuringu eesmärk on uurida kuidas tehisintellekti lahendusi kasutavad riigid päriselus jõuavad inimkesksete lahendusteni ning mis praktilisi meetmeid nad selle jaoks kasutavad. Teiste riikide praktikate väljaselgitamine on vajalik Eesti enda inimkeskse lähenemisviisi hindamiseks ja puudujääkide eemaldamiseks. Läbiviidud uuring näitab märkimisväärset huvi inimkeskse lähenemisviisi vastu, mida võib seostada tehisintellekti kiire arengu ja integratsiooniga nii era- kui ka avalikus sektorites. Eetiliste printsiipide rakendamiseks riigid välja töötasid mitmeid meetmeid, mis hõlmavad õigusnorme, praktilisi tööriistu ja muid toetusmeetmeid. Eesti siiski erineb oma tsentraliseeritud lähenemisega tehisintellekti rakendamise suhtes, mis soodustab nii innovatsiooni kui ka loob kontrollmehhanisme juurutavate tehisintellekti lahenduste üle. Samal ajal uuring näitab, et riigil on veel palju realiseerimata potentsiaali hindamismehhanismide ning läbipaistvust tagavate meede loomise vaates. Samuti jääb arenguruumi teadmiste edendamise ja tervikliku inimkeskse tehisintellekti kontseptsiooni loomiseks.

Märksõnad: tehisintellekt, inimkeskne tehisintellekt, avalik sektor, eetika
Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 75 leheküljel, 6 peatükki, 4 tabelit.

List of abbreviations and terms

AI	*Artificial intelligence*
AI HLEG	*High-Level Expert Group on Artificial Intelligence*
BE	*Belgium*
CDEI	*Centre for Data Ethics and Innovation of The United Kingdom*
CA	*Canada*
DIGG	*Agency for Digital Government of Sweden*
EU	*European Union*
FI	*Finland*
GB	*United Kingdom*
GDPR	*General Data Protection Regulation*
MKM	*Ministry of Economic Affairs and Communications of Estonia*
NL	*Netherlands*
OECD	*Organisation for Economic Co-operation and Development*
PET	*Privacy-enhancing technologies*
RIA	*Information System Authority of Estonia*
RQ	*Research question*
SE	*Sweden*
UNESCO	*United Nations Educational, Scientific and Cultural Organization*

Table of contents

1 Introduction 10
2 Theoretical framework 13
2.1 Setting the scene: background to AI 13
2.2 AI in the public sector of Estonia 14
2.3 Importance of AI ethics 16
2.4 Ensuring AI ethics through soft and hard law 17
2.5 Achieving human-centric AI through ethical principles 18
2.6 Critique of AI ethics: lack of practical implementation 20
3 Methodology 22
3.1 Data collection methods 22
3.2 Study sample 23
3.3 Data analyses methods 25
3.4 Limitations 25
4 Research outcomes and findings 27
4.1 Defining human-centric AI approach 27
4.1.1 Defining AI 27
4.1.2 Use of AI in the public sector 29
4.1.3 Defining human-centric AI 31
4.1.4 Interest towards human-centric AI approach 33
4.2 Translating human-centric AI approach into life 35
4.2.1 Regulations 35
4.2.2 Practical tools 37
4.2.3 Other measures 42
4.2.4 Challenges and interest 45
4.3 Human-centric AI approach in Estonia 46
4.3.1 Human-centric AI measures 46
4.3.2 Further development 52

5 Results and discussion 54
6 Summary 62
References 66
Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis 73
Appendix 2 – Interview questions 74

List of tables

Table 1. The list of the interviewees. Source: Author 24

Table 2. Defining human-centric AI: categorized actions aligned with ethical principles mentioned by the experts. Source: Author 55

Table 3. Tools and measures for translating human-centric AI principles into practice. Source: Author..... 58

Table 4. Comparison of tools and measures for translating human-centric AI principles into practice with adoption status in Estonia. Source: Author 61

1 Introduction

In recent years public governance and technology tend to become inseparable. The use of technological tools to improve governance efficiency is becoming more widespread. While some countries are just getting started with e-governance, others have been investing in the technology for decades. Estonia is well known for its pioneering role in the field of e-governance, having over two decades of experience and a strong commitment to sustain innovation. One of the future advancements outlined in Estonia's Digital Agenda 2030 is the development of a government system powered by artificial intelligence (AI) [1]. This includes enhancing the adoption of AI-based solutions to increase the efficiency of the public sector, simplifying communication between citizens and the government, and achieving a higher level of automation in public services. Approximately 120 AI solutions have been implemented in the Estonian public sector within the past five years [2]. However, widespread adoption of AI technology is not only a source of pride, but also places a great responsibility on the state, which seeks to automate its work. The Estonia's Digital Agenda 2030 emphasises the adoption of AI solutions in line with a human-centric approach. These solutions should serve as a means to generate public value and safeguard the fundamental rights of individuals, as well as uphold democracy and the rule of law [1]. The absence of a human-centric approach can result in various forms of discrimination, undermine citizens' trust in public institutions, and violate their rights [3]. Additionally, there are concerns regarding the ability of AI solutions to repeat and strengthen social biases, as well as to change the perception of human role, agency, and self-perception [4]. The absence of ethical principles in AI solutions is a complex issue that can originate from both societal and technical factors [5]. This complexity makes it challenging to pinpoint the exact cause of bias, thereby reducing the likelihood of developing AI tools that are completely centred around human needs.

Despite acknowledging the importance of ethics and a human-centric approach, there is still a lack of agreement on universal AI ethics due to the diversity of cultural and societal norms that define it [6]. Despite the numerous ethical guidelines published by various

organisations to establish a framework for advancing towards more human-centric AI, there remains a gap in understanding how to effectively implement existing ethical principles [7]–[12].

The challenge of transitioning from ethical considerations to practical implementation shouldn't discourage countries from utilising AI or creating tools and practices to ensure the development of AI solutions that prioritise human-centric values. Presently, there is a noticeable tendency towards the establishment of legal frameworks for regulating AI solutions. Legislative measures that establish a framework of acceptable and unacceptable instruments are typically seen as a solution to guarantee the human-centricity of AI. However, while it is difficult to argue that the development of the legal field is an important step towards human-centric approach, this step should be accompanied by the emergence of the common practices, standards, and tools required to support the practical implementation and assessment of human-centric principles in AI solutions. Estonia has developed multiple tools to ensure the human-centeredness of AI, recognising the significance of implementing practical measures to prevent, assess, and eliminate the risks associated with these systems [13], [14]. Nevertheless, there is still a lack of a complex and methodical approach that supports ethics in AI solutions. Many other countries are working on the same problem of inventing practical measures to assist in the development, deployment, and use of AI solutions in a human-centric manner. Therefore, it is probable that certain solutions may be applicable in an Estonian context and utilised to enhance the ethical standards of AI.

The main goal of this research is to learn how countries move from ethics to action when developing, deploying, and using AI tools in public services. So that Estonia's human-centric AI approach can be compared to other countries' approaches in order to make it more complete.

As a first step, the author intends to analyse the perception of the human-centric AI approach. Next, examine the various strategies employed by other countries such as Finland, Sweden, the Netherlands, Belgium, the United Kingdom, and Canada to promote a human-centric AI approach. Lastly, to compare the Estonian approach with approaches employed in mentioned countries, in order to identify present solutions that might support the advancement of the human-centric AI domain in Estonia if implemented.

It is important to note that this work focuses on finding the fine line between protecting fundamental human rights and embracing AI's full potential for public good.

The main research questions of this study are the following:

- RQ1: How do countries define a human-centric AI approach?
- RQ2: How do countries translate human-centric AI approach into life?
- RQ3: How can Estonia's human-centric AI approach be compared to other countries' approaches in order to make it more complete?

The thesis is structured into six primary sections. The initial section provides an introductory overview of the research topic by presenting the background, identifying the problem, and stating the goal of the study. The second section provides an overview of the literature on AI in the public sector, as well as an ethical approach to the AI field. The third section focuses on the research methodology, providing more information on data collection and analysis methods. The fourth section presents the outcomes and findings of this qualitative research. The fifth section is dedicated to the analysis of the findings and offers responses to the three research questions. The final section offers a brief summary of the research.

2 Theoretical framework

The theoretical framework's goal is to set a foundation for the study by providing an overview of the literature on AI in the public sector as well as an ethical approach to the AI field. The purpose of this section of the thesis is to outline the major steps in the rise of AI technology and its subsequent implementation in the public sector; show, using Estonia as an example, how AI can be utilised to achieve public good; explore how ethical principles in several international guidelines outline human-centric AI, as well as why AI ethics are critiqued in order to determine what human-centric approach is missing today and how to address these gaps while translating AI ethics into practice.

2.1 Setting the scene: background to AI

Alan Turing [15] was among the first to propose the concept of an intelligent machine, followed by, John McCarthy in 1956 establishing AI as a new research discipline [16]. Although AI emerged as a field of study in the middle of the twentieth century, a lack of computational power and training data slowed its rapid development and widespread adoption [17]. In 2010, social demand for better services, combined with the availability of technical prerequisites, pushed a large number of large corporations to develop and use AI in their products [18], paving the way for governments eager to adopt new technologies. AI began to be viewed as a tool for improving public service delivery and civic engagement in the second half of 2010 [19]. AI has found application in a variety of public sector domains, including public governance, education, transportation, health, communication, security, and armed forces [20].

While it may appear reasonable from a technological standpoint not to distinguish between the use of AI in the private and public sectors, seeing it as a unified technological development expanding into all areas of life. Meanwhile, from a socioeconomic standpoint, AI implementation places government agencies in a unique position. The goal of the public sector is to provide public good. However, not only does it need to promote well-being for all groups of society, but it also regulates citizens' lives, making public

authorities' decisions important for a large number of people by affecting their rights, interests, and legal status [21]. Taking that consideration, the automation of processes in the public sector should be thoroughly examined in order to eliminate the risks of violations of law, rights, and ethics. It is important to note that when used correctly, AI can improve policymaking decision-making processes and outcomes, improve public service delivery along with the interaction between government and citizens, optimise internal management, and support operational, political, and social public values [21].

2.2 AI in the public sector of Estonia

Estonia follows the European Commission's definition of AI, which defines it as " *systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals*" [22]. However, according to the report of Estonia's AI taskforce, the definition becomes limited in terms of technology, with only narrow AI solutions based on machine learning qualifying as AI [23]. In the context of ethical AI, Estonia employs the term "human-centric AI" [1], [14], [24]. Therefore, this study mainly addresses the concept of human-centric AI rather than AI ethics.

In 2018, Estonia took its initial actions to implement AI in the public sector by establishing the first national AI Strategy [25]. This strategy encompasses various initiatives aimed at promoting the adoption of AI, enhancing skills, fostering research and development, and establishing a legal framework. The primary motivation behind launching the nation's own AI Strategy was to address demographic challenges, increase productivity, and allocate more human resources to value-creating activities rather than routine tasks [23]. After successfully implementing AI in the public sector and acquiring initial expertise, Estonia has released its current strategy, which provides a strategic outlines and action plan for the years 2022-2023 [24]. The first strategy primarily focused on increasing the adoption of AI, while the following strategy featured more initiatives to ensure a human-centric AI approach. The national strategy for 2022-2023 prioritises the development of new AI solutions and the promotion of AI usage in public institutions. Furthermore, it strongly emphasises other actions such as improving digital services through AI, advancing principles that prioritise human well-being and trustworthiness in AI, improving the quality and accessibility of data, and establishing a legal framework to

regulate the development and utilisation of AI in a manner that prioritises human well-being and trustworthiness [24].

By the end of 2023, more than 120 AI solutions and 40 reusable AI components have been implemented in Estonia's public sector. None of the AI solutions developed are used for automated decision-making; however, some solutions are used for decision-making support [2]. It should be noted that the development of AI solutions is of interest not only at the level of individual public institutions, but also at the national level, where could be seen a larger engagement and support for AI adoption. The coalition agreement signed in 2023 fully supports the development of artificial intelligence in the public sector. The Government of the Republic of Estonia [26] wants to 1) *develop the digital state in a user-centric way; enable more service-related and decision-making processes characteristic of the personalised state and the integration of artificial intelligence into data-based decision-making;* 2) *create proactive solutions or solutions which are based on event services;* 3) *promote the widespread use of artificial intelligence and machine learning;* 4) *renew and add to all data-related legislation in order to guarantee a lawful basis for the crossover use and housing of data and for the use of open data and artificial intelligence.*

Regarding the Estonian citizens' viewpoint on the utilisation of AI in the public sector, the Ministry of Economic Affairs and Communications of Estonia conducted a study to gather insights on AI-related matters during the development of a new national AI strategy for the years 2024-2026. The research revealed that individuals are cautious regarding the utilisation of AI, with a significant majority of respondents expressing their lack of support for the implementation of AI in public governance when discussing it in a broad sense [27]. However, survey participants are in favour of adopting AI for delivering public services. They see the automation of repetitive tasks, particularly those that don't require decision-making, as a beneficial trend. Additionally, they recognise the significant potential of AI in fields such as medicine, document management, translation, reporting, and customer service [27].

2.3 Importance of AI ethics

Numerous countries have implemented narrow AI solutions to automate routine bureaucratic processes and improve governance efficiency, as well as to offer personalised and proactive public services [21].

In Estonia, AI solutions are employed in the public sector to advance the field of language technology (e.g., real-life subtitles, translation, voice recognition, anonymisation), making data-based predictions (e.g., identifying the severity of the patient's health condition, forecasting the economic results of companies, hazard assessment), identification purposes (e.g., border control automation, digitalization and classification of archive materials, stroke identification), creation of self-driving cars and robots [2]. All the mentioned use cases assist in accomplishing highly specific tasks with the objective of generating public good.

However, the utilisation of certain AI solutions raises a greater number of ethical concerns compared to others. It is crucial to thoroughly evaluate solutions that directly impact people's lives during the design and implementation stages in order to minimise the potential for undesirable outcomes and eliminate the risks of discrimination and human rights violations. According to the European Parliament [28] certain AI systems have been identified as posing an unacceptable level of risk. For example, in the public sector, AI systems such as real-time biometric identification systems and social scoring solutions should be prohibited.

When discussing the potential dangers of AI, it is important to distinguish between the risks associated with narrow AI systems currently used in the public sector to perform specific tasks, and concerns regarding the unpredictable consequences of developing general AI that possesses human-like cognitive abilities and can apply knowledge across various domains. However, this does not preclude the ethical AI approach from being applied to all current and future AI solutions. The current level of AI capacity and further spread of technology was achieved in a very short timeframe [29], so we must think a few steps ahead to ensure that we maintain the human-centric AI approach while developing new solutions in both the public and private sectors. Not only must we be cautious when upgrading technology, but the simple spread of already used technology and the large number of use cases creates more room for AI incidents [30].

Furthermore, the importance of an ethical approach stems from the fact that AI solutions are embedded in search engines, applications, digital assistants, smart cars, e-commerce,

and many other aspects of our daily lives outside of the public sector [31]. The fact that AI solutions are mostly black boxes raises the ethical question even more. Since AI models are trained using data and desired results, neither the AI system's user nor the AI model itself can explain the reasoning behind a particular decision [32]. Hence, in order for such technology to gain acceptance, it is crucial to establish trust in it and ensure that AI is solidly grounded in principles of human dignity and privacy protection [3]. One method for ensuring human-centric AI development is to establish clear ethical principles as a foundation for future development.

2.4 Ensuring AI ethics through soft and hard law

Establishing clear principles on human-centric AI can be accomplished using both a soft and a hard law approach. Soft law is a non-legislative policy instrument such as guidance or a set of principles, whereas hard law is a legally binding regulation that defines allowed and prohibited measures [33]. Several public and private organisations have formed expert groups and developed guidelines or principles to address emerging challenges in the AI field in recent years. The objective of implementing soft and hard law measures is to encourage "good" or favourable outcomes for people, or at the very least, prevent "bad" unfavourable outcomes for people, caused by AI systems [34].

Furthermore, the human-centric AI approach is inextricably linked with the issue of responsibility, but scientists have differing perspectives on the allocation of agency. Ihde [35] believes that the meaning of technology is determined solely by the contexts in which it is used, and that without the context of use and the user's intentions, technology has no specific value. Van de Poel [36], on the other hand, believes that values are incorporated into technology from the beginning, emphasising that the design of technology comes before the context in which it is used. According to Winner [37], not only does conscious political will leading to the creation of technology have an impact on the technology and the consequences of its adoption, but technology itself, even if created without any political intent, may have politics, influencing the further development of society over time. In this sense, he compares the impact of technology adoption to the law, emphasising the importance of thoughtful innovation creation and adoption.

The question of agency serves to recognise that the purpose of soft and hard law is to facilitate the implementation of a human-centric AI approach, both during the design phase and AI utilisation. Furthermore, the relevance lies in the fact that ethical dilemmas

in the field of AI represent a combination of social and technological challenges that emerge at various stages of AI implementation in the public sector [6].

Soft law can be considered an initial measure to establish a human-centric approach to AI. Various international organisations [4], [38], public entities [22], private entities [39], [40], and professional associations [41] have established guidelines or released ethical principles for AI. Typically, these documents provide a broad overview of ethical principles but do not provide detailed information on specific practical steps that can be implemented immediately [9]. On the contrary, they aid in describing the positive outcomes that AI should produce. As a result, outline the components of a human-centric solution and develop a framework for further discussion and a starting point for human-centric AI development and implementation.

Meanwhile, hard law contributes to the creation of legal certainty. Although it may appear that AI is currently unregulated, this perception is incorrect, because there are several overarching legislative norms directly affecting the AI field, such as fundamental human rights or the General Data Protection Regulation (GDPR) in the European Union (EU), and the purpose of specific AI regulations would be to fill in the legal gaps [22]. The AI Act will serve this purpose within the European Union [42]. Nevertheless, certain countries, such as the USA, opt to refrain from imposing additional regulations on the field of AI and instead adopt soft law measures as opposed to more stringent legal regulations [43].

2.5 Achieving human-centric AI through ethical principles

Guidelines on AI ethics outline various principles that are intended to aid in the development of a human-centric AI approach. According to Hagendorff [9], the guidelines commonly mention accountability, privacy, and fairness as the most prevalent principles. He believes that these principles can be regarded as a fundamental prerequisite for developing what is known as *ethically sound AI*. In addition to these three components, there are numerous other principles that contribute to defining what constitutes a "good" AI system and how to effectively develop, deploy, and utilise it.

The High-Level Expert Group on AI (AI HLEG) set up by the European Commission outlines four ethical principles: *(1) respect for human autonomy, (2) prevention of harm, (3) fairness, (4) explicability;* and seven key requirements for trustworthy AI: *(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 [22].

The Organisation for Economic Co-operation and Development (OECD) Council at Ministerial level have adopted the recommendation on AI, which include following value-based principles: *(1) inclusive growth, sustainable development, and well-being, (2) human-centred values and fairness, (3) transparency and explainability, (4) robustness, security, and safety, (5) accountability* [38].

The United Nations Educational, Scientific and Cultural Organization (UNESCO) has proposed recommendation on the ethics of AI, which contain next ten principles: *(1) proportionality and do no harm, (2) safety and security, (3) Fairness and non-discrimination, (4) sustainability, (5) right to privacy, and data protection, (6) human oversight and determination, (7) Transparency and explainability, (8) responsibility and accountability, (9) awareness and literacy, (10) multi-stakeholder and adaptive governance and collaboration* [4].

Although the organisations recognised that principles can be abstract and may lead to tension, these guidelines are intended to encourage countries to incorporate these principles into their own action plans, strategies, and legislation. In addition, countries must operationalize ethical principles by developing practical instruments and measures to ensure the implementation of ethical AI. This can be challenging due to the abstract nature of ethical values. The previously mentioned principles are applicable to all AI solutions, encompassing various domains and institutions, regardless of whether they belong to the public or private sector. Although the author only provides three examples of guidelines, there are many other documents proposing their own vision on ethical AI principles, such as other guidelines, national strategies, charters, rules of conduct, and others.

The novelty of the AI field, coupled with the rapid dissemination of technology, led to numerous debates on the ethics of AI solutions. The pursuit of more efficient governance presents an opportunity to automate daily bureaucratic procedures, offer personalised services, and improve efficiency. Nevertheless, there is a potential danger of violating fundamental human rights, so it is crucial to recognise the potential consequences that

may arise from the absence of an ethical approach. Therefore, it is imperative to ensure that human-centric principles are integrated during the development of AI solutions and remain solid during the implementation and utilisation stages of AI.

2.6 Critique of AI ethics: lack of practical implementation

The recent increase in the number of ethical guidelines has also fuelled criticism of stipulating ethical principles through soft law documents, claiming that AI ethics in this form lacks practical implication and control mechanisms [9]–[11], [44]. Although experts evaluate AI ethics more critically in the context of the private sector, some of the critical points may also apply to the public sector, especially given that, while the public sector orders and uses AI solutions, the creation of such solutions is frequently outsourced to private companies.

First, AI ethics is criticised for focusing primarily on the description of ethical principles, in other words, answering the question "what is ethical?" rather than addressing "how to make AI solutions ethical?" [12]. Second, while many soft laws are intended to describe the best ways to develop, deploy, and use ethical AI solutions, created guidelines often lack clarity and can be accused of being too high-level, making them difficult to implement in practice [7]–[9]. Furthermore, the use of high-level terms such as safety or privacy is highly context dependent and may have multiple explanations [10]. Plus, ethical principles suffer from a lack of reinforcement. Due to their reliance on soft law, many of these stipulations lack effective mechanisms for control and enforcement [10].

The problem of AI ethics is not limited to the form in which we adopt it. In other words, the problem stems not only from choosing soft law measures over hard law or from a lack of practical implementation, but it extends beyond that. Human-made AI exists in a world ruled by humans, which is by no means ethical. Munn [10] emphasises that technology is influenced by existing practices and structures, whether we mean the culture of the organisation designing the solution or the bias the AI solution may contain due to unfair behaviour of people in the real world. This also raises the issue of a lack of ethical education among workers across all domains [10], so both public servants ordering the solution and private company workers developing it may lack an overall understanding of ethics.

Despite the fact that ethical principles may often generate more questions than they can resolve, and that the world will not achieve perfection, this does not imply that we should

abandon the human-centric approach to AI. Many countries and organisations comply to ethical AI principles and take concrete steps to develop, deploy, and use AI solutions that are human-centric. The following part of the study will concentrate on the manner in which nations delineate a human-centric approach and the concrete actions they undertake to guarantee the development of AI solutions aligns with ethical principles in AI.

3 Methodology

The main goal of this research is to learn how countries move from ethics to action when developing, deploying, and using AI tools in public services, so that Estonia's human-centric AI approach can be compared to other countries' approaches to make it more complete. The main research questions of this work are the following:

RQ1: How countries define human-centric AI approach?
RQ2: How countries translate human-centric AI approach into life?
RQ3: How human-centric AI approach used in Estonia can be compared to other countries approach, with an aim to make it more complete?

To achieve the aim of research author used the qualitative research approach as the main methodology. The purpose of qualitative research is to uncover the true nature of a phenomenon by delving into its core, uncovering its hidden elements, and making them apparent to the public [45]. The data was collected through the use of semi-structured interviews, which follow a predetermined interview format but also offer flexibility in the sequence of questions and provide an opportunity to ask clarifying questions [46]. For data analyses was used qualitative content analysis, the objective of which is to present a summary of the main concepts and findings of the analysed text, focusing on the research questions [47]. However, this method also allows the examination of uncommon or unique phenomena in the text and takes into account the viewpoints expressed by study participants, even if they have no strong connections to the initial topics proposed by the author.

3.1 Data collection methods

The author conducted nine individual semi-structured interviews with public sector representatives from seven countries: Estonia, Finland, Sweden, the Netherlands, Belgium, the United Kingdom, and Canada. Three interviews were conducted with Estonian representatives in the Estonian language, while the remaining six interviews

were conducted in English. The interviews were conducted between November 7th and November 27th, 2023. The interviews had a duration ranging from 30 to 70 minutes.

The author used semi-structured interviews consisting of a set of 15 questions aiming to find out how different countries define and use AI in the public sector; addressing the importance of AI ethics and measures used to support human-centric AI approach by translating ethics into practise; as well as touching upon challenges and interests countries have in the field of human-centric AI (Appendix 2).

Three experts were provided with the questions in advance, while the other participants received the questions only at the time of the interview. The interviews were conducted online through using the Microsoft Teams application. Interviews were conducted remotely to accommodate experts located in different countries or due to time constraints. The experts participated in the research on a voluntary basis and were provided with information regarding the objective and content of the study. The consent to record the interviews was given by eight out of nine participants. During the interviews, experts were asked for their consent to be quoted and referenced in the research. Those interviewees who choose not to remain anonymous disclosed their names, organisation and/or position. To protect the confidentiality of those participants who preferred to maintain some level of anonymity, their roles and organisations are portrayed in a simplified manner.

3.2 Study sample

The study sample included public sector officials from seven countries. The selection of experts was based on their affiliation with the AI field in the public sector, their involvement in the development or implementation of measures related to human-centric AI approach, as well as on recommendations provided by Estonia's public sector officials. First, the author approached three experts from Estonia's public sector. All Estonian experts have a strong connection to the AI field, development of AI solutions in the public sector or/and have direct influence on the course of Estonian public sector politics in terms of AI. The individuals mentioned are Ott Velsberg, the Chief Data Officer from the Ministry of Economic Affairs and Communications; Kristel Kriisa, the AI in the Public Sector Project Manager from the Information System Authority; and Henrik Trasberg, the Legal Advisor on New Technologies from the Ministry of Justice.

As a next step, the author approached Estonia's Chief Data Officer for advice on other potential interviewees from digitally advanced countries using AI solutions in the public sector. The author was given contact information for representatives from Finland, Sweden, the Netherlands, Belgium, the United Kingdom, Canada, New Zealand, South Korea, Singapore, Portugal, Uruguay, and Denmark. The author contacted all representatives, but only experts from Finland, Sweden, the Netherlands, Belgium, the United Kingdom, and Canada consented to the interview taking place between November 7th and November 27th, 2023. Other experts who received the inquiry and provided responses, but we have not proceeded with the interview, include experts from New Zealand, South Korea, Singapore, Portugal, and Uruguay. Expert from Denmark was contacted but did not provide a response. A list of the interviewees participated in the researched is brough in the Table 1.

Table 1. The list of the interviewees. Source: Author

ID	Country	Position	Organisation
E1	Estonia	Chief Data Officer	Ministry of Economic Affairs and Communications
E2	Estonia	AI in the Public Sector Project Manager	Information System Authority
E3	Estonia	Legal Advisor on New Technologies	Ministry of Justice
E4	Finland	Director of Digital Engagement & Customer Experience	Finnish Digital Agency
E5	Sweden	Trend Analyst	Agency for Digital Governance
E6	Netherlands	Policy Officer	Ministry of the Interior and Kingdom Relations of Netherlands
E7	Belgium	Expert	Public sector of Belgium
E8	United Kingdom	Public Official	Centre for Data Ethics and Innovation
E9	Canada	Senior Official	Federal Government of Canada

3.3 Data analyses methods

A total of eight interviews out of nine, were transcribed. The speech recognition tool *Tekstiks* was used to transcribe the interviews conducted in Estonian [48]. Google Recorder was used to transcribe the five interviews that were conducted in English. As a next step, each transcription was carefully reviewed to guarantee that it corresponded precisely to the recordings. On one of the interviews, the author was only permitted to take notes, which were later used in analyses.

During the interviews, the author noticed several recurring topics, which were later confirmed through qualitative content analysis of the transcriptions. These topics were then evaluated based on their relevance to the research questions. As a result, the author identified three main topics and ten subtopics:

Defining human-centric AI

- Defining AI
- Use of AI in the public sector
- Defining human-centric AI
- Interest towards human-centric AI approach

Translating human-centric AI approach into life

- Regulations
- Practical measures and tools
- Other measures
- Challenges and interest

Human-centric AI approach in Estonia

- Human-centric AI measures
- Potential development

3.4 Limitations

There are several limitations to the study that was conducted. In the first place, the research is strictly directed to the AI solutions used in public sector and describes only tools and measures which are used for ensuring human-centric AI approach in the public sector. Second, research is restricted to the tools and measures that were exclusively

discussed during the interviews. Consequently, the range of tools and measures in practice may be more extensive, and solutions that were mentioned during the interviews may also be implemented in countries where the experts did not mention them. Additionally, even though the research's experts are highly knowledgeable about both AI and human-centric practices, their expertise may not be comprehensive due to the size of the participating countries' public sector, the decentralisation of AI use in public sectors, and the decentralised development of tools and measures that ensure the use of AI in a human-centric manner.

Third, the study's relatively small sample size, consisting of only six foreign experts and three experts from Estonia, further hinders the achievement of an in-depth examination of all potential solutions existing today. While the research author aimed to obtain a comprehensive understanding of the tools and measures employed to ensure the practical implementation of human-centric AI by focusing on a large number of countries, the final sample consists primarily of European nations. As a result, it cannot include practices in other regions of the world that may vary from those of the countries where the interviewed experts were situated.

Fourth, because AI is a rapidly evolving field, the described views and practices may also be subject to rapid change.

Fifth, the opinions expressed by the experts in this work might reflect their personal perspectives on the subject of human-centric AI. Thus, such perspectives cannot be portrayed as the official position of their organisation or the nations they represent.

Lastly, the study's results offer a comprehensive summary of the findings and may incorporate the author's subjective opinion, despite the absence of any deliberate bias.

4 Research outcomes and findings

Interviews were conducted to collect data on the implementation of AI solutions in the public sector and the understanding of human-centric AI practices in Estonia and other countries. Initially, the author conducted interviews to gather a broad viewpoint regarding the utilisation of AI in the public sector. This involved gaining an understanding of the AI concept and the underlying motivations for its implementation. Furthermore, interviews were conducted to gather data related to the human-centric AI approach. This included collecting explanations of the human-centric AI approach within various countries, identifying risks associated with AI implementation in the public sector, assessing interest in the human-centric approach, and uncovering the legal, practical, and other measures employed by countries to ensure that AI solutions prioritise human interests. Furthermore, the author conducted interviews to acquire insights into the utilisation of AI in the public sector of Estonia. The objective was to obtain a comprehensive understanding of both the existing practices and the areas where improvements are needed to ensure a human-centric approach to AI. This section specifically addresses the data obtained during the interview process.

4.1 Defining human-centric AI approach

This section addresses the first research question, "RQ1: How do countries define a human-centric artificial intelligence approach?" In order to address the research question, this section also discusses a definition of AI and the use of AI in the public sector, as a prerequisite for addressing the topic of a human-centric AI approach.

4.1.1 Defining AI

Properly defining the scope of what can be classified as AI is important in developing regulations that prioritise a human-centric approach to AI. This is because it enables the identification of potential risks and supports the implementation of measures for specific

algorithmic tools that fall under the AI category. Experts claim that reaching a consensus on a common understanding of AI is exceedingly challenging, even within a single country (E6, E8). AI is often perceived as a collection of algorithms, but it is important to note that not all algorithms can be classified as AI (E6). AI is often perceived as an advanced set of algorithms that possess a certain level of autonomy or a more complex way of functioning (E4, E6).

"Artificial intelligence is… I think we can say we agree to a certain extent these are more complex algorithms and what makes them more complex is that they tend to have a certain degree of autonomy, or they tend to be more complex in how they work as opposed to simple "if-when" rule base algorithms." *(E6 - NL)*

The definition of AI is also determined by the description of the underlying technology it relies on. During the interviews, experts discussed technologies such as machine learning (E1, E4) and robotic process automation (E4).

Furthermore, countries rely on AI definitions established by international organisations for example the OECD. They also indicate that they are awaiting the AI Act to determine what qualifies as AI (E7).

Nevertheless, considering the focus on human-centric approaches, some experts suggest that it may be necessary to expand the range and ensure that human-centric practices are applicable to all algorithmic solutions that have the ability to impact the lives of people. Basic algorithms can have a significant impact on people through relatively simple automation (E3, E6). The AI Act definition is subject to critique for this reason as well (E6).

"The scope of AI Act doesn't address all the challenges with all the algorithms that we have. If we look at impact, a simple algorithm can have a major impact on rights of citizens, so therefore, we think it's very important to also focus on more simple algorithms. And we also see a lot of problems especially in government, where more simple algorithms are used." (E6 - NL)

The absence of a clear definition of AI is also perceived as an obstacle to the adoption of human-centric practices or gaining a comprehensive overview of the AI field in the country (E6, E8). This is because it becomes more difficult to identify the solutions that can be classified as AI and consequently apply human-centric practices.

"They have a site argument - what is AI, right? Because they've used complex data analytical systems for many years and at what point you draw the line to say something is AI." (E8 - GB)

"It [having full overview of all algorithms used in country] is really hard. Where to start? What's the definition of an algorithm?" (E6 - NL)

4.1.2 Use of AI in the public sector

AI solutions in the public sector are utilised to enhance both internal organisational processes and services provided to citizens. AI solutions are employed internally to enhance the efficiency of the public sector (E1, E6, E8, E9) and aid in performing repetitive tasks, thereby allowing public servants to allocate more time towards responsibilities that are not easily automated (E1, E2, E6, E8, E9). AI solutions are employed to preserve financial resources (E3) and accomplish tasks that require more human resources than the public sector possesses (E3, E5).

" Of course, the primary objective is simply to save money more efficiently, we can just do everything faster and cheaper, and given that in Estonia we have very and very limited resources, just as we have limited human resources and financial resources, we just need to automate" (E3 - EE)

From the perspective of the citizen, AI is utilised to enhance the quality of public services provided to citizens (E1, E7), optimise service delivery by making intelligent decisions that benefit citizens the most (E1), and enable more efficient engagement with the government (E7). Furthermore, AI can enhance the provision of personalised services (E8) and assist in delivering proactive services (E1).

"So, that a person can get more benefits. In addition to this, in fact, in order to make more informed decisions, liker based on forecasts… Of course, another important aspect is actually improving the quality of public services. Even just taking the idea that a person does not need to know which government agency to contact in order to receive a service, services can be provided proactively." (E1 - EE)

Experts from Estonia have noted that the motivation to adopt AI solutions might be driven by the increasing popularity of AI as a topic, as well as a desire to maintain innovation (E2). Furthermore, the integration of AI in public services can be seen as a next step in automating the public sector, especially in digitally advanced countries (E3).

"AI is just a natural next step of the automation that we have had for decades." (E3 - EE)

Nevertheless, an expert from the Netherlands expressed an alternative viewpoint, emphasising that the utilisation of AI in the public sector should not be pursued as a goal in itself.

"You shouldn't use algorithms as a full per se, it really should be a tool, not a goal." (E6 - NL)

AI has a wide range of applications in various areas of the public sector (E2, E6). The mentioned domains include service operations such as automated response to customers (E4) and chatbots (E1, E8), as well as healthcare (E6), mobility (E6), tax services (E6), fraud detection (E7), and border control (E7, E9).

The majority of the countries included in the research lack comprehensive overview of all AI applications used by their public sectors (E4, E5, E6, E7). Experts have identified several reasons that make it challenging to obtain a full picture. Complications in certain countries arise from the decentralised structure of their public sector (E5, E6, E7). As more authority is given to the federal level, the landscape becomes increasingly fragmented. Decentralisation is evident even within large public institutions, where there may be a lack of oversight regarding the AI solutions employed by organisations itself (E6). Mapping all AI solutions can be particularly challenging in digitally advanced countries with a large number of AI use cases (E6). Other challenges may arise due to the resistance of organisations to disclose information regarding the use of AI (E6). In addition, certain countries may adopt an approach that do not have a specific focus on AI use-cases. For instance, Sweden primarily focuses on monitoring big data analyses (E5). Moreover, obtaining a comprehensive overview is difficult because it requires an understanding of how AI solutions are implemented and utilised within institutions (E5).

"So, what we have seen is a potential increase in the use of big data analysis and AI implementations, but we don't have granular data to see if these are own developments if they are bought solution, shelf solutions from private sector. And we are not sure how they're being implemented" (E5-SE)

The majority of experts shared a wish to obtain a comprehensive overview of AI solutions utilised in their public sectors. Hence, in their countries were launched projects with the

objective of mapping the current AI solutions (E5, E6, E7). However, experts notice that the task is quit challenging.

"[…] Each time, it's like a piece of the puzzle. You get to see this piece or this piece, but no one has puzzle together yet. So that is a problem." (E5-SE)

4.1.3 Defining human-centric AI

In Estonia, the concept of "human-centric AI" is used to refer to AI systems that follow ethical principles [14]. The human-centric AI approach in Sweden and Belgium is referred to as "AI for good" (E5, E7). Canada has the term "Responsible use of AI" (E9), while the United Kingdom names human-centric approach as "Responsible innovation" or "ethics" (E8).

"I don't think the kind of term "human-centred AI" is one that's commonly used, but I think responsible innovation or ethics is how we generally how we refer to the idea. I know that it is slightly different, but that sort of more what people think about." (E8 - GB)

According to experts, the terms associated with AI ethics are overly broad. They admit that it can be challenging to determine the precise definition of these broad terms, particularly when considering the practical implementation of a human-centric approach to AI (E5, E6).

"It can be an umbrella term. People can have different ideas what it means. We tend to use the term "Responsible AI", quite a lot also in our policy papers and communicating with the citizens. But what it really means?" (E6 - NL)

"I think it's quite hard to define what we mean by human-centric in this area. Like I said before, we have this general approach to not to damage or not to harm, so "AI for good". I think it's being used by many private companies. We have similar approach to it, but it's very hard to codify what does that mean in the implementation phase." (E5-SE)

The human-centric approach to AI gets described through international principles, regulations, and concrete actions. From the international guidelines experts have mentioned several ethical principles, including transparency (E1, E3, E5, E6, E7, E8, E9), explainability (E6, E7, E8), accountability (E5, E8, E9), robustness (E5, E8), fairness (E6, E9), safety (E6, E8), security (E8), responsibility (E6), trustworthiness (E1), and

accessibility (E3). The principle of transparency was mentioned most frequently, expert from Estonia defined its importance in the following way:

"Transparency is actually a prerequisite for us to understand these systems. And understanding, in turn, is a necessary condition so that we can actually notice if something is wrong with the system. Beyond this, transparency is also a value in itself. After all, if a state uses artificial intelligence systems, the state, in fact, should have an obligation to be transparent in order to be able to explain to a person how a particular decision was made" (3 - EE)

However, when it comes to the direct correlation between the emergence of real-life actions and high-level ethical principles, some experts argue that it is difficult to determine which came first (E6).

"It's a bit of a chicken and egg story" (E6 - NL)

Multiple countries suggested that the concept of a human-centric approach can be defined through international frameworks and principles (E1) such as AI HLEG guidelines (E1)[22], UNESCO guidelines (E5)[4], OECD principles (E8) [38], and UN sustainable development goals (E7). As well as in relation to justice (E6), human rights (E3, E7, E9), and democratic values (E5). Furthermore, experts proposed specific measures that they believe promote a human-centric or ethical approach to AI. These measures include: respecting individual autonomy (E1, E3, E5), providing efficient public services (E1, E4, E5), reducing bias in AI solutions (E3, E5), following legislative frameworks during AI development and use (E4, E7), protecting personal data (E1, E5), involving humans in the decision-making process (E6, E7), granting individuals the right to object AI decisions (E3), informing people about the use of AI (E1), creating AI solutions that are ethical by design (E7), protecting integrity (E5), considering people's needs (E1), maintaining trust in the public sector (E5), conducting risk assessments (E1), offering inclusive services (E3), and adopting a life event-based approach (E5).

"Human-centric means that the service is implemented, so to speak, putting the needs of the person in the first place." (E1 - EE)

4.1.4 Interest towards human-centric AI approach

Experts from the countries involved in the study reported a significant level of interest in the topic of human-centric AI within their respective public sectors (E1, E2, E4, E6, E7, E8). Discussions among policymakers were started by the rapid development of AI and the need for utilizing a human-centric AI approach (E2). Part of the discussions led to the development of practical measures (E4) or triggered a debate on potential practical tools for promoting human-centric values in AI (E6). The topic generates interest not only at the governmental level, but also among regions in more decentralised countries (E7). In certain countries human-centric AI approach gets incorporated in the strategies (E1, E6) or government programmes (E4). However, experts note that in some cases, there may be a greater focus on theoretical interest rather than a desire to put efforts into real-world applications (E1, E3).

While the topic of human-centricity is recognised as being important, certain countries tend to apply it more extensively beyond just AI solutions (E5, E8).

"The so-called "human-centric approach to AI in the public sector", like I said before, this is not specifically developed for AI, this could be for anything you want to do in the public sector. So, for example, I don't know, if you were to implement a policy tomorrow where everything must run on green energy for example, then again, the same human-centric points would be made that green energy is not allowed to harm democracy or trust in the system or so on. So, this is not specifically for AI." (E5-SE)

The human-centric AI approach is attracting attention for various reasons. AI is currently experiencing the peak of the hype cycle (E1, E4), which in turn generates significant public interest in the subject of AI (E6). According to experts, there has been significant interest in AI this year, which is indicated by the amount of AI-related content published by various media platforms (E4, E6). In addition, regular citizens had the opportunity to engage with AI systems (E6, E7), especially through the use of ChatGPT (E6, E8). By gaining direct experience, AI became more tangible for general public (E6), leading to a better understanding of AI concept (E8). The popularity of AI-related events is another indication of the high level of interest in this field (E6, E8) .

"What we did, we had different types of sessions to collect input and one of those types of sessions was citizen session and yeah, the number of citizens that came to these sessions to

share their thoughts, to ask questions, to address serious concerns that they had, I think was very impressive. So that shows you empirically that there is more interest." (E6 - NL)

According to the Finnish expert, AI even have helped in initiating a broader discussion on the ethics of the public sector:

"From my point of view, there has not been ethical questions and discussion regarding an ethical behaviour before AI. So, I think that AI has brought up this ethical aspect into our everyday life. So, in that sense, I think it's very positive that everybody is talking about the ethical questions. Before I haven't witnessed any questions or discussion regarding the ethics of different public sector activities." (E4 - FI)

Another factor that contributes to the interest in the human-centric AI topic is the potential risks associated with AI implementation and the long-term consequences of AI utilisation.

„Let's just say that if you look at international studies, the more any area has a direct impact on people, the greater the potential risks and consequences. And the more, in fact, these consequences, so to speak, get thought through and evaluated much more carefully." (E1 - EE)

The citizens are highly concerned about AI due to the potential job losses and the lack of transparency of AI solutions (E1). At the same time, the Finnish expert thinks that people in the Nordic and Baltic countries show higher curiosity towards new technology rather than fearfulness (E4).

According to experts, the public sector recognises the risks associated with artificial intelligence. An expert from the Netherlands has observed a shift from an overly optimistic approach towards rapidly developing data-based technologies to a more realistic and cautious approach (E6). The growing interest in addressing long-term existential risks (E8) can be seen as a reaction to innovation (E6, E8). Furthermore, significant influence on the development of human-centric AI practices have AI incidents. According to experts from Sweden, the Netherlands, and Belgium, AI incidents have occurred in the countries they represent (E5, E6, E7). In addition, the risks associated with AI are also being discussed in relation to the rapidly advancing private sector solutions (E6, E7, E9) and the influence from non-European Union countries (E6).

While acknowledging that addressing the risks associated with AI is a key motivator for adopting a human-centric AI approach, certain experts warn that concentrating only on the risks could hinder innovation (E2). Therefore, the public sector should try to find a

balance between regulation and innovation (E2, E3, E8). At the same time, it's critical to remember that innovation can only fully benefit you if it's executed responsibly (E6).

"So, we should not forget the potential that it has, but I think if you really want to profit from the full-scale potential, then you really need to do it responsible. Otherwise, it will not be fully responsible and… Uh, the chance is that it will only benefit small group of societies. Don't think only about backside, but not forget about the responsible way of AI implementation." (E6 - NL)

4.2 Translating human-centric AI approach into life

This section addresses the second research question, "RQ2: How do countries translate human-centric AI approach into practice?" This section presents information on the legal, practical, and other measures that contribute to ensuring a human-centric AI approach in the public sector of Finland, Sweden, The Netherlands, Belgium, The United Kingdom, and Canada.

4.2.1 Regulations

One of strategies for promoting human-centric values in AI solutions utilised by the public sector is the application of hard law. The legal measures mentioned by the experts fall into four categories: laws that are already adopted and directly apply to AI solutions (E9); laws that are not yet adopted but directly apply to AI solutions (E2, E3, E4, E5, E6, E7); laws that are technology neutral (E1, E8); and laws that have an impact on certain fields that have a close correlation with AI (E1, E3, E5, E6, E7), like personal data field or human rights. The study was carried out involving delegates from both EU and non-EU countries to explore the differences between the regulatory requirements introduced in EU countries and the approaches taken by the United Kingdom and Canada. Representatives from EU member states (E2, E3, E4, E5, E6, E7) pointed out that the primary legislation addressing AI in the EU, also referred to as the Artificial Intelligence Act or AI Act [28] is yet to come. EU countries expect the implementation of the AI Act to establish the legal framework for AI solutions (E2, E3, E4, E5, E6, E7). Despite some concerns that the AI Act's scope does not fully address the issues with all the algorithms that they currently use, EU countries hope that it will help to define human-centric principles (E6).

"Well, of course we are all waiting that The European Commission to publish the AI Act" (E4 - FI)

"We are waiting the AI Act for guidance" (E5-SE)

The second regulation mentioned by EU countries (E1, E5, E6) is the General Data Protection Regulation or GDPR [49]. This regulation safeguards individuals when their data is being processed by both the private and public sectors, and also empowers individuals to maintain greater control over their personal data. While the GDPR does not directly address all aspects of the development, deployment, and use of AI solutions in a human-centric manner, it is already implemented in EU countries and helps to cover various data-related aspects that are relevant to the creation and implementation of AI in the public sector.

"We have the GDPR rulings about data protection, like all EU countries. […] A lot of what considered ethical use of AI, I would say in Sweden it comes under GDPR protection, the privacy and personal data, and security aspects of that." (E5-SE)

The expert from the Netherlands emphasises the significance of implementing AI-related legal measures that have a global reach and incorporating control mechanisms:

"What is also difficult when it comes to AI, it's always a matter of... not always, but a lot of times it's a matter of different regulators, because AI is not something that sticks to boundaries, it goes over domains, cuts through domains, sectors and countries." (E6 - NL)

One recognised framework found in many countries is the International Bill of Human Rights (E3, E7) [50]. This framework includes a number of rights such as freedom from discrimination, equality between men and women, right to privacy, and other freedoms that may be relevant to the human-centric application of AI in the public sector. All countries involved in the research have ratified at least of 13 out of 18 International Human Rights Treaties [51], indicating that they are committed to the legal responsibilities outlined in the International Bill of Human Rights.

Both the United Kingdom and Canada, as non-European Union countries, have their own legislation that establishes a framework for AI, although this legislation may not explicitly focus on AI as an independent field. As a former member of the European Union, the United Kingdom has incorporated the GDPR into its own laws through the adoption of The Data Protection Act [52]. This act establishes the principles of data protection that

must be followed by anyone who handles personal data. The British expert emphasised that although AI is not currently under any specific legal regulations, the possibility of establishing laws related to AI cannot be dismissed, particularly given the rapid developments in the field of AI (E8).

Canada does not regulate specifically for AI, however, it has laws that could be applied to this field (E9). Canada has recently introduced the Artificial Intelligence and Data Act (E9), which aims to establish the necessary framework for the ethical development, construction, and application of AI systems that have an impact on the daily lives of Canadians [53]. However, the legislation specifically targets AI solutions in the private sector (E9).

Experts also highlighted a few problems associated with the implementation of strict legal practices. They highlighted the difficulty of predicting the risks associated with the rapid development of AI technology (E8). Moreover, there is a need for a more clear framework that addresses the human-centric AI approach (E3). In addition, experts mention that courts require time to establish procedures that arise in response to the utilisation of AI by the public sector (E5).

*"But it (AI incident) went to court and so far, as far as I know, this is only decision we have to see how ready we are for these kinds of problems. And the answer is we're not ready at all, the court decided that it's the city *** did not have to disclose the algorithm as evidence, let's say to be surveyed by the neutral third party. [...] So it would have been a very useful case if it come further, but because the judicial system didn't know how to handle a city being sued for the use of an algorithm, how to handle the algorithm part- what is it and who is responsible, it's been dismissed, as far as I know. But that's the only indication we have so far and it's not a very good one unfortunately." (E5-SE)*

4.2.2 Practical tools

The author categorised practical tools for ensuring a human-centric AI approach as measures that involve active participation of the organisations developing or utilising AI solutions, with the aim of promoting human-centric values in AI. The measures discussed in this section involve the tools that public sector institutions can or are required to utilise during or after the development of an AI solution in order to maintain a focus on human-centered principles. The author identified seven categories of practical measures employed by the countries involved in the study: human-centric AI design practices (E1,

E2, E5, E9), institutions and facilities associated with AI (E1, E4, E5, E6, E8), privacy-enhancing technologies (PET) (E1, E8), regulatory sandboxes (E2, E5), AI assessment tools (E1, E5, E6, E9), AI transparency standard (E1, E8), and AI registers (E1, E5, E6).

Human-centric AI design practices (E1, E2, E5, E9) define how countries are approaching the development of AI solutions by focusing on the needs of their citizens and incorporating human-centric values during the creation phase of the public services. This approach has the potential to cover all public services and can also be extended to AI solutions. Service design that incorporates the experiences people have is an essential component of Sweden's strategy (E5). Hence, they are preparing to launch a project with the aim of exchanging practices centred around people within Nordic countries (E5).

"There's a focus on human-centric design of services. [...] Tomorrow (in the middle of November 2023), I think we are sending our application together with Finland and Norway to have a project at the Nordic Council for human-centric design public services, where we can learn from each other. Hopefully this will be a way for Sweden to start this process of showing the benefits for the citizens, if we have services designed from the point of view of citizens." (E5-SE)

In Canada, the human-centric approach to AI involves the need to test AI solutions for bias both during their development and after their implementation. This is done to ensure that the solutions have the least amount of bias possible (E9). Furthermore, Canada employs the Gender-based Analysis Plus tool to assess the potential risks of AI initiatives (E9). This tool helps to identify the risks of initiatives by providing a framework for contextualizing a variety of human characteristics such as sex, gender, race, ethnicity, religion, age, and mental or physical impairment to guarantee that these characteristics do not impede success and inclusion [54].

The institutions and facilities associated with AI (E1, E4, E5, E6, E8) mainly serve as support systems for organisations wanting to adopt AI solutions. The institutions mentioned during the interviews had diverse functions, including the publication of guidelines on human-centric AI (E4, E5), provision of information on AI-related topics offering AI training and testing facilities (E5, E6), and providing assistance in the development of AI solutions (E2, E5, E8).

Finland's Digital and Population Data Services Agency serves as both an authority that advises other agencies in the field of AI and provides guidelines on AI practices that

prioritise human needs (E4) [55]. Sweden lacks a centralised government institution to fulfil the role of a national competency centre (E5). However, they have a national centre for applied AI in Sweden that operates as a hybrid of private and public entities [56]. This centre is responsible for managing the national data platforms and publishing AI guidelines. The United Kingdom established The Centre for Data Ethics and Innovation (CDEI) with the goal to safeguard the public and core values while promoting safe and ethical innovation and investment in AI technology (E8)[57].

"The CDEI was set up in 2018 and then continued to play an important role across government, on practical implementation and thinking through how you go further than the legal minimums, when you're working on AI products" (E8 - GB)

The Netherlands intend to establish several institutions dedicated to supporting the advancement of AI in both technical and ethical aspects. An institution of this kind would serve as an advisory council, offering feedback to policy makers on innovations in AI and providing assistance to government organisations dealing with AI-related matters (E6). The "National Test Facility for AI" is an organisation that would provide computational power for the creation and testing of AI models (E6).

The United Kingdom implements **privacy-enhancing technologies (PET)** (E1, E8). PET are digital solutions that enable to share and use data in a way that protects confidentiality. These technologies allow the development of AI and data-driven solutions with a focus on human-centric approach. Privacy-enhancing technologies offer both decent access to data and help to respect individual privacy. The United Kingdom launched a programme called the "Responsible Data Access Work Programme" to encourage the ethical utilisation of data, with a specific a focus on the PET (E8) [58].

"Technologies need to use a lot of data. How can we create the right mechanisms to ensure that data is accessed in a responsible way? So what privacy or protection technologies can you use to make sure that data can be shared between organisations, that need it to create AI, but do that in a way that protects individual privacy and meets data protection requirements. […] We have Responsible data access program and majority of the work there is on PET which are the technologies which enable people to share and use data in a way that protects confidentiality" (E8 - GB)

The United Kingdom included PET into their portfolio of AI assurance techniques, which consists of a number of methods used throughout the country to support the development of trustworthy AI [59].

Regulatory sandboxes (E2, E5) serve as experimentation tools that provide a regulated environment for testing and scaling up AI algorithms. The expert from Sweden highlighted that the development of AI solutions is slowed by a lack of public agencies having the technical capacity to create their own AI models within the organisation mainly due to a lack of infrastructure. Furthermore, the exchange of data between agencies is legally prohibited. In order to combine data and train an algorithm for the development of AI, it is necessary to set up regulatory sandboxes and supporting infrastructure that allows legal experimentation (E5).

AI assessment tools (E5, E6, E9) are currently being developed in Belgium and are already in use in Canada, Sweden, and the Netherlands. Assessment tools offer institutions creating AI solutions the chance to assess their own systems and ensure they adhere ethical standards. Canada has created the Algorithmic Impact Assessment (E9), which is a mandatory tool in the form of a questionnaire. This questionnaire is designed to align with administrative law, ethics, and Canadian policy regarding automated decision-making. The tool aims to assist agencies and departments to increase their understanding and management of the risks associated with automated decision systems [60]. The Algorithmic Impact Assessment serves as a self-evaluation tool for institutions, providing them with a score that reflects the impact and risk levels of their AI solutions. Sweden uses a trust Model for AI assessment (E5):

> *"The trust model is I think the best thing that we have done so far this area. So, the trust model is a self-evaluation tool. It follows sort of the same headings as the AI guide - the data management, ethical considerations, legal considerations and so on. So, there are questions related to ethical AI - Have you thought about this? Have you done this? How have you tackled this program? How did you ensure that you don't have bias? Who is responsible for this dataset? It has these very specific questions and then you can answer 'Yes' or 'No' or you can also free text." (E5-SE)*

The Swedish assessment tool aims to proactively address the potential evaluation procedures associated with the implementation of the AI Act. The current version of the assessment is still unfinished. According to a Swedish expert, in order for this tool to be

truly valuable, it should be integrated with the risk analysis tool and also include automatic feedback and recommendations (E5).

The Netherlands have developed a tool called "AI Impact Assessment" to help with the development of responsible AI projects (E6). Assessment is used to promote discussions regarding AI systems. It examines issues related to data, systems, and algorithms while keeping to relevant rules and regulations. It is a tool for discussion and recording thought processes, which promotes accountability, quality, and consistency [61]. The use of solutions in Canada and the Netherlands is mandatory for public institutions, whereas the Swedish solution is not enforced. However, the Canadian approach focuses on evaluating the risks of solutions that have already been developed, whereas the assessment methods employed by Sweden and the Netherlands are designed to be used during the development phase of the AI project.

Belgium is currently in the phase of developing their AI assessment tool (E7). Belgium adopted the AI HLEG [22] guidelines as a foundation for their assessments and intends to develop a user-friendly application consisting of 140 questions. This app will function as a self-assessment tool to evaluate compliance to a human-centric approach. Additionally, users will have the option to view their final score on a radar chart containing various ethical indicators (E7).

AI transparency standard (E1, E8) aims to enhance transparency in the public sector's use of AI tools by providing full descriptions of the tools and detailed information on their application. The main difference between the AI transparency standard and AI assessment tools lies in their final goals. The transparency standard is not designed to evaluate the solution itself, rather, its purpose is to provide the public with complete and understandable information regarding the AI solution used in the public sector. The United Kingdom has developed the Algorithmic Transparency Recording Standard to be open about the AI tools and algorithm-assisted decisions [62].

"It aims to translate the principle of transparency into practice. It particularly focuses on public sector organisations making decisions using AI and data-driven technologies in ways, that impact the public. So, the public is owed information about how these decisions are made, but it also kind of creating a mechanism which forces people to think about how they use those tools. What process do they have in development to make sure that they are mitigating risks? […] It's been iterated, and it's been tested around the public sector in the UK through piloting and that's all available online." (E8 - GB)

AI registers (E6) serve as central repositories which create the collection of information on AI solutions employed in the public sector, aiming to enhance transparency towards citizens. The Netherlands launched The Algorithm Register of the Dutch government in 2022 to support responsible algorithm use by making publicly available information on the algorithms used in the public sector (E6) [63]. Although the use of the register is not mandatory yet, public authorities have already published information on 254 algorithmic use-cases. A Dutch expert highlights that the register includes not only AI solutions, but also basic algorithms (E6).

"We also have a so-called algorithm register. It also includes simple algorithm, so if your scope is really AI, then not all algorithms that are in there are AI. But I think it's worth mentioning because that tells that we're already working on transparency regarding AI for quite a while. Algorithm register really has some, interesting fields that give you information on what variables are used in a model, what type of model is it, are there any sort of legal steps being taken, for example, a the GDPR or a human rights impact assessment. There is more about the legal ethical considerations, but also technical." (E6 - NL)

4.2.3 Other measures

This section includes measures that can be characterised as supportive measures that promote human-centric AI approach but are not directly related to legal regulations or practical tools used by public institutions in the field of AI. Additional measures represent various strategies and policies (E1, E2, E6, E7, E8, E9), guidelines (E1, E2, E4, E5), educational programmes for public sector workers (E5, E7), events (E4, E5, E8), higher education programmes and an AI oath (E7), research (E7, E8), civic engagement (E1, E2, E5, E6, E7, E8, E9), as well as control mechanisms within the public sector combined with public sector work ethics (E1, E2, E3, E5, E6, E8, E9).

Nearly all of the experts mentioned **strategies and policies** as a means of promoting the nation's human-centric AI approach (E1, E2, E4, E6, E7, E8, E9). The Netherlands have the Value-Driven Digitalisation Work Agenda [64], Belgium has implemented the national plan for the advancement of artificial intelligence [65] and is currently formulating guidelines for the responsible use of AI, while the United Kingdom has produced multiple documents such as the AI white paper [66], AI strategy [67], and the model for responsible innovation. Canada created the Directive on policy automation to guarantee that AI is utilised in a manner that upholds the fundamental principles of

administrative law, including accountability, transparency, procedural fairness, and legality [68].

Guidelines (E1, E2, E4, E5) are used to offer detailed recommendations for the application of AI in the public sector. Finland recently released its first guide for developers of digital services, providing guidance on responsible use of AI [69]. Similarly, Sweden has developed a national AI guide called 'Offentlig AI' [70]. The guideline, which is based on the principles of the EU High-Level Expert Group on Artificial Intelligence, addresses various aspects of AI, including data management, ethical principles, organisational issues, definitions of AI, and various technical applications of AI (E5). Canada has developed the Guide on the utilisation of Generative AI [71] in an attempt to provide some classification how generative AI can be used and help to understand the risks related to it (E9).

Educational programmes (E5, E7) were mentioned by the experts from Sweden and Belgium. Sweden created courses in human-centric studies in collaboration with widely recognised professors (E5). Belgium has recently introduced a course on AI discrimination that was jointly developed with the Council of Europe (E7).

Countries organise **events** on topics related to human-centric AI in addition to their educational programmes (E4, E5, E8). The United Kingdom recently organised the AI safety Summit (E8). The purpose of the Summit was to address the risks AI may pose and discuss their mitigation options [72]. Finland and Sweden organise events that promote AI guidelines prioritising a human-centric approach (E4, E5).

Although not specifically related to the public sector, **the higher education programmes and AI oath** mentioned by Belgium were added to the list of other human-centric AI practices. These programmes promote the broad acceptance of human-centric AI and educate specialists on the ethical considerations associated with AI. The Urban Engaged University in Brussels provides a postgraduate programme called "AI for the common good" that covers the holistic perspective on AI important for responsible digital transformation [73]. Furthermore, Belgium aims to implement an oath for students pursuing AI studies, similar to the Hippocratic Oath followed by medical students (E7).

Research (E1, E5, E8, E9) with an aim to form a better understanding of human-centric AI field, exploring the tools for human-centric practices or understanding a public viewpoint can be a strong mechanism to support development of the human-centric AI field. Sweden conducts a survey called "The Internet Citizen" to gather citizens'

perspectives on various aspects related to the digital transformation of society (E5). The United Kingdom conducts an annual tracker survey with the aim of comprehending the public's attitudes towards data and AI [74].

"Every year they publish a tracker survey where they ask the same set of questions, so year-on-year, you can see how people's understandings of AI, how people's fears about use of data, use of AI and decision-making changes. They also have some questions about different AI use-cases and what kind of risks and mitigations people would find valuable" (8 - GB)

Civic engagement (E1, E2, E5, E6, E7, E8, E9) or individual and communal acts aimed at identifying and addressing issues of public concern might be an important instrument for development of human-centric AI approach. Direct involvement of citizens in discussion related to AI use in the public sector happens not so often. In Canada, citizens were involved in the development of The Directive on Automated Decision-Making. Academics, representatives of unions, and international cooperations were invited to provide feedback for the document (E9). The national AI Coalition in the Netherlands has initiated a project called "AI parade" with the objective of increasing awareness and encouraging a dialogue about AI (E6). The main platform for the "AI parade" is provided by public libraries. In addition, Sweden also uses public libraries as venues for engaging in discussions related to AI subjects (E5).

"I would say the focus that I've seen for public square kind of solutions are the libraries. The libraries getting a lot of attention lately for their role in society, their role in the future society as gathering place for information sharing and for expressing opinions." (E5-SE)

Another crucial factor in the human-centric AI approach is the **presence of control mechanisms within the public sector, along with the work ethic of the public sector employees** (E1, E2, E3, E5, E6, E8, E9). Several experts have noted that there exist fundamental principles of work ethics in the public sector that are applicable to all public servants, including those working in the field of artificial intelligence. For example, in Sweden there is a public sector ethos:

"Ethos has some general guidelines on how you should work as a public servant. Basically, you should be trying to promote democratic values, for example, or should not discriminate based on age, or sex, or gender and so on. So, these guidelines were taken to cover

all ethical aspects that we needed, no matter if we are talking about ethical AI or something else." (E5-SE)

Public agencies lack the will to undertake initiatives that could potentially harm their reputation (E2, E8). Consequently, they avoid high-risk projects and tend to be careful when considering the implementation of complex AI solutions (E2, E6). Furthermore, experts believe that solid governance systems, which effectively oversee AI solutions, are already in place within various departments (E8). Countries inform citizens about the utilisation of AI in public services to enhance transparency and ensure oversight (E9). They also offer citizens the chance to express their objections to decisions made by AI systems and provide explanations on how these decisions were generated (E9).

4.2.4 Challenges and interest

While many countries express their interest in adopting measures that promote a human-centric approach to AI, they also keep concerns and face challenges associated with it. Experts have noted that the development and implementation of human-centric measures require significant investments of time and financial resources. However, not all countries have the necessary resources allocated for such purposes (E4), especially if the implementation of the practices includes the development of specific tools and practices (E4). Moreover, the field of human-centric AI remains confusing as there is a lack of agreement on the mandatory tools that countries should employ. The absence of a clear strategy adopted by countries creates challenges for expected public institutions to follow a human-centric approach (E5, E6). Furthermore, there is a so-called race among technologically advanced countries that are promoting the adoption of artificial intelligence to improve the effectiveness of governance. However, the establishment of excessive regulations may impede innovation within the public sector (E2, E7).

Meanwhile, countries show a high interest towards practices that are discussed and implemented by other countries. Canada is deeply interested in the negotiations of The Council of Europe (E9). Additionally, Canada is paying close attention to the G7 recommendations, The Hiroshima Process International Guiding Principles for Organisations Developing Advanced AI Systems, the Digital Nations shared approach on AI, and the experiences of the USA and UK (E9). Finland has shown interest in the

OECD's initiatives regarding human-centric AI practices and is closely monitoring the advancements made by other Nordic countries (E4).

4.3 Human-centric AI approach in Estonia

This section deals with the information related to the third research question "RQ3: How can Estonia's human-centric AI approach be compared to other countries' approaches in order to make it more complete?". In this section, the author discusses legal, practical, and other measures that contribute to ensuring a human-centric AI approach in Estonia's public sector. Furthermore, it collects the perspectives of Estonian experts on the missing practices.

4.3.1 Human-centric AI measures

This section covers legal, practical, and other measures that contribute to ensuring a human-centric AI approach in Estonia's public sector.

Regulations

There are no laws in Estonia that are specific to technology, so there are none that are specific to AI. Like other countries involved in the research, Estonia has already passed laws that have a significant influence on specific areas closely related to AI, such as GDPR (E2, E3), the Administrative Procedure Act (E1, E3), and the Constitution of the Republic of Estonia (E3).

"We have not separately regulated the technology in Estonia. Our law today is purely technology neutral. Let's say, regardless of whether we are talking about artificial intelligence, or whether we are talking about conventional IT developments." (E1 - EE)

As an EU member state, Estonia looks forward to the implementation of the AI Act, which is set to become the primary regulatory framework for the field of artificial intelligence in Europe (E1, E2, E3). Over 5 years ago, Estonia made the decision to refrain from establishing any distinct regulations regarding AI, anticipating that inevitably the field of AI would be covered by EU legislation. Hence, fragmenting the united market of the European Union with country-specific regulations would be unreasonable (E1). Furthermore, legal analyses conducted by Estonia in 2018 demonstrated that all the key legal requirements had already existed (E1).

"We didn't really see the need [to additionally regulate AI] at that time, because we did a very scrupulous legal analysis. All these principles, from data quality, transparency and so on, from impact assessment… we actually already had all this in the legislation today." (E1 - EE)

One such measure is the GDPR, which safeguards privacy and establishes a structure for processing data. This regulation can be directly applicable to AI solutions (E2, E3). The Estonian expert mentioned that personal data cannot be processed unless there is a specific necessity (E1).

"Regarding the protection of private life and the right to privacy, the General Data Protection Regulation or GDPR is crucial. It lays out the general principles for when and how data can be processed, how can we combine data, of course, also in the context of artificial intelligence, and it sets out the protective measures people must be subject to." (E3- EE)

Another overarching rule is the Constitution of the Republic of Estonia. It establishes the fundamental rights that are applicable to the use of AI (E3). Experts emphasise that while the content may be at a high level, the Constitution is still important for AI field as it provides fundamental principles.

„Well, the central aspect simply is the Constitution of the Republic of Estonia. It establishes the basic rights of our people, and there is no need for any additional applications or basic laws for institutions to follow the Constitution when applying artificial intelligence. So, the framework of fundamental rights set by the Constitution is crucial and will continue to hold a central position. Additionally, field-specific legislation also holds great importance." (E3 - EE)

Estonia has implemented the Administrative Procedure Act, which permits the use of automated decision-making by public administrations (E3). However, there is still a need to legally specify the conditions under which automation is permitted in order to enhance clarity and address various protective measures:

"So, where we want to get, is to clarify when we allow automation. And if we were to implement automation, what considerations should be made regarding decisions? What safeguards would need to be put in place?... Well, in some cases, would it be crucial to have some form of human supervision or transparency? Or, for example, are there any cases there institutions should conduct audits to ensure the system's suitability before its implementation? That, well... we don't have such answers for the Administrative Procedure Act today." (E3- EE)

Practical tools

Estonia employs a range of measures that can be categorised as practical tools aimed at ensuring that AI is human-centric. That includes the institutions supporting the AI uptake within public organisations (E1, E2), human-centric AI design practices (E1, E2), privacy-enhancing technologies (E1), data panels and AI sandboxes (E2), data protection assessments (E1), and the making publicly available information regarding nearly all AI solutions used in the public sector (E1, E2). In addition, Estonia is currently working on implementing the AI transparency standard (E1).

Estonia is the only country that participated in the research and practising a centralised approach to AI development. This approach means that nearly all AI solutions implemented in the public sector were created with straight involvement from the Ministry of Economic Affairs and Communications (MKM) or the Information System Authority (RIA) (E2). These two agencies offer consultations to other public institutions and support in order to obtain the base capacity and skills required to carry out AI projects (E1).

"What was my personal principle - if you support one-two organisations, these people with a knowledge remain. The knowledge base of the institution itself grows significantly. As a result of this growth, the institution no longer requires as much support and help as before, so you can move on to the next organisation. In other words, its focus has been very clichéd - the cultivation of basic competencies." (E1 - EE)

Both mentioned organisations have competences in AI field, so they provide other public institutions with consultation and services such as seminars, workshops, brainstorming sessions, and other resources outlined in the AI Support Toolbox [13]. Additionally, they promote and oversee the adoption of human-centric AI design practices by organisations (E2). This approach includes a careful evaluation of all potential risks associated with the AI solution throughout the entire development process, ensuring that it is designed to be both safe and ethical. Emphasis is placed on the data processing procedures, which must be carried out in compliance with the law and without violating individuals' privacy. It is recommended for organisations to initiate pilot projects and focus on developing the minimum viable product. Prioritise thorough planning and careful preparation of all necessary documents for the procurement process. Additionally, organisations implementing AI projects are required to continuously monitor the AI solution even after its implementation to identify any potential risks it may pose.

"I guess we [MKM and RIA] have this approach - "safe and ethical by design", right from the beginning of any AI project. We ensure that oversight is present right from the start to uphold these principles. It appears to me that the lack of solutions causing significant issues or dramatic consequences indicates that we might be on the right track." (E2 - EE)

Estonia actively employs privacy-enhancing technologies in its data processing (E1). For example, MKM has been developing an anonymization tool to remove personal data from datasets. Additionally, there is a current project focused on federated learning, which aims to enable the training of AI models independently within each public agency, without the need for combining personal data. In addition, Estonia has initiated several projects focused on synthetic data. To gain more insight into potential technologies for use in the public sector, MKM this year also carried out PET analysis and already received the results [75].

As a part of AI Support Toolbox services, MKM and RIA carries out data panels and AI sandbox meetings with the goal of ensuring ethical and responsible data processing, improving AI capabilities, executing projects successfully, and accomplishing business goals (E2).

"The experts on the data panel will listen to what the project might be about, what risks are there, what data problems might come up, what data protection issues might come up, and what preventive measures might be possible or should be thought about. Then, the sandbox is more oriented towards practical implementation, playing it through." (E2- EE)

As a part of each data-related project, institutions are obliged to carry out data protection assessment (E1). Comparable to a start-up company's SWOT analysis, the impact assessment is conducted across the spectrum of personal data processing to minimise any potential risks to individuals' privacy and threats to personal data in the digital sphere [76].

In order to provide information and promote transparency regarding the utilisation of AI in the public sector, MKM and RIA have made the decision to publish information on AI projects on their website (E2). Currently, there is information available on approximately one hundred solutions, including a brief overview, the names of the organisations responsible for the projects, and other relevant data [2]. Institutions receiving financial assistance from MKM to develop AI solutions may be required to

publicly disclose the source code of the developed AI solutions in a code repository for e-governance solutions (E1).

"One approach in particular - the entire public code repository for e-governance solutions. When we fund AI projects, we are increasingly making it an obligation from our end. This also applies to research projects conducted through, say, a language technology programme, so you must always disclose the source code of a solution." (E1 - EE)

In addition, Estonia is currently adopting the AI transparency standard developed by the United Kingdom to offer a more comprehensive explanation of AI solutions in the public sector and educate the general public about the reasons for their utilisation. In the upcoming year, the MKM and RIA webpage dedicated to data and AI solutions will feature detailed descriptions of several AI projects (E1).

Other measures

Not all measures mentioned by Estonian experts can be classified as regulations or practical tools. Some of these measures, like those observed in other countries involved in the research, can be categorised as "other measures." This encompasses strategies and policies (E1, E2), guidelines (E1, E2), civic engagement (E1, E2), as well as control mechanisms within the public sector combined with public sector work ethics (E1, E2, E3).

Regarding **strategies and policies**, Estonia's Digital Agenda 2030, the primary strategy centred on the digitalization of the public sector, places a strong emphasis on human-centric practices (E1, E3). According to the Agenda, all public services must prioritise the needs and preferences of users, ensuring their fundamental rights are protected [1]. This means that these services should be designed and provided with a human-centric approach. The National AI Strategy is another significant strategic document in the field of AI in Estonia (E2). According to experts, the upcoming strategy will place an even greater emphasis on a human-centric approach to artificial intelligence (E1, E2).

Author: Currently the next AI strategy is being developed. To what extent does this strategy intend to pay attention to the aspect of human-centric AI approach?

E1: Yeah, very much in that sense. It is natural that we have an AI strategy in which evolution occurs gradually. So, in a situation where you have nothing done in the country, to speak of a human-centric approach, where do you apply it? If you haven't taken any actions and there are no AI solutions in use. It would be too early to discuss this kind of topic. Now we

have experience and have learned from what had been done, know where the risks and opportunities are. So, in the new strategy, we will definitely continue to put more attention on the human-centric AI approach. (E1 - EE)

In order to provide assistance to government agencies MKM and RIA have produced a range of **guiding materials** related to AI and other data-centric disciplines. While the content of the materials may not directly address the topic of human-centric AI, it highlights procedures that eventually aim to develop such technology (E1, E2). For example, through the high quality of the data that organisations might use for AI initiatives (E1).

"Certainly, what has been done, these are all possible guiding materials. There is, for example, a guide that assists in determining whether there is sufficient amount of data for the project. This is one of the guidelines that I've created a while ago. Or, how to tag data to improve data quality. Once again, very specific. I think that because we have not regulated artificial intelligence as such, these materials are not artificial intelligence specific." (E1 - EE)

Furthermore, **civic engagement** occurs in Estonia as a precondition for delivering services in a human-centric manner. Engagement primarily serves the purpose of gaining a deeper understanding of the needs of the end client. However, experts have observed that the level of engagement is greatly influenced by the specific approach taken by public institutions. Some institutions involve citizens to a greater extent than others (E2). Despite experts acknowledging that the level of engagement is lower than might be expected, they still observe that this topic receives more attention than it did previously (E1). For example, recently MKM and RIA ordered a survey to assess the level of awareness and opinions of Estonian residents regarding artificial intelligence (E1).

The expert emphasises the crucial role played by **control mechanisms and public sector work ethics** in the public sector (E1, E2, E3). The general principles of public sector work, including work ethics, are applicable to all areas of the public sector (E2, E3). In addition, numerous aspects of the public sector's work have an impact on the level of the human-centric approach to AI (E1). AI development and implementation involves various components, for example enhancing data quality in the public sector, ultimately leading to a more human-centric approach in AI (E1). Despite the presence of numerous factors that influence the quality of AI, both directly and indirectly, AI can serve as an

excellent illustration to advocate for a more comprehensive human-centric approach, extending beyond the boundaries of this particular field (E1).

"We need to move towards ensuring that these principles are truly applied everywhere. Whether we agree or not that this is... this is artificial intelligence... When we process personal data, when we provide services to people, what principles do we all agree on? what technological solutions do we use? How we have actually regulated in our legal space the topic of data, its storage, processing, management, as well as archiving, and so on." (E1 - EE)

Furthermore, AI is a new discipline, and it is perfectly acceptable for organisations just to begin exploring this field and establishing boundaries as they enter it (E3). Nevertheless, there are already existing organisational measures in place, for example engaging layers into the AI project team (E2). Additionally, another safeguard mechanism is for organisations to abstain from implementing high-risk projects (E2). When discussing risky AI solutions, it is crucial to understand that there is no rational reasons for intentionally developing problematic solutions or ignoring their problematic nature (E2). Utilising such solution will not yield any advantages. In fact, it may escalate the negative effects. The potential consequences for one's reputation are too significant to undertake projects that lack ethical standards (E2). Moreover, even if we observe any errors in the AI solution, it is necessary to consider the pre-existing practices that were in place prior to its implementation. AI can be a superior solution in certain cases, as its work often yields more accurate results compared to those produced by public sector workers (E1, E2).

"Requiring a model to work 100% of the time is unrealistic. In other words, again, this must be a so-called business decision, in essence, you need to decide what is optimal. If we allow people to make mistakes, why don't we allow technology to make mistakes? If humans are more fallible than technology, then technology is inherently the better solution. In this context, the fact that the AI is making mistakes by a certain percentage is not so bad." (1 - EE)

4.3.2 Further development

Estonian experts have observed a significant increase in interest regarding human-centric AI, indicating that it is an opportune moment to engage in discussions about this topic more extensively than in the past (E1, E2, E3). Estonia is being actively involved in

numerous initiatives and is frequently invited to participate in discussions on AI ethics (E2, E3).

Estonia demonstrates a practical and realistic approach towards the application of AI and measures supporting a human-centric approach. It endeavours to take concrete actions and carefully consider the next course of action (E2). Furthermore, it is essential to achieve a balance between implementing adequate protective measures and allowing room for innovation. The excessive enforcement of irrational limitations would significantly inflate the expenses associated with deploying AI for the delivery of public services (E1, E2).

As potential supplementary measures to strengthen the human-centric AI approach, experts have proposed the following: advancing PET solutions, such as anonymization and the utilisation of synthetic data (E1, E2); further enhancing sandboxes allowing freer data processing (E1, E3); establishing educational programmes to foster knowledge (E1, E3); increasing organisations' engagement with their end users (E1); developing tools to evaluate bias (E3); and implementing AI passports (E2). While experts acknowledge the potential for growth in the field of human-centric AI, they emphasise the importance of a careful and systematic approach to evaluate the actual requirements and abilities of the public sector and citizens (E1).

5 Results and discussion

This section provides answers to the research questions of this study outlining the main findings and correlations.

RQ1: How do countries define a human-centric AI approach?

The term human-centric AI approach gets defined by both overarching principles outlined in international guidelines and specific actions taken by countries to establish a human-centric AI approach.

From an ethical AI guidelines standpoint expressed by the experts participated in the research, human-centric AI is regarded as a system that must align with the **ethical principles** of transparency, explainability, accountability, robustness, fairness, safety, security, responsibility, trustworthiness, and accessibility.

By outlining **concrete actions**, it is possible to categorise human-centric AI into three distinct categories:

1) *Ethical AI design and compliance*, encompassing actions like adhering to legislative frameworks, mitigating bias in AI solutions, incorporating ethical considerations into AI design, performing risk assessments, safeguarding personal data, and upholding public trust in the public sector.
2) *Human-centric and efficient services*, including actions like considering people's needs, offering inclusive services, providing efficient public services, and adopting a life event-based approach.
3) *Transparent use of AI and engagement*, involving actions like informing people about the use of AI, granting the right to object AI decisions, respecting individual autonomy, protecting integrity, involving human in the loop of AI decision-making process.

The author attempted to establish a correlation between ethical principles and actual actions both mentioned by the experts by aligning them. The findings of this alignment can be observed in Table 2.

Table 2. Defining human-centric AI: categorized actions aligned with ethical principles mentioned by the experts. Source: Author

Purpose	Actions	Correlating ethical principles
Ethical AI design and compliance	○ following legislative frameworks ○ reducing bias in AI solutions ○ creating AI solutions ethical by design ○ conducting risk assessments ○ protecting personal data ○ maintaining trust in the public sector	Accountability Robustness Fairness Safety Security Responsibility
Human-centric and efficient services	○ considering people's needs ○ offering inclusive services ○ providing efficient public services ○ adopting a life event-based approach	Fairness Accessibility
Transparent use of AI and engagement	○ informing people about the use of AI ○ granting the right to object AI decisions ○ respecting individual autonomy ○ protecting integrity ○ involving humans in the AI decision-making process	Transparency Explainability Responsibility Trustworthiness

A closer look at the idea of a human-centric AI approach reveals that the goal of using the AI solutions may have an even greater impact than how AI is developed and utilised. According to the research, AI solutions in the public sectors of the countries that participated in the study are used to improve both internal organisational processes and services provided to citizens. Internally, AI solutions are used to improve public sector efficiency and aid in repetitive tasks, as well as to conserve financial resources and complete tasks that require more human resources than the public sector has. From the standpoint of the citizen, AI is used to improve the quality of public services provided to citizens, optimise service delivery by making intelligent decisions that benefit citizens the most, enable more efficient engagement with the government, improve the provision of personalised services, and aid in the delivery of proactive services.

It should be noted that the entire field of human-centric AI approaches is currently in a very hectic and still developing stage due to several factors combining - high public interest in AI as an innovation as well as a source of potential threats, high rate of

developing AI solutions in the public sector, awaiting legal regulation of the AI field, and developing each country's own tools and measures ensuring human-centric AI. As a result, combination of the aforementioned factors brings a lack of a methodical and mandatory approach to developing ethically sound solutions. However, the author believes that this problem will be resolved naturally as the topic of AI overcomes the peak of the hype cycle and countries gain more experience through already implemented AI solutions and human-centric AI tools and measures.

RQ2: How do countries translate human-centric AI approach into life?

There is a significant level of interest in the human-centric AI approach across all countries. The significant interest has arisen mainly the last years as a result of the rapid advancement of AI technology and its broader integration in both the private and public sectors. Consequently, there is now a pressing requirement to adopt a human-centric AI approach. The political debate stimulates a discussion regarding the development of effective tools to promote human-centric values in AI. This leads to the creation of new tools and the integration of the human-centric AI approach into government strategies and programmes. Nevertheless, it has been observed that in certain instances, there might be a heightened emphasis on theoretical curiosity rather than a willingness to invest efforts into practical implementations. However, the practical implementation of human-centric measures still gets achieved through the utilisation of three distinct categories of measures: regulations (law), practical tools, and other measures. All tools and measures discovered via research are described in the Table 3.

Taking a look at the first category of measures **- regulations** (law), it can be noticed that the domain of artificial intelligence continues to be largely unregulated by legislation across many countries. However, while many countries lack explicit regulations specifically addressing AI, they often have other laws that are closely related to the field of AI. These laws can be referred to as technology-neutral and are applicable to various domains, including AI. An example of technology-neutral legislation can be presented through the framework of the International Bill of Human Rights, which requires countries to ensure rights on freedom from discrimination, equality between men and women, right to privacy, and other freedoms that not directly aimed, but might be relevant to the human-centric application of AI in the public sector. In the case of the EU, one of the primary pieces of legislation that applies to AI today is the GDPR, which aids in the

protection of individuals' personal data. Moreover, there is an apparent distinction in the approaches of the EU and non-EU countries included in the study. EU member states are currently developing stricter regulations for the field of AI in the face of AI Act, whereas non-EU countries are adopting a more technology-neutral approach and would like to have less regulations of AI field. Overall, although the hard law on AI can be viewed as one of the primary means of guaranteeing a human-centric approach by defining specific measures to ensure ethics or by prohibiting certain, riskier solutions, it also has drawbacks, such as the challenge of anticipating the risks connected to the quick development of AI technology.

The second category of measures identified during the research is **practical tools** that countries use to ensure human-centric AI. Practical tools are those that involve the active participation of agencies developing or deploying AI solutions in order to promote human-centric values in AI. There are two types of tools that have been identified.

1. *Human-centric AI design and development tools* including practices, institutions, and facilities that guide AI agencies towards human-centric AI, privacy-enhancing technologies, and regulatory AI sandboxes.
2. *Already developed AI solutions' assessment and transparency enchantment tools* including AI transparency standards, AI registers, and AI assessment tools.

Third category or **other measures** are those that fall into a set of supportive measures that advance the human-centric AI approach, but not directly connected to any laws or practical tools that public institutions employ in the field of AI. According to research, the presence of control mechanisms within the public sector, as well as the work ethic of public sector employees, plays a key role in ensuring the human-centric AI approach. Many countries rely heavily on public sector work ethics and technology-neutral legislation to ensure that AI solutions are human-centric. Meaning that employees engaged in the public sector's AI development should uphold the core values of public sector work ethics, which state that public servants must act in the public interest.

In addition to the significant impact of work ethics, other measures include promoting knowledge, conducting research, and incorporating ethical principles into strategies and policies.

Table 3. Tools and measures for translating human-centric AI principles into practice. Source: Author

	Purpose	Active tools and measures	Explanation and relation to human-centric AI approach
Law	Establish legal certainty and responsibilities	Direct AI legislation	AI-specific legislation intended to directly regulate the AI industry, aims to prohibit specific solutions, or establish specific practices.
		Technology-neutral legislation	Although not explicitly designed for AI, legislation that establishes specific regulations and has a substantial connection to the field of AI.
Practical tools	Design and develop human-centric AI solutions	Human-centric AI design practices	Methods of developing AI solutions with an emphasis on citizen needs and the incorporation of human-centric values throughout the AI development process.
		Institutions and facilities that guide AI agencies towards human-centric AI	Institutions that serve as a support system for agencies seeking to adopt AI solutions, in addition to advising on human-centric AI approaches and assisting with solution testing.
		Privacy-enhancing technologies	Digital solutions that enable to share and use data in a way that protects confidentiality.
		Regulatory AI sandboxes	Experimentation tools that provide a regulated environment for testing and scaling up AI algorithms in a safe way.
	Assess AI solutions and enhance transparency	AI assessment tools	Evaluation instruments that enable agencies developing AI solutions to oversee the ethical compliance of their own systems.
		AI transparency standard	Standard enhancing transparency of AI tools by providing full descriptions of the tools and detailed information on their application.
		AI registers	Central repositories which create the collection of information on AI solutions employed in the public sector, aiming to enhance transparency towards citizens.
Other measures	Educate and promote the adoption of a human-centric AI approach	AI and related field guidelines	Detailed recommendations for the AI in the public sector and other related fields.
		Educational programmes	Human-centric AI approach-promoting courses and programmes that explain how to implement such an approach.
		Events	Events functioning as platform for initiating discussions and spreading information regarding human-centric AI approaches.
		Programmes of higher education	Academic programmes educating future specialists on the ethical development and implications of AI.
	Discover new practices and citizens' viewpoint	Research	Research to gain a public perspective, construct a deeper understanding of the human-centric AI field, and investigate the tools for human-centric practices.
		Civic engagement	Direct citizen participation in discussions concerning the use of AI in the public sector in order to comprehend the public's perspective.
	Establish political will, address strategic steps	Strategies and policies	Strategic plans and high-level documents used to advocate for the country's AI approach that prioritises human needs.
	Maintain loyalty to ethical principles	AI Oath	Oath to act in the best interests of humanity within the domain of AI.
		Public sector work ethics and control mechanisms	Adhering to the most fundamental principles of public sector work ethics, which dictate that public officials should act in the best interests of the public.

RQ3: How can Estonia's human-centric AI approach be compared to other countries' approaches in order to make it more complete?

Estonia is showing a strong interest in both AI development and the field of human-centric AI. Table 4 contains the full set of tools and measures used in Estonia, as well as a comparison of the approaches mentioned by other countries.

In terms of **regulations**, Estonia, as an EU country, is awaiting direct regulation of the AI field through the AI Act. However, many technology-neutral regulations, such as GDPR, the Administrative Procedure Act, and the Constitution of the Republic of Estonia, are already in use in Estonia to support human-centric values in AI solutions. Estonia uses a variety of measures that can be classified as practical tools aimed at ensuring that AI is human-centric.

Furthermore, Estonia has quite unique approach to developing AI solutions. Almost all AI solutions implemented in the public sector were developed with direct involvement from MKM and RIA - agencies which provided consultations and support to other public institutions in order to obtain the base capacity and skills required to carry out AI projects. It should be noted that these organisations are also in charge of AI policy. As a result, Estonia has a highly centralised approach to the AI field in the public sector, allowing for both innovation and control over agencies implementing AI solutions, with a focus on the AI development stage.

In terms of concrete **practical tools**, Estonia is implementing privacy-enhancing technologies, data panels and AI sandboxes, data protection assessments, and making public information about nearly all AI solutions used in the public sector.

In terms of **other measures**, Estonia, like many other countries, heavily relies on public sector work ethics as the primary component of ensuring human-centric AI. Furthermore, Estonia promotes AI through strategies and policies, develops guidelines, and attempts to engage more citizens in the discussion of AI-related topics.

When we **compare Estonia approach to the other countries** (Table 4), we observe that it mainly lacks measures related to knowledge promotion on human-centric AI approach through events and different courses. Due to Estonia's strong emphasis on the incorporation of human-centric values during the development stage of AI solutions, we can see that it has a slightly weaker practice aimed at assessing already developed AI solutions and increasing their transparency. Estonia did, however, previously invest in the adoption of the Canada AI assessment tool and is currently working to improve the

situation by adapting the UK AI transparency standard. When looking at the overall picture of the already implemented solutions, Estonia stands out as having made a significant contribution to the field of human-centric AI. It should be noted that none of the countries involved in the research use all of the tools and measures listed in Table 3 and Table 1Table 4. As a result, gaps in the Estonian approach should not be interpreted as a lack of investment in the human-centric AI field. From the viewpoint of the Estonian experts, more work can be done to strengthen existing measures and make them more comprehensive. Estonia believes it is critical to have a realistic and practical approach to the AI field, taking into account both innovation and a focus on the interests of their citizens.

Considering Estonia's approach to innovation and the comparative analyses conducted during this research, the **following steps can be proposed to make Estonia's human-centric AI approach more complete**:

1. Continue to strengthen existing tools and measures and, if necessary, scale them to meet the increasing pace of AI development in the public sector.
2. Continue collaboration with other countries in order to obtain best practices rather than developing own solutions from the ground up, if solutions suitable for accommodation in the Estonian context already exist.
3. Determine the amount and order of mandatory measures that should be implemented by public agencies during the development and implementation of AI solutions, with the objective of establishing a systematic approach.
4. Maintain emphasis on the development phase of AI solutions, including control over both the goal of the coming AI solutions and the process by which they are designed and developed.
5. Ensure that, while the main focus stays on AI development, there are also tools to assess AI solutions that have already been developed, as well as tools that help to improve AI solutions' transparency for the public.
6. Increase the emphasis on knowledge promotion in the human-centric AI field and approaches that lead to such ways of innovation via courses, guidelines, events, and educational systems.

Table 4. Comparison of tools and measures for translating human-centric AI principles into practice with adoption status in Estonia. Source: Author

	Purpose	Tools and measures	Countries mentioning tools or measures	Presence in Estonia (X)
Law	Establish legal certainty and responsibilities	Direct AI legislation	* FI, SE, NL, BE, CA	X*
		Technology-neutral legislation	FI, SE, NL, BE, GB, CA	X
Practical tools	Design and develop human-centric AI solutions	Human-centric AI design practices	SE, CA	X
		Institutions and facilities that guide AI agencies towards human-centric AI	FI, SE, NL, GB	X
		Privacy-enhancing technologies	GB	X
		Regulatory AI sandboxes	SE	X
	Assess AI solutions and enhance transparency	AI assessment tools	SE, NL, CA	
		AI transparency standard	GB	X*
		AI registers	NL	X
Other measures	Educate and promote the adoption of a human-centric AI approach	AI and related field guidelines	FI, SE	X
		Educational programmes	SE, BE	
		Events	FI, SE, GB	
		Programmes of higher education	BE	
	Discover new practices and citizens' viewpoint	Research	SE, GB, CA	X
		Civic engagement	SE, NL, BE, GB, CA	X
	Establish political will, address strategic steps	Strategies and policies	FI, NL, BE, GB, CA	X
	Maintain loyalty to ethical principles	AI Oath	SE, NL, BE, GB, CA	
		Public sector work ethics and control mechanisms	SE, NL, BE, GB, CA	X

*Tools and measures marked " * " are under development*

6 Summary

The main goal of this research was to learn how countries move from ethics to action when developing, deploying, and using AI tools in public services; with a purpose to compare Estonia's human-centric AI approach to other countries in order to make it more complete. The whole study provides answers to the three research questions.

The theoretical framework of the thesis, based on the literature review, outlines the major steps in the rise of AI technology and its subsequent implementation in the public sector, using an example of Estonia to showcase how AI can be utilised for achieving public good. It also aims to explore how ethical principles in several international guidelines outline human-centric AI, as well as why AI ethics get critiqued, revealing the gaps in real-life implementation of AI ethics.

To achieve the aim of the research, the author used a qualitative research approach, including nine semi-structured interviews with experts from Estonia, Finland, Sweden, the Netherlands, Belgium, the United Kingdom, and Canada who have an affiliation with the AI field in the public sector of their countries and are involved in the development or implementation of measures related to the human-centric AI approach. Using qualitative content analysis to present a summary of the main concepts and findings, focusing on the research questions and other important viewpoints expressed by the interviewees, the author identified three main topics and ten subtopics directly related to the aim of the research.

Research revealed that the term human-centric AI approach indeed gets defined by principles outlined in international guidelines. However, it is more often described by concrete actions associated with the ethical behaviour of public officials while developing, deploying, and using AI solutions. Such actions can be divided into three categories: 1) serving the purpose of ethical AI design and compliance; 2) helping to achieve human-centric and efficient public services; 3) contributing to transparent use of AI and engagement. Moreover, research shows that an important role plays not only *how*

AI solutions are developed and utilised, but even greater impact might have *what* AI solutions get deployed. Meaning that for an AI solution to be ethically complaint, its initial purpose must also be strictly human-centric.

There is a significant level of interest in the human-centric AI approach across all countries, especially in light of the rapid advancement of AI technology and its broader integration in both the private and public sectors. However, in certain instances, this interest might place a heightened emphasis on theoretical curiosity rather than a willingness to invest efforts into practical implementations. When it comes to the real-life application of AI ethics, countries achieve human-centric AI by utilising measures that can be divided into three distinct categories.

Legal regulations or laws that serve a purpose to establish legal certainty and responsibilities, including AI-specific legislation intended to directly regulate the AI industry, which aims to prohibit specific solutions or establish certain legal practices, and, on the other hand, technology-neutral legislation, which is not explicitly designed for AI but establishes specific regulations and has a substantial connection to the field of AI.

Practical tools or tools that involve the active participation of agencies developing or deploying AI solutions in order to promote human-centric values help to design and develop human-centric AI solutions through design practices, institutions and facilities, privacy-enhancing technologies, and regulatory sandboxes. As well as aim to assess AI solutions and enhance their transparency through assessment tools, transparency standards, and the use of AI registers.

Other measures fall into a set of supportive measures that advance the human-centric AI approach but are not directly connected to any laws or practical tools that public institutions employ in the field of AI. These supportive measures have the purpose of educating individuals and promoting the adoption of a human-centric AI approach via guidelines, educational programmes, and events. They also help to discover new practices and citizens' viewpoints by conducting research and supporting civic engagement, as well as establish political will and address strategic steps using strategies and policies. But the most prominent impact other measures have on the maintenance of loyalty to ethical principles through public sector work ethics and control mechanisms, which often play a key role in ensuring the human-centric AI approach.

Compared to other countries, Estonia also has a strong interest in the further development of human-centric AI approach. Estonia stands with its highly centralised approach to the AI field in the public sector, allowing for both innovation and control over agencies implementing AI solutions, with a focus on the AI development stage. From a legal perspective, Estonia relies mostly on technology-neutral regulations, which help support the presence of human-centric values in developed AI solutions. From the perspective of practical tools, Estonia is quite advanced in the design and development of human-centric AI solutions but has unrealized potential for assessment of already developed solutions and enhancing their transparency. In terms of other measures, Estonia, like many other countries, heavily relies on public sector work ethics but also promotes a human-centric AI approach through strategies and policies, develops guidelines, and attempts to engage more citizens in the discussion of AI-related topics. Meanwhile, it lacks measures to educate and promote the adoption of a human-centric AI approach.

With a purpose to make Estonia's human-centric AI approach more complete, the author proposed six steps:

1. Continue to strengthen existing tools and measures and, if necessary, scale them to meet the increasing pace of AI development in the public sector.
2. Continue collaboration with other countries in order to obtain best practices rather than developing your own solutions from the ground up if solutions suitable for accommodation in the Estonian context already exist.
3. Determine the amount and order of mandatory measures that should be implemented by public agencies during the development and implementation of AI solutions, with the objective of establishing a systematic approach.
4. Maintain emphasis on the development phase of AI solutions, including control over both the goal of the coming AI solutions and the process by which they are designed and developed.
5. Ensure that, while the main focus stays on AI development, there are also tools to assess AI solutions that have already been developed, as well as tools that help to improve AI solutions' transparency for the public.
6. Increase the emphasis on knowledge promotion in the human-centric AI field and approaches that lead to such ways of innovation via courses, guidelines, events, and educational systems.

In conclusion, the research showed that countries take active steps from AI ethics to action, which reflects in the number of tools and measures created with a purpose to ensure that AI solutions are human-centric. However, many countries, including Estonia, still lack a methodical and mandatory approach to developing ethically sound solutions due to the novelty and rapid development of the AI field. The author believes that this problem will be resolved naturally as the topic of AI overcomes the peak of the hype cycle and countries gain more experience through already implemented AI solutions and human-centric AI tools and measures.

References

[1] Ministry of Economic Affairs and Communications of Estonia, “Estonia’s Digital Agenda 2030,” 2021. Accessed: Oct. 22, 2023. [Online]. Available: https://www.mkm.ee/en/e-state-and-connectivity/digital-agenda-2030

[2] Ministry of Economic Affairs and Communications of Estonia, “AI kasutuslood.” Accessed: Oct. 22, 2023. [Online]. Available: https://www.kratid.ee/kasutuslood-kratid

[3] European Commission, “White Paper on Artificial Intelligence A European approach to excellence and trust,” 2020. Accessed: Oct. 22, 2023. [Online]. Available: https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[4] UNESCO, “Recommendation on the Ethics of Artificial Intelligence,” 2022, Accessed: Oct. 23, 2023. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000381137

[5] S. Bird *et al.*, “Fairlearn: A toolkit for assessing and improving fairness in AI,” 2020.

[6] A. Campolo, M. Sanfilippo, M. Whittaker, and K. Crawford, “AI Now 2017 Report,” 2017, Accessed: Oct. 23, 2023. [Online]. Available: https://ainowinstitute.org/publication/ai-now-2017-report-2

[7] J. Morley, L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, and L. Floridi, “Operationalising AI ethics: barriers, enablers and next steps,” *AI Soc*, vol. 38, no. 1, pp. 411–423, Feb. 2023, doi: 10.1007/s00146-021-01308-8.

[8] J. Zhou and F. Chen, “AI ethics: from principles to practice,” *AI Soc*, Nov. 2022, doi: 10.1007/s00146-022-01602-z.

[9] T. Hagendorff, “The Ethics of AI Ethics: An Evaluation of Guidelines,” *Minds Mach (Dordr)*, vol. 30, no. 1, pp. 99–120, Mar. 2020, doi: 10.1007/s11023-020-09517-8.

[10] L. Munn, “The uselessness of AI ethics,” *AI and Ethics*, vol. 3, no. 3, pp. 869–877, Aug. 2023, doi: 10.1007/s43681-022-00209-w.

[11] I. Strümke, M. Slavkovik, and V. I. Madai, “The social dilemma in artificial intelligence development and why we have to solve it,” *AI and Ethics*, vol. 2, no. 4, pp. 655–665, Nov. 2022, doi: 10.1007/s43681-021-00120-w.

[12] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, “From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices,” *Sci Eng Ethics*, vol. 26, no. 4, pp. 2141–2168, Aug. 2020, doi: 10.1007/s11948-019-00165-5.

[13] Ministry of Economic Affairs and Communications of Estonia, “ AI Support Toolbox.” Accessed: Dec. 12, 2023. [Online]. Available: https://www.kratid.ee/en/kratitoe-portfell

[14] Ministry of Economic Affairs and Communications of Estonia, “Human-centric AI.” Accessed: Oct. 24, 2023. [Online]. Available: https://www.kratid.ee/en/inimkeskne-kratt

[15] A. M. Turing, “COMPUTING MACHINERY AND INTELLIGENCE,” *Computing Machinery and Intelligence. Mind*, vol. 49, pp. 433–460, 1950.

[16] J. Moor, “The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years.,” *AI Mag*, vol. 27, Jan. 2006, Accessed: Nov. 04, 2023. [Online]. Available: https://www.researchgate.net/publication/220605256_The_Dartmouth_College_Artificial_Intelligence_Conference_The_Next_Fifty_Years

[17] A. Rockwell, “The History of Artificial Intelligence.” Accessed: Nov. 04, 2023. [Online]. Available: https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

[18] Y. Pan, “Heading toward Artificial Intelligence 2.0,” *Engineering*, vol. 2, no. 4, pp. 409–413, Dec. 2016, doi: 10.1016/J.ENG.2016.04.018.

[19] H. Mehr, “Artificial Intelligence for Citizen Services and Government,” 2017. Accessed: Nov. 04, 2023. [Online]. Available: https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

[20] W. G. de Sousa, E. R. P. de Melo, P. H. D. S. Bermejo, R. A. S. Farias, and A. O. Gomes, “How and where is artificial intelligence in the public sector going? A literature review and research agenda,” *Gov Inf Q*, vol. 36, no. 4, p. 101392, Oct. 2019, doi: 10.1016/j.giq.2019.07.004.

[21] European Commission *et al.*, “AI Watch, road to the adoption of artificial intelligence by the public sector : a handbook for policymakers, public administrations and relevant stakeholders,” 2022. Accessed: Nov. 05, 2023. [Online]. Available: https://data.europa.eu/doi/10.2760/288757

[22] AI HLEG, “Ethics Guidelines for Trustworthy AI,” 2019. Accessed: Nov. 25, 2023. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[23] Ministry of Economic Affairs and Communications of Estonia and Government Office of Estonia, “Report of Estonia’s AI Taskforce,” 2019, Accessed: Oct. 26, 2023. [Online]. Available:

https://www.kratid.ee/_files/ugd/980182_681757534b4a444caf6b7bd8796cfc4c.pdf

[24] Ministry of Economic Affairs and Communications of Estonia, *Estonia's National Artificial Intelligence Strategy or Kratt Strategy for 2022–2023*. 2022. Accessed: Oct. 26, 2023. [Online]. Available: https://www.kratid.ee/_files/ugd/980182_4434a890f1e64c66b1190b0bd2665dc2.pdf

[25] Ministry of Economic Affairs and Communications of Estonia, "Estonia's national artificial intelligence strategy 2019-2021," 2019. Accessed: Oct. 25, 2023. [Online]. Available: https://www.kratid.ee/_files/ugd/980182_8d0df96fd41145739dff2595e0ab3e8d.pdf

[26] Government of the Republic of Estonia, *Coalition agreement 2023-2027*. 2023. Accessed: Oct. 26, 2023. [Online]. Available: https://valitsus.ee/en/coalition-agreement-2023-2027

[27] Ministry of Economic Affairs and Communications of Estonia and Emor AS, "'Eesti elanike teadlikkus ja arvamused tehisintellektist' uuringu alusandmed." Accessed: Nov. 26, 2023. [Online]. Available: https://avaandmed.eesti.ee/datasets/%22eesti-elanike-teadlikkus-ja-arvamused-tehisintellektist%22-uuringu-alusandmed

[28] European Parliament, "EU AI Act: first regulation on artificial intelligence," News European Parliament.

[29] OECD.AI, "Public AI projects worldwide over time," OECD.AI Policy Observatory. Accessed: Nov. 26, 2023. [Online]. Available: https://oecd.ai/en/data?selectedArea=ai-software-development

[30] OECD.AI, "OECD AI Incidents Monitor," OECD.AI Policy Observatory. Accessed: Nov. 26, 2023. [Online]. Available: https://oecd.ai/en/incidents?search_terms=%5B%5D&and_condition=false&from_date=2014-01-01&to_date=2023-11-26&properties_config=%7B%22principles%22:%5B%5D,%22industries%22:%5B%5D,%22harm_types%22:%5B%5D,%22harm_levels%22:%5B%5D,%22harmed_entities%22:%5B%5D%7D&only_threats=false&order_by=date&num_results=20

[31] H. Devlin, R. Cousins, and A. Amitrano, "A day in the life of AI," *The Guardian*, 2023. Accessed: Nov. 26, 2023. [Online]. Available: https://www.theguardian.com/technology/ng-interactive/2023/oct/25/a-day-in-the-life-of-ai

[32] L. Blouin, "AI's mysterious 'black box' problem, explained," *University of Michigan-Dearborn*, 2023. Accessed: Nov. 26, 2023. [Online]. Available: https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained

[33] A. Jobin, M. Ienca, and E. Vayena, “The global landscape of AI ethics guidelines,” *Nat Mach Intell*, vol. 1, no. 9, pp. 389–399, Sep. 2019, doi: 10.1038/s42256-019-0088-2.

[34] N. Smuha, “AI between Ethics and Law: Guidelines for Trustworthy AI of the EC High-Level Expert Group on AI,” 2021. Accessed: Nov. 27, 2023. [Online]. Available: https://www.youtube.com/watch?v=d3P-SBoGtxk&t=1711s

[35] D. Ihde, “Technology and prognostic predicaments,” *AI Soc*, vol. 13, no. 1–2, pp. 44–51, Mar. 1999, doi: 10.1007/BF01205256.

[36] I. van de Poel, “Design for Values,” *Social responsibility and science in innovation economy*, pp. 115–165, 2015, Accessed: Nov. 11, 2023. [Online]. Available: https://e.kul.pl/files/10351/public/Social_respon_srodek_DRUK2.pdf#page=115

[37] L. Winner, “Do Artifacts Have Politics?,” *Daedalus*, vol. 109(1), pp. 121–136, 1980, Accessed: Nov. 09, 2023. [Online]. Available: http://www.jstor.org/stable/20024652

[38] OECD, “Recommendation of the Council on Artificial Intelligence,” 2019. Accessed: Nov. 27, 2023. [Online]. Available: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449#mainText

[39] Google, “Responsible AI practices,” Google AI. Accessed: Nov. 27, 2023. [Online]. Available: https://ai.google/responsibility/responsible-ai-practices

[40] IBM, “AI Ethics.” Accessed: Nov. 27, 2023. [Online]. Available: https://www.ibm.com/impact/ai-ethics

[41] IEEE, “ Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition,” 2019. Accessed: Nov. 27, 2023. [Online]. Available: https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html

[42] European Commission, “Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts.” Accessed: Nov. 27, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

[43] The White House, “Blueprint for an AI bill of rights making automated systems work for the american people.” Accessed: Nov. 27, 2023. [Online]. Available: https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[44] J. Morley, A. Elhalal, F. Garcia, L. Kinsey, J. Mökander, and L. Floridi, “Ethics as a Service: A Pragmatic Operationalisation of AI Ethics,” *Minds Mach (Dordr)*, vol. 31, no. 2, pp. 239–256, Jun. 2021, doi: 10.1007/s11023-021-09563-w.

[45] L. Õunapuu, *Kvalitatiivne ja kvantitatiivne uurimisviis sotsiaalteadustes*. Tartu: Tartu Ülikool, 2014.

[46] K. Lepik, H. Harro-Loit, K. Kello, M. Linno, M. Selg, and J. Strömpl, “Intervjuu.” Accessed: Dec. 16, 2023. [Online]. Available: https://samm.ut.ee/intervjuu

[47] V. Kalmus, A. Masso, and M. Linno, “Kvalitatiivne sisuanalüüs.” Accessed: Dec. 16, 2023. [Online]. Available: https://samm.ut.ee/kvalitatiivne-sisuanalyys

[48] A. Olev and T. Alumäe, “Estonian Speech Recognition and Transcription Editing Service,” *Baltic Journal of Modern Computing*, vol. 10, no. 3, 2022, doi: 10.22364/bjmc.2022.10.3.14.

[49] European Union, “General Data Protection Regulation.” 2016. Accessed: Dec. 09, 2023. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[50] United Nations, “International Bill of Human Rights.” 1948.

[51] United Nations Human Rights Office of The High Commissioner, “Ratification of 18 International Human Rights Treaties,” 2020.

[52] Government of the United Kingdom, “The Data Protection Act.” Accessed: Dec. 09, 2023. [Online]. Available: https://www.gov.uk/data-protection

[53] Government of Canada, “Artificial Intelligence and Data Act.” Accessed: Dec. 09, 2023. [Online]. Available: https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act

[54] Government of Canada, “Government of Canada’s approach on Gender-based Analysis Plus.” Accessed: Dec. 09, 2023. [Online]. Available: https://women-gender-equality.canada.ca/en/gender-based-analysis-plus/government-approach.html

[55] Digital and Population Data Services Agency, “About the Digital and Population Data Services Agency (The Finnish Digital Agency).” Accessed: Dec. 09, 2023. [Online]. Available: https://dvv.fi/en/digital-and-population-data-services-agency

[56] AI Sweden, “About AI Sweden.” Accessed: Dec. 09, 2023. [Online]. Available: https://www.ai.se/en/about

[57] Centre for Data Ethics and Innovation, “About the Centre for Data Ethics and Innovation.” Accessed: Dec. 09, 2023. [Online]. Available: https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation/about

[58] Centre for Data Ethics and Innovation, “Introducing our responsible data access work programme.” Accessed: Dec. 09, 2023. [Online]. Available: https://cdei.blog.gov.uk/2022/06/13/introducing-our-responsible-data-access-programme/#:~:text=The%20responsible%20data%20access%20programme,of%20the%20National%20Data%20Strategy.

[59] Centre for Data Ethics and Innovation, “CDEI portfolio of AI assurance techniques.” Accessed: Dec. 09, 2023. [Online]. Available: https://www.gov.uk/guidance/cdei-portfolio-of-ai-assurance-techniques

[60] Government of Canada, “Algorithmic Impact Assessment tool,” 2023, Accessed: Dec. 09, 2023. [Online]. Available: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html

[61] Government of the Netherlands, “AI Impact Assessment.” Accessed: Dec. 09, 2023. [Online]. Available: https://www.government.nl/documents/publications/2023/03/02/ai-impact-assessment

[62] Central Digital and Data Office and Centre for Data Ethics and Innovation, “Algorithmic Transparency Recording Standard Hub.” Accessed: Dec. 10, 2023. [Online]. Available: https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub

[63] The Algorithm Register, “About the Algorithm Register.” Accessed: Dec. 10, 2023. [Online]. Available: https://algoritmes.overheid.nl/en/footer/over

[64] Government of the Netherlands, “Value-Driven Digitalisation Work Agenda.” Accessed: Dec. 10, 2023. [Online]. Available: https://www.government.nl/documents/reports/2022/11/30/value-driven-digitalisation-work-agenda

[65] M.-C. Benoit, “Belgium adopts a national plan for the development of artificial intelligence.” Accessed: Dec. 10, 2023. [Online]. Available: https://www.actuia.com/english/belgium-adopts-a-national-plan-for-the-development-of-artificial-intelligence/

[66] Government of the United Kingdom, “A pro-innovation approach to AI regulation.” Accessed: Dec. 10, 2023. [Online]. Available: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

[67] Government of the United Kingdom, “National AI Strategy.” Accessed: Dec. 10, 2023. [Online]. Available: https://www.gov.uk/government/publications/national-ai-strategy

[68] Government of Canada, “Directive on Automated Decision-Making.” Accessed: Dec. 10, 2023. [Online]. Available: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592

[69] Digital and Population Data Services Agency, “Guide for digital services developers using AI responsibly,” 2023, Accessed: Dec. 10, 2023. [Online]. Available: https://www.suomi.fi/guides/responsible-ai

[70] DIGG, “Offentlig AI.” Accessed: Dec. 10, 2023. [Online]. Available: https://beta.dataportal.se/offentligai

[71] Government of Canada, “Guide on the use of Generative AI.” Accessed: Dec. 10, 2023. [Online]. Available: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-generative-ai.html

[72] Government of the United Kingdom, “AI Safety Summit 2023.” Accessed: Dec. 10, 2023. [Online]. Available: https://www.gov.uk/government/topical-events/ai-safety-summit-2023

[73] VUB, “Postgraduate ‘AI for the common good.’” Accessed: Dec. 10, 2023. [Online]. Available: https://www.vub.be/en/studying-vub/all-study-programmes-vub/bachelors-and-masters-programmes-vub/ai-for-common-good

[74] Government of the United Kingdom, “Public attitudes to data and AI: Tracker survey (Wave 3),” 2023, Accessed: Dec. 10, 2023. [Online]. Available: https://www.gov.uk/government/publications/public-attitudes-to-data-and-ai-tracker-survey-wave-3/public-attitudes-to-data-and-ai-tracker-survey-wave-3

[75] Ministry of Economic Affairs and Communications of Estonia, “Analyses and studies.” Accessed: Dec. 12, 2023. [Online]. Available: https://www.kratid.ee/en/analuusid-ja-uuringud

[76] Andmekaitse inspektsioon, “Mõjuhinnangu tegemine.” Accessed: Dec. 12, 2023. [Online]. Available: https://www.aki.ee/et/eraelu-kaitse/mojuhinnangu-tegemine

Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Sofia Paes

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis "From Ethics to Action: A Study of Human-Centric AI Implementation in Public Services, Comparing the Estonian Approach with Approaches Used in Other Countries", supervised by Innar Liiv
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

18.12.2023

1 The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.

Appendix 2 – Interview questions

1) Introduction
2) How do you define AI in your country?
3) What is the primary reason for the use of AI solutions in your country?
4) Do you have a complete picture of AI solutions currently in use in the public sector?
5) How would you rate your country's public-sector experience with AI solutions? *(For example, positive/negative)*
6) Is AI solution development centralised or decentralised?
    a) Do you have any public authorities in place to assist public institutions as they develop AI solutions? *(For example, a support or competence centre)*
7) Is the use of AI-assisted decision-making permitted in the public sector?
8) How important is the topic of human-centric (ethical) AI on a national level?
    a) Does it have any reflection in strategies/policies?
    b) How would you assess public sector organisations' interest in the topic of human-centric (ethical) AI? Are those developing AI solutions focusing solely on the technical side or also on the ethical side?
    c) Does your country have any AI implementation fields where human-centric (ethical) AI is particularly relevant? *(For example, fields with higher risks)*
9) How does your country define human-centric (ethical) AI?
    a) What principles/values should it adhere to? *(For example, fairness, accountability, and so on.)*
10) Does your country have any legal practices in place to ensure human-centric (ethical) AI creation and use in the public sector? *(For example, an AI law)*
11) Do you have any practical solutions in your country to ensure human-centric (ethical) AI creation and use in the public sector?
    a) How did these solutions emerge?

b) How do these solutions function?
c) What stages of AI creation/use are they concerned with?
d) Are public agencies required to use these solutions?
e) How would you describe the use of these solutions, organised or hectic?

12) Do ordinary citizens participate in the discussion of human-centric (ethical) AI?
    a) Does their opinion affect future actions of the government/public institutions?
13) Do you feel that there are any practical solutions or practices for ensuring human-centric (ethical) AI that are currently lacking?
14) Are there any concerns in the public sector about implementing human-centric (ethical) practices? *(For example, high cost)*
15) Are there any trends/initiatives in the international community involving/interesting your country in human-centric AI?