Artificial Intelligence and Automated Decision Making systems adoption by the public sector:

Insights from AI and ADM Projects by the Swedish public sector organisations

Hertie School

Anna Kubaszewska

Master of Public Policy, 2024

Supervisor: Prof. Dr. Kai Wegrich

Practice Partner: Reza Mousavi Director General of AI Centre Sweden

Submitted: 28.04.2024

Word count: 13170

Acknowledgements

I would like to thank my thesis supervisor, Prof. Dr. Kai Wegrich, and my practice partner, Reza Mousavi, for their support in writing this thesis. I would also like to acknowledge the contribution of my interview partners, whose input was pivotal to my research.

Table of Contents

Executive Summary

The thesis investigates the Swedish public sector as an Artificial Intelligence (AI) and Automated Decision Making (ADM) systems user. It contributes to the under-researched area of AI and ADM adoption within specific contexts of public sector organisations (PSOs). The aim is to understand how public sector AI and ADM projects can and shall be realised and whether specific guidance exists on achieving the desired outcome while remaining transparent and accountable.

The thesis performs a two-phase comparative case study of the Skatti Chatbot adopted by the Swedish Tax Agency (Skatteverket) and the Robotic Process Automation (RPA) by the Trelleborg Municipalitiy's Labour Market Agency (LMA). A developed theoretical framework serves as a base for examining the requirements PSOs should fulfil to meet algorithmic transparency, the conceptualisation of transparency and accountability, and possible relationships between the two principles in a PSO's setting.

The thesis reveals that despite Sweden's leading role in promoting AI ethics, there is a lack of consistency in organisations' approaches at the different levels of the Swedish public sector in addressing the requirements of algorithmic transparency. Further, there is no uniform relationship between transparency and accountability between organisations. The thesis further identifies necessary management tools to promote the organisational fulfilment of algorithmic transparency and institutional capabilities for achieving a direct and positive relationship between transparency and accountability.

The findings imply that although attempts are made to explain the built-in absence of transparency introduced by AI and ADM systems, more consistent practices across different levels of the Swedish public sector are needed to optimise the transparent and accountable use of those systems.

List of Abbreviations

ADM	Automated Decision Making
AI	Artificial Intelligence
FAccT	Fairness, Accountability, and Transparency in Computing
GDPR	General Data Protection Regulation
HLEG	The High-Level Expert Group on AI
LMA	Trelleborg Municipality’s Labour Market Agency
PSOs	Public sector organisations
RPA	Robotic Process Automation

List of Figures

Introduction

Whether Artificial Intelligence (AI) is appropriate for the public sector is being debated. Nevertheless, despite its critics, the use of AI is growing worldwide, allowing for enhanced performance. Public sector organisations (PSOs) are increasingly adopting AI and Automated Decision Making (ADM) systems, which are often considered as sub-categories of AI (Kaun, 2022), to facilitate and automate their decision-making processes (Berman et al., 2024; De Bruijn et al., 2022; Sousa et al., 2019).

Nevertheless, in contrast to other IT developments, AI use raises ethical concerns where the importance of transparency and accountability has been reaffirmed in the most recent worldwide reviews of AI ethics (Corrêa et al., 2023; Fjeld et al., 2020). PSOs try to address those principles through the way they adopt AI (Selten & Klievink, 2024; Busuioc, 2021; Grimmelikhuijsen & Meijer, 2022; Wirtz et al., 2019). Nevertheless, little attention has been given to those processes, as the public sector is primarily seen as an enabler or regulator of AI rather than a user (Kuziemski & Misuraca, 2020).

This led to a lack of knowledge on how AI adoption facilitates the principles of transparency and accountability, with the public sector as an AI user, considering organisations' respective contexts and processes (Chapinal-Heras & Díaz-Sánchez, 2023; Kaun, 2022b; Kuziemski & Misuraca, 2020; Madan & Ashok, 2023; Medaglia et al., 2023; Zuiderwijk et al.,2021).

Therfore, this thesis aims to contribute to the under-researched area of AI adoption in specific contexts of PSOs by examining selected AI projects in the Swedish public sector. Sweden has been chosen as the country of interest due to its focus on the responsible use of AI for societal benefit and the Swedish public sector's role in constructing national AI policies (European Commission, 2019b; 2023, OECD, 2018).

The thesis will perform a comparative case study of two purposively selected AI and ADM projects in the Swedish public sector: Skatti Chatbot, adopted by the Swedish Tax Agency (Skatteverket), and Robotic Process Automation (RPA), adopted by the Trelleborg Municipality's Labour Market Agency (LMA). This will allow the comparison of relatively similar projects in different organisational contexts.

A transparency and accountability theoretical framework has been created and applied to answer the following main research question and the sub-question:

1) **How do Swedish public sector organisations address the requirements of algorithmic transparency when adopting AI systems?**
    a) **What is the relationship between transparency and accountability created by public sector organisations when adopting AI systems?**

The thesis starts by conceptualising the term AI to include ADM systems, which will be referred to throughout the thesis. This is followed by outlining the acquired definition of transparency and accountability. Next, the background and literature review sections are presented, culminating in identifying Sweden as the focus of the thesis. Subsequently, a theoretical framework conceptualising transparency and accountability and the relationship between them is presented. Following this, the methodological approach is outlined, describing the case selection process and the methods used to examine them. Subsequently, the contexts of the two selected cases are presented, followed by the analysis section. The findings from the analysis section are later explored in the discussion segment, where policy implications are outlined. Next, policy recommendations are presented, focused on improving PSOs' achievement of algorithmic transparency and facilitating a positive and direct relationship between transparency and accountability. Afterwards, the recognised limitations of the thesis are outlined, suggesting avenues for further research. The thesis ends with a conclusion.

Conceptualisation of AI

The European Commission's Joint Research Center, in its 2020 report (p. 7), recognised that a universally agreed-upon definition of what constitutes AI does not exist; four years later, this is still the case. The High-Level Expert Group on AI (HLEG) tries to overcome the necessary simplification of the concept of intelligence needed to define AI by concentrating on rational AI, measuring against a standard of optimal performance (ibid).

HLEG defines AI as: "**software (and possibly also hardware) systems designed by humans(2) that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions**" (ibid, p.9).

This broad definition captures the various techniques PSOs employ when applying AI tools. Based on the system's capabilities, it categorises AI into reasoning and decision-making and learning and perception (ibid, p.11). The first category involves converting data into knowledge by transforming real-world information into formats machines can understand, use and develop decisions through a structured planning process (ibid). The second category operates without any rules, focusing on learning (ibid).

Further, this study considers ADM systems beyond the general concept of AI, **which could act as a category under the first area of AI as distinguished by the AI Watch**

(Cobbe et al., 2021). In defining ADM systems, this thesis mainly refers to Diakopoulos (2020, p. 198), who describes those tools as using algorithmic procedures to come up with a decision. He further argues that those systems display characteristics of AI to the extent that they assist in decision-making processes typically performed by humans, recognising them as “composites of nonhuman actors woven together with human actors into complex sociotechnical assemblages” (ibid).

Therefore, considering the multidisciplinary umbrella of AI research (Corrêa et al., 2023), this thesis understands AI to include technologies designed to automate decision-making processes and replicate intelligent behaviour, encompassing fields such as software development and robotics. Defining AI strictly as “systems capable of learning” is avoided not to exclude a whole category of non-learning (rule-based) systems that can still operate intelligently and autonomously (Corrêa et al., 2023, p.10).

Definition of Transparency and Accountability

Following scholars such as Bovens et al. (2014) and McGee (1980), the thesis recognises no universally agreed-upon definition of transparency or accountability.

For the thesis’s research needs, transparency and accountability are approached from the perspective of a relation. Here, transparency is described as “**an institutional relation between an actor and a forum**” (Bovens et al., 2014, p.512) where **“(…) an actor is rendered transparent to another actor”** (Boven, 2010, p. 946). Consequently, in terms of PSOs, transparency involves revealing internal processes, otherwise hidden to external observers, thereby validating the effective operation of an organisation (Moser, 2001, p. 3).

Considering accountability, the thesis will follow the approach of public administration scholars of an “accountability by mechanism”(Boven, 2010; 2014; Romzek & Dubnick, 1987). This mechanism considers accountability primarily as an administrative mechanism where, **through an institutional relation, one agent and institution can hold another agent or institution accountable** (ibid).

Background and Literature Review

The literature review initially provides an overview of the public sector's AI adoption landscape. Subsequently**,** it justifies why the thesis focuses on the principles of transparency and accountability. This is followed by a discussion of the various strategies public sector entities employ in adopting AI technologies within their organisations. It culminates with identifying Sweden as the focus of the thesis's research, where the justification for its choice as an ideal country for case study selection to investigate how public AI projects can and should be realised is provided.

AI Adoption Landscape by the Public Sector

The public sector progressively relies on algorithms for decision-making processes (Kaun, 2022a; Kaun & Taranu, 2020). AI has started to be seen as a symbol of efficiency and progress within its operations (ibid). This marks a significant change since the period of "AI winter", when, between 1987 and 1993, there was no interest in exploring the possibilities of AI (Corrêa et al., 2023, p. 1). Following the recent OECD report (2019), we are entering a transformative phase in which the public sector is committed to integrating AI into policymaking and service planning. The report predicts that in the upcoming years, almost one-third of public servants' tasks will be substituted by technology (ibid, p. 3).

Academics, however, warn that uncontrolled and unmonitored delegation of public authority to machines could erode human rights and the principles of legal governance (Liu et al., 2019). The challenge comes from AI systems' built-in absence of transparency, as decision-making rules automatically emerge often without a possibility to trace their origin, introducing a **"technical black box problem"** (ibid, p.134).

To counteract this, De Bruijn et al. (2022, p.1) recognise efforts to make AI decisions more acceptable to the public by exploring **Explainable AI (XAI)**, where the **"black box" challenge** is addressed through accessible explanations of the process behind AI's workings.
Nevertheless, it comes with challenges. It assumes that the public has the necessary expertise and depends on a particular societal context, which often leads to offering unsatisfactory explanations and failing to enhance transparency (ibid). Therefore, what may at first appear as a straightforward task for the public sector to clarify the functioning of an algorithm to the public is usually more complex (ibid).

Further, as algorithms gain more autonomy and fade into the background, achieving "trustworthy" AI in various processes becomes more challenging, highlighting the necessity of identifying ethical principles that should guide its adoption (Corrêa et al., 2023, p. 9; European Commission, 2019a; Liu et al., 2019).

Why Transparency and Accountability?

Following academic works on ethical principles to guide the development of "trustworthy AI," an overall trend can be identified: accountability emerged as the primary focus of documents released in 2017(Corrêa et al., 2023, p. 9), while transparency (including explainable AI/XAI) surfaced in 2018 as the central area of interest (ibid). The importance of those two principles is confirmed in the most recent worldwide review of guidelines and recommendations for AI ethics by Corrêa et al. (2023), who observed that governmental bodies globally are primarily concerned with the demand for transparency in systems (89.5%), with transparency mentioned 165 times. On the other hand, accountability, listed in the first five most common principles, was mentioned 134 times (ibid). **Organisations try to promote the principles of transparency and accountability through the way they adopt AI** (Busuioc, 2021; Grimmelikhuijsen & Meijer, 2022; Selten & Klievink, 2024; Wirtz et al., 2019).

Different Modes of AI Adoption

AI adoption involves incorporating new and varied knowledge by developing fresh capabilities, technologies and training programs (Ashok et al., 2016, p. 1008).

For AI adoption, organisations create separate departments for data science teams or integrate data science teams into already existing operational departments (Selten & Klievink, 2024). In the case of the first approach, the organisation's technical proficiency and capacities are enhanced (ibid). However, their ability to scrutinise new technologies or explain the processes behind a particular decision remains limited (ibid). In the latter's case, the conformity between AI and the primary processes is better, but technological advances are compromised (ibid). Further, the public sector can act as a smart buyer or co-developer of existing AI solutions or collaborate as a co-designer through public-private partnerships to develop new or customised AI remedies (Hello, World, 2019, p. 24). Nevertheless, such a cross-sectoral collaboration creates a challenging environment to craft a detailed contract that effectively manages the acquisition of services while minimising risk due to the absence of well-established markets and standards (ibid, p. 129).

Therefore, for PSOs, this uncertainty introduced by AI leads to a trade-off between balancing the efficient management of existing operational procedures and the challenges of integrating technological advancements (Selten & Klievink, 2024). Public organisations depend on formal and standardised bureaucratic mechanisms, emphasising short-term time frames and gradual enhancement to guarantee legitimacy in decision-making processes (Bekkers et al., 2011; Borins, 2001; Gopalakrishnan & Damanpour, 1994). Consequently, they need to balance their existing identity with the requirements posed by AI innovation (Selten & Klievink, 2024).

Despite the existing guidelines and frameworks on ethical principles establishing general limits for the use of AI introduced by supernational bodies and governments, their adoption at meso and micro levels of the public sector remains challenging to capture (Madan & Ashok, 2023). **AI scholars emphasise the necessity of translating the practical aspects of adopting the widely accepted principles of AI into the specifics of "what" and "how"** (Madan & Ashok, 2023, p. 2; Samuel, 1960).

Further, recent literary reviews summarising the latest academic writings on the implications of adopting AI in the realm of the public sector demonstrate a **lack of comprehensive understanding of how AI systems take place in the public sector, with the public sector as an AI user** (Chapinal-Heras & Díaz-Sánchez, 2023; Kaun, 2022b; Kuziemski & Misuraca, 2020; Madan & Ashok, 2023; Medaglia et al., 2023; Zuiderwijk et al.,2021). **To enhance the understanding of organisations' respective contexts and processes facilitating the ethical principles must be considered** (ibid).

Therefore, further empirical research exploring different methods of AI adoption in specific contexts of PSOs, as well as how they manage the risks introduced by AI, is needed to make the processes more transparent and accountable to the public (Ben Rjab & Mellouli, 2019; Dignum, 2018; Zuiderwijk et al., 2021).

Case Study: Sweden

The thesis focuses on Sweden for its case study selection. Sweden is considered one of the European leaders of digital innovation, with extensive use of AI, and it has utilised its findings to advocate for robust social welfare (European Commission, 2023; OECD, 2018, p.2). With its decentralised model of government, where both the government and the public sector hold a pivotal position in modelling AI policies and regulations, Sweden strives for responsible utilisation of AI technology for societal benefit while minimising potential risks (ibid).

This creates an environment that encourages a shared understanding of AI with a multistakeholder approach (European Commission, 2019b). **It allows the individual actors to assume responsibility for AI's ethical development and application while following governmental frameworks and incentives** (ibid). **This approach makes Sweden an ideal case study for investigating how public AI projects can and should be realised** (ibid).

Continuous transparency and openness are fundamental to Swedish democratic processes (Erkkilä, 2012; Swedish Institute, 2024). The country pioneered the adoption of legislation for access to information in 1766 and has a well-established tradition of providing public access to authority and public sector-generated data (ibid; Vestin et al., 2023).

Therefore, the thesis will investigate the Swedish public sector as an AI user, focusing on public service delivery. **The aim is to understand how the public sector AI projects can and shall be realised and whether specific guidance exists on achieving the desired outcome while remaining transparent and accountable.**

The thesis will now introduce a developed theoretical framework to guide the analysis of AI projects within the Swedish public sector.

Theoretical Framework[1]

This section lays out the theoretical framework used to analyse how AI adoption in the public sector **should address the challenge of organising transparency and, to a lesser extent, accountability**.

Firstly, algorithmic transparency is identified as particular to AI systems. Subsequently, requirements for achieving algorithmic transparency are outlined, and a normative framework is introduced where the requirements are structured into layers, claiming the basic standard needed to ensure transparency when adopting AI systems in a public sector setting. Next, the theoretical framework introduces different types of transparency information followed by transparency information pathways, which influence the type and quality of information disclosed by PSOs (Diakopoulos, 2020). Afterwards, the understanding of accountability is further elaborated, with accountability forums and forms introduced, followed by a section explaining the possible relationships between transparency and accountability in a public sector setting.

Transparency in AI Systems

Considering AI's "technical black box problem" (Liu et al., 2019, p. 134), there is no such thing as "full transparency"; all we can achieve is context-specific and carefully engineered algorithmic transparency (Diakopoulos, 2020, p.199)**. Algorithmic transparency is, therefore, the most basic level of transparency concerning AI systems** (ibid). Through it, sufficient information can be generated for the algorithms to be governed, facilitating accountability while allowing the user to understand the process behind a model's workings (Barredo Arrieta et al.,2020; Diakopoulos, 2020; Haresamudram et al., 2023).

Decisions about which information should be disclosed and how to deliver it to various audiences should be based on a **human-centred design process that allows for adjustment to particular AI projects, overcoming a communication challenge** (Diakopoulos, 2020). The focus should be to ensure that those who are supposed to get the necessary information

[1] The structure of the theoretical framework was inspired by Lajla Fetic's (2021) work. While drawing on this foundation, its structure was adapted to investigate the specific AI projects within the public sector entities.

on AI systems receive and understand it (ibid). **Ethical concerns determine what needs to be disclosed, establishing layers of transparency** (ibid).

Following Diakopoulos (2020), for algorithmic transparency, four layers of transparency must be disclosed:

a) information that an algorithmic process is in operation,
b) the level and nature of human involvement,
c) data used in training and operating the system, and
d) the algorithmic model and its inferences.

The following section explains the meaning behind each transparency layer and what they entail regarding AI system adoption in a public sector setting.

Layers of Algorithmic Transparency

a) Algorithmic Process in Operation

The first transparency layer mandates disclosing that the AI system is in use to the user (Diakopoulos, 2020). Services using AI should be transparent about it, even if AI is used only for some parts of their decision-making (ibid).

b) The Level and Nature of Human Involvement

Secondly, transparency should help identify human influence and responsibility within AI systems to lead to accountability (Diakopoulos, 2020). Individuals can and should be held accountable for the actions and decisions of AI systems (Johnson & Verdicchio, 2019). Academic literature distinguishes four subcategories when analysing the level and nature of human involvement in AI systems (Diakopoulos, 2020).

Firstly, the design decisions and intentions should be acknowledged. This includes the possible bias introduced by the individuals responsible for creating algorithms (Kraemer et al., 2011). Their views on what is ethical and unethical are mirrored in their decisions, influencing the design process of an AI system (Diakopoulos, 2020; ibid). To enhance transparency, algorithms should explicitly communicate their error bias, disclosing their tendency for false positives or false negatives (Kraemer et al., 2011). Additionally, if applicable, thresholds used by the models should be disclosed to enhance transparency and prevent further bias (Ranard et al., 2024).

Secondly, human involvement in a system's design, operation and management should be transparent (Diakopoulos, 2020). Automation does not simply replace human efforts but transforms them (Parasuraman et al., 2000). Transparency concerning human involvement can support a human-centred automation approach as individuals are often involved throughout AI's operations, reviewing its suggestions or stepping in during automation failures (Billings, 1997; Diakopoulos, 2020). Additionally, AI systems may behave in unexpected ways (ibid). Therefore, even in the case of full automation, the human role should be considered and disclosed (ibid).

Thirdly, the organisation's goal and intent behind adopting an AI system and recognising uses beyond its intended scope should be transparent to mitigate unforeseen biases as the system's context develops (Diakopoulos, 2020).

Lastly, sharing the contact details of individuals tasked with the engineering, upkeeping, and supervision of an AI system could heighten their sense of responsibility and accountability (Diakopoulos, 2020). Following Diakopoulos and Friedler (2016), every AI system requires an individual with the power to address its adverse impacts on people and society. Sharing individual contact details establishes mechanisms for correction, open conversation and capacity for internal improvement (ibid).

c) Data Used in Training and Operating the System

The third transparency layer concerns the transparency of the data used in training and operating an AI system, which is essential to avoid potential bias (Diakopoulos, 2020). If data is biased, the model is biased too (ibid).

Important features that should be disclosed include the standards organisations adopt to document and disclose the data they use (ibid). Diakopoulos (2020, p. 202) further highlights the importance of the disclosed datasets' quality, including their accuracy, completeness, and update frequency as the crucial features.

Additionally, details about the personal data used to customise a system for an individual should be communicated, ensuring that the person has given explicit consent to use their data. (Diakopoulos, 2020).

Similarly, the identity of the entity in charge of data set maintenance, how it is being updated, and the clarification of the data's nature—whether it is public, private, or subject to any distributional licences or copyrights—should be disclosed (ibid).

Depending on how organisations adopt AI, data access and ownership should be specified in the contract (ibid). This is relevant in the case of external providers and in the event of separate departments for data science teams, as in those cases, the ability to scrutinise new technologies is limited (Hello, World, 2019; Selten & Klievink, 2024).

d) The Algorithmic Model and its Inferences

For the fourth layer of transparency, following Diakopoulos (2020), several aspects of the algorithmic models and their inferences could be disclosed to support transparency. Mitchell et al. (2019) recommend that the released AI models should be accompanied by documentation of their performance. The documentation should include descriptions of the model's features, type, justification for its selection and any assumptions or constraints concerning it, including the process and performance metrics (Diakopoulos, 2020; Mitchell et al., 2019).

Transparency Information Types

Apart from transparency layers, there are also different types of transparency information, which can be distinguished into the transparency of the outcome and the transparency of a system's process (Diakopoulos, 2020).

The first type examines AI systems' outputs—“the what,” while the second analyses their processes and governance—"the how" (ibid, p. 199). **Human-centred design methods are necessary to consider the end user and their requirements for the disclosed transparency information type**, as the chosen type can either enhance or constrain their ability to understand and question the processes behind an AI system in a given context (ibid). Additionally, the type and quality of available information are influenced by different pathways of transparency (ibid).

Transparency Information Pathways

Transparency information pathways can be distinguished into **demand-driven**, **proactive** and **forced transparency** (Bovens et al., 2014; Diakopoulos, 2020, p. 200; Fox, 2007). The different pathways influence what kind of information the organisation reveals, which can either support or constrain transparency.

Demand-driven transparency represents an institutional promise to meet users’ requests for a particular type of information that would not be available otherwise (Fox, 2007, p. 665). It can expose shortcomings or mismanagement of an organisation, facilitating user expectations and tools enabling public access to information (Diakopoulos, 2020; Bovens et al., 2014; Fox, 2007). Nevertheless, following this pathway, public organisations can still strategically select which information to disclose (Bovens et al., 2014).

Proactive transparency involves PSOs actively publishing their actions and achievements, providing information on how transparency is achieved by following their rules and focusing on accuracy (Diakopoulos, 2020; Fox, 2007, p. 665). It may be susceptible to manipulation, compromising its accountability support (Diakopoulos, 2020). However, it can still prompt ethical considerations otherwise overlooked by the actors involved, providing a tool to access information (ibid).

Forced transparency refers to situations where information about an organisation becomes publicly available without the organisation's intention, either through leaks or external audits (Diakopoulos, 2020; Fox, 2007). This means that details they might have preferred to keep private are exposed to the public (ibid).

The thesis will now introduce the concepts of accountability forums and forms before explaining the possible relationships between transparency and accountability in a public sector setting.

The Accountability Forums and Forms

Public officials and organisations encounter diverse accountability contexts, each dictating a different form of accountability based on the specific forum they must answer to (Bovens et al., 2014). Accountability forums are, therefore, actors or institutions, internal or external, to whom an account is reported by the PSOs or its employees (ibid).

Scholars such as Bovens et al. (2014) and Romzek and Dubnick (1987) distinguish five accountability forums with which PSOs can be faced: **political, hierarchical**, **administrative, professional and social**.

In examining the Swedish public sector, this study will explore four forums of accountability. Political accountability does not apply to these organisations, as they do not undergo electoral processes (OECD, 2023).

A hierarchical accountability forum constitutes a component of a hierarchical structure within a bureaucratic organisation, where the more senior individuals in an organisation are part of the forums and the organisation's focus is aligned with the priorities of the senior leadership (Bovens et al., 2014; Romzek & Dubnick, 1987). They often entail close supervision dependent on the relationship between the superior and a subordinate, based on **vertical accountability**, where higher authorities oversee lower-level officials (Reddick et al., 2020**;** Romzek & Dubnick, 1987**)**.

In an **administrative accountability forum,** an external party controls the organisation (Romzek & Dubnick, 1987). The forums can consist of administrative entities and regulatory bodies (Bovens et al., 2014). Agreements between parties are based on contracts, and expectations might also be grounded in legal standards and regulations (Romzek & Dubnick, 1987). It entails a **vertical form of accountability** that also introduces legal accountability, where the forums can constitute judicial entities, including courts, prosecutors, or judges (Bovens et al., 2014).

As part of **the professional accountability forums**, the process of account giving is directed towards peers in an organisation, as well as to "professional bodies of oversight" (Bovens et al., 2014, p.11). Organisation's management trusts their employees and expects them to take full responsibility for their actions (Romzek & Dubnick, 1987). The basis for authority comes from within the agency, and expectations are based on professional norms and standards (ibid). It is categorised as a **horizontal form of accountability**, ensuring stakeholders mutually regulate and respond to each other (Bovens et al., 2014; Reddick et al., 2020).

Lastly, **social accountability forums** create an opportunity for stakeholders affected by the doings of PSOs to respond through actions such as creating interest groups or providing information to other regulatory bodies, creating **horizontal accountability** (ibid).

After exploring the concepts of transparency and accountability, the section will investigate the potential relationships between them.

The Relationship Between Transparency and Accountability

Transparency is expected to enhance accountability (Bovens et al., 2014). Nevertheless, the academic knowledge of how this works is limited (ibid). Following Diakopoulos (2020) and Fox (2007), **even though a discussion about an AI system is only possible when there is transparency**, **transparency, on its own, does not ensure accountability.** Therefore, a positive relationship between transparency and accountability cannot be assumed, and it should be investigated through empirical research (Bovens et al., 2014; Hood, 2010). To analyse this, the thesis will use the insights from Meijer (Bovens et al., 2014, pp. 507-524) to **identify theoretical relationships between transparency and accountability**, classifying those relationships as **direct**, **indirect**, and **inverse**. Further, Fox's (2007, p. 668) **findings will be employed to match those relations with specific institutional capabilities facilitating them**. This will allow to determine how PSOs create particular relationships between transparency and accountability.

Direct: Promoting Horizontal Accountability

A direct and positive relationship promoting horizontal accountability, perceived as the most effective form of a relationship, results from the accountability of an organisation to citizens and stakeholders (Bovens et al., 2014; Fox, 2007; Meijer, 2007). It indicates a positive relationship between **proactive or demand-driven transparency** and **accountability**, where citizens and stakeholders are offered chances to judge the organisation's actions (Bovens et al., 2014). It is characterised by no formal structures for providing information and no official sanctions (Meijer, 2007). Following Fox (2007, p. 668), this kind of relation is facilitated by **institutional "answerability"**, where PSOs can disclose existing data, conduct investigations and gather information on actual institutional conduct.

Nevertheless, limited evidence supports citizen accountability in states with advanced accountability structures, as few citizens use the information to hold public organisations accountable (Bovens et al., 2014).

Indirect: Promoting Vertical Accountability

An indirect relationship is initiated by a third party alerting the vertical accountability forums upon noticing any inappropriate behaviour by public officials or organisations (McCubbins & Schwartz, 1984). It implies a positive relationship between **proactive and demand-driven** transparency and accountability (Bovens et al., 2014). Similarly to the direct relationship, it is **facilitated by institutional "answerability"** (Fox, 2007, p. 668).

Inverse: Transparency Reduces the Need for Accountability

In an inverse relationship, transparency simplifies accountability to sharing the public organisation's performance, often limited to specific areas of the organisation's activity, with the expectation that public scrutiny will motivate the desired conduct (Bovens et al., 2014; Erkkilä, 2012; Meijer, 2007). It encourages the perception that "the numbers tell the full story," where the need for performance discussion between the actor and the accountability forums disappears (Bovens et al., 2014, pp. 513-514). Following Fox (2007, p. 668), this kind of relationship is **facilitated by the institutional capability of "disseminating and accessing information"** by the PSOs, where only selected information is disclosed.

The presented theoretical framework will guide the investigation of AI adoption in the public sector in Sweden to answer the following research question and sub-question:

1. **How do Swedish public sector organisations address the requirements of algorithmic transparency when adopting AI systems?**
   a) **What is the relationship between transparency and accountability created by public sector organisations when adopting AI systems?**

The insights from Diakopoulos (2020), Fox (2007), and Bovens et al. (2014) were suitable as the basis for the theoretical framework of this thesis, as the academic works are concerned with challenges regarding transparency and accountability in AI systems in the public sector.

The thesis will now move on to outline the methodological approach.

Methodological Approach

This section introduces the methodological approach implemented by the thesis, explaining how the data for analysis has been obtained. Firstly, the application of a case study method is justified. This is followed by outlining the data collection strategy, which involves conducting desktop research before selecting AI projects for a detailed analysis. Subsequently, a justification for incorporating semi-structured interviews as part of the detailed analysis is provided, followed by an outline of the semi-structured interview guide.

Adoption of Case Study Method

To answer the research question, the thesis employed a qualitative comparative approach cross-sectional design where practical insights were obtained by examining real-world AI projects in the Swedish public sector. The thesis is, therefore, case-oriented, incorporating a la Weber approach to comprehend complex units of PSOs (Della Porta & Keating, 2008). To achieve this, the thesis followed a two-phase qualitative method. Initially, existing public sector AI projects were outlined by conducting desktop research. This allowed for identifying a smaller number of cases, which were then investigated in greater detail through semi-guided expert interviews.

Data Collection

Desktop Research

In its initial stages, the research looked at different AI and ADM projects in Sweden. For this purpose, the EU AI Watch, OECD AI Policy Observatory and the IPS Survey databases were investigated as they contain the most up-to-date and comprehensive information on the current AI projects. The most noteworthy source was the explorable dataset of 686 AI and ADM projects in the public sector across Europe provided by IPS Survey (2024), which outlined 23 AI and ADM cases realised by the Swedish public sector, of which 16 were selected for further analysis. The source was valuable, comprising information on the technology used and the responsible PSOs (IPS Survey, 2024).

The search was performed by selecting Sweden as the country of interest in the mentioned databases and reviewing Swedish PSOs' outlined examples of AI and ADM projects. The search was restricted to the projects listed in the mentioned databases, and projects done by the central government, local governments, NGOs, and academic research in Sweden were considered. **An inconsistency was noted when researching the databases, as they often failed to distinguish between AI and ADM projects, potentially resulting from a lack of a universally agreed-upon definition of what constitutes AI.**

The initial 16 cases (see Appendix I) were pre-selected based on their brief descriptions of their use and role in supporting public service delivery processes. The author qualitatively assessed their relevance to the stated research question.

Selection of Cases for Detailed Analysis

The 16 selected cases were processed for further analysis. For each case, a systemic survey was conducted, analysing a wide range of available primary and secondary sources, including the organisations' websites, news articles, press releases, official testimonies, and compulsory data analysis reports, to select a small number of cases for a detailed investigation.

As the study aims to contribute to an under-researched area of the public sector as an AI user, further analysis was aimed at capturing the adoption of AI at the meso and micro levels by the Swedish public sector. In case of limited, old information or no further findings, the cases were no longer considered for further analysis.

The further analysis resulted in two purposively selected cases which have been chosen following Gerring and Seawright's (2008) most similar approach where the first one being- a Skatti Chatbot adopted by the Swedish Tax Agency (Skatteverket) operating on a national--meso level and the second Robotic Process Automation (RPA) by the Trelleborg Municipalitiy's Labour Market Agency (LMA), operating on the local-micro level. As the subsequent analysis will reveal, both projects are described as "co-workers", supporting PSOs' daily processes, and are adopted via a vendor solution, allowing for the comparison of relatively similar projects in a different organisational context.

Semi-Structured Expert Interviews

In the second analysis phase, the data collection was based on expert interviews for a detailed investigation of the two purposively selected cases. The interviews were semi-structured with pre-defined open-ended questions, enabling the interviewer to explore particular themes more deeply (Kallio et al., 2016). The developed interview guides (see Appendix V) provided a framework facilitating comparisons across different interviews, allowing for exploring complex topics where nuanced views of the participants were desired (ibid). The sample was selected using purposive sampling logic, selecting participants with particular knowledge and experiences relevant to the thesis's research (ibid). The engagement of different stakeholders in both cases allowed for a higher level of data validity (Walsham, 2006).

To investigate Chatbot Skatti's AI case (overview of the interviews: see Appendix VI), interviews were conducted with the Chatbot Skatti product owner (Interview F), the Chief Data Scientist (Interview E), and an employee at Skatteverket (Interview D). For examination of the RPA case of Automated Social Welfare Decisions, interviews were conducted with a Digital Business Developer (Interview A), a Unit Manager at the Department of Welfare (Interview B) from the Municipality of Trelleborg, and an academic engaged in research on automation in the public sector (Interview C). Overall, six expert interviews (video interviews) lasted around 30 minutes. The interviews were recorded and transcribed.

Semi-Structured Interview Guide

The interviews were comprised of five parts (see Appendix V):

1. Firstly, experts described their role in the AI adoption process to facilitate the interpretation and classification of their statements.
2. Secondly, the process of AI adoption within the organisation was enquired. This section also investigated the technology's intended use, including the rationale behind its design and adoption decisions.
3. Thirdly, the organisation's transparency practices were investigated. The section also examined requirements for algorithmic transparency.
4. Subsequently, experts discussed the accountability mechanisms within their organisations.
5. Lastly, details on the data governance mechanisms were investigated to explore data disclosure practices and the policies governing data use.

Data Analysis

For analysis purposes, a reference table on algorithmic transparency and accountability (ATA), based on the theoretical framework, was created to support the analysis of the selected AI projects (see Appendix II). The aim behind developing the tool was multifaceted. First, it allowed to differentiate between different AI adoption modes incorporated by the organisations and the potential reasons behind them. Secondly, the table facilitated the assessment of if and how the organisations fulfilled the requirements for algorithmic transparency, allowing the investigation of whether the four layers of transparency, as outlined by the theoretical framework, were present. Thirdly, it allowed to investigate and compare the transparency information types and pathways created by the organisations, followed by the investigation of the adopted accountability forums and forms. Fourthly, the outlined information allowed for assessing the institutional capabilities exercised by the organisations, which in turn facilitated the assessment of the relationship between transparency and accountability created within the organisations' specific contexts.

Before moving to the data analysis section, the thesis introduces the organisational context and the specifics of the two selected projects. Firstly, Skatteverket and its Chatbot Skatti are presented, followed by the LMA's RPA.

Introduction to the Cases

Skatteverket

Organisational Context

Skatteverket is an autonomous public authority operating on a national level that has been working with AI for over twenty years (Int. E, personal communication, March 2024). The organisation follows the principle of "ensuring sustainable AI"(Skatteverket, 2021b, p.1). It aims to build confidence in the technology among stakeholders while reducing ethical concerns associated with AI (ibid). Its practices comply with accountability and transparency outlined by the FAccT[2] community. Further, it proactively incorporates ethical principles into its operational frameworks, following recent developments in the field, including the AI Act (Jfokus, 2023).

To oversee AI's sustainable development and application, the organisation has established an internal Council for Sustainable AI (Skatteverket, 2023). Before adoption into the organisation's operations, all AI systems are inspected for sustainability, security, legality, and ethics and to understand the implications of adopting AI into the tax authority's operations (ibid).

There is a preference within the organisation to develop AI projects in-house over external providers (int. D, personal communication, March 2024). Nevertheless, it remains open to collaboration based on organisational standards and resource constraints (Skatteverket, 2021a, p. 4).

Project

Skatti is a text-to-text system resembling a chat (see. Fig 1 and 2). It is intended to handle non-personal FAQs with a repository for many prepared responses to customer intentions (Skatteverket, 2024a). It was first adopted in 2017 and is continuously used by the agency and trained by contact centre agents in tax-specific dialogue (ibid).

[2] FAccT, stands for Fairness, Accountability, and Transparency in Computing, it is a research community focused on ensuring a fair and transparent AI (ACM, 2024).

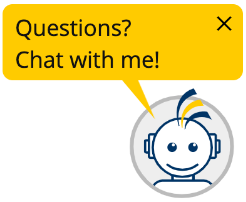

Fig.1 Screenshot of Skatteverket's Chatbot (Skatteverket, 2024a).

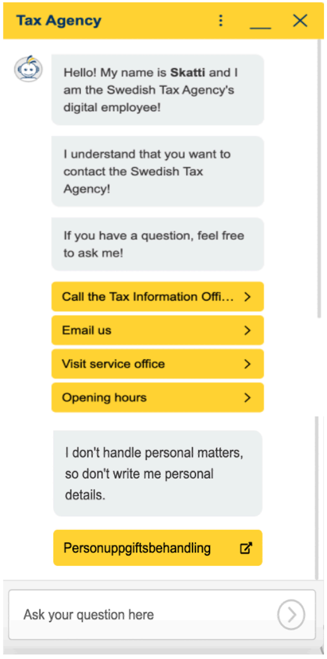

Fig.2 Screenshot of Skatteverket's Chatbot Skatti after clicking the chatbot's icon (Skatteverket, 2024a).

It is a complex model designed to categorise questions to match them with predefined answers (ibid). Skatti uses probability thresholds to assess the likelihood that a given question corresponds to a predefined answer (ibid).

Adoption Mode

Skatti has been adopted through a vendor solution (Jfokus, 2023). The vendor, Boost.ai, facilitates the training platform and Skatteverket acts as a smart buyer of technology from the private sector (ibid).

Following interviewee F (personal communication, March 2024), the motivation behind opting for a vendor solution was to: "gain experience and try out the technology so that Skatteverket could remain at the forefront when it comes to digitalisation".

Skatteverket oversees all data and chatbot training, while Boost AI provides consulting, training, and tools for optimal platform use (ibid). Nevertheless, following Skatteverket's Chief Data Scientist (personal communication, March 2024), this solution led to a lack of full transparency regarding the algorithm in the early stages of adoption.

LMA

Organisational Context

LMA is a local PSO of the Trelleborg Municipality. In 2017, it adopted a fully automated ADM system, making Trelleborg the first municipality in Sweden to adopt this technology (European Commission et al., 2020; Kaun, 2022; Ranerup & Henriksen, 2022).

The organisation faced regulatory uncertainty when introducing the system following its Digital Business Developer's comment: **"The Swedish Law was not ready for it"** (personal communication, March 2024).

Further, LMA lacks specific internal guidelines for transparent and accountable use of the ADM system, adhering to existing national legal standards within Sweden, which introduced uncertainty at the time of the system's adoption (ibid).

Consequently, the system was short in operation due to various challenges, including legal and ethical issues (Lind, 2020), a lack of internal expertise and increased maintenance costs (ibid).

Project

The RPA model optimised the re-application process for social benefits (European Commission et al., 2020; Kaun, 2022). It reviewed applications on the organisation's case management system, ProCapita, to assess the citizens' financial status and action plan (Ranerup & Henriksen, 2022). The system operated on a decision tree model, functioning similarly to a flow chart (Kaun, 2022; Kaun & Velkova, 2019). It allowed for the automation of structured and repetitive digital tasks by imitating human interaction with software applications (ibid). It compared specific variables against databases to make informed decisions (ibid).

Even though the system was broadly considered as a "fully automated decisions on applications for social benefits" (Kaun, 2022, p. 2048), the LMA's employees referred to it as "automated handling of applications and considered rule-based algorithms as decision support systems rather than automated robots to which tasks are fully delegated"(Kaun, 2022, p. 2053).

The process yielded a verdict that was either positive, partially positive or negative (ibid). However, the procedure justifying negative decisions, which were more complicated, was fully established only in 2019, after the robot had already been in operation- showcasing its rather complex nature (ibid).

Adoption Mode

Due to a lack of in-house expertise, the organisation initially depended on an external provider, Valcon Consultants. The external provider managed the system's deployment, maintenance and update on the organisation's online platform (Int. A, personal communication, March 2024). A contract and regular follow-ups regulated close collaboration (ibid).

In 2020, this shifted to relying on a specialist internal co-worker within the department for welfare support to manage the RPA (ibid). Nevertheless, as the co-worker left the organisation in September 2022, LMA found it challenging to organise a replacement, leading to an eventual shutdown of the RPA model (ibid). This showcased insufficient internal knowledge even after the system was operated internally. Additionally, the technology was not as effective as the organisation had primarily anticipated, however, it lacked sufficient know-how to assess that when adopting the model (ibid).

After introducing the organisational context and the specifics of the selected AI projects, the thesis will now move to the analysis section.

Analysis[3]

This section presents a comprehensive analysis of the two selected AI projects following the theoretical framework created in the previous sections. Insights from semi-structured interviews and primary and secondary sources reviewed serve as a base for the analysis and to address the research question. Firstly, the theoretical framework is applied to each case, analysing how and if the organisations addressed the requirements of algorithmic transparency when adopting AI. Subsequently, it examines the type of transparency and accountability created to assess the relationship between the two principles. A previously developed ATA table was used to support the analysis. Firstly, the analysis of the Skatti Chatbot adopted by Skatteverket will be presented, followed by the RPA model adopted by LMA.

Applying the Theoretical Framework: Chatbot Skatti

This section provides an in-depth analysis of how the process of AI adoption in Skatteverket addressed the challenges of algorithmic transparency. Subsequently, it examines the types and pathways of transparency created, followed by accountability forums and forms. Lastly, the relationship between transparency and accountability is assessed following the developed ATA.

Algorithmic Transparency

Chatbot Saktti appears to encompass the majority of the essential layers for algorithmic transparency, revealing that AI is in use, aspects of the level and nature of human involvement, elements of the information on the data used, and, to some extent, the details on the algorithmic model in use.

The information communicating the use of the algorithmic process is disclosed firstly on Skatteverket's website (see Fig. 3) and secondly when clicking on the chat (see Fig. 2) to start a conversation with Skatti.

[3] In the initial stage, the interviews were analysed with the help of Deep-L response to: "What were the underlying intentions and contextual reasoning behind this answer," to reduce bias in the results' interpretation. OpenAI, April 10, 2024.

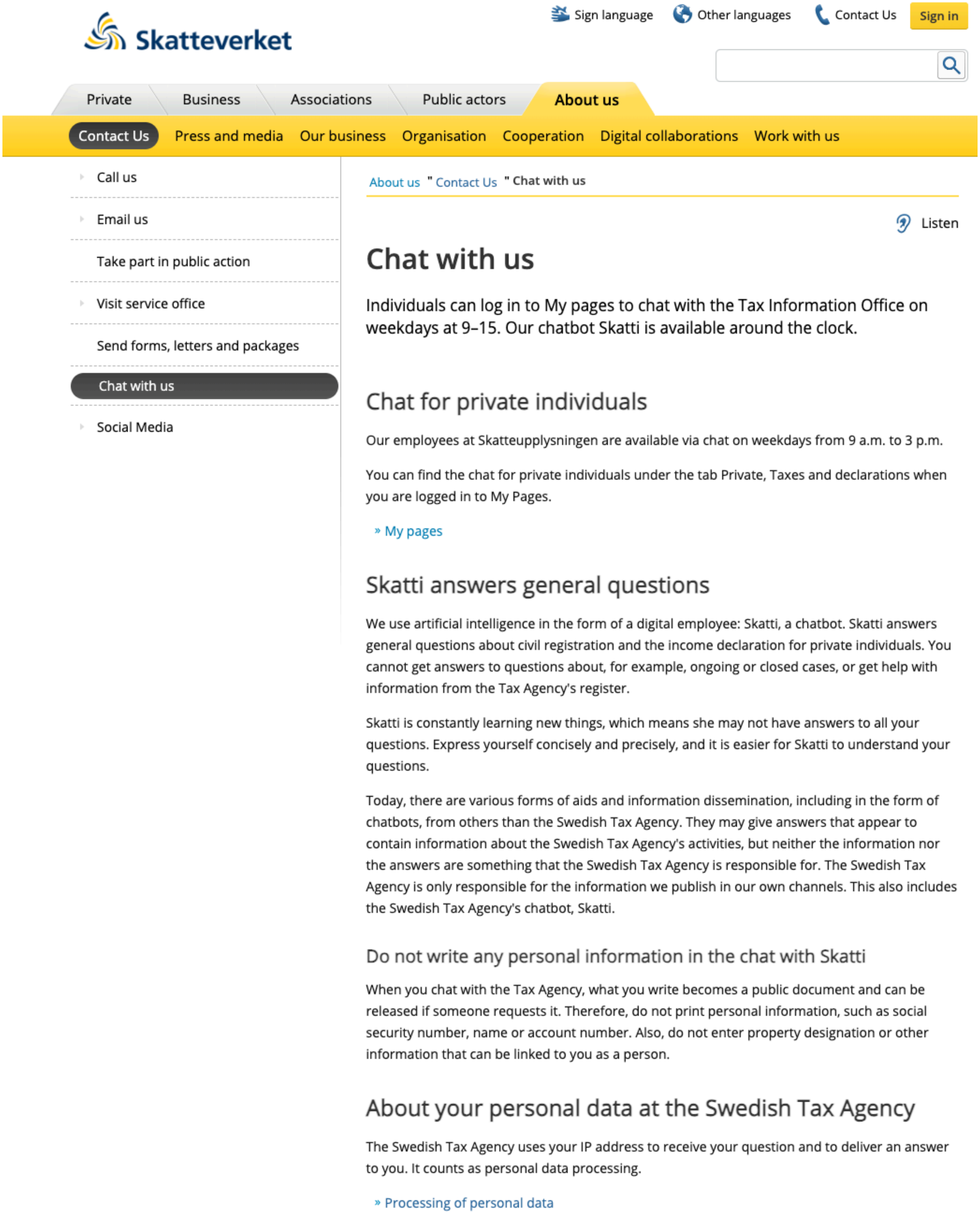

Fig. 3 Screenshot of Skatteverket's website, "chat with us" webpage (Skatteverket, 2024a).

Regarding the organisation's design decisions and intentions, Skatteverket's guidelines recognise that AI solutions may intentionally or unintentionally exhibit biases resulting from organisational culture or the developers' actions (Skatteverket, 2021b). The organisation commits to transparently reporting the identified biases and measures to control them (ibid). Following the Product Owner of Skatti (personal communication, March 2024), the chatbot is constructed to be able to recognise its limitations by responding with “I don’t know” in case a question asked does not confidently match its responses, and no bias has been recognised in its work. Nevertheless, the interviewee recognises that personal queries may result in accidental biases, which could cause incorrect responses (ibid). Despite the risk, the organisation does not disclose the model’s probability thresholds (Int. E, personal communication, March 2024), potentially limiting the recognition of biases these thresholds may cause (Ranard et al., 2024).

Concerning human involvement in the system's design, operation, and management, interviewees E and F emphasised that humans control the data training process through active selection, analysis, and division of training data across different themes (personal communication, March 2024). This ensures a manually initiated training process that is subject-oriented (ibid). Therefore, human-in-the-loop is an integral part of the system to ensure compliance with legal standards, suggesting that humans may step in when automation is insufficient. Nevertheless**, the extent of human involvement in Skatii's operations has not been disclosed.** Users may, however, enquire about it, although this is not commonly done (Int. E & F, personal communication, March 2024).

The organisation’s goal and intent behind the model were first designed to aid Skatteverket's customer support by becoming an FAQ system (AI Sweden, 2019; Bertrand & Phillips, 2022). However, its focus has shifted to enhancing accessibility and addressing the needs of broader audiences (Int. E, personal communication, March 2024). This enabled the organisation to engage with new customer groups (ibid). Following its guidelines, Skatteverket **informs about the scope of the AI system and warns against undesired uses outside of its scope** to prevent inappropriate uses (Skatteverket, 2021b, 2024a, 2024c, 2024e)(see Fig 2 and 3).

Regarding the responsible individuals for the model, interview partners highlight that the first point of contact is the organisation itself (Int E, F and D personal communications, March 2024). Following a sent inquiry, it is decided internally who should address the question and how (ibid). **No contact is provided to an identified individual responsible for the AI model;** however, on Skatteverket’s website, a selection of contact forms to streamline inquiries is available (Skatteverket, 2024b). Each contact form is tailored to specific requests. However, no contact form particular to AI-related inquiries has been identified.

Concerning the standards Skatteverket adopts to document and disclose the data it uses, Skatteverket’s Chief Data Scientist (personal communication, March 2024) explained that collected data is anonymised to protect user privacy and meet the organisation's data protection standards. The Product Owner of Chatbot Skatti complemented it by adding that conversations can be reviewed for 366 days, after which they are automatically deleted according to data retention protocols (personal communication, March 2024). However, based on ongoing observations and analyses of actual use, The Chief Data Scientist (personal communication, March 2024) highlights that there is room for improvement as the data could be anonymised “more effectively”, possibly to prevent cross-referencing incidents. While Skatteverket prioritises transparency, legal constraints limit access to personal and sensitive data when externally requested (ibid). This includes data classified as secret or sensitive (ibid). Granting potential access would require a case-by-case evaluation based on the data type and relevant laws (ibid)

Acknowledging the importance of the dataset’s quality, including its accuracy, completeness, and update frequency, Skatteverket’s employees frequently update the dataset through interactive manual processes, ensuring its high standard (Int. E, personal communication, March 2024).

The users are informed about personal data used on the organisation’s website (Int. E & D, personal communication, March 2024). The specific types of personal data collected are outlined, along with the reasons for their collection, the legal framework governing their use, and how the principle of public disclosure influences the agency's handling of personal data (Skatteverket, 2024c).

Following Skatteverket’s data policy, specific personal data processing agreements are in place for external vendors tasked with handling personal data for the organisation (ibid). **The organisation retains accountability for the personal data and sets the terms for how these external service providers manage the data** (ibid). Nevertheless, some ambiguity arises concerning data access regulations of the external vendors as the data is hosted on the Swedish cloud service "Qleura," which is outside of the organisation’s direct control (Int. E, personal communication, March 2024).

When disclosing the algorithmic model and its inferences, Skatteverket guidelines mandate to inform about how an AI solution was developed, its purpose, description of the model’s features and justification for the model’s selection and design (Skatteverket, 2021b). **Nevertheless, as noted by interviewees E and D, the process and performance metrics are not openly communicated to the public** (personal communication, March 2024). However, internally, detailed documentation on the AI system is available (ibid). The AI's deployment versions are documented, enabling internal traceability that captures the model's evolution and training data (ibid). This enables the tracking of changes over time (ibid). **Nevertheless, in the case of external providers, interviewees E, F and D indicate uncertainty concerning the specifics of disclosing AI information, as Skatteverket’s data scientists are limited in what they can fully disclose** (personal communication, March 2024).

Transparency

Following Skatteverket's guidelines for sustainable AI development, transparency is defined as a multifaceted concept (Skatteverket, 2021a). It ties the use of AI to transparency, accountability, and fairness (ibid). The interviewees highlight a commitment to transparency, ensuring that **all stakeholders can comprehend the purpose of the AI system and are aware of its errors and biases** (Int. E & F personal communication, March 2024). **The information is adjusted depending on the expertise of the enquiring party** (ibid).
The focus is on **proactive transparency**, as the information is shared openly rather than when requested (Diakopoulos, 2020; Fox, 2007). Although the organisation provides general contact details, its transparency is mainly realised through its active guidance in disclosing information on its website.

The organisation's transparency focuses **on the process ("the how") as well as on the outcome ("the what")** (Diakopoulos, 2020, p. 199). This is visible through its commitment to reporting recognised biases and means to mitigate them, available information on the scope of the AI system and uses outside of its scope, and available information on how the AI solution is developed and justification for its selection. However, regarding Saktti, their ability to disclose specifics on AI information may be compromised due to reliance on the external provider.

Accountability

Following its guidelines, Skatteverket takes responsibility as an organisation for any AI being developed and assigns officials to manage AI ethical and social risks (Skatteverket, 2021a). Even in the case of a vendor solution, the agency assumes the ultimate responsibility (ibid). Internally, accountability for AI systems is assigned to specific departments and leaders who act as the organisation's accountability forum, such as the head of the "customer meeting department" responsible for Skatti (Int F, personal communication, March 2024). An "AI product owner" is identified as the key individual in charge of an AI product like Skatti, reporting and overseeing its success and data management standards to the forum (ibid).

This arrangement incorporates elements of professional and social accountability forums, where the "professional bodies of oversight" in an organisation are part of the forums—as head of the departments—and where the organisation's management trusts its employees and expects them to take full responsibility for their actions, in the case of the AI product owners (Bovens et al., 2014; Romzek & Dubnick, 1987). Further, stakeholders affected by the chatbot Skatti have the possibility to respond to it, which introduces the social accountability forum. There are no sanctions, compensations or remediations in place as, in the case of Skatti, the relationship with the vendor is regulated through a contract introducing another mechanism to resolve issues (Int. F personal communication, March 2024). **Through different accountability forums, a horizontal form of accountability is created**.

Relationship Between Transparency and Accountability

The analysed case indicates **a positive relationship between proactive transparency and accountability**, as citizens and stakeholders are offered opportunities to judge the organisation's actions and outcomes via information available on its website. **The mechanism of institutional answerability** is in place, as the organisation can disclose the existing data on the AI system to the users who can enquire via the general contact information provided, indicating **a direct relationship promoting horizontal accountability** (Bovens et al., 2014; Fox, 2007).

Applying the Theoretical Framework: RPA Model

This section provides an in-depth analysis of how LMA addressed the challenges of organising algorithmic transparency when adopting the ADM system. Subsequently, it examines the types and pathways of transparency created, followed by accountability forums and forms. Lastly, the relationship between transparency and accountability is assessed following the developed ATA.

Algorithmic Transparency

The adoption process of RPA did not encompass all the essential layers for algorithmic transparency: it failed to disclose that an algorithmic process was in use, and some aspects of the level and nature of human involvement, as well as elements of the algorithmic model in use, were not fully disclosed.

Following interview partner C (personal communication, April 2024**), the organisation's website did not initially communicate that an algorithmic process was in use**. This omission might have been based on regulatory uncertainty and a lack of internal expertise—especially when collaborating with the external provider (ibid). At some point, the algorithm was disclosed; however, there have been challenges in making it understandable or accessible to the different stakeholders (ibid). Additionally, due to the perception that humans were involved in the decision-making process and the RPA did not have the final say, there was no perceived need for additional disclosure beyond the information released by the organisation to the press, implying acceptance of some lack of transparency in the workings of the RPA (ibid; Ranerup & Henriksen, 2019). Therefore, applicants were often unaware of the automated processing of their applications, except that they received a decision quicker (Kaun, 2022).

Regarding the organisation's design decisions and intentions, Runerup and Henriksen (2022), who studied the RPA model, discussed the system's tendency to provide partially positive verdicts when they should have been positive- revealing a tendency for false negatives. **The bias, however, has not been communicated, likely because the human**

element in the decision process was seen as a check against such errors (Int. A personal communication, March 2024).

Concerning human involvement in the system's design, operation, and management, interviewees highlighted a **human-in-the-loop approach** (Int. A, B and C personal communication, March and April 2024**)**. Although the RPA model aimed to enhance efficiency, it still relied on human oversight to ensure the integrity of the final decisions (ibid). The system continued to be monitored daily to ensure it was functioning correctly- it generated reports of failed applications, which caseworkers would then review (ibid).
However, as the system became well-established by 2019, the caseworkers were mainly engaged only with more complex applications, allowing the more straightforward cases to be fully automated (Int. C, personal communication, April 2024). The Unit Manager at the Department of Welfare in Trelleborg (personal communication, March 2024) further clarified that if the decision was positive, the process was fully automated; if there were issues with some decisions, a case worker initially reviewed those applications. Over time, the need for human intervention at the end of the process decreased as the RPA became more reliable in handling applications, including those turned down (ibid). **The level and nature of human involvement in the RPA model's work were not disclosed to the applicants**. The extent of human involvement in the process has been seen as superior to the technology used (Ranerup & Henriksen, 2019). There was no perceived need to disclose the ADM system's role in the process, nor how individuals were involved in operating it, even after RPA became well established and conducted the final decisions independently (ibid).

The organisation's goal and intent behind the model were part of a broader initiative to enhance the management of social assistance cases (Int. C personal communication, March 2024). The introduction of the technology was supposed to improve interactions between clients and caseworkers, allowing for more time to be dedicated to job-seeking support (ibid). Additionally, one of the primary goals of introducing RPA was cost efficiency (ibid). **However, due to a lack of algorithmic process disclosure and insufficient internal knowledge, the system's intended and unintended scope was not communicated**.

Regarding the responsible individuals for the model, interviewees A and B **(**personal communication, March 2024) highlight the Department of Welfare Support and Customer Service role, which could be contacted in case of inquiries about the RPA **system**. The communication between applicants and caseworkers occurred via mail, phone, or the organisation's online platforms (Ranerup & Henriksen, 2022).

Following interviewee B (personal communication, March 2024**), the data used in training and operating the system, as commonly understood in the context of AI, was not relevant to it as the model was "rule-based" rather than "learning-based"** (ibid). Although the RPA processed data gathered from individual applicants using the organisation’s online platform, it was designed only to handle data pertinent to rule-oriented procedures (European Commission et al., 2020).

The applicants were informed about the extent of personal data used- they had to consent for their details to be reviewed during the reapplication process (Int. B, personal communication, March 2024). This step was a compliance measure with General Data Protection Regulation (GDPR) (ibid).
As noted by interviewee B (ibid), the contract allowed the external provider **t**o access the system where applicants’ information was stored. This access was conditional and tied to the duration of the contract (ibid). Therefore, ultimately, LMA was responsible for the data (ibid). The cooperation with the external provider was disclosed (ibid), however, it was not depicted as an entity responsible for data management. Nevertheless, despite being responsible for the dataset management, LMA did not have sufficient expertise to provide details regarding the data used and relied on the external provider (int. A, personal communication, March 2024). When the management of the system shifted to operate internally, the company relied on the one qualified employee in this area (ibid). **Therfore, while the municipality had oversight over the data and the algorithm, it lacked the in-house expertise to maintain control over the data and the system**. This was reflected when experts, including scholars and specialists, were denied access to the system’s data, programming, or algorithms to evaluate its performance (AIAAIC, 2018). Freedom of information requests were also denied, citing trade secrecy grounds (ibid).

In the case of disclosing the algorithmic model and its inferences, the organisation had an annual review of goals (Int. C, personal communication, April 2024). The results of these reviews were made available externally (ibid). These included information on how many people were assisted by the application system and how effective it was in achieving its objectives, allowing anyone to understand the basis on which decisions for reapplications were made. However, details about the model's functioning or the model itself were not disclosed (ibid). Interviewee C (ibid) emphasises that while technical details were not part of the public disclosure, technical information could have been obtained from the LMA upon request. The LMA’s Digital Business Developer (personal communication, March 2024), however, further clarifies that the organisation did not have sufficient knowledge to fully describe how the algorithm functioned in detail, even when the RPA was managed internally, as only a single individual oversaw the whole process. Similarly, constraints were noted regarding a detailed description of the model and justification for its selection and design.

Transparency

Due to a lack of specific internal guidelines, LMA does not have a definition of transparency. It follows European and Swedish law, according to which the decisions concerning the ADM system should be open and transparent[4] (Int. B, personal communication March 2024).

The applicants were provided an understandable explanation of their application results, disclosing the decision and its rationale (ibid). The RPA's rules were considered complex, and only basic explanations for the decision and next steps were readily offered (ibid). **However, in none of the cases, including the organisation's annual reviews, algorithm-specific functions were readily disclosed** (Int C, personal communication April 2024).

This practice illustrates adherence to the transparency of outcomes, where the information on what is required to be eligible for welfare benefits is publicly accessible through the organisation's website (Int. B and C, personal communication March and April 2024)- **indicating a form of proactive transparency, where LMA made an attempt to make the outcome transparent and understandable to individuals via initiating the disclosure of information**.

Accountability

Within the organisation, each department has a board responsible for its actions (Int. B, personal communication, March 2024). Therefore, when collaborating with the external provider and later when the municipality's employee developed the expertise to operate RPA, a board within the LMA's specific department was always responsible for the ADM system (ibid). The organisation disclosed its board's reports as a form of institutional accountability- where various processes and decisions were made public and could be scrutinised (Int. C, personal communication April 2024).

This structure is analogous to the concept of **hierarchical and social accountability forums**. There is a hierarchical structure within the organisation, introducing **vertical accountability**, with the departments' boards bearing the ultimate responsibility for the actions of its units. Although there was no available information on aligning the goals with the priorities of the senior leadership, there were examples of close supervision through regular follow-ups by LMA's board when collaborating with the external provider. When the model was operated internally, the actions of the responsible co-worker were overseen by the board of the specific department, where higher-level authorities provided oversight over a lower-level official.

[4] The GDPR includes provisions on ADM systems; providing for the individuals to have the right to understand how algorithms make decisions (Goodman & Flaxman, 2016) and to have the possibility for these decisions to be checked and regulated by a human (Dreyer et al., 2019; Kaun, 2022).

Additionally, as the reports were publicly disclosed, the social accountability forum was introduced, creating **horizontal accountability**.

Relationship between transparency and accountability

Primarily, the case indicates an **inverse relationship between transparency and accountability**. The understanding of human-in-the-loop and oversight over the RPA's model's decisions was perceived as a sufficient accountability measure, making the need to disclose RPA by the organisation redundant. Additionally, the applications' outcomes were generally accepted by the public, reducing the need for deeper accountability. **The organisation's institutional capability of disseminating and accessing information** was perceived as sufficient, eliminating the need for discussion between an actor and the accountability forum and restricting transparency to selected information disclosed (Bovens et al., 2014).

Nevertheless, the interest in this case, being the first fully automated ADM system among the Swedish Municipalities, resulted in various stakeholders requesting access to the information on the system (AIAAIC, 2018), which could lead to the relationship between transparency and accountability to **transform into an indirect relationship where the demand-driven information could create accountability**. However, this would require the institution to develop the institutional capability of answerability. Nevertheless, as the system was the first of its kind, there seemed to be insufficient knowledge about how to work with the system to create such a capability, leading to the refusal of such requests or incorrect sharing of information (AIAAIC, 2018).

The thesis will now move on to discuss the findings derived from the analysis and their implications.

Discussion

The thesis investigated the Swedish public sector as an AI user, examining in detail two AI projects, the Chatbot Skatti and RPA Welfare Support, to answer the main research question: 1) **How do Swedish public sector organisations address the requirements of algorithmic transparency when adopting AI systems?**, and the following sub-question: **a) What is the relationship between transparency and accountability created by public sector organisations when adopting AI systems?** The section provides a discussion of the findings and their implications.

1) **How do Swedish public sector organisations address the requirements of algorithmic transparency when adopting AI systems?**

Applying ATA uncovered that complete algorithmic transparency was not achieved by the investigated AI projects. However, attempts were made to explain the built-in absence of transparency introduced by the AI systems. The analysis reveals **variation in outcome** as significant differences were noted between the micro- and meso-level organisations in providing information on the disclosure of the algorithmic processes, the level and nature of human involvement, the data, and the algorithmic model in use.

This makes it challenging to compare the two cases as elements of algorithmic transparency, such as data, were not considered by the micro-level organisation. Therefore, not every layer of transparency required for algorithmic transparency could have been compared. **The results suggest a need for a better exchange of best practices between PSOs operating on different levels.**

Possible reasons behind Skatteverket's better fulfilment of the required transparency layers in comparison to LMA were the existence of internal organisational guidelines and guiding principles of assuring sustainable AI. Further, the organisation also established an internal Council for Sustainable AI to oversee AI's development and application. This resulted in open communication about whether an algorithmic model was in use, potential bias, scope and out-of-scope uses of the model, bias communication, and information on the AI model in use.

However, the analysis revealed that despite Skatteverket's commitment to disclose bias, thresholds used by the model were not revealed, potentially limiting its integrity. Through thresholds, algorithms determine how and when a specific decision is made (Ranard et al., 2024). Public organisations might, therefore, increase their transparency by making these kinds of measures publicly available and improve public engagement by making it easier for the public to understand how their inquiries are categorised, allowing them to phrase their questions more effectively. Additionally, in the case of an external provider, this could make the cooperation more transparent.

The analysis further supported academic claims that reliance on an external provider limits the organisation's ability to oversee and fully understand a system's workings. In both cases, although a detailed contract was in place, the service's acquisition still introduced a degree of unpredictability due to the lack of internal knowledge. Further, the case of LMA shows that even when the model operation moved to be realised in-house, through reliance on a single employee, the lack of internal knowledge remained, shifting the dependency from an external provider to a single co-worker.

Additionally, despite the uncertainty, both organisations maintained accountability over personal data when cooperating with an external provider. However, to avoid ambiguity regarding data management, the case of Skatteverket highlighted the importance of understanding how exactly data is stored, which was not considered in the case of LMA. Therefore, the analysis suggests that in addition to detailed contracts, organisations need to build their know-how before engaging with an external provider to oversee their activities meaningfully to avoid uncertainty regarding data management and the model in operation.

Both cases failed to disclose the level and nature of human involvement to the public. Disclosing such information would allow for a more human-centred automation approach, where the stakeholders could understand the process behind adopted AI systems more easily. Additionally, neither organisation provided specific details to identify individuals responsible for the AI systems. Including such information could facilitate inquiries regarding the systems and allow to address potential issues more promptly. Further, to aid transparency, the analysis also shows that the performance metrics should be made available for the general public in addition to the model's description, features, and selection criteria.

Therefore, when considering the research question of How do Swedish PSOs address the requirements of algorithmic transparency when adopting AI systems?, **the findings evidence that there is a lack of consistency in the approach of organisations at the different levels of the public sector. Nevertheless, establishing tools such as internal guidelines and an internal body to oversee AI development and applications significantly supports the fulfilment of algorithmic transparency, facilitating accountability.**

a) **What is the relationship between transparency and accountability created by public sector organisations when adopting AI systems?**

The analysis revealed, in both of the studied cases, the **information pathway of proactive transparency**, **where those organisations disclosed selected information proactively rather than when requested**. This resulted in varying levels of algorithmic transparency achieved by the two organisations, implying that focus on other types of informational pathways should be considered to ensure uniform results among the organisations, as following Diakopulos (2020), proactive transparency is susceptible to manipulation, compromising its accountability support.

Further, the organisations were characterised by varying accountability contexts, where Skatteverket created professional and social accountability forums with a horizontal form of accountability and LMA created hierarchical and social accountability forums with vertical and horizontal forms of accountability.

With different institutional capabilities in place, where Skatteverket developed the capability of answerability and LMA of access and dissemination of information, this culminated in various forms of relationships between transparency and accountability within the organisations.

The analysis recognised Skatteverket's direct and positive relationship with horizontal accountability in place. However, only a limited number of interactions with the public were recorded, and more data is needed to prove this relationship. Nevertheless, it may also support Bovens et al.'s (2014) claim that in states with advanced accountability structures, few citizens use the available information to hold public organisations accountable.

The analysis classified LMA's relationship between transparency and accountability as inverse, as in this case, the human oversight over the RPA's model's decisions eliminated the need for discussion between an actor and the accountability forum. An indirect relationship was initiated as various stakeholders started to request access to information regarding the model. However, the organisation failed to develop the capability of answerability due to the lack of internal know-how to facilitate the indirect relationship.

Therefore, the analysis suggests that the ability of an organisation to develop the capability of answerability is crucial to promote a direct and positive relationship between transparency and accountability, which, following Boven et al. (2014), is the most effective form of a relationship.

To answer the sub-question of What is the relationship between transparency and accountability created by public sector organisations when adopting AI systems?**, there is no uniform relationship between transparency and accountability between the organisations, it is dependent on the institutional capabilities of the organisations.**

The next section presents selected policy recommendations derived from the discussion.

Policy recommendations

Based on the discussion, policy recommendations can be suggested for improving algorithmic transparency when adopting AI by PSOs and facilitating a positive and direct relationship between transparency and accountability.

1) **Introduce public-sector innovation labs.**

To minimise disparities in knowledge of AI adoption between the meso and micro levels of PSOs, support should be provided so that organisations can experiment with AI and run tests before introducing new technologies. The Swedish Agency for Digital Government (DIGG) could oversee this as part of the exchange of knowledge and promotion of best practices concerning AI in the public sector, enabling organisations to develop the answerability capability.
Further, the public-sector innovation labs could complement the regulatory sandboxes proposed by the AI Act, expected to be in effect at the end of 2024 (European Parliament & Council of the European Union, 2021). The innovation labs could provide opportunities for

AI research and development, supporting experimentation with new models of AI technologies that are not yet ready for regulatory testing. This could further enhance the development of appropriate regulatory approaches for the new technologies to avoid instances when the law is unclear about how to manage them.

2) **Develop internal ethical guidelines for the use of AI.**

The development of internal guidelines could support internal expertise on what requirements AI adoption processes should fulfil and would allow for more uniform practices between different levels of the public sector. Further, they would make it easier for an organisation to stay up-to-date with new regulations and policies regarding the use of AI. To ensure a human-centred approach and the fulfilment of the requirements for algorithmic transparency, the guidelines should ensure that:

a) The level and nature of human involvement behind an AI solution are disclosed.
b) Contact details to specific individuals responsible for the AI system are provided.
c) The model's performance metrics and thresholds (when applicable), in addition to the model's description, features and selection criteria, are disclosed.

3) **Establishment of an internal body to oversee AI development and application, as well as the organisation's decision to engage with an external technology provider.**

The establishment of an internal body to oversee AI would allow for improved scrutiny of AI systems developed in-house or through a vendor solution before their adoption in the workings of an organisation. This would support the AI system's ethics, legality and conformity with the organisation's guidelines.

4) **Ensure internal know-how on the technology adopted via a vendor solution and a detailed contract.**

Organisations need to build their know-how before engaging with an external provider to ensure they can oversee and fully understand the workings and data management mechanisms of a system adopted. This will allow for meaningful learning from this experience and the possibility of realising the AI solution in-house, ensuring the expertise is not within a single employee. This will further support assessing whether a particular AI solution is adequate for a problem at hand.
Internal expertise could be further facilitated through DIGG's enhanced initiatives to promote already existing knowledge-sharing projects on AI, such as AI Sweden (AI Sweden, n.d.), and the creation of new ones particular to PSOs operating on different levels.

5) **Apart from proactive information pathways, demand-driven and forced transparency should be encouraged to support a direct and positive relationship between transparency and accountability.**

Demand driven transparency could be introduced by encouraging the public to learn more about AI and making the information easily accessible to them through the introduction of tools such as online request portals and readily available information on the organisations' websites. Forced transparency could be introduced via external audits focused on ensuring uniform standards in AI adoption across the public sector. Audits could be performed by DIGG. Such initiatives could further encourage transparency of the process and outcome of an AI project.

Limitations

This section critically assesses the study's chosen approach, theoretical framework and methodology to ensure the appropriate understanding of its findings. Subsequently, avenues for further research, building on this study, are identified.

The study's approach to generalising the term AI to include ADM systems may limit its potential, introducing generalisability and ignoring differences between the two types of systems.

The theoretical framework provides a strong foundation for understanding the dynamics of transparency and accountability in a public sector setting. However, the selected standards to define the terms of transparency and accountability introduce simplification and potential bias, which may disturb the thesis's conclusions. Similarly, distinguishing between different types and forms of transparency and accountability may introduce complexity, making applying the framework to real-life examples challenging. Additionally, focusing on transparency and accountability simultaneously limits the capacity to investigate each concept in greater detail.

The qualitative approach integrating semi-structured interviews could have introduced bias as the interpretive nature of the approach and the purposive selection of the case studies were based on the researcher's own reasoning, potentially negatively affecting the impartiality of the study (Della Porta & Keating, 2008). The selection of the case studies and further focus on specific AI projects in particular context settings hinder the potential transferability of the results. It also introduces bias when trying to generalise the findings to the Swedish public sector as a whole. Additionally, selecting the cases chosen for a detailed analysis may introduce bias based on the researcher's own interpretation. Moreover, the chosen approach to compare meso- and micro-level organisations and comparing cases where one AI project was no longer in operation could have introduced differences not accounted for by the research.

Therefore, further studies evaluating the AI projects by the public sector in Sweden and Europe will likely be useful in assessing how these organisations address the ethics of transparency and/ or accountability through their adoption of AI systems and exploring the relationship between these principles on a larger scale. Additionally, future studies could distinguish between AI and ADM systems and focus on examining them separately. Similarly, micro- and macro-level organisations could be approached individually to account for the potential differences between them. Finally, future research could introduce mixed methodology approaches to minimise interpretation bias. This could be done by combining case studies and a content analysis approach, where the context of particular organisations could be investigated together with the content analysis used to examine public communications and policy reviews concerning AI or ADM initiatives.

Conclusion

The thesis contributed to the under-researched area of AI adoption in specific contexts of PSOs and their management of risks posed by AI, aiding in making the process more transparent and accountable to the public. Further, the thesis supported exploring the principles of transparency and accountability and their relationship in the AI adoption process. It augmented the understanding of how PSOs' contexts may limit the enhancement of the principles when adopting AI technologies.

The created transparency and accountability framework allowed to answer the main research question of 1) how the Swedish public sector organisations address the requirements of algorithmic transparency when adopting AI systems, **the thesis revealed a lack of consistency in organisations' approach at the different levels of the Swedish public sector. Nevertheless, establishing tools such as internal AI guidelines and an internal body to oversee AI development and applications–significantly supports the fulfilment of algorithmic transparency, facilitating accountability.**

Additionally, by addressing the sub-question of a) What is the relationship between transparency and accountability created by public sector organisations when adopting AI systems, **the thesis demonstrated that there is no uniform relationship between transparency and accountability between organisations; it depends on their institutional capabilities.**

Based on the analysed data, policy recommendations were derived to support PSOs' achievement of algorithmic transparency and to facilitate a positive and direct relationship between transparency and accountability. Finally, the thesis acknowledged its limitations and suggested avenues for future research.

References

ACM. (2024). *Association for Computing Machinery*. https://www.sigcas.org/events/conferences/

AI Sweden. (n.d.). *AI Sweden*. https://www.ai.se/en

AI Sweden. (2019, April 29). *Artificial Intelligence improves the Swedish Tax Agency's customer service*. https://www.ai.se/en/news/artificial-intelligence-improves-swedish-tax-agencys-customer-service

Ashok, M., Narula, R., & Martinez-Noya, A. (2016). How do collaboration and investments in knowledge management affect process innovation in services? *Journal of Knowledge Management*, *20*(5), 1004–1024. https://doi.org/10.1108/JKM-11-2015-0429

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bekkers, V., Edelenbos, J., & Steijn, B. (2011). An Innovative Public Sector? Embarking on the Innovation Journey. In V. Bekkers, J. Edelenbos, & B. Steijn (Eds.), *Innovation in the Public Sector* (pp. 197–221). Palgrave Macmillan UK. https://doi.org/10.1057/9780230307520_10

Ben Rjab, A., & Mellouli, S. (2019). Artificial Intelligence in Smart Cities: Systematic Literature Network Analysis. *Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance*, 259–269. https://doi.org/10.1145/3326365.3326400

Berman, A., De Fine Licht, K., & Carlsson, V. (2024). Trustworthy AI in the public sector: An empirical analysis of a Swedish labor market decision-support system. *Technology in Society*, *76*, 102471. https://doi.org/10.1016/j.techsoc.2024.102471

Bertrand, A., & Phillips, A. (2022, April 19). *How can harnessing tax data create value for all?* https://www.ey.com/en_cz/government-public-sector/how-can-harnessing-tax-data-create-value-for-all

Billings, C. E. (1997). *Aviation Automation: The Search for a Human-Centered Approach*. Mahwah, NJ. http://sunnyday.mit.edu/16.355/Billings.pdf

Borins, S. (2001). Public Management Innovation: Toward a Global Perspective. *The American Review of Public Administration*, *31*(1), 5–21. https://doi.org/10.1177/02750740122064802

Bovens, M. (2010). Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism. *West European Politics*, *33*(5), 946–967. https://doi.org/10.1080/01402382.2010.486119

Bovens, M., Goodin, R. E., Schillemans, T., & Meijer, A. (2014). Transparency. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford Handbook of Public*

Accountability. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199641253.013.0043

Busuioc, M. (2021). Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review*, *81*(5), 825–836. https://doi.org/10.1111/puar.13293

Chapinal-Heras, D., & Díaz-Sánchez, C. (2023). A review of AI applications in Human Sciences research. *Digital Applications in Archaeology and Cultural Heritage*, *30*, e00288. https://doi.org/10.1016/j.daach.2023.e00288

Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 598–609. https://doi.org/10.1145/3442188.3445921

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & De Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, *4*(10), 100857. https://doi.org/10.1016/j.patter.2023.100857

De Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*, *39*(2), 101666. https://doi.org/10.1016/j.giq.2021.101666

Della Porta, D., & Keating, M. (Eds.). (2008). *Approaches and methodologies in the social sciences: A pluralist perspective*. Cambridge University Press.

Diakopoulos, N. (2020). Chapter 10 Transparency. In *The Oxford Handbook of Ethics of AI*.

Diakopoulos, N., & Friedler, S. (2016). *How to Hold Algorithms Accountable*. https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/.

Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Information Technology*, *20*(1), 1–3. https://doi.org/10.1007/s10676-018-9450-z

Erkkilä, T. (2012). *Government Transparency: Impacts and Unintended Consequences*. New York: Palgrave Macmillan. https://books.google.de/books?hl=de&lr=&id=EwRaCwAAQBAJ&oi=fnd&pg=PP1&ots=dhjw1r0J31&sig=3NRm0KgHCQBLjN4Dx6cg-M_EGXI&redir_esc=y#v=onepage&q&f=false

European Commission. (2019a). *Ethics guidelines for trustworthy AI*. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission. (2019b). *National strategies on Artificial Intelligence A European perspective in 2019 Country report – Sweden*. https://knowledge4policy.ec.europa.eu/sites/default/files/sweden-ai-strategy-report.pdf

European Commission. (2023). *European Innovation Scoreboard*. https://research-and-innovation.ec.europa.eu/statistics/performance-indicators/european-innovation-scoreboard_en#:~:text=European%20innovation%20scoreboard

%202023,-Based%20on%20their&text=Other%20Innovation%20Leaders%20are%20Sweden,performing%20above%20the%20EU%20average.
European Commission. Joint Research Centre. (2020). *AI watch: Defining Artificial Intelligence : towards an operational definition and taxonomy of artificial intelligence.* Publications Office. https://data.europa.eu/doi/10.2760/382730
European Commission. Joint Research Centre. (2022). *Glossary of human-centric artificial intelligence.* Publications Office. https://data.europa.eu/doi/10.2760/860665
European Commission, Joint Research Centre, Misuraca, G., & Noordt, C. (2020). *AI watch, artificial intelligence in public services: Overview of the use and impact of AI in public services in the EU*. Publications Office. https://data.europa.eu/doi/10.2760/039619
European Parliament, & Council of the European Union. (2021). *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts [2021/0106 (COD)].* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
Fetic, L. (2021). *Public AI Registers Analyzing Mechanisms for the Transparent and Accounable Use of AI and Algorithmic Decision-Making in the Public Sector.* Hertie School of Governance.
Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3518482
Fox, J. (2007). The uncertain relationship between transparency and accountability. *Development in Practice*, *17*(4–5), 663–671. https://doi.org/10.1080/09614520701469955
Gerring, J. (2017). Qualitative Methods. *Annual Review of Political Science*, *20*(1), 15–36. https://doi.org/10.1146/annurev-polisci-092415-024158
Goodman, B., & Flaxman, S. (2016, June 23). *EU regulations on algorithmic decision-making and a"right to explanation"* [[Paper presentation]].
Gopalakrishnan, S., & Damanpour, F. (1994). Patterns of generation and adoption of innovation in organizations: Contingency models of innovation attributes. *Journal of Engineering and Technology Management*, *11*(2), 95–116. https://doi.org/10.1016/0923-4748(94)90001-9
Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of Algorithmic Decision-Making: Six Threats and the Need for a Calibrated Institutional Response. *Perspectives on Public Management and Governance*, *5*(3), 232–242. https://doi.org/10.1093/ppmgov/gvac008
Haresamudram, K., Larsson, S., & Heintz, F. (2023). Three Levels of AI Transparency. *Computer*, *56*(2), 93–100. https://doi.org/10.1109/MC.2022.3213181
Hello, World: Artificial intelligence and its use in the public sector (OECD Working Papers on Public Governance 36; OECD Working Papers on Public Governance, Vol. 36). (2019). https://doi.org/10.1787/726fd39d-en

Hood, C. (2010). Accountability and Transparency: Siamese Twins, Matching Parts, Awkward Couple? *West European Politics*, *33*(5), 989–1009. https://doi.org/10.1080/01402382.2010.486122

IPS Survey. (2024). *IPS-X Cases* [dataset].

Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, *33*(3), 371–377. https://doi.org/10.1016/j.giq.2016.08.011

Jfokus (Director). (2023). *Is the Swedish Tax Agency ready for the age of AI? by Jens X Augustsson and Andreas Voxberg*. https://www.youtube.com/watch?v=WwNiQ7_5gLI

Johnson, D. G., & Verdicchio, M. (2019). AI, agency and responsibility: The VW fraud case and beyond. *AI & SOCIETY*, *34*(3), 639–647. https://doi.org/10.1007/s00146-017-0781-9

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

Kaun, A. (2022). Suing the algorithm: The mundanization of automated decision-making in public services through litigation. *Information, Communication & Society*, *25*(14), 2046–2062. https://doi.org/10.1080/1369118X.2021.1924827

Kaun, A., & Taranu, G. (2020). Sweden / Research. In *Automating society 2020. A Report by AlgorithmWatch in Cooperation withBertelsmann Stiftung*. https://automatingsociety.algorithmwatch.org/

Kaun, A., & Velkova, J. (2019). *Automating society: Taking stock ofautomated decision-making in the EU.* (In M. Spielkamp (Ed.)). A Report by AlgorithmWatch in Cooperation withBertelsmann Stiftung, supported by Open Society Foundations. https://www.algorithmwatch.org/automating-society

Kraemer, F., Van Overveld, K., & Peterson, M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology*, *13*(3), 251–260. https://doi.org/10.1007/s10676-010-9233-7

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, *44*(6), 101976. https://doi.org/10.1016/j.telpol.2020.101976

Lind, K. (2020, March 17). Central authorities slow to react as Sweden's cities embrace automation of welfare management. *Algorithm Watch*. https://algorithmwatch.org/en/trelleborg-sweden-algorithm/

Liu, H.-W., Lin, C.-F., & Chen, Y.-J. (2019). Beyond State v Loomis: Artificial intelligence, government algorithmization and accountability. *International Journal of Law and Information Technology*, *27*(2), 122–141. https://doi.org/10.1093/ijlit/eaz001

Madan, R., & Ashok, M. (2023). AI adoption and diffusion in public administration: A systematic literature review and future research agenda. *Government Information Quarterly*, *40*(1), 101774. https://doi.org/10.1016/j.giq.2022.101774

McCubbins, M. D., & Schwartz, T. (1984). Congressional Oversight Overlooked: Police Patrols versus Fire Alarms. *American Journal of Political Science*, *28*(1), 165. https://doi.org/10.2307/2110792

McGee, M., Calvin. (1980). *The 'Ideograph': A Link between Rhetoric and Ideology. Quarterly Journal of Speech, v66 n1 p1-16 Feb 1980.*

Medaglia, R., Gil-Garcia, J. R., & Pardo, T. A. (2023). Artificial Intelligence in Government: Taking Stock and Moving Forward. *Social Science Computer Review*, *41*(1), 123–140. https://doi.org/10.1177/08944393211034087

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596

Moser, C. (2001). *How open is' open as possible'? Three different approaches to transparency and openness in regulating access to EU documents*.

OECD. (2018). *Digital Government Review of Sweden Towards a Data-driven Public Sector*. https://www.oecd.org/gov/digital-government/digital-government-review-of-sweden-2018.pdf

OECD. (2023). *Public Governance Monitor of Sweden*. OECD. https://doi.org/10.1787/086f9b89-en

OPSI. (2016). *New ways of providing social support through process automation*. https://oecd-opsi.org/innovations/9131-2/

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Ranard, B. L., Park, S., Jia, Y., Zhang, Y., Alwan, F., Celi, L. A., & Lusczek, E. R. (2024). Minimizing bias when using artificial intelligence in critical care medicine. *Journal of Critical Care*, *82*, 154796. https://doi.org/10.1016/j.jcrc.2024.154796

Ranerup, A., & Henriksen, H. Z. (2019). Value positions viewed through the lens of automated decision-making: The case of social services. *Government Information Quarterly*, *36*(4), 101377. https://doi.org/10.1016/j.giq.2019.05.004

Ranerup, A., & Henriksen, H. Z. (2022). Digital Discretion: Unpacking Human and Technological Agency in Automated Decision Making in Sweden's Social Services. *Social Science Computer Review*, *40*(2), 445–461. https://doi.org/10.1177/0894439320980434

Reddick, C. G., Demir, T., & Perlman, B. (2020). Horizontal, Vertical, and Hybrid: An Empirical Look at the Forms of Accountability. *Administration & Society*, *52*(9), 1410–1438. https://doi.org/10.1177/0095399720912553

Romzek, B. S., & Dubnick, M. J. (1987). Accountability in the Public Sector: Lessons from the Challenger Tragedy. *Public Administration Review*, *47*(3), 227. https://doi.org/10.2307/975901

Samuel, A. L. (1960). Some Moral and Technical Consequences of Automation—A Refutation. *Science*, *132*(3429), 741–742. https://doi.org/10.1126/science.132.3429.741

Seawright, J., & Gerring, J. (2008). Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options. *Political Research Quarterly*, *61*(2), 294–308. https://doi.org/10.1177/1065912907313077

Selten, F., & Klievink, B. (2024). Organizing public sector AI adoption: Navigating between separation and integration. *Government Information Quarterly*, *41*(1), 101885. https://doi.org/10.1016/j.giq.2023.101885

Skatteverket. (2021a). *Skatteverkets AI-policy [Translation from Swedish to English provided by Deep-L]*. Skatteverket.

Skatteverket. (2021b). *Riktlinje för hållbar AI [Translation from Swedish to English provided by Deep-L]*. Skatteverket.

Skatteverket. (2023). *Råd för hållbar AI - Uppdragsbeskrivning [Translation from Swedish to English provided by Deep-L]*. Skatteverket.

Skatteverket. (2024a). *Chat with us*. https://www.skatteverket.se/omoss/kontaktaoss/chattamedvardigitalamedarbetare.4.2cf1b5cd163796a5c8b403b.html

Skatteverket. (2024b). *Email us*. https://skatteverket.se/omoss/kontaktaoss/mejlaoss.4.71004e4c133e23bf6db800066246.html

Skatteverket. (2024c). *More on Skatteverket*. https://skatteverket.se/servicelankar/otherlanguages/inenglishengelska/moreonskatteverket.4.7856a2b411550b99fb780009630.html#:~:text=The%20work%20of%20collecting%20taxes,affairs%20of%20individuals%20or%20businesses.

Skatteverket. (2024d). *Processing of personal data*. Processing of Personal Data. https://skatteverket.se/omoss/varverksamhet/behandlingavpersonuppgifter.4.18e1b10334ebe8bc80005971.html

Skatteverket. (2024e). *Skatteverket: Our View of AI*. Skatteverket. https://skatteverket.se/omoss/varverksamhet/skatteverketssynpa/aiartificiellintelligens.4.1997e70d1848dabbac9363f.html

Sousa, W. G. D., Melo, E. R. P. D., Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly*, *36*(4), 101392. https://doi.org/10.1016/j.giq.2019.07.004

Swedish Institute. (2024). Free speech, free press and overall openness and transparency are key to Swedish society. *Openness in Sweden*. https://sweden.se/life/democracy/openness-in-sweden

Van Noordt, C., Medaglia, R., & Tangi, L. (2023). Policy initiatives for Artificial Intelligence-enabled government: An analysis of national strategies in Europe.

Public Policy and Administration, 09520767231198411. https://doi.org/10.1177/09520767231198411
Vestin, E., Sundberg, C., & Ribenfors, A. (2023). AI, Machine Learning & Big Data Laws and Regulations. In *AI, Machine Learning & Big Data Laws and Regulations*. Global Legal Insights. https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/sweden
Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems*, *15*(3), 320–330. https://doi.org/10.1057/palgrave.ejis.3000589
Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector—Applications and Challenges. *International Journal of Public Administration*, *42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103
Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, *38*(3), 101577. https://doi.org/10.1016/j.giq.2021.101577

Appendix

Appendix I: The 16 Initially selected cases of AI and ADM projects by the Swedish public sector. The projects were extracted from the IPS Survey (2024).

Code	**AI Project**	**Responsible Agency**	**AI Typology**
1.	AIDA - Interpreting detailed plan provisions	Örebro Municipality (Local Government)	Natural language processing
2.	Automated Social Welfare decisions	Municipality of Trelleborg (Local Government)	Cognitive Robotics, Process Automation and Connected and Automated Vehicles; Robotic Process Automation (RPA)
3.	Automation to facilitate internal processes	Södertälje municipality (Local Government)	Cognitive Robotics, Process Automation and Connected and Automated Vehicles; Robotic Process Automation (RPA)
4.	AI to facilitate internal processes	Uddevalla municipality (Local Government)	Process Automation

5.	Decentralised Energy Trading Marketplace	Business Area Markets (Central-Government)	Blockchain and AI
6.	Financial Assistance Automation	Nacka Municipality (Local Government)	Cognitive Robotics, Process Automation and Connected and Automated Vehicles; Robotic Process Automation (RPA)
7.	Kari- a chatbot for local governments	Local Governments	Conversational AI
8.	Self-driving buses roll into Gothenburg	City of Gothenburg (Local Government)	Combination of several technologies.
9.	Service: Checks for financial support - AI for checking companies requesting support	Public Employment Service (Central-Government)	Automated decision making
10.	Service: Detection causes early school leavers Kungsbacka - Analysing data of early school leavers	Kungsbacka Municipality (Local Government)	Data mining and analytics
11.	Skatti: answering system about population registration and income tax return	Swedish Tax Agency (Skatteverket) (Central-Government)	Chatbot
12.	Swedish Employment Agency using AI to better understand the labour market	Swedish Employment Agency (Central Government)	Natural language processing
13.	Swedish Land Registry (SLR)- Fostering efficiency when dealing with land registry requests	Swedish Land Registry (Central-Government)	Natural Language Processing
14.	Tengai- Robot in Recruitment Processes.	Municipality of Upplands-Bro (Local Government)	Voice and facial expression analysis

15.	The Land Registry in the blockchain – testbed	Mapping, Cadaster and Land Registration Authority (Central Government)	Blockchain and AI
16.	Tillfälligt Arbete i Sverige (TAIS) Project: automated AI and machine learning supporting tax registration for non-Swedish customers	Swedish Tax Agency (Skatteverket) (Central-Government)	Automated AI and Machine Learning

Appendix II: The reference table on algorithmic transparency and accountability (ATA) was based on insights from Diakopoulos (2020), Fox (2007), and Bovens et al. (2014) adjusting concepts related to AI adoption, transparency and accountability for the purpose of this thesis.

Code	Factor	Application
Mode of AI Adoption		
A.1	External provider.	A) Yes B) No
A.2	Separate department for the data science team.	A) Yes B) No
A.3	Integration of data science team into an already existing department.	A) Yes B) No
Requirements for algorithmic transparency		
B.1	Disclosure that an algorithmic process is in place.	A) Yes B) No
B.2	Disclosure of the level and nature of human involvement.	The design decisions and Intentions (communication of the error bias).

		Human involvement in the design, operation and management of a system. A) Is the level and nature of human involvement disclosed? B) Are humans in the loop? Do they step in during automation failure?
		Organisation's goal and intent. A) Scope of AI system adopted. B) Have the recognised uses beyond its scope been disclosed?
		Are contact details of the responsible individuals disclosed? A) How could the end-user enquire about the system?
B.3	Data used in traning and operating the system.	What are the standards adopted to document and disclose the data used? A) Can individuals enquire about obtaining access to raw data?

		What is the quality of the disclosed data? A) Data accuracy. B) Data completeness. C) Update frequency.
		Are details disclosed on the personal data used?
		Is the identity of the entity in charge of the data set maintenance disclosed (e.g., who was allowed to access individuals' data)?
B.4	Algorithmic model and its inferences.	Is information about the AI model disclosed? A) Is a detailed description of the model's features/ type disclosed? B) Has the justification for the model's selection and design been disclosed?

		Is the model self contained enough for a person to understand and reason about it?
		Are details concerning the institutional performance of AI disclosed (including model's assumptions use or constraints)?
Transparency		
C.1	Is the used definition of AI adjusted to different audiences (communication challenge)?	A) Yes B) No
C.2	What are the transparency information types?	A) Transparency of the outcomes. B) Transparency of the process. C) Not clear.
C.3	What are the transparency information pathways?	A) Demand driven. B) Proactive. C) Forced.
Accountability		

D.1	What form does the accountability take?	A) Horizontal. B) Vertical. C) Hybrid.
D.2	What are the accountability forums?	A) Hierarchical B) Administrative C) Professional. D) Social.
Institutional capabilities		
E.1	What are the institutional capabilities in place?	A) Dissemination and access to information. B) Institutional answerability. C) Sanctions, compensations and/or remediations.
Relationship between transparency and accountability		
F.1	What is the type of interconnetcedness between transparency and accountability?	A) Direct. B) Indirect. C) Inverse.

Appendix III: Applied ATA for the Skatti Chatbot adopted by the Swedish Tax Agency (Skatteverket)

Code	Factor	Application	
Mode of AI Adoption			
A.1	External provider.	A) **Yes** B) No	Motivation: to gain experience and test the technology.
A.2	Separate department for the data science team.	A) Yes B) No	Not applicable.
A.3	Integration of data science team into an already existing department	A) Yes B) No	Not applicable.
Requirements for algorithmic transparency			

B.1	Disclosure that an algorithmic process is in place.	A) **Yes** B) No	On the website, no need to enquire.
B.2	Disclosure of the level and nature of human involvement.	The design decisions and Intentions (communication of the error bias).	Recognises that there may be bias, but thresholds are not disclosed.
		Human involvement in the design, operation and management of a system. A) Is the level and nature of human involvement disclosed? B) Are humans in the loop? Do they step in during automation failure?	Human-in-the-loop is an integral part of the system- but the disclosure about human involvement is not readily available.
		Organisation's goal and intent. A) Scope of AI system adopted. B) Have the recognised uses beyond its scope been disclosed?	It provides information about the scope of the AI system and recognises undesired uses outside its scope (personal inquiries).
		Are contact details of the responsible individuals disclosed? A) How could the end-user enquire about the system?	No contact details to an individual. Following a sent inquiry, it is decided internally who should address the question and how
B.3	Data used in traning and operating the system.	What are the standards adopted to document and disclose the data used? A) Can individuals enquire about obtaining access to raw data?	Interactive manual processes were adopted to document and disclose the data. Interviewees recognise that data could be better anonymised. Individuals' potential access to data is reviewed on a case-by-case basis.

		What is the quality of the disclosed data? A) Data accuracy. B) Data completeness. C) Update frequency.	Data is accurate, complete and updated frequently.
		Are details disclosed on the personal data used?	Users are informed about the extent of the personal data used.
		Is the identity of the entity in charge of the data set maintenance disclosed (e.g., who was allowed to access individuals' data)?	Following its guidelines, Skatteverket is responsible for all the data. Specific data processing agreements regulate the access to data by external vendors. Some ambiguity arises concerning the regulations of the external vendors' data access as the data is hosted on the Swedish cloud service "Qleura," which is outside of the organisation’s direct control.
B.4	Algorithmic model and its inferences.	Is information about the AI model disclosed? A) Is a detailed description of the model’s features/ type disclosed? B) Has the justification for the model’s selection and design been disclosed?	The model is disclosed. A detailed description of the model’s features/ type and justification for model selection can be enquired about.

		Is the model self contained enough for a person to understand and reason about it?	No, It is a complex model.
		Are details concerning the institutional performance of AI disclosed (including model's assumptions use or constraints)?	Institutional performance of the model is only available internally. In the case of external providers there is uncertainty concerning the specifics of disclosing AI information, as Skattevrket's data scientists are limited in what they can fully disclose
Transparency			
C.1	Is the used definition of AI adjusted to different audiences (communication challenge)?	A) **Yes** B) No	
	What are the transparency information types?	A) **Transparency of the outcomes.** B) **Transparency of the process.** C) Not clear.	
C.2	What are the transparency information pathways?	A) Demand driven. B) **Proactive.** C) Forced.	
Accountability			

D.1	What form does the accountability take?	A) **Horizontal.** B) Vertical. C) Hybrid.	
D.2	What are the accountability forums?	A) Hierarchical. B) Administrative. C) **Professional.** D) **Social.**	
Institutional capabilities			
E.1	What are the institutional capabilities in place?	A) Dissemination and access to information. B) **Institutional answerability.** C) Sanctions, compensations and/or remediations.	
Relationship between transparency and accountability			
F.1	What is the type of interconnectedness between transparency and accountability?	A) **Direct.** B) Indirect. C) Inverse.	

Appendix IV: Applied ATA for the Robotic Process Automation (RPA) adopted by the Trelleborg Municipalitiy's Labour Market Agency (LMA)

Code	Factor	Application	
Mode of AI Adoption			
A.1	External provider.	**A) Yes** B) No	Motivation: lack of internal expertise.
A.2	Separate department for	A) Yes B) No	Not applicable.

	the data science team.		
A.3	Integration of data science team into an already existing department	A) Yes B) No	Not applicable.
Requirements for algorithmic transparency			
B.1	Disclosure that an algorithmic process is in place.	A) Yes B**) No**	It was initially limited to the press release. Later, it was disclosed; however, there were issues with the format. There was some acceptance of a lack of transparency in the workings of RPA.
B.2	Disclosure of the level and nature of human involvement.	The design decisions and Intentions (communication of the error bias).	Bias tended to false negatives, not communicated to the users.
		Human involvement in the design, operation and management of a system. A) Is the level and nature of human involvement disclosed? B) Are humans in the loop? Do they step in during automation failure?	The level and nature of human involvement were not disclosed. It was a human-in-the-lopp approach. Nevertheless, as the system became more established, there was less human involvement.
		Organisation's goal and intent. A) Scope of AI system adopted. B) Have the recognised uses beyond its scope been disclosed?	The scope of the AI system, as well as the potential uses outside of its scope, were not communicated due to a lack of algorithmic process disclosure and a lack of sufficient internal knowledge.

		Are contact details of the responsible individuals disclosed? A) How could the end-user enquire about the system?	There were no contact details to a particular individual disclosed. The Department of Welfare Support and Customer Service could be contacted via email, phone, or the organisation's online platform.
B.3	Data used in traning and operating the system.	What are the standards adopted to document and disclose the data used? A) Can individuals enquire about obtaining access to raw data?	Data use and training, as commonly understood in AI, were irrelevant to this system. Individuals could enquire about access to raw data; however, in the past, there were issues with its disclosure.
		What is the quality of the disclosed data? A) Data accuracy. B) Data completeness. C) Update frequency.	Not applicable.
		Are details disclosed on the personal data used?	Applicants were informed about the extent of personal data used.
		Is the identity of the entity in charge of the data set maintenance disclosed (e.g., who was allowed to access individuals' data)?	LMA was ultimately in charge of the data set maintenance. An external partner was not disclosed as an entity responsible for the dataset management. However, LMA did not have sufficient expertise to provide details regarding the data.

B.4	Algorithmic model and its inferences.	Is information about the AI model disclosed? A) Is a detailed description of the model's features/ type disclosed? B) Has the justification for the model's selection and design been disclosed?	Technical information could have been obtained upon request; however, the organisation did not have sufficient knowledge to describe how the algorithm functioned in detail. Similar constraints exist regarding a detailed description of the model and justification for its choice.
		Is the model self-contained enough for a person to understand and reason about it?	No, It is a complex model.
		Are details concerning the institutional performance of AI disclosed (including model's assumptions use or constraints)?	The organisation's annual review was disclosed, allowing anyone to understand the basis on which decisions were made—information about the application process rather than anything about the model itself.
Transparency			

C.1	Is the used definition of AI adjusted to different audiences (communication challenge)?	A) Yes **B) No**	An easy-to-understand, non-technical explanation of the reasons for the denial and possible remedies was provided. The algorithm was eventually disclosed; however, there have been challenges in making it understandable or accessible to the different stakeholders (ibid).
C.2	What are the transparency information types?	A) **Transparency of the outcomes.** B) Transparency of the process. C) Not clear.	
C.3	What are the transparency information pathways?	A) Demand driven. B) **Proactive.** C) Forced.	
Accountability			
D.1	What form does accountability take?	A) **Horizontal.** B) **Vertical.** C) Hybrid.	Vertical internally and horizontal externally.
D.2	What are the accountability forums?	A) **Hierarchical.** B) Administrative. C) Professional. D) **Social.**	
Institutional capabilities			

E.1	What are the institutional capabilities in place?	A) **Dissemination and access to information.** B) Institutional answerability. C) Sanctions, compensations and/or remediations.	The institution did not manage to create the institutional capability of answerability due to the lack of internal knowledge about the system.
Relationship between transparency and accountability			
F.1	What is the type of interconnectedness between transparency and accountability?	A) Direct. B) Indirect. C) **Inverse.**	The institution did not manage to develop the indirect relationship between transparency and accountability due to the lack of internal knowledge about the system.

Appendix V: Semi-structured interview guide

Introduction

My name is Anna Kubaszewska, and I am a Public Policy student at the Hertie School of Governance in Berlin. My master's thesis is on AI and ADM adoption in the public sector, supervised by Prof. Kai Wegrich and supported by Reza Mousavi, the Director General of AI Centre Sweden.

I would like to examine how the process of AI/ ADM adoption in the public sector addresses the challenge of organising transparency and, to some extent, accountability. The aim is to understand if the requirements of algorithmic transparency are met and what kind of transparency and accountability are created, as well as the possible relationships between them.

The interview will last around 30 minutes. I would like to record the interview for the purpose of the thesis. However, as per your request, your anonymity is guaranteed. If I directly quote selected sentences, I will ask for your authorisation of the version beforehand.

<u>Role:</u>

1. What was your role in adopting the AI/ ADM system in the organisation?

<u>AI adoption:</u>

1. How was the AI/ADM system adopted within the organisation?

a. In the case of external providers, what was the nature of the relationship?
b. Was there a contact provided to individuals responsible for the system?

2. How do you define the scope of the AI/ADM system adopted?
   a. What were the design decisions and intentions?

Transparency:

3. How do you define transparency? What are the organisation's guidelines on transparency?
   a. What type of transparent information was used?

(e.g., Transparency of the outcomes of the system (the what) vs. transparency of the processes (the how), Demand-driven vs. forced transparency, Proactive transparency; Upward vs Downward).

4. Is the fact that an algorithmic process is in use disclosed to the end users?
   a. What is the model used?
   b. Is the error bias of algorithms communicated when determining whether to minimise false positives or false negatives in decision making?

5. What is the level and nature of human involvement? - is this disclosed to the user?
   a. Are humans in the loop, or do they step in during an automation failure?

6. What data has been used in training and operating the system?
   a. What was the quality of the data used, including its accuracy, completeness, update frequency, and uncertainty?
   b. Was the system transparent about the extent of personal data used?

7. Were there potential problem/s concerning algorithmic transparency?

Form of the information disclosed:

8. Was the information about the AI/ ADM system disclosed differently depending on to whom it was directed?

9. Was the model self-contained enough for a person to understand and reason?

10. Were accurate details concerning the institutional performance of AI disclosed?

11. How could the end-user enquire about the system?

Accountability:

12. What is the accountability mechanism in place?

13. Are there sanctions, compensation and/or remediation in place?

<u>Data Governance:</u>

14. What are the data governance mechanisms?

15. Who was allowed to access individuals' data?

(Various policy decisions about the use of data).

   a. Is it possible for people to access raw data on request?
   b. Do the external providers have access to this data?

Appendix VI: Interview Codes

Interview Code	**Position**	**Organisation**	**AI Use Case**
Interview A	Digital Business Developer Labour Market Administration Trelleborg Municipality	Municipality of Trelleborg	Automated Social Welfare Decisions, RPA model
Interview B	Unit Manager at the Department of Welfare in Trelleborg	Municipality of Trelleborg	Automated Social Welfare Decisions, RPA model
Interview C	An academic engaged in research on automation in the public sector	Municipality of Trelleborg	Automated Social Welfare Decisions, RPA model
Interview D	An employee at Skatteverket	Swedish Tax Agency	Skatti Chatbot
Interview E	Chief Data Scientist	Swedish Tax Agency	Skatti Chatbot
Interview F	Product Owner of Chatbot Skatti	Swedish Tax Agency	Skatti Chatbot

Transcripts of the interviews can be requested.

Statement of Authorship

I hereby confirm and certify that this master thesis is my own work. All ideas and language of others are acknowledged in the text. All references and verbatim extracts are properly quoted and all other sources of information are specifically and clearly designated. I confirm that the digital copy of the master thesis that I submitted on 28.04.2024 is identical to the printed version I submitted to the Examination Office on 29.04.2024.

DATE: 28.04.2024

NAME: Anna Kubaszewska

SIGNATURE:

Anna Kubaszewska