ALGORITHMIC TRANSPARENCY IN GOVERNMENT

A multi-level perspective on transparency of and trust in algorithm use by governments

Esther Nieuwenhuizen

Algorithmic transparency in government

A multi-level perspective on transparency of and trust in algorithm use by governments

Esther Nieuwenhuizen

Algorithmic transparency in government: A multi-level perspective on transparency of and trust in algorithm use by governments.

ISBN: 978-94-6522-431-2
DOI: https://doi.org/10.33540/2985

Illustrations: Rosa Snijders | www.rosasnijders.com
Lay-out and design: Robin Weijland | www.persoonlijkproefschrift.nl
Printing: Ridderprint | www.ridderprint.nl

Algorithmic transparency in government

A multi-level perspective on transparency of and trust in algorithm use by governments

Algoritmische transparantie bij de overheid

Een multi-level perspectief op transparantie over en vertrouwen in het gebruik van algoritmen door overheden

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. ir. W. Hazeleger,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

vrijdag 19 september 2025 des middags te 12.15 uur

door

Esther Netanna Nieuwenhuizen

geboren op 5 december 1996
te Utrecht

Promotoren:

Prof. dr. A.J. Meijer
Prof. dr. F.J. Bex

Copromotor:

Dr. S.G. Grimmelikhuijsen

Beoordelingscommissie:

Prof. dr. J.G. van Erp
Prof. dr. M.L.M. Hertogh
Prof. dr. ir. M. Janssen
Dr. S. Kempeneer
Prof. dr. T. Schillemans

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) voor het onderzoeksproject 'ALGOPOL. Value-Sensitive and Transparent Algorithmization: Key to Building Citizen Trust?' [406.DI.19.011].

Contents

List of tables and figures

Figure **Page**

1

Introduction

In September 2020, the cities of Amsterdam and Helsinki became the first public organizations globally to launch an algorithm register. These registers offer an overview of the algorithms used by the municipalities (Floridi, 2020). Today, an increasing number of public organizations worldwide are adopting similar registers, aiming to provide citizens and interested stakeholders with transparency about their use of algorithms. The prevailing assumption is that this transparency could enhance trust in governmental use of algorithms, by making their functioning and utilization more understandable. The algorithm register is a concrete example of a means to provide insight into the use of algorithms by governments across multiple levels.

Through the register, an organization can provide information not only about how algorithms are embedded within an organization ("meso-level transparency") but also explain their technical functionality ("micro-level transparency") and include information about compliance with legal requirements, such as Data Protection Impact Assessments, and ongoing monitoring practices ("macro-level transparency"). This dissertation develops and applies a multi-level framework to analyze transparency and trust in the government's use of algorithms. The findings demonstrate whether and how algorithmic transparency contributes to citizen trust in the government's use of algorithms.

1.1 Algorithms, transparency and trust

Algorithms are playing an increasingly prominent role in many public service domains, ranging from healthcare to criminal justice (Meijer & Grimmelikhuijsen, 2020; Wirtz et al., 2019). For instance, a virtual health assistant can help care for patients by providing virtual consultation, health monitoring and treatment reminders (Chavali et al., 2024). Several hospitals in India make use of these virtual health assistants to reduce costs in the healthcare industry and to improve healthcare quality for individuals who cannot visit the hospital (Kaur et al., 2023). Another example is an algorithmic recommendation system that is used by the Netherlands Police to assist citizens in reporting online fraud (Borg & Bex, 2021; Odekerken et al., 2020, 2022; Soares 2024). The police developed this tool because traditional online intake forms often fall short, as citizens struggle to determine which facts are legally relevant or whether their case is civil or criminal. The recommendation system addresses this by guiding citizens to provide all relevant legal details and advising them on the best course of action based on the legal context. Moreover, this tool can help reduce costs the police face in evaluating fraud reports. These examples highlight the growing role of algorithms in public services. Before discussing the value and concerns related to the use of algorithms in the public sector, it is important to first define what algorithms are.

A widely used definition was developed by Cormen et al. (2022, p. 10): "Informally, an algorithm is any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output. An algorithm is thus a sequence of computational steps that transform the input into the output". For instance, when using a navigation app, the GPS system applies routing algorithms to determine the best route (output) based on your starting location and destination (input).

There are numerous ways to categorize algorithms in public administration (see Wirtz et al., 2019 for a comprehensive overview); for the purposes of this dissertation, I focus on two simplified relevant dimensions to describe and illustrate the algorithms I examine. The first dimension is the level of *complexity*, distinguishing, for example, simple rule-based algorithms from more advanced algorithms like deep neural networks for machine learning. Artificial intelligence (AI) can be seen as a subset of these advanced algorithms that can perform tasks typically requiring human intelligence, such as learning from data, recognizing patterns and making decisions (Wirtz et al., 2019).

The second dimension involves the degree to which an algorithm's output is *outcome determinative*, which is "the extent to which the output of an algorithm corresponds directly to governmental action eventually taken" (Coglianese & Lehr, 2019, pp. 6-7). In other words, this refers to how much the result produced by an algorithm directly influences or determines the final decision or action taken by a government entity. An algorithm can be outcome determinative when it either directly triggers an action or makes a decision, such that human intervention is largely excluded by design, particularly in automated government processes or when integrated with other systems. On the other hand, the algorithm's output might simply serve as one of several factors considered by a human decision-maker, who retains full authority over the final action (Coglianese & Lehr, 2019).

This dissertation starts from the premise that algorithms are now an integral part of our society, as Kitchin (2021, p. 221) points out: "As policy makers and law makers have discovered, trying to put the genie back in the bottle after it has been released and changed a situation dramatically is incredibly difficult to do". As algorithms become more integrated into decision-making processes, determining how we use these algorithms in a responsible and trustworthy way is becoming an increasingly urgent issue. This is illustrated by the recent adoption of the Artificial Intelligence (AI) Act by the European Union. This AI Act, among other things, prohibits certain technologies, declares different requirements for algorithmic systems based on their risk levels, such as the obligation to register high-risk algorithms in an algorithm register (Veale & Zuiderveen Borgesius, 2021). As the integration of algorithms becomes more widespread, alongside efforts such as the EU's AI Act to promote responsible use of algorithms, it is important to address the concerns they raise (see also Grimmelikhuijsen & Meijer, 2022).

These *concerns* include, but are not limited to, worries about reduced privacy from increased surveillance, job security due to automation and the risk of social inequities through profiling and control (Kitchin, 2021, p. 221). They can make the government more intrusive and powerful, potentially leading to unintended consequences such as reduced accountability and trust (Cordella & Gualdi, 2024; Katzenbach & Ulbricht, 2019). The inherent opacity of many algorithms further complicates efforts to ensure transparency and contestability in government decision-making (Aoki et al., 2024). Moreover, the misuse of algorithms can introduce or exacerbate bias and reinforce historical discrimination, raising fairness concerns in public decision-making (Wang et al., 2023). In the Netherlands, for instance, several local governments used an algorithmic system called *Systeem Risicoindicatie* (SyRI) to identify potential welfare fraud among benefit recipients. However, a Dutch court ruled the system as lacking transparency (Rechtbank Den Haag, 2020). Finally, citizens are becoming more concerned about the influence of AI on their everyday lives (Ingrams et al., 2022). These concerns must be addressed to ensure their responsible and effective implementation and use (Wirtz et al., 2019).

While the use of algorithms raises some critical questions, it is also important to acknowledge that algorithms can bring substantial *value* to the public sector, offering a range of advantages such as enhanced efficiency, productivity, safety and competitiveness (Kitchin, 2021). Furthermore, algorithms can offer faster information processing, greater accuracy and stronger predictive capabilities than human-decision makers (Wang et al., 2023). They are also seen as a way to make governance more inclusive and responsive, by reducing human subjectivity in decision-making (Giest & Grimmelikhuijsen, 2020; Katzenbach & Ulbricht, 2019).

It is important that the use of algorithms is conducted in a responsible and trustworthy manner to mitigate these concerns and make use of the value that algorithms can offer to the public sector (Meijer & Grimmelikhuijsen, 2020). One of the key aspects of the AI Act is its emphasis on *transparency* about algorithms, which is often mentioned as an important part of the solution for ensuring trustworthy and responsible algorithm use (Meijer & Grimmelikhuijsen, 2020). This resonates with the "call for transparency", where scholars and practitioners emphasize the importance of openness in algorithmic processes, recognizing that transparency about algorithms is necessary to ensure and possibly even strengthen citizen trust in algorithm use by public sector organizations (Datta et al., 2016; de Laat, 2018; Pasquale, 2015).

Transparency is believed to enhance accountability by making government actions and decision-making processes involving algorithms visible and open to public scrutiny (Cobbe et al., 2021). This visibility allows other governmental parties, such as inspectorates and data protection authorities, to detect and address potential negative impacts of algorithms, as efforts to identify and mitigate such harms become more feasible (Camilleri et al., 2023). Additionally, transparency offers "other nongovernmental organizations—the media, non-profit advocacy organizations, academic researchers, law firms and businesses—the opportunity to monitor what the government is doing" (Coglianese & Lehr, 2019, p. 18). Moreover, transparency can help build an informed citizenry, as it provides a basis for more meaningful public participation in government decision-making (Coglianese & Lehr, 2019). By granting citizens access to information about data collection, storage and use, transparency can foster *trust* as it increases their understanding of how the government operates in the digital realm (de Fine Licht & de Fine Licht, 2020; Jangoan et al., 2024). Finally, it can demonstrate the government's commitment to ethical practices and its dedication to serving the public's best interest (Floridi, 2020; Meijer & Grimmelikhuijsen, 2020).

It is important to note the nuance that other factors also play a role in citizens' acceptance of and trust in algorithms. For example, research on concepts related to trust, in particular AI acceptance, has shown that in certain situations, people's acceptance of AI is more strongly influenced by the costs and accuracy of the technology than by having a human in the loop or providing transparency about the technology (Horvath et al., 2023). While transparency is not the only relevant factor in ensuring and enhancing citizens' trust in algorithms, various studies indicate that it is indeed an important factor (e.g., Coglianese & Lehr, 2019; Grimmelikhuijsen, 2023; Grimmelikhuijsen & Meijer, 2022; Meijer & Grimmelikhuijsen, 2020). However, there remains considerable debate about what exactly this transparency regarding algorithms entails and what its effects are on citizen trust. This dissertation therefore seeks to clarify how transparency about algorithm use by public sector organizations influences citizen trust. The following research question is central to this dissertation: *"How does transparency affect citizen trust in algorithm use by government organizations?"*. More specifically, I aim to tackle four different *gaps* in the literature with this research question, leading to four sub-questions that will be described below. These research gaps were identified by analyzing the existing academic literature at the intersection of transparency, trust and algorithmization. This process involved reviewing both theoretical and empirical contributions within public administration and relevant other disciplines.

1.2 Addressing four gaps in the academic literature

A disconnect between micro, meso and macro level research on algorithmic transparency

The first gap relates to the analytical lens: how to study the relationship between transparency and citizen trust in the use of algorithms by governments? A widely recognized method for categorizing research in public administration and management is the micro, meso, macro distinction, which reflects the specific level of analysis employed in a study (Roberts, 2020). The micro level usually includes a focus on individuals, such as civil servants, citizens, etc. The meso level focuses on the organizational context, such as organizational policies and practices. The macro level, finally, zooms in on the institutional context, covering, for instance, legislations, regulations and national strategies. These three levels thus provide insights regarding a topic at different levels. Research about the relationship between transparency and trust suggests that the three levels are very relevant for studying this relationship (Porumbescu et al., 2022). Similarly, research in the context of algorithmic transparency argues that these three analytical levels are relevant when examining the concept of transparency in the context of algorithm use by governments (Giest & Grimmelikhuijsen, 2020). However, most of the (scarce) research on algorithmic transparency focuses on the micro level (e.g., Aoki, 2020; Grimmelikhuijsen, 2023; Schiff et al., 2022). This broader, multi-level perspective on transparency in the context of trust in the use of algorithms by governments has not been explored so far.

Chapter 2 addresses this gap by conceptualizing how the relationship between transparency and citizen trust in the use of algorithms by governments is shaped across different levels. In line with previous research advocating for a multi-level perspective (e.g., Giest & Grimmelikhuijsen, 2020; Porumbescu et al., 2022), the following sub question (RQ1) is central to this chapter: *"How can we understand the relationship between trust and algorithmic transparency at the micro, meso and macro levels?"*. I use a literature review to address this first sub question. This results in a multi-level framework which focuses on **micro** (specific algorithms), **meso** (organizational) and **macro** (institutional) level transparency challenges and on how these are linked to citizen trust, which I further elaborate on in Chapter 2. The answer to the first sub question provides a relevant theoretical and conceptual basis for addressing the next three sub questions. To refine the framework, I use insights from the empirical studies (Chapters 3-5) regarding the relationship between trust and algorithmic transparency across different levels.

Mixed evidence on the impact of explaining algorithmic outputs on citizen trust

The second gap relates to providing micro-level transparency about the functioning of an algorithm. This concerns algorithmic transparency, which involves the act of making a system knowable or visible (Ananny & Crawford, 2018). One way to provide algorithmic transparency is by giving explanations about the output of an algorithm (Grimmelikhuijsen, 2023). Research into trust in algorithmic outputs is scarce, non-conclusive and mostly focused on the private sector. Some studies show an increase in trust by providing explanations (e.g., Kizilcec, 2016; Nothdurft et al., 2014; Wang & Benbasat, 2007) while others show a negative or non-existent effect of explanations on trust (e.g., Cramer et al., 2008; Tintarev & Masthoff, 2012). Only recently has some empirical evidence on trust in algorithmic outputs in the public sector been provided (e.g., Aoki, 2020; Grimmelikhuijsen, 2023; Schiff et al., 2022). These studies, however, do not specify what kinds of explanations are needed for citizens to trust these algorithmic outputs.

Chapter 3 addresses this gap by examining the relationship between providing explanations and citizen trust in a specific algorithm. In particular, this chapter focusses on a rapidly emerging trend in digital public service delivery: algorithmic recommender systems, such as chatbots and decision support systems that provide suggestions to citizens in various service domains. The second sub question (RQ2) addressed in this chapter is: *"What are the effects of explanations on citizen trust in algorithmic recommendations?"*. This question is answered based on a quantitative design including two representative survey experiments with N=717 and N=1005 Dutch citizens, respectively.

Unresolved questions on the organizational embedding of algorithms in government

The third research gap relates to providing meso-level transparency about the organizational embedding of algorithms. As explained before, the process of algorithmization causes organizations to adapt their work processes and routines around algorithms (Meijer & Grimmelikhuijsen, 2020). Therefore, a growing number of researchers is emphasizing that information about algorithms should not only focus on their technical functioning but also pay attention to the organizational context in which these algorithms operate (De Bruijn et al., 2022; Grimmelikhuijsen & Meijer, 2022). In such an approach, not only information about specific algorithms should be presented, but also about their usage, governance and evaluation, and not just to citizens but also to intermediaries, such as NGOs, journalists, independent experts and regulators (Grimmelikhuijsen & Meijer, 2022).

In this vein, the Dutch government initiated the implementation of algorithm registers to publicly disclose information about algorithms and their integration within public organizations. The early scholarly debate about this new phenomenon shows that the value of these registers is questioned: some claim that governments only disclose politically non-sensitive algorithms and provide information with limited usefulness for accountability purposes (Cath & Jansen, 2022). In contrast, a more optimistic view is also put forward: registers could help organizations to be transparent to the public, which could increase public trust in governmental algorithm use (Floridi, 2020). However, the academic debate on this topic is still in its early stages, making it difficult to understand what algorithm registers are and what impact they might have.

Chapter 4 addresses this gap by investigating the nature and implications of algorithm registers. Given that algorithm registers represent a novel development in both societal and academic debates, a qualitative approach was employed to explore what algorithm registers are, the objectives they aim to achieve, how they are designed, and their implications. This exploration is guided by the sub question (RQ3): *"What is the nature of algorithm registers, and what are their implications?"*. Note that the question does not directly address trust, as algorithm registers are still a relatively new phenomenon. Therefore, I formulated an exploratory question to openly investigate what these algorithm registers are and the impact they may have. The answer to the sub question is based on interviews with N=27 developers (from public organizations) and key users (oversight authorities and societal watchdogs) of algorithm registers. This is complemented by a document analysis of N=33 policy documents.

Limited exploration of how institutional transparency impacts citizen trust

The fourth gap relates to providing macro-level transparency about the institutional embedding of algorithms. Several scholars argue and show that transparency about the institutional context in which algorithms operate is very important (Grimmelikhuijsen & Meijer, 2022; New & Castro, 2018; Wenzelburger et al., 2024). They highlight the importance of institutional guardrails, such as laws and oversight bodies, in promoting trust in an era of algorithmic governance (Grimmelikhuijsen & Meijer, 2022). In legal studies, research has been conducted on the institutional guardrails for algorithm use by governments, such as the GDPR and the AI Act. These studies in particular highlight the importance of improved legal requirements as the current ones lack effectiveness and oversight (Mantelero, 2024; Temme, 2017; Wachter, 2024).

To date, the legal literature has largely overlooked the citizen perspective, with limited attention given to ensuring transparency about the institutional context in which algorithms operate. One notable exception is the study by Grimmelikhuijsen and Meijer (2022), which offers a public administration perspective on these institutional guardrails for algorithms. They argue that revising institutional arrangements is essential to enhancing the (perceived) legitimacy of algorithmic decision-making. This work emphasizes the importance of institutional guardrails for legitimacy, which is linked to trust, but the research is conceptual and needs further empirical validation. This highlights, among others, the need for empirical research on information about institutional guardrails, such as regulations and monitoring, for citizen trust in algorithm use by governments.

Chapter 5 addresses this gap by examining how transparency regarding the institutional embedding of algorithms impacts citizen trust in the use of algorithms by public organizations. In this chapter, I zoom in on predictive policing, a widely used algorithmic system by police organizations, to examine how transparency regarding legislation and compliance monitoring impacts citizen trust. The fourth sub question (RQ4) is formulated as: *"To what extent does institutional transparency affect citizen trust in predictive policing?"*. This question is answered by the means of a representative survey experiment with N=877 citizens from the Netherlands.

1.3 Academic relevance of this dissertation

Building on the identified research gaps, I make both theoretical and empirical contributions to advancing academic discussions on government transparency and algorithmization literature. Providing government transparency is becoming increasingly complex in the algorithmic age due to the complexity and opacity of algorithmic decision-making processes (Aoki et al., 2024; Burrell, 2016; Schuilenburg & Peeters, 2024). The general government transparency literature primarily focuses on providing citizens with information about government decision-making, policies and policy outcomes (Grimmelikhuijsen, 2012b). However, there is a growing need for information about how algorithms are used by governments, as they can influence their decision-making procedures, policies and policy outcomes (Meijer & Grimmelikhuijsen, 2020). With this dissertation, I aim to make important theoretical and empirical contributions to the understanding of how transparency can be achieved in the context of algorithm use by public sector organizations, and whether such transparency might impact citizen trust.

On the *theoretical* side, this dissertation conceptualizes a multi-level framework on algorithmic transparency and trust, an approach that has been introduced but not applied to the transparency and trust relationship in the context of government use of algorithms (Giest & Grimmelikhuijsen, 2020; Porumbescu et al., 2022). The framework integrates micro, meso and macro perspectives on transparency and trust in the use of government algorithms (Chapter 2). This framework moves beyond the traditional focus on providing transparency about individual algorithms, incorporating the organizational and institutional contexts in which these algorithms are embedded. As a result, this dissertation contributes to a better theoretical understanding of how transparency at various levels relates to public trust in algorithmic systems.

On the *empirical side*, the dissertation provides *innovation*, *refinement* and *consolidation*. Currently, there are conceptual debates about the nature and implications of algorithm registers as a tool for providing transparency about government use of algorithms (Cath & Jansen, 2022; Floridi, 2020). This dissertation innovates by conducting the first empirical investigation into algorithm registers (Chapter 4). Furthermore, while some studies focus on explainability, they often lack nuance or refinement in exploring different approaches to explaining algorithmic outputs (e.g., Grimmelikhuijsen, 2023; Schiff et al., 2022) or do not address algorithmic decision-making processes in the public sector (e.g., Kizilcec, 2016; Rader et al., 2018). The findings in this dissertation refine these existing insights into the relationship between algorithmic transparency and trust by testing different types of explanations for algorithmic outputs in the public sector context (Chapter 3). Finally, several authors discuss the importance of institutional guardrails, such as legislation and monitoring related to the use of algorithms by public organizations, for government legitimacy (Grimmelikhuijsen & Meijer, 2022) and argue that transparency about the institutional context in which algorithms operate is very important (New & Castro, 2018). The insights from this dissertation consolidate theoretical assumptions about the importance of institutional guardrails in the context of algorithm use by governments by providing an empirical foundation (Chapter 5). Altogether, these empirical insights enrich the growing body of literature on government transparency and algorithmization. This approach is consistent with one of the recommendations made by Zuiderwijk et al. (2021, p. 15) to prioritize empirical research over speculation about the societal impact of AI on public governance.

1.4 Research strategy and context

This paragraph discusses the mixed-methods approach used in this dissertation, combining both quantitative and qualitative research to examine the relationship between transparency and trust in algorithm use by governments. It also addresses how the research is firmly embedded in practice, focusing on real-world algorithmic tools within public sector organizations.

Mixed-methods approach

This dissertation consists of positivist research, which assumes the existence of an objective reality independent of the researcher, allowing it to be known objectively, from an external perspective (Haverland & Yanow, 2012). This is evident in both the empirical quantitative and qualitative studies presented in this dissertation. The first empirical study in Chapter 3 develops hypotheses that are rigorously tested through two survey experiments. A distinctive aspect of this study is its focus on measuring actual human behavior (Grimmelikhuijsen et al., 2017), rather than solely capturing people's trusting attitudes. By having individuals interact with a mock version of a real algorithmic recommender system, we measure their trust-related behaviors. The second empirical study in Chapter 4, although exploratory, also uses theory as a foundation for qualitative data collection. This theoretical framework is further enriched through abductive coding, incorporating insights from practical experiences. Finally, the third empirical study in Chapter 5 places significant emphasis on the theoretical and methodological development of a new concept in the context of algorithmic use by governments: institutional transparency. Each empirical chapter includes a detailed discussion of the research strategy employed for that specific study.

The police serve as a recurring context to explore broader phenomena related to algorithm use within the public sector. Chapter 2 examines the relationship between transparency and trust in algorithmic governance, using a case study involving an algorithmic recommender system employed by the police. This *Intelligent Crime Reporting Tool* assesses citizen complaints of online fraud related to, for example, fake web shops and malicious second-hand traders on platforms like eBay.[1] On the basis of citizens' stories and several follow-up questions, the algorithm recommends citizens whether it is feasible to file a report and whether they should officially report their case or not (Odekerken et al., 2020). Chapter 3 investigates how different types of explanations influence the level of trust individuals place in the recommendations generated by the same algorithmic tool. Chapter 5 explores how institutional transparency can enhance public trust in predictive algorithms, with a focus on a separate case study involving the police, specifically in the context of predictive policing. This *Crime Anticipation System* employs algorithms and data analysis to predict criminal activities and anticipate

[1] Politie (n.d.) *Keuzehulp internetoplichting.* Retrieved January 21, 2025, from https://aangifte.politie.nl/iaai-preintake/#/.

potential offenses. By processing historical data alongside real-time information, it provides insights into potential high-risk areas and timeframes, enabling the police to take proactive measures to prevent crime (Meijer et al., 2021).

Firmly embedded in practice

This research was part of a larger interdisciplinary research project called Algorithmic Policing (ALGOPOL), which explores how the Netherlands Police can use algorithms in a responsible and trustworthy manner. This project, funded for four years by the Dutch Research Council (NWO), included several subprojects focused on the responsible and trustworthy use of algorithms by the Netherlands Police, with mine being one of them. The Netherlands Police are increasingly adopting data-driven approaches to enhance public safety and combat crime more effectively. To achieve this, the police are making greater use of big data and algorithms, particularly for risk assessments and predictive analytics (den Hengst & Wijsman, 2023; Meijer et al., 2021; Schuilenburg & Soudijn, 2023). A key priority for the Netherlands Police in the use of big data in algorithms is "trust as the goal, transparency as the means" (Politie, 2022, p. 9). As the technologies used by the police become increasingly difficult for citizens to understand, while citizens themselves become more transparent through their data production, an imbalance in transparency arises, which can undermine trust in the police. Therefore, the police strive to be, according to their strategy, "as transparent as possible" (Politie, 2022, p. 9), while simultaneously seeking a balance between confidentiality and openness. This key priority demonstrates that the police are actively engaged with issues of transparency and trust in the context of algorithms, highlighting the practical relevance of focusing on this context in this dissertation.

The independent academic research conducted within the ALGOPOL project is carried out in collaboration with the Netherlands Police, which greatly facilitated research access to the police. Following a comprehensive security screening, I was granted access to police facilities through an access pass, along with entry to various systems and an official email address. This arrangement facilitated my interactions with police personnel, enabling me to conduct on-site conversations related to my research. Furthermore, I gained access to employees working on and with AI within the police through my membership in the Netherlands Police Lab AI. This is a collaboration between several Dutch universities and the police. This lab develops state-of-the-art AI techniques aimed at improving public safety in a socially, legally and ethically responsible manner (National Police Lab AI, n.d.). Many members of the Netherlands Police Lab AI are employed by the police while simultaneously pursuing a PhD through a university. During monthly meetings, lab members present their research to one another and engage in discussions on the design and application of AI for the Netherlands Police.

Practically, this access provided me with insights into the algorithms being developed and utilized. By engaging with the developers and product owners of various algorithmic tools, I could comprehend their functionalities.[2] I applied this knowledge in my investigations of two existing police algorithms: the Intelligent Crime Reporting Tool (Chapters 2 and 3) and the Crime Anticipation System (Chapter 5). With my knowledge of how these algorithmic tools work, I was able to create a mock version and a hypothetical scenario, respectively, of them for my research. This enabled me to thoroughly examine the existing algorithmic tools used by the police (see also Camilleri et al., 2023). This approach aligns with the recommendation made by Zuiderwijk et al. (2021, p. 15) in their research agenda on the implications of AI use for public governance. They highlight the necessity for more domain-specific studies (such as safety), particularly in relation to specific countries (such as the Netherlands) and at particular levels of government concerning AI (such as executive agencies like the police).

The research on algorithm registers (Chapter 4) is an exception to this focus on the police context. This chapter centers on transparency regarding the organizational embedding of algorithms. An algorithm register, which provides insight into the algorithms utilized by public organizations (Haataja et al., 2020), is an ideal means to investigate this meso level transparency. I chose to adopt a broader focus in this study, examining various public organizations across the Netherlands, for two reasons. First, this exploratory study aimed to address the phenomenon of algorithmic registers. Aside from two conceptual studies (Cath & Jansen, 2022; Floridi, 2020), there had been no empirical research available on this topic. Consequently, I could not build upon previous work on algorithmic registers that could be used to investigate the context of the police. Second, and building on the first point, the police did not have its own algorithm register or published algorithms in the national algorithm register at the time of my research. To avoid conducting a study based on a hypothetical or speculative scenario, I opted to focus on organizations that already had established algorithm registers. This approach aligns with one of the recommendations made by Zuiderwijk et al. (2021, p. 15) to conduct more empirical research rather than relying on speculations about the societal impact of AI on public governance.

I discuss the broader implications of my findings for organizations in the public sector throughout all chapters. By drawing on public administration literature and applying it to the context of policing, the theoretical mechanisms identified are also

[2] I am aware of the potential risks, such as conflicts of interest or comprised objectivity, associated with collaborating with the police. In Chapter 7, I reflect on my positioning as a researcher both within and outside the police organization, and how I have maintained the independence and integrity of my research.

relevant across other domains. An example of this is presented in Chapter 2, where I use public administration literature on the relationship between transparency and trust to develop a multi-level framework regarding this relationship in algorithmic governance. The various levels are illustrated using a case study of an existing algorithmic recommender system employed by the police. This illustration provides a concrete context for understanding how these theoretical mechanisms operate in practice. As a result, this dissertation provides valuable implications not only for the police but also for other public organizations more broadly.

It is important to note that a fundamental difference between police organizations and other public organizations lies in, most notably, the police's monopoly on violence, which means that the consequences of algorithmic decisions can have far-reaching implications beyond those of other public organizations (Wessels, 2024). Decisions made by algorithms within the police context can directly impact the freedom, safety and rights of citizens (e.g., Brayne, 2020; Jansen, 2022), thereby making the necessity for transparency and trust even more urgent compared to other government agencies (Wessels, 2024). That being said, there are also decisions made by other government organizations, which do not have a monopoly on violence, that can have significant consequences for citizens. One example of this is the childcare benefits scandal, where the Dutch Tax Authority used biased algorithms to make decisions about citizens, resulting in many individuals being wrongly accused of fraud, forced to repay money and facing financial problems (Bouwmeester, 2023).[3] Thus, while there are differences between the police context and other government organizations, they share the commonality that decisions made within these institutions can have far-reaching impacts on citizens. This highlights the need for transparent and trustworthy algorithm use by these institutions.

1.5 Societal relevance of this dissertation

This dissertation makes three important societal contributions. First, the insights from this dissertation can contribute to the responsible use of algorithms by the *police*. As mentioned earlier, several studies were conducted within the context of policing. Given that the police face challenges in offering transparency about their algorithmic practices (Wessels, 2024), my efforts to provide insights into what kind of transparency about their algorithm use can enhance trust, can help them tackle this challenge (Camilleri et al., 2023). The insights from this dissertation have both direct and indirect implications for the police. The specific lessons derived from the police's algorithmic systems examined in this research can be directly applied within the organization. Additionally, the broader

[3] For a detailed discussion of the childcare benefits scandal, see Section 4.2.

findings on algorithmic transparency can inform and enrich wider discussions within the police about the responsible and ethical use of their algorithms.

Other *public organizations* can also benefit from the findings in this dissertation. Many public organizations struggle with how best to implement transparency regarding their algorithms (van Vliet et al., 2024), and this dissertation offers valuable guidance on addressing this challenge. Public organizations can use these insights about algorithmic transparency for their design and development of algorithmic processes within their organization (Felzmann et al., 2020) and for the organizational and institutional embedding of their algorithms. For example, they can use the findings on algorithm registers to inform and guide their approach to making transparent to the public how they integrated algorithms in their organizational work processes and routines.

Finally, the knowledge gained from this dissertation contributes to broader *societal debates* surrounding algorithmic transparency. For instance, while transparency has gained prominence within the AI Act in the European Union (Panigutti et al., 2023), it remains largely unclear what steps countries will take to meet the various transparency obligations. This research offers practical insights that can inform these discussions, shedding light on which forms of transparency are more or less valuable. Furthermore, this connects directly to concrete policy issues concerning the increasing influence of algorithms on society. Several reports by national and international policy advisors and oversight bodies, such as those by the Netherlands Scientific Council for Government Policy (WRR, 2021) and the European Court of Auditors (2024) emphasize the need to critically evaluate the societal impact of algorithms, underscoring a broader demand for evidence-based insights. In this light, the findings from this research can support politicians in holding the government accountable, because they can help to evaluate whether algorithmic practices are used in a just and responsible manner.

1.6 Dissertation overview

This dissertation consists of seven chapters, including this introduction. The next chapter, **Chapter 2**, develops the theoretical framework for the dissertation, proposing a multi-level approach to analyzing transparency and trust. It argues for the need to study transparency and trust not only at the micro level (focused on individual algorithms) but also at the meso level (within the organizational context) and the macro level (within the broader institutional framework). This chapter synthesizes existing literature on transparency and trust, identifying gaps that the dissertation seeks to address through conceptualizing a multi-level perspective on the relationship between trust and transparency in algorithmic governance.

Chapter 3 focuses on the *micro level* analysis, exploring how transparency related to specific algorithms influences citizen trust. This chapter presents the findings from two experimental studies that test the effects of different types of algorithmic explanations (procedural, rationale, combined and directive) on trust. The experiments reveal that providing explanations significantly enhances trust, and the chapter discusses the implications of these findings for designing transparent algorithmic systems in government.

Chapter 4 shifts the focus to the *meso level*, examining the nature of algorithm registers, an example of an organizational transparency practice and their implications. This chapter is based on qualitative research, including in-depth interviews with public organizations, oversight authorities and societal watchdogs, as well as a detailed analysis of policy documents. The findings provide an overview and understanding of the design (process) of algorithm registers, their benefits and limitations, and the chapter suggests ways to enhance their effectiveness (in building trust).

Chapter 5 explores transparency and trust at the *macro level*, focusing on the broader institutional context, including legal frameworks and external oversight mechanisms. This chapter presents the results of a survey experiment that investigates how transparency about legislation and external monitoring influences public trust in predictive policing algorithms. The findings emphasize both the importance and limits of institutional transparency in fostering trust and offer insights for policymakers and regulators seeking to enhance the legitimacy of their algorithmic governance.

The dissertation continues with a conclusion and discussion in **Chapter 6**, in which the central research questions will be answered and several academic and societal implications of the dissertation will be presented. Finally, the dissertation ends with an epilogue in **Chapter 7**, where I reflect on my role and position as a researcher, discuss the various impact activities I engaged in throughout the course of this research and share three lessons learned about "making impact" in a PhD trajectory.

Table 1.1 provides an overview of the chapters of the dissertation. It includes the questions that are central in each chapter, the research approach that has been used to investigate these questions, the publication status and a main takeaway message.

Please note that chapters 2-5 were written as independent manuscripts. Because of this, the introductions of these chapters may overlap. In addition, the numbering and formatting of the article manuscripts have been standardized to ensure consistency across the chapters in this dissertation.

Table 1.1 Overview of the dissertation

Chapter	Research question	Research approach	Publication status	Main takeaway message
2 *Multi-level framework*	How can we understand the relationship between trust and algorithmic transparency at the micro, meso and macro levels?	Literature review	Published chapter in *Handbook on Trust in Public Governance* (Nieuwenhuizen, 2025b)	This chapter proposes a framework to study the relationship between transparency and trust in algorithmic governance at micro (individual algorithms), meso (organizational embedding of algorithms) and macro (institutional embedding of algorithms) levels.
3 *Micro-level*	What are the effects of explanations on citizen trust in algorithmic recommendations?	Experimental sequential factorial design with two survey-experiments. Study 1 with N=717 citizens from the Netherlands, Study 2 with N=1005 citizens from the Netherlands.	Published paper in *Public Performance & Management Review* (Nieuwenhuizen et al., 2024)	Two survey-experiments in this chapter reveal that various types of explanations increase citizen trust in algorithmic recommendations.
4 *Meso-level*	What is the nature of algorithm registers, and what are their implications?	Qualitative design with interviews with N=27 developers and users of algorithm registers and a document analysis of N=33 policy documents.	Published paper in *Information Polity* (Nieuwenhuizen, 2025a)	Interviews and a document analysis in this chapter reveal that algorithm registers, intended for transparency, often fall short in providing meaningful information for outsiders. At the same time, they create organizational awareness and initiate critical debates about the use of algorithms within government.
5 *Macro-level*	To what extent does institutional transparency affect citizen trust in predictive policing?	Experimental design with N=877 citizens from the Netherlands.	Paper under review (co-authored with Vishal Trehan and Gregory Porumbescu)	A survey-experiment in this chapter demonstrates that transparency about external monitoring, rather than legislation, significantly boosts citizen trust in predictive policing algorithms.

2

Conceptualizing a multi-level framework of algorithmic transparency and trust

This chapter is based on the following published book chapter: Nieuwenhuizen, E. (2025b). Trust and transparency in algorithmic governance. In F. Six, J. Hamm, D. Latusek, E. Zimmeren, van, & K. Verhoest (Eds.), *Handbook on Trust in Public Governance* (pp. 116 - 135). Edward Elgar Publishing.

Abstract

Transparency is presented as a key element for maintaining trust in the use of algorithms by governments. The current literature focuses mainly on the need for explainable algorithms and accessible source code. A narrow focus on specific algorithmic applications fails to acknowledge the organizational and institutional context of algorithmic governance. This chapter argues that we need to broaden our understanding of the relationship between trust and algorithmic transparency by not only studying this at the *micro level* (trust in and transparency of specific algorithms), but also at the *meso level* (trust in and transparency about the organizational embedding of algorithms) and the *macro level* (trust in and transparency about the institutional embedding of algorithms). This comprehensive three-level framework creates a deeper understanding of whether and how citizens trust a government that uses algorithms.

2.1 Introduction

The way governments present themselves at the point of contact with citizens has radically changed over the past years due to the introduction and use of algorithms by governments, and this has important consequences for citizens' trust. Let me illustrate this transformation in the case of reporting fraud to the police in the Netherlands. Before algorithms emerged as a way to automate government decision-making and services, Dutch citizens would have regular physical encounters with government officials. For instance, citizens would go to a police station in cases where they believed they had been scammed. They would discuss the case with a police officer behind the counter and, if necessary, file a report of fraud. Nowadays, citizens can report online fraud cases online. On the website of the police, they click on "file a report" and use an interface of an algorithmically aided system: the Intelligent Crime Reporting Tool of the Netherlands Police. This system assesses citizen complaints of online fraud related to, for instance, fake web shops and malicious second-hand traders on platforms like eBay. On the basis of their story and several follow-up questions, the algorithm then recommends to the citizens whether it is feasible to file a report and whether they should officially report their case or not (Odekerken et al., 2020). Yet, how do citizens know whether this recommendation can be trusted?

In the past, at the police station, citizens could see the police official in person and could assess whether they found their judgements and actions trustworthy. With the emergence of algorithmically aided systems, citizens engage with an impersonal interface that utilizes algorithms to evaluate their reports. This transition raises questions about the trustworthiness of the output of the algorithm. Citizens may be uncertain as to how the algorithm makes its decisions and whether it can be truly relied upon. The absence of a visible human being may lead to feelings of uncertainty and detachment. This example highlights how the dynamic between governments and citizens has significantly changed with the rise of algorithm use by governments. Do citizens still trust government organizations based on digital encounters? In view of the shift from physical to digital encounters, it is important to study how trust in algorithmic governance can be realized and safeguarded. This chapter will explore how different forms of transparency can play a key role in generating trust in the use of algorithmic governance.

Governments are slowly transforming by introducing algorithms to automate decision-making and service delivery (Meijer & Grimmelikhuijsen, 2020). Algorithms are increasingly used to facilitate digital encounters. For instance, chatbots are initiated to improve communication between government and citizens. Virtual assistants respond to questions in human language to provide real-time access to information and support, such as in the delivery of public safety or citizens' security and immigration services (Androutsopoulou et al., 2019, p. 360). Furthermore, algorithmic recommender systems are used by public organizations to support decision-making by citizens (Smets et al., 2020). They are used, for example, to support citizens filling out their online tax forms or, as illustrated above, to support citizens in filing a report of online fraud to the police (Odekerken et al., 2020). Algorithmic governance can be considered a new method of organization around the use of an algorithm and not just the use of a new tool within the existing organization (Grimmelikhuijsen & Meijer, 2022).

The interaction between citizens and governments and the object of trust—going from physical to digital encounters—changes considerably in algorithmic governance (Gritsenko & Wood, 2022). This transformation raises questions about citizens' trust in government. One significant concern is the digital divide, where certain segments of the population, particularly those with limited access to technology or digital literacy skills, may be left behind, exacerbating existing inequalities and eroding trust (Myeong et al., 2014). With algorithms taking a prominent place in government to inform policy or support decision-making, researchers argue that it is necessary to examine how citizens' trust in algorithmic governance can be realized in the first place, safeguarded, and, if possible, even strengthened (Grimmelikhuijsen & Meijer, 2022; Meijer & Grimmelikhuijsen, 2020). As algorithmic governance is a relatively new phenomenon, the focus extends beyond maintaining and strengthening trust; it also involves the challenge of building trust in the first place.

It is important that citizens trust digital encounters in the context of algorithmic governance. Trust is regarded as an essential element in democracies. If government organizations are not trusted by the citizens they serve, these organizations are unable to function properly (Warren, 1999). For instance, interactions become slower and less efficient, cooperation becomes scarce, and overall service delivery suffers. This not only impacts the dependent service users, who experience delays and frustration, but also undermines the government's ability to fulfil its duty of providing essential services to its citizens. Without trust, the foundation for a harmonious relationship between the government and its citizens weakens, impeding progress and hindering societal development. Moreover, since citizens mostly have no realistic 'exit' option for public services (Hirschman, 1970), they are dependent on a single service provider. It should be noted that there are cases in which alternative service providers may be available, particularly in private provider-based systems (e.g., Diphoorn, 2013; Warner & Hefetz, 2008), or cases in which citizens decide not to engage or access (non-essential) public services (e.g., Dowding & John, 2008; Hirschman, 1993). Nevertheless, in most cases citizens cannot choose between different service providers in contexts of digital public service delivery; they are 'locked into' the system selected by the government. Therefore, this chapter starts from the assumption that no realistic exit exists.

Besides, trust plays a pivotal role as an instrumental value (Hoff & Bashir, 2015). In algorithmic decision-making processes, the reliance on automated systems necessitates the establishment of trust between the governing entities and the individuals affected by algorithmic outcomes. When citizens trust the algorithms employed in public decision-making, they are more likely to accept and comply with the decisions and recommendations generated by these systems (Nieuwenhuizen et al., 2021). Consequently, trust in algorithmic governance acts as a driving force behind the successful implementation and acceptance of algorithmic systems in public administration.

It is argued that algorithmic transparency is crucial for building, maintaining and strengthening trust in government in an information age. However, algorithms use various datasets and can be so complex that the logic of decision-making and possible biases are hard to detect and understand (Janssen & van den Hoven, 2015). A lack of transparency can negatively affect citizens' trust (Giest & Grimmelikhuijsen, 2020), especially when these algorithms are "black boxes" that do not provide understandable explanations as to why or how an algorithmic recommendation is given (Lepri et al., 2018; Rader et al., 2018) or when the source codes are not accessible to external actors for monitoring (Diakopoulos, 2016). Algorithmic transparency thus requires accessible and explainable algorithms so that their users can monitor and understand the outcomes (Kroll et al., 2017).

So far, most studies have investigated algorithmic transparency and trust in a rather narrow manner. For instance, the studies by Grimmelikhuijsen (2023) and Nieuwenhuizen et al. (2021) focus on the transparency of the algorithmic application itself, whereas general transparency literature suggests that transparency also plays a role at higher-order levels, such as at the organizational and institutional levels (Porumbescu et al., 2022). This includes transparency about organizational policies and institutional rules on the appropriate implementation and use of algorithms in government (Grimmelikhuijsen & Meijer, 2022). How we can use this broader understanding of transparency in the context of trust in algorithmic governance has not been examined so far. This chapter shows that in order to build, maintain and strengthen trust in algorithmic governance, it is not only necessary for algorithms themselves to be transparent, but the way in which they are used by government organizations and the rules and regulations regarding government organizations' algorithm use should also be transparent.

In this chapter, I first conceptualize trust in algorithmic governance. Then, I discuss how transparency can be used to build, maintain and strengthen citizens' trust in algorithmic governance, leading to a conceptual framework that can be used to empirically examine the relationship between transparency and trust in algorithmic governance. This is a multi-level framework which focusses on micro (specific algorithms), meso (organizational) and macro (institutional) level transparency challenges (Giest & Grimmelikhuijsen, 2020, p. 411) and on how these are linked to citizen trust.

I will use the case of the Intelligent Crime Reporting Tool (ICRT) of the Netherlands Police, briefly described above, to illustrate the three different levels of transparency and to show how the framework can be used. Illustrated by this concrete example, this chapter shows which actions can be used to build, maintain and strengthen citizens' trust in governments that increasingly rely on algorithmic governance.

2.2 Trust in algorithmic governance

Trust has been analyzed on several levels: trust between people, trust in teams, trust in organizations and trust in systems and institutions (Grimmelikhuijsen, 2012b, p. 29). Rousseau et al. (1998, p. 395) tried to capture "trust" in a multidisciplinary definition: "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another". This definition underscores trust as the perceived trustworthiness of another (Grimmelikhuijsen, 2012b). Many scholars, recognizing the complexity of trust, have identified its various dimensions of perceived trustworthiness (e.g., Grimmelikhuijsen & Knies, 2017; McKnight et al., 2002; Rousseau et al., 1998). Importantly, this multidimensional understanding of trust is not restricted to traditional government-citizen interactions but extends to novel contexts such as algorithmic governance (Ingrams et al., 2022).

While recent studies have primarily zoomed in on examining trust in specific algorithms at the micro-level (Grimmelikhuijsen, 2023; Hobson et al., 2023; Lee, 2018; Nieuwenhuizen et al., 2021), a holistic understanding of trust in algorithmic governance and its implications for public organizations requires a broader perspective. In addition to analyzing trust at the micro-level, it is important to explore the establishment of organizational and institutional trust in the utilization of algorithms. Exploring these broader dimensions of trust provides deeper insights into the dynamics of algorithmic governance and offers valuable insights into how public organizations can effectively address trust-related challenges.

In this chapter, I discuss the three levels of trust in algorithmic governance separately. However, it is important to note that the literature acknowledges the interdependence between different levels of trust (Kramer & Tyler, 1995). In other words, trust is shaped by a combination of factors at the institutional, organizational and individual levels (Misztal, 2013; Porumbescu et al., 2022). This recognition of the interplay between different levels of trust adds nuance to our understanding of the relation between trust in the algorithm itself and trust in institutions using the algorithm, and highlights the complex nature of trust dynamics in algorithmic governance.

Micro-level trust

At the micro level, algorithmic trust refers to trust in a specific algorithm by its user. This can refer to various types of systems: an algorithmic decision-making system, an algorithmic recommender system, a chatbot, etc. To understand this level of trust, I draw upon the framework of human-computer trust (HCT) theory. HCT is "the extent to which a user is confident in, and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent decision aid" (Madsen & Gregor, 2000, p. 1). The object of trust at this level is thus the algorithmic outcome of an intelligent decision-aid system. By using insights from the literature on e-commerce and information systems, and its emphasis on the central role of trust, I further define algorithmic trust at the micro-level.

The literature on e-commerce extensively elaborates on the central role trust plays in making consumers comfortable acting on algorithmic outcomes and making purchases. For example, research conducted by Wang and Emurian (2005) examines the impact of trust in e-commerce transactions, revealing that trust significantly influences consumers' willingness to engage in transactions and make purchases based on algorithmic outcomes. McKnight and Chervany (2001) explore the role of trust in fostering customer confidence in algorithmic recommendations, emphasizing its positive influence on consumers' acceptance and utilization of algorithmic outcomes. Additionally, Lee et al. (2015) investigate the impact of trust on consumers' acceptance of personalized recommendations generated by algorithms, finding that trust in the algorithm's competence, transparency and privacy safeguards greatly shapes consumers' trust perceptions and their acceptance of algorithmic recommendations. Collectively, these studies underscore the central importance of trust in e-commerce settings, influencing consumers' comfort levels and their propensity to act on algorithmic outcomes when making purchasing decisions.

Shifting from the literature on e-commerce, the influential model of trust in information systems, formulated by McKnight et al. (2002), provides valuable insights into the conceptualization of algorithmic trust as trusting beliefs, trusting intentions and trust-related behaviors. This multidimensional understanding of trust has been further developed by Dietz (2011). He distinguishes between the assessment of trustworthiness, the actual decision to trust and trust-informed actions (Dietz, 2011; Dietz & Den Hartog, 2006). The assessment of trustworthiness, which aligns with evaluating trust beliefs, involves perceiving attributes in the trustee that are beneficial to the trustor. Similarly, the actual decision to trust corresponds to trusting intentions, reflecting the trustor's secure willingness to depend on the trustee. Trust-informed actions, in turn, influence trust-related behaviors.

In the subsequent sections, I will delve into the three dimensions of McKnight and colleagues' (2002) model, explaining their significance in understanding algorithmic trust. This model is particularly relevant as it helps us comprehend the complex dynamics between users and algorithmic systems. Exploring these dimensions enables us to comprehensively investigate trust in specific algorithms within governmental contexts, uncovering the factors influencing user behavior and decision-making processes.

Trusting beliefs refer to the perception that the trustee (the one who is to be trusted, e.g., an algorithmic recommendation system) has attributes that are beneficial to the trustor (the one who trusts, e.g., a citizen). These attributes include competence (the ability of the trustee to do what the trustor needs), benevolence (the trustee's caring and motivation to act in the trustor's interests) and integrity (the trustee's honesty and promise-keeping) (Mayer et al., 1995; McKnight et al., 2002, p. 337). It is assumed that these trusting beliefs lead to trusting intentions, as Vidotto et al. (2012, p. 576) describe: "Trusting beliefs are a solid conviction that the trustee has favorable attributes to induce trusting intentions". In case of the ICRT, this would mean that the algorithmic recommendation has attributes that are beneficial to the citizen who is filing their report.

Trusting intentions refer to the intention to engage in trust-related behaviors. It means that the trustor is "securely willing to depend, or intends to depend, on the trustee" (McKnight et al., 2002, p. 337). In this chapter, this includes the willingness to follow the algorithmic outcome. In the case of the ICRT, trusting intentions are reflected in the willingness of citizens to follow the advice given as to whether the citizen should file a fraud report or not.

Trust-related behaviors refer to actual behavior: acting according to an algorithmic recommendation (McKnight et al., 2002). If trusting beliefs and intentions are present, trust-related behaviors are likely to follow (e.g., Matook et al., 2015; Moody et al., 2014). With respect to the example of the ICRT, trust-related behavior refers to the situation where the citizen follows the algorithmic recommendation.

Meso-level trust

The second, meso level goes beyond trust in a specific algorithmic application; *organizational trust* in algorithmic governance refers to trust in the organizational embedding of algorithms. Algorithms are not stand-alone products; they are part of organizational processes. Organizations structure their work routines around the use of algorithms for their actions and decisions. This process is what Meijer and Grimmelikhuijsen (2020, p. 3) call "algorithmization", in which they distinguish six components. These components serve as a basis for citizens' trust in the organizational embedding of algorithms.

First, algorithmization requires the introduction of *technology* into organizational processes. The algorithm can be a single decision-support system, but it can also be a system that is well integrated into the organization's infrastructure. Second, the use of algorithms in an organization requires a variety of *expertise*. Experts who know how to work with the system are needed, as well as experts who can maintain the algorithm and ensure that it is properly installed in the organization's information environment. Third, the algorithm will generally build upon existing information in the organization but also produces new types of information. This means that the algorithm affects *information relations* within the organization and often outside of it when information from other actors is used. Fourth, the use of algorithms will often result in a new *organizational structure* because new collaborations between different departments could emerge. The algorithm can also result in new forms of organizational control when the algorithm dictates the implementation of processes. Fifth, *organizational policies* can be developed for the use of the algorithm in the organization. These policies will touch upon issues such as the transparency of the algorithm and responsibilities for usage or maintenance. Finally, organizations could develop methods and systems for *monitoring and evaluating* the foreseen and unforeseen outcomes of the use of algorithms in terms of output and effects (Meijer & Grimmelikhuijsen, 2020, pp. 6-7).

These six components form a guideline for analyzing the organizational process regarding algorithm use by governments. They also provide a starting point for considering ways to increase citizen trust in the use of algorithms. Meso-level trust pertains to the trust in the embedding of algorithms within organizational structures and policies, rather than focusing on trust in a specific algorithm alone. It acknowledges the complexity and interdependence of algorithms within organizations (Grimmelikhuijsen & Meijer, 2022; Zouridis et al., 2020), emphasizing that the analysis should not solely revolve around trust in individual algorithms. Instead, it encompasses trust in how algorithms are integrated into the organizational framework.

Drawing from the literature on trusting beliefs at the micro-level, I adopt a comprehensive perspective that considers competence, benevolence and integrity as key factors shaping perceptions of organizational trustworthiness (Grimmelikhuijsen & Knies, 2017; McKnight et al., 2002). Perceived competence pertains to the citizen's perception of the government's capabilities, effectiveness, skills and professionalism. Perceived benevolence captures the extent to which citizens believe that the government genuinely cares about the welfare of the public and is motivated to act in the public interest. Perceived integrity revolves around the citizen's perception of the government's sincerity, truthfulness and ability to fulfil its promises. It is important to highlight that while competence relates to the functional aspect of trust, benevolence and integrity represent ethical dimensions, reflecting the motives and ethical traits of the government (Meijer & Grimmelikhuijsen, 2020). Examining these attributes by focusing on the six components of algorithmization provides a nuanced understanding of trust in algorithmization within the context of government operations.

In the case of the ICRT, citizens' trust encompasses not only their trust in the algorithm itself but also extends to their perception of the competence, benevolence and integrity of the police regarding the various elements of the algorithmization process. Meso-level trust involves evaluating how the ICRT is integrated into the overall framework of the police organization. This integration encompasses factors such as its interaction with existing information, structures, and policies, as well as the expertise involved in its implementation and maintenance, and the systems in place for monitoring and evaluation. While the ICRT is embedded in the policy and workflow of police administrators, in practice its current use in assessing a fraud report is still limited (Soares et al., 2024).

Macro-level trust

The third and last level of analysis focusses on *institutional trust* in algorithmic governance, which refers to trust in the institutional embedding of algorithms. To date, not much attention has been paid to institutional trust in the context of algorithmic governance. Standard definitions of institutional trust refer to a trustor placing trust in the rules, roles and norms of an institution, independent of the people occupying those roles. These definitions emphasize the reliance on formal mechanisms rather than personal characteristics or organizational processes. For instance, according to Zucker (1986), institutional trust emerges when formal mechanisms are employed to establish trust, irrespective of personal attributes or prior exchanges. Similarly, O'hara (2004) argues that trust in institutions arises from a specific and often inflexible framework of codes of conduct and rules, backed by credible sanctions enforced by powerful and authoritative institutions. Cook et al. (2005) perceive trust in institutions as a reflection of individuals' confidence in the quality of the institutional arrangements within which they operate. Additionally, McKnight et al. (2011) suggest that trust based on institutions is fostered through the provision of supportive structures and mechanisms, which facilitate technology adoption by overcoming barriers. Overall, these definitions underscore the notion that institutional trust is predominantly founded on impersonal and formal institutional mechanisms rather than interpersonal relationships (Smith, 2010).

Institutional trust is especially relevant in uncertain situations, such as when algorithms are employed, as government organizations face limitations in fully anticipating or managing potential outcomes of algorithms. Beneficial structures that are established by institutions, such as supportive frameworks and mechanisms, contribute to the development of trust (Bruckes et al., 2019). They provide individuals and organizations with a sense of confidence and assurance, helping to mitigate uncertainties and enhance the overall trustworthiness of the institution. McKnight and colleagues (2011) found that legal protection is an important aspect in the context of institution-based trust. By establishing clear rules and regulations, legal protections provide a framework that guides the actions of individuals and organizations, promoting transparency and accountability. Institutional legal frameworks increase predictability and reduce the risk of unwanted outcomes. Moreover, these legal frameworks provide the basis for the punishment of inappropriate or opportunistic behaviors, further strengthening trust by deterring misconduct and ensuring compliance with ethical standards (Zucker, 1986).[4]

In the context of algorithmic governance, the object of trust is the legal embedding of algorithms. More specifically, it is about trust in the rules and regulations for embedding algorithms. However, it is important to acknowledge that the legal framework regarding artificial intelligence (AI) is currently characterized by fragmentation, ambiguity and uncertainty, despite the increasing involvement of various regulators (Ulbricht & Yeung, 2022). This raises a valid question about how trust can be established within such a context. The fragmented nature of the legal framework may create challenges in terms of transparency, accountability and the protection of individual rights (Barocas & Selbst, 2016; De

[4] In this chapter I do not elaborate on the complex relation between trust and control. For a comprehensive discussion, see Bentzen, Six and Op De Beeck (2025).

Hert & Papakonstantinou, 2016). Unclear guidelines and inconsistent regulations can undermine trust in algorithmic systems, as they may lead to unforeseen outcomes, potential biases, or violations of ethical norms (Jobin et al., 2019). Therefore, addressing the legal gaps and ensuring a robust and coherent legal framework are essential to foster trust in algorithmic governance and promote the responsible and ethical use of algorithms in a manner that safeguards individual rights and societal well-being (Etzioni & Etzioni, 2017; Paul, 2024).

Despite the existing fragmentation and uncertainty in the legal framework, there are still notable rules and regulations that guide the use of algorithms by governments. One example is the General Data Protection Regulation (GDPR) in Europe, which sets rules for the processing of personal data and emphasizes the importance of privacy and data protection. Additionally, the European Commission has issued Ethics Guidelines for Trustworthy AI, providing principles and recommendations for the development and deployment of AI systems in a manner that is transparent, accountable and respects fundamental rights (Larsson, 2020). Moreover, it is worth mentioning that, recently, the AI Act was unanimously approved by the Council of Ministers of the European Union (EU). This legislation seeks to establish a harmonized and robust regulatory framework for AI, adopting a risk-based approach, encompassing aspects such as transparency, accountability and human oversight (Floridi, 2021). These existing and forthcoming regulations serve as necessary first steps towards addressing the legal gaps and ensuring the responsible and trustworthy use of algorithms in governmental practices, contributing to the establishment of a more secure and trusted algorithmic governance landscape (Grimmelikhuijsen & Meijer, 2022).

For the ICRT case, institutional trust could be fostered by focusing on the institutional embedding of algorithms. Legal protection, clear rules and regulations are crucial in establishing a framework that guides the actions of individuals and organizations involved in the ICRT, promoting transparency, accountability and predictability. However, the fragmented nature of the legal framework surrounding algorithmic governance poses challenges, requiring efforts to address legal gaps and ensure a robust and coherent framework. By respecting existing (and forthcoming) regulations, such as the GDPR and Ethics Guidelines for Trustworthy AI, the police can work towards establishing a more secure and trusted algorithmic governance landscape that safeguards individual rights and promotes responsible and ethical use of the ICRT.

This section provided an overview of the three analytical levels to understand citizens' trust in algorithmic governance. It showed that for building, maintaining and strengthening citizens' trust in governments that increasingly use algorithms, it is not only necessary to assess how to safeguard trust in specific algorithms but also to take into account how to organizationally and institutionally embed algorithms in a trustworthy manner. The next section describes how transparency in algorithmic governance can be realized at the micro, meso and macro levels.

2.3 Transparency of algorithmic governance

To understand what transparency in algorithmic governance is and how it could play a role in building, maintaining and strengthening citizens' trust, I use mainstream government transparency literature. In general, transparency in government is denoted as "The availability of information about an organization or actor allowing external actors to monitor the internal workings or performance" (Grimmelikhuijsen & Meijer, 2014, p. 141). Following this definition, transparency can be analyzed on three dimensions. First, transparency as an *institutional relation* refers to one actor being the object of transparency (he or she can be monitored), while the other actor is the subject of transparency and monitors the first actor. Second, transparency as *information exchange* refers to the fact that transparency is often a representation of reality. For instance, actors may selectively document and report information with a specific strategic agenda in mind. Consequently, this highlights the complex nature of information dissemination and the potential impact it has on shaping the constructed reality. Third, transparency of *workings and performance* refers to the domains of government activity that are rendered transparent (Meijer, 2013, p. 430). According to Heald (2006, p. 30), actors can be transparent about what they achieve (i.e., their performance) and how they achieve these results (i.e., workings of their organization).

In addition, government transparency can be studied on multiple levels. Porumbescu et al. (2022) point out inconsistencies in the effects of transparency policies when only studying them at a single level. They show how insights between micro- (individuals), meso- (organization) and macro-level (government) transparency are interrelated. Government transparency in the context of algorithms means embedding the algorithm in the organization in such a way that outsiders can monitor how algorithms work and perform (cf. Grimmelikhuijsen & Meijer, 2014, p. 139). Transparency in algorithmic governance is generally described at the micro level (Giest & Grimmelikhuijsen, 2020) as promoting *explainability* by providing explanations (Kim & Routledge, 2018) and *accessibility* by making the data, source codes and decision trees of a specific algorithmic tool accessible to external actors (Mittelstadt et al., 2016). At the organizational and institutional levels, however, there is less information as to how transparency of algorithmic governance can take shape. This section provides a brief overview of the multi-level understanding of government transparency and applies it to transparency in algorithmic governance specifically.

Micro-level transparency

A micro-level perspective on transparency focusses on the effects and determinants of government transparency from the point of view of individuals. These can be individual citizens or other individual stakeholders such as journalists, government officials, or activists (Porumbescu et al., 2022). Thus, it focusses on the individual perceptions and attitudes of those affected by transparency. The effects of transparency on trust and legitimacy have been a focal point of many studies, examining its impact on individual citizens or civil servants (e.g., de Fine Licht, 2014; Porumbescu, 2015). Additionally, research has explored the relationship between transparency and (perceived) performance (e.g., de Boer et al., 2018; Fox, 2007), corruption (e.g., Azfar & Nelson, 2007; Park & Blenkinsopp, 2011) and accountability (e.g., Bauhr & Grimes, 2014). While micro-level transparency deals with the effects and determinants of transparency from an individual perspective, it focusses predominantly on individual perceptions and attitudes of *citizens* affected by transparency. Limited attention is paid to the effects on civil servants or to the determinants and factors that shape transparency at the micro level (Porumbescu et al., 2022).

I argue that, similarly to what was stated above for micro-level trust, it is important to assess transparency regarding the algorithm itself (as a technological artefact) in the context of the algorithm used by public organizations. *Algorithmic transparency* involves the act of making a system knowable or visible (Ananny & Crawford, 2018). The promotion of transparency in algorithmic processes involves disclosing information on the inner workings of a particular algorithmic tool, enabling interested parties to actively monitor, evaluate, criticize, or intervene, as advocated by Diakopoulos and Koliska (2017). Additionally, Nieuwenhuizen and colleagues (2021) emphasize the importance of providing explanations for algorithmic decisions or recommendations, enabling stakeholders to understand the reasoning behind such outcomes. This way, algorithmic transparency can be viewed as a mechanism that could create changes in the behavior of its users (Rader et al., 2018).

For the ICRT, algorithmic transparency could be realized by providing information to external actors such as data protection authorities to monitor the data and source codes used by internal developers of the police *(accessibility)* and by providing explanations for the recommendation procedure of the ICRT *(explainability)* (i.e., information about the algorithms and input data that were used in developing and providing the recommendation) or the motivation behind the recommendation (i.e., the individual circumstances and specific reasons that led to the recommendation). Integrating transparency right from the start of the development of an algorithm (e.g., transparency-by-design) could be a promising direction to further strengthen accessibility and explainability (see the work of Felzmann et al., 2020 for a comprehensive discussion on the concept of transparency-by-design).

Meso-level transparency

The meso- or organizational-level studies the effects and antecedents of government transparency from the point of view of an organization in its institutional and stakeholder environment. Here, transparency is understood as a set of organizational actions which may have happened in response to legal requirements or societal expectations and norms but may also be based on intrinsic motivation to disclose information to the public (Porumbescu et al., 2022). Organizational structures, funding, procedures, policies and interorganizational networks are relevant in this respect (cf. Bolman & Deal, 2017; Scott & Davis, 2015) to understand why and how organizations produce transparency.

In the context of algorithmic governance, this includes providing information about the organizational process around the use of algorithms, rather than just information about the algorithm itself. Building upon the work on algorithmization (Meijer & Grimmelikhuijsen, 2020), as discussed in the previous section, organizational transparency refers to the disclosure of how algorithms are embedded in an organization. This can include information about the type of technology used, the experts involved in the design, use and maintenance of the algorithm, the (transforming) information relations within the organization, the organizational structure, the organizational policies for algorithm use, and the monitoring and evaluation of the use of algorithms.

Whereas subdimensions of algorithmic transparency—accessibility and explainability—are commonly mentioned in studies on the micro level that focus on a specific algorithm, I argue that these dimensions are applicable to the meso level too. Public organizations could, for instance, provide information about the structural checks of the algorithmic codes that are performed by independent external actors or offer public explanations about how they monitor and evaluate algorithm use by their organizations.

For the ICRT, organizational transparency could be realized by publicly explaining which actors have been involved in the design process of this recommender system *(explainability)* or by providing access to a supervisory organization that can check and monitor the ICRT for its compliance with the police's internal organizational policies such as the Big Data Quality Framework[5] *(accessibility)*.

[5] Politie (2020). *Kwaliteitskader Big Data.* https://open.overheid.nl/repository/ronl-ac9d5901-7d8d-44f7-ae89-3d305a2506d0/1/pdf/tk-bijlage-2-kwaliteitskader-big-data.pdf.

Macro-level transparency

At the macro level, the institutional perspective is important, which relates to “systems of rules that structure the course of actions that a set of actors may choose” (Scharpf, 1997, p. 38). Institutions have both formal and informal rules (Skoog, 2005). *Formal rules* are set out in legislation, regulations, or guidelines, and *informal rules* emerge spontaneously and unintentionally over time through human interaction and take the form of unwritten conventions, routines, customs, codes of conduct and behavioral norms. Transparency, then, is viewed as an information flow to establish a basis for accountability by contributing to a hierarchical relationship that promotes surveillance of these rules (Heald, 2006). Institutional transparency emphasizes the idea of answerability because a subordinate actor (an agent) is coerced to disclose specific information to a superior (a principal) (Bovens, 2010).

When examining the translation of these concepts of answerability and the promotion of surveillance within the realm of algorithmic governance, there is an emphasis on the provision of information pertaining to rules and regulations encompassing aspects such as responsibility, accountability, privacy, or safety (Wirtz et al., 2019). These rules can be formal and explicit, codified in legislation such as the GDPR or in guidelines such as the European Commission’s Ethics Guidelines for Trustworthy AI. However, they may also be informal rules used by actors involved in the development and implementation of algorithmic governance (Grimmelikhuijsen & Meijer, 2022), such as professional norms regarding the development of responsible algorithms for public organizations. A significant distinction between the meso and macro level lies in the fact that institutional transparency pertains to external rules and regulations, while the meso level focusses on internal organizational policies. *Institutional transparency* thus addresses information about the embedding of algorithms in external rules and regulations.

For the example of the ICRT, I use the GDPR to illustrate how institutional transparency could be produced, as this is a formal rule. The GDPR codified the “right to an explanation” for users of platforms in which data is collected by mandating that (public) organizations explain what data is collected and how it is used (de Laat, 2018; Lee et al., 2019). Communicating about the existence of rules and regulations, such as the GDPR, *(explainability)* and providing access to an external actor who can monitor the organization’s compliance with the GDPR, in this case, the Dutch Data Protection Authority, *(accessibility)* is how institutional transparency can be achieved.

Table 2.1 provides a summary of the three levels of transparency in algorithmic governance that have been discussed above. The next section will link the three levels of transparency to citizens' trust to illustrate how we can use these insights to empirically study citizens' trust in algorithmic governance.

Table 2.1 Three levels of transparency of algorithmic governance

Level	Type of transparency	Definition	Operationalization
Micro	Algorithmic transparency	Information about how an algorithm functions	▪ Providing explanations about why or how an algorithmic outcome came about. ▪ Disclosing information to external actors about the internal workings of an algorithm.
Meso	Organizational transparency	Information about the organizational embedding of algorithms	Providing explanations about and disclosing information to external actors regarding these aspects of an organizations' use of algorithms: ▪ the technology; ▪ the experts involved; ▪ the information relations; ▪ the organizational structures; ▪ the organizational policies; ▪ the monitoring and evaluation.
Macro	Institutional transparency	Information about the institutional embedding of algorithms	▪ Explaining the external rules and regulations the organization has to comply with regarding their algorithm use. ▪ Disclosing information to external actors about an organization's compliance with external rules and regulations regarding their algorithm use.

2.4 Linking transparency to trust in algorithmic governance

The relationship between transparency and trust is not always straightforward. Factors such as cultural norms, information overload, complexity, a lack of understanding, prior knowledge of the topic and previous experiences can influence how transparency affects trust (Grimmelikhuijsen, 2012a; Grimmelikhuijsen & Meijer, 2014; Grimmelikhuijsen et al., 2013). Additionally, the context in which transparency and trust are examined plays a significant role in shaping their relationship (Wang & Guan, 2023). Considering these complexities, understanding the importance of transparency in building, maintaining and strengthening trust in digital encounters with governments becomes crucial.

Transparency holds significance in building trust through two compelling mechanisms in the context of digital encounters with governments. Firstly, transparency may foster accountability by ensuring that government actions and decision-making processes are visible and subject to scrutiny (Cobbe et al., 2021). By providing citizens with access to information about data collection, storage and usage, transparency enables the government to be held accountable for its actions, promoting trust by showcasing a commitment to ethical practices and acting in the best interest of citizens (Meijer & Grimmelikhuijsen, 2020). Secondly, transparency not only demystifies algorithms but also fosters a deeper understanding of digital systems. Through access to information about algorithms, data usage and decision-making procedures, citizens could gain a clearer understanding of how the government operates in the digital realm, potentially instilling confidence in its ability to make fair and informed decisions (de Fine Licht & de Fine Licht, 2020). This increased understanding could empower citizens to engage more actively in digital encounters with governments, strengthening trust (Floridi, 2020).

Building upon the empowerment of citizens through increased understanding in digital encounters with governments, it is crucial to recognize that trust in algorithmic governance extends beyond trust in the technology. It also encompasses trust in how algorithms are integrated into organizational processes, as well as trust in the rules and regulations that govern an organization's use of algorithms. In order to fully comprehend the dynamics of trust in algorithmic governance, it is essential to approach the subject from a multi-level perspective. Similarly, transparency in algorithmic governance can be considered at three levels of analysis. Micro- (algorithm), meso- (organizational) and macro-level (institutional) insights on trust and transparency in algorithmic governance are brought together to develop a framework. This framework provides directions for concrete actions necessary to build, maintain and strengthen citizens' trust in governments, which increasingly rely on algorithmic governance. Using the concrete example of the ICRT of the Netherlands Police, I illustrate below how the framework can be used for concrete actions at the three different levels.

Micro

If Dutch citizens want to file a report of online fraud, they have to use the ICRT. After reporting the facts, the ICRT analyses these facts and asks necessary follow-up questions. As soon as sufficient information is collected to determine whether or not the facts amount to a potential fraud case, the ICRT provides a recommendation to file—or not—a report of online fraud to the police. The recommendation is followed by an explanation as to the reasons for the specific recommendation. On the basis of this information, citizens can then assess whether they agree with the decision or recommendation, which in turn could foster trust in this algorithmic tool (Nieuwenhuizen et al., 2021).

Providing public access to, for example, the data used, source code, or the argumentation module of an algorithm is not always possible for the police for several reasons, such as the risk of "gaming the system" or privacy and safety concerns. Specifically, the risk of gaming the system is at stake with respect to the ICRT, which makes it difficult to provide public information about the algorithm itself. However, internal actors (ethics, data scientists) and external actors (the Inspectorate of Justice and Security) can access the source codes and check the specific indicators used in the tool, which could strengthen citizens' trust in this specific algorithmic tool.

Meso

In 2020, Amsterdam and Helsinki became the first cities in the world to launch open algorithm registers explaining how they use algorithms (Floridi, 2020). While there is a wide variety in the algorithms that are registered and those which are not, an algorithm register can be described as "a log of algorithmic decision-making systems used by a public authority that have some level of direct impact on its citizens" (Murad, 2021, p. 16). These registers thus showcase how algorithms are being used in public organizations. For now, the Netherlands Police have not released such an algorithm register. This would, however, be a way to explain how the ICRT is embedded in their organization by describing which (external) experts are involved in the design and/or maintenance of their algorithms or when, how and by whom their algorithms are evaluated. These explanations could maintain and strengthen citizens' trust in the organizational embedding of algorithms in government organizations.

Without sufficient monitoring of how algorithms are embedded in an organization, Cath and Jansen (2022) argue that it is difficult for outsiders to assess whether meso-level transparency is trustworthy. Providing access to external actors regarding, for instance, the evaluation of the ICRT, is therefore a first step in strengthening citizen trust. Ideally, detailed information about the monitoring outcomes should be disclosed, but this appears to be a complex undertaking due to the lack of resources in monitoring agencies and potential concerns related to privacy, security, or proprietary considerations regarding police algorithms. A preliminary measure involves granting external actors access to internal documents and information, enabling them to assess the embedding of ICRT within the police organization.

Macro

At the national and international levels, an observable trend is emerging wherein an increasing number of regulations, guidelines, frameworks and policies are being established to govern the use of algorithms by public and private organizations. Most well-known is the GDPR, standardizing rules for the processing of personal data by public and private actors active in the EU. The Netherlands Police are obliged to comply with regulations governing the handling of personal data in the course of their duties. These regulations are specifically outlined within the framework of the Police Data Law in the Netherlands.[6] Strengthening citizens' trust in the institutional embedding of algorithms can be achieved by providing an explanation that emphasizes the ICRT's compliance with the Police Data Law and its commitment to just and fair data processing practices. This includes ensuring that data collected by the tool is deleted or destroyed within the specified timeframes, thus promoting transparency and accountability in algorithmic decision-making processes. By adhering to principles such as data minimization and privacy by design in the case of the ICRT, the police underscore their commitment to safeguarding individuals' privacy rights and promoting responsible data handling practices in compliance with the Police Data Law (e.g., Biega et al., 2020; Schaar, 2010).

The public lacks access to information on police compliance with the Police Data Law, depriving them of control over vital aspects such as the collection of their information, the duration of data retention and the adherence to just and fair data collection practices. However, this information is crucial for citizens to trust government organizations' algorithm use. One potential approach to address this issue is by providing access to external actors who can monitor and check the police's compliance with external rules and regulations. In the example of the ICRT, the Dutch Data Protection Authority is able to check whether the police complied with the GDPR, which can contribute to citizens' trust in the institutional embedding of their algorithm use.

2.5 Conclusion

With the shift from traditional face-to-face interactions to digital encounters mediated by algorithmic tools, we need to reconceptualize our understanding of citizens' trust in government. Building on multi-level insights on government transparency (Porumbescu et al., 2022) and algorithmic transparency (Giest & Grimmelikhuijsen, 2020), this chapter shows that transparency is not only about making computer code accessible to ensure citizens' trust. In fact, transparency

[6] *Wet politiegegevens* can be accessed through https://wetten.overheid.nl/BWBR0022463/2022-10-01.

may play a role in building, maintaining and strengthening citizen trust on three different levels.

The multi-level framework developed in this chapter of transparency and trust in algorithmic governance can be used in two ways. First, since most empirical work focusses on trust in specific algorithms rather than taking a broader view to analyze how citizen trust is shaped by the organizational and institutional embedding of algorithms, this framework can provide guidance for examining the organizational and institutional aspects of the (causal) relationship between trust and transparency. Moreover, empirical work can also shed light on possible interaction effects. A nuanced perspective on these interaction effects may acknowledge the potential interconnections between different levels of trust and transparency in algorithmic governance, which entails recognizing that trust and transparency are not isolated phenomena occurring at distinct levels. Instead, there can be intricate relationships and feedback loops between micro, meso and macro levels of analysis (Porumbescu et al., 2022). This perspective acknowledges that transparency and trust can mutually shape each other across different levels. Future research can explore the interconnections between different levels and consider the potential reciprocal influences, amplifications and constraints that exist within the complex dynamics of trust and transparency in algorithmic governance.

Second, this framework may guide governments in considering the role of transparency at different levels for maintaining citizen trust in algorithmic governance. The example of the Dutch ICRT, used throughout this chapter, illustrates how government organizations may adopt transparency measures at the algorithmic, organizational and institutional levels to build, maintain, or strengthen citizens' trust in algorithmic governance. To conclude, this chapter argues that building, maintaining and strengthening citizens' trust in algorithmic governance requires a combination of technological, organizational and institutional interventions. A narrow focus on the algorithm itself fails to acknowledge the organizational and institutional context of its use. A combination of the knowledge of lawyers, engineers, organizational scientists and institutional designers is needed to develop the transparency mechanisms that are necessary to build, maintain and strengthen citizens' trust when citizen engagement with government means interacting with an algorithm rather than a human bureaucrat.

3

Micro-level: Explaining algorithmic recommendations enhances citizen trust

This chapter is based on the following published article: Nieuwenhuizen, E., Meijer, A., Bex, F., & Grimmelikhuijsen, S. (2024). Explanations increase citizen trust in police algorithmic recommender systems: Findings from two experimental tests. *Public Performance & Management Review, 48*(3), 590–625. https://doi.org/10.1080/15309576.2024.2443140

Abstract

A long-standing question in e-government research is how to maintain citizen trust in digital encounters with the government. This question is even more pertinent as algorithmic recommender systems (such as chatbots) are now becoming an integral part of digital public service delivery. The literature suggests that the explanations that these systems provide for their recommendations are crucial to maintaining citizen trust in digital encounters, but so far the empirical research into this relationship is limited. To test the effects of various explanations provided by algorithmic recommender systems on citizen trust, we conducted two experimental studies. We developed a mock version of an actual algorithmic recommender system used by the Netherlands Police and tested it in two representative survey experiments. Study 1 (n = 717) tested the effects of *procedural, rationale* and *combined explanations*. We found that providing any explanation increased trust and made citizens more likely to follow an algorithmic recommendation. Study 2 (n = 1005) investigated whether providing a *directive explanation*—specific instructions for achieving a desired service outcome—increases trust, building a more nuanced understanding of the relationship between explanations and trust in algorithmic recommendations. We conclude that explaining algorithmic recommendations—in any form—strengthens trusting beliefs, trusting intentions and trust-related behavior in citizens receiving digital public services. This may suggest that trust in algorithmic recommendations increases when citizens see that governments make an effort to provide an explanation, regardless of the nature of this explanation.

3.1 Introduction

As more interactions between government and citizens become “digital encounters” (Lindgren et al., 2019), the question of how the digitalization of public services affects citizen trust in these services becomes more urgent (Welch et al., 2005). Here, we contribute to answering this question by focusing on a rapidly emerging trend in digital public service delivery: algorithmic recommender systems such as chatbots and decision support systems that provide suggestions to citizens in various service domains. In particular, we present experimental evidence for how the explanations provided by these systems can improve citizen trust in digital services.

Algorithmic recommender systems are increasingly introduced to facilitate digital encounters and, more specifically, to support decision-making by citizens (Androutsopoulou et al., 2019; Chen & Gasco-Hernandez, 2024). They are used, for example, to support citizens when they fill out online tax forms or file a report of online fraud to the police. The literature suggests that algorithmic systems can reduce administrative burden (Fatima et al., 2020) and improve decision making (Ojo et al., 2019). However, these algorithms can also pose challenges to public organizations. A key challenge is that these systems can negatively affect citizen trust (Giest & Grimmelikhuijsen, 2020), especially when the algorithms are “black boxes” that do not provide understandable explanations of why or how a recommendation is given (Lepri et al., 2018; Rader et al., 2018). Following the literature, we expect that the transparency of the digital encounter—the interaction between citizen and algorithmic recommender system—is important for building trust. Specifically, to foster trust in algorithmic recommendations, it seems crucial that recommender systems provide citizens explanations for their recommendations (Grimmelikhuijsen, 2023).

In spite of the rapid rise of algorithmic recommender systems in government, research into trust in algorithmic recommendation is scarce, non-conclusive and primarily focused on the private sector. The few empirical studies on recommender systems for products or holidays (Tintarev & Masthoff, 2012; Wang & Benbasat, 2007) and art, books, or movies (Cramer et al., 2008; Tintarev & Masthoff, 2012) provide contradicting results. Some studies show an increase in trust by providing explanations (e.g., Kizilcec, 2016; Nothdurft et al., 2014; Wang & Benbasat, 2007) while others show a negative or non-existent effect of explanations on trust (e.g., Cramer et al., 2008; Tintarev & Masthoff, 2012).

Only recently has some empirical evidence on trust in algorithmic recommendations in the public sector been provided. Aoki (2020) found that citizens’ trust in algorithmic recommendations depends on the area or type of service. Schiff et al. (2022) found that citizens lose trust in automated decisions when there is “transparency failure”, meaning that a decision is not understood by government officials themselves. Grimmelikhuijsen (2023) distinguished accessibility and explainability as components of algorithmic transparency and showed that explaining decisions trumps accessibility in terms of generating citizen trust. These three studies, however, do not specify what kinds of explanations are needed for citizens to trust algorithmic recommendations.

Two recent papers offer relevant, but contrasting, insights into what kind of explanations strengthen (citizen) trust in algorithmic outcomes. According to a conceptual work by de Fine Licht and de Fine Licht (2020), providing *justifications for individual algorithmic decisions* is central to generating legitimacy and trust. Kizilcec (2016), on the other hand, showed that providing information to students about *the algorithmic decision procedure* regarding their grades was most effective in strengthening trust in the algorithmic outcome. Altogether, the type of explanation that is most effective in strengthening citizen trust has only been investigated to a limited extent and remains contested.

To fill this gap, we tested the effects of different types of explanations for algorithmic recommendations in the public sector on citizen trust. We aimed to answer the following research question: *What are the effects of explanations on citizen trust in algorithmic recommendations?* Our research builds on theories about government transparency and explainable artificial intelligence (XAI) to conceptualize algorithmic explanations and on psychological and e-commerce theories of trust to understand the effects of different types of explanations on citizen trust in algorithmic recommendations. For the empirical testing of the different explanations, we use an innovative *sequential factorial design* (Sniderman, 2018).

This study contributes to the literature on algorithmic transparency (de Fine Licht & de Fine Licht, 2020; Grimmelikhuijsen, 2023) by showing that citizen trust in algorithmic recommendations increases for all explanations. This indicates that trust increases when citizens see that governments make an effort to provide an explanation, regardless of the nature of this explanation. We conclude that maintaining citizen trust does not seem to depend on the exact content of the algorithmic explanation, but rather on the fact that some sensible explanation has been provided to citizens. We reflect on what this finding means for the relationship between governments and citizens in democratic societies.

3.2 Algorithmic recommendations

Explaining algorithmic recommendations

How digitization of public services affects citizen-state interactions has long been a question of interest to public administration scholars (Dunleavy et al., 2006; Welch et al., 2005), with a recent increased interest in the adoption of new (AI) technologies in public service delivery (Lindgren et al., 2019). For example, there has been a rapid rise in the use of algorithmic recommender systems (Androutsopoulou et al., 2019). Many of these recommender systems, especially in government services, rely on algorithms that operate based on predefined rules and instructions, and do not adapt to the user or learn from data (Quijano-Sánchez et al., 2020). Other recommender systems use a broader range of modern AI technologies and capabilities. These systems utilize, for example, machine learning to adapt to the user over time, or natural language processing and generation to allow the user to interact with the system in a more natural way (Makasi et al., 2021; Nai et al., 2023).

There are various specific transparency challenges associated with the use of algorithmic recommender systems. One of these is the increasing use of complex, data-driven AI technologies, such as deep learning for natural language processing and generation (Klopfenstein et al., 2017). This reliance makes the reasoning and behavior of such systems hard to understand and predict even for their own designers (Burrell, 2016), let alone citizen users. XAI is an active research field working to address the issue of (algorithmic) explainability (Arrieta et al., 2020; Miller, 2019), with techniques being developed to explain not only modern deep learning algorithms but also traditional rule-based algorithms (Lacave & Diez, 2004).

There are, however, transparency challenges that have nothing to do with the complexity or the inner workings of the algorithmic system itself. For instance, some government agencies may hesitate to explain the reasoning behind an algorithmic decision, fearing that people might attempt to "game the system", as seen in fraud prevention scenario's (Mittelstadt et al., 2016). Similarly, commercial vendors may withhold information about how a system works due to business interests, as seen in cases like COMPAS recidivism prediction (Rudin et al., 2020). Unexplained algorithmic systems may erode citizen trust because they lack basic transparency (Grimmelikhuijsen, 2023; Meijer & Grimmelikhuijsen, 2020), which is an important element in creating impartial and trustworthy institutions (Bauhr & Grimes, 2012; Rothstein & Teorell, 2008).

Government transparency research indicates that governments can be transparent about the decision process—transparency in process—and about the motivation of a decision—transparency in rationale—or a combination of the two (de Fine Licht, 2014; Mansbridge, 2009). Transparency in rationale concerns information about the substance of the decision, such as the facts and reasons on which it was based. In contrast, transparency in process refers to transparency about the decision-making process, for example, which procedures were followed and which parties were heard. Following this, we apply the distinction between process and rationale transparency to explanations of algorithmic recommendations by distinguishing "procedural explanations", "rationale explanations" and "combined explanations".

A *procedural explanation* of algorithmic recommendations contains information about the process that results in recommendations. De Fine Licht and de Fine Licht (2020) equate such a procedural explanation with what Wachter et al. (2017) called a "system functionality" explanation, that is, information about the general functionality of a system. Kizilcec (2016) tested this type of explanation in a study about the grading process of students' work. The procedural explanation contained information about how someone's grade had been established—a combination of a peer review by fellow students and an algorithm. In this study, we take a similar view, considering a procedural explanation to contain general information about the algorithm and the steps that were taken in the process leading up to the recommendation.

A *rationale explanation* contains information about the reasoning behind a specific decision, such as the weighing of specific characteristics or individual circumstances (Kim & Routledge, 2018; Wachter et al., 2017). The main aim of a rationale explanation is to provide the underlying arguments for a specific decision (de Fine Licht & de Fine Licht, 2020). For example, Grimmelikhuijsen (2023) included vignettes in which specific reasons were provided for an algorithmic decision about the rejection of a visa application and a call for a welfare fraud investigation. Here, we see a rationale explanation as information about the individual circumstances and specific reasons behind the decision by the recommender system.

Finally, both types of explanations can be put together in a *combined explanation*, entailing information about both the decision procedure and the underlying arguments for a specific decision. Kizilcec (2016), for instance, provided students with a combined explanation, including information about the grading process as well as information about the reasoning behind a specific grade. In our research, the combined explanation consists of a procedural and a rationale explanation.

Note that the above types of explanations should be understood as being on a continuum rather than as completely separate explanations. For example, as de Fine Licht and de Fine Licht (2020, pp. 918–919) argue, it is difficult to present the process leading up to a decision without giving some insight into the reasons on which the decision is based, in the same way that it is difficult to provide the reasons for a decision without mentioning something about the process that led up to the decision.

Although far from conclusive, the literature suggests that these different types of explanation may result in different effects on citizen trust. To investigate these effects, we will present a conceptualization of citizen trust in the next section and develop hypotheses about the expected effects of the different explanations on citizen trust.

Trusting algorithmic recommendations

Citizens' trust in government in general and in digital public encounters specifically is an important goal in itself and also has instrumental value. First, citizen trust is an important public value to be pursued by government organizations (Moore, 1995) and can be seen as a goal in itself. Since citizens have no realistic "exit" option for public services (Hirschman, 1970), they are dependent on a single service provider. Ensuring citizen trust in service providers they are "stuck with" should therefore be at the core of (digital) public service delivery. Second, trust may serve an instrumental value (Hoff & Bashir, 2015; Mei & Zheng, 2024). Public organizations have initiated algorithmic recommendations, for instance, to increase the efficiency or quality of public service delivery. Increased trust will also increase citizens' willingness to follow algorithmic advice, such as acting according to a recommendation (McKnight et al., 2002). In the following sections, we conceptualize trust in algorithmic recommendations and describe how different explanations might affect trust in algorithmic recommendations, leading to four hypotheses.

Researchers in computer and information sciences often use the concept of human-computer trust (HCT) when investigating users' trust in algorithmic recommendations. HCT is "the extent to which a user is confident in, and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent decision aid" (Madsen & Gregor, 2000, p. 1). Two elements of trust are captured in this definition. First, the *confidence* of a user in the recommendations, actions and decisions of the decision aid. Second, the *willingness* of the user *to act* on those recommendations, actions and decisions of the decision aid. Literature on e-commerce elaborates extensively on the central role trust plays in making consumers comfortable acting on vendor advice and making purchases. Similar ideas about trust—as confidence and willingness to act—can be found in this research area.

McKnight and colleagues (2002) developed an influential model of trust in information systems, consisting of trusting beliefs, trusting intentions and trust-related behaviors. *Trusting beliefs* refer to the perception that the trustee (the one who is to be trusted, e.g., an algorithmic recommendation system or a public organization) has attributes that are beneficial to the truster (the one who trusts, e.g., a citizen). The most cited trusting beliefs are competence (ability of the trustee to do what the truster needs), benevolence (trustee's caring and motivation to act in the truster's interests) and integrity (trustee's honesty and promise keeping) (McKnight et al., 2002, p. 337). Eventually, trusting beliefs lead to trusting intentions, as Vidotto et al. (2012, p. 576) describe: "Trusting beliefs are a solid conviction that the trustee has favorable attributes to induce trusting intentions". In our context, this would mean that an algorithmic recommendation by a public organization has attributes that are beneficial to a citizen, leading to trusting intentions.

Trusting intentions refer to the intention to engage in trust-related behaviors. It means that the truster is "securely willing to depend, or intends to depend, on the trustee" (McKnight et al., 2002, p. 337). In this paper, this includes citizens' willingness to follow the advice in an algorithmic recommendation.

Trust-related behaviors refer to actual behavior: acting according to an algorithmic recommendation (McKnight et al., 2002). If trusting beliefs and intentions are present, trust-related behaviors are likely to follow (e.g., Matook et al., 2015; Moody et al., 2014). In this paper, we see trust-related behaviors as the actual behavior of citizens when they follow an algorithmic recommendation. We adopt these three dimensions to investigate trust in algorithmic recommendations in digital public service delivery.

While recognizing that trust in government algorithmic recommender systems is influenced by various factors, as shown by Aoki (2020), we specifically focus on understanding the impact of explanations. Aoki's work demonstrates that clear communication of purposes that directly benefit citizens, such as ensuring consistent response quality and timely interactions, plays a significant role in building trust in chatbots. Additionally, her study reveals interesting variations in initial trust levels when a government announces the use of AI chatbots in different areas (parental support and waste separation). While these factors contribute to trust, our paper's primary aim remains to thoroughly investigate how explanations influence trust in algorithmic recommendations. The next section delves into various ways in which explanations may affect citizen trust in algorithmic recommendations.

How explanations might affect trust in algorithmic recommendations

Many scholars claim that transparency mechanisms, such as providing explanations, should be implemented to increase user trust in algorithmic recommendations. On the basis of the few empirical studies on algorithmic transparency (Grimmelikhuijsen, 2023; Kizilcec, 2016) and the variety of conceptual publications on the topic (Ananny & Crawford, 2018; de Fine Licht & de Fine Licht, 2020; Meijer & Grimmelikhuijsen, 2020), we formulate four hypotheses regarding the relationships between different types of explanations and citizen trust in algorithmic recommendations. Grimmelikhuijsen (2023) found that explaining algorithmic decisions generates citizen trust, but he did not specify what types of explanations are needed for trust to increase. Kizilcec (2016), on the other hand, tested the effects of different explanations but did so outside the public sector context. Providing explanations has been found to be effective only when expectations have been violated (Kizilcec, 2016; Rader et al., 2018). We take this notion into account by only looking into decisions that violate citizens' expectations.

Recent studies in public administration have found positive effects of algorithmic transparency, specifically explainability, on citizen trust (Grimmelikhuijsen 2023; Schiff et al., 2022), but they don't examine or specify types of explanations. A longer tradition of researching different types of explanations for algorithmic outcomes exists in computer science. We can draw on insights from this field to understand what effects different explanations might have on citizen trust in algorithmic recommendations. Nothdurft et al. (2014) have investigated the effects of two types of explanations on trust in human-computer interactions. First, the authors looked at "transparency explanations", which focus on increasing the understanding of users regarding how a technical system works. This is similar to a procedural explanation. Second, they looked at "justifications", which include a motivation for a specific decision. The goal of this explanation is "to provide support for and increase confidence in a given system advices or actions" (Nothdurft et al., 2014, p. 53). This type of explanation corresponds with the rationale explanation. They found both explanations to be effective in strengthening user trust.

These positive relationships also appeared in other studies. We will discuss findings for the procedural explanations first and then elaborate on the rationale explanations. Studies by Kizilcec (2016), Nothdurft et al. (2014) and Rader et al. (2018) found support for a positive relationship between a procedural explanation and trust in an algorithmic decision but provide different reasons for the mechanism behind this effect. Kizilcec (2016, p. 2391) referred to procedural justice theory (Tyler, 1990), which posits that individuals can be satisfied with negative outcomes as long as the underlying procedure is considered to be just.

Nothdurft et al. (2014) and Rader et al. (2018) provided alternative reasonings for the positive effects on user trust of explaining the decision procedure. The mechanism, as Nothdurft et al. (2014) explained, is that a procedural explanation helps users to understand how a system works and reasons and thereby changes the user's perception of the process from a black-box model to a comprehensible system. This allows the users to build a mental model of the system, including its underlying reasoning processes (Nothdurft et al., 2014, p. 53). Rader et al. (2018) followed a similar line of reasoning. They argued that knowing how an algorithmic recommender system produces a recommendation helps users to understand and act upon the recommendation. A procedural explanation helps to fill the gap between users' intuitions about a system and the actual internal processes of a system. In fact, this makes explanations "white box" descriptions of the inputs, outputs and steps a system takes to arrive at a specific outcome (Rader et al., 2018, p. 2), which increases user trust in the system. Despite disagreement about the underlying mechanisms, all studies show the positive effects that a procedural explanation can have on trust in algorithmic outcomes. We therefore expect a positive relationship between a procedural explanation and trust in an algorithmic recommendation. Trust will be higher with a procedural explanation than when no explanation is provided.

> ***Hypothesis 1:*** *Citizen trust (i.e., trusting beliefs, intentions, and trust-related behaviors) in algorithmic recommendations with a procedural explanation is higher than in algorithmic recommendations without an explanation.*

To investigate the effects of the rationale explanation, we draw on studies from Hind (2019) and Kim and Routledge (2018), who argued that a rationale explanation could lead to an increase in trust in algorithmic decisions because people understand *why* they receive what they receive. In addition, de Fine Licht and de Fine Licht (2020, p. 924) argued that if citizens know the reasons behind algorithmic decisions, they should have a greater understanding of why the decisions were made. A rationale explanation allows citizens to form an opinion about the desirability of the decision. Similarly, Rader et al. (2018, p. 2) described how users feel more comfortable and satisfied with a recommendation and are more willing to accept it when they believe they understand why the recommendation was made. Our second hypothesis therefore expects trust to be higher with a rationale explanation than when no explanation is provided about the algorithmic recommendation.

> ***Hypothesis 2:*** *Citizen trust (i.e., trusting beliefs, intentions, and trust-related behaviors) in algorithmic recommendations with a rationale explanation is higher than in algorithmic recommendations without an explanation.*

While these two hypotheses provide expectations for procedural and rationale explanations separately, the literature provides mixed insights into the comparative strengths of the effects. Some scholars argue that a rationale explanation ensures a better understanding of a decision outcome compared to a process explanation (de Fine Licht & de Fine Licht, 2020). For a procedural explanation to be understandable, they argue, the code would likely need to be much simpler than it is in current systems, assuming that the public needs a chance to truly grasp it. A simplified code would potentially make it easier to manipulate the system and give fewer incentives to innovate (e.g., Lepri et al., 2018; Zarsky, 2016). Demands for transparency would therefore likely result in decisions of inferior quality and, as a result, less trust (de Fine Licht & de Fine Licht, 2020, p. 921).

Empirical work on algorithmic trust from a non-public context paints a different picture, however. Rader et al. (2018) found no statistical differences between procedural and rationale explanations about Facebook's News Feed algorithm but acknowledged that this might be the result of too little difference between the two types of explanations in the operationalization of their study. In another study, Kizilcec (2016) examined the effect of different explanations (no, procedural and combined) on trust in an algorithmic interface. He found that an explanation outlining how an algorithm worked (i.e., procedural explanation) had the strongest positive effect on trust in the outcome. This can be understood using procedural justice theory, which posits that individuals can be satisfied with negative outcomes as long as they consider the underlying procedure to be just (Tyler, 1990). Opaque algorithmic procedures, therefore, may erode trust in algorithmic recommendations (Rudin, 2019). In sum, this means that there is scarce, and contradictory, evidence for which type of explanations work better in terms of trust. Since Kizilcec's experiment more closely aligns with the context of algorithmic recommendations, we hypothesize that procedural explanations are most effective and therefore trump rationale explanations.

> ***Hypothesis 3:*** *Citizen trust (i.e., trusting beliefs, intentions, and trust-related behaviors) in algorithmic recommendations with a procedural explanation is higher than in algorithmic recommendations with a rationale explanation.*

Our fourth and final hypothesis concerns combining procedural and rationale explanations, which could perhaps provide the best of both worlds. At the same time, there is a risk that people will receive so much information that they do not read the message properly, which leaves them dissatisfied with the final recommendation or decision. This is also referred to as an *information overload* (Hosseini et al., 2015). An information overload in an explanation can lead to a decrease in trust. Finding the relevant information would be, as de Fine Licht and de Fine Licht (2020, p. 922) noted, "as difficult as finding the proverbial needle in a haystack". Making all information available (full transparency) would therefore not be more beneficial than providing a single explanation (partial transparency). This expectation is supported by Kizilcec's experiment (2016), which found that providing a combination of rationale and procedural explanations decreased trust. This has led to the fourth and final hypothesis, which expects a combination of a procedural and a rationale explanation to be less effective in strengthening citizen trust in algorithmic recommendations than a single explanation.

> ***Hypothesis 4:*** *Citizen trust (i.e., trusting beliefs, intentions, and trust-related behaviors) in algorithmic recommendations with a combined explanation is lower than in algorithmic recommendations with only a procedural or a rationale explanation.*

In the next section, Study 1, we test these four hypotheses.

3.3 Study 1: Initial test of the four hypotheses

The Intelligent Crime Reporting Tool

We developed a mock version of an actual algorithmic recommender system, namely the Intelligent Crime Reporting Tool (ICRT) of the Netherlands Police (Odekerken et al., 2022). Our research therefore addresses a recent call by Zuiderwijk et al. (2021) to move beyond the generic focus on AI in public administration research, focusing on a specific AI tool (ICRT) in a specific domain (police), in a specific country (the Netherlands).

The ICRT can be accessed via the police website[7] by citizens who believe they have been victims of online fraud, such as scams involving fake web shops or malicious secondhand traders on platforms like eBay, where they order and pay for a product but never receive it. Through the ICRT, citizens can describe what happened to them, with the ICRT asking follow-up questions when necessary. The ICRT then automatically assesses whether it is likely that the citizen has been a victim of fraud, and, if so, recommends that they proceed with filing an official report online. If the ICRT assesses that fraud did not occur, it can provide recommendations for other actions the citizen can take (e.g., contacting the trading platform). Our mock version of the ICRT asked participants the same questions as the real ICRT of the Netherlands Police.

The ICRT primarily relies on a rule-based legal model of the fraud domain, in combination with some basic natural language processing for analyzing the

[7] See https://aangifte.politie.nl/iaai-preintake/#/.

free-text descriptions of (alleged) fraud, provided by citizens. It is an example of digital public services, defined as "public services provided using internet-based technologies wherein a citizen's interaction with a public organization is mediated partly or completely by an IT-system" (Lindgren et al., 2019, p. 429). With online shopping being increasingly commonplace, the Netherlands Police receive tens of thousands of complaints regarding online trade fraud every year. In these complaints, it is not always clear whether a case is fraud or not, particularly since citizens do not know the nuances of the law and hence tend to have a hard time identifying which facts are relevant from a legal standpoint. So, in addition to allowing citizens to file a report of a crime online, the ICRT also acts as an official source of information, explaining why a citizen's case is (or is not) a case of fraud and recommending further actions to take besides filing an official report. We believe that asking for advice on and possibly reporting a crime to the police through the ICRT is a very relevant case to investigate when learning about trust in algorithmic recommender systems because it sheds light on a digital public service that many citizens are likely to use.

Materials and methods

Experimental setting and procedure

To examine the previously stated research question, we designed an online survey experiment in which we randomly varied the type of explanation that followed the recommendation that participants received. The design is shown schematically in Figure 3.4 in the Appendix. This fully randomized experiment has the advantage of high internal validity (Shadish et al., 2001). Therefore, we can draw firm cause-effect conclusions about the effects of different types of explanations on trust in algorithmic recommendations.

As illustrated by Figure 3.4 in the Appendix, the experiment started by asking all participants three demographic questions in order to ensure a representative sample of the Dutch population in terms of gender, age and education. Next, participants were told that they would read a hypothetical case of possible online fraud and that they should use the Intelligent Crime Reporting Tool (ICRT) to file a report of this case.

After using the ICRT (going through all the steps and questions illustrated in Appendix 3D) to report their case of online fraud, all participants received a recommendation not to file an official report of online fraud. This recommendation went against their expectations, as they were told in the hypothetical situation that they suspected to be a victim of online fraud. Participants were randomly assigned to a group in which an experimentally varied explanation was presented. The potential fraud situation, the recommendation and the experimental vignettes used for the explanation can be found in the appendix.

After reading the recommendation and the explanation, participants were first asked whether they wanted to file a report of online fraud, which allowed us to measure their actual behavior. Then, they were asked about their trusting beliefs and intentions regarding the recommendation they received.

Sample and data collection

We conducted the survey experiment online using Qualtrics in June 2021. Prior to data collection, we preregistered the experiment using the Open Science Framework format (Bowman et al., 2020), and the study received ethical approval from the institutional Ethical Review Committee.[8] Data was collected using the sample-only service of *Dynata*, a renowned global recruitment firm. Dynata has a large respondent pool, which they use for distributing surveys. Respondents are free to choose whether they want to apply for Dynata's participant pool and could therefore decide for themselves whether they wanted to participate in our experiment. Dynata provided a sample of 717 respondents for our experiment with the following parameters: 1) Dutch speakers living in the Netherlands, and 2) participants that represented the country's population in terms of age, gender and level of education. Respondents were reimbursed for their participation in the survey-experiment upon completion by Dynata. Using the software program G*Power, we conducted an *a priori* power analysis to calculate the estimated sample size (Power = .9 and $\alpha = .008$).[9] The sample for the experiment was $n = 717$. We used stratified sampling methods to ensure we had a representation of the Dutch population. The background variables of the sample are reported in Appendix 3E. The sample resembles the Dutch population regarding three key background variables: education, gender and age. We took these background variables into account as control variables to carry out balance checks. Participants in our sample were somewhat older compared to the Dutch population (see Appendix 3E).

[8] Open Science Framework Registration: https://osf.io/d862z/?view_only=59da5dfefa934dccb39ad-73a0cdcae74.

[9] We used *p*-values with a significance level of alpha = 0.05 for all tests. However, we used a Bonferroni adjustment to correct for an inflated chance of Type 1 errors due to multiple testing. For each outcome variable, we conducted six comparisons between groups. Therefore, in our analysis we multiplied every *p*-value by six. G*Power, however, doesn't allow adjustment of the *p*-value, so we performed the power analysis with the equivalent adjustment of alpha (.05) divided by six.

Experimental conditions

We have four different explanation conditions in our experiment (see Appendix 3G for the exact wording of the explanation manipulations). Participants in the control condition received no explanation for the statement that, based on their story, they did not need to file a report to the police. The procedural condition included information about the decision procedure: how the ICRT used various algorithms to analyze their text and arrive at the conclusion that the webshop was trustworthy. Participants in the rationale condition were told the specific reason why the webshop was trustworthy: it was affiliated with a quality mark, which guaranteed an extensive screening procedure for all associated webshops.[10] The combined condition included both the procedural and the rationale explanations.

Measures

The trusting beliefs of participants were measured after the experiment by means of a questionnaire, using Gulati et al.'s (2019) human-computer trust scale that investigates user trust in human-computer interactions. We used cognitive interviews to test the scale (DeVellis, 2017). By asking eight people unfamiliar with the topic what they understood the items to be about and how they would formulate responses to the items, we were able to identify confusion about vocabulary and concepts as well as misunderstandings related to response options that we had overlooked. This led to using four items that measure general feelings of trust. We have adapted these items according to the research needs of the current study. Trusting beliefs were thus measured using four items (alpha = .96) on a scale from 1 ("totally disagree") to 7 ("totally agree").

Similar to participants' trusting beliefs, trusting intentions were measured using a questionnaire. The scale for the intention to act according to the recommendation was derived from the scale measuring intention to follow vendor advice by McKnight et al. (2002). We adapted the scale to the context of the ICRT and shortened the scale based on the cognitive interviews. Trusting intentions were measured using four items (alpha = .96) on a scale from 1 ("totally disagree") to 7 ("totally agree"). The items used for measuring trusting beliefs and intentions can be found in Appendix 3A.

Before questioning the participants about their trusting beliefs and intentions, participants were asked whether or not they wanted to report the crime (trust-

[10] The quality mark that the ICRT refers to, claims to be the biggest web shop quality mark in the Netherlands and Europe (WebshopKeurmerk, n.d.). It is therefore likely to assume that participants have heard of this quality mark before.

related behaviors), measured as “Don’t file a report”/“File a report”. This allowed us to record the actual behavior of participants based on the recommendation.

The survey was extensively pre-tested prior to its implementation in three ways. First, we asked two developers of the real ICRT from the Netherlands Police to revise the experimental vignettes in order to increase mundane realism. Second, to ensure measurement validity, we carried out eight face-to-face cognitive interviews to ensure survey questions were well understood, which led to some revisions of the survey items. Third, we conducted a pilot study with 82 people to test the reliability of the new scales of trusting beliefs and intentions. Furthermore, we used a factual manipulation check (FMC) in the pilot study to test the experimental vignettes. Findings from the FMC in the pilot study resulted in some changes in the vignettes to make them more distinctive in Study 1. The results of the FMC of both the pilot study and Study 1 can be found in Appendix 3H.

Analyses

For the outcome variables *trusting beliefs* and *trusting intentions* we used independent samples *t*-tests to compare means between groups. For *trust-related behaviors* we used *z*-tests to compare proportions between groups. We used *p*-values with a significance level of alpha = 0.05 for all tests. However, we used a Bonferroni adjustment to correct for an inflated chance of Type 1 errors due to multiple testing. For each outcome variable, we conducted six comparisons of means or proportions between groups. Therefore, we multiplied every *p*-value by six. To calculate effect sizes, we used Cohen’s *d* for the outcome measures trusting beliefs and intentions, and Cohen’s *h* for the outcome measure trust-related behaviors (Cohen, 1988).

Results

Descriptive statistics

Means and standard deviations for trusting beliefs and intentions can be found in Table 3.1. Participants that received no explanation for the recommendation by the intelligent crime reporting tool scored on average between 3 (fairly disagree) and 4 (neither agree nor disagree) on a scale from 1-7 for their trusting beliefs and intentions. Participants that received a procedural, rationale, or combined explanation scored on average between 4 (neither agree nor disagree) and 5 (fairly agree) on a scale from 1-7 for their trusting beliefs and intentions. Furthermore, for trust-related behaviors we reported the percentages of participants that did not report the crime in Table 3.1, thus the percentage of participants that followed the algorithmic recommendation not to report the crime. Less than half of the participants that received no explanation followed the recommendation not to report the crime, compared to approximately three quarters of the participants that received any type of explanation.

Table 3.1 Descriptive statistics of trusting beliefs, intentions, and trust-related behaviors in the algorithmic recommendation for different types of explanations

	n	**Trusting beliefs**		**Trusting intentions**		**Trust-related behaviors**
		Mean	**SD**	**Mean**	**SD**	**% that did not report the crime**
No explanation	177	3.59	1.54	3.61	1.67	42.29%
Procedural explanation	177	4.75	1.48	4.84	1.52	74.58%
Rationale explanation	179	4.86	1.30	4.88	1.50	75.98%
Combined explanation	184	4.87	1.41	4.92	1.58	76.09%

Analyses

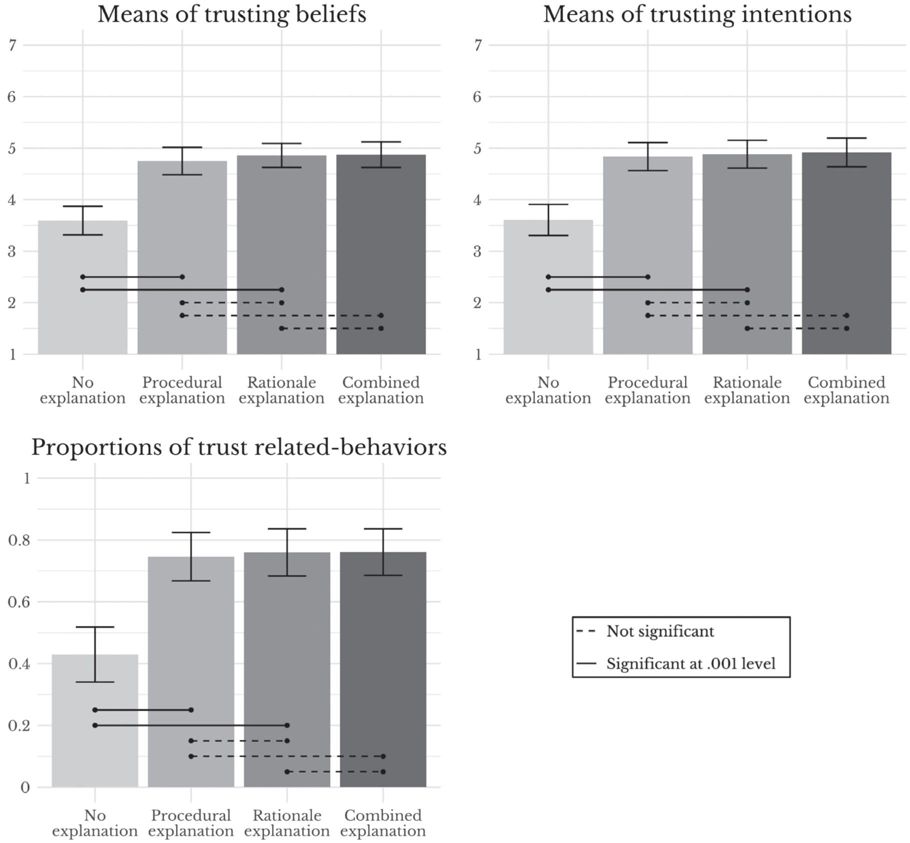
We tested four hypotheses, the results of which are visualized in Figure 3.1. We found support for the first hypothesis (procedural > no explanation). Participants that received a procedural explanation had significantly higher *trusting beliefs* than participants that received no explanation, $t(351.49) = 7.18$, $p < .001$. This is a medium effect ($d = .76$). In addition, participants that received a procedural explanation had significantly higher *trusting intentions* than participants that received no explanation, $t(348.76) = 7.25$, $p < .001$, with a medium effect ($d = .77$). Finally, the proportion of participants with a procedural explanation that followed the recommendation not to report the crime (*trust-related behaviors)* was significantly higher than the proportion of those with no explanation, $z = 6.05$, $p < .001$, with a medium effect ($h = .66$).

We also found support for the second hypothesis (rationale > no explanation). Participants who received a rationale explanation had significantly higher *trusting beliefs* than participants that received no explanation, $t(343.28) = 8.34$, $p < .001$, with a large effect ($d = .89$), and significantly higher *trusting intentions,* $t(349.22) = 7.58$, $p < .001$, with a large effect ($d = .80$). Finally, for *trust-related behaviors*, the proportion of participants with a rationale explanation that followed the recommendation not to report the crime was significantly higher than the proportion of those with no explanation, $z = 6.35$, $p < .001$. This is a medium effect ($h = .68$).

We did not find any support for the third hypothesis (procedural > rationale). Participants with a procedural explanation did not have significantly higher *trusting beliefs* than participants with a rationale explanation, $t(347.25) = -.74$, *p* = 1, nor did they have significantly higher *trusting intentions,* *t*(353.58) = -.30, *p* = 1, or *trust-related behaviors, z* = .31, *p* = 1.

Lastly, we did not find support for the fourth hypothesis (combined < procedural or rationale). Participants with a combined explanation did not have significantly lower *trusting beliefs* (*t*(356.2) = .81, *p* = 1), *trusting intentions* (*t*(359) = .50, *p* = 1), or *trust-related behaviors* (*z* = .33, *p* = 1) than participants with a procedural explanation. Similarly, participants with a combined explanation did not have significantly lower *trusting beliefs* (*t*(360.05) = .10, *p* = 1), *trusting intentions* (*t*(360.83) = .20, *p* = 1) or *trust-related behaviors* (*z* = .02, *p* = 1) than participants with a rationale explanation. A sensitivity analysis excluding participants that failed the factual manipulation check did not alter the results of our analyses (i.e., no significant effect became nonsignificant, and no nonsignificant effect became significant).

Figure 3.1 Analyses of differences between explanation conditions



Note. Error bars are confidence intervals with Bonferroni-adjusted standard errors.

Conclusion study 1

To answer our research question "What are the effects of different types of explanations on citizen trust in algorithmic recommendations?" we examined four theoretical expectations. In line with the first two hypotheses, our findings show that citizen trust in algorithmic recommendations is significantly higher with a procedural or a rationale explanation than without an explanation. Our results showed that providing any type of explanation will increase trust in algorithmic recommendations compared to providing no explanation. Regarding the third and fourth hypotheses, there were no significant differences in trust between a procedural and a rationale explanation, nor were there any significant differences in trust between providing a combined explanation instead of merely a procedural or a rationale explanation. Using sequential factorials, which are "a model of a sequence of experiments to progressively deepen and draw out the implications of a line of reasoning" (Sniderman, 2018, p. 266), we tested a possible interpretation for these null findings in Study 2. This means that the results of the first experiment determined the design and input of the follow-up experiment, which helped to deepen our understanding of which explanations increase citizen trust in algorithmic service provision. In line with our sequential factorial design, we identified a new concept, a directive explanation, that seemingly plays a key role in interpreting findings that cannot be understood by concepts previously discussed in the literature. This concept led to a fifth hypothesis that was tested in Study 2.

3.4 Study 2: Follow-up test with a directive explanation as an additional manipulation

In line with the principles of sequential factorials design (Sniderman, 2018), we explored the literature to identify a new concept which could help us to better understand the relationship between explanations and citizen trust. In contrast to the hypotheses formulated in Study 1, we did not find any significant differences between different types of explanations. While this could be a true null effect (i.e., there is no effect of explanation type), it is also possible that this lack of differentiation is an artifact of our experimental design and the way the explanations were formulated. Specifically, every treatment contained the following sentence: "Check the website of the quality mark (www.webshopqualitymark.nl) to see what you can do to get your money back". Providing such a sentence offers citizens a concrete path forward with actions to undertake in order to achieve their desired outcome, i.e., to get their money back. After all, an important reason for victims of online fraud to report their case is to get their money back (Cross, 2018).

According to Singh et al. (2021a), explaining which steps to take in order to achieve the desired result is covered by a so-called directive explanation. A *directive explanation* lists specific actions or interventions an individual needs to take to achieve a desired outcome (Singh et al., 2021a, p. 1). If a recommendation is detrimental to a citizen (e.g., advising against filing a report of fraud due to its low likelihood of success), then a directive explanation provides information about how the citizen could obtain their desired outcome, if possible.

A directive explanation can be seen as a layer of explanation on top of other types of explanations to justify algorithmic decisions that affect people personally (Singh et al., 2021b). It increases an individual's sense of control over a decision outcome, which could strengthen trust. In the context of our research, this could mean that citizen trust in algorithmic recommendations is higher when a directive explanation is provided, because it potentially brings a recommendation more in line with an individual's preference. We therefore hypothesize that providing a *directive explanation* is a specific element in an explanation that may foster trust.

> ***Hypothesis 5:*** *Citizen trust in algorithmic recommendations with a directive explanation is higher than in algorithmic recommendations without a directive explanation.*

We cannot formulate hypotheses about the effect of a directive explanation on specific types of explanations, such as a procedural, rationale, or combined explanation because it has not been empirically examined before. Thus, in the next section and main analysis, we look at the effect of a directive explanation in general. This allows us to contribute to theory development about explaining public sector algorithmic recommendations.

Materials and methods

Experimental setting and procedure

Our second study builds on what we learned from the findings and limitations of our first survey experiment. The experimental setting and procedure are the same as the first experiment, except for the experimental treatments. As illustrated in Figure 3.5 in the Appendix, Study 2 has eight experimental groups instead of the four in Study 1. Study 2 investigates whether citizen trust in algorithmic recommendations is higher if a directive explanation is provided.

The *directive explanation* is operationalized as follows: "Check the website of the quality mark (www.webshopqualitymark.nl) to find out what you can do to get your money back". This element provides participants with concrete steps forward to achieve their desired outcome (i.e., get their money back). In our study we use the element of the directive explanation, as operationalized above, to test a greater degree of directiveness versus little to no directiveness. See Appendix 3J for the experimental vignettes used in Study 2. A subjective manipulation check (SMC) shows that participants experienced the treatment as we intended (see Appendix 3L).

Sample and data collection

We conducted the survey experiment online on Qualtrics in September 2021 using the sample-only service of the recruitment firm *Dynata*. Prior to data collection, we preregistered the experiment using the Open Science Framework format (Bowman et al., 2020).[11],[12] Additionally, the study was approved by the institutional Ethical Review Committee. We conducted an *a priori* power analysis to calculate the estimated sample size (Power = .9 and α = .05) using the software program G*Power.[3] The sample for the experiment was *n* = 1005. The sample resembles the Dutch population regarding level of education, gender and age, although the participants in our study were slightly older in comparison to the Dutch population (see Appendix 3E).

Analyses

Based on the hypothesis, we compared the means of trusting beliefs and intentions of Groups 1-4 with the means of trusting beliefs and intentions of Groups 5-8 using *t*-tests with planned contrasts. For trust-related behaviors we used a *z*-test to compare proportions between groups. We compared the proportion of trust-related behaviors of Groups 1-4 with the proportion of trust-related behaviors of Groups 5-8 using planned contrasts. We used *p*-values with a significance level of α = .05. To calculate effect sizes, we used Cohen's *d* for the outcome measures trusting beliefs and intentions and Cohen's *h* for the outcome measure trust-related behaviors (Cohen, 1988).

Results

Descriptive statistics

Table 3.2 shows descriptive statistics for trusting beliefs, intentions, and trust-related behaviors. The descriptive statistics for the two main sets of participants

[11] Open Science Framework Registration: https://osf.io/hyuxm/?view_only=377a4f29c7ab414d897652fd2f789180.

[12] The terminology has been changed from "action perspective" to "directive explanation".

(with and without a directive explanation) are shown in bold. All participants, receiving a directive explanation or not, scored on average between 4 (neither agree nor disagree) and 5 (fairly agree) on a scale from 1-7 for their trusting beliefs and intentions. Furthermore, more than two-thirds of all participants followed the recommendation not to report the crime. Table 3.2 also shows the descriptive statistics for the explanation subsets (Groups 1 to 4) that received no directive explanation and the subsets (Groups 5-8) that did receive a directive explanation.

Table 3.2 Descriptive statistics of trusting beliefs, intentions, and trust-related behaviors in the algorithmic recommendation with or without a directive explanation

	n	**Trusting beliefs**		**Trusting intentions**		**Trust-related behaviors**
		Mean	**SD**	**Mean**	**SD**	**% that did not report the crime**
No directive explanation	**497**	**4.60**	**1.47**	**4.69**	**1.57**	**69.22%**
G1: No explanation	125	4.03	1.52	3.96	1.60	55.20%
G2: Procedural explanation	123	4.44	1.46	4.43	1.57	64.23%
G3: Rationale explanation	123	4.98	1.28	5.15	1.35	79.67%
G4: Combined explanation	126	4.95	1.42	5.22	1.40	77.78%
Directive explanation	**508**	**4.69**	**1.41**	**4.75**	**1.53**	**72.24%**
G5: No explanation	129	4.25	1.48	4.25	1.62	58.91%
G6: Procedural explanation	125	4.92	1.29	4.99	1.35	74.40%
G7: Rationale explanation	124	4.98	1.20	5.11	1.34	83.06%
G8: Combined explanation	130	4.64	1.53	4.66	1.63	73.08%

Analyses

We did not find support for our fifth hypothesis that citizen trust in algorithmic recommendations with a directive explanation is higher than in algorithmic recommendations without a directive explanation. Participants that received the directive explanation manipulation did not have significantly higher *trusting beliefs* than participants that received no directive explanation ($t(998.64) = 1.06$, $p = .144$). In addition, no significant effect was found on *trusting intentions* ($t(1000.6) = .64$, $p = .260$) or *trust-related behaviors* ($z = 1.06$, $p = .854$). Overall, the main analysis showed that participants that received an explanation with a directive explanation did not have significantly higher trust in algorithmic recommendations than participants that received an explanation without a directive explanation.

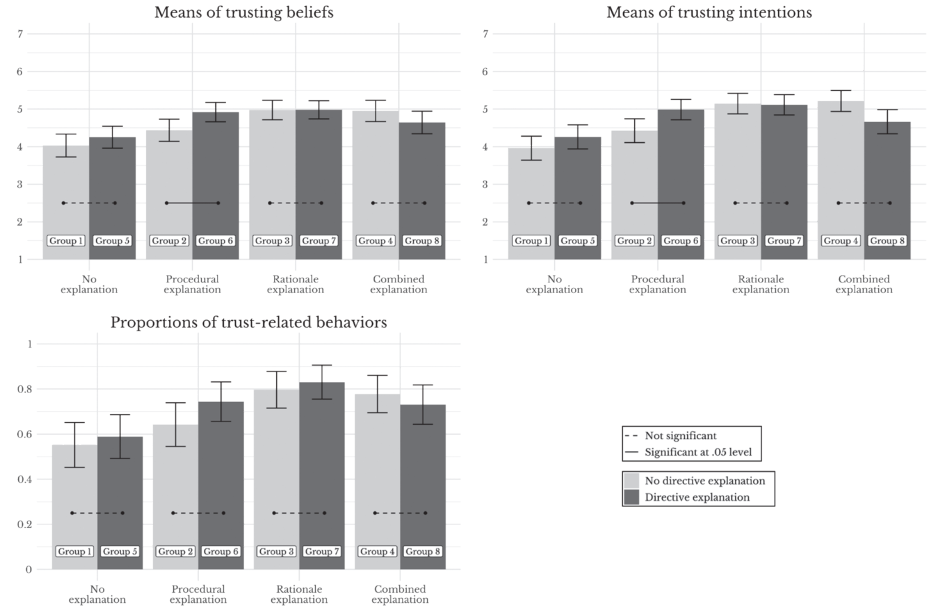
Post-hoc analyses

We introduced a directive explanation as a new concept that could be an important element in explaining algorithmic decisions. Due to its novelty, we could not build on theoretical foundations to develop specific theoretical expectations. In other words, we were only able to formulate a hypothesis for the main effect of a directive explanation on citizen trust. To examine whether a directive explanation has an interaction effect, meaning different effects on specific explanations, we conducted four post-hoc analyses.

Following H5, we expected that the groups with a directive explanation had higher trust in algorithmic recommendations than the groups without a directive explanation. We compared all groups with equal explanation treatments separately, so we compared Groups 1 and 5, 2 and 6, 3 and 7, and 4 and 8. All comparisons were directional, expecting that the groups with a directive explanation would have higher trust scores than the groups without a directive explanation, as stated in H5. As in Study 1, to adjust for the inflated chance of Type 1 errors, we used a Bonferroni adjustment. For these post-hoc analyses we multiplied the *p*-value by four because we conducted four comparisons between groups, with an alpha of .05. The full results of all four post-hoc comparisons can be found in Appendix 3M.

Overall, the post-hoc analyses indicated that a directive explanation had an effect only on the procedural explanation. In other words, only the comparison between Group 6 (procedural explanation *with a directive explanation*) and Group 2 (procedural explanation *without a directive explanation*) resulted in significant results, as illustrated in Figure 3.2.

Figure 3.2 Post-hoc analyses of directive explanation differences within equal explanation conditions



Means of trusting beliefs

Means of trusting intentions

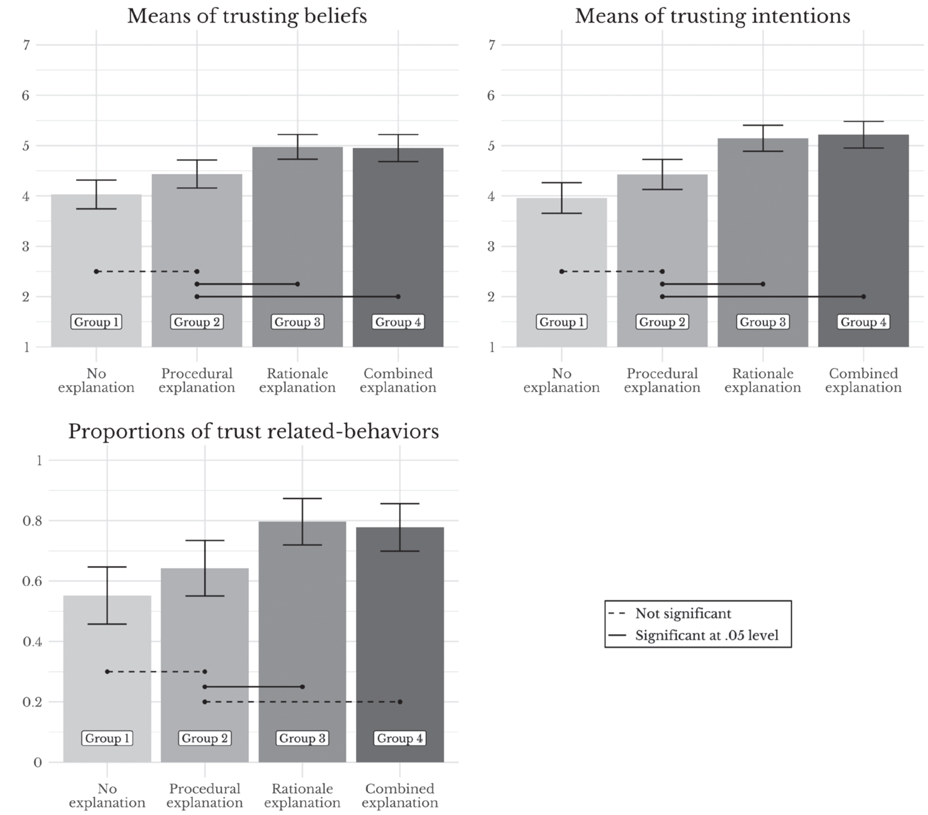
Proportions of trust-related behaviors

Note. Error bars are confidence intervals with Bonferroni-adjusted standard errors.

Because a directive explanation has an interaction effect on the procedural explanation *only*, it is interesting to see if the effect of a directive explanation has mitigated possible differences between the procedural explanation and other types of explanations without a directive explanation. After all, this was a potential reason for not finding differences between types of explanations in Study 1. To test this, we compared the group of participants that received a procedural explanation without a directive explanation with all the other groups without a directive explanation, so we compared Groups 2 and 1, 2 and 3, and 2 and 4. The comparisons were nondirectional since we did not have theoretical expectations for these comparisons. To adjust for the inflated chance of Type 1 errors, we used a Bonferroni adjustment of three times the *p*-value, with an alpha of .05. The full results of these comparisons can be found in Appendix 3N.

These post-hoc analyses resulted in three important findings that are visualized in Figure 3.3. First, we found no significant differences in trust between participants that received a procedural explanation without a directive explanation (Group 2) and participants that received no explanation without a directive explanation (Group 1). Second, the results showed that participants receiving a rationale explanation without a directive explanation (Group 3) had significantly higher trust in the algorithmic recommendation than participants with a procedural explanation without a directive explanation (Group 2). Third, participants with a combined explanation without a directive explanation (Group 4) had higher trust in algorithmic recommendations than participants with a procedural explanation without a directive explanation (Group 2).

Figure 3.3 Post-hoc analyses of differences between the procedural explanation condition without a directive explanation and all the other explanation conditions without a directive explanation



Note. Error bars are confidence intervals with Bonferroni-adjusted standard errors.

Conclusion study 2

In Study 2, we focused on a novel element in explaining algorithmic recommendations, namely a directive explanation that lists specific actions a citizen needs to take to achieve a desired outcome. We examined the main effect of a directive explanation on trust in algorithmic recommendations. We did not find support for our fifth hypothesis, which expected that citizen trust in algorithmic recommendations would be higher when a directive explanation was provided compared to when no directive explanation was provided. Overall, there was no main effect of a directive explanation on citizen trust in algorithmic recommendations.

However, a post-hoc analysis suggested that adding the element of a directive explanation to a procedural explanation does, in fact, increase trust in algorithmic recommendations. In addition, we found that a directive explanation mitigated the differences between a procedural explanation and the other three experimental groups that received no directive explanation in two ways. First, the analyses showed that participants that received no explanation with no directive explanation (Group 1) did not have significantly higher or lower trust than participants with a procedural explanation without a directive explanation (Group 2). Statistically, this means that when no directive explanation is provided, a procedural explanation is as effective as no explanation in strengthening citizen trust in algorithmic recommendations.

Second, we found that participants that received a procedural explanation without a directive explanation had less trust in the algorithmic recommendation than participants with a rationale or a combined explanation without a directive explanation. Thus, when no directive explanation is provided, a rationale and a combined explanation are more effective than a procedural explanation in strengthening citizen trust in algorithmic recommendations. Table 3.3 summarizes the results of the main findings of both Study 1 and Study 2.

Table 3.3 Overview of results per hypothesis of Study 1 and Study 2

Hypothesis				**Result**
H1:	Procedural	>	No explanation	Supported
H2:	Rationale	>	No explanation	Supported
H3:	Procedural	>	Rationale	Rejected
H4a:	Combined	<	Procedural	Rejected
H4b:	Combined	<	Rationale	Rejected
H5:	Directive explanation	>	No directive explanation	Rejected

3.5 Discussion

Our research has important implications for academic and societal debates on algorithmic transparency in digital service delivery and raises a new set of questions for further research into algorithmic explanations.

Academic implications

We found that the type of explanation matters, but only to an extent. Indeed, while some argue that a rationale explanation (explaining *why*) leads to the most pertinent increase in citizen trust (e.g., de Fine Licht & de Fine Licht, 2020) and others show that a procedural explanation (explaining *how*) is most effective in strengthening trust (e.g., Kizilcec, 2016), we found that generally similar effects can be achieved using all types of explanations. Our manipulation checks showed that, generally, people were able to distinguish and recognize the different explanation types. It is therefore unlikely that the lack of differences is an artifact of our research design. One way to explain these findings is by looking at social psychology literature on what constitutes a persuasive message. One of the most-used persuasion models is the Elaboration Likelihood Model (ELM) (Petty & Briñol, 2011; Petty & Cacioppo, 1986). The ELM describes two routes to persuasion. On the central route, persuasion will likely occur based on someone's careful and thoughtful consideration of the benefits of the information. On the peripheral route, on the other hand, persuasion is not based on the intrinsic merits of the information but on peripheral cues such as format and visual presentation.

People are likely to take the peripheral route when the stakes are low and when there is little involvement with the given information. Individuals generally have limited time and capacity for information processing and simplify the options and information provided in order to make decisions. Therefore, the choice to trust a government agency may not necessarily be conscious or rational. Based on the ELM, Grimmelikhuijsen and Meijer (2014) argue that if citizens have high prior knowledge about a topic, providing information will not affect a government organization's perceived trustworthiness. Citizens' trust is "cognition based", meaning that their prior knowledge becomes the main driver for perceived trustworthiness along with the information provided to them (Grimmelikhuijsen & Meijer, 2014, p. 153).

In addition, Alon-Barkat (2020) shows that citizens do not rationally decide to trust the government based on the information provided to them. Using the ELM, he found that symbols increase citizens' trust by making citizens pay less attention to logically unpersuasive information, and thus counteracting its negative effect. Even though they were carried out in other contexts, these studies show that the ELM is relevant for explaining unexpected and non-rational trusting attitudes and behaviors of citizens based on the information provided to them.

Despite our expectations that differences would exist between explanations, our studies showed that all types of explanation led to similar levels of trust, which could suggest that participants were taking the peripheral route to process the information they received. The experiments involved a hypothetical fraud case in which shoes were not delivered after being purchased online. This could be an example of a case where the stakes were low for most participants, meaning that results of attitude or behavioral changes between the types of explanations were limited, but between no explanation and any explanation, the results were significant. Future research is necessary to better understand how the ELM model can be used in theorizing the relationship between explanations (transparency) and trust in the context of algorithmic recommendations by public services. The model could be used to examine whether higher stakes cause participants to take the central route, which might result in differences between the types of explanations. Still, we must pay attention to the implications of this interpretation.

From a democratic point of view it might be harmful that citizens do not critically evaluate the explanations that are provided to them, but that they find *any* explanation satisfactory. As explained above, accepting any explanation contrasts with the normative expectation that citizens in a democracy should form their opinions on the government based on rational critical thinking and healthy skepticism (Norris, 2022). Uncritical citizens may be more vulnerable to manipulation and government abuse in public service delivery. Therefore, we argue that having correct and understandable explanations is not enough. There should be checks and balances in place for the use and explanation of algorithmic recommendations for public organizations. The latter concern, if valid, necessitates the development of algorithmic scrutiny or accountability mechanisms (see also Grimmelikhuijsen & Meijer, 2022; Wieringa, 2020).

Societal implications

Based on the finding that explanations are highly important for trustworthy digital public services, we argue for including explanations as a design requirement for algorithmic recommendation systems in public services. This is in line with the Transparency by Design principles of Felzmann et al. (2020), which highlight areas where system designers need to address transparency concerns regarding artificial intelligence during the design process. Moreover, requiring explanations for algorithmic recommendations connects with the work of Rudin (2019), who states that the way forward is to design algorithmic models that are inherently interpretable, rather than creating methods for explaining black box models. Including explanations as a core element of the design process will ensure the development of algorithms that contribute to citizen trust. This requires tool designers to ensure that a recommender system not only provides a recommendation but also (procedural and rationale) explanations for it. This is crucial for a representative government that is, or can be, called to account for its actions. Moreover, it stimulates the designers to think about the pros and cons of the tool, thus leading to higher-quality decisions that are reflected in an AI tool (de Fine Licht & de Fine Licht, 2020). While embracing greater openness in digital public service delivery, it is important to recognize that openness might incur substantial costs by diminishing operational efficiency, posing a delicate balancing act for public managers (Halachmi & Greiling, 2013). However, this paper showed that considering explanations as an integral part of the AI design process in the public domain is crucial for fostering citizen trust in digital encounters and should therefore be prioritized.

Research limitations and future research

This study is subject to some limitations that provide future research directions. An important point of discussion is to what extent our findings are generalizable to other contexts. This question touches upon three elements of our study: the design, the sample and the experimental scenario. First, regarding the design, we used a mock version of a real algorithmic recommender system with, in contrast to Kizilcec's (2016) experiment, understandable text for all explanation conditions. Recall that the actual Intelligent Crime Reporting Tool (ICRT) of the Netherlands Police, on which our experiment is based, arrives at recommendations via an inherently interpretable, rule-based algorithm, which is based on legal and policy rules of the police (Borg & Bex, 2021; Odekerken et al., 2022). The rule-based algorithm makes it possible to provide rationale explanations in terms of these rules and the input provided by the citizen, making these explanations understandable for the average citizen. This might be a reason why our results differ from those of Kizilcec (2016), whose rationale explanation consisted of more complex statistical calculations of students' grades and possible biases. Thus, the challenge of explanation will become more pronounced in systems that use less transparent AI techniques such as deep learning or machine learning (Grimmelikhuijsen & Meijer, 2022). The type of explanations for such systems (if any) are very often based on statistical patterns that are difficult to understand for lay users such as citizens.

Furthermore, as our explanations were designed with the police based on real-world situations, they might have been less distinctive as explanations used in experiments with fictional AI tools. This could have been a reason why not all participants correctly identified the type of explanation they read. A sensitivity analysis, however, showed that excluding participants that failed the factual manipulation check did not alter our results. Alternatively, and in line with our interpretation of the results, participants' incorrect responses could have also been due to the fact that individuals generally have limited time and capacity for information processing and therefore simplify the information provided to them in order to make decisions. Having any type of explanation (in comparison with no explanation) would have been enough to change their attitudes and behaviors. Moreover, it has been proven difficult to make a strict distinction between what is a directive explanation and what is not.

In the second study, we targeted what we believed were the desired outcomes of citizens who reported their case of online fraud, that is, specific actions they needed to take to achieve financial compensation. Nevertheless, there was an element in the rationale explanation that could have been interpreted as a directive, since it mentioned that the quality mark mediated the dispute with the web shop. To the best of our knowledge, this research is the first to empirically examine directive explanations for public sector algorithmic systems. Future researchers may learn from the limitations of our study and further specify and test directive explanations in other settings. In any case, examining these different types of explanations for public sector algorithmic recommender systems is necessary to better understand their effects in different settings.

Second, we used a mock version of the tool in our experiments with a random sample of citizens, rather than those who had actually been victims of internet fraud and used the police's tool. Despite this, we consistently focused on safeguarding high mundane realism by closely mimicking a real chatbot in our experimental scenario. This experimental scenario was developed together with police employees to make the vignettes as realistic as possible. In addition, we asked participants how well they could empathize with the fraud situation they were given. Participants in both experiments scored relatively high on this question: M = 5.46 in Study 1 and M = 5.49 in Study 2 on a 7-point Likert-scale. This suggests that even though hypothetical, our experiment was highly realistic to participants. Evaluations of the actual crime reporting tool performed by the police both support and counter some of our findings: while a large number of citizens trusted the recommendation not to file a report, there is an equally large group that ignored the (explained) recommendation and still filed an official report even when recommended not to. Of course, these people paid actual money for a product they did not receive or did not like, so it can be imagined that they were more emotionally invested in the case than our participants. In any case, in future research we may study real end-users of an algorithmic recommender system in, for instance, a field experimental setting.

Third, we focused on one specific system in one organization, so the question of how well our findings transfer to other contexts remains. However, we expect that other public organizations deal with similar questions. Usually, citizens do not have a realistic exit option for public services, and recommendations may therefore be much more forceful than in private-sector organizations. In addition, public organizations are generally subject to greater demands for transparency and accountability than commercial enterprises, which leads to stronger demands on explainability of algorithms in government (Busuioc, 2021; Meijer & Grimmelikhuijsen, 2020). The findings of this study may therefore be helpful for public organizations that deal with similar explainability questions. The type of service the ICRT provides—allowing citizens to seek justice, i.e., get advice on and report on a possible crime—is perhaps more abstract than in cases where the citizen directly requests something of more tangible value from the government, such as a residence permit or welfare benefits. Additionally, the type of algorithmic public service examined in this paper, assisting citizens in detecting fraudulent activities of online vendors, is a beneficial service, instead of a service that could potentially violate civil rights, such as a risk assessment tool used for bail decisions. It is important to consider the potential implications of the findings when applying them to different situations or contexts (Aoki, 2020). Nevertheless, upholding the law in an effective and transparent way and giving its citizens access to justice are among the core tasks of good government (Holmberg et al., 2009), and the police have every reason not to damage citizens' trust in them. That said, we encourage scholars to empirically examine the effects of providing explanations for algorithmic recommendations in other scenarios where other aspects of government (e.g., access to health care) and other type of services (e.g., requesting welfare benefits or predictive policing) play a role.

Furthermore, and expanding on the points made in the preceding paragraph, there might be a context-specific nature to the effectiveness of different types of explanations in enhancing trust (Aoki, 2020; Grimmelikhuijsen, 2023; Schiff et al., 2022). It is important to acknowledge that trust dynamics can vary significantly across countries and regions. In the Netherlands, known for its high levels of trust in both law enforcement and government overall (European Commission, 2024), this elevated baseline trust might impact the reception of explanations of police algorithms differently compared to countries with lower levels of baseline trust in the police. Nevertheless, in countries with low trust in the police, procedural justice becomes crucial, as it shapes attitudes toward the entire criminal justice system (Nix et al., 2015). While procedural explanations for police algorithms may help in low-trust regions, the evidence about the effectiveness of procedural justice in police work is mixed, with some studies suggesting potential adverse effects (Murphy, 2017). This fits with our findings from Study 2, where procedural explanations had a less pronounced effect compared with rationale explanations. Hence, additional research is necessary to comprehend the applicability of our findings in various contexts, particularly those characterized by low trust in police.

3.6 Conclusion

We investigated the question *What are the effects of explanations on citizen trust in algorithmic recommendations?* by conducting a sequential factorial design with two consecutive survey experiments. On the basis of our findings, we draw two main conclusions. First, explanations—in general—have a significant and positive effect on all dimensions of citizen trust. Explanations increase citizens' trusting beliefs, enhance their intention to act accordingly, and even increase the likelihood that they will change their behavior to follow up on algorithmic recommendations in digital public service delivery. The second main conclusion is that the type of explanation does not seem to matter in terms of the level of citizen trust. Providing information about a decision procedure, the rationale behind a decision, or a combination of these two had no distinguishable effect. This may suggest that trust increases when citizens see that governments make an effort to provide an explanation, regardless of the nature of this explanation.

At the same time, the post-hoc analyses add some nuance to the conclusions. These exploratory analyses showed that the devil might be in the details. If no directive explanation—listing which specific actions an individual needs to take to achieve their desired outcome—is provided, having a rationale element, explaining why a specific action has been recommended, seems to be more effective in strengthening citizen trust. Thus, a rationale explanation or a combined explanation are possibly more powerful than a procedural explanation when no directive explanation is provided.

Finally, the post-hoc analyses provided evidence that a directive explanation has an effect only on a procedural explanation. Adding a directive explanation, that specifies which actions the individual should perform to obtain their desired outcome (if possible), to a procedural explanation leads to an increase in trust. If explaining why a citizen receives a specific recommendation is not possible or feasible—for example, to prevent "gaming the system" (such as tax avoidance) or to protect national security—providing a procedural explanation with a directive element can increase citizen trust.

4

Meso-level: Transparency as a meaningful box-ticking exercise

This chapter is based on the following published article:
Nieuwenhuizen, E. (2025a). Algorithm registers: A box-ticking exercise or meaningful tool for transparency? *Information Polity*, *29*(4), 415-433.
https://doi.org/10.1177/15701255241297107

Abstract

Governments increasingly implement algorithm registers—openly accessible overviews of algorithms used in decision-making—to promote transparency and accountability. However, the contribution of these registers is questioned: some scholars claim that governments only disclose politically non-sensitive algorithms and provide information with limited usefulness for accountability purposes. In contrast, a more optimistic view is also put forward: registers help organizations to be transparent to the public, which increases public trust in algorithm use. In response to this academic debate, this paper aims to provide a theoretical understanding and an empirical mapping of the different factors that shape algorithm registers and their implications. I conducted in-depth interviews with 27 respondents, both developers (public organizations) and key users (oversight authorities and societal watchdogs) of algorithm registers in the Netherlands, and I analyzed 33 policy documents. The findings nuance the current criticisms. While I find that public organizations indeed selectively disclose information and registers are currently not found useful by societal watchdogs and oversight authorities, there are also dynamics that could contribute to more responsible use of algorithms by public organizations, such as a potential *disciplining effect* of registers. This study highlights the importance of going beyond the rhetoric on algorithm registers, by establishing a deep understanding of the *indirect* and *unexpected* dynamics they create. The paper concludes that algorithm registers are currently not a meaningful tool for transparency, but a *meaningful box-ticking exercise* for public organizations.

4.1 Introduction

Scholars and societal watchdogs are concerned about the increasing use of algorithms in government decisions and services. A core criticism is that algorithms lack accountability and transparency (Busuioc et al., 2022). In response, (local) governments all over the world are experimenting with setting up algorithm registers, and the European Union recently implemented the Artificial Intelligence (AI) Act, mandating the registration of high-risk algorithms (Wörsdörfer, 2023). An *algorithm register* is a database or record-keeping system that shows the deployment and use of algorithms within an organization. It typically includes information about the type of algorithm, its purpose, the data it uses and any potential risks associated with its deployment (Haataja et al., 2020). These registers could offer insight into the specific algorithms used for public tasks to intermediaries, including NGOs, journalists, independent experts and regulators (Grimmelikhuijsen & Meijer, 2022), thereby easing the burden on individual citizens.

Despite the hope that algorithm registers will solve the issue of algorithmic opacity, there is a scarcity of empirical research about what algorithm registers are, and what these registers yield. Most of the existing research on algorithm registers has a conceptual approach, rather than presenting actual data (Cath & Jansen, 2022; Floridi, 2020). The few scholars discussing this new phenomenon have contrasting opinions. There are "optimists" who argue that algorithm registers help both government organizations and the public to comprehend the societal impact of algorithms, thereby increasing public trust in algorithm use by governments (Floridi, 2020). "Pessimists", on the other hand, argue that disclosure of algorithms through these registers is limited, because public organizations have substantial discretion regarding the information they communicate about an algorithm (Cath & Jansen, 2022).

Beyond the few conceptual studies, little is known about what algorithm registers are and how they can be designed. This is important to know, especially considering the requirement outlined in the AI Act to register high-risk algorithms (Wörsdörfer, 2023). Therefore, this paper seeks to empirically investigate the current use of algorithm registers by answering the following research question: *What is the nature of algorithm registers, and what are their implications?* In doing so, we can complement the conceptual studies, and generate new concepts and theories about the promises and limitations of algorithm registers for governments and relevant stakeholders.

To investigate the nature and implications of algorithm registers, I applied two different methods. First, I conducted in-depth, semi-structured interviews with

27 respondents: developers (civil servants responsible for algorithm registers) and users (oversight authorities and societal watchdogs) of algorithm registers. Second, I conducted a comprehensive analysis of 33 policy documents to explore considerations underlying the design choices for algorithm registers and their implications. This mixed-methods approach allows for an in-depth investigation of the nature of algorithm registers and their perceived implications.

This research was conducted in the Netherlands. The Dutch context is interesting because government organizations have embraced digitalization and they are pioneers in openly disclosing the utilization of algorithms in the public sector (Floridi, 2020). The Netherlands and Finland were the first countries to publish an algorithm register for public organizations to address transparency concerns and promote the responsible use of algorithms (Floridi, 2020). Several other countries increasingly recognize the importance of such registers. For instance, a collaborative effort among nine major European cities, including Barcelona, Brussels, Sofia and Mannheim, has resulted in the development of an Algorithmic Transparency Standard—a shared data schema for algorithm registers that is validated, open-source, publicly available, and ready for implementation.[13] Finally, algorithm registers are not limited to Europe. Cities like New York and Toronto have expressed their intention to establish such registers in the near future[14] and the city of San Jose in California (United States) has already published several algorithms in their municipal algorithm register.[15]

This study makes three significant contributions to the literature on and the field of AI governance and transparency. This paper sheds light on how public organizations sometimes partially, or even selectively, disclose information about algorithms. This practice highlights how formal transparency requirements may seem adequate in theory but often fail to offer real insights into the practical realities of algorithmic decision-making, showing a form of *decoupling* in the context of algorithmic accountability. At the same time, the study reveals a *disciplining effect* within public organizations, through comprehensive engagement in algorithm registration. This process compels critical reflection on algorithm use, illustrating the *relational* aspect of transparency, emphasizing mutual understanding and learning among stakeholders involved in the deployment and use of algorithmic decision-making processes. Finally, it offers a conceptual

[13] Algorithm register (2022). *Algorithmic Transparency Standard*. Retrieved December 19, 2023, from https://www.algorithmregister.org/.

[14] Cities Today (2022, December 14). *Cities confront the challenges of algorithm transparency*. https://cities-today.com/cities-confront-the-challenges-of-algorithm-transparency/.

[15] City of San Jose (n.d.) *AI Reviews and Inventory*. https://www.sanjoseca.gov/your-government/departments-offices/information-technology/digital-privacy/ai-reviews-algorithm-register.

contribution by providing a nuanced mapping of the different factors that play a role in shaping algorithm registers and their broader societal and internal implications. The detailed insights into algorithm registers provide a deeper understanding how they potentially shape transparency, accountability and trust in government organizations. As such, the research serves as a valuable resource for practitioners and researchers, facilitating the establishment of effective algorithm registers globally.

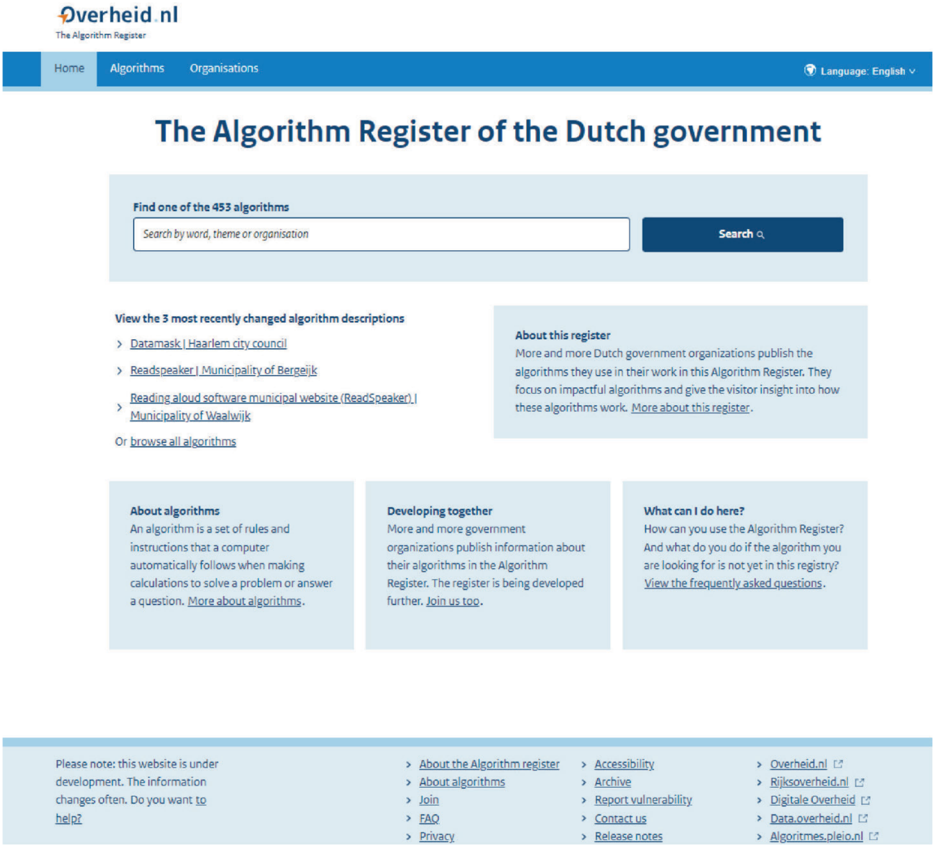
4.2 Research context: The Netherlands as forerunner

This exploratory qualitative study was conducted in the Netherlands, a leading country in the adoption of algorithm registers. To understand the context in which this study took place, two important elements should be described. First, the Netherlands is an early adopter of new technologies, and there is widespread use of algorithms in digital public services (Floridi, 2020). The Netherlands ranks among the top 5 countries in the DiGix index, a multidimensional index of digitization, showcasing a robust digital infrastructure and progressive approach to technology integration (Cámara, 2022).

Second, there have been a few major scandals related to the use of algorithms by the government in the Netherlands. One of the most well-known is the childcare benefits scandal, also known as the "Toeslagenaffaire", which involved the wrongful accusation of thousands of parents of fraudulently claiming childcare benefits between 2013 and 2019. Many of these parents, often from immigrant backgrounds, were forced to repay large sums of money, leading to severe financial and personal distress (Konaté & Pali, 2023). The scandal was intensified by the use of algorithms within the Dutch Tax Authority, which contributed to discriminatory practices and aggressive enforcement policies (Alon-Barkat & Busuioc, 2023). This crisis led to widespread public anger, government apologies, and the resignation of the entire Dutch cabinet in January 2021, sparking calls for greater transparency and accountability in the use of governmental algorithms (Bouwmeester, 2023).

One approach to enhancing such transparency is through the implementation of an algorithm register. In 2020, the City of Amsterdam launched one of the first algorithm registers in the world, followed by a national algorithm register in 2022 that has been launched by the Ministry of the Interior and Kingdom Relations in the Netherlands. Public organizations can decide whether to create their own registers or use the national register. Within the Dutch context, the national algorithm register thus coexists alongside decentralized algorithm registers of public organizations. Figure 4.1 displays the homepage of the national algorithm register, offering an illustration of its appearance.

Figure 4.1 Homepage of the national algorithm register of the Ministry of the Interior and Kingdom Relations (the Netherlands)[16]



The Algorithm Register of the Dutch government

Find one of the 453 algorithms

Search by word, theme or organisation

View the 3 most recently changed algorithm descriptions

- Datamask | Haarlem city council
- Readspeaker | Municipality of Bergeijk
- Reading aloud software municipal website (ReadSpeaker) | Municipality of Waalwijk

Or browse all algorithms

About this register
More and more Dutch government organizations publish the algorithms they use in their work in this Algorithm Register. They focus on impactful algorithms and give the visitor insight into how these algorithms work. More about this register.

About algorithms
An algorithm is a set of rules and instructions that a computer automatically follows when making calculations to solve a problem or answer a question. More about algorithms.

Developing together
More and more government organizations publish information about their algorithms in the Algorithm Register. The register is being developed further. Join us too.

What can I do here?
How can you use the Algorithm Register? And what do you do if the algorithm you are looking for is not yet in this registry? View the frequently asked questions.

Please note: this website is under development. The information changes often. Do you want to help?

- About the Algorithm register
- About algorithms
- Join
- FAQ
- Privacy

- Accessibility
- Archive
- Report vulnerability
- Contact us
- Release notes

- Overheid.nl
- Rijksoverheid.nl
- Digitale Overheid
- Data.overheid.nl
- Algoritmes.pleio.nl

Notably, at the time of the research, there were no regulations stipulating the design of or content requirements for these registers. This lack of standardization led to a broad diversity in algorithm registers, rendering the Netherlands a very suitable context for studying the various approaches and challenges associated with algorithm registers. Although the Dutch context is unique in terms of the widespread use of algorithms by the government and the significant call for transparency following several scandals, the global attention to algorithm transparency is also rapidly increasing (Watson & Nations, 2019). There are various forms, ways and

[16] Overheid (n.d.). *Algorithm Register of the Dutch government.* Retrieved July 15, 2024, from https://algoritmes.overheid.nl/en.

shapes to establish transparency about algorithms beyond an algorithm register, and this study provides interesting results in those respects as well.

4.3 Algorithm registers: Nature and implications

Scholarly attention toward the transparency and accountability of algorithms has intensified over the past few years. Several recent studies have shown that providing transparency about algorithmic decisions or recommendations can increase public trust in these algorithms (e.g., Grimmelikhuijsen, 2023; Nieuwenhuizen et al., 2021; Schiff et al., 2022). Much of this current research focuses on algorithmic transparency towards individual citizens. While valuable, this individual approach to algorithmic transparency requires significant effort from individuals to interpret and critically assess algorithmic decision-making. In addition, this approach neglects the broader institutional context in which they operate (Grimmelikhuijsen & Meijer, 2022; Nieuwenhuizen, 2025b).

A broader organizational or institutional approach is required to increase accountability (De Bruijn et al., 2022; Grimmelikhuijsen & Meijer, 2022). In such an approach, not only information about specific algorithms should be presented, but also about their usage, governance and evaluation, and not just to citizens but also to intermediaries, such as NGOs, journalists, independent experts and regulators (Grimmelikhuijsen & Meijer, 2022). This could lead to more public trust in how governments embed algorithms in their organization (Nieuwenhuizen, 2025b). Algorithm registers have been proposed to realize such an approach, yet there is still limited knowledge about these registers.

The first section focuses on the definitions of algorithm registers, their intended goals and underlying design considerations. The second section discusses the potential implications of algorithm registers. These sections rely on conceptual studies on algorithm registers and a few empirical studies. This literature review establishes the foundation for developing sensitizing concepts that will guide this qualitative study.

What are definitions, goals and design considerations regarding algorithm registers?

Definitions

Algorithm registers have emerged as a critical instrument in the pursuit of effective and responsible integration of artificial intelligence (AI) within public organizations (Floridi, 2020). But what are these registers? Murad (2021, p. 16) is one of the first to define an algorithm register as "a log of algorithmic decision-making systems used by a public authority that have some level of direct impact on its citizens". This includes not only technical specifications but also governance structures that regulate their application. This definition emphasizes the dual role of algorithm registers in shedding light on both the algorithms themselves and the surrounding decision-making processes. Cath and Jansen (2022) further underscore the significance of algorithm registers in providing transparency about decision-making mechanisms that rely on algorithms. They stress that algorithm registers offer insights into the rationale behind algorithmic decisions as well as potential biases inherent in such technologies. Algorithm registers, at their core, thus serve as comprehensive databases that catalog detailed information about the algorithms employed by public organizations. Cath and Jansen (2022) also include the broader context of how these algorithms are utilized in decision-making processes within the public sector. I start with these broad notions of algorithm registers. Then, I aim to get a better understanding of how algorithm registers are defined in practice and to what extent this aligns with the theoretical definition.

Goals

Although the scientific studies on algorithm registers are of a conceptual nature, several students have empirically investigated the design, objectives or effects of these registers (Bottenbley, 2023; de Meijer, 2023; Murad, 2021). They illustrate that algorithm registers serve multiple purposes, primarily focusing on promoting transparency, offering avenues for sanctions and redress, and fostering knowledge sharing and learning in the public sector.

The primary objective of algorithm registers is to promote transparency about the utilization of algorithms (Murad, 2021). This entails making the inner workings of algorithms and the decision-making processes behind them transparent to stakeholders. This allows stakeholders to gain insights into the rationale and potential biases embedded within algorithmic decision-making (Kaminski, 2020). Transparency can be categorized into two levels: first- and second-order. First-order transparency primarily focuses on unveiling the technical aspects of an algorithmic system, encompassing its design, implementation and functioning. This allows stakeholders to understand the logic of the system, ideally exposing any biases and negative impacts, if any (Kaminski, 2020). Second-order transparency goes beyond the technicalities of algorithms and delves into the governance structures and decision-making processes surrounding algorithms. The purpose is to ensure that those responsible for governing algorithms are transparent and accountable for their actions. Second-order transparency acts as a safeguard, preventing power abuse and promoting responsible AI governance (Kaminski, 2020).

Second, building upon the first goal, algorithm registers could play a significant role in providing opportunities for sanctions and redress in the case of misuse (Busuioc, 2021). By maintaining a comprehensive record of algorithm deployments and their associated governance structures, algorithm registers can enable the identification of instances where algorithms are utilized unethically or harmfully (Cath & Jansen, 2022). This not only serves as a deterrent against misconduct, but also offers a mechanism for citizens and stakeholders to seek remedies and hold public organizations accountable for any wrongdoings (Busuioc, 2021).

Third, algorithm registers can foster knowledge exchange and learning among public organizations. These registers could serve as comprehensive repositories of information concerning algorithm implementation, facilitating public organizations at all levels to gain insights and learn from each other's experiences (Murad, 2021). This collaborative approach aims to prevent redundancy, promote efficiency and encourage widespread adoption of best practices in the domain of AI governance (Floridi, 2020). I will use these three initial objectives from the literature to explore existing goals and determine if any additional goals arise.

Design considerations

When creating algorithm registers, there are several aspects to think about regarding how they should look and work. This section describes the key considerations drawn from existing research. The first decision is to determine the *unit of disclosure*. Organizations must choose what they define as an algorithm, whether to register all algorithms in use, or focus solely on those deemed high-risk (Murad, 2021). Cath and Jansen (2022) argue that the unit of disclosure is inherently a political choice, impacting the register's comprehensiveness and reflecting the organization's values and priorities in fostering transparency and responsible algorithm use.

A second important consideration in the design of an algorithm register is determining its *intended audience*. These registers can cater to distinct user groups, including citizens, public-interest representatives, or government and policy officials (Floridi, 2020; Murad, 2021). Each group has different needs and levels of technological sophistication: citizens want to know how algorithms have impacted specific government decisions. Public interest representatives, including researchers, investigative journalists and advocacy groups, seek information about algorithms to safeguard public or stakeholder interests (Murad, 2021). Public servants and policymakers constitute another user group, as registers offer insights into which algorithms in government can help improve policy formulation and decision-making (Floridi, 2020; Murad, 2021).

A third decision concerns the type of information, the *relevant disclosures*, to be made transparent in the algorithm register. These disclosures, however, are not one-size-fits-all but rather intricately linked to the intended audience and objectives of the register. To avoid a "false sense of transparency", it is vital to recognize that sharing complex technical details, such as source code and raw data, with the public is often impractical (Murad, 2021). By contrast, governments can decide only to disclose information relevant to users, such as how data input leads to a specific recommendation or decision (Cheng et al., 2019).

Finally, organizations should decide on the *disclosure modality*, which relates to the presentation of information and compatibility of data (Murad, 2021). Designing the interface requires practical consideration of accessibility factors, such as language selection and font size, and easy navigation and location of relevant information for different target audiences. According to Murad (2021), an interactive interface with layered information is the best approach to address the diverse needs of stakeholder groups. This involves presenting basic information that is accessible to all user categories upfront. More technical information can then be presented in specific fields to address the information requirements of other intended user groups. These four design steps will serve as a lens that structures the data gathering on design considerations for algorithm registers.

What are potential implications of algorithm registers?

Algorithm registers are generally expected to have several positive impacts. Foremost, algorithmic transparency carries the promise of enhanced accountability and citizen trust (Busuioc, 2021; Floridi, 2020). The idea is that by offering a comprehensive and accessible repository of information on algorithms and their utilization, these registers lay the groundwork for citizens to scrutinize, evaluate, and ultimately trust the ethical and responsible use of algorithms in government operations (Haataja et al., 2020; Seppälä et al., 2021). Secondly, algorithm registers could play an important role in promoting the responsible use of algorithms within organizations, aligning the register closely with ethical principles and regulatory frameworks (Harish et al., 2022). As Murad (2021) suggests, these registers serve as a "cornerstone" to ensure that algorithms are employed in a responsible and ethical manner within the public sector. Overall, the promise of algorithm registers is their ability to enhance transparency and accountability, foster citizen trust, and promote responsible algorithm use within government organizations.

On the other hand, algorithm registers also face fundamental challenges that may prevent such positive outcomes. First, there is a risk that registers create what has been called "ethics theater", where the importance of algorithm registers is exaggerated, potentially leading to the creation of false narratives about the risks and benefits of algorithms (Cath & Jansen, 2022). Public organizations might establish algorithm registers to merely appear ethical and transparent, essentially using them as a facade to project responsible algorithm use while sidestepping deeper ethical and practical issues surrounding algorithms. A second criticism concerns the absence of transparency obligations for algorithm registers (Busuioc et al., 2022). While the recently adopted AI Act includes a registration requirement for high-risk algorithms, it remains unclear how EU member states will implement this transparency obligation. In the absence of (legal) requirements regarding the adoption and filling of algorithm registers, these registers can become a token gesture rather than a robust tool to ensure transparency and accountability. These discussed potential positive and negative implications serve as sensitizing concepts in the empirical research (Bowen, 2020).

In sum, while expectations are high, scholars also criticize algorithm registers as potentially ineffective and even risky. So far, however, there have been hardly any empirical accounts on the nature of algorithm registers and their implications. Therefore, this article presents an in-depth empirical account to map the characteristics of algorithm registers and their perceived implications.

4.4 Methods

The context in which this study has been conducted, is extensively described in the "Research context" section. The next sections make the choices related to the data collection and analysis explicit, following the recommendations of Ashworth et al. (2019) to be transparent in qualitative research reporting.

Data collection

This study involves in-depth semi-structured interviews with 27 respondents (developers and users of algorithm registers) and an analysis of 33 relevant policy documents. The data collection took place between May 2023 and January 2024. The next section begins with a discussion of the semi-structured interviews, followed by a review of the document analysis.

This research seeks to explore the extent to which the discussed conceptual frameworks regarding algorithm registers are reflected in practice. Therefore, the literature section serves as a starting point for the topics covered in the interviews. To accommodate potential new findings related to definitions, goals and design considerations of algorithm registers, semi-structured interviews were considered most suitable. This approach allowed for deviations from the initial questions if respondents introduced new insights (Bryman, 2016). Further details on the emergence of new topics are provided in the "Data analysis" section.

The guiding questions for the interviews can be found in the supplementary materials (Appendices 4A and 4B). Each interview was conducted by the author and lasted for an average of one hour. Eleven respondents were interviewed in their workplace, while 16 respondents were interviewed online using MS Teams, according to their preferences.

Recordings were made of each interview with the permission of the interviewees. The recordings were transcribed and then anonymized. The transcription process was conducted in two steps. Initially, the transcription software Amberscript was employed to automatically generate a *verbatim* transcript. This software, provided by the author's institution, adheres fully to GDPR regulations, ensuring the highest level of data protection. Subsequently, either the author or a research assistant listened to the interview and made necessary corrections to the generated transcript.

To gain insights into algorithm registers from different perspectives, a purposive sampling strategy for selecting respondents was used in this study (Robinson, 2014). The author identified two categories of respondents that were crucial for understanding the current practices around algorithm registers. First, the developers of these registers: public organizations. Second, important users of algorithm registers: oversight authorities and societal watchdogs. Table 4.1 presents an overview of the respondents. The author gained access to the field through two methods: by requesting colleague-researchers to facilitate contact with individuals from the selected organizations, and by searching for suitable respondents from the selected organizations on LinkedIn and directly messaging them to invite them to participate in the research.

The author chose the public organizations by using two criteria. First, the organization needed to be at the forefront of the registration process, with multiple algorithm registrations in their register. This criterion was essential to enable the author to inquire about various stages of the design process and to explore its implications. Second, the author aimed for diversity within the category of frontrunners. This meant differences in 1) the design of the register, and 2) the type of organization responsible for the register (ministry, administrative agency and municipality). This diversity in the type of organization and algorithm registers allows for a comprehensive examination of best practices, challenges and variations in the implementation of algorithm registers across different contexts.

The following five registers met these criteria: the algorithm register of the Ministry of the Interior and Kingdom Relations, the Netherlands Employees Insurance Agency ("*UWV*"), the Dutch Social Insurance Bank ("*SVB*"), the Municipality of Amsterdam and the Municipality of Utrecht. Although all are considered frontrunners, these organizations vary in their types and approaches to algorithm registration. For instance, the Municipality of Amsterdam extensively registered a few algorithms through an interactive webpage. In contrast, the Municipality of Utrecht stated that they registered all deployed algorithms, but provided only basic information about all these algorithms in an Excel sheet. The interviewed respondents held partial or full responsibility for the algorithm register in their respective organizations, with job titles ranging from public managers to policy advisors. Throughout this study, this group will be called public managers.

The oversight authorities were chosen based on their formal and informal roles in overseeing public organizations and their use of algorithms. Each authority approaches its mandate uniquely; for example, the Court of Audit scrutinizes the legality and efficiency of Dutch government expenditures, whereas the Data Protection Authority supervises the application of the General Data Protection Regulation. Despite these diverse focuses, all seven selected authorities share a common task: they monitor aspects of the work of public organizations that are progressively becoming more intertwined with algorithms.

The term "societal watchdogs" in this study refers to organizations, groups, or individuals within society that monitor or oversee certain aspects of social, political, or ethical behavior. This includes entities such as investigative journalism and civil society organizations (Karadimitriou et al., 2022; Trägårdh et al., 2013). Societal watchdogs play a role in keeping a check on government organizations to ensure transparency, accountability and adherence to societal norms or values. Four NGOs and two investigative journalists were identified as relevant societal watchdogs for inclusion in this study due to the policy documents, news articles and statements they had written about transparency about public algorithms in the Netherlands. During the interviews, another journalist, an advocacy organization for Dutch municipalities and a critical citizen were proposed by respondents as important experts regarding the topic of algorithm registers. This resulted in interviews with 10 respondents from nine societal watchdogs.

The 33 documents used in this study were policy documents addressing the design, use or implications of algorithm registers. Initially, relevant public policy documents from all organizations listed in Table 4.1 were gathered, totaling 24 documents. Subsequently, all interviewees were consulted to identify any potentially overlooked relevant documents, resulting in the inclusion of an additional nine documents. In the results section, passages from documents are linked to "(DX)", where X represents the document number as outlined in the supplementary materials (Appendix 4C).

Table 4.1 Overview of respondents

Category	**Type of organization**	**Number of respondents**
Public organization (algorithm register)	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*	2
	Employees Insurance Agency *(Uitvoeringsinstituut Werknemersverzekeringen)*	1
	Social Insurance Bank *(Sociale Verzekeringsbank)*	1
	City of Amsterdam *(Gemeente Amsterdam)*	1
	City of Utrecht *(Gemeente Utrecht)*	1
Oversight authority	Netherlands Court of Audit *(Algemene Rekenkamer)*	1
	(Municipal) Court of Audit Rotterdam *(Rekenkamer Rotterdam)*	2
	Data Protection Authority *(Autoriteit Persoonsgegevens)*	1
	National Ombudsman *(Nationale Ombudsman)*	2
	(Municipal) Ombudsman Metropolis Amsterdam *(Ombudsman Metropool Amsterdam)*	3
	National Inspectorate of Justice and Security *(Inspectie Justitie en Veiligheid)*	1
	Netherlands Institute for Human Rights *(College voor de Rechten van de Mens)*	1
Societal watchdogs	Association of Dutch Municipalities (*Vereniging van Nederlandse Gemeenten)*	2
	Expert citizen	1
	Investigative journalism	3
	Non-governmental organization (NGO)	4
Total		**27**

Data analysis

The interviews and documents were coded and analyzed based on the two themes in the theoretical framework (Boeije & Bleijenbergh, 2019). First, the algorithm registers itself: how do stakeholders perceive these registers, what are their goals, and what are the design considerations? Second, the implications: what are positive and negative perceived implications of algorithm registers?

The coding process started by coding interviews. Five interviews (24%) were coded by the author and two other researchers to ensure inter-coder reliability in the use and expansion of the code tree (Bryman, 2016). A detailed description of the coding process can be found in the supplementary materials (Appendix 4D). In assessing inter-coder reliability, the Kappa scores were computed for three coder pairs. The agreement between coders A and B yielded a Kappa score of 0.85. For coders A and C, the Kappa score was 0.77. Coders B and C demonstrated an agreement with a Kappa score of 0.88. The overall inter-coder reliability, represented by the average Kappa score, was computed as 0.83, with the calculation taking into account the weighted average based on the number of codes in each comparison. The author coded the remaining 76% of the interviews and all documents.

The qualitative data analysis software NVivo 14 was used for coding. The coding tree can be found in the supplementary materials (Appendices 4E and 4F). Only themes that were mentioned by six or more different sources (interviews or documents) were included in the coding scheme, to limit the analysis to the important findings and ensure a focused exploration of the key outcomes. After coding the data on algorithm registers and their implications, the author identified several patterns in the data. Subsequently, the author collaborated with three senior researchers, experts in the field of responsible and trustworthy use of algorithms in government, to review and make necessary modifications to the identified patterns based on the coding scheme.

4.5 Results

This results section addresses the two leading research questions: What empirical insights can we gain about 1) the nature of algorithm registers, and 2) their implications?

Nature of algorithm registers

The insights gathered from interviews and documents regarding algorithm registers will be presented in a structured manner through three steps. Firstly, I discuss how public organizations, oversight authorities and societal watchdogs perceive algorithm registers. Secondly, the focus shifts to identifying the goals that algorithm registers intend to achieve. The third step involves a thorough examination of four important design considerations related to algorithm registers.

Perspectives

The analyses revealed three perspectives regarding what an algorithm register represents: a starting point, an endpoint, or a governance mechanism. In a public organization, the chosen perspective significantly dictates the objectives and the design of the register. Simultaneously, the perspective on algorithm registers plays a key role in defining the expectations of oversight authorities and societal watchdogs, outlining what they anticipate a register to encompass.

The *starting point perspective* views the algorithm register as a basic reference or initial stage in the process. It primarily serves as a repository, listing algorithms with minimal details about them. The emphasis is on providing stakeholders with an overview of algorithms in use, without delving deeply into their specifics. The objectives and design of the register are influenced by the need for simplicity and accessibility, focusing on a straightforward catalog of deployed algorithms. Most oversight authorities and a few public organizations adhered to this perspective. A public manager illustrates this view on the register:

> It is the table of contents of your book and the back, and people who want to read it, buy it and explore further.

Conversely, the *end point perspective* positions the algorithm register as the last step in the algorithm lifecycle. Here, transparency and accountability take center stage. The register becomes a means for the public and relevant parties to access detailed information about algorithms, emphasizing the inner workings and characteristics. Respondents, mostly public organizations themselves, express the wish for publishing algorithms in a register to become a standard and obligatory part of a public organization's service or product process when algorithms are used. This would require providing updates to the algorithm register when changes occur in these processes. This perspective underscores the importance of openness in algorithmic processes, serving as a tool for scrutiny and oversight.

Finally, the *governance mechanism perspective* sees the algorithm register as a tool for managing and regulating algorithm usage within a public organization. Both societal watchdogs as well as several public organizations perceived algorithm registers in this manner. In addition to extensive information about algorithms, this perspective incorporates processes and policies for updating, reviewing and assessing algorithms. The register becomes an integral part of an organization's broader governance framework, aiming to ensure compliance, monitor performance and mitigate risks associated with algorithmic processes. Some public organizations mention taking various measures to handle algorithms more responsibly and consciously, such as designing an algorithm policy for the organization or establishing an internal ethical committee that evaluates algorithms, as this quote from an organizational policy document of a public organization illustrates:

> The SVB [Social Insurance Bank] establishes a quality management system for the deployment of the algorithm, encompassing its development, use, and impact. This involves considering the context of the process in which the algorithm will be embedded and determining the measures necessary to ensure a proper alignment. For example, it is important to assess whether the knowledge level of the employees who will be working with the algorithm aligns well with how the algorithm is intended to be used. Effective quality management for algorithm deployment requires attention to both the technical aspects of implementing this technology and handling data, as well as consideration for the ethical, organizational, and human aspects of algorithm use. (D30)

Goals

Following from the different perspectives on what algorithm registers are, there are also different views on the goal of these registers. Firstly, many public organizations believe that using algorithm registers can make how their internal algorithms work, more transparent, ultimately aiming to build *trust in the government*. A policy paper from the Ministry of the Interior and Kingdom Relations outlines a goal tree for their algorithm register, identifying seven key objectives. The top priority is enhancing trust in the government:

> The government is there, among other things, to provide social added value. The government can only do this effectively if there is trust in the government and people feel heard and involved. The algorithm register should contribute to improving trust in the government. (D18)

Secondly, *transparency* serves not only to enhance trust but also as an important objective in itself. Respondents indicate several types of transparency an algorithm register should aim for. This distinction closely aligns with the previously discussed difference between first-order and second-order transparency. On the one hand, respondents that favor first-order transparency primarily focus on disclosing the technical aspects of an algorithmic system itself, encompassing its design, implementation and functioning. According to these respondents, this type of transparency aims to understand the rules and logic of the algorithm. This level of transparency is instrumental in comprehending which algorithms are implemented and how civil servants utilize them in their work. On the other hand, respondents that aim for second-order transparency argue that algorithm registers should disclose the governance structures and decision-making processes surrounding algorithms. The following quote from an investigation report by a municipal court of audit illustrates this perspective:

> For a proper safeguarding of the ethics and quality of algorithms, insight into the usage and nature of these algorithms is necessary, especially since existing GDPR legislation is insufficient. The initial version of an algorithm register does not yet provide this insight adequately. As a result, there is no understanding of where ethical risks may arise and what corresponding control measures need to be implemented. The nature of the algorithms determines the severity of the control measures. (D28)

Respondents that aim for this type of transparency with an algorithm register explain that its purpose is to ensure that those responsible for governing algorithms are transparent and accountable for their actions.

Finally, respondents have identified additional goals (in policy documents), reflecting *downstream effects of transparency*. These include enhancing government accountability, promoting responsible algorithm usage, creating a dialogue with society about algorithms, fostering knowledge exchange among public organizations, and empowering citizens by strengthening their position relative to the government.

Design considerations

Next to more strategically laden considerations about the perspectives on and goals of registers, I found important differences in the operationalization of the design of algorithm registers. Four important design considerations guide public organizations in the process of instituting an algorithm register.

First, a public organization must decide on what they want to be transparent about, known as the *unit of disclosure*. This decision is influenced by several choices. Public organizations must define what an algorithm is to determine which data applications are included in the register. The interpretation of an algorithm varies among organizations, with some adopting a narrow definition, while others opt for a broader scope, as this public manager illustrates:

> We do not want only self-learning algorithms or only high-impact algorithms, but rather a very broad spectrum. Well, even an Excel file with just one formula behind it, so to speak, is already a set of rules, and we consider it as an algorithm.

The debate further centers around the choice between including all algorithms a public organization deploys versus focusing solely on high-risk or high-impact algorithms in the register. Although many public organizations express a preference in policy documents for only registering algorithms with a significant impact on citizens, the lack of clear definitions of what these high-risk or high-impact algorithms are, often leads oversight authorities and societal watchdogs to advocate for the inclusion of all algorithms in the register. In situations where public organizations include all algorithms in the register, it is often observed by stakeholders that there is limited information provided about these algorithms. The interviews and documents show a trade-off between the level of completeness (from zero to all algorithms) and the depth of the information (from no to all information) disclosed. Lastly, there is a consideration regarding the relation between an external register for the public and an internal register for the organization. While understanding the necessity for non-disclosure in the realm of justice and security, oversight authorities and societal watchdogs argue that for most public organizations, there are no justified reasons for limited transparency, as explained by this NGO researcher:

> What is the reason to keep something non-public and not disclose anything about it at all? I generally find that quite objectionable. Of course, there may be a reason why, for example, the police's wiretapping room works with certain algorithms that are not disclosed. […] However, when it comes to the municipality, I don't quite see why the municipality of Amsterdam would have algorithms that need to be kept secret. […] So, if you say, yes, this can be very painful if it comes out, well, that might be precisely the reason it should be made public.

A second consideration is the *audience* for the register. Respondents identify various intended users, with "the citizen" consistently ranked as the primary audience. A policy officer from an NGO briefly summarizes this argument:

> For people who don't know much about algorithms but are interested in how decision-making affects them. So that they can see at a glance: oh yes, here an algorithm was used, I was selected because I meet XYZ criteria, and that algorithm is trained in this way, these were the risks, and they are mitigated in this way; so the most basic information.

In addition to citizens, journalists, NGOs, businesses, researchers, legislators and oversight authorities are also mentioned as intended users. Some organizations distinguish in policy documents between users without technical backgrounds and those with varying degrees of technical knowledge. It is important to note that the intended users do not completely align with the signals organizations receive about the usage of their registers. When public organizations with an algorithm register were asked by the researcher about who provides feedback or inquiries about the register, they typically mention journalists, NGOs and researchers. Employees of public organizations indicate that they rarely, if ever, receive questions from citizens, although they cannot precisely trace the actual visitors to the registers.

Thirdly, organizations decide what type of information is made transparent, also referred to as the *relevant disclosures*. Due to different perspectives on what a register is, what goals it should achieve and who the intended audience is, there is a discrepancy between what public organizations offer and what significant users, such as oversight authorities and societal watchdogs, expect. This mismatch between demand and supply is reflected in what public organizations publish—the current relevant disclosures—which, exceptions aside, mainly consist of information about the technical functioning of an algorithm. However, some of the users interviewed in this study indicated that they wish to gain insight into identified risks and mitigation strategies applied to specific algorithms and, if possible, the training and input data and the source code, so that they can replicate the algorithm. Additionally, users indicate that a relevant disclosure is information on how technology functions in its context. This oversight authority policy officer explains why this is important:

> It is mainly about: what are you going to do with it afterwards? And those human components, I think, played a significant role in the Childcare Benefits Scandal as well, because what do civil servants do with the information they receive? And what you do not want is for a civil servant to lean back and accept everything the system produces as the truth. You also do not want the system to constantly crash because various peculiarities and exceptions are made. So, it is almost impossible to separate the use of an algorithm from how a bureaucracy deals with it. They are so closely intertwined, and in that sense, the algorithm register does not actually say much now. It is much more interesting to know: what happens in-house, behind the scenes?

A final design consideration relates to how the information is presented, i.e., the *disclosure modality*. Here, organizations make three choices: regarding the format, language and level of differentiation of information. In the Netherlands, there is a national algorithm register where all government organizations are allowed to list their algorithms, but organizations are also free to create their own register. This leads to a significant degree of differentiation in terms of format, language and information provided. For all three choices, the intended end user is a crucial factor. If a public organization primarily focuses on citizens, respondents indicate that, concerning the format, more effort is generally put into creating an accessible, interactive website rather than an Excel list. The language is also adjusted accordingly, as explained by this public manager:

> And then you have an extensive discussion with the communication department about what understandable language is. So, we have to present that in B1, Dutch level B1, on the website, so a lot of time was actually spent on that.

On the other hand, if the citizen is not the primary intended user, and the register is mainly intended for journalists and oversight authorities, a format is chosen that allows for easy comparison of data, such as an Excel file or a website with filtering capabilities. In this case, technical language and jargon are employed, providing detailed explanations of the characteristics of the algorithms. Finally, there is also an intermediate option, offering differentiation of information. In this case, a format is chosen where users without technical knowledge, such as many citizens, can see information about algorithms in simple language in an accessible manner. Users who want more detailed information, such as journalists, can click through and open other tabs with additional information, such as the legal basis or the data sources used.

Table 4.2 offers a comprehensive overview, mapping out the factors that shape algorithm registers. While the table suggests that there are different distinct characteristics of algorithm registers, it is important to highlight that many of these characteristics are interconnected. The perspective a public organization holds towards a register has consequences for both the goals they intend to achieve with it and the design considerations involved. For example, an organization viewing the register as a starting point will aim to meet minimal transparency requirements. This means that such a register will provide limited information about algorithms, requiring interested parties to seek additional details if they wish to know more. Nevertheless, all public organizations had one goal in common. They explicitly stated that the overarching goal of the algorithm register is to strengthen citizen trust in algorithm use by public organizations. For external parties, such as journalists or NGOs, transparency, rather than trust, is a goal in itself. To fulfill this goal, different standards for information are necessary, including greater detail, completeness and second-order transparency, where the algorithm's risks and mitigation strategies are explicitly explained.

Table 4.2 Overview of characteristics of algorithm registers

Perspective	Starting point
	End point
	Governance mechanism
Goals	Trust
	Transparency
	Downstream effects of transparency (e.g., accountability, responsible use of algorithms, knowledge exchange, dialogue about algorithms, etc.)
Design considerations	*Unit of disclosure* ▪ Definition of algorithm ▪ Choice which algorithms in register (all vs. high-risk algorithms) ▪ Internal vs. external register
	Audience ▪ Intended users (citizens) ▪ Actual users (societal watchdogs, oversight authorities and other public organizations)
	Relevant disclosures ▪ Technical information ▪ Potential risks and mitigation strategies ▪ How algorithms are used by civil servants
	Disclosure modality ▪ Format ▪ Language ▪ Differentiation in type of information

Implications

The in-depth analysis of interviews and documents revealed that several challenges led to negative implications of algorithm registers. These challenges can be broadly categorized into regulatory and oversight issues, transparency concerns, conceptual issues, resource constraints and user-related challenges.

One of the *regulatory and oversight challenges* is the absence of comprehensive legislation governing algorithm registration. The voluntary nature of this task undermines the impact of algorithm registers, leading to incomplete and inaccurate representations within these registers. Furthermore, the lack of an overseeing organization dedicated to ensuring the accuracy and completeness of registrations fosters perverse effects. Several respondents mention that examining similar registers, like the record of processing activities under the General Data Protection Regulation, reveals that regulatory frameworks alone are insufficient to prompt organizations to diligently register their data processing activities. Active monitoring and oversight are necessary for the register to evolve into a meaningful tool. Furthermore, for some public organizations, this regulatory and oversight vacuum fosters a “checklist mentality”, wherein there is a tendency to view the registration process merely as a checklist to be ticked off, as this NGO researcher explains:

> And now we often see that there is a desire to have it, but the means are not provided, and the commitment to do it well and also maintain it properly is lacking. Because it also takes time to keep it up to date, and then you end up with it becoming a kind of token gesture.

Consequently, these organizations typically register algorithms only once, lacking incentives for subsequent updates or comprehensive field completion. The perception of the register as a mere checklist thus results in sporadic updates and incomplete information. Another example of a perverse effect is that frontrunner organizations, early adopters of algorithm registers, are under intense scrutiny. Societal watchdogs and oversight authorities closely follow and observe their algorithm registrations. This way, intensified scrutiny and media attention of existing algorithm registers act as obstacles for organizations considering starting algorithm registration, leading to a collective action problem.

In addition, *transparency challenges* significantly hinder the potential value that algorithm registers may have. Incomplete registrations and (essential) information being withheld by public organizations lead to “deceptive transparency”. Societal watchdogs and oversight authorities explain that, in many cases, public organizations may present a facade of openness while omitting crucial details or providing vague descriptions about algorithms. An inspector from an oversight authority illustrates this:

> At some point you can imagine that it is, of course, not pleasant if what you are doing as a civil servant seems to be leading to very high risks. So, those descriptions for citizens, are indeed, concealing is not the right word, but yes, they are described very generally.

Furthermore, respondents indicate that there are many inconsistencies in the way public organizations register their algorithms. It is, therefore, difficult to make meaningful comparisons, between organizations or algorithms, about the information that *is* made transparent.

Conceptual challenges further impede the value of algorithm registers. The lack of a clear goal for these registers introduces ambiguity, with varying objectives that are often challenging to integrate within a single register. For instance, one organization outlines eight different goals for its algorithm register, encompassing goals such as enhancing trust in government, increasing explainable decision-making and empowering both citizens and businesses. Additionally, the absence of clear definitions for what an algorithm is, and specifically what "high-risk algorithms" or "high-impact algorithms" are, contributes to confusion regarding what should be registered and leads to inconsistencies in registering within and across public organizations. Finally, some public organizations suggest that the registration process motivates them to adapt rather than deactivate algorithms. The discussions that emerge in the registration process are merely about how to align an organization's algorithms with expected norms, rather than engaging in a critical discussion about their necessity. This public manager illustrates how unusual it is to deactivate algorithms:

> It is, I must honestly say, my experience that no one ever says: "this is not ethical" or "we won't do it".

Resource challenges, such as the time and financial investment required for organizations to register algorithms, are perceived as negative implications of algorithm registers. The registration process, which can take from a few weeks to one or even several years for one algorithm, requires not only considerable time but also substantial financial costs. Many individuals are involved in completing such registrations. In certain cases, third-party reluctance to cooperate in the registration process further extends the time and financial investment required for the registration process.

User challenges, lastly, become visible as algorithm registers in their current form do not appear to yield the intended benefits for its primary intended user, namely, the citizen. Societal watchdogs and oversight authorities explain this using two reasons. Firstly, the language employed exceeds the comprehensibility level of the general population, as very technical terms and jargon are used in describing algorithms, despite efforts to use simple, understandable language. Moreover, suboptimal marketing strategies by certain public organizations contribute to a lack of visibility of algorithm registers, hampering citizens' access to these registers, as described in a municipal policy document:

> Additionally, the algorithm register is not sufficiently accessible for residents and businesses, as it cannot be accessed through the Municipality of Utrecht's website http://www.utrecht.nl. (D31)

This lack of visibility and awareness regarding algorithm registers has also been confirmed by public organizations. These organizations mentioned that they had surveyed citizens or their so-called "customer panel" to assess awareness of their algorithm registers, with the results showing that most people were unfamiliar with them. Additionally, some public managers noted that users could submit questions about algorithms or the register through their website. However, they mentioned that inquiries came exclusively from journalists and researchers, not from ordinary citizens, further highlighting the general lack of public awareness of these registers.

Secondly, there is no connection between individual decisions and algorithm registers, as was mentioned in several interviews and documents. Individual decisions often fail to indicate the involvement of an algorithm in the decision-making process, leaving citizens unaware of its existence. Several oversight authorities and societal watchdogs advocate for public organizations to inform citizens when algorithms are part of the process. In these individual decisions, a reference to the algorithm register could be made for those citizens seeking more information about the algorithm that affected the decision.

In addition to these challenges, the results also highlight several perceived positive implications of algorithm registers. Distinct *external positive implications* emerge when considering other users, such as oversight authorities and societal watchdogs, interacting with the register. First, interested parties, including societal watchdogs and oversight authorities, gain valuable insights into the algorithms deployed by public organizations. This way, the register raises awareness about public algorithms, providing an overview of algorithmic applications within these organizations. Second, respondents acknowledge the positive aspect of the register in conveying ethical considerations related to algorithms. When filled in completely and accurately, algorithm registers could offer stakeholders not only a list of deployed algorithms but also insights into an organization's approach to responsible algorithmic decision-making. Third, according to oversight authorities and societal watchdogs, one of the most significant positive implications is that the register serves as a crucial starting point for further investigations. The information, or its absence, acts as a point of departure for in-depth examinations, particularly when potential risks are identified that may require further mitigation, as this journalist explains:

> And then for journalists and like other societal watchdogs, I mean, it gives us the mechanisms to, you know, more quickly identify whether there is a system that may, you know, need to be investigated or held accountable, because right now we all spend a ton of time just trying to figure out the basics of how one of these systems works.

Furthermore, the implementation of an algorithm register brings about three significant *internal positive implications* for public organizations. First, the process of establishing and filling an algorithm register generates awareness among civil servants about the use of algorithms within their organization. This is particularly important since many civil servants were previously unaware of the presence of (certain) algorithms in their work processes. Second, the registration process serves as a catalyst for organizational learning. As mentioned before, the Netherlands is a pioneering country in terms of algorithm registers, leading to a lot of experimentation and trial and error in the process of embedding algorithm registers in organizations. From creating an inventory of all algorithms in an organization to publishing them in understandable language on a website, organizations are doing this for the first time and cannot build on acknowledged best practices. This means that organizations are undergoing a significant learning process in dealing with this process. For example, public organizations increasingly employ mechanisms like e-learnings to educate employees about the risks and responsible usage of algorithms. Feedback from society further contributes to organizational learning by shedding light on potential risks or weaknesses in their algorithmic processes. Third, the completion of various information fields in an algorithm register introduces a disciplining effect within public organizations. This policy officer of the Ministry explains:

> And now you have to go through it all again and discuss it with each other once more. […] then you actually have the conversation: do we really want to use this at all? Because you have to be able to justify it. What is your legal basis? Why are you doing this? What are the risks? How do you manage them? And if you can't answer those questions, well, then you might need to think twice. In that sense, it also has a purifying effect. By going through this exercise, you may decide not to do something anymore.

The registration process nudges organizations to critically assess their algorithmic decision-making processes, identify associated risks, develop mitigation strategies, and ensure adherence to existing rules and regulations such as the General Data Protection Regulation. Algorithm registers, therefore, not only enhance internal awareness but also foster a culture of continuous learning and improved decision-making within public organizations.

Table 4.3 summarizes the most important implications of algorithm registers. The results indicate that there are still many challenges that limit the extent of the positive implications. This finding aligns with the concerns raised by Cath and Jansen (2022) that algorithm registers have limited value for external stakeholders. It remains questionable whether the predefined goals, such as increasing public trust, will indeed be achieved through these registers. However, there are also surprising positive implications, which mostly benefit public organizations themselves. They experienced increased organizational awareness, opportunities for learning, and a disciplining effect from algorithm registers, aligning with some of the positive expectations of Floridi (2020). Over time, these internal positive implications could potentially contribute to the overarching goal of enhancing public trust.

Table 4.3 Overview of implications of algorithm registers

Negative implications	*Regulatory and oversight challenges* ▪ Voluntary nature leads to incompleteness ▪ No incentive to keep it up-to-date (checklist mentality) ▪ Collective action problem
	Transparency challenges ▪ Deceptive transparency ▪ Inconsistencies in registrations hinder meaningful comparisons between algorithms or organizations
	Conceptual challenges ▪ Lack of clear goal algorithm register ▪ Lack of clear definition of (high-risk) algorithm ▪ Adapting instead of deactivating algorithms
	Resource challenges ▪ Large financial and time investment
	User challenges ▪ Mismatch between algorithm register and intended users due to type of language used and visibility of algorithm register ▪ Lack of connection between individual decisions to citizens and algorithm registers
Positive implications	*External* ▪ Awareness within society about algorithms used by public organizations ▪ Starting point for investigation
	Internal ▪ Awareness within public organization about algorithms ▪ Organizational learning ▪ Disciplining effect

4.6 Conclusion and discussion

The main research question of this study was what the nature of algorithm registers is and what their implications are. Based on data from in-depth interviews with 27 respondents and 33 policy documents, we can provide an answer.

Experiences from the Netherlands, a leading country in adopting public algorithm registers, show that a great diversity in algorithm registers exists. This is primarily attributed to varying perspectives on what an algorithm register is: a starting point, end point or governance mechanism, and to various design choices. This becomes visible in the unit of disclosure (e.g., registering all algorithms versus only high risk algorithms), the audience (e.g., citizens as intended users versus oversight authorities and societal watchdogs as actual users), relevant disclosures (e.g., technical information versus how civil servants use algorithms) and the disclosure modality (e.g., interactive webpage versus excel list) of an algorithm register. Despite these differences, a common goal unites all registers: the cultivation of citizen trust through transparency in the deployment of algorithms by public organizations.

However, the findings regarding the implications of algorithm registers showed that it is unlikely that the goal of enhancing public trust will be achieved. At first glance, the register appears to be mainly a paper reality. It seems to be a checklist, with no incentive to complete all fields or maintain up-to-date information. Societal watchdogs and oversight authorities indicate that it merely offers a glimpse into the inner workings of public organizations. Nevertheless, results of this study also show a hidden value.

The process of registering algorithms fosters organizational awareness and learning, as it brings to light pertinent questions and concerns related to an organization's use of algorithms. This could lead to a disciplining effect, as the registration process nudges organizations to critically assess their algorithmic decision-making processes, identify associated risks, develop mitigation strategies and ensure adherence to existing rules and regulations. Thus, while capturing complex phenomena, such as algorithms, within a tool, like an algorithm register, is challenging, it can hold significance for an organization (Noordegraaf, 2008).

This research offers a nuanced mapping of the different factors that play a role in shaping algorithm registers. While earlier research mainly discusses algorithm registers in the context of their (non)existence (Cath & Jansen, 2022; Floridi, 2020), this study demonstrates that such categorization is not a simple dichotomous choice. The nuanced mapping of algorithm registers' characteristics reveals that there are many difficult choices to make regarding what an algorithm register is, what objectives it tries to reach and how it should be designed. To create a useful and valuable algorithm register, these choices must be aligned. For example, if one sees an algorithm register as a starting point for oversight authorities and societal watchdogs, the primary goal should not be to increase citizen trust directly. Instead, the goal of the register should be to foster accountability, and design considerations should be tailored accordingly. However, because of the developmental (trial and error) phase in which many public organizations find themselves regarding the implementation of algorithm registers, choices are mostly taken out of pragmatic reasons and not so much out of alignment reasons.

The nuanced mapping in this study provides an overview of the different choices, which can help organizations critically consider how they want to align the different characteristics of their algorithm registers.

In sum, considering the primary goal of the register—building trust—initially, it does not lead to an increase in trust by stakeholders using the register. Partial or selective disclosure of information may even negatively impact the trust of stakeholders. However, eventually, the registration process could contribute to the responsible use of algorithms through its learning and disciplining effect, thereby fostering better algorithmic governance and potentially enhancing public trust. To understand the potential of algorithm registers, we need to look for indirect and unexpected dynamics, instead of direct and linear effects outcomes such as greater transparency, accountability and trust. As public organizations are the ones primarily reaping the benefits, algorithm registers are (currently) not a meaningful tool for transparency, but a *meaningful box-ticking exercise*.

Limitations and directions for future research

The field of algorithm registers in the Netherlands is undergoing rapid development, with frequent modifications to registers occurring weekly. This study, conducted between May 2023 and January 2024, reflects this evolution. For instance, the last interview in January 2024 revealed respondents presenting well-founded arguments for certain design considerations. This contrasts with the initial interviews in May 2023, where design choices were largely based on intuition. This shift can be attributed to extensive inter-organizational discussions and learning from best practices within the field. Looking ahead, future research could explore how the line of reasoning, specifically regarding the nature of algorithm registers, changes over the course of the next couple of years, providing valuable insights into the ongoing evolution of algorithm registers.

It is important to note that the nature and implications of algorithm registers are discussed separately but are not independent of each other. This study offers several indications about the interconnectedness of these aspects. For example, an organization's perspective on an algorithm register has implications for its eventual structure and the information it discloses, thereby influencing the outcomes of the register. Future research should examine how the nature and implications of algorithm registers are interconnected. Conducting a quantitative study involving numerous public organizations that register algorithms or employing a process tracing study to assess whether certain design considerations lead to distinct implications would be approaches to enrich the discourse around algorithm registers.

Theoretical contributions

This study contributes to the ongoing debate about the need for a broader organizational and institutional approach to transparency in the use of algorithms by public organizations (Grimmelikhuijsen & Meijer, 2022). The findings regarding the implications of algorithm registers show a nuanced reality: arguments of both transparency optimists and pessimists find support to an extent, aligning with earlier research that emphasizes transparency is not a silver bullet in fostering citizen trust (Grimmelikhuijsen & Meijer, 2014).

On the one hand, public organizations indeed selectively disclose information, which can come across as a false sense of transparency (Cath & Jansen, 2022). This selective approach, also known as "openwashing", creates an impression of openness without presenting a complete or accurate picture of the organization's actions (Heimstädt, 2017). This phenomenon of selective disclosure and openwashing exemplifies *decoupling* in accountability literature (Power, 1997), where formal transparency requirements may appear to fulfill their intended function on paper but fail to provide genuine insight into the operational realities of algorithmic decision-making. Additionally, in line with earlier research, this study has shown that transparency about algorithms in a register could normalize their use, decontextualizing and depoliticizing the discussion by focusing on creating conditions for deployment rather than critically debating whether algorithms should be deployed at all (Cath & Jansen, 2022; Wang, 2022). This could create a perception that algorithms are inherently neutral tools, potentially masking their complex socio-political consequences. Moreover, the assumption that citizens, or even developers and civil servants using the algorithm, fully comprehend the disclosed information may be overly optimistic. Therefore, this study advocates that transparency intermediaries like oversight authorities and societal watchdogs should be the targeted audience, emphasizing the need for a critical audience in achieving algorithmic accountability (Kemper & Kolkman, 2019).

On the other hand, this study identifies a disciplining effect demonstrating that the more civil servants engage comprehensively in registering algorithms, the more thoroughly the use of algorithms is analyzed, leading to a more responsible use of algorithms in public organizations. Regardless of specific outcomes or practical benefits, transparency through algorithm registers is considered valuable because it contributes to the responsible use of algorithms in public organizations, enhancing the legitimacy of technological solutions within a democratic framework (Loi & Spielkamp, 2021). The act of making something transparent forces organizations to discuss and critically evaluate their use of algorithms. This shows the *relational* aspect of transparency that could lead to mutual understanding and learning (Ananny & Crawford, 2018; Valentinov et al., 2019) rather than a descriptive type of transparency that involves one-sided algorithm disclosure (e.g., Kempeneer, 2021).

Moreover, these insights align with previous research on the use of measurement tools and accountability mechanisms (e.g., Kempeneer & Van Dooren, 2021; Meijer, 2007; Noordegraaf, 2008), that demonstrates that the tools themselves are not inherently valuable. Rather, the requirement for organizations to reflect on the practices they need to measure or account for, and to make these practices explicit, leads to improved internal processes. Similarly, an algorithm register provides a framework for reflecting on how and why algorithms are used by the government. This research demonstrates that the value of such a transparency tool lies in the discussion about algorithms in use. Transparency tools in other forms and formats might have the same relational and learning effect.

Practical recommendations

This research is of practical relevance given the growing attention and adoption of algorithm registers worldwide and the recent enactment of the AI Act which mandates registration of high-risk algorithms for EU member states. By investigating algorithm registers' characteristics and implications, this research informs the discourse on AI governance, offering two practical recommendations for future developments regarding algorithm registers.

I recommend public organizations to reassess the target audience for algorithm registers. This suggestion stems from the findings revealing a lack of awareness among citizens regarding these registers. One potential adjustment entails shifting the focus of the register from ensuring transparency for citizens to providing transparency for intermediaries. These intermediaries can leverage the register's information for accountability purposes (Grimmelikhuijsen & Meijer, 2022). However, should the primary objective remain citizen-centric, efforts should be directed towards establishing a clear connection between individual decisions involving algorithms, such as speeding tickets, and the register. This allows citizens to navigate and monitor algorithmic decision-making processes that affect them.

In addition, I advise regulators to prioritize establishing mechanisms for robust oversight, as this study indicates that the effectiveness of regulation regarding algorithm registration (AI Act) relies on this critical foundation (see also Enqvist, 2023; Laux, 2023; Tutt, 2017). The findings showed that experiences with similar registers, such as the register of data processing activities under the GDPR, highlight that legislation alone is insufficient. Active oversight is necessary for comprehensive and up-to-date registrations. Oversight authorities responsible for monitoring these registers must be adequately resourced to effectively investigate algorithm registrations.

5

Macro-level: Monitoring transparency increases citizen trust

This chapter is based on the following manuscript:
Nieuwenhuizen, E., Trehan, V., & Porumbescu, G. Does institutional transparency affect citizen trust in predictive policing? Evidence from a survey-experiment in the Netherlands. *Under review.*

Abstract

Police organizations increasingly use predictive algorithms for surveillance, raising privacy concerns and potentially decreasing citizen trust in these prominent policing tools. This paper investigates whether two dimensions of institutional transparency, transparency about legislation and external monitoring, can influence trust in predictive policing. Using a representative survey-experiment with 877 Dutch citizens, we find that disclosing how police's use of predictive algorithms is monitored by an external oversight agency significantly increased citizen trust, while providing details about predictive policing legislation did not. The findings highlight the importance of external monitoring in safeguarding citizen trust regarding police use of predictive algorithms, offering insights for policymakers and law enforcement. Furthermore, this paper provides a theoretical and methodological framework for conceptualizing and investigating institutional transparency in the context of algorithmic governance.

5.1 Introduction

Big data and sophisticated algorithms are increasingly viewed by governments as essential tools to address long-standing public service challenges (Janssen & Kuk, 2016). One area where this is particularly true is crime prevention. Indeed, police departments around the world are increasingly using algorithms to predict when and where future crimes are likely to occur (Meijer & Wessels, 2019). One such application, "predictive policing", is touted for its potential to increase efficiency in crime prevention and allocate limited law enforcement resources more effectively (e.g., Levine et al., 2017; Mohler et al., 2015). For example, in one of the most compelling illustrations of the power of predictive policing to date, a recently developed algorithm claims to predict crimes up to one week *before* they occur with 90% accuracy (Rotaru et al., 2022).

Nevertheless, these impressive advances in computing power have increased the risk of biased predictions, the misuse of power, and inappropriate handling of sensitive personal data by police departments (e.g., Andrews, 2019; Brayne, 2020; Burgess, 2018). These practices can, in turn, perpetuate existing inequalities (Richardson et al., 2019). Although several scholars have raised these concerns, Meijer and Wessels' (2019) systematic review of the literature on predictive policing suggests that there is limited empirical evidence on this issue. These concerns, whether well-founded or not, are contributing to low public trust in algorithmic policing (Crawford & Schultz, 2014; Hobson et al., 2023).

Algorithmic transparency is seen as one important approach to addressing such concerns. As defined by Grimmelikhuijsen (2023, p. 244), this form of transparency is established when "external actors can access the underlying data and code of an algorithm and the outcomes produced by it are explainable in a way a human being can understand". Algorithmic transparency can not only demystify the predictive policing process but also enable public scrutiny of these algorithms. By allowing external entities to review core components of these algorithms, the potential for uncovering biases or discriminatory patterns increases (Ferguson, 2016). Furthermore, it sets the stage for the public to comprehend the decision-making processes involved in predictive policing. Ultimately, algorithmic transparency holds the promise of elevating public awareness regarding how algorithms shape law enforcement decisions, thereby fostering greater trust (Nieuwenhuizen, 2025b).

To date, research on the role of algorithmic transparency in fostering trust has primarily focused on public access to information about how algorithms work (e.g., Grimmelikhuijsen 2023; Nieuwenhuizen et al., 2024). While algorithmic transparency is important and can increase citizen trust, we argue that it can also impose a substantial cognitive burden, as individuals must interpret and assess complex algorithmic decision-making processes that require domain-specific expertise (Nieuwenhuizen, 2025a). Additionally, algorithmic transparency neglects the broader institutional context in which algorithms operate, such as the legal and regulatory frameworks that shape their use (Grimmelikhuijsen & Meijer, 2022; Nieuwenhuizen, 2025b). Several scholars argue that transparency about this institutional context is just as important as transparency about the functioning of individual algorithms, as these algorithms do not function in isolation but are embedded within regulatory constraints and oversight structures (New & Castro, 2018; Wenzelburger et al., 2024). Indeed, without an understanding of these constraints, it is difficult to fully grasp the societal impact of algorithms or ensure meaningful accountability (Schlehahn et al., 2015; Wirtz et al., 2019).

Given the limitations of algorithmic transparency as a mechanism for fostering trust, we explore the role of institutional transparency in the context of predictive policing. Institutional transparency includes information about the institutional context predictive policing is used in, such as regulations and cultures of algorithm use (Giest & Grimmelikhuijsen, 2020; Wenzelburger et al., 2024). In other words, it communicates institutional guardrails that prescribe how organizations should or should not behave (cf. Moe, 2012). Institutional transparency is established when external actors can readily access and understand the rules, regulations and oversight mechanisms that govern how an algorithm is deployed and used within a public organization. A key objective of this paper is to theoretically explore the implications of institutional transparency for algorithmic trust, focusing on transparency about legislation on monitoring police's use of predictive algorithms. Studying institutional transparency and trust in the context of policing offers valuable insights about a public sector context which involves a direct and often personal impact on citizens' rights, safety and well-being. The stakes are particularly high in this context, as the misuse or perceived misuse of predictive algorithms can lead to significant consequences, including biased law enforcement, erosion of civil liberties and increased societal inequalities (Brayne, 2020). By focusing on policing, this study addresses a gap in understanding how institutional transparency about algorithm use influences citizen trust when power asymmetries and ethical stakes are pronounced. With this in mind, we address the following research question: *To what extent does institutional transparency affect citizen trust in predictive policing?*

We examine our research question using a survey experiment (*N* = 877) in which we test the effect of transparency about legislation that shapes police's use of predictive algorithms (*ex-ante* control) and transparency about efforts by external oversight agencies to monitor police's use of these tools (*ex-post* control) on citizen trust in the Netherlands. The Netherlands is an interesting context to study the relationship between institutional transparency and citizen trust in predictive policing. The country has a strong tradition of government transparency and has adopted a well-established European legal framework surrounding the use of data and technology in public administration, including the General Data Protection Regulation (GDPR) and the recently adopted European AI Act. Additionally, the Netherlands police are increasingly incorporating data-driven approaches to improve public safety and enhance crime-fighting effectiveness, with a particular emphasis on leveraging big data and predictive analytics for risk assessments (den Hengst & Wijsman, 2023; Meijer et al., 2021; Schuilenburg & Soudijn, 2023). The Netherlands also provides a relevant case for studying the use of algorithms by public organizations, given the recent public backlash over high-profile scandals, such as the childcare benefits scandal[17], which have heightened calls for greater transparency and accountability in the use of algorithms within public institutions (Bouwmeester, 2023).

We find that institutional transparency significantly influences citizen trust in police use of predictive algorithms. Specifically, our survey-experiment reveals that disclosure of information about external monitoring of police's algorithm use enhances citizen trust in the use of predictive algorithms by the police, whereas information about legislation, either alone or in combination with monitoring, does not contribute to increased trust.

While our study focuses on predictive policing, our findings offer valuable insights for the broader field of public administration in view of the increasing use of algorithms across various public-sector contexts. Specifically, these findings underscore the importance of institutional transparency for public administration scholars, as they not only inform governance strategies in the era of algorithmic governance, but also respond to pressing concerns about the social legitimacy of algorithmic tools used in public services governance (Grimmelikhuijsen & Meijer, 2022). In doing so, they offer two important contributions to public administration research on algorithm use by public agencies. First, they shed light on the importance of institutional guardrails, such as laws and oversight bodies, in promoting trust in an era of algorithmic governance (Grimmelikhuijsen & Meijer, 2022). Practically, our findings provide actionable insights for policymakers and public managers on the important role of external oversight in fostering citizen trust.

Second, while the access to information about algorithms approach to promoting trust is important, this information, even when simplified, is very difficult to

[17] For more information about the Dutch childcare benefits scandal, please refer to Hadwick and Lan (2021).

understand for the average resident, thus raising concerns over the practicality of implementing such solutions in the real-world (Kemper & Kolkman, 2019). Thus, our overarching contribution here is that we interrogate the utility of another form of transparency that does not require domain-based expertise as a means of promoting trust. In doing so, this paper provides a conceptual and methodological framework for institutional transparency about predictive policing. This framework can serve as a foundation for further research about and the application of institutional transparency in the context of algorithm use by governmental organizations.

5.2 Theoretical background

Trust and transparency in predictive policing: Threats and potential solutions

Predictive policing involves collecting and processing historical crime data using sophisticated algorithms to forecast criminal activity as well as design policing interventions and preventive strategies (Meijer & Wessels, 2019, p. 1033). Trust in predictive policing is crucial as it helps to ensure the public has confidence in the criminal justice system and its ability to fairly and effectively enforce the law. Further, since citizens generally have no realistic "exit" option for public services such as policing (Hirschman, 1970), they depend on the police as their single service provider. That said, low trust can contribute to forms of partial exit, such as citizens choosing not to contact police when they observe suspicious activity or are in need, thereby contributing to less effective policing (Porumbescu et al., 2019). All told, if the police are not trusted by the citizens they serve, they are unable to function properly.

Threats to trust in police introduced by predictive policing include concerns over biased data, a lack of accountability for algorithmic decisions and the potential for algorithmic systems to perpetuate existing systemic inequalities (Hobson et al., 2023). Algorithmic transparency is widely viewed as a potential solution for increasing trust in algorithmic policing as it aims to promote explainability (Kim & Routledge, 2018) and accessibility, by making data, source codes and decision trees accessible for a specific algorithm to external actors (Mittelstadt et al., 2016). Algorithmic transparency can help address key threats to trust in predictive policing, such as concerns over the use of biased data and the lack of accountability for algorithmic decisions, by embedding algorithms in an organization in ways that allow outsiders to monitor how algorithms work and perform (cf. Grimmelikhuijsen and Meijer 2014). This information allows for the examination and assessment of algorithmic systems and the data that drives them. Offering insights into the decision-making processes of algorithmic systems through algorithmic transparency can increase public understanding and confidence in these systems, thereby building or increasing trust in predictive policing.

Transparency about algorithmic performance is important, but several scholars argue and show that transparency about the context in which algorithms operate is equally important (Grimmelikhuijsen & Meijer, 2022; New & Castro, 2018; Wenzelburger et al., 2024). Indeed, the effects of algorithms on society are difficult to fully grasp without a clear understanding of the contextual constraints on algorithmic systems, such as laws and regulations (Schlehahn et al., 2015; Wirtz et al., 2019). This is particularly relevant for policing, where algorithmic systems have been found to have different effects in different contexts. For example, Meijer, Lorenz and Wessels (2021) found that a predictive policing system in the Netherlands was viewed as an "algorithmic colleague" while in Germany it was viewed as an "algorithmic cage". These differences were attributed to administrative cultures and organizational patterns, rather than the technological features of the algorithmic system. What these findings highlight is the need for transparency not only about how a specific algorithm works, but also about the context in which it operates, including its organizational embedding and legal and regulatory frameworks.

Institutional forms of algorithmic transparency

The importance of accounting for institutional context when evaluating administrative tools has recently garnered attention from public administration scholars, who emphasize a multi-level perspective on transparency for algorithmic governance. For example, Giest and Grimmelikhuijsen (2020, p. 410), highlight that transparency challenges do not only concern making algorithms—and the datasets that feed them—accessible and explainable, but also include issues regarding how algorithms interact with existing rules and norms in the justice system.

Building on these perspectives, we focus on institutional transparency in this paper. By institutions, we refer to "systems of rules that structure the course of actions that a set of actors may choose" (Scharpf, 1997, p. 38). Both formal rules, set out in legislation, regulations or guidelines, and informal rules, which emerge spontaneously and unintentionally over time through human interaction, and take the form of unwritten conventions, routines, customs, codes of conduct and behavioral norms, are at the core of institutions (Skoog, 2005). Institutional transparency, then, is viewed as an information flow to establish a hierarchical relationship that promotes accountability to formal and informal rules of conduct (Heald, 2006). In other words, institutional transparency highlights how features of the regulatory context a public organization operates in, such as legal frameworks and regulatory agencies, constrain the behavior of that agency.

With this understanding of institutional transparency, we apply it to the realm of algorithmic governance, emphasizing the importance of providing information about rules, regulations and oversight bodies, tasked with ensuring responsible and accountable use of autonomous intelligence systems by public organizations (Wirtz et al., 2019). These rules can be codified in legislation such as the General Data Protection Regulation or in guidelines such as the European Commission's Ethics Guidelines for Trustworthy AI (Grimmelikhuijsen & Meijer, 2022). In this setting, *institutional transparency* is established when external actors can readily access and understand the rules, regulations and oversight mechanisms that govern how an algorithm is deployed and used within a public organization. This includes clarity about the legal frameworks (e.g., legislation) that prescribe how and when the algorithm may be used, as well as any processes in place for monitoring compliance.

To further refine our discussion, we distinguish between two important mechanisms that ensure institutional transparency in practice: *ex-ante* and *ex-post* control. The first, *ex-ante* control, refers to legislation that guides the action of actors before actions are taken (Verhoest et al., 2004). Through legislation, oversight authorities exert preemptive control over what agencies can and cannot do with respect to inputs or processes (Thompson et al., 1993). The second, *ex-post* control, concerns whether and how intended organizational goals have been realized and determining if any corrective actions are necessary in the future (Verhoest et al., 2004). An important purpose of *ex-post* control is to ensure agency performance aligns with *ex-ante* forms of control, such as legislation, and to hold agencies accountable in instances of misalignment (Thompson et al., 1993). A key tool used to exert *ex-post* control are external audits, which monitor an agency's compliance to the norms, rules and legislations (Verhoest et al., 2004).

However, it can be challenging to determine what is legally required (*ex-ante* control) in the context of algorithmic governance, due to the rapid pace of technological change and innovation. This can lead to unexpected developments that are hard to foresee and legally anticipate (Schlehahn et al., 2015). In these cases, *ex-post* control may be more effective, but this comes with the downside that monitoring and auditing can be costly and time-consuming (Bertelli, 2006). With the costs and benefits in mind, public agencies usually draw on a mix of *ex-ante* and *ex-post* controls, with the emphasis determined by the political system and the characteristics of the topic in question (Andeweg & Thomassen, 2005).

Bringing these insights together, we argue that institutional transparency—by explaining the regulatory guardrails in place—can play a critical role in promoting trust in predictive policing. We distinguish between two dimensions of institutional transparency, focusing on transparency about *ex-ante* control mechanisms (i.e., legislation and legal frameworks) and *ex-post* control mechanisms (i.e., independent regulatory agencies) tasked with regulating and monitoring police organizations' use of algorithms and personal data (Amicelle, 2022; Brown et al., 2019; Raji et al., 2022; Wirtz et al., 2019).

Linking institutional transparency to trust in predictive policing

While there is a significant body of research documenting the effects of algorithmic transparency on citizen trust in government in general, studies exploring this relationship within the specific context of policing and/or from an institutional perspective of transparency are relatively scarce. The scarcity of research in this area is a concern due to the important role that trust holds in the connection between the police and the public. In studies that do focus on the effects of transparency on trust in algorithmic governance, transparency is often considered and examined from the perspective of algorithmic transparency (i.e., disclosure of information about how these tools work) (e.g., Grimmelikhuijsen, 2023; Kizilcec, 2016; Nieuwenhuizen et al., 2024; Schiff et al., 2022) or the work is conceptual (e.g., Ananny & Crawford, 2018; de Fine Licht & de Fine Licht, 2020; Meijer & Grimmelikhuijsen, 2020). Our paper is one of the first to provide empirical evidence for the relation between institutional transparency and citizen trust in predictive policing.

Trust plays an important role in shaping the relationship between the police and the public, especially in sensitive areas like predictive policing. There are various ways to conceptualize trust in predictive policing, but we draw upon a well-established framework that conceptualizes trust as comprising three dimensions: competence, benevolence and integrity (Grimmelikhuijsen, 2012). This framework provides a multidimensional understanding of trust and the three dimensions are particularly relevant to our study, as they cover the core aspects of public trust in government action. Researchers studying trustworthy algorithm use argue that these dimensions of trust are relevant for understanding how public trust is shaped in the context of algorithm use by governments (Meijer & Grimmelikhuijsen, 2020). This framework has also been successfully applied in studies examining trust in the use of algorithms by government organizations (Ingrams et al., 2022), making it relevant for our research on predictive policing. When we translate the three dimensions to the context of predictive algorithms in policing, competence reflects trust in the police's ability to effectively implement these algorithms in a way that produces accurate and beneficial outcomes. Benevolence addresses the belief that the police use predictive tools with the public's best interests at heart, rather than for self-serving purposes. Integrity concerns the belief that the police are honest and sincere when using predictive algorithms. By conceptualizing trust in this way, we can investigate how institutional transparency can influence citizens' trust in the use of predictive algorithms by the police.

As discussed in the previous section, we focus on transparency about *ex-ante* and *ex-post* control as two key dimensions of institutional transparency. How can they be used to strengthen citizen trust in the use of predictive algorithms by police organizations? One way institutional transparency is believed to affect citizen trust in algorithm use by public organizations is by increasing access to information about official rules governments should adhere to (*ex-ante* control). As Grimmelikhuijsen and Meijer (2022) argue, legal structures are critical institutional mechanisms to maintain legitimacy of algorithmic government (see also Costanzo et al., 2015).

Legal structures are important because they help to define procedural fairness, which refers to the perceived fairness of the process used to make decisions, regardless of the outcome of those decisions (Lind & Tyler, 1988). As past work has shown, procedural justice is positively associated with citizen acceptance of government decisions (Grimes, 2006). Procedural justice provides a valuable framework for understanding how institutional transparency can enhance citizen trust in predictive policing. Procedural justice emphasizes the importance of fairness in the processes through which decisions are made, including transparency, impartiality and respect for human dignity, as key drivers of legitimacy (Tyler & Mentovich, 2023). Transparency, as a component of procedural justice, allows citizens to see and understand how decisions are made, ensuring that the process is perceived to be fair. In the context of predictive policing, institutional transparency about how the use of algorithms by the police is being regulated and monitored, could address citizens' concerns about fairness by demonstrating safeguards against potential misuse.

In the context of predictive policing, procedural justice can be understood as providing residents with information about the guardrails in place to ensure accountable and ethical use of these powerful decision-making tools. Disclosing such information and, in turn fostering greater perceived procedural justice, can build public confidence in algorithmic policing and increase the legitimacy of the law, as citizens are more likely to trust and support systems that they perceive as fair and impartial (Tyler, 2006). Conversely, failure to disclose information about the rules and legislation in place to regulate, for instance, the utilization of big data and the privacy of citizens when it comes to predictive policing, can have profound negative consequences (e.g., Inayatullah, 2013; Schlehahn et al., 2015). Related, studies have also shown that, if citizens are uncertain about the legal limitations of police surveillance and monitoring, they may develop a deep mistrust towards the government (Meijer & Wessels, 2019, p. 1036). As such, we predict that explaining the rules and regulations the police have to comply with regarding their algorithm use will increase citizen trust in predictive policing.

> ***Hypothesis 1:*** *Disclosure of information about legislation on algorithms (ex-ante control) the police must comply with increases citizen trust in predictive policing.*

Another way in which institutional transparency can play a role in enhancing citizen trust is through disseminating information about how algorithms are monitored by external actors once in use (*ex-post* control). There are two reasons for this. First, by drawing the public's attention to oversight agencies that carry out audits and evaluations of how police are using predictive policing tools, the government can communicate to the public that protections are in place to ensure algorithms are being used in ways that are appropriate and effective once implemented (Schlehahn et al., 2015). Further, studies also show that agencies can reassure the public that when problems do exist, audits carried out by oversight agencies will help in identifying and remediating issues that cause adverse consequences attributable to the use of algorithms and personal data (Mittelstadt et al., 2016). In sum, the first path suggests raising public awareness of independent oversight agencies helps to communicate that guardrails are in place to ensure that algorithms are being used for the benefit of citizens, rather than for the benefit of the government itself, thereby increasing citizen trust in the government's use of these technologies (Purves & Davis, 2022, p. 25).

The second reason highlights the role oversight agencies play in lending legitimacy to the legal frameworks that guide the use of algorithms and personal data by police. As research from procedural justice demonstrates, legal frameworks (*ex-ante* control mechanisms) are most effective when paired with enforcement mechanisms, such as independent regulatory agencies tasked with enforcing the law (Tyler, 2006). Past research has demonstrated that, without impartial regulatory agencies to ensure compliance, laws may be perceived as symbolic rather than substantive, doing little to reassure the public or foster trust (Rothstein & Teorell, 2008). When citizens are aware of oversight agencies actively monitoring compliance, they are more likely to believe laws arebeing enforced (Tyler & Huo, 2002). Regulatory agencies play an important role in reassuring the public that the government is taking steps to ensure difficult-to-understand laws are being implemented in ways that advance public well-being (Jackson et al., 2012).

Building on these insights from procedural justice and public accountability research, we expect that information about regulatory agencies is not just likely to foster trust in predictive policing on its own, but also strengthen the effect of information about legislation on trust in predictive policing. This leads to the following hypotheses:

> ***Hypothesis 2:*** *Disclosure of information about external monitoring of the police's use of algorithms (ex-post control) increases citizen trust in predictive policing.*
>
> ***Hypothesis 3:*** *The impact of disclosing information about legislation on algorithms (ex-ante control) on citizen trust in predictive policing is stronger when citizens are also made aware of external monitoring of police's use of this tool (ex-post control).*

5.3 Materials and methods

Research design

To address our research question, we designed an online survey experiment, as survey experiments enable causal conclusions and allow for generalization to a broader population (Thomas, 2024). We used a 2x2 between-subjects factorial design (yes/no legislation transparency, yes/no monitoring transparency), where we randomly vary the type of transparency in the experimental vignettes participants receive. The design is shown schematically in Appendix 5A. This fully randomized experiment has the advantage of high internal validity, meaning that we are able to more precisely estimate the effect of our treatments on our outcome variable: trust in algorithmic recommendations (Shadish et al., 2001).

The flow of our experiment is as follows: First, we brief participants on the study's objectives and request their consent. Subsequently, we collect demographic information. Participants are then randomly assigned to one of the four transparency conditions. Next, we present participants with a short text asking them to imagine they see police patrolling their neighborhood and questioning a neighbor. Subsequently, participants were asked to imagine that they contacted the police and received an explanation about the police's behavior they witnessed earlier. The explanation randomly differed for each experimental condition. Appendix 5B includes the experimental vignettes. Afterward, all participants were directed to the same survey that asked participants about their trust in the use of predictive algorithms by the police, as well as several other questions, such as manipulation and attention checks. At the end of the survey-experiment, participants were debriefed about the goal of the experiment and the manipulated content.

Sample and data collection

In November 2023, we conducted the online survey experiment using Qualtrics. Before collecting data, we preregistered the experiment following the Open Science Framework format (Bowman et al., 2020)[18], and obtained ethical approval from the institutional Ethical Review Committee. Data was collected using the sample-only service of *Dynata*, a well-known global recruitment firm with a substantial respondent pool that can be used for survey distribution. Respondents are free to decide whether they want to join Dynata's participant pool and could choose whether or not they wanted to participate in our experiment. Dynata provided a respondent sample for our experiment with the following parameters: 1) individuals who are Dutch speakers residing in the Netherlands, and 2) participants reflecting the demographic composition of the country in terms of gender, age and highest level of educational attainment. Dynata compensated respondents for their engagement in the survey-experiment upon its successful completion.

Using the software program G*Power, we conducted an *a priori* power analysis to calculate the estimated sample size (Power = .9 and a = .05). The final sample for the experiment was n = 877. We used stratified sampling to ensure we had a representation of the Dutch population. The sample resembles the Dutch population regarding three key background variables: gender, age and education. We took these background variables into account to carry out balance checks (see Appendix 5D).

Experimental vignettes

In this study, we investigate institutional transparency in the context of predictive policing. Institutional transparency is conceptualized along two dimensions: legislation (*ex-ante* control) and external monitoring (*ex-post* control). For the *legislation transparency* condition, we chose to focus on the Dutch Police Data Act (Wet politiegegevens (Wpg)) in our manipulation. The Wpg outlines the conditions for processing personal data, not only of criminals and suspects but also of other individuals involved, such as witnesses, which is necessary for the police to perform their duties. In the context of predictive policing, the Wpg prescribes when and for which tasks the predictive policing algorithm may use personal information of citizens (Alkemade & Toet, 2021). In the experimental vignette, the legislation transparency condition thus includes information about an existing, relevant act that safeguards responsible use of data for predictive policing.

[18] Open Science Framework Registration: https://osf.io/pshjt/?view_only=8fa76d1524d2483692cde-7b24e1426ec.

For the *external monitoring transparency* condition, we used the Dutch Data Protection Authority (Autoriteit Persoonsgegevens (AP)) in our manipulation. The AP is an external, independent government institution that oversees compliance of the police with the Wpg (Hirsch Ballin & Oerlemans, 2023). In the context of predictive policing, the AP can monitor the police's use of personal data for their predictive policing tool. In the experimental vignette, the monitoring condition thus includes information about an existing external monitoring agency that oversees the police's use of personal information for the predictive policing algorithm.

Predictive policing has become a standard practice in the Netherlands, making it an ideal research context for investigating realistic scenarios of its application. The adoption of the Crime Anticipation System (CAS), a predictive policing tool designed to forecast crime locations, has risen significantly since its nationwide implementation in 2019. This makes the Netherlands the first country globally to utilize predictive policing at a national level (Strikwerda, 2021). The experimental vignettes (see Appendix 5B) are based on the functionalities and inner workings of CAS, indicating, for example, an increased risk of burglaries in a specific neighborhood (for more information on CAS, see Strikwerda, 2021).

Overall, this allows us to investigate a realistic scenario of the two dimensions of institutional transparency in a realistic context of predictive policing in the Netherlands, fostering high mundane realism of the survey-experiment (Difonzo et al., 1998). The full text of the two manipulations can be found in Appendix 5B.

The operationalization of institutional transparency and the survey experiment were extensively pre-tested prior to its implementation in three ways. First, we asked two Dutch police employees for input on the experimental vignettes in order to increase mundane realism (Difonzo et al., 1998). Second, to ensure measurement validity, we carried out four face to face cognitive interviews to ensure survey questions were well understood (DeVellis, 2017). We identified a few minor discrepancies in the translation from English to Dutch, leading to subtle revisions to ensure the survey items accurately conveyed the intended meaning across languages. Third, following the recommendation by Ejelöv and Luke (2020), we conducted two pilot studies to test the validity of manipulations in an environment where the dependent measure is not influenced by the construct the manipulation aims to target.

The first pilot study was conducted on March 8th 2023 through the online platform Amazon Mechanical Turk (MTurk) and included a sample of 296 participants. This pilot study aimed to test our preliminary implementation of the manipulations of institutional transparency. It revealed that the distinction between *ex-ante* and *ex-post* control was not clear enough. Participants seemed to struggle with differentiating between the manipulation conditions. As a result, we refined the experimental vignettes to make this distinction more apparent. We accomplished this by significantly shortening the text (by approximately 60 words) and by highlighting the two conditions (legislation and monitoring) more clearly through a question in the vignette (see Appendix 5B). We then tested these adjustments in a second pilot study. The second pilot was conducted on May 29th 2023 through Dynata and included a representative sample of 201 participants from the Netherlands. We preregistered this pilot at Open Science Framework.[19] The results of this pilot study showed that participants were better able to identify the experimental condition they were placed in. Chi-squared tests confirmed that participants who received information about a specific condition were more likely to correctly indicate receiving this information compared to those in other conditions. Therefore, we proceeded with the full study using these refined experimental vignettes, which were discussed extensively above and can be found in Appendix 5B.

We deliberately designed our manipulations to highlight the existence of legislation and monitoring mechanisms, focusing on the act of disseminating information about these mechanisms rather than the detailed content of those laws and oversight practices. This approach reflects three considerations. First, we wanted participants to recognize the presence of regulatory frameworks and external oversight without overwhelming them with legal intricacies. Including too many details risked detracting from the study's main objective by directing participants' attention away from the broader question of whether knowing about safeguards affects their trust. Second, we were mindful of how overly complex information might shift participants from a compliance-oriented mindset to confusion or disengagement. By keeping the descriptions concise and clear, we aimed to maintain participants' focus on the core elements of the intervention. Finally, during the pre-testing phase, we discovered that longer, more detailed vignettes reduced participants' ability to retain the information provided. Through multiple iterations, we refined our materials to ensure that participants could readily distinguish between different conditions and clearly identify the nature of the treatment they received. This balance allowed us to highlight institutional safeguards in a manner that was easy to comprehend, without sacrificing the study's focus on how awareness of such frameworks impacts trust.

[19] Open Science Framework Registration: https://osf.io/tjmec/?view_only=390acf47cbb14087ae90de75d9c66cf5.

Measures

Our outcome variable is citizen trust in the use of predictive algorithms by the police, which was measured using a modified version of the 'Citizen Trust in Government Organizations' scale, validated by Grimmelikhuijsen and Knies (2017). This modification involved adapting the scale by specifying the relevant domain (the use of predictive algorithms) and organization (the police), as proposed by Grimmelikhuijsen and Knies (2017, p. 592). The scale operationalizes trust using three dimensions (competence, benevolence and integrity), with 3 items for each dimension, resulting in 9 items in total (see Appendix 5C), measured on a 5-point Likert scale (strongly disagree, disagree, neutral, agree, and strongly agree). In the survey experiment, we randomized the order of the nine items to avoid possible ordering effects (De Jong et al., 2010).

Because Grimmelikhuijsen and Knies' (2017) scale has previously been used by only one study on algorithms (Ingrams et al., 2022), we ran a confirmatory factor analysis with a measurement model to validate and assess the reliability and validity of the scale in the context of our research to ensure its applicability and robustness for our study on algorithms. The model consists of three common factors (competence, benevolence and integrity) that explain nine observed variables. We consider the model fit to be satisfactory if RMSEA < .10 (Browne & Cudeck, 1992) and CFI > .90 (Bentler & Bonett, 1980), indicating respectively exact to mediocre fit and good fit. As the model fit was satisfactory, RMSEA = .046, RMSEA 90% CI [.033, .059], CFI = .993, the measurement model and therefore the common factor structure was accepted. Appendix 5E includes a detailed description of the model fit results. We ran our analyses separately for competence, benevolence and integrity, using the predicted common factor scores.

Finally, we used two manipulation checks to identify respondents' thoughts regarding the independent variable being manipulated. We had one question for the legislation transparency manipulation and one for the monitoring transparency manipulation. This allowed us to test whether the treatment groups differ, on average, from the control group. The questions for and results of the manipulation checks can be found in Appendix 5F.

Data analysis

We use linear multiple regression to test how transparency about legislation and monitoring affects citizen trust in the use of predictive algorithms by the police. We opted for a multiple linear regression analysis in our research over a 2x2 factorial ANOVA because regression allows us to examine each predictor individually as regression coefficients, assessing the significance of each variable in predicting the outcome. This approach provides a better and more detailed understanding of the relationships we are interested in, revealing if specific predictors are significant while others may not be. In contrast, a 2x2 factorial ANOVA employs an omnibus test that collectively assesses all predictors, preventing the separate examination of each variable. Since our interest lies in understanding the effects of different types of institutional transparency and the magnitudes of their impacts, a regression analysis is more appropriate for measuring the relationships and variations between these predictors and the outcome variable.

First, we fit Model 1, the intercept-only model. Second, we fit Model 2, using only the main effects of legislation transparency and monitoring transparency to predict participants' trust scores. If this model explains the data significantly better than Model 1, we continue with the next step. Third, we fit Model 3, by adding the interaction effect (interaction between legislation and monitoring transparency) to Model 2. Again, only if Model 3 explains the data significantly better than Model 2, we continue with the next step. Lastly, we fit Model 4, by adding the personal and general covariates to Model 3. We use these hierarchical steps to determine the effect of the main predictors first, before testing how these effects interact with each other and controlling for covariates (Teo, 2013).

Hypothesis 1 is supported when the *t*-test of the coefficient for the dummy variable legislation transparency is significant. Hypothesis 2 is supported when the *t*-test of the coefficient for the dummy variable monitoring transparency is significant. Hypothesis 3 is supported when the *t*-test of the interaction coefficient of legislation and monitoring transparency is significant. For all tests in our data analysis, we use an alpha of .05.

Furthermore, we use a chi-squared test to analyze the responses to the manipulation check. This tells us whether there is a significant association or difference between the experimental groups and the likelihood of people choosing the correct option over the incorrect option. In other words, it assesses whether the observed distribution of responses is different from what we would expect by chance.

5.4 Results

Descriptive statistics for each dimension of our measure of trust is as follows. On average, participants scored 3.25 out of 5 on the competence dimension (*SD* = 1.00), 3.48 out of 5 on the benevolence dimension (*SD* = 0.96), and 3.25 out of 5 on the integrity dimension (*SD* = 1.01). See Appendix 5E for the summary statistics of the common factor scores. The regression analyses were performed using the common factor scores for competence, benevolence and integrity. A score of 0 corresponds to the average score on each dimension. We discuss the standardized regression coefficients for all significant models to aid in the interpretability of the results. Results for each dimension of trust are outlined below, starting with perceived competence.

Competence. Table 5.1 shows the regression results for competence. The main effects model (Model 2) explained significantly more variance than the intercept-only model (Model 1), *F*(2, 874) = 4.26, *p* = .014. The interaction effect model (Model 3) did not explain significantly more variance than the main effects model, *F*(2, 873) = .04, *p* = .809. Therefore, we focus our discussion on the main effects model, which does not include an interaction term (Model 2). Model 2 explained $R^2 = 0.01$ proportion of all variance in the common factor scores. Participants who received information about legislation did not have significantly higher perceived competence scores than participants who did not receive this information, *b* = 0.08, *p* = .141, SE = 0.06, 95% CI = [-0.028, 0.194]. Participants who received information about monitoring were expected to have β = 0.08 standard deviations higher perceived competence scores than participants who did not receive this information, *b* = 0.14, *p* = .012, SE = 0.06, 95% CI = [0.032, 0.254]. This finding suggests that citizens' perceptions that police competently use predictive policing tools are improved when governments disclose information about monitoring mechanisms, but not when they explain legislative steps in place to guide the usage of these tools. In other words, the public seems to be more sensitive to some types of institutional transparency (i.e., information about monitoring) than they are to others (i.e., legislative transparency) when it comes to evaluating how competent police are when using predictive policing tools.

Benevolence. Table 5.2 presents the regression results for benevolence. The main effects model (Model 2) explained significantly more variance than the intercept-only model (Model 1), *F*(2, 874) = 6.49, *p* = .002. This means that the addition of the treatments to our regression model significantly improved model fit. The interaction effect model (Model 3) did not explain significantly more variance than the main effects model, *F*(2, 873) = 1.14, *p* = .285, thus indicating that the inclusion of the interaction term did not improve the explanatory power of our regression model. Therefore, we again focus our discussion on the main effects model (Model 2). Model 2 explained $R^2 = 0.01$ proportion of all variance in the common factor scores. Participants who received information about legislation did not have significantly higher perceived benevolence scores than participants who did not receive this information, *b* = 0.09, *p* = .066, SE = 0.05, 95% CI = [-0.006, 0.196]. Participants who received information about monitoring were, on average, found to have β = 0.10 standard deviations higher benevolence scores than participants who did not receive this information, *b* = 0.16, *p* = .002, SE = 0.05, 95% CI = [0.059, 0.260]. Thus, as with perceptions of competence, citizens seem to be more responsive to information about monitoring mechanisms than legislation when evaluating police benevolence in using predictive algorithms.

Integrity. Table 5.3 presents the regression results for integrity. The main effects model (Model 2) explained significantly more variance than the intercept-only model (Model 1), *F*(2, 874) = 5.56, *p* = .004. The interaction effect model (Model 3) did not explain significantly more variance than the main effects model, *F*(2, 873) = .03, *p* = .861. Therefore, as with the competence and benevolence dimensions, we focus our discussion on the main effects model (Model 2). Model 2 explained R^2 = 0.01 proportion of all variance in the common factor scores. Participants who received information about legislation did not have significantly higher perceived integrity scores than participants who did not receive this information, *b* = 0.10, *p* = .074, SE = 0.05, 95% CI = [-0.009, 0.204]. Participants who received information about monitoring were expected to have β = 0.09 standard deviations higher integrity scores than participants who did not receive this information, *b* = 0.15, *p* = .005, SE = 0.05, 95% CI = [0.047, 0.260]. These results imply the perceptions of police integrity regarding their use of predictive algorithms are also positively influenced by information about monitoring mechanisms, whereas information about legislation does not influence these perceptions. However, it is also important to note that this effect is more modest when compared to the other dimensions of perceived trustworthiness.

Cumulatively, the regression analyses show consistent results for all three dimensions of perceived trust in police use of predictive policing tools. The disclosure of information about data protection legislation the police must comply with when using predictive algorithms does not significantly increase any aspect of citizen trust in the use of these algorithms by the police. Hypothesis 1 is therefore rejected. However, we found support for hypothesis 2. The disclosure of information about external monitoring of the police's use of algorithms significantly increased citizen trust in the use of predictive algorithms by the police. Lastly, we did not find support for our hypothesis 3, as the interaction effect model did not significantly explain the data better than the main effects model. The impact of disclosing information about data protection legislation on citizen trust in the use of predictive algorithms by the police is not significantly stronger when citizens are also made aware of external monitoring of police's use of this tool.

We conducted a sensitivity analysis to check for differences in the main effects between participants with and without a migration background (i.e. one or both parents were born outside of the Netherlands). The sensitivity analysis showed no substantive alteration to our findings (i.e., no significant effect became nonsignificant, and no nonsignificant effect became significant).

Table 5.1 Regression coefficients for competence

	Estimate	***SE***	***t***	***p***	**95% CI**
Model 1: Intercept only					
Intercept	0.00	0.00	0.00	1.000	[-0.056, 0.056]
Model 2: Main effects					
Intercept	-0.11	0.05	-2.31	.021*	[-0.209, -0.017]
Legislation transparency	0.08	0.06	1.48	.141	[-0.028, 0.194]
Monitoring transparency	0.14	0.06	2.52	.012*	[0.032, 0.254]

* *p* < .05 ** *p* < .01

Table 5.2 Regression coefficients for benevolence

	Estimate	***SE***	***t***	***p***	**95% CI**
Model 1: Intercept only					
Intercept	0.00	0.00	0.00	1.000	[-0.051, 0.051]
Model 2: Main effects					
Intercept	-0.13	0.04	-2.86	.004**	[-0.215, -0.040]
Legislation transparency	0.09	0.05	1.84	.066	[-0.006, 0.196]
Monitoring transparency	0.16	0.05	3.10	.002**	[0.059, 0.260]

* *p* < .05 ** *p* < .01

Table 5.3 Regression coefficients for integrity

	Estimate	***SE***	***t***	***p***	**95% CI**
Model 1: Intercept only					
Intercept	0.00	0.00	0.00	1.000	[-0.054, 0.054]
Model 2: Main effects					
Intercept	-0.13	0.05	-2.67	.008**	[-0.218, -0.033]
Legislation transparency	0.10	0.05	1.79	.074	[-0.009, 0.204]
Monitoring transparency	0.15	0.05	2.82	.005**	[0.047, 0.260]

* *p* < .05 ** *p* < .01

5.5 Discussion

Implications for scholars

This study contributes to public administration scholarship by addressing important considerations related to the use of algorithms in public decision-making, as emphasized by scholars such as McDonald et al. (2022). Specifically, this research advances our understanding of how institutional transparency influences citizen trust in the use of predictive algorithms by public agencies. Our findings emphasize that transparency regarding the external monitoring of algorithm use—rather than simply disclosing information about legal frameworks—is essential for fostering trust. For public administration, these insights suggest that transparency initiatives should focus on ensuring that citizens are informed about the active monitoring of an organization's use of algorithmic systems to reinforce accountability and build trust. These findings align with prior research, such as Wenzelburger et al. (2024), which found that in the context of predictive policing, there is a strong demand for institutional transparency that can prevent abuse of state power.

Our results also show that not all forms of institutional transparency contribute equally, aligning with established research on the dynamics between transparency and trust, which finds that different types of transparency do not all have the same positive impact on trust (e.g., Grimmelikhuijsen, 2012; Wang & Guan, 2023). These findings underscore the importance of disclosing information about monitoring algorithms, whereas information about legislation is less important in strengthening citizen trust. Therefore, this research offers important implications for the broader discourse on the relationship between algorithmic accountability and transparency. Specifically, while earlier research regarding algorithmic accountability distinguishes between direct public accountability via public transparency and indirect public accountability via transparency to auditors in public organizations (Loi & Spielkamp, 2021), our findings point to the importance of a third path: accountability via public transparency about auditors monitoring the use of predictive algorithms by police. Taken together, these insights highlight how public managers can strengthen trust in algorithmic governance by emphasizing visible oversight mechanisms and communication, thus clarifying and reinforcing the role of external monitors in safeguarding the public interest.

In terms of implications for police organizations, in contrast to the hypothesized positive effect of disclosing information about legislation governing police's use of predictive algorithms, our study reveals that this form of transparency does not matter much to citizen's trust in police's use of predictive policing. Neither providing information about legislation alone nor in combination with monitoring transparency seem to affect citizen trust. One way of interpreting these null effects is that citizens may have specific and nuanced information needs when it comes to evaluating the trustworthiness of predictive policing tools in particular, and government's use of artificial intelligence more generally. In turn, if efforts to leverage transparency to foster greater trust in predictive algorithms are to succeed, more attention has to be given to understanding just what information citizens want. This interpretation dovetails with existing transparency research, which demonstrates significant mismatches are present in terms of the information governments publicly disclose and the types of information citizens need (Cucciniello & Nasi, 2014). Extending these insights within the context of this study, our findings suggest that citizens are less interested in knowing about the legal frameworks used to guide the use of artificial intelligence tools, such as predictive policing. Instead, information outlining what happens when these tools do things they are not supposed to do appears to be more salient to citizen evaluations of government's use of artificial intelligence.

Further, related to the need to consider citizen information needs, this study highlights how citizens seem to distinguish between information about proactive institutional control mechanisms that are implemented before the deployment of algorithms (*ex-ante*) and reactive measures that are deployed after these tools are in use (*ex-post*). Research on algorithmic transparency, as well as government transparency more generally, has highlighted the importance of different aspects of transparency to citizen decision-making. For example, Grimmelikhuijsen (2012) divides types of transparency into decision-making, policy and policy outcome transparency. Research on algorithmic transparency highlights the importance of explainability, i.e., the extent to which the public can understand how an algorithm reached a particular decision, and accessibility, i.e., the extent to which the public is able to access information about how an algorithm was made (Shin & Park, 2019). Our findings build upon these conceptualizations by highlighting the limitations of algorithmic transparency and exploring additional dimensions of transparency that researchers must account for when attempting to understand public evaluations of government's use of complex technologies such as algorithms and artificial intelligence, namely information about proactive (*ex-ante*) and reactive (*ex-post*) institutional control mechanisms. This distinction is important because it offers additional insight into factors that help to explain the long-debated relationship between transparency and trust. While algorithmic transparency can contribute to increased citizen trust in specific algorithms that are used by public organizations such as the police (Nieuwenhuizen et al., 2024), institutional transparency, particularly when it involves providing information about external monitoring, can prove effective in strengthening trust in the use of algorithms by these organizations.

Finally, the discussion within the field of regulation explores the relationship between control and trust in the context of algorithmic governance (Tamò-Larrieux et al., 2024). It often appears that a trade-off exists: if there is already sufficient trust, additional regulation and control might seem redundant. However, legislation alone is insufficient; adherence to regulations and being transparent about this is crucial for building trust (see also Grimmelikhuijsen et al., 2024). This study demonstrates that trust can be enhanced through information about effective enforcement of regulations, suggesting that the trade-off between trust and control is not necessarily present. In other words, information about robust control and adherence to regulations can even contribute to a heightened sense of trust in the use of algorithms by public agencies.

Limitations

While our study contributes valuable insights into the relationship between institutional transparency and citizen trust in the use of predictive algorithms by police organizations, it is essential to acknowledge several limitations that might influence the generalizability and robustness of our findings. Firstly, the results revealed small standardized regression coefficients, indicating that the observed effect of monitoring transparency on citizen trust is relatively modest. This finding aligns with the nature of survey-experiments examining the complex relationship between transparency and trust, as demonstrated in previous research by Wang and Guan (2023). It is crucial to recognize that survey-based methodologies inherently capture perceptions, and the small effect size might be indicative of the complex nature of trust-building processes in public administration. Future research could investigate additional factors that may mediate or moderate this relationship, such as cultural differences, previous experiences with law enforcement, or the perceived effectiveness of predictive algorithms in crime prevention. Moreover, longitudinal studies could provide a better understanding of how citizen trust evolves over time in response to changes in institutional transparency.

Secondly, the context of our study poses a limitation to the external validity of our findings. The legislative and monitoring context investigated in our research is specific to the Netherlands, as are the predictive policing practices. While our study offers valuable lessons that may be relevant to similar (police) domains where algorithms are employed, it is important to recognize that the unique features of each context require careful consideration. Variations in institutional structures, cultural norms and legal frameworks may shape the dynamics of the transparency-trust relationship (Wenzelburger et al., 2024).

However, the broader implications of this study extend to other contexts where public organizations use algorithms. This emphasizes the important role of external monitoring transparency in fostering trust. Although rooted in the domain of predictive policing in the Netherlands, the demonstrated preference for *ex-post* accountability mechanisms may indicate a more universal principle: citizens prioritize assurances that the use of algorithms by organizations is actively monitored for compliance with legislation. This principle likely holds relevance across countries and sectors where algorithms are being employed, such as healthcare, social services and education.

That said, the Netherlands is a forerunner in both the adoption of algorithms by public institutions and the development of legislative frameworks to regulate their use. Given this context, generalizing directly to other countries may be challenging, as regulatory and institutional contexts can vary significantly. Nevertheless, the insights derived from this study, particularly regarding the importance of monitoring transparency, remain very relevant and can offer valuable lessons for other contexts. Future research could explore the applicability of our insights to other settings, taking into account the contextual variations and nuances inherent in different regulatory and policing environments. Additionally, investigating how the findings of this research apply to other algorithms, both within the policing sector and across different government domains, would be highly interesting.

Societal implications and recommendations

This research highlights the importance for police organizations and, more broadly, public organizations to proactively explain how oversight mechanisms monitor their algorithm use. They should put effort in developing communication strategies to inform the public about the external audits that took place and, if possible, elaborating on the outcomes of these audits.

Currently, oversight authorities and regulatory agencies face resource constraints when auditing the use of algorithms by public organizations. This is in line with earlier research on *ex-ante* and *ex-post* control mechanisms in which Verhoest et al. (2004) see a preference *for ex-ante* control mechanisms, because *ex-post* control is very hard to accomplish. External monitoring is often underdeveloped and ineffective, characterized by insufficient staff, lack of expertise or inadequate mandates. Addressing this capacity gap is necessary for meaningful audits to take place, as indicated by Raji et al. (2022). Our study underscores the significance of external monitoring as a crucial component in fostering trustworthy predictive policing. This, in turn, contributes to a functioning democracy, thereby enhancing the legitimacy of law enforcement. Moreover, external monitoring not only has a positive impact on citizens, it could also help public organizations in fostering responsible use of algorithms as these audits establish and signal the competency of institutions in using algorithmic systems (Purves & Davis, 2022).

5.6 Conclusion

In this paper, we aimed to answer the question "To what extent does institutional transparency affect citizen trust in predictive policing?". We conceptualized two dimensions of institutional transparency. The first is disclosure of information about *ex-ante* control, i.e., information about legislation governing police's use of predictive algorithms. The second is disclosure of information about *ex-post* control, i.e., information about external monitoring of police's use of predictive algorithms. The findings from our survey-experiment show that only disclosure of information about how the use of predictive algorithms by the police is monitored significantly contributes to greater citizen trust. Moreover, we find that this effect is consistent across all three dimensions of trust. This means that citizens have higher trust in police's competence, benevolence and integrity regarding the use of predictive algorithms when they are told that police's use of these tools is being monitored by an external agency. Information about legislation alone or in combination with information about monitoring did not increase citizen trust. Moving forward, this study's conceptual and methodological contributions can help scholars and practitioners better understand the relationship between institutional transparency and trust in algorithm use by governments and guide policy discussions on the oversight of algorithmic decision-making in law enforcement.

6

Conclusion and discussion

The previous chapters present an analysis of the relationship between transparency and citizen trust in algorithm use by government organizations. The empirical chapters analyzed a specific algorithm for citizen recommendations, but also the organizational embedding of algorithms in an algorithm register as well as the institutional embedding of algorithms in legislation and monitoring. Altogether, these chapters provide a rich overview of how transparency about algorithm use by governments at different levels impacts citizen trust. Below, I first provide a summary of the answers to the four sub questions that guided this research. Then, I answer the central research question of this dissertation, resulting in three key messages. Next, I discuss the broader contributions to theory and reflect on the methodological approach and the limitations of the studies, leading to an agenda for future research. Finally, I offer several recommendations for practice and conclude with a closing remark regarding the future of algorithmization.

6.1 Answers to the research questions

RQ1: *How can we understand the relationship between trust and algorithmic transparency at the micro, meso and macro levels?*

The literature on algorithmic transparency and trust shows that most studies have investigated this relationship in a rather narrow manner. For instance, the studies by Grimmelikhuijsen (2023) and Nieuwenhuizen et al. (2024) focus on transparency of the algorithmic application itself, whereas general transparency literature suggests that transparency also plays a role at higher-order levels, such as at the organizational and institutional levels (Porumbescu et al., 2022). This includes transparency about organizational policies and institutional rules on the appropriate implementation and use of algorithms in government (Grimmelikhuijsen & Meijer, 2022). How this broader understanding of transparency can be applied to enhance trust in algorithmic governance remains unexplored in the literature.

Chapter 2 addresses this gap by conceptualizing trust in algorithmic governance and discussing how transparency can be used to build, maintain and strengthen citizens' trust in algorithmic governance. This results in a conceptual framework that can be used to empirically examine the relationship between transparency and trust in algorithmic governance. This is a multi-level framework that focuses on micro (specific algorithms), meso (organizational) and macro (institutional) level transparency challenges (Giest & Grimmelikhuijsen, 2020, p. 411) and on how these are linked to citizen trust. This chapter shows that to build, maintain and strengthen trust in algorithmic governance, it is not only necessary for algorithms themselves to be transparent, but that the way in which they are used by government organizations and the regulations regarding government organization's algorithm use should also be transparent. Additional insights about the relationship between trust and algorithmic transparency at the micro, meso and macro levels emerged from the three empirical studies. These insights are presented in Table 6.1 and discussed in the answer to the central research question of this dissertation.

One important insight that I want to highlight here is that empirical studies have demonstrated that the term "algorithmic governance", as used in Chapter 2, is less suitable for addressing the issues examined in this dissertation than "algorithm use by governments". This is due to the fact that algorithmic governance has a narrower meaning than algorithm use by governments. According to Katzenbach and Ulbricht (2019, p. 1), algorithmic governance "highlights the idea that digital technologies produce social ordering in a specific way". This definition limits the analysis of public sector algorithms to "governance by algorithms" (Katzenbach & Ulbricht, p. 2), emphasizing the ways in which algorithms themselves shape and influence governance structures. In contrast, algorithm use by governments has a broader meaning. This dissertation argues that this concept includes both the use of algorithms to support decision-making and the ways in which they are embedded within broader institutional and organizational frameworks. Consequently, in addressing the central research question, this dissertation conceptualizes algorithm use by governments as the object of trust.

RQ2: *What are the effects of explanations on citizen trust in algorithmic recommendations?*

Research has shown contrasting insights into what kind of explanations strengthen (citizen) trust in algorithmic outcomes. De Fine Licht and de Fine Licht (2020) argue in a conceptual paper that providing justifications for individual algorithmic decisions is crucial for generating legitimacy and trust. Kizilcec (2016), on the other hand, showed that providing information to students about the algorithmic decision procedure regarding their grades was most effective in strengthening trust in the algorithmic outcome. Altogether, the type of explanation that is most effective in strengthening citizen trust has only been investigated to a limited extent and has remained contested.

Chapter 3 addresses this gap by testing the effects of various explanations provided by algorithmic recommender systems on citizen trust in two survey-experiments. The first experiment builds upon the concepts of procedural, rationale and combined explanations (Kizilcec, 2016); the second experiment focuses on directive explanations (Singh et al., 2021a). While nuances exist, the findings show that explaining algorithmic recommendations—in any form—

strengthens trusting beliefs, trusting intentions, and trust-related behavior in citizens who receive digital public service deliveries. This may suggest that trust in algorithmic recommendations increases when citizens see that governments make an effort to provide an explanation, regardless of the nature of this explanation.

RQ3: *What is the nature of algorithm registers, and what are their implications?*

The early scholarly debate about the new phenomenon of algorithm registers shows that the value of these registers is questioned: some scholars claim that governments only disclose politically non-sensitive algorithms and provide information with limited usefulness for accountability purposes (Cath & Jansen, 2022). In contrast, a more optimistic view is also put forward: registers could help organizations to be transparent to the public, which increases public trust in algorithm use (Floridi, 2020). However, the academic debate on this topic is still in its early stages, making it difficult to understand what algorithm registers are and what impact they might have.

To address this gap, Chapter 4 provides a theoretical understanding and an empirical mapping of the different factors that shape algorithm registers and their implications. Experiences from the Netherlands, a leading country in adopting public algorithm registers, show that a great diversity in algorithm registers exists. The nuanced mapping of algorithm registers' characteristics reveals that there are many difficult choices to make regarding what an algorithm register is, what objectives it tries to reach and how it should be designed. To create a useful and valuable algorithm register, these choices must be aligned. Furthermore, the findings show that the reality is nuanced regarding the implications of these registers. While public organizations indeed selectively disclose information and registers are currently not found useful by societal watchdogs and oversight authorities, there are also dynamics that could contribute to more responsible use of algorithms by public organizations. For instance, the process of registering algorithms forces organizations to critically evaluate their algorithmic processes, potentially creating a disciplining effect. The chapter argues that algorithm registers are currently not a meaningful tool for transparency, but a meaningful box-ticking exercise for public organizations.

RQ4: *To what extent does institutional transparency affect citizen trust in predictive policing?*

Research shows that algorithmic transparency is viewed as a potential solution for increasing trust in algorithmic policing as it can help address key threats to foster trust in predictive policing (Alikhademi et al., 2022; Bakke, 2018). These threat include concerns over the use of biased data and the lack of accountability for algorithmic decisions, which can be mitigated by embedding algorithms in an organization in ways that allow outsiders to monitor how algorithms work and perform (cf. Grimmelikhuijsen and Meijer 2014). While empirical studies mainly focus on providing transparency about how specific algorithms function, several scholars argue and show that transparency about the context in which algorithms operate is equally important (Grimmelikhuijsen & Meijer, 2022; New & Castro, 2018; Wenzelburger et al., 2024). However, research on providing transparency about the institutional context in which algorithms are situated is lacking.

Chapter 5 addresses this gap by investigating whether two dimensions of institutional transparency, transparency about legislation and external monitoring, can strengthen trust in predictive policing. By doing so, the chapter provides a theoretical and methodological framework for conceptualizing and investigating institutional transparency in the context of algorithm use by governments. The findings in Chapter 5 show that disclosing how police's use of predictive algorithms is monitored by an external oversight agency, significantly increases citizen trust, while providing details about predictive policing legislation does not. This highlights the importance of transparency about external monitoring in safeguarding citizen trust regarding police use of predictive algorithms, offering insights for policymakers and law enforcement.

Table 6.1 presents the multi-level approach to transparency in the context of government use of algorithms (RQ1). It also highlights the main findings on how algorithmic transparency (RQ2), organizational transparency (RQ3) and institutional transparency (RQ4) influence citizen trust in governments' use of algorithms.

Table 6.1 Main findings regarding the influence of transparency on citizen trust in the use of algorithms by governments

Level	Type of transparency	Definition	Main findings
Micro	Algorithmic transparency	Information about how an algorithm functions	▪ Explanations about how an algorithm functions and/or how it comes to its output significantly increase citizen trust in the algorithm.
Meso	Organizational transparency	Information about the organizational embedding of algorithms	▪ Partial or selective disclosure of information about the organizational embedding of algorithms can negatively impact trust in the use of algorithms by governments. ▪ The requirement to provide organizational transparency can foster a disciplining effect, as organizations have to critically asses their algorithmic decision-making practices.
Macro	Institutional transparency	Information about the institutional embedding of algorithms	▪ Information about the rules and regulations regarding an organization's use of algorithms does *not* increase citizen trust. ▪ Information about the monitoring of an organizations' compliance to rules and regulations regarding their algorithm use significantly increases citizen trust.

By combining the main conclusions of the different sub studies, I can formulate an answer to the central research question of this dissertation:

***RQ**: How does transparency affect citizen trust in algorithm use by government organizations?*

This research highlights the limitations of viewing transparency solely as an information process. The findings from the studies reveal two mechanisms through which algorithmic transparency can foster citizen trust: a communicative (direct) route and a disciplining (indirect) route. Both routes have been discussed in the transparency literature before, but this dissertation shows how they coexist in the context of algorithmic transparency and interact with each other at different levels. This illustrates that the relationship between algorithmic transparency and trust occurs not only within individual levels but also through an interconnectedness across the micro, meso and macro levels.

The first route concerns the *communicative function* of transparency. This involves what organizations communicate externally about their algorithms: explaining how they work and produce their outputs, and how they are embedded within organizational and institutional frameworks. However, an overemphasis on the communicative function of transparency carries the risk of it being used manipulatively, by sharing only information that fosters trust. On the other hand, complete disclosure risks causing information overload. As is evident from the answers to the various sub-questions, more transparency does not always lead to more trust.

At the micro, meso and macro levels, providing transparency requires a tailored approach. An important aspect of this tailored approach is identifying your target audience—in other words, who you are addressing. When public organizations increase transparency about their algorithm use for citizens, this does not lead to an increase in public trust in all situations. One reason is that citizens may not always read or critically engage with the information provided, as discussed in Chapter 3.

Additionally, the information organizations share about their algorithms may be too complex for citizens to understand (Chapter 4) or may not align with what citizens actually want to know (Chapter 5). Therefore, having a critical audience (such as transparency intermediaries) is essential to review the information these organizations provide, ensuring proper checks and balances on government use of algorithms (Kemper & Kolkman, 2019; Scholtes, 2012).

Currently, public organizations try to serve all audiences (regulators, journalists, NGOs, citizens, etc.) with their transparency efforts, as shown in Chapter 4. However, the insights from Chapters 4 and 5 indicate that different audiences have distinct information needs and varying levels of technical understanding (see also van Vliet et al., 2024). For the communicative function to be effective in strengthening citizen trust, algorithmic transparency efforts must be tailored to meet the specific needs, interests and capacities of each audience. This ensures that the right information is delivered in a format that resonates with and is accessible to them.

The second route involves the *disciplining function* of transparency. The requirement to provide transparency encourages public organizations to engage in critical reflections, which can lead to improved algorithmic decision-making processes and operations (a disciplining effect), as shown in Chapter 4, and to better substantiated algorithmic decisions, as argued in Chapter 3. By considering in advance how to communicate the algorithm and its integration within organizational and institutional contexts, these processes become more carefully designed and refined in preparation for potential public exposure (see also de Fine Licht & de Fine Licht, 2020; Levy et al., 2021; Ripamonti, 2024; van Vliet et al., 2024).

This proactive approach contrasts with the reactive transparency that is still common today, where organizations are required to disclose information about their algorithms and their implementation only after receiving Freedom of Information (FOI) requests. In these cases, post-hoc reasoning and justification often prevail over ex-ante reasoning, which is encouraged by proactive transparency (Wessels, 2024). With reactive transparency, organizations may "justify what has already been done", potentially presenting a skewed or overly favorable view. In contrast, proactive transparency may foster a process of reflection and follow-up actions that have intrinsic value for the organizations themselves, serving as a disciplining effect on their transparency practices. This proactive transparency forces organizations to critically evaluate their internal processes. Thus, while transparency may not yield immediate positive effects for external audiences, especially given that proactive transparency is often voluntary, with governments often only publishing positive information proactively, it remains a valuable mechanism in its own right for public organizations.

This dissertation empirically demonstrates that the two routes above often operate independently, as they serve distinct functions. Prescriptively, this implies that public organizations should attend to the different implications of each route. In the short term, the communicative function of transparency could *directly* foster trust by informing citizens and other relevant stakeholders about how the algorithms function and produce their outputs, and how they are organizationally and institutionally embedded. In the long term, the disciplinary function of transparency could *indirectly* strengthen trust by encouraging public organizations to continually evaluate and refine their algorithmic practices. This dual focus—on both effective communication and rigorous internal reflection—can contribute to trustworthy use of algorithms by government organizations.

Finally, the findings in this dissertation demonstrate that the relationship between transparency and trust occurs not only within individual levels but also through an *interconnectedness* across the micro, meso and macro levels. Distinguishing between micro, meso and macro levels is relevant, as it offers both theoretical and practical insights into how transparency can be operationalized in theory and implemented in practice, as shown in Chapter 2. However, the relationship between transparency and trust exists not only within individual levels but also extends through interconnections across micro, meso and macro levels (see also Porumbescu et al., 2022).

For example, Chapter 5 demonstrates that transparency at the macro level, specifically by providing information on the monitoring of legal compliance in algorithm use, can enhance meso trust in the organizational embedding of algorithms, particularly in how the algorithm is used by the police. Furthermore, at the macro level, transparency around legal compliance promotes public trust, but at the meso level, providing transparency about the organizational embedding of algorithms proves challenging and monitoring practices are very limited thus far. Here, a tension arises: while monitoring transparency is necessary to build trust, effectively achieving it in practice can be complicated. Another example shows that, at the micro level, the ability to explain decisions significantly influences trust. However, findings at the meso level reveal that civil servants often lack a clear understanding of what algorithms are or how to communicate about them. Thus, while algorithms may be listed in a register, without a proper explanation, they fail to enhance trust.

6.2 Key messages

The answer to the central research question leads to the following three key messages:

- Transparency can serve a **communicative function**: external communication about algorithms, tailored to the targeted audience, can directly enhance citizen trust by explaining how algorithms work and how they are integrated within organizational and institutional frameworks.
- Transparency can serve a **disciplining function**: the need to provide transparency encourages public organizations to critically evaluate and refine their algorithmic decision-making processes, leading to better-justified and embedded decisions, which may indirectly enhance public trust.
- The relationship between transparency and trust occurs not only within individual levels but also through an **interconnectedness** across the micro, meso and macro levels. For instance, macro-level transparency impacts meso-level trust in algorithm use by governments.

6.3 Scientific contributions

In academic and societal debates, there seems to be an overly optimistic view that making government algorithms transparent will automatically lead to trust (Floridi, 2020). This belief is rooted in the broader discourse on transparency, where advocates of algorithmic transparency assume that once the inner workings of algorithms are made visible and understandable, public skepticism and mistrust will decrease or even disappear. This optimism is further fueled by a growing demand for transparency as public value and therefore goal in itself, especially in response to concerns about algorithmic bias, fairness and accountability (Diakopoulos, 2020). However, this assumption tends to overlook certain complexities: transparency may overwhelm or confuse citizens with technical details or even backfire if transparency exposes flaws without sufficient mitigation efforts. This dissertation challenges and refines the naive, optimistic view, offering a more nuanced understanding of algorithmic transparency, by investigating how transparency works across three levels: algorithmic, organizational and institutional. These insights contribute to two significant bodies of literature: the literature on government transparency and the literature on algorithmization.

Contributions to the government transparency literature

The body of literature on government transparency is extensive.[20] The insights from this dissertation contribute to two prominent questions that often recur in transparency literature: How to provide transparency and what does it achieve? The first contribution highlights that the act of providing meaningful transparency can be very challenging, requiring efforts at multiple levels. The second contribution underscores that transparency does not always lead to trust, as its effects depend on the type of the information shared.

Algorithms add another layer of complexity to government transparency

In general government transparency literature, research often focuses on providing citizens with information about government decision-making, policies and policy outcomes (Grimmelikhuijsen, 2012b). For instance, Grimmelikhuijsen (2012a) examines how citizens are presented with either a negative performance outcome (highly polluted air) or a positive performance outcome (clean air) and explores the effects of these types of transparency on citizen trust. While such research is very important for understanding general transparency-trust dynamics, it does not fully address the complexities of today's algorithmic decision-making. For example, in the context of air pollution, algorithms might predict pollution levels, allocate resources, or develop policies based on complex data. Unlike traditional transparency, which offers clear outcomes like air quality metrics, algorithmic transparency involves revealing the (often) sophisticated processes, such as data sources, modeling techniques and decision criteria. The insights from this dissertation demonstrate that providing algorithmic transparency is highly complex across all three analytical levels.

At the micro level, we have seen that algorithms can be very complex. As the level of complexity increases, it becomes more difficult to trace how an algorithmic decision or recommendation was made. Previous research indicates

[20] For a comprehensive overview of the literature on government transparency, see Porumbescu et al. (2022).

that more complex algorithms, such as machine learning algorithms, are often harder to explain than simple rule-based systems (Katzenbach & Ulbricht, 2019; Wang et al., 2023), though not impossible (Coglianese & Lehr, 2019; Giest & Grimmelikhuijsen, 2020). Chapter 3 focuses on micro-level transparency, where explanations are provided for a rule-based algorithmic recommender system. While in this case it is clear how a particular recommendation is derived, explaining this in an understandable and comprehensible manner is not simple. For more complex algorithms, however, it is more complicated to trace how the algorithm arrived at a particular outcome, even though there are increasingly technical solutions (e.g., XAI) available for this (Coglianese & Lehr, 2019).

At the meso level, the findings from the dissertation have shown that determining the extent to which algorithmic outcomes are definitive and how civil servants interpret and use these results is challenging to articulate (see also Schuilenburg & Peeters, 2024). The situation rarely presents a simple dichotomy of "humans in the loop" versus "humans out of the loop" (Katzenbach & Ulbricht, 2019). For example, the outcomes of the same predictive recommendation algorithm might be used by one municipality as just one of many factors in deciding whether to investigate welfare fraud, while another municipality may rely strongly on the algorithm's recommendations, with minimal deviation. This variation can be even more pronounced among individual civil servants. Some find it challenging to deviate from an algorithmic recommendation (even when it's only advisory), while others prefer to trust their own judgment and disregard the algorithm's suggestions, or merely follow algorithmic recommendations that confirm their own judgement (e.g., Selten et al., 2023). This makes it extremely difficult to clearly communicate how algorithms are being applied by civil servants. In addition, Chapter 4 highlights that mapping the organizational embedding of an algorithm and communicating it in a way that is understandable to outsiders is extremely time-consuming and costly. Organizations invest considerable resources in accurately documenting these processes, which also necessitates making strategic political decisions.

Finally, at the macro level, what complicates matters further is that providing insights into algorithmic decision-making processes and the organizational and institutional embedding of algorithms can invite scrutiny from external parties, potentially resulting in negative media coverage, as illustrated in Chapter 4. This stakeholder accountability, or media accountability, trough transparency can, in turn, work as a "fire alarm" for vertical accountability fora such as the Parliament (Meijer, 2014; Schillemans, 2008, 2011). Moreover, providing transparency regarding the technological, organizational and institutional integration of algorithms is a relatively recent development, presenting a collective-action problem. As demonstrated in Chapter 4, organizations that take the lead in being transparent about their algorithm usage often face heightened scrutiny from journalists. According to Camilleri et al. (2023), negative media attention may create a reluctance to be transparent about algorithms in the future, leading to a vicious cycle that erodes the opportunities for external scrutiny.

The relationship between algorithmic transparency and trust is complex

The three empirical studies in this dissertation demonstrate that providing transparency is important to build trust. However, the relationship between transparency and trust in algorithm use by governments is not straightforward. The findings from this dissertation advance the general government transparency literature by empirically demonstrating that the relationship between transparency and trust in governmental algorithm use is complex, showing that not all types or amounts of transparency contribute equally to trust and that transparency strategies can sometimes undermine trust, especially when perceived as selective or incomplete.

First of all, not all types of transparency contribute (equally) to building trust. For example, Chapters 3 and 5 show that transparency can indeed increase public trust, though not all forms of transparency have the same impact. The post-hoc analyses in Chapter 3 show that offering a rationale element (i.e., explaining why a particular action is recommended) appears more effective for strengthening trust than a procedural explanation, in cases when no directive explanation is given. Chapter 5 further demonstrates that information about monitoring the police's compliance with legislation on algorithm use is more effective than merely providing information about the legislation itself. These findings resonate with Aoki et al. (2024) who examine other types of explanations for algorithmic decisions (see Section 6.5) and find that not all explanations are equally effective and that the effects might depend on the specific decision-context. Additionally, the effect sizes are larger at the micro level (Chapter 3) than at the macro level (Chapter 5), indicating that the effects of transparency on trust are stronger at the micro level than at the macro level. This may imply that the greater the impact on an individual (such as a decision that directly affects them) the stronger the effects of transparency on trust.

Secondly, Chapter 4 shows that if recipients feel that critical information is being withheld in the transparency provided to them, this transparency can potentially undermine trust. This finding adds to the notion of "openwashing", which describes a mismatch in information that the public expects to be shared and the information that organizations actually make available to the public (Heimstädt, 2017). Heimstädt (2017, p. 77) finds three approaches how organizations strategically respond to transparency expectations: they select the disclosed information to exclude parts of the data or parts of the audience; they bend the information in order to retain some control over its representative value; and they orchestrate new information for a particular audience. The findings of this dissertation add to the discussion on openwashing, by showing how public organizations select, bend and orchestrate information in the process of algorithm registration, and by highlighting what the implications of these strategies for information recipients can be. Recipients may become suspicious about why the government is withholding certain information, which could reinforce their beliefs that the government is using harmful or biased algorithms. This way, transparency could lead to increased skepticism, thereby having negative implications for public trust.

Thirdly, the findings in this dissertation illustrate that providing more information is not always better. This highlights the nuanced balance between the benefits of transparency, described by Heald (2003) as the "value of sunlight" and the potential drawbacks, referred to as the "danger of over-exposure". For instance, Chapter 3 reveals that combining two types of explanations does not necessarily lead to more trust than providing either explanation individually. Chapter 5 paints a similar picture: offering information about both legislation regarding police's use of algorithms, and monitoring police's compliance thereof does not foster greater trust than providing just one type of transparency. These findings are in line with (conceptual) arguments about full transparency not being realistic as information will always be missing (Fung et al., 2007) and that the complexity of algorithms rarely allows for complete disclosure (Glikson & Woolley, 2020). This highlights the significance of this dissertation, which has taken important steps toward empirically examining which types of transparency effectively increase citizen trust in algorithm use by governments. It underscores the importance of nuance: not all types and amounts of transparency foster trust.

Contributions to the algorithmization literature

The body of literature on algorithmization is emerging.[21] The insights from this dissertation contribute to two key themes in the literature on algorithmization: how algorithmic accountability can be strengthened, and how to balance openness and confidentiality in algorithmic transparency. The first contribution highlights the need for meaningful transparency in algorithmic governance to avoid risks such as "decoupling" and "ethics theater" where accountability and ethical practices are overshadowed by superficial compliance. The second contribution demonstrates that while transparency about specific algorithms is important for

[21] For a detailed discussion on the concept of algorithmization, refer to Meijer and Grimmelikhuijsen (2020) and Chapter 2 of this dissertation.

building trust, operational practices may limit substantive disclosures. In these cases, procedural transparency could be used to maintain trust.

From box-ticking to trustworthy algorithm use: The role of meaningful transparency and continuous monitoring

As discussed earlier in this chapter, transparency can serve a disciplining function when the requirement to provide transparency, such as that mandated by the AI Act, becomes a meaningful box-ticking exercise for public organizations. This occurs, for example, when organizations critically evaluate their algorithmic decision-making processes as they are required to register their algorithms. However, there is also a risk that transparency requirements become a meaningless box-ticking exercise, which introduces two significant challenges.

Firstly, when algorithmic transparency becomes merely a box-ticking exercise (without making genuinely meaningful contributions to organizational practices) it risks leading to "decoupling", where actual ethical implementation and accountability are disconnected from the visible external reporting (Meijer, 2014; Metcalf et al., 2021; Power, 1997). In other words, while formal transparency requirements may technically be met, genuine accountability practices may lag behind. A similar risk is what Kitchin (2021) describes as "numbers theater" and what Cath and Jansen (2022) refer to as "ethics theater". In both cases, it involves the strategic selection of figures, data and ethical considerations presented to create an image of a responsible, efficient and accurate government—without necessarily reflecting actual practice. This can create the illusion of responsible algorithm use, appearing to meet requirements while potentially ignoring the core ethical challenges and failing to provide meaningful transparency.

Secondly, the box-ticking exercise loses its potential when it is not properly monitored and when there are no sanctions for organizations that fail to complete all the required boxes or update them when the algorithm changes. As Chapter 4 shows, if transparency is not meaningful, it may undermine trust. Chapter 5 further highlights that continuous oversight is essential for ensuring trustworthy algorithm use. This is in line with conceptual research on algorithmization that emphasizes the importance of monitoring of algorithms for responsible and trustworthy use of algorithms by public organizations (Busuioc, 2020; Grimmelikhuijsen & Meijer, 2022; Meijer & Grimmelikhuijsen, 2020). This dissertation adds to this discussion that ongoing evaluations and monitoring of algorithms and their organizational embedding are necessary for *meaningful* transparency. This includes a baseline of information, such as completing the mandatory categories in an algorithm register and ensuring that this information remains accurate and up to date.

Balancing transparency and confidentiality through procedural transparency

At the micro level, insights from this dissertation demonstrate that transparency about specific algorithms is not only possible but *essential* for building public trust. Providing a rationale explanation (i.e., the reasoning behind a specific decision, including the factors or criteria that influenced it) has been shown to be highly effective in building trust. However, this can be challenging for the police and other public organizations. At times, sharing detailed reasons behind decisions is simply not feasible, as such disclosures could undermine operational security, compromise ongoing investigations, or expose sensitive information (Wessels, 2024). These constraints make it difficult to offer substantive transparency without jeopardizing the investigation or the safety of involved parties. This tension between the need for openness and the necessity of confidentiality demands careful consideration and strategies that foster trust without undermining the core missions of the police and other organizations.

The dissertation highlights procedural transparency as a valuable alternative: providing insight into the decision-making process rather than revealing specific outcomes. By showing the steps and safeguards built into the algorithmic process, the public can gain an understanding of how decisions are made without revealing the exact criteria or details of each outcome. This approach prevents manipulation of the system through the exploitation of known criteria ("gaming the system") and can build trust in the fairness of the decision-making process without compromising, for instance, the integrity or safety of a police investigation. The criteria used and the rationale behind algorithmic decisions can be monitored by oversight bodies, without being externally published, ensuring accountability while protecting sensitive information. Needless to say, this approach should be applied with restraint. Otherwise, there is a risk that, under the pretext of security, this reasoning may be used to withhold substantive information about algorithms even in cases where disclosure is feasible.

6.4 Methodological reflections

This dissertation has many strong methodological aspects. For example, I dedicated substantial time and attention to *manipulation checks*. In the two survey experiments (Chapters 3 and 5), I rigorously pre-tested the experimental manipulations in pilot studies, aligning with recommendations from the scientific literature (Kane & Barabas, 2019). Based on the pilot study results, I adjusted the manipulations where necessary to ensure they more accurately reflected the intended manipulation goals. Additionally, I *preregistered* the hypothesis-testing studies in the spirit of Open Science. Preregistration not only enhances the transparency and rigor of the research process but also strengthens the credibility and replicability of a study's findings (Simmons et al., 2021).

Nonetheless, there are several methodological reflections and limitations that provide input for future research. A first reflection pertains to design trade-offs of *survey experiments*. The three survey experiments in Chapters 3 and 5 were carefully designed to ensure mundane realism and construct validity. All experimental scenarios were developed in consultation with the police to ensure they were both relevant and aligned with the actual functioning of the algorithms in use. While this approach enhances realism, it can come at the expense of the "cleanliness" of the experimental scenarios. The inclusion of context-specific information may have introduced noise into the experimental treatments, which could reduce their precision. Although I acknowledge this limitation, survey experiments are often criticized for lacking realism, as treatments are typically presented in "clean" survey environments that fail to capture real-world effects (Barabas & Jerit, 2010; Gaines et al., 2007). Thus, designing a survey experiment inevitably involves trade-offs. While the inclusion of realistic scenarios might introduce some noise, the insights from these studies better reflect the real-world application of police algorithms.

A second reflection concerns the way the main concepts in this study were *operationalized and measured*. Although I lay the foundations in Chapter 2 for investigating the relationship between macro transparency and macro trust, I did not actually examine this relationship. Instead, Chapter 5 focuses on meso trust in the study on macro transparency, as this better suited the daily work practices of the police, as well as the goals and design of the research. As a result, the intended relationship between macro transparency and trust was not empirically measured. The suggestions provided in Chapter 2 for this relationship can be further developed in future research. Additionally, instead of examining various degrees of transparency (Levy et al., 2021), I focused on the absence or presence of specific transparency mechanisms. Chapters 3 and 5 tested micro and macro transparency across different dimensions of transparency, each involving different types of information. An exception to this is the combined dimensions, where in Chapter 3, procedural and rationale explanations are combined into a third type of explanation, and in Chapter 5, both transparency about legislation and about monitoring compliance with this legislation are provided. These combined dimensions can be seen as representing a greater degree of transparency than the single dimensions. The insights from Chapter 4 highlight that providing transparency is highly complex and that, in practice, it is never a binary choice of full transparency versus none (absence vs. presence); rather, it is always a matter of degree. Future research could therefore approach and explore transparency as a continuum, examining varying degrees and combinations of transparency mechanisms to better reflect the complexities of real-world practices.

A third reflection concerns the *generalizability* of the findings. The participants in the survey experiments (Chapters 3 and 5) and the respondents in the interview study (Chapter 4) were all from the Netherlands: a western, educated, industrialized, rich and democratic (WEIRD) country. There has been increasing attention to the limitations of "WEIRD" research, which refers to studies conducted with participants from these populations and highlights how these samples often do not reflect global diversity in human behavior (Henrich et al., 2010). Like much research on the fairness, accountability and transparency of AI systems, this dissertation risks introducing potential biases by not accounting for the cultural norms, characteristics and expectations of much of the world's population (Septiandri et al., 2023). Additionally, the survey experiments in this dissertation rely on sample pools. Although these sample pools include real citizens, they may inadvertently exclude individuals who are less comfortable with online surveys or who have concerns about privacy, potentially leading to an underrepresentation of those who are less familiar with or less trusting of digital platforms. This limitation could result in a bias toward participants who are more digitally engaged, thereby impacting the representativeness of the findings (Bethlehem, 2010). Future research should thus aim to include a broader range of cultural contexts and reach participants who may be less digitally engaged or have privacy concerns. This approach would enhance the representativeness of findings and provide a more globally inclusive understanding of the relationship between algorithmic transparency and citizen trust.

6.5 Limitations and agenda for future research

Below, I discuss three limitations that lead to directions for future research, which pertain to the need for further exploration of the substantive findings, the need for empirical testing of proposed relationships between levels and the need for further research into other types of algorithmic transparency.

Exploration of substantive findings: Contexts, moderators and algorithm types

Firstly, as mentioned in the methodological reflections section, the research presented in this dissertation has been conducted in the Netherlands, an institutional environment where citizens generally hold a high level of trust in their government. Most of the chapters focus on the Dutch National Police, with a specific emphasis on policy contexts such as fraud management and predictive policing. An exception to this is Chapter 4, in which I examine the Dutch government more broadly, addressing various policy contexts in which algorithms are used. However, previous studies have indicated that the effects of transparency on trust may vary depending on the *institutional or policy context* in which the research is conducted (de Fine Licht, 2014; Meijer, 2013). Emerging research also suggests that public perceptions of the trustworthiness, fairness and accuracy of (government's use of) algorithms may depend on the specific context in which the algorithm operates (Aoki et al., 2024; Giest & Grimmelikhuijsen, 2020; Schuilenburg & Peeters, 2024; Wang et al., 2023). It would therefore be valuable for future research to examine how different institutional and policy contexts affect the dynamics between transparency and trust in the governmental use of algorithms. For instance, examining institutional environments where public trust in government is relatively low, such as the United States (Pew Research Center, 2024), could provide insights into how reduced public trust interacts with the implementation of algorithmic transparency. Furthermore, investigating other policy contexts within the police domain—such as crowd control management—or within other government sectors—such as social welfare eligibility or immigration and asylum processes—could provide valuable insights into the robustness of the effects of algorithmic transparency on public trust across various policy contexts. Building on this, in contexts where algorithms are used to identify or prosecute specific individuals, and where the police's monopoly on violence plays a significant role, the impact of algorithmic decisions on citizens can be profound (Brayne, 2020). Transparency in these contexts is likely to be even more critical for the individuals affected by such decisions. However, it also poses challenges for police organizations, such as the previously mentioned risk of "gaming the system". It would therefore be highly relevant for future research to explore the role of algorithmic transparency in these sensitive contexts.

Secondly, this dissertation mainly examines the broader, population-wide effects of algorithmic transparency practices. While this approach provides valuable insights, it primarily focuses on general trends and outcomes, leaving certain *moderators* unexamined or not systematically explored. Research suggests that individual characteristics—such as age, gender, personality traits and perceptions of whether the use of the technology may help or hinder the government to achieve goals that are particularly important to individuals—may also influence citizens' acceptance of AI (Horvath et al., 2023, p. 11). Furthermore, studies show that the relationship between government transparency and trust is significantly moderated by individuals' preexisting attitudes toward transparency (Ripamonti, 2024). Future research should therefore aim to develop a more structured and theoretical approach to investigating these interaction effects. In particular, it would be relevant to systematically measure and analyze the potential moderating roles of individual-level characteristics in the relationship between transparency and trust in governmental algorithm use.

Thirdly, this dissertation examines particular types of transparency that are not universally applicable to or relevant for all *algorithm types*. In terms of complexity, this dissertation does not cover highly complex algorithms, such as neural networks or reinforcement learning. This choice affects the degree to which we can trace how inputs are transformed into outputs, as prior research suggests that complex algorithms—such as those based on machine learning techniques—tend to be more challenging to interpret than simpler, rule-based algorithms (Katzenbach & Ulbricht, 2019; Wang et al., 2023). However, while challenging, interpretability is not impossible for these more complex systems (Coglianese & Lehr, 2019; Giest & Grimmelikhuijsen, 2020). In terms of the extent to which an algorithm is outcome determinative, this dissertation focuses solely on algorithmic recommendations, such as those provided by the Intelligent Crime Reporting Tool and the Crime Anticipation System rather than on automated decision making algorithms where there is no possibility to deviate from the algorithmic output. In the case of the Intelligent Crime Reporting Tool, the citizen is free to disregard the algorithmic advice, just as police employees can with the Crime Anticipation System. Previous research shows that public trust in AI systems varies across types of AI and is influenced by the system's level of machine intelligence and representation (Glikson & Woolley, 2020). Future research could extend the findings of this dissertation by examining the applicability of the different transparency mechanisms to more complex and outcome-determinative algorithms, providing a broader understanding of transparency in diverse algorithmic contexts.

Empirical testing of proposed relations between levels

In this dissertation I primarily focused on the relationship between transparency and trust within each individual level, with the exception of Chapter 5, where I link macro-level transparency to meso-level trust. This chapter demonstrates that macro-level transparency, such as providing information about the institutional embedding of algorithms, impacts trust at the meso level, specifically regarding how the police use algorithms. Previous research also highlights the importance of examining the interconnectedness between levels (Criado et al., 2024; Giest & Grimmelikhuijsen, 2020; Porumbescu et al., 2022). In the section on scientific contributions, I discussed potential interconnections based on my findings, but future research should empirically test these connections more thoroughly. For example, future research could explore whether explaining specific algorithmic decisions (micro-level transparency) impacts institutional (macro-level) trust, or whether providing information about how algorithms are embedded in an organization (meso-level transparency) influences micro-level trust in specific algorithmic outputs.

Researching other types of algorithmic transparency

In this dissertation, I focused on several types of transparency. However, there are many other types of algorithmic transparency that I did not investigate, but which could be interesting for future research. A concrete example of this is that Chapter 3 focuses on three types of explanations for algorithmic recommendations, while recent research shows that other types of explanations are relevant, too (e.g., Aoki et al., 2024; Ripamonti, 2024). For instance, Aoki et al. (2024) distinguish between input-based, group-based, case-based and counterfactual explanations, each of which highlights different aspects of transparency. These four types of explanations are more focused on the underlying data and logic of the algorithm: input-based explanations clarify the impact of individual inputs on the output, group-based explanations compare outcomes for similar demographic groups, case-based explanations provide comparisons to past similar cases, and counterfactual explanations highlight the minimal changes needed in inputs to achieve a different outcome (Aoki et al., 2024).

The explanations examined in Chapter 3 (procedural, rationale and combined explanations) differ from the input-based, group-based, case-based and counterfactual explanations in their focus. Procedural explanations provide information about the process leading to a recommendation, focusing on the steps taken in the algorithmic decision-making process, whereas rationale explanations explain the reasoning behind a specific decision, such as the factors or criteria that influenced it. Notably, rationale explanations closely resemble input-based explanations, as both emphasize the role of individual inputs in influencing the decision. However, while rationale explanations frame this influence in terms of the reasoning or criteria behind the decision, input-based explanations focus more directly on quantifying the specific impact of inputs on the output. Combined explanations include both the procedural details and the rationale for a decision, offering a more comprehensive account.

These different explanation types serve complementary roles: procedural, rationale and combined explanations emphasize process and reasoning, whereas input-based, group-based, case-based and counterfactual explanations focus on data-driven aspects of algorithmic decision-making. Future research could explore, for example, how these different types of explanations interact and their relative effectiveness in fostering trust and understanding.

Another example is from Chapter 4, where I examined an algorithm register as a meso-level algorithmic transparency practice. While this register could be a relevant mechanism for communicating transparency about the organizational embedding of algorithms, other transparency options could also be considered. For instance, a publicly accessible dashboard or a citizen forum could play a role in providing insights into the use of algorithms by governments. These mechanisms would not only contribute to public understanding of algorithmic systems but could also increase engagement and accountability by creating opportunities for dialogue and scrutiny from diverse stakeholders. Future research could explore multiple types of algorithmic transparency simultaneously by investigating the information needs of individuals—what they *want* or *prefer*—rather than focusing on testing only one or a few types of transparency. This could be achieved through methods such as conjoint experiments or surveys, which allow for a good understanding of how different transparency mechanisms are valued by the public.

6.6 Recommendations for practice

Algorithms are playing an increasingly prominent role in public service delivery, a trend that is unlikely to be reversed. The genie is out of the bottle, so the question here is: how can we use algorithms in a trustworthy manner? This dissertation highlights that not all forms of transparency inherently foster greater trust, and achieving effective transparency can often be challenging. Yet, understanding what *does* work is very important and relevant for practice. In the following sections, I present a series of actionable recommendations tailored to enhance algorithmic transparency across micro, meso and macro levels.

Micro level: Transparency-by-design and user focused explanations

To ensure micro level transparency, I recommend that *public organizations* proactively consider how to explain the output of algorithms and engage in discussions on this topic, which highlights the communicative function of algorithmic transparency. This proactive approach aligns with transparency-by-design principles (Felzmann et al., 2020), where transparency is embedded in the algorithm's design. This practice encourages organizations to be mindful of their decisions (= increasing fishbowl transparency) and to clarify why these decisions are made (= increasing reasoned transparency), fostering more careful decision-making and more conscientious use of algorithms (Coglianese & Lehr, 2019). In addition, they should proactively publish the information about the workings of an algorithm rather than only disclose this in response to a FOI request. For algorithms already in use, it is also important to consider how to make their operations understandable to the citizens affected by their decisions. While embedding transparency from the outset is ideal, many algorithms are already active within public organizations. For these existing systems, it is also important that public organizations ensure that the impacted citizens gain insight into the underlying reasoning behind algorithmic decisions and the structure of the decision-making process. The insights from Chapter 3 (see Section 3.5) offer several concrete measures to best explain algorithmic outputs, such as algorithmic recommendations. This may require adjusting procurement regulations if algorithms are owned by private parties (Coglianese & Lehr, 2019).

Meso level: Targeted transparency strategies and algorithmic accountability

At the meso level, I advise *public organizations* to align algorithmic transparency initiatives with the specific needs of the audience, as they ultimately determine how transparency is perceived and valued (Ripamonti, 2024; van Vliet et al., 2024). This calls for tailored transparency: initiatives designed to meet the distinct expectations of different audience groups. Attempting to create a "jack of all trades" transparency approach—one that tries to satisfy every possible demand—can backfire, resulting in a mismatch between expectations from the transparency receivers and what is realistically provided. Such a misalignment can undermine the effectiveness of the communicative function of transparency efforts. Furthermore, for audiences that are especially critical of algorithmic decision-making, transparency is particularly valuable as it empowers individuals to understand and, if necessary, challenge the decisions made by algorithms. Without sufficient transparency, public trust in algorithmic processes may erode, ultimately jeopardizing the broader societal acceptance of these technologies. To address these diverse needs effectively, public organizations should create targeted transparency strategies. Transparency, therefore, is not a one-size-fits-all solution.

Building upon this, I recommend that *policy makers* implementing algorithms in public services pay close attention to establishing a clearly defined accountability forum, as this contributes to the disciplining function of algorithmic transparency. Currently, public organizations focus their algorithmic transparency efforts primarily on citizens. However, as Meijer (2014, p. 521) aptly observes: "The assumption that citizens find government so interesting and important that they want to scrutinize its records simply does not hold true. At the same time, transparency does open up the playing field of accountability by giving more prominent roles to the media and stakeholders". This suggests that algorithmic transparency strategies should consider the important roles played by media and other stakeholders in fostering accountability. These accountability fora are essential for ensuring that algorithmic transparency efforts—such as algorithmic explanations or algorithm registers—are not merely symbolic but genuinely functional (Wieringa, 2020). Without such structures, there is a significant risk of undermining public trust rather than strengthening it (Metcalf et al., 2021).

Finally, I advise *politicians* to actively make use of the meso-level transparency measures provided by public organizations. For example, local council members could make more frequent use of algorithm registers to pose critical questions. They can hold municipalities accountable when information about a high-risk algorithm is missing or incomplete. Such scrutiny can create a disciplining effect, because the organization has to critically re-examine the algorithm and investigate why the information is absent. This process can lead to follow-up actions, which can improve algorithmic decision-making practices , as previously discussed in this chapter.

Macro level: Continuous oversight and explicitly defined transparency obligations

I recommend that *regulators and oversight authorities* adopt a proactive approach to ensure meaningful algorithmic transparency. The findings of this dissertation at the macro level highlight the necessity of continuous oversight to prevent transparency from becoming a mere checkbox exercise. Without incentives to be complete or to keep information up-to-date, especially given the dynamic nature of algorithmic decision-making processes, transparency risks losing its significance and undermining public trust. Two risks might emerge here: first, that algorithmic transparency is reduced to a superficial compliance task, rather than being seen as a richer process that can enhance algorithmic decision-making; and second, that public organizations fail to apply even the checkbox approach properly, due to the absence of sanctions. Ongoing, iterative evaluations and monitoring of algorithms and their organizational embedding are essential to mitigate these risks and ensure the disciplining function of algorithmic transparency (see also Grimmelikhuijsen & Meijer, 2022; Levy et al., 2021; Schuilenburg & Peeters, 2024).

This aligns with McCubbins and Schwartz's (1984) distinction between "police patrol" and "fire alarm" accountability. A "police patrol" model, characterized by continuous and proactive monitoring, is particularly well-suited to algorithmic transparency. It ensures that standards are consistently upheld and allows for regular adjustments, fostering trust and accountability more effectively than the reactive "fire alarm" approach. In this case, forced transparency is crucial for accountability to succeed (see also Meijer, 2014).

How to do this effectively is a distinct research direction and falls largely outside the scope of this dissertation. However, one recommendation that follows from the findings of this dissertation is the need for clear and well-defined frameworks specifying the required types of transparency. Without such frameworks, the risk of incomplete or meaningless transparency increases greatly. Consequently, regulators and oversight authorities must define clear criteria and indicators for algorithmic transparency against which public organizations are evaluated. As Fung et al. (2007, p. i) note: “To be successful, transparency policies must be accurate, must keep ahead of disclosers’ efforts to find loopholes, and, above all, must focus on the needs of ordinary citizens”. Such an approach ensures that transparency initiatives are meaningful and aligned with public interests, reducing the likelihood of superficial compliance or obfuscation.

Furthermore, the research by Lorenz (2024, pp. 182-185) provides several relevant recommendations on how to regulate and oversee the use of algorithms by public organizations. He emphasizes, among others, that political and societal actors need to adopt new perspectives that account for the complex institutional dynamics and organizational processes shaping algorithmization outcomes, in order to oversee the use of algorithms effectively. This means that oversight must go beyond focusing on algorithms alone and extend to all involved actors, organizational policies, and the broader institutional context in which algorithmization occurs. This aligns with the multi-level transparency framework, which underscores the importance of disclosing not only information about algorithms themselves but also their organizational and institutional context.

Finally, the findings highlight that *public organizations* must actively work to raise citizen awareness of the oversight mechanisms in place. This dissertation demonstrates that when citizens are aware that governments are being monitored for their compliance with rules and regulations governing algorithm use, it leads to an increase in public trust. Here, the communicative and disciplining functions of transparency converge. Monitoring transparency tools that provide insight into the organizational embedding of algorithms can strengthen the disciplining function. Meanwhile, informing citizens that public organizations are being monitored supports the communicative function. Both approaches can contribute to enhancing public trust in government use of algorithms.

6.7 Sketching the horizon of algorithmization

Looking ahead, the use of algorithms in government processes is expected to grow, not only in the Netherlands (Algemene Rekenkamer, 2024), but also worldwide (Rashid & Kausik, 2024). This algorithmic way of governance, in which supra-organizational data flows dominate, is also referred to as data bureaucracy (Schuilenburg & Peeters, 2024). This growing reliance on algorithmic systems highlights the need for safeguards at multiple levels (macro, meso and micro) to ensure that these systems are both transparent and trustworthy. Without strong institutional safeguards, such as comprehensive legal frameworks, independent oversight bodies and mechanisms to ensure accountability in the development and use of algorithms, it will be difficult to achieve meaningful algorithmic transparency at the organizational and technological levels. In this context, democracy and democratic decision-making are indispensable in shaping and sustaining these safeguards. However, in many countries, democracy is increasingly under threat, facing challenges such as rising authoritarianism, eroding trust in public institutions and the spread of misinformation (Repucci & Slipowitz, 2021).

In an increasingly authoritarian regime, for example, it is possible that insights into how explaining algorithmic decisions can enhance citizen trust might also be exploited for manipulation (see also Yang & Roberts, 2023; Pearson, 2024). The findings presented in this dissertation on how trust in the use of algorithms by the government can be strengthened could, therefore, be misused. This use of the communicative route of algorithmic transparency might, in the short term, foster greater (albeit misplaced) trust. However, in such an authoritarian regime, meaningful checks and balances are rare, preventing the disciplining function of algorithmic transparency from being effectively utilized.

Another possible path forward is the rise of a technocracy, in which the so-called "coding elite"—a predominantly white, male and heterosexual group of data professionals who make critical decisions in the design and development of algorithms—could evade both internal oversight and public accountability (Schuilenburg, 2024, pp. 120-121). This presents a risk to the disciplining function of algorithmic transparency, as these data professionals might act without acknowledging their own privileged position or the potential negative consequences of the algorithms they design and implement (Schuilenburg & Peeters, 2024). Furthermore, there is a growing risk that the algorithms they develop, will become increasingly complex due to the intra-organizational nature of the data, rendering them almost impossible to explain to the public, which could obstruct the communicative function of algorithmic transparency.

These trends can undermine efforts to establish strong governance structures for the use of algorithms by governments, making it even more important to advocate for democratic resilience. Therefore, ensuring transparency is crucial for preserving the integrity of democratic systems in an increasingly algorithm-driven world. Nevertheless, it is important to acknowledge that within the data bureaucracy, controlling and overseeing these supra-organizational data flows and (self-learning) algorithms will become increasingly challenging. As a result, it is crucial to focus on designing *inclusive* algorithms, ensuring that algorithms are developed with careful consideration of the public values at stake and a deliberate balancing of conflicting or incompatible values (Schuilenburg & Peeters, 2024; see also Fest et al., 2023). Additionally, when designing algorithmic systems for the public sector, it is essential to carefully consider how to explain the use of the algorithm before its implementation, aligning with the earlier discussed principles of transparency-by-design (Felzmann et al., 2020), where transparency is integrated into the algorithm's design. If algorithmic decisions cannot be explained to citizens, due to factors such as self-learning algorithms or supra-organizational data flows, a critical discussion must occur regarding the *desirability* and ethical and practical implications of using such algorithms. Building on this, ensuring transparency can not only foster greater public trust in the use of algorithms by governments through the communicative, direct route, but also promotes more responsible algorithmic practices by holding systems accountable through the indirect, disciplining effects of algorithmic transparency. Efforts to ensure algorithmic transparency at all levels will be crucial for preserving the integrity of democracy in a world of algorithms.

7

Epilogue: Impact portfolio

In this epilogue, I would like to go beyond the set of papers and chapters that I wrote and present in this thesis, and share with you some of the impact highlights of my PhD trajectory. Alongside academic contributions and recommendations for practice, I have made a number of practical contributions during the process of writing this dissertation, focusing on academic outreach aimed at the police, government organizations and broader public discussions. Before detailing these *contributions*, I will first reflect on my *role as a researcher*, to situate my position within my research project and discuss how my interactions with practitioners both influenced my research and shaped the impact of my findings. After that, I reflect on my lessons learned while trying to make impact in *doing a PhD*.

7.1 Role as a researcher

I conducted the research in this dissertation as part of the Algorithmic Policing research project (ALGOPOL).[22] Through collaboration with the police within this research project, I was able to gain access to the police. This allowed me put my research to better use (Orr & Bennett, 2012). I viewed my role within this context as a "critical friend" where, based on insights from literature and my empirical research, I encouraged them to reconsider certain transparency practices concerning their use of algorithms. This position allowed me to both offer support and provide constructive feedback. Furthermore, the nature of the research project enabled me to be a member of the National Policelab AI, where I attended monthly meetings with lab members. During these meetings, members of the lab presented their research and shared important insights. This ensured that I was in regular contact with individuals working on developing responsible and trustworthy algorithms for the police.

However, I also experienced a situation in which the police tried to involve me more deeply in a transparency project by inviting me to become a member of their project team and to participate in inter-organizational meetings involving ministries and other executive agencies related to this project. This blurred the boundary between researcher and practitioner, rendering my role more ambiguous (Orr & Bennett, 2012). After consulting with colleagues at the university, I decided to maintain a greater degree of distance and refrain from participating in these meetings as if I were a police staff member. This decision allowed me to better fulfill my role as a critical friend, ensuring that I could provide support and constructive feedback without compromising my objectivity or independence.

[22] See Section 1.4 for a detailed description of the Algorithmic Policing research project.

The funding from the Dutch Research Council further contributed to my independent position as a researcher. The agreements made as part of this funding emphasize the freedom to shape my scientific research independently. This has allowed me to formulate my own research questions and shape my study in a way that I considered most appropriate for the issue at hand. The police, or other organizations, had no influence on this process, except for providing inspiration through existing challenges within the organization related to the use of algorithms.

As a researcher, a very important goal for me in reporting my research findings is to contribute to society.[23] This goal can be achieved by making my knowledge public and allowing others to use it, or by directly contributing to political or social debates and agenda-setting (Van Thiel, 2014, p. 158). This approach closely resembles action research, but it differs in that, as a researcher, I do not play an active role within the situation being studied during the empirical research phase (Van Thiel, 2014, p. 17). Through my engagement with practitioners regarding my research findings, I could make practical contributions in various areas, as I explain in detail below. At the same time, this interaction enriched my own knowledge and ideas about transparency and trust in government algorithms. Additionally, since developments in the field of algorithms are advancing so quickly, it was important to embed my results in the organizations for which my findings are relevant in a timely manner. Therefore, my impact activities often took place shortly after the completion of the research.

7.2 Contributions to the police

The results of my research led me to rewrite all types of explanations for the police's Intelligent Crime Reporting Tool.[24] Chapter 3 highlights the importance of providing explanations for algorithmic recommendations made to citizens regarding whether or not they should report online fraud. The findings showed that offering a rationale element (i.e., explaining why a particular action is recommended) is more effective in strengthening trust than a procedural explanation, in cases when no directive explanation is given. After my research concluded, the police requested that I rewrite all decision tree outcomes in the algorithm to provide rationale-based explanations. This was an interesting experience and challenge for me, as it allowed me to gain firsthand experience in implementing my own recommendations within an organizational context.

[23] This goal also aligns with the global Open Science movement and the Recognition and Rewards movement.

[24] https://www.uu.nl/en/news/research-contributes-to-crime-reporting-tool-national-police.

While it is challenging to establish direct causality, the police noted a reduction in the number of citizens disregarding the tool's advice following these updates.

In addition to this work, I contributed through several presentations and meetings for police staff, aimed both at developers and at more strategic levels focused on AI strategy and governance. These sessions allowed me to pitch my research ideas, gather valuable feedback to align my work with their practices and challenges, and share findings so that the police could actively implement them. For example, I held multiple meetings with the project leader responsible for the police algorithm register at various stages of my research. In these sessions, I presented insights from both theoretical and empirical findings, discussing how they could be practically applied within the police context. I also provided input on specific, practical challenges the project leader faced in developing the algorithm register. Another example is that I presented all my three empirical papers to the members of the National Policelab AI. In addition to doing their PhD research, most of the members of this lab work for the police as data scientists, which provided a good opportunity to bridge the gap between my academic insights and on-the-ground applications of the police.

Finally, I co-authored two articles for *Cahiers Politiestudies*, a peer-reviewed journal for police professionals widely read among Dutch and Flemish police staff. In the first article, alongside two members of the National Police Lab AI, I illustrate how transparency at the micro, meso and macro levels can foster trust in autonomous AI systems within law enforcement (Testerink et al., 2023). In the second article, co-authored with two researchers from the Algorithmic Policing (ALGOPOL) research group, we demonstrate how explaining algorithmic recommendation system outcomes is important both for citizens using this system and for police officers processing reports submitted through it (Nieuwenhuizen et al., 2023). With these efforts I tried to bridge the gap between my research insights and real-world algorithmic applications, practices and challenges of the police.

7.3 Contributions to other government organizations

I have also shared insights from this dissertation with various government organizations on multiple occasions. For instance, I designed and conducted a workshop as part of a training program for employees from various governmental bodies focused on making algorithmic decisions more interpretable. In this session, participants experienced the impact of receiving different types of explanations and practiced techniques for enhancing the interpretability of algorithmic outcomes. They could then apply these insights within their own organizations.

Additionally, I presented my work at several conferences and workshops aimed at bridging the gap between science and practice. For example, I spoke at a conference on Governing AI, the Utrecht AI Labs Event and a workshop on algorithms and evidence in criminal proceedings, where I discussed the opportunities and challenges associated with algorithm registries.

Finally, I participated in multiple meetings organized by the Ministry of the Interior and Kingdom Relations regarding the development of a national algorithm register. These meetings, informed by input from public organizations and the wider community, focused on establishing guidelines for the register's structure and the type of information it should include. I contributed to these discussions based on my research findings and took part in a national event on algorithms ("*Expeditie Algoritmes*"), organized by the Ministry. This event brought together employees from diverse government organizations, including ministries, municipalities and executive agencies. In debates and workshops, I shared insights from my research with attendees and learned about the practical struggles they encountered regarding algorithmic transparency, which further expanded my own ideas on the topic.

7.4 Contributions to public debates

Finally, I have contributed to broader public discussions about transparency and trust in the use of algorithms by governments. I have given guest lectures in both primary and secondary schools, where I engaged with children about what algorithms are, how government organizations use them, and why it can be difficult to make algorithms transparent.[25] Inspired by Ziewitz's (2016) "Algorithmic Walk", I had the pupils create their own walking algorithms. In pairs, the students created two movement rules, such as "take two steps forward, then three steps to the right". They then executed their walking algorithms on the schoolyard, with one partner acting as the rule executor and the other as the rule overseer. The goal was to give high fives to five other pairs. Back in the classroom, the students shared their experiences. Some got stuck when they ran into a wall or a slide, leading them to sometimes sneak in a few extra steps to still give someone else a high five. They encountered, as Ziewitz (2016) shows in his experiment, circumstances and obstacles they did not expect, which caused them to either adapt the rules or circumvent them. Through this practical exercise, the students were able to experience that algorithms do not always work as intended in practice. Rules often take on different meanings in real-

[25] https://algopol.sites.uu.nl/2022/07/11/esther-vertelt-als-slimme-gast-over-haar-onderzoek-op-een-basisschool/.

world situations. People create rules, but they do not always interpret them the same way as others, or they encounter challenges in their work that lead them to adjust or bypass the rules to achieve the desired outcome. This showed in a very practical way how providing information about algorithms can be complicated.

In addition, I have participated in public debates about responsible algorithm use. For example, I was one of the key speakers at a political café in the municipality of Utrecht, where I engaged with the audience about the value of algorithm registers.[26] Furthermore, I was selected as one of the Faces of Science by the Royal Netherlands Academy of Arts and Sciences. As a Face of Science, I have made my research accessible to a wider audience in various ways. I have written several blogs on a popular platform[27], recorded a video explaining my research in simple terms[28], and appeared on national radio where I spoke about the necessity of transparent algorithm use by governments.[29]

7.5 *Doing* a PhD

Making an impact with my scientific findings was a source of great energy and fulfillment throughout my PhD journey. In line with the evolving perspective on recognition and rewards in academia, it is important to acknowledge that making impact may not hold the same significance for all PhD researchers as it did for me. Impact is only one of the five basic principles of the TRIPLE Model for recognizing and rewarding academic staff at Utrecht University (n.d.), and not every researcher is expected to excel in all these principles. For those PhD researchers who, like me, find motivation in making an impact with their work, I am happy to share my experiences of *doing* a PhD. Reflecting on my own journey, I have identified three lessons learned that I believe could be valuable for any PhD researcher aiming to make a tangible impact.

One of the most important lessons I learned was the value of *engaging with practitioners and stakeholders early* in the research process. While the academic community sometimes operates in silos, working directly with practitioners not only informed my research but also ensured that my findings were relevant and actionable for practice. I learned that feedback from the police, government organizations and other stakeholders helped refine my research and gave it

[26] https://www.youtube.com/live/i8jfm7gJp58?si=G4FXREDEPlaiWT5I&t=1064.

[27] https://www.nemokennislink.nl/facesofscience/profielen/esther-nieuwenhuizen/.

[28] https://youtu.be/fLw8Yx-cuhg?si=Z0nfwOMTzRzPFahE.

[29] https://www.nporadio1.nl/fragmenten/de-nacht-is-zwart/10c6f242-10e6-4ee7-9a15-1e9c223a2274/2-023-12-04-wetenschapper-esther-nieuwenhuizen-over-haar-onderzoek-naar-algoritmen.

practical utility. By incorporating real-world challenges into my academic work, I was able to strengthen the impact of my findings. For PhD researchers, I would recommend finding ways to involve practitioners throughout the research process—not just at the end. This iterative engagement can help to bridge the gap between theory and practice and makes sure that their work addresses the (most urgent) needs of those applying it.

Furthermore, I had the opportunity to work closely with the police during my research, which presented both valuable insights and some ethical dilemmas. I learned that while collaboration with practitioners can improve the relevance of your work, it is important to *maintain a critical distance*. At times, the line between researcher and practitioner can blur, and I had to make difficult decisions about where to draw that boundary to ensure my objectivity and independence. By maintaining my position as a "critical friend" and resisting pressures to become fully involved in the practical aspects of the work, I believe that I was able to provide objective, evidence-based recommendations without compromising my integrity. I would advise PhD researchers to keep this balance in mind when trying to make an impact.

Finally, I learned that academic impact is not limited to publishing papers in journals. Engaging with a broader audience, including policymakers, practitioners and the public, is important for making impact in the "real world". I was fortunate to be able to share my findings in various settings: from workshops for government employees to public debates on responsible algorithm use. Through these activities, I saw firsthand how important it is to *proactively make your research accessible and relatable to those outside of academia*. Individuals and groups interested in your research may not always know you, but you can make an effort to become known by proactively approaching them. Writing for non-academic audiences, engaging in public speaking and contributing to policy discussions helped my research resonate with those who could benefit from it most. I would therefore recommend PhD researchers to adopt a proactive stance while exploring different avenues of disseminating their work, making it both accessible and impactful for a broader audience.

Reference list

Algemene Rekenkamer (2024). *Focus op AI bij de rijksoverheid*. https://www.rekenkamer.nl/publicaties/rapporten/2024/10/16/focus-op-ai-bij-de-rijksoverheid

Alikhademi, K., Drobina, E., Prioleau, D., Richardson, B., Purves, D., & Gilbert, J. E. (2022). A review of predictive policing from the perspective of fairness. *Artificial Intelligence and Law*, *30*(1), 1-17. https://doi.org/10.1007/s10506-021-09286-4

Alkemade, G., & Toet, J. (2021). Data protection regulation in the Netherlands. In E. Kiesow Cortez (Ed.), *Data Protection Around the World: Privacy Laws in Action* (pp. 165-188). T.M.C. Asser Press.

Alon-Barkat, S. (2020). Can government public communications elicit undue trust? Exploring the interaction between symbols and substantive information in communications. *Journal of Public Administration Research and Theory, 30*(1), 77-95. https://doi.org/10.1093/jopart/muz013

Alon-Barkat, S., & Busuioc, M. (2023). Human–AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory, 33*(1), 153-169. https://doi.org/10.1093/jopart/muac007

Amicelle, A. (2022). Big data surveillance across fields: Algorithmic governance for policing & regulation. *Big Data & Society*, *9*(2). https://doi.org/10.1177/20539517221112431

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973-989. https://doi.org/10.1177/1461444816676645

Andeweg, R. B., & Thomassen, J. J. a. (2005). Modes of political representation: toward a new typology. *Legislative Studies Quarterly*, *30*(4), 507-528. https://doi.org/10.3162/036298005X201653

Andrews, L. (2019). Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data'. *Public Administration, 97*(2), 296–310. https://doi.org/10.1111/padm.12534

Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*, *36*(2), 358-367. https://doi.org/10.1016/j.giq.2018.10.001

Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. *Government Information Quarterly*, *37*(4), 101490. https://doi.org/10.1016/j.giq.2020.101490

Aoki, N., Tatsumi, T., Naruse, G., & Maeda, K. (2024). Explainable AI for government: Does the type of explanation matter to the accuracy, fairness, and trustworthiness of an algorithmic decision as perceived by those who are affected? *Government Information Quarterly*, *41*(4), 101965. https://doi.org/10.1016/j.giq.2024.101965

Arrieta, A., Barredo, 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

Ashworth, R. E., McDermott, A. M., & Currie, G. (2019). Theorizing from qualitative research in public administration: Plurality through a combination of rigor and richness. *Journal of Public Administration Research and Theory*, *29*(2), 318-333. https://doi.org/10.1093/jopart/muy057

Azfar, O., & Nelson, W. R. (2007). Transparency, wages, and the separation of powers: An experimental analysis of corruption. *Public Choice*, *130*(3), 471-493. https://doi.org/10.1007/s11127-006-9101-5

Bakke, E. (2018). Predictive policing: The argument for public transparency. *NYU Annual Survey of American Law*, *74* (1), 131-171.

Barabas, J., & Jerit, J. (2010). Are survey experiments externally valid? *American Political Science Review*, *104*(2), 226-242. https://doi.org/10.1017/S0003055410000092

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*(3), 671-732. https://doi.org/10.15779/Z38BG31

Bauhr, M., & Grimes, M. (2012). What is government transparency? New measures and relevance for quality of government. *QOG the Quality of Government Institute, Working Paper Series,* 16, 1-27. https://gupea.ub.gu.se/handle/2077/38960

Bauhr, M., & Grimes, M. (2014). Indignation or resignation: The implications of transparency for societal accountability. *Governance*, *27*(2), 291-320. https://doi.org/10.1111/gove.12033

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bentzen, T. Ø., Six, F., & Beeck, S. O. de. (2025). Trust, control and motivation in public organizations. In F. Six, J. Hamm, D. Latusek, E. Zimmeren, van, & K. Verhoest (Eds.), *Handbook on Trust in Public Governance* (pp. 408 - 424). Edward Elgar Publishing.

Bertelli, A. M. (2006). Delegating to the quango: Ex ante and ex post ministerial constraints. *Governance*, *19*(2), 229-249. https://doi.org/10.1111/j.1468-0491.2006.00313.x

Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, *78*(2), 161-188. https://doi.org/10.1111/j.1751-5823.2010.00112.x

Biega, A. J., Potash, P., Daumé, H., Diaz, F., & Finck, M. (2020). Operationalizing the legal principle of data minimization for personalization. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, 399-408. https://doi.org/10.1145/3397271.3401034

Boeije, H., & Bleijenbergh, I. (2019). *Analyseren in kwalitatief onderzoek: Denken en doen.* Boom Onderwijs.

Bolman, L. G., & Deal, T. E. (2017). *Reframing organizations: Artistry, choice, and leadership.* John Wiley & Sons.

Borg, A., & Bex, F. (2021). Explaining arguments at the Dutch National Police. In V. Rodríguez-Doncel, M. Palmirani, M. Araszkiewicz, P. Casanovas, U. Pagallo, & G. Sartor (Eds.), *AI approaches to the complexity of legal systems XI-XII: AICOL international workshops 2018 and 2020: AICOL-XI@JURIX 2018, AICOL-XII@JURIX 2020, XAILA@JURIX 2020, revised selected papers* (pp. 183-197). (Lecture Notes in Computer Science; Vol. 13048). Springer. https://doi.org/10.1007/978-3-030-89811-3_13

Bottenbley, A. (2023). *The algorithm agenda: A discourse analysis of the Dutch algorithm register* [Master Thesis, Utrecht University]. UU Theses Repository. https://studenttheses.uu.nl/handle/20.500.12932/45374

Bouwmeester, M. (2023). System failure in the digital welfare state: Exploring parliamentary and judicial control in the Dutch childcare benefits scandal. *Recht Der Werkelijkheid*, *44*(2), 13-37. https://doi.org/10.5553/RdW/138064242023044002003

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

Bowen, G. A. (2020). *Sensitizing concepts.* SAGE Publications Limited.

Bowman, S., DeHaven, A., Errington, T., Hardwicke, T. E., Mellor, D. T., Nosek, B. A., & Soderberg, C. K. (2020). *OSF Prereg Template.* MetaArXiv. https://doi.org/10.31222/osf.io/epgjd

Brayne, S. (2020). *Predict and surveil: Data, discretion, and the future of policing.* Oxford University Press.

Brown, A., Chouldechova, A., Putnam-Hornstein, E., Tobin, A., & Vaithianathan, R. (2019). Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Article 41, pp. 1–12). ACM. https://doi.org/10.1145/3290605.3300271

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, *21*(2), 230-258. https://doi.org/10.1177/0049124192021002005

Bruckes, M., Grotenhermen, J. G., Cramer, F., & Schewe, G. (2019). Paving the way for the adoption of autonomous driving: Institution-based trust as a critical success factor. In *Proceedings of the 27th European Conference on Information Systems (ECIS)*, Stockholm & Uppsala, Sweden, June 8–14, 2019. Association for Information Systems. https://aisel.aisnet.org/ecis2019_rp/87

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951715622512

Burgess, M. (2018, May 4). Facial recognition tech used by UK police is making a ton of mistakes. *WIRED.* https://www.wired.co.uk/article/face-recognition-police-uk-south-wales-met-notting-hill-carnival

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, *81*(5), 825-836. https://doi.org/10.1111/puar.13293

Busuioc, M., Curtin, D., & Almada, M. (2022). Reclaiming transparency: Contesting the logics of secrecy within the AI Act. *European Law Open*, 1-27. https://doi.org/10.1017/elo.2022.47

Cámara, N. (2022, juni 3). *DiGiX 2020 update: A multidimensional index of digitization*. https://www.bbvaresearch.com/wp-content/uploads/2022/06/DiGiX_2022_Update_A_Multidimensional_Index_of_Digitization.pdf

Camilleri, H., Ashurst, C., Jaisankar, N., Weller, A., & Zilka, M. (2023). Media coverage of predictive policing: Bias, police engagement, and the future of transparency. In *EAAMO '23: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Article 28, pp. 1–19). ACM. https://doi.org/10.1145/3617694.3623249

Cath, C., & Jansen, F. (2022). Dutch comfort: The limits of AI governance through municipal registers. *Techné: Research in Philosophy and Technology*, *26*(3), 395-412. https://doi.org/10.5840/techne202323172

Chavali, D., Dhiman, V. K., & Katari, S. C. (2024). AI-powered virtual health assistants: Transforming patient engagement through virtual nursing. *International Journal of Pharmaceutical Sciences*, *2*(2), 613-624. https://doi.org/10.5281/zenodo.10691495

Chen, T., & Gasco-Hernandez, M. (2024). Uncovering the results of AI Chatbot use in the public sector: Evidence from US State Governments. *Public Performance & Management Review*, 1-26. https://doi.org/10.1080/15309576.2024.2389864

Cheng, H. F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F. M., & Zhu, H. (2019). Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. In *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Paper No. 559, pp. 1–12). ACM. https://doi.org/10.1145/3290605.3300789

Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable automated decision-making: A framework for accountable algorithmic systems. In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 598–609). ACM. https://doi.org/10.1145/3442188.3445921

Coglianese, C., & Lehr, D. (2019). Transparency and algorithmic governance. *Administrative Law Review*, *71*(1), 1–56. https://www.jstor.org/stable/27170531

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cook, K. S., Hardin, R., & Levi, M. (2005). *Cooperation without trust?* Russell Sage Foundation.

Cordella, A., & Gualdi, F. (2024). Algorithmic formalization: Impacts on administrative processes. *Public Administration*, 1-26. https://doi.org/10.1111/padm.13030

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). *Introduction to algorithms*. MIT press.

Costanzo, P., D'Onofrio, F., & Friedl, J. (2015). Big data and the Italian legal framework: Opportunities for police forces. In B. Akhgar, G. B. Saathoff, H. R. Arabnia, R. Hill, A. Staniforth, & P. S. Bayerl (Eds.), *Application of big data for national security* (pp. 238–249). Butterworth-Heinemann. https://doi.org/10.1016/B978-0-12-801967-2.00016-1

Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., Aroyo, L., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, *18*(5), 455-496. https://doi.org/10.1007/s11257-008-9051-3

Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review, 55*(1), 93-128. https://ssrn.com/abstract=2325784

Criado, J. I., Sandoval-Almazán, R., & Gil-Garcia, J. R. (2024). Artificial intelligence and public administration: Understanding actors, governance, and policy from micro, meso, and macro perspectives. *Public Policy and Administration, 0*(0). https://doi.org/10.1177/09520767241272921

Cross, C. (2018). Victims' motivations for reporting to the 'fraud justice network'. *Police Practice and Research*, *19*(6), 550-564. https://doi.org/10.1080/15614263.2018.1507891

Cucciniello, M., & Nasi, G. (2014). Transparency for trust in government: How effective is formal transparency? *International Journal of Public Administration*, *37*(13), 911-921. https://doi.org/10.1080/01900692.2014.949754

Datta, A., Sen, S., & Zick, Y. (2016). Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *2016 IEEE Symposium on Security and Privacy (SP)* (pp. 598–617). IEEE. https://doi.org/10.1109/SP.2016.42

de Boer, N., Eshuis, J., & Klijn, E.-H. (2018). Does disclosure of performance information influence street-level bureaucrats' enforcement style? *Public Administration Review*, *78*(5), 694-704. https://doi.org/10.1111/puar.12926

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

de Fine Licht, J. (2014). Policy area as a potential moderator of transparency effects: An experiment. *Public Administration Review*, *74*(3), 361-371. https://doi.org/10.1111/puar.12194

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI & SOCIETY*, *35*(4), 917-926. https://doi.org/10.1007/s00146-020-00960-w

de Hert, P., & Papakonstantinou, V. (2016). The new General Data Protection Regulation: Still a sound system for the protection of individuals? *Computer Law & Security Review*, *32*(2), 179-194. https://doi.org/10.1016/j.clsr.2016.02.006

de Jong, M. G., Pieters, R., & Fox, J.-P. (2010). Reducing social desirability bias through item randomized response: An application to measure underreported desires. *Journal of Marketing Research*, *47*(1), 14-27. https://doi.org/10.1509/jmkr.47.1.14

de Laat, P. B. (2018). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy & Technology*, *31*(4), 525-541. https://doi.org/10.1007/s13347-017-0293-z

de Meijer, J. (2023). *From explainability to trust: A conjoint analysis to explore governmental algorithm registers' positive and negative effects on citizens' trust in government decisions* [Master Thesis, TU Delft]. Repository TU Delft. https://repository.tudelft.nl/islandora/object/uuid%3A04e819fb-e993-445a-84e5-451448f3c7c8 den Hengst, M. & Wijsman, O.L. (2023). Datagedreven politiewerk: Een organisatorisch en juridisch perspectief. In T. Snaphaan, W. Hardyns, A. J. van Dijk, R. Spithoven & R. Van Brakel (Eds.), *Big data policing* (pp. 71–90). Gompel&Svacina.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, *59*(2), 56-62. https://doi.org/10.1145/2844110

Diakopoulos, N. (2020). Accountability, transparency, and algorithms. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 197–213). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.11

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, *5*(7), 809-828. https://doi.org/10.1080/21670811.2016.1208053

Dietz, G. (2011). Going back to the source: Why do people trust each other? *Journal of Trust Research*, *1*(2), 215-222. https://doi.org/10.1080/21515581.2011.603514

Dietz, G., & Den Hartog, D. N. (2006). Measuring trust inside organisations. *Personnel Review*, *35*(5), 557-588. https://doi.org/10.1108/00483480610682299

Difonzo, N., Hantula, D. A., & Bordia, P. (1998). Microworlds for experimental research: Having your (control and collection) cake, and realism too. *Behavior Research Methods, Instruments, & Computers*, *30*(2), 278-286. https://doi.org/10.3758/BF03200656

Diphoorn, T. (2013). The emotionality of participation: Various modes of participation in ethnographic fieldwork on private policing in Durban, South Africa. *Journal of Contemporary Ethnography*, *42*(2), 201-225. https://doi.org/10.1177/0891241612452140

Dowding, K., & John, P. (2008). The three exit, three voice and loyalty framework: A test with survey data on local services. *Political Studies*, *56*(2), 288-311. https://doi.org/10.1111/j.1467-9248.2007.00688.x

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New public management is dead—Long live digital-era governance. *Journal of Public Administration Research and Theory*, *16*(3), 467-494.https://doi.org/10.1093/jopart/mui057

Ejelöv, E., & Luke, T. J. (2020). “Rarely safe to assume”: Evaluating the use and interpretation of manipulation checks in experimental social psychology. *Journal of Experimental Social Psychology*, *87*, 103937. https://doi.org/10.1016/j.jesp.2019.103937

Enqvist, L. (2023). ‘Human oversight’ in the EU Artificial Intelligence Act: What, when and by whom? *Law, Innovation and Technology*, *15*(2), 508-535. https://doi.org/10.1080/17579961.2023.2245683

Etzioni, A., & Etzioni, O. (2017). Incorporating ethics into artificial intelligence. *The Journal of Ethics*, *21*, 403-418. https://doi.org/10.1007/s10892-017-9252-2

European Commission (2024). *Eurobarometer 92.4 (2019)* [Data file Version 2.0.0]. GESIS. https://doi.org/10.4232/1.14381

European Court of Auditors (2024). *EU ambition on artificial intelligence: Time to speed up* (Special Report No. 08/2024). European Court of Auditors. https://www.eca.europa.eu/nl/publications/sr-2024-08

Fatima, S., Desouza, K. C., & Dawson, G. S. (2020). National strategic artificial intelligence plans: A multi-dimensional analysis. *Economic Analysis and Policy*, *67*, 178-194. https://doi.org/10.1016/j.eap.2020.07.008

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics*, *26*(6), 3333-3361. https://doi.org/10.1007/s11948-020-00276-4

Ferguson, A. G. (2017). Policing predictive policing. *Washington University Law Review, 94(*5), 1109-1190. https://ssrn.com/abstract=2765525

Fest, I., Schäfer, M., van Dijck, J., & Meijer, A. (2023). Understanding data professionals in the police: a qualitative study of system-level bureaucrats. *Public Management Review, 25*(9), 1664–1684. https://doi.org/10.1080/14719037.2023.2222734

Floridi, L. (2020). Artificial intelligence as a public service: Learning from Amsterdam and Helsinki. *Philosophy & Technology*, *33*(4), 541-546. https://doi.org/10.1007/s13347-020-00434-3

Floridi, L. (2021). The European legislation on AI: A brief analysis of its philosophical approach. *Philosophy & Technology, 34*, 215–222. https://doi.org/10.1007/s13347-021-00460-9

Fox, J. (2007). Government transparency and policymaking. *Public Choice*, *131*, 23-44. https://doi.org/10.1007/s11127-006-9103-3

Fung, A., Graham, M., & Weil, D. (2007). *Full disclosure: The perils and promise of transparency*. Cambridge University Press.

Gaines, B. J., Kuklinski, J. H., & Quirk, P. J. (2007). The logic of the survey experiment reexamined. *Political Analysis*, *15*(1), 1-20. https://doi.org/10.1093/pan/mpl008

Giest, S., & Grimmelikhuijsen, S. (2020). Introduction to special issue algorithmic transparency in government: Towards a multi-level perspective. *Information Polity*, *25*(4), 409-417. https://doi.org/10.3233/IP-200010

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627-660. https://doi.org/10.5465/annals.2018.0057

Grimes, M. (2006). Organizing consent: The role of procedural fairness in political trust and compliance. *European Journal of Political Research*, *45*(2), 285-315. https://doi.org/10.1111/j.1475-6765.2006.00299.x

Grimmelikhuijsen, S. (2012a). Linking transparency, knowledge and citizen trust in government: An experiment. *International Review of Administrative Sciences, 78(1)*, 50-73. https://doi.org/10.1177/0020852311429667.

Grimmelikhuijsen, S. (2012b). *Transparency and trust. An experimental study of online disclosure and trust in government* (Doctoral dissertation, Utrecht University). Utrecht University Repository. https://dspace.library.uu.nl/handle/1874/218113

Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, *83*(2), 241-262. https://doi.org/10.1111/puar.13483

Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral public administration: Combining insights from public administration and psychology. *Public Administration Review*, *77*(1), 45-56. https://doi.org/10.1111/puar.12609

Grimmelikhuijsen, S., & Knies, E. (2017). Validating a scale for citizen trust in government organizations. *International Review of Administrative Sciences*, *83*(3), 583-601. https://doi.org/10.1177/0020852315585950

Grimmelikhuijsen, S., & Meijer, A. (2014). Effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment. *Journal of Public Administration Research and Theory*, *24*(1), 137-157. https://doi.org/10.1093/jopart/mus048

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

Grimmelikhuijsen, S., Porumbescu, G., Hong, B., & Im, T. (2013). The effect of transparency on trust in government: A cross-national comparative experiment. *Public Administration Review*, *73*(4), 575-586. https://doi.org/10.1111/puar.12047

Grimmelikhuijsen, S., de Vries, F., & Bouwman, R. (2024). Regulators as guardians of trust? The contingent and modest positive effect of targeted transparency on citizen trust in regulated sectors. *Journal of Public Administration Research and Theory, 34*(1), 136–149. https://doi.org/10.1093/jopart/muad010

Gritsenko, D., & Wood, M. (2022). Algorithmic governance: A modes of governance approach. *Regulation & Governance*, *16*(1), 45-62. https://doi.org/10.1111/rego.12367

Gulati, S., Sousa, S., & Lamas, D. (2019). Design, development and evaluation of a human-computer trust scale. *Behaviour & Information Technology*, *38*(10), 1004-1015. https://doi.org/10.1080/0144929X.2019.1656779

Haataja, M., van de Fliert, L., & Rautio, P. (2020). *Public AI Registers. Realising AI transparency and civic participation in government use of AI.* https://algoritmeregister.amsterdam.nl/wp-content/uploads/White-Paper.pdf

Hadwick, D., & Lan, S. (2021). Lessons to be learned from the Dutch childcare allowance scandal: A comparative review of algorithmic governance by tax administrations in the Netherlands, France, and Germany. *World Tax Journal, 13*(4), 609–645. https://doi.org/10.59403/27410pa

Halachmi, A., & Greiling, D. (2013). Transparency, e-government, and accountability: Some issues and considerations. *Public Performance & Management Review*, *36*(4), 572-584. https://doi.org/10.2753/PMR1530-9576360404

Harish, V., Samson, T. G., Diemert, L., Tuite, A., Mamdani, M., et al. (2022). Governing partnerships with technology companies as part of the COVID-19 response in Canada: A qualitative case study. *PLOS Digital Health, 1*(12), e0000164. https://doi.org/10.1371/journal.pdig.0000164

Haverland, M., & Yanow, D. (2012). A hitchhiker's guide to the public administration research universe: Surviving conversations on methodologies and methods. *Public Administration Review*, *72*(3), 401-408. https://doi.org/10.1111/j.1540-6210.2011.02524.x

Heald, D. A. (2006). Varieties of transparency. In C. Hood & D. Heald (Eds.), *Transparency: The key to better governance?* (Proceedings of the British Academy 135, pp. 25–43). Oxford University Press. http://ukcatalogue.oup.com/product/9780197263839.do

Heimstädt, M. (2017). Openwashing: A decoupling perspective on organizational transparency. *Technological Forecasting and Social Change, 125,* 77–86. https://doi.org/10.1016/j.techfore.2017.03.037

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hind, M. (2019). Explaining explainable AI. *XRDS, 25*(3), 16–19. https://doi.org/10.1145/3313096

Hirsch Ballin, M., & Oerlemans, J.-J. (2023). Datagedreven opsporing verzet de bakens in het toezicht op strafvorderlijk optreden. *Delikt en Delinkwent*, *2023*(1), 18-38. https://new.navigator.nl/document/id489ed3b4a24a4fe2a2f9f501714fa5d3?ctx=WKNL_CSL_32&tab=tekst

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states* (Vol. 25). Harvard University Press.

Hirschman, A. O. (1993). Exit, voice, and the fate of the German Democratic Republic: An essay in conceptual history. *World Politics*, *45*(2), 173-202. https://doi.org/10.2307/2950657

Hobson, Z., Yesberg, J. A., Bradford, B., & Jackson, J. (2023). Artificial fairness? Trust in algorithmic police decision-making. *Journal of Experimental Criminology*, *19*(1), 165-189. https://doi.org/10.1007/s11292-021-09484-9

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407-434. https://doi.org/10.1177/0018720814547570

Holmberg, S., Rothstein, B., & Nasiritousi, N. (2009). Quality of government: What you get. *Annual Review of Political Science, 12,* 135–161. https://doi.org/10.1146/annurev-polisci-100608-104510

Horvath, L., James, O., Banducci, S., & Beduschi, A. (2023). Citizens' acceptance of artificial intelligence in public services: Evidence from a conjoint experiment about processing permit applications. *Government Information Quarterly*, *40*(4), 101876. https://doi.org/10.1016/j.giq.2023.101876

Hosseini, M., Shahri, A., Phalp, K., & Ali, R. (2015). Towards engineering transparency as a requirement in socio-technical systems. In *2015 IEEE 23rd International Requirements Engineering Conference (RE)* (pp. 268–273). IEEE. https://doi.org/10.1109/RE.2015.7320435

Inayatullah, S. (2013). The futures of policing: Going beyond the thin blue line. *Futures*, *49*, 1-8. https://doi.org/10.1016/j.futures.2013.01.007

Ingrams, A., Kaufmann, W., & Jacobs, D. (2022). In AI we trust? Citizen perceptions of AI in government decision making. *Policy & Internet*, *14*(2), 390-409. https://doi.org/10.1002/poi3.276

Jackson, J., Bradford, B., Stanko, B., & Hohl, K. (2012). *Just authority?: Trust in the police in England and Wales* (1st ed.). Willan. https://doi.org/10.4324/9780203105610

Jangoan, S., Krishnamoorthy, G., Muthusubramanian, M., & Sharma, K. K. (2024). Demystifying explainable AI: Understanding, transparency, and trust. *International Journal For Multidisciplinary Research*, *6*(2), 1-13. https://doi.org/10.36948/ijfmr.2024.v06i02.14597

Jansen, F. (2022). *Data-driven policing: Negotiating the legitimacy of the police*. Cardiff University.

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

Janssen, M., & van den Hoven, J. (2015). Big and open linked data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly, 32*(4), 363-368. https://doi.org/10.1016/j.giq.2015.11.007

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2

Kaminski, M. E. (2020). Understanding transparency in algorithmic accountability. In W. Barfield (Ed.), *The Cambridge Handbook of the Law of Algorithms* (pp. 121-138). Cambridge University Press. https://doi.org/10.1017/9781108680844.006

Kane, J. V., & Barabas, J. (2019). No harm in checking: Using factual manipulation checks to assess attentiveness in experiments. *American Journal of Political Science*, *63*(1), 234-249. https://doi.org/10.1111/ajps.12396

Karadimitriou, A., von Krogh, T., Ruggiero, C., Biancalana, C., Bomba, M., & Lo, W. H. (2022). Investigative journalism and the watchdog role of news media: Between acute challenges and exceptional counterbalances. In *Success and failure in news media performance: Comparative analysis in the Media for Democracy Monitor 2021* (pp. 101–125). https://doi.org/10.48335/9789188855589-5

Katzenbach, C., & Ulbricht, L. (2019). Algorithmic governance. *Internet Policy Review*, *8*(4), 1-18. https://doi.org/10.14763/2019.4.1424

Kaur, M., Sharma, R., & Bansal, N. (2023). 24/7 healthcare services with AI-powered virtual assistants: An update. *Journal of Advanced Research in Embedded System, 10*(1), 6-9.

Kempeneer, S. (2021). A big data state of mind: Epistemological challenges to accountability and transparency in data-driven regulation. *Government Information Quarterly*, *38*(3), 101578. https://doi.org/10.1016/j.giq.2021.101578

Kempeneer, S., & Van Dooren, W. (2021). Using numbers that do not count: How the latent functions of performance indicators explain their success. *International Review of Administrative Sciences*, *87*(2), 364-379. https://doi.org/10.1177/0020852319857804

Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, *22*(14), 2081-2096. https://doi.org/10.1080/1369118X.2018.1477967

Kim, T. W., & Routledge, B. R. (2018). Informational privacy, a right to explanation, and interpretable AI. In *2018 IEEE Symposium on Privacy-Aware Computing (PAC)* (pp. 64-74). https://doi.org/10.1109/PAC.2018.00013

Kitchin, R. (2021). *Data lives: How data are made and shape our world*. Bristol University Press.

Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2390–2395). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858402

Klopfenstein, L. C., Delpriori, S., Malatini, S., & Bogliolo, A. (2017). The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (pp. 555-565). ACM. https://doi.org/10.1145/3064663.3064690

Konaté, S., & Pali, B. (2023). You have to talk with us, not about us": Exploring the harms of wrongful accusation on those affected in the case of the Dutch 'childcare-benefit scandal'. *Revista de Victimología/Journal of Victimology*, *16*, 139-164. https://hdl.handle.net/11245.1/04050d04-c728-42fe-b3c0-29c8a88d1e8f

Kramer, R. M., & Tyler, T. R. (1995). *Trust in organizations: Frontiers of theory and research*. Sage Publications.

Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, *165*(3), 633-706. https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3

Lacave, C., & Diez, F. J. (2004). A review of explanation methods for heuristic expert systems. *The Knowledge Engineering Review*, *19*(2), 133-146. https://doi.org/10.1017/S0269888904000190

Larsson, S. (2020). On the governance of artificial intelligence through ethics guidelines. *Asian Journal of Law and Society*, *7*(3), 437-451. https://doi.org/10.1017/als.2020.19

Laux, J. (2024). Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the European Union AI Act. *AI & Society, 39,* 2853–2866. https://doi.org/10.1007/s00146-023-01777-z

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 1-16. https://doi.org/10.1177/2053951718756684

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., & Kusbit, D. (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. In *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 182. https://doi.org/10.1145/3359284

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1603–1612). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702548

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy & Technology*, *31*(4), 611-627. https://doi.org/10.1007/s13347-017-0279-x

Levine, E. S., Tisch, J., Tasso, A., & Joy, M. (2017). The New York City police department's domain awareness system. *Interfaces*, *47*(1), 70-84. https://doi.org/10.1287/inte.2016.0860

Levy, K., Chasalow, K. E., & Riley, S. (2021). Algorithms and decision-making in the public sector. *Annual Review of Law and Social Science*, *17*, 309-334. https://doi.org/10.1146/annurev-lawsocsci-041221-023808

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. Springer Science & Business Media.

Lindgren, I., Madsen, C. Ø., Hofmann, S., & Melin, U. (2019). Close encounters of the digital kind: A research agenda for the digitalization of public services. *Government Information Quarterly*, *36*(3), 427-436. https://doi.org/10.1016/j.giq.2019.03.002

Loi, M., & Spielkamp, M. (2021). Towards accountability in the use of artificial intelligence for public administrations. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 757–766). Association for Computing Machinery. https://doi.org/10.1145/3461702.3462631

Lorenz, L. C. (2024). *The Myth of Algorithmic Regulation: An ethnographic exploration of algorithms, actors, and institutions* (Doctoral dissertation, Utrecht University). Utrecht University Repository. https://dspace.library.uu.nl/handle/1874/438148

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. In G. Gable & M. Vitale (Eds.), *Proceedings of the 11th Australasian Conference on Information Systems* (p. 53).

Makasi, T., Tate, M., Desouza, K. C., & Nili, A. (2021). Value-based guiding principles for managing cognitive computing systems in the public sector. *Public Performance & Management Review*, *44*(4), 929-959. https://doi.org/10.1080/15309576.2021.1879883

Mansbridge, J. (2009). A "selection model" of political representation. *Journal of Political Philosophy, 17*(4), 369–398. https://doi.org/10.1111/j.1467-9760.2009.00337.x

Mantelero, A. (2024). The fundamental rights impact assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. *Computer Law & Security Review, 54,* 106020. https://doi.org/10.1016/j.clsr.2024.106020

Matook, S., Brown, S. A., & Rolf, J. (2015). Forming an intention to act on recommendations given via online social networks. *European Journal of Information Systems*, *24*(1), 76-92. https://doi.org/10.1057/ejis.2013.28

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709-734. https://doi.org/10.5465/amr.1995.9508080335

McCubbins, M. D., & Schwartz, T. (1984). Congressional oversight overlooked: Police patrols versus fire alarms. *American Journal of Political Science, 28*(1), 165–179. https://doi.org/10.2307/2110792

McDonald, B. D. III, Hall, J. L., O'Flynn, J., & van Thiel, S. (2022). The future of public administration research: An editor's perspective. *Public Administration, 100*(1), 59–71. https://doi.org/10.1111/padm.12829

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), 1-25. https://doi.org/10.1145/1985347.1985353

McKnight, D. H., & Chervany, N. L. (2001). Trust and distrust definitions: One bite at a time. In R. Falcone, M. Singh, & Y.-H. Tan (Eds.), *Trust in cyber-societies* (Vol. 2246, pp. 27–54). Springer. https://doi.org/10.1007/3-540-45547-7_3

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334-359. https://doi.org/10.1287/isre.13.3.334.81

Mei, H., & Zheng, Y. (2024). How m-government services build relative trust? The mediating roles of value creation and risk perception. *Public Performance & Management Review*, *47*(6), 1327-1355. https://doi.org/10.1080/15309576.2024.2370935

Meijer, A. (2007). Publishing public performance results on the Internet: Do stakeholders use the Internet to hold Dutch public service organizations to account? *Government Information Quarterly*, *24*(1), 165-185. https://doi.org/10.1016/j.giq.2006.01.014

Meijer, A. (2013). Understanding the complex dynamics of transparency. *Public Administration Review*, *73*(3), 429-439. https://doi.org/10.1111/puar.12032

Meijer, A. (2014). Transparency. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford Handbook of Public Accountability* (pp. 507–524). Oxford University Press.

Meijer, A., & Grimmelikhuijsen, S. (2020). Responsible and accountable algorithmization: How to generate citizen trust in governmental usage of algorithms. In *The Algorithmic Society* (pp. 53-66). Routledge.

Meijer, A., Lorenz, L., & Wessels, M. (2021). Algorithmization of bureaucratic organizations: Using a practice lens to study how context shapes predictive policing systems. *Public Administration Review*, *81*(5), 837-846. https://doi.org/10.1111/puar.13391

Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration*, *42*(12), 1031-1039. https://doi.org/10.1080/01900692.2019.1575664

Metcalf, J., Moss, E., Watkins, E. A., Singh, R., & Elish, M. C. (2021). Algorithmic impact assessments and accountability: The co-construction of impacts. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 735–746). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445935

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1-38. https://doi.org/10.1016/j.artint.2018.07.007

Misztal, B. (2013). *Trust in modern societies: The search for the bases of social order.* John Wiley & Sons.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716679679

Moe, T. M. (2012). Delegation, control, and the study of public bureaucracy. *The Forum*, *10*(2). https://doi.org/10.1515/1540-8884.1508

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association, 110*(512), 1399–1411. https://doi.org/10.1080/01621459.2015.1077710

Moody, G. D., Galletta, D. F., & Lowry, P. B. (2014). When trust and distrust collide online: The engenderment and role of consumer ambivalence in online consumer behavior. *Electronic Commerce Research and Applications*, *13*(4), 266-282. https://doi.org/10.1016/j.elerap.2014.05.001

Moore, M. H. (1995). *Creating public value: Strategic management in government.* Harvard University Press.

Murad, M. (2021). *Beyond the "black box": Enabling meaningful transparency of algorithmic decision-making systems through public registers* [Master Thesis, Massachusetts Institute of Technology]. DSpace@MIT https://dspace.mit.edu/bitstream/handle/1721.1/139092/murad-mmurad-sm-idm-2021-thesis.pdf?sequence=1&isAllowed=y

Murphy, K. (2017). Challenging the 'invariance' thesis: Procedural justice policing and the moderating influence of trust on citizens' obligation to obey police. *Journal of Experimental Criminology*, *13*(3), 429-437. https://doi.org/10.1007/s11292-017-9298-y

Myeong, S., Kwon, Y., & Seo, H. (2014). Sustainable e-governance: The relationship among trust, digital divide, and e-government. *Sustainability, 6*(9), 6049-6069. https://doi.org/10.3390/su6096049

Nai, R., Meo, R., Morina, G., & Pasteris, P. (2023). Public tenders, complaints, machine learning and recommender systems: A case study in public administration. *Computer Law & Security Review*, *51*, 105887. https://doi.org/10.1016/j.clsr.2023.105887

National Police Lab AI (n.d.). *National Police Lab AI - AI Labs—Utrecht University*. https://www.uu.nl/en/research/ai-labs/our-labs/national-police-lab-ai

New, J., & Castro, D. (2018, May 21). *How policymakers can foster algorithmic accountability*. Center for Data Innovation. https://www2.datainnovation.org/2018-algorithmic-accountability.pdf

Nieuwenhuizen, E. (2025a). Algorithm registers: A box-ticking exercise or meaningful tool for transparency? *Information Polity, 29*(4), 415-433. https://doi.org/10.1177/15701255241297107

Nieuwenhuizen, E. (2025b). Trust and transparency in algorithmic governance. In F. Six, J. Hamm, D. Latusek, E. Zimmeren, van, & K. Verhoest (Eds.), *Handbook on Trust in Public Governance* (pp. 116 - 135). Edward Elgar Publishing.

Nieuwenhuizen, E. N., Meijer, A. J., Bex, F. J., & Grimmelikhuijsen, S. G. (2021, September 8). Citizen trust in algorithmic recommendations: Evidence from two survey experiments. Paper presented at the EGPA Conference, Brussels, Belgium.

Nieuwenhuizen, E. N., Meijer, A. J., Bex, F. J., & Grimmelikhuijsen, S. G. (2024). Explanations increase citizen trust in police algorithmic recommender systems: Findings from two experimental tests. *Public Performance & Management Review, 48*(3), 590–625. https://doi.org/10.1080/15309576.2024.2443140

Nieuwenhuizen, E., Soares, C., & Meijer, A. (2023). Strategisch perspectief op de dubbele digitale transformatie van de politie. Over de uitlegbaarheid van algoritmen voor politiemedewerkers en burgers. *Cahiers Politiestudies*, *3*(68), 29-43.

Nix, J., Wolfe, S. E., Rojek, J., & Kaminski, R. J. (2015). Trust in the police: The influence of procedural justice and perceived collective efficacy. *Crime & Delinquency*, *61*(4), 610-640. https://doi.org/10.1177/0011128714530548

Noordegraaf, M. (2008). Meanings of measurement. *Public Management Review*, *10*(2), 221-239. https://doi.org/10.1080/14719030801928672

Norris, P. (2022). *In praise of skepticism: Trust but verify*. Oxford University Press.

Nothdurft, F., Richter, F., & Minker, W. (2014). Probabilistic human-computer trust handling. In K. Georgila, M. Stone, H. Hastie, & A. Nenkova (Eds.), *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)* (pp. 51–59). Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-4307

Odekerken, D., Borg, A., & Bex, F. (2020). Efficient argument-based inquiry at the Dutch police. In *Applications of AI to Forensics 2020 (AI2Forensics 2020)* (pp. 22-23). https://research.nii.ac.jp/~ksatoh/AI2Forensics/AI2Forensics_proceedings.pdf#page=26

Odekerken, D., Bex, F., Borg, A., & Testerink, B. (2022). Approximating Stability for Applied Argument-based Inquiry. *Intelligent Systems with Applications,* 16, Article 200110. https://doi.org/10.1016/j.iswa.2022.200110

O'hara, K. (2004). *Trust: From Socrates to Spin*. Icon Books.

Ojo, A., Mellouli, S., & Ahmadi Zeleti, F. (2019). A realist perspective on AI-era public management. In *Proceedings of the 20th Annual International Conference on Digital Government Research* (pp. 159–170). Association for Computing Machinery. https://doi.org/10.1145/3325112.3325261

Orr, K., & Bennett, M. (2012). Public administration scholarship and the politics of coproducing academic–practitioner research. *Public Administration Review*, *72*(4), 487-495. https://doi.org/10.1111/j.1540-6210.2011.02522.x

Panigutti, C., Hamon, R., Hupont, I., Fernandez Llorca, D., Fano Yela, D., Junklewitz, H., Scalzo, S., Mazzini, G., Sanchez, I., Soler Garrido, J., & Gomez, E. (2023). The role of explainable AI in the context of the AI Act. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1139–1150). Association for Computing Machinery. https://doi.org/10.1145/3593013.3594069

Park, H., & Blenkinsopp, J. (2011). The roles of transparency and trust in the relationship between corruption and citizen satisfaction. *Active Learning in Higher Education, 77*(2), 77-88. https://doi.org/10.1177/1469787415616720

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Paul, R. (2024). European artificial intelligence "trusted throughout the world": Risk-based regulation and the fashioning of a competitive common AI market. *Regulation & Governance, 18*(4), 1065–1082. https://doi.org/10.1111/rego.12563.

Pearson, J. S. (2024). Defining digital authoritarianism. *Philosophy & Technology, 37*(2), 73. https://doi.org/10.1007/s13347-024-00754-8

Petty, R. E., & Briñol, P. (2012). The elaboration likelihood model. In P. A. M. Van Lange, A. Kruglanski, E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (Vol. 1, pp. 224-245). Sage. https://doi.org/10.4135/9781446249215.n12

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and persuasion* (pp. 1-24). Springer. https://doi.org/10.1007/978-1-4612-4964-1_1

Pew Research Center (2024, June 24). *Public trust in government: 1958-2024.* Pew Research Center. https://www.pewresearch.org/politics/2024/06/24/public-trust-in-government-1958-2024/

Politie (n.d.) *Keuzehulp internetoplichting.* Retrieved January 21, 2025, from https://aangifte.politie.nl/iaai-preintake/#/.

Politie (2022). *Datastrategie politie 2022-2025.* https://www.politie.nl/binaries/content/assets/politie/wet-open-overheid/00-landelijk/woo-documenten/voor-2024/it-systeem-politie/09---2022-10-19-datastrategie-politie-v1.0_geredigeerd.pdf

Porumbescu, G. A. (2015). Using transparency to enhance responsiveness and trust in local government: Can it work? *State and Local Government Review*, *47*(3), 205-213. https://doi.org/10.1177/0160323X15599427

Porumbescu, G., Meijer, A., & Grimmelikhuijsen, S. (2022). *Government transparency: State of the art and new perspectives.* Cambridge University Press. https://doi.org/10.1017/9781108678568

Porumbescu, G. A., Neshkova, M. I., & Huntoon, M. (2019). The effects of police performance on agency trustworthiness and citizen participation. *Public Management Review*, *21*(2), 212-237. https://doi.org/10.1080/14719037.2018.1473473

Power, M. (1997). *The audit society: Rituals of verification.* Oxford University Press.

Purves, D., & Davis, J. (2022). Public trust, institutional legitimacy, and the use of algorithms in criminal justice. *Public Affairs Quarterly*, *36*(2), 136-162. https://doi.org/10.5406/21520542.36.2.03

Quijano-Sánchez, L., Cantador, I., Cortés-Cediel, M. E., & Gil, O. (2020). Recommender systems for smart cities. *Information Systems*, *92*, 101545. https://doi.org/10.1016/j.is.2020.101545

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3173574.3173677

Raji, I. D., Xu, P., Honigsberg, C., & Ho, D. (2022). Outsider oversight: Designing a third-party audit ecosystem for AI governance. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society,* (pp. 557–571). Association for Computing Machinery. https://doi.org/10.1145/3514094.3534181

Rashid, A. B., & Kausik, M. A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*, *7*, 100277. https://doi.org/10.1016/j.hybadv.2024.100277

Rechtbank Den Haag (2020). ECLI:NL:RBDHA:2020:1878. https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBDHA:2020:1878

Repucci, S., & Slipowitz, A. (2021). *Democracy under siege*. Freedom House. https://freedomhouse.org/sites/default/files/2021-03/FIW2021_Abridged_03112021_FINAL.pdf

Richards, L. (2015). *Handling qualitative data: A practical guide* (3rd ed.). Sage.

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review Online*, *94*, 15-55. https://ssrn.com/abstract=3333423

Ripamonti, J. P. (2024). Does being informed about government transparency boost trust? Exploring an overlooked mechanism. *Government Information Quarterly*, *41*(3), 101960. https://doi.org/10.1016/j.giq.2024.101960

Roberts, A. (2020). Bridging levels of public administration: How macro shapes meso and micro. *Administration & Society*, *52*(4), 631-656. https://doi.org/10.1177/0095399719877160

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative research in psychology*, *11*(1), 25-41. https://doi.org/10.1080/14780887.2013.801543

Rotaru, V., Huang, Y., Li, T., Evans, J., & Chattopadhyay, I. (2022). Event-level prediction of urban crime reveals a signature of enforcement bias in US cities. *Nature Human Behaviour*, *6*(8), 1056-1068. https://doi.org/10.1038/s41562-022-01372-0

Rothstein, B. O., & Teorell, J. A. (2008). What is quality of government? A theory of impartial government institutions. *Governance*, *21*(2), 165-190. https://doi.org/10.1111/j.1468-0491.2008.00391.x

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *The Academy of Management Review*, *23*(3), 393-404. https://doi.org/10.5465/amr.1998.926617

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Rudin, C., Wang, C., & Coker, B. (2020). The age of secrecy and unfairness in recidivism prediction. *Harvard Data Science Review*, *2*(1), 1-53. https://doi.org/10.1162/99608f92.6ed64b30

Schaar, P. (2010). Privacy by design. *Identity in the Information Society*, *3*(2), 267-274. https://doi.org/10.1007/s12394-010-0055-x

Scharpf, F. W. (1997). Games real actors play: Actor-centered institutionalism in policy research (1st ed.). Routledge. https://doi.org/10.4324/9780429500275

Schiff, D. S., Schiff, K. J., & Pierson, P. (2022). Assessing public value failure in government adoption of artificial intelligence. *Public Administration*, *100*(3), 653-673. https://doi.org/10.1111/padm.12742

Schillemans, T. (2008). Accountability in the shadow of hierarchy: The horizontal accountability of agencies. *Public Organization Review, 8*(2), 175–194. https://doi.org/10.1007/s11115-008-0053-8

Schillemans, T. (2011). Does horizontal accountability work? Evaluating potential remedies for the accountability deficit of agencies. *Administration & Society, 43*(4), 387–416. https://doi.org/10.1177/0095399711412931

Schlehahn, E., Aichroth, P., Mann, S., Schreiner, R., Lang, U., Shepherd, I. D. H., & Wong, B. L. W. (2015). Benefits and pitfalls of predictive policing. In *Proceedings of the 2015 European Intelligence and Security Informatics Conference* (pp. 145-148). IEEE. https://doi.org/10.1109/EISIC.2015.29

Scholtes, H. H. M. (2012). *Transparantie, icoon van een dolende overheid* (Doctoral dissertation, Tilburg University). https://pure.uvt.nl/ws/portalfiles/portal/1440922/Scholtes_Transparantie_16-04-2012_emb_tot_16-10-2012.pdf

Schuilenburg, M. (2024). *Making surveillance public: Why you should be more woke about AI and algorithms.* Eleven.

Schuilenburg, M., & Peeters, R. (2024). Voorbij de system-level bureaucratie. *Beleid en Maatschappij, 51*(3), 278. https://doi.org/10.5553/BenM/138900692024051003006

Schuilenburg, M., & Soudijn, M. (2023). Big data policing: The use of big data and algorithms by the Netherlands Police. *Policing: A Journal of Policy and Practice, 17*, paad061. https://doi.org/10.1093/police/paad061

Scott, W. R., & Davis, G. F. (2015). *Organizations and organizing: Rational, natural and open systems perspectives* (1st ed.). Routledge.

Selten, F., Robeer, M., & Grimmelikhuijsen, S. (2023). "Just like I thought": Street-level bureaucrats trust AI recommendations if they confirm their professional judgment. *Public Administration Review, 83*(2), 263–278. https://doi.org/10.1111/puar.13602

Seppälä, A., Birkstedt, T., & Mäntymäki, M. (2021). From ethical AI principles to governed AI. In *Proceedings of the Forty-Second International Conference on Information Systems*, Austin 2021 (pp. 1-17).

Septiandri, A. A., Constantinides, M., Tahaei, M., & Quercia, D. (2023). WEIRD FAccTs: How Western, educated, industrialized, rich, and democratic is FAccT? In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 160–171). Association for Computing Machinery. https://doi.org/10.1145/3593013.3593985

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior, 98*, 277-284. https://doi.org/10.1016/j.chb.2019.04.019

Simmons, J., Nelson, L., & Simonsohn, U. (2021). Pre-registration: Why and how. *Journal of Consumer Psychology, 31*(1), 151–162. https://doi.org/10.1002/jcpy.1208

Singh, R., Dourish, P., Howe, P., Miller, T., Sonenberg, L., Velloso, E., & Vetere, F. (2021a). Directive explanations for actionable explainability in machine learning applications. *arXiv*. https://arxiv.org/abs/2102.02671

Singh, R., Ehsan, U., Cheong, M., Riedl, M. O., & Miller, T. (2021b). LEx: A framework for operationalising layers of machine learning explanations. *arXiv*. https://arxiv.org/abs/2104.09612

Skoog, G. E. (2005). *Supporting the development of institutions—Formal and informal rules: An evaluation theme: basic concepts*. Department for Evaluation and Internal Audit, Sida. https://cdn.sida.se/publications/files/sida23418en-supporting-the-development-of-institutions---formal-and-informal-rules-an-evaluation-theme-basic-concepts.pdf

Smets, A., Walravens, N., & Ballon, P. (2020). Designing recommender systems for the common good. In *Adjunct publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 276–278). Association for Computing Machinery. https://doi.org/10.1145/3386392.3399570

Smith, M. L. (2010). Building institutional trust through e-government trustworthiness cues. *Information Technology & People*, *23*(3), 222-246. https://doi.org/10.1108/09593841011069149

Sniderman, P. M. (2018). Some advances in the design of survey experiments. *Annual Review of Political Science*, *21*(1), 259-275. https://doi.org/10.1146/annurev-polisci-042716-115726

Soares, C., Grimmelikhuijsen, S., & Meijer, A. (2024). Screen-level bureaucrats in the age of algorithms: An ethnographic study of algorithmically supported public service workers in the Netherlands Police. *Information Polity*, *29*(3), 277-292. https://doi.org/10.3233/IP-220070

Strikwerda, L. (2021). Predictive policing: The risks associated with risk assessment. *The Police Journal*, *94*(3), 422-436. https://doi.org/10.1177/0032258X20947749

Tamò-Larrieux, A., Guitton, C., Mayer, S., & Lutz, C. (2024). Regulating for trust: Can law establish trust in artificial intelligence? *Regulation & Governance*, *18*(3), 780-801. https://doi.org/10.1111/rego.12568

Temme, M. (2017). Algorithms and transparency in view of the new General Data Protection Regulation. *European Data Protection Law Review (EDPL)*, *3*(4), 473-485. https://doi.org/10.21552/edpl/2017/4/9

Teo, T. (Ed.). (2013). *Handbook of quantitative methods for educational research*. Sense Publishers. https://doi.org/10.1007/978-94-6209-404-8

Testerink, B., Nieuwenhuizen, E. N., & Bex, F. J. (2023). Wat doet het er toe dat je mens bent? In T. Snaphaan, W. Hardyns, A. J. van Dijk, R. Spithoven & R. Van Brakel (Eds.), *Big data policing* (pp. 121-134). Gompel&Svacina.

Thomas, K. (2024). The advent of survey experiments in politics and international relations. *Government and Opposition, 59*(1), 297–320. https://doi.org/10.1017/gov.2022.36

Thompson, S. C., Sobolew-Shubin, A., Galbraith, M. E., Schwankovsky, L., & Cruzen, D. (1993). Maintaining perceptions of control: Finding perceived control in low-control circumstances. *Journal of Personality and Social Psychology*, *64*(2), 293. https://doi.org/10.1037/0022-3514.64.2.293

Tintarev, N., & Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems: Methodological issues and empirical studies on the impact of personalization. *User Modeling and User-Adapted Interaction*, *22*(4-5), 399-439. https://doi.org/10.1007/s11257-011-9117-5

Trägårdh, L., Witoszek, N., & Taylor, B. (2013). *Civil society in the age of monitory democracy*. Berghahn Books.

Tutt, A. (2017). An FDA for algorithms. *Administrative Law Review*, *69*(1), 83-124. https://www.jstor.org/stable/44648608

Tyler, T. R. (1990). *Why people obey the law: Procedural justice, legitimacy, and compliance.* Yale University Press.

Tyler, T. R. (2006). Restorative justice and procedural justice: Dealing with rule breaking. *Journal of Social Issues*, *62*(2), 307-326. https://doi.org/10.1111/j.1540-4560.2006.00452.x

Tyler, T. R., & Huo, Y. J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. Russell Sage Foundation.

Ulbricht, L., & Yeung, K. (2022). Algorithmic regulation: A maturing concept for investigating regulation of and through algorithms. *Regulation & Governance*, *16*(1), 3-22. https://doi.org/10.1111/rego.12437

Utrecht University. (n.d.). *UU vision: Recognition and rewards*. Utrecht University. https://www.uu.nl/sites/default/files/UU%20Vision%20Recognition%20and%20Rewards_2023.pdf

Valentinov, V., Verschraegen, G., & Van Assche, K. (2019). The limits of transparency: A systems theory view. *Systems Research and Behavioral Science*, *36*(3), 289-300. https://doi.org/10.1002/sres.2591

Van Thiel, S. (2014). *Research methods in public administration and public management: An introduction*. Routledge.

Van Vliet, M., Schuitemaker, N., Espana, S., Van de Weerd, I., & Brinkkemper, S. (2024). Defining and implementing algorithm registers: An organizational perspective. In *ECIS 2024 Proceedings*. https://aisel.aisnet.org/ecis2024/track04_impactai/track04_impactai/5

Veale, M., & Zuiderveen Borgesius, F. Z. (2021). Demystifying the draft EU Artificial Intelligence Act: Analyzing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97-112. https://doi.org/10.9785/cri-2021-220402

Verhoest, K., Verschuere, B., Peters, B. G., & Bouckaert, G. (2004). Controlling autonomous public agencies as an indicator of New Public Management. *Management International*, *9*(1), 25-35. https://api.semanticscholar.org/CorpusID:159795100

Vidotto, G., Massidda, D., Noventa, S., & Vicentini, M. (2012). Trusting beliefs: A functional measurement study. *Psicologica*, *33*(3), 575-590. https://psicologicajournal.com/trusting-beliefs-a-functional-measurement-study/

Wachter, S. (2024). Limitations and loopholes in the EU AI Act and AI liability directives: What this means for the European Union, the United States, and beyond. *Yale Journal of Law & Technology, 26*(3). https://doi.org/10.2139/ssrn.4924553

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. *Science Robotics*, *2*(6), eaan6080. https://doi.org/10.1126/scirobotics.aan6080

Wang, G., Guo, Y., Zhang, W., Xie, S., & Chen, Q. (2023). What type of algorithm is perceived as fairer and more acceptable? A comparative analysis of rule-driven versus data-driven algorithmic decision-making in public affairs. *Government Information Quarterly*, *40*(2), 101803. https://doi.org/10.1016/j.giq.2023.101803

Wang, H. (2022). Transparency as manipulation? Uncovering the disciplinary power of algorithmic transparency. *Philosophy & Technology*, *35*, 1-25. https://doi.org/10.1007/s13347-022-00564-w

Wang, Q., & Guan, Z. (2023). Can sunlight disperse mistrust? A meta-analysis of the effect of transparency on citizens' trust in government. *Journal of Public Administration Research and Theory*, *33*(3), 453-467. https://doi.org/10.1093/jopart/muac040

Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems, 23*(4), 217–246. https://doi.org/10.2753/MIS0742-1222230410

Wang, Y. D., & Emurian, H. H. (2005). An overview of online trust: Concepts, elements, and implications. *Computers in Human Behavior, 21*(1), 105-125. https://doi.org/10.1016/j.chb.2003.11.008

Warner, M. E., & Hefetz, A. (2008). Managing markets for public service: The role of mixed public–private delivery of city services. *Public Administration Review, 68*(1), 155–166. https://doi.org/10.1111/j.1540-6210.2007.00845.x

Warren, M. E. (1999). *Democracy and trust*. Cambridge University Press.

Watson, H. J., & Nations, C. (2019). Addressing the growing need for algorithmic transparency. *Communications of the Association for Information Systems, 45*, 488-510. https://doi.org/10.17705/1CAIS.04526

Welch, E. W., Hinnant, C. C., & Moon, M. J. (2005). Linking citizen satisfaction with e-government and trust in government. *Journal of public administration research and theory*, *15*(3), 371-391. https://doi.org/10.1093/jopart/mui021

Wenzelburger, G., König, P. D., Felfeli, J., & Achtziger, A. (2024). Algorithms in the public sector. Why context matters. *Public Administration*, *102*(1), 40-60. https://doi.org/10.1111/padm.12901

Wessels, M. (2024). Algorithmic policing accountability: Eight sociotechnical challenges. *Policing and Society*, *34*(3), 124-138. https://doi.org/10.1080/10439463.2023.2241965

Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1-18. https://doi.org/10.1145/3351095.3372833

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

Wörsdörfer, M. (2023). The E.U.'s Artificial Intelligence Act: An ordoliberal assessment. *AI and Ethics*. https://doi.org/10.1007/s43681-023-00337-x

WRR (2021). *Opgave ai. De nieuwe systeemtechnologie*. WRR-Rapport 105. WRR.

Yang, E., & Roberts, M. E. (2023). The Authoritarian Data Problem. *Journal of Democracy, 34(4), 141-150*. https://doi.org/10.1353/jod.2023.a907695

Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, *41*(1), 118-132. https://doi.org/10.1177/0162243915605575

Ziewitz, M. (2016). Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Values*, *41*(1), 3-16. https://doi.org/10.1177/0162243915608948

Zouridis, S., van Eck, M., & Bovens, M. (2020). Automated discretion. In T. Evans & P. Hupe (Eds.), *Discretion and the quest for controlled freedom* (pp. 307-326). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-19566-3_20

Zucker, L. G. (1986). Production of trust: Institutional sources of economic structure, 1840–1920. *Research in Organizational Behavior, 8*, 53–111.

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

Appendices

Appendices Chapter 3

Appendix 3A: Measures

Table 3.4 Items of trusting beliefs and intentions

Trusting beliefs	Trusting intentions
I believe that the recommendation by the intelligent crime reporting tool has all the characteristics I would expect from a recommendation about reporting online fraud.	I would feel comfortable acting upon the recommendation by the intelligent crime reporting tool.
If I read the recommendation by the intelligent crime reporting tool, I think I would be able to depend on it completely.	I would confidently act on the recommendation by the intelligent crime reporting tool.
I can always rely on the intelligent crime reporting tool for a recommendation about reporting online fraud.	I would not hesitate acting upon the recommendation by the intelligent crime reporting tool.
I can trust the recommendation by the intelligent crime reporting tool.	I accept the recommendation by the intelligent crime reporting tool.

Appendix 3B: Factual manipulation check

The reason that the intelligent crime reporting tool indicated for 'not reporting' was:

A. The intelligent crime reporting tool used an **algorithm** to analyze my story. This text analysis showed that the web shop www.sneakerwinkel.nl is **very unlikely to be fraudulent**.
B. The intelligent crime reporting tool mentioned that the web shop www.sneakerwinkel.nl was affiliated with the **quality mark** and has gone through an **extensive screening procedure**.
C. The intelligent crime reporting tool provided both reasons mentioned under A and B.
D. The intelligent crime reporting tool did not provide a clear explanation for its recommendation.

Appendix 3C: Potential fraud situation

Imagine the following situation:

Three weeks ago you bought new shoes via the web shop www.sneakerwinkel.nl. You have paid € 150,- for these shoes. The website indicated that the sneakers would be delivered to your house within 5 working days.

After three weeks, you still do not have your package and you did not receive a confirmation of shipment.

Because you think you have been scammed, you go to the website of the police to report a case of online fraud. To file a report, you must first fill in some information.

Appendix 3D: Mock version Intelligent Crime Reporting Tool

Explanation about the intelligent crime reporting tool

You want to report a case of online fraud.

The police wants to advise you on whether filing a report would be useful in your case. Therefore, you will go through the different steps of the intelligent crime reporting tool:

- You start by telling what happened.
- The intelligent crime reporting tool will look at your story and will ask some questions, if necessary.
- Finally, the intelligent crime reporting tool advises whether you should file a report of online fraud or if it would be better to take other steps.

Step 1: What happened?

Here you can tell what happened.

It is important that you are as complete as possible. When describing, consider:

- whether you paid for what you wanted to buy
- whether the seller delivered something
- how long you waited before filing a report
- at which (web) store you made a purchase

Describe in your own words what happened below.

Step 2: Findings of the intelligent crime reporting tool

Based on your input, the intelligent crime reporting tool came with the following findings. Tick the finding(s) that apply to your situation:
☐You have bought a product.
☐You waited a reasonable time before you wanted to file a report.
☐You were expecting a package.
☐A product has been delivered.

Step 3: Questions about what happened

Based on your story, the intelligent crime reporting tool still has a few questions.

Did you pay the seller for the product or make a deposit?
☐Yes
☐No

Did you order at a web shop?
☐Yes
☐No

Enter the name or URL (internet address) of the web shop here:

Appendix 3E: Background variables of samples

Table 3.5 Background variables of two experiments and the Dutch population

	Study 1 **(*n* = 717)**	**Study 2** **(*n* = 1005)**	**Dutch population** **(*N* = 17.505.934)**
Sex (female)	53.1%	52.3%	50%
Age (40 and higher)	60,6%	56.4%	53%
Education (bachelor's degree or higher)	42,1%	40.9%	41%

Balance checks for background variables Study 1

The proportion of female participants in Study 1 did not significantly differ from the proportion of females in the Dutch population (z = 1.68, p = .06). The proportion of participants of 40 years old and higher in Study 1 was significantly lower from the proportion of participants of 40 years old and higher in the Dutch population (z = 4.11, p < .01). The proportion of participants with a bachelor's degree or higher in Study 1 did not significantly differ from the proportion of people with a bachelor's degree or higher in the Dutch population (z = .60, *p* = .54). We carried out a sensitivity analysis to check whether the results for participants over 40 years old were different for participants under 40 years old. The sensitivity analysis showed no substantive alteration to our results (i.e., no significant effect became nonsignificant, and no nonsignificant effect became significant).

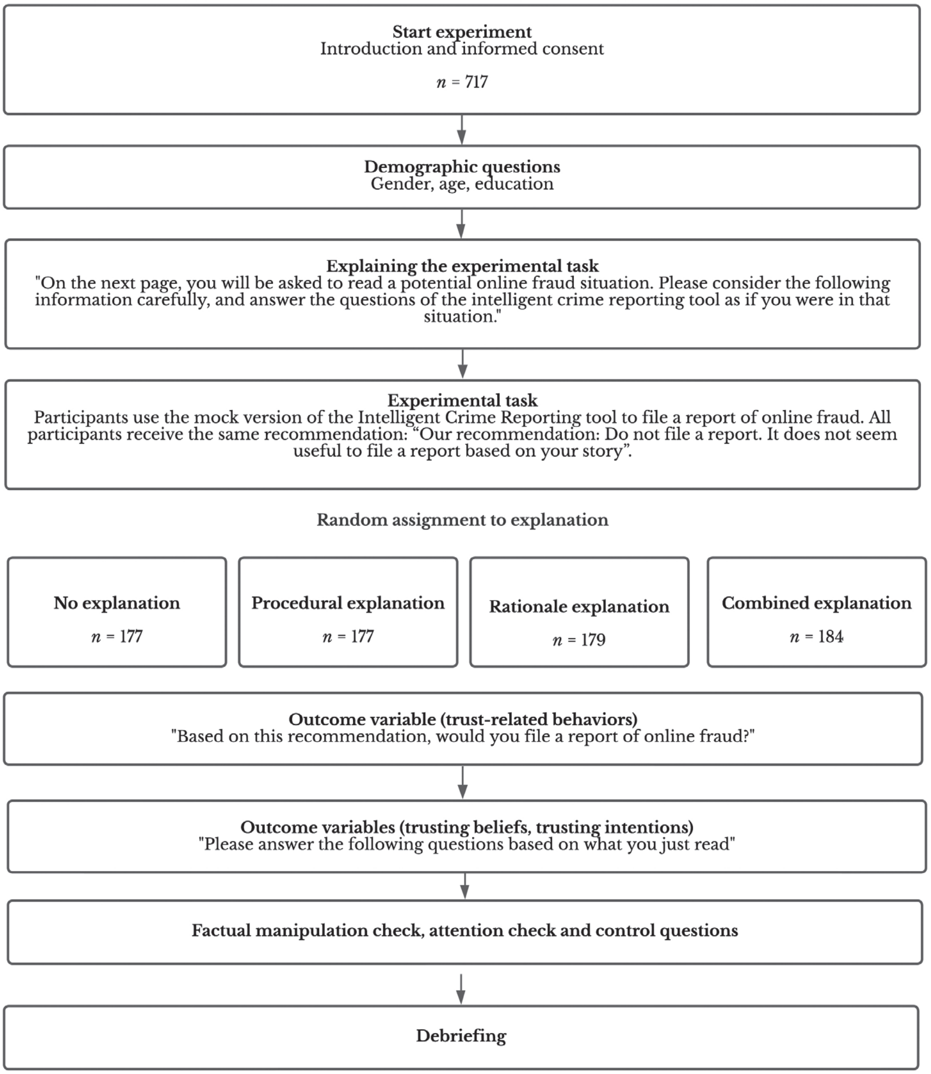
Balance checks for background variables Study 2

The proportion of female participants in Study 2 did not significantly differ from the proportion of females in the Dutch population (*z* = 1.48, *p* = .13). The proportion of participants of 40 years old and higher in Study 2 was significantly lower from the proportion of participants of 40 years old and higher in the Dutch population (*z* = 2.17, *p* = .02). The proportion of participants with a bachelor's degree or higher in Study 2 did not significantly differ from the proportion of people with a bachelor's degree or higher in the Dutch population (*z* = .06, *p* = .94). We carried out a sensitivity analysis to check whether the results for participants over 40 years old were different for participants under 40 years old. The sensitivity analysis showed no substantive alteration to our results (i.e., no significant effect became nonsignificant, and no nonsignificant effect became significant).

Appendices Study 1

Appendix 3F: Flowchart Study 1

Figure 3.4 Flowchart experimental design Study 1



Start experiment
Introduction and informed consent

n = 717

Demographic questions
Gender, age, education

Explaining the experimental task
"On the next page, you will be asked to read a potential online fraud situation. Please consider the following information carefully, and answer the questions of the intelligent crime reporting tool as if you were in that situation."

Experimental task
Participants use the mock version of the Intelligent Crime Reporting tool to file a report of online fraud. All participants receive the same recommendation: "Our recommendation: Do not file a report. It does not seem useful to file a report based on your story".

Random assignment to explanation

No explanation *n* = 177

Procedural explanation *n* = 177

Rationale explanation *n* = 179

Combined explanation *n* = 184

Outcome variable (trust-related behaviors)
"Based on this recommendation, would you file a report of online fraud?"

Outcome variables (trusting beliefs, trusting intentions)
"Please answer the following questions based on what you just read"

Factual manipulation check, attention check and control questions

Debriefing

Appendix 3G: Experimental vignettes

No explanation

> **Step 4: Our recommendation**
> Our recommendation: Do not file a report.
>
> Based on your story it does not seem useful to file a report.

Procedural explanation

> **Step 4: Our recommendation**
> Our recommendation: Do not file a report.
>
> Based on your story it does not seem useful to file a report.
>
> **Why am I recommended not to file a report?**
> The web shop www.sneakerwinkel.nl is known to the police as a **trustworthy** web shop and such web shops hardly ever commit online fraud.
>
> **How do we come to this recommendation?**
>
> - The intelligent crime reporting tool analyzes your text using various **algorithms**.
> - Among other things, a **text analysis** is used to search for specific terms and combinations of words that could indicate a fraud case.
> - Based on this text analysis, it was examined whether there are still **relevant questions**.
> - Your answers to these questions showed that you have made a purchase at www.sneakerwinkel.nl.
> - This web shop usually **does not commit fraud**.
> - Check the website of the quality mark (www.webshopqualitymark.nl) to see what you can do to get your money back.

Rationale explanation

Step 4: Our recommendation
Our recommendation: Do not file a report.

Based on your story it does not seem useful to file a report.

Why am I recommended not to file a report?
The web shop www.sneakerwinkel.nl is known to the police as a **trustworthy** web shop and such web shops hardly ever commit online fraud.

Why is this a trustworthy web shop?

- The web shop has the **quality mark** 'Web Shop Quality Mark'.
- A web shop that is affiliated with this quality mark has gone through an **extensive screening procedure**.
- The quality mark **mediates** in a dispute with a web shop.
- If this mediation does not lead to the desired result, you can submit the dispute to an **independent disputes committee**.
- Check the website of the quality mark (www.webshopqualitymark.nl) to see what you can do to get your money back.

Combined explanation

Step 4: Our recommendation
Our recommendation: Do not file a report.

Based on your story it does not seem useful to file a report.

Why am I recommended not to file a report?
The web shop www.sneakerwinkel.nl is known to the police as a **trustworthy** web shop and such web shops hardly ever commit online fraud.

How do we come to this recommendation?

- The intelligent crime reporting tool analyzes your text using various **algorithms**.
- Among other things, a **text analysis** is used to search for specific terms and combinations of words that could indicate a fraud case.
- Based on this text analysis, it was examined whether there are still **relevant questions**.
- Your answers to these questions showed that you have made a purchase at www.sneakerwinkel.nl.
- This web shop usually **does not commit fraud**.

Why is this a trustworthy web shop?

- The web shop has the **quality mark** 'Web Shop Quality Mark'.
- A web shop that is affiliated with this quality mark has gone through an **extensive screening procedure**.
- The quality mark **mediates** in a dispute with a web shop.
- If this mediation does not lead to the desired result, you can submit the dispute to an **independent disputes committee**.
- Check the website of the quality mark (www.webshopqualitymark.nl) to see what you can do to get your money back.

Appendix 3H: Descriptive statistics of factual manipulation check

Table 3.6 Descriptive statistics of the factual manipulation check in the pilot study

Answer chosen to FMC	***n***	**A (procedural)**	**B (rationale)**	**C (combined)**	**D (control)**
No explanation	20	2	3	0	15
Procedural explanation	21	12	5	4	0
Rationale explanation	21	3	9	6	3
Combined explanation	20	8	7	4	1

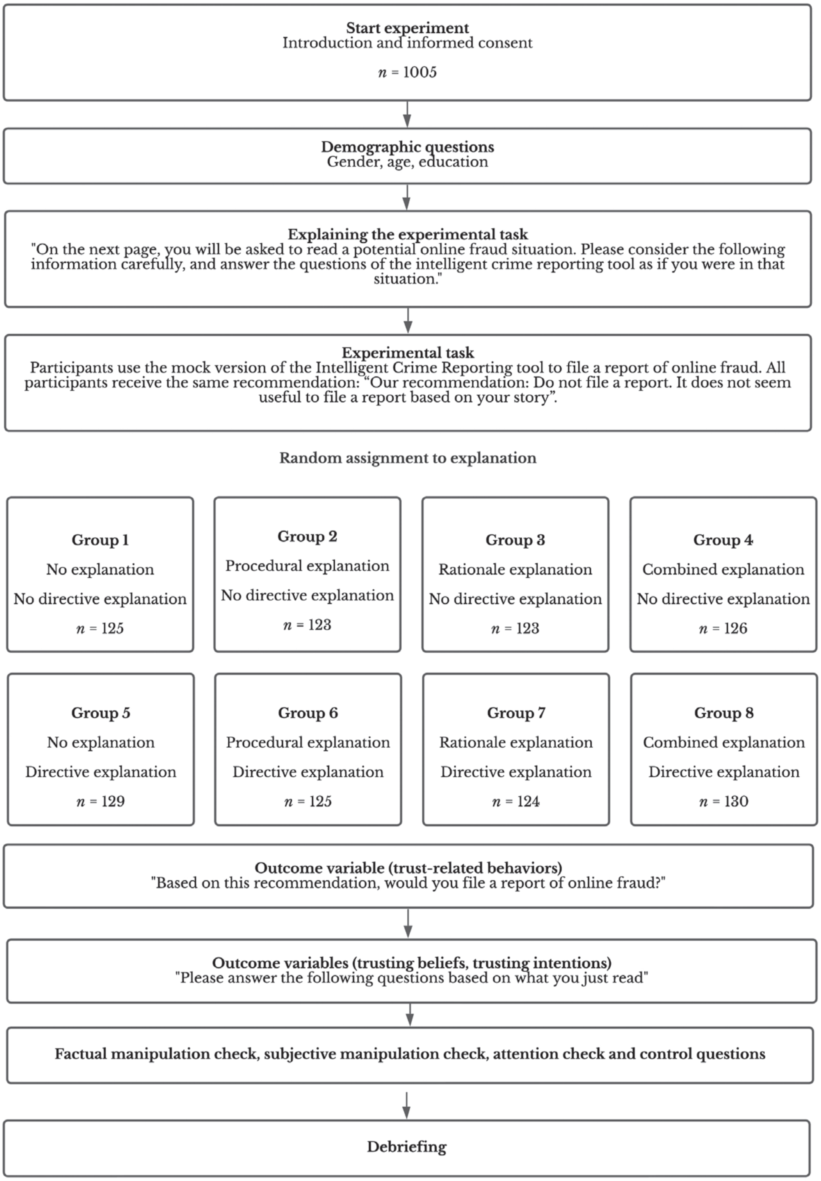
Table 3.7 Descriptive statistics of the factual manipulation check in Study 1

Answer chosen to FMC	***n***	**A (procedural)**	**B (rationale)**	**C (combined)**	**D (control)**
No explanation	177	22	12	12	131
Procedural explanation	177	98	26	47	6
Rationale explanation	179	28	63	78	10
Combined explanation	184	47	41	88	8

Appendices Study 2

Appendix 3I: Flowchart Study 2

Figure 3.5 Flowchart experimental design Study 2


Start experiment
Introduction and informed consent
n = 1005
Demographic questions
Gender, age, education
Explaining the experimental task
"On the next page, you will be asked to read a potential online fraud situation. Please consider the following information carefully, and answer the questions of the intelligent crime reporting tool as if you were in that situation."
Experimental task
Participants use the mock version of the Intelligent Crime Reporting tool to file a report of online fraud. All participants receive the same recommendation: "Our recommendation: Do not file a report. It does not seem useful to file a report based on your story".
Random assignment to explanation
Group 1
No explanation
No directive explanation
n = 125
Group 2
Procedural explanation
No directive explanation
n = 123
Group 3
Rationale explanation
No directive explanation
n = 123
Group 4
Combined explanation
No directive explanation
n = 126
Group 5
No explanation
Directive explanation
n = 129
Group 6
Procedural explanation
Directive explanation
n = 125
Group 7
Rationale explanation
Directive explanation
n = 124
Group 8
Combined explanation
Directive explanation
n = 130
Outcome variable (trust-related behaviors)
"Based on this recommendation, would you file a report of online fraud?"
Outcome variables (trusting beliefs, trusting intentions)
"Please answer the following questions based on what you just read"
Factual manipulation check, subjective manipulation check, attention check and control questions
Debriefing

Appendix 3J: Experimental vignettes

Group 1: No explanation, without a directive explanation

Step 4: Our recommendation
Our recommendation: Do not file a report.

Based on your story it does not seem useful to file a report.

Group 2: Procedural explanation, without a directive explanation

Step 4: Our recommendation
Our recommendation: Do not file a report.

Based on your story it does not seem useful to file a report.

Why am I recommended not to file a report?
The web shop www.sneakerwinkel.nl is known to the police as a **trustworthy** web shop and such web shops hardly ever commit online fraud.

How do we come to this recommendation?

- The intelligent crime reporting tool analyzes your text using various **algorithms**.
- Among other things, a **text analysis** is used to search for specific terms and combinations of words that could indicate a fraud case.
- Based on this text analysis, it was examined whether there are still **relevant questions**.
- Your answers to these questions showed that you have made a purchase at www.sneakerwinkel.nl.
- This web shop usually **does not commit fraud**.

Group 3: Rationale explanation, without a directive explanation

Step 4: Our recommendation
Our recommendation: Do not file a report.

Based on your story it does not seem useful to file a report.

Why am I recommended not to file a report?
The web shop www.sneakerwinkel.nl is known to the police as a **trustworthy** web shop and such web shops hardly ever commit online fraud.

Why is this a trustworthy web shop?

- The web shop has the **quality mark** 'Web Shop Quality Mark'.
- A web shop that is affiliated with this quality mark has gone through an **extensive screening procedure**.
- The quality mark **mediates** in a dispute with a web shop.
- If this mediation does not lead to the desired result, you can submit the dispute to an **independent disputes committee**.

Group 4: Combined explanation, without a directive explanation

Step 4: Our recommendation
Our recommendation: Do not file a report.

Based on your story it does not seem useful to file a report.

Why am I recommended not to file a report?
The web shop www.sneakerwinkel.nl is known to the police as a **trustworthy** web shop and such web shops hardly ever commit online fraud.

How do we come to this recommendation?

- The intelligent crime reporting tool analyzes your text using various **algorithms**.
- Among other things, a **text analysis** is used to search for specific terms and combinations of words that could indicate a fraud case.
- Based on this text analysis, it was examined whether there are still **relevant questions**.
- Your answers to these questions showed that you have made a purchase at www.sneakerwinkel.nl.
- This web shop usually **does not commit fraud**.

Why is this a trustworthy web shop?

- The web shop has the **quality mark** ‘Web Shop Quality Mark’.
- A web shop that is affiliated with this quality mark has gone through an **extensive screening procedure**.
- The quality mark **mediates** in a dispute with a web shop.
- If this mediation does not lead to the desired result, you can submit the dispute to an **independent disputes committee**.

Group 5: No explanation, with a directive explanation

Step 4: Our recommendation
Our recommendation: Do not file a report.

Based on your story it does not seem useful to file a report.

Check the website of the quality mark (www.webshopqualitymark.nl) to see what you can do to get your money back.

Group 6: Procedural explanation, with a directive explanation

Step 4: Our recommendation
Our recommendation: Do not file a report.

Based on your story it does not seem useful to file a report.

Why am I recommended not to file a report?
The web shop www.sneakerwinkel.nl is known to the police as a **trustworthy** web shop and such web shops hardly ever commit online fraud.

How do we come to this recommendation?

- The intelligent crime reporting tool analyzes your text using various **algorithms**.
- Among other things, a **text analysis** is used to search for specific terms and combinations of words that could indicate a fraud case.
- Based on this text analysis, it was examined whether there are still **relevant questions**.
- Your answers to these questions showed that you have made a purchase at www.sneakerwinkel.nl.
- This web shop usually **does not commit fraud**.

Check the website of the quality mark (www.webshopqualitymark.nl) to see what you can do to get your money back.

Group 7: Rationale explanation, with a directive explanation

Step 4: Our recommendation
Our recommendation: Do not file a report.

Based on your story it does not seem useful to file a report.

Why am I recommended not to file a report?
The web shop www.sneakerwinkel.nl is known to the police as a **trustworthy** web shop and such web shops hardly ever commit online fraud.

Why is this a trustworthy web shop?

- The web shop has the **quality mark** ‘Web Shop Quality Mark’.
- A web shop that is affiliated with this quality mark has gone through an **extensive screening procedure**.
- The quality mark **mediates** in a dispute with a web shop.
- If this mediation does not lead to the desired result, you can submit the dispute to an **independent disputes committee**.

Check the website of the quality mark (www.webshopqualitymark.nl) to see what you can do to get your money back.

Group 8: Combined explanation, with a directive explanation

Step 4: Our recommendation
Our recommendation: Do not file a report.

Based on your story it does not seem useful to file a report.

Why am I recommended not to file a report?
The web shop www.sneakerwinkel.nl is known to the police as a **trustworthy** web shop and such web shops hardly ever commit online fraud.

How do we come to this recommendation?

- The intelligent crime reporting tool analyzes your text using various **algorithms**.
- Among other things, a **text analysis** is used to search for specific terms and combinations of words that could indicate a fraud case.
- Based on this text analysis, it was examined whether there are still **relevant questions**.
- Your answers to these questions showed that you have made a purchase at www.sneakerwinkel.nl.
- This web shop usually **does not commit fraud**.

Why is this a trustworthy web shop?

- The web shop has the **quality mark** ‘Web Shop Quality Mark’.
- A web shop that is affiliated with this quality mark has gone through an **extensive screening procedure**.
- The quality mark **mediates** in a dispute with a web shop.
- If this mediation does not lead to the desired result, you can submit the dispute to an **independent disputes committee**.

Check the website of the quality mark (www.webshopqualitymark.nl) to see what you can do to get your money back.

Appendix 3K: Descriptive statistics of factual manipulation check

Table 3.8 Descriptive statistics of the factual manipulation check in Study 2

Answer chosen to FMC	*n*	A (procedural)	B (rationale)	C (combined)	D (control)
No directive explanation					
G1: No explanation	125	15	8	6	96
G2: Procedural explanation	123	80	12	16	15
G3: Rationale explanation	123	22	50	49	2
G4: Combined explanation	126	29	20	67	10
Directive explanation					
G5: No explanation	129	16	25	15	73
G6: Procedural explanation	125	67	18	35	5
G7: Rationale explanation	124	23	59	36	6
G8: Combined explanation	130	27	30	66	7

Appendix 3L: Descriptive statistics of subjective manipulation check

Table 3.9 Descriptive statistics of the subjective manipulation check in Study 2

"The recommendation by the intelligent crime reporting tool made clear what I could do to get my money back.", measured from 1 (totally disagree) to 7 (totally agree).

	n	Mean	SD
No directive explanation	**497**	**3.25**	**1.91**
G1: No explanation	125	2.23	1.61
G2: Procedural explanation	123	2.85	1.74
G3: Rationale explanation	123	4.11	1.81
G4: Combined explanation	126	3.72	1.96
Directive explanation	**508**	**4.37**	**1.79**
G5: No explanation	129	3.94	1.86
G6: Procedural explanation	125	4.43	1.77
G7: Rationale explanation	124	4.55	1.77
G8: Combined explanation	130	4.64	1.69

Appendix 3M: Post-hoc analyses of similar groups (1-5, 2-6, 3-7, 4-8)

No explanation with versus without a directive explanation. Participants that received no explanation in combination with a directive explanation (*M* = 4.25, SD = 1.48) did not have significantly higher *trusting beliefs* than participants that received no explanation without a directive explanation (*M* = 4.03, SD = 1.52); *t*(251.14) = 1.18, *p* = .477. In addition, participants that received no explanation with a directive explanation (*M* = 4.26, SD = 1.62) did not have significantly higher *trusting intentions* than participants that received no explanation without a directive explanation (M = 3.96, SD = 1.60); *t*(251.93) = 1.49, *p* = .278. Finally, the proportion of participants that received no explanation with a directive explanation that acted according to the recommendation to not report the crime (58.91%) was not significantly higher than without a directive explanation (55.20%), *z* = .60, *p* = 1. Thus, participants that received no explanation with a directive explanation did not have significantly higher *trust-related behaviors* than participants that received no explanation without a directive explanation.

Procedural explanation with versus without a directive explanation. Participants that received a procedural explanation with a directive explanation (*M* = 4.92, SD = 1.29) had significantly higher *trusting beliefs* than participants that received a procedural explanation without a directive explanation (*M* = 4.44, SD = 1.46); *t*(241.06) = 2.76, *p* = .012, with a small effect *d* = .35. In addition, participants that received a procedural explanation with a directive explanation (*M* = 4.99, SD = 1.35) had significantly higher *trusting intentions* than participants that received a procedural explanation without a directive explanation (M = 4.43, SD = 1.57); *t*(239.53) = 3.01, *p* = .006 , with a small effect *d* = .38. Finally, the proportion of participants that received a procedural explanation with a directive explanation that acted according to the recommendation to not report the crime (74.40%) was not significantly higher than without a directive explanation (64.23%), *z* = 1.74, *p* = .165. Thus, participants that received a procedural explanation with a directive explanation did not have significantly higher *trust-related behaviors* than participants that received a procedural explanation without a directive explanation.

Rationale explanation with versus without a directive explanation. Participants that received a rationale explanation with a directive explanation (*M* = 4.98, SD = 1.20) did not have significantly higher *trusting beliefs* than participants that received a rationale explanation without a directive explanation (*M* = 4.98, SD = 1.28); *t*(243.76) = .04, *p* = 1. In addition, participants that received a rationale explanation with a directive explanation (*M* = 5.11, SD = 1.34) did not have significantly higher *trusting intentions* than participants that received a rationale explanation without a directive explanation (M = 5.14, SD = 1.35); *t*(244.95) = -.18, *p* = 1. Moreover, the proportion of participants that received a rationale explanation with a directive explanation that acted according to the recommendation to not report the crime (83.06%) was not significantly higher than without a directive explanation (79.67%), *z* = .68, *p* = .988. Thus, participants that received a rationale explanation with a directive explanation did not have significantly higher *trust-related behaviors* than participants that received a rationale explanation without a directive explanation.

Combined explanation with versus without a directive explanation. Participants that received a combined explanation with a directive explanation (*M* = 4.64, SD = 1.53) did not have significantly higher *trusting beliefs* than participants that received a combined explanation without a directive explanation (*M* = 4.95, SD = 1.42); *t*(253.56) = 2.76, *p* = 1. In addition, participants that received a combined explanation with a directive explanation (*M* = 4.66, SD = 1.63) did not have significantly higher *trusting intentions* than participants that received a combined explanation without a directive explanation (M = 5.22, SD = 1.40); *t*(250.47) = -2.91, *p* = 1. Moreover, the proportion of participants that received a combined explanation with a directive explanation that acted according to the recommendation to not report the crime (73.08%) was not significantly higher than without a directive explanation (77.78%), *z* = .002, *p* = 1. Thus, participants that received a combined explanation with a directive explanation did not have significantly higher *trust-related behaviors* than participants that received a combined explanation without a directive explanation.

Appendix 3N: Post-hoc analyses of procedural without a directive explanation compared to other groups without a directive explanation (2-1, 2-3, 2-4)

Participants that received a procedural explanation without a directive explanation (*M* = 4.44, SD = 1.46) did not have significantly higher or lower *trusting beliefs* than participants that received no explanation without a directive explanation (*M* = 4.03, SD = 1.52); *t*(245.91) = 2.15, *p* = .097. In addition, participants that received a procedural explanation without a directive explanation (M = 4.43, SD = 1.57) did not have significantly higher or lower *trusting intentions* than participants that received no explanation without a directive explanation (M = 3.96, SD = 1.60); *t*(246) = 2.32, *p* = .063. Finally, the proportion of participants that received a procedural explanation without a directive explanation that acted according to the recommendation to not report the crime (64.23%) was not significantly higher than participants that received no explanation without a directive explanation (55.20%), *z* = 1.45, *p* = .442. Thus, participants that received a procedural explanation without a directive explanation did not have significantly higher *trust-related behaviors* than participants that received no explanation without a directive explanation.

Participants that received a procedural explanation without a directive explanation (*M* = 4.44, SD = 1.46) had significantly lower *trusting beliefs* than participants that received a rationale explanation without a directive explanation (*M* = 4.98, SD = 1.28); *t*(239.77) = -3.07, *p* = .007, with a small effect *d* = .39. In addition, participants that received a procedural explanation without a directive explanation (M = 4.43, SD = 1.57) had significantly lower *trusting intentions* than participants that received a rationale explanation without a directive explanation (M = 5.15, SD = 1.35); *t*(238.74) = -3.86, *p* < .001 , with a small effect *d* = .49. Finally, the proportion of participants that received a procedural explanation without a directive explanation that acted according to the recommendation to not report the crime (64.23%) was significantly lower than participants that received a rationale explanation without a directive explanation (79.67%), *z* = 2.70, *p* = .021, with a medium effect *h* = .35. Thus, participants that received a procedural explanation without a directive explanation had significantly lower *trust-related behaviors* than participants that received a rationale explanation without a directive explanation.

Participants that received a procedural explanation without a directive explanation (*M* = 4.44, SD = 1.46) had significantly lower *trusting beliefs* than participants that received a combined explanation without a directive explanation (*M* = 4.95, SD = 1.42); *t*(246.34) = -2.81, *p* = .016, with a small effect *d* = .36. In addition, participants that received a procedural explanation without a directive explanation (M = 4.43, SD = 1.57) had significantly lower *trusting intentions* than participants that received a combined explanation without a directive explanation (M = 5.22, SD = 1.40); *t*(242.54) = -4.19, *p* < .001 , with a medium effect *d* = .53. Finally, the proportion of participants that received a procedural explanation without a directive explanation that acted according to the recommendation to not report the crime (64.23%) was not significantly higher or lower than participants that received a combined explanation without a directive explanation (77.78%), *z* = 2.36, *p* = .055. Thus, participants that received a procedural explanation without a directive explanation did not have significantly higher or lower *trust-related behaviors* than participants that received a combined explanation without a directive explanation.

Appendices Chapter 4

Appendix 4A: Guiding questions for public organizations

Table 4.4 Guiding questions for public organizations

Theme	Topic	Guiding questions or information
Introduction	Introduction researcher	▪ Introduction of the researcher
	Goal of study	▪ Insight into what algorithm registers are, and what the implications of these registers are.
	Confidentiality, anonymity, recording	▪ Informed consent letter (confidentiality, personal data, anonymity, duration of interview, etc.) ▪ May I record the conversation? · *If yes: start recording* ▪ Formally asking about the informed consent on the recording: · Have you read and understood the information letter? · Do you have any questions about it?
	Introduction respondent	▪ What is your role (in this organization)?
Algorithm registers	Definition	▪ How do you define an algorithm register? ▪ What is seen as an 'algorithm' in your organization?
	Goal	▪ What is the main goal of your algorithm register? ▪ What kind of transparency do you achieve with your algorithm register?
	Design	▪ Which algorithms do you include in your algorithm register? · Who decides that? · How do you decide which algorithms are high-risk? ▪ How did you determine which categories to provide information about in the register? ▪ For whom is your algorithm register intended? · What are their needs in terms of accessibility and usability? ▪ How did you arrive at what the algorithm register looks like now *(excel/website/etc.)*?

Table 4.4 Continued

Theme	Topic	Guiding questions or information
		▪ What do you know about the actual usage of your algorithm register? · How often is it consulted? · Has it been covered by the media? · Who are the actual users or your register? · From whom do you receive questions about your register? ▪ Did you notice a difference in the amount of FOIA-requests since the publication of the register? ▪ Could you describe the process of registering an algorithm from beginning to end? ▪ In a perfect world without any limitations, how does the perfect algorithm register look like?
Implications	Positive implications	▪ What are positive implications of your algorithm register?
	Negative implications	▪ Did you experience any negative implications of your algorithm register?
Closing	Questions	▪ Would you like to address a topic that we did not cover? ▪ Do you have any questions?
	Documents	▪ Do you have any relevant documents that I could use for this research?
	Thank you	▪ I would like to thank you very much for your participation in this study.
	Follow-up	▪ After completion of the research I can send you the end result. Are you interested in that?

Note. This table encompasses the general guiding questions posed to respondents. For each participant, supplementary questions were included, aligning with the policy documents they authored or the research they conducted. These tailored questions aimed to capture nuanced insights and contribute to a comprehensive understanding of the diverse perspectives represented in the dataset.

Appendix 4B: Guiding questions for oversight authorities and societal watchdogs

Table 4.5 Guiding questions for oversight authorities and societal watchdogs

Theme	Topic	Guiding questions or information
Introduction	Introduction researcher	▪ Introduction of the researcher
	Goal of study	▪ Insight into what algorithm registers are, and what the implications of these registers are.
	Confidentiality, anonymity, recording	▪ Informed consent letter (confidentiality, personal data, anonymity, duration of interview, etc.) ▪ May I record the conversation? · *If yes: start recording* ▪ Formally asking about the informed consent on the recording: · Have you read and understood the information letter? · Do you have any questions about it?
	Introduction respondent	▪ What is your role (in this organization)?
Algorithm registers	Experiences	▪ Have you ever consulted an algorithm register? *If yes, ask the following:* · How was that experience? · *ask follow-up questions (content, design, etc.)* · Which information did you find in the register? · What did it bring you? · What did you miss? · How often do you use it?
	Definition	▪ How would you describe an algorithm register?
	Goal	▪ What do you think is the main goal of an algorithm register? ▪ What kind of transparency is being achieved with an algorithm register?
	Design	▪ Which algorithms should be disclosed in an algorithm register? ▪ What type of information (data) about an algorithm should be disclosed and why? ▪ Who are the intended users of algorithm registers? · What are their needs in terms of accessibility and usability?

Table 4.5 Continued

Theme	Topic	Guiding questions or information
		▪ Who are the actual users of algorithm registers? ▪ What are their needs in terms of accessibility and usability? ▪ How should an algorithm register look like? ▪ In a perfect world without any limitations, how does the perfect algorithm register look like?
Implications	Positive implications	▪ What are positive implications of algorithm registers? ▪ How can an algorithm register contribute to the improvement of the algorithms used by public organizations?
	Negative implications	▪ What are negative implications of algorithm registers? ▪ How can these negative implications be prevented?
Closing	Questions	▪ Would you like to address a topic that we didn't cover? ▪ Do you have any questions?
	Documents	▪ Do you have any relevant documents that I could use for this research?
	Thank you	▪ I would like to thank you very much for your participation in this study.
	Follow-up	▪ After completion of the research I can send you the end result. Are you interested in that?

Note. This table encompasses the general guiding questions posed to respondents. For each participant, supplementary questions were included, aligning with the policy documents they authored or the research they conducted. These tailored questions aimed to capture nuanced insights and contribute to a comprehensive understanding of the diverse perspectives represented in the dataset.

Appendix 4C: Overview of documents

Table 4.6 Overview of documents

Document reference ID	Document name	Author
D1	Xenophobic machines *(Xenofobe machines)*	Amnesty International
D2	Agenda Digital City Interim report 2019 – 2020 *(Agenda Digitale Stad Tussenrapportage 2019 – 2020)*	City of Amsterdam *(Gemeente Amsterdam)*
D3	A Digital City for and from everyone *(Een Digitale Stad voor én van iedereen)*	City of Amsterdam *(Gemeente Amsterdam)*
D4	Amsterdam Intelligence *(Amsterdamse Intelligentie)*	City of Amsterdam *(Gemeente Amsterdam)*
D5	Agenda Digital City 2021 | 2022 Interim report *(Agenda Digitale Stad 2021 | 2022 Tussenrapportage)*	City of Amsterdam *(Gemeente Amsterdam)*
D6	Supervision of AI & Algorithms *(Toezicht op AI & Algoritmes)*	Data Protection Authority *(Autoriteit Persoonsgegevens)*
D7	Report on algorithm risks in the Netherlands *(Rapportage algoritmerisico's Nederland)*	Data Protection Authority *(Autoriteit Persoonsgegevens)*
D8	An audit of algorithms *(Algoritmes getoetst)*	Netherlands Court of Audit *(Algemene Rekenkamer)*
D9	Information needs of citizens about the use of algorithms by governments *(Informatiebehoeften van burgers over de inzet van algoritmes door overheden)*	Het PON & Telos
D10	Letter to Parliament Response to report 'an audit of algorithms' *(Kamerbrief Reactie op rapport 'Algoritmes getoetst')*	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*
D11	Letter to Parliament State of affairs Algorithm register *(Kamerbrief Stand van zaken Algoritmeregister)*	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*

Table 4.6 Continued

Document reference ID	Document name	Author
D12	Implementation framework ‘Responsible use of algorithms’ Version 1.0 *(Implementatiekader ‘Verantwoorde inzet van algoritmen’ Versie 1.0)*	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*
D13	Algorithm Register Guide Getting started with the Algorithm Register Version 0.8.4 *(Handreiking Algoritmeregister Aan de slag met het Algoritmeregister Versie 0.8.4)*	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*
D14	Bingo Reasons not to publish Reasons to publish *(Bingo Redenen om niet te publiceren Redenen om te publiceren)*	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*
D15	Letter to Parliament Collection letter ‘regulating algorithms’ *(Kamerbrief Verzamelbrief ‘algoritmen reguleren’)*	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*
D16	Newsletter Algorithms Nominate your algorithm for the “Best Described Algorithm” election! *(Nieuwsbrief Algoritmes Nomineer je algoritme voor de “Best beschreven Algoritme” verkiezing!)*	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*
D17	Newsletter Algorithm Register Facts and figures | Tips to quickly get out of the starting blocks *(Nieuwsbrief Algoritmeregister Feiten en cijfers | Tips om snel uit startblokken te komen)*	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*
D18	Algorithm Register Guide Getting started with the Algorithm Register Version 1.0 *(Handreiking Algoritmeregister Aan de slag met het Algoritmeregister Versie 1.0)*	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*

Table 4.6 Continued

Document reference ID	Document name	Author
D19	Algorithm Register Manual Version 1.0.0a *(Handleiding Algoritmeregister Versie 1.0.0a)*	Ministry of the Interior and Kingdom Relations *(Ministerie van Binnenlandse Zaken en Koninkrijksrelaties)*
D20	Digital rights governance framework	Cities Coalition for Digital Rights
D21	In all openness: transparent algorithm use by the government *(In alle openheid: transparant algoritmegebruik door de overheid)*	Netherlands Institute for Human Rights *(College voor de Rechten van de Mens)*
D22	Open Algorithms in France, the Netherlands and New Zealand	Open Government Partnership
D23	Registration of algorithms and sensors, for transparency and public accountability *(Registratie van algoritmen en sensoren, voor transparantie en publieke verantwoording)*	Association of Dutch Municipalities (*Vereniging van Nederlandse Gemeenten)*
D24	Call for development of national sensor register *(Oproep ontwikkeling landelijk sensorenregister)*	Association of Dutch Municipalities (*Vereniging van Nederlandse Gemeenten)*
D25	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 COM(2021) 206 final *(Voorstel voor een verordening van het Europees Parlement en de Raad tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie (wet op de artificiële intelligentie) en tot wijziging van bepaalde wetgevingshandelingen van de unie COM(2021) 206 final)*	European Commission *(Europese Commissie)*
D26	A citizen is not a data set *(Een burger is geen dataset)*	National Ombudsman *(Nationale Ombudsman)*
D27	A performance review with your algorithm? *(Een functioneringsgesprek met je algoritme?)*	Arjan Widlak and Ola Al Khatib

Table 4.6 Continued

Document reference ID	**Document name**	**Author**
D28	Colored technology Exploring the ethical use of algorithms *(Gekleurde technologie Verkenning ethisch gebruik algoritmes)*	(Municipal) Court of Audit Rotterdam *(Rekenkamer Rotterdam)*
D29	Research design Follow-up research into algorithms *(Onderzoeksopzet Vervolgonderzoek algoritmes)*	(Municipal) Court of Audit Rotterdam *(Rekenkamer Rotterdam)*
D30	Algorithm policy SVB *(Algoritmebeleid SVB)*	Social Insurance Bank *(Sociale Verzekeringsbank)*
D31	Report Digital audit algorithm management and accountability *(Rapport Digitale audit algoritmebeheer en verantwoording)*	City of Utrecht *(Gemeente Utrecht)*
D32	UWV Compass Data Ethics *(UWV Kompas Data Ethiek)*	Employees Insurance Agency *(Uitvoeringsinstituut Werknemersverzekeringen)*
D33	Principles for the Digital Society *(Principes voor de Digitale Samenleving)*	Association of Dutch Municipalities (*Vereniging van Nederlandse Gemeenten)*

Appendix 4D: Detailed description of the coding process

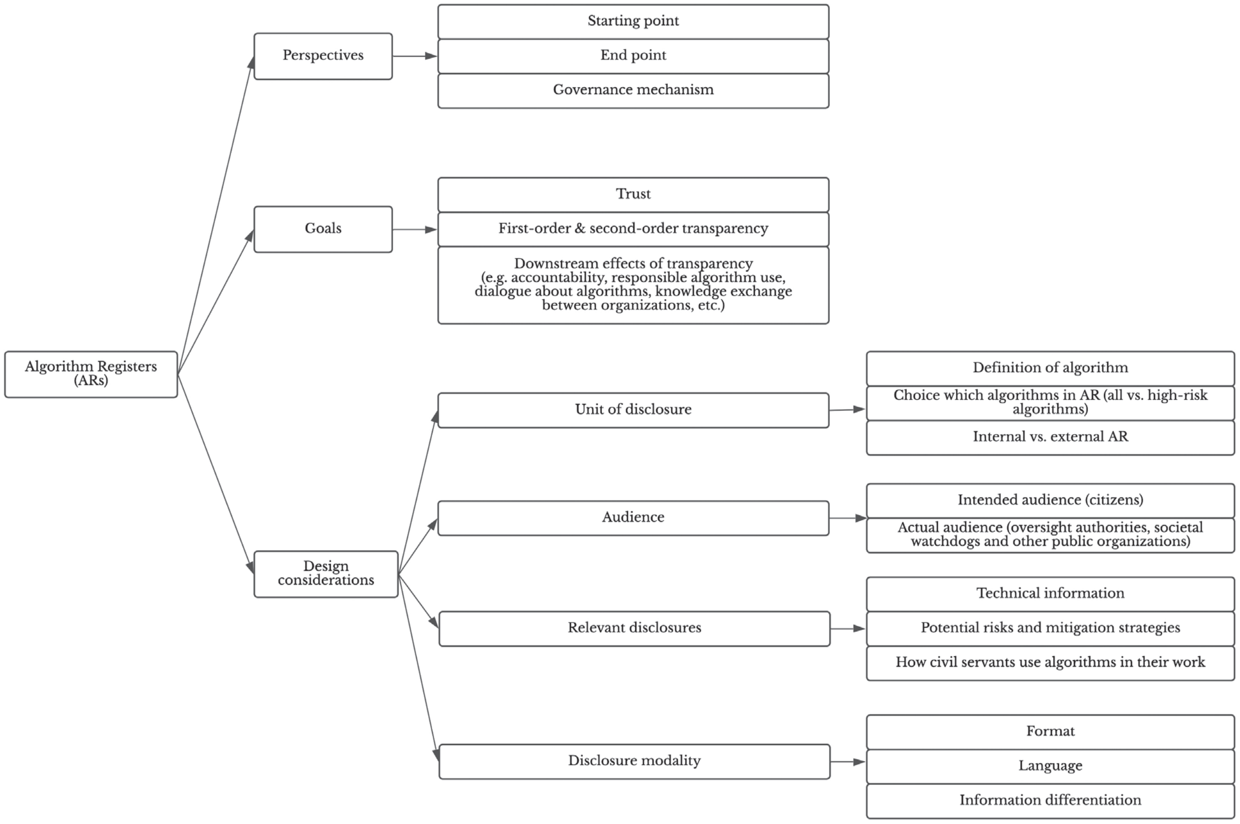
The coding process started by coding interviews. Five interviews (24%) were coded by the author and two other researchers to ensure inter-coder reliability in the use and expansion of the code tree (Bryman, 2016). Employing abductive coding, the researchers utilized a coding scheme based on the theoretical framework (86 codes), allowing for the emergence of new codes from the data. Subsequently, the researchers scrutinized the existing coded data to confirm the presence of any newly identified codes, resulting in 107 new codes. Following the initial coding of the five interviews, a series of three 1.5-hour sessions was convened, wherein the researchers engaged in an iterative dialogue. This process involved scrutinizing the coded interviews collaboratively. Consensus was sought particularly in instances where variations in coding emerged among the researchers. In cases where disparities in coding emerged, the researchers engaged in a dialogue to deliberate on the most appropriate code for a given text segment, ensuring a shared understanding of the coding choices and a coherent coding structure. This not only ensured inter-coder reliability but also allowed for the exploration of different perspectives, thereby contributing to the depth of the analysis.

In assessing inter-coder reliability, the Kappa scores were computed for three coder pairs. The agreement between coders A and B yielded a Kappa score of 0.85. For coders A and C, the Kappa score was 0.77. Coders B and C demonstrated an agreement with a Kappa score of 0.88. The overall inter-coder reliability, represented by the average Kappa score, was computed as 0.83, with the calculation taking into account the weighted average based on the number of codes in each comparison.

The author coded the remaining 76% of the interviews and all documents. After completing this coding of interviews and documents, the author and another researcher participated in the iterative process known as "coding on". Following Richards' (2015, p. 116) recommendation, we meticulously reviewed all codes that had combined have more than 40 references to transcribed texts. This involved thoughtful reflection on whether a category required reformulation or if the emergence of a new (sub)category was necessary. In instances of the latter, we went through all the references in the category to see if they belonged to the newly created (sub)category or if they could stay in their original category. This process of "coding on" resulted in creating one or more new codes under these 10 subcategories.

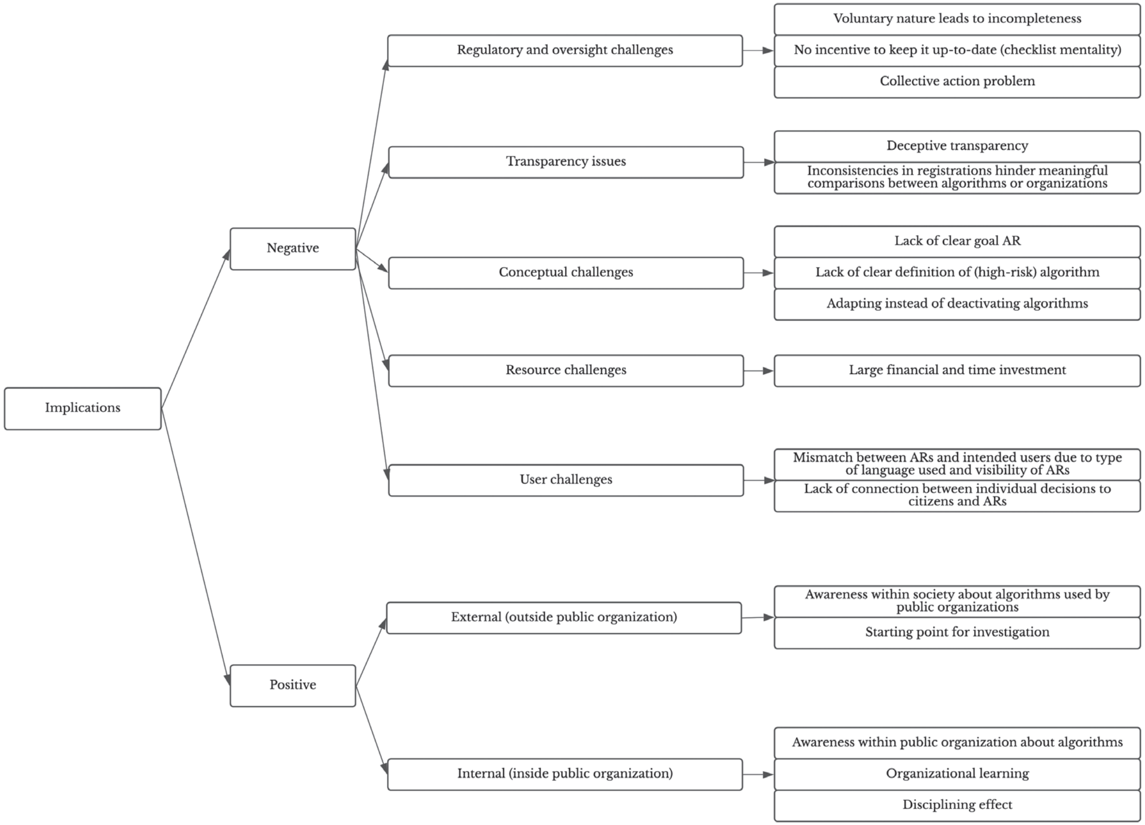
Appendix 4E: Coding tree algorithm registers

Figure 4.2 Coding tree: Algorithm registers


Algorithm Registers (ARs)
Perspectives
Starting point
End point
Governance mechanism
Goals
Trust
First-order & second-order transparency
Downstream effects of transparency (e.g. accountability, responsible algorithm use, dialogue about algorithms, knowledge exchange between organizations, etc.)
Design considerations
Unit of disclosure
Definition of algorithm
Choice which algorithms in AR (all vs. high-risk algorithms)
Internal vs. external AR
Audience
Intended audience (citizens)
Actual audience (oversight authorities, societal watchdogs and other public organizations)
Relevant disclosures
Technical information
Potential risks and mitigation strategies
How civil servants use algorithms in their work
Disclosure modality
Format
Language
Information differentiation

Appendix 4F: Coding tree implications

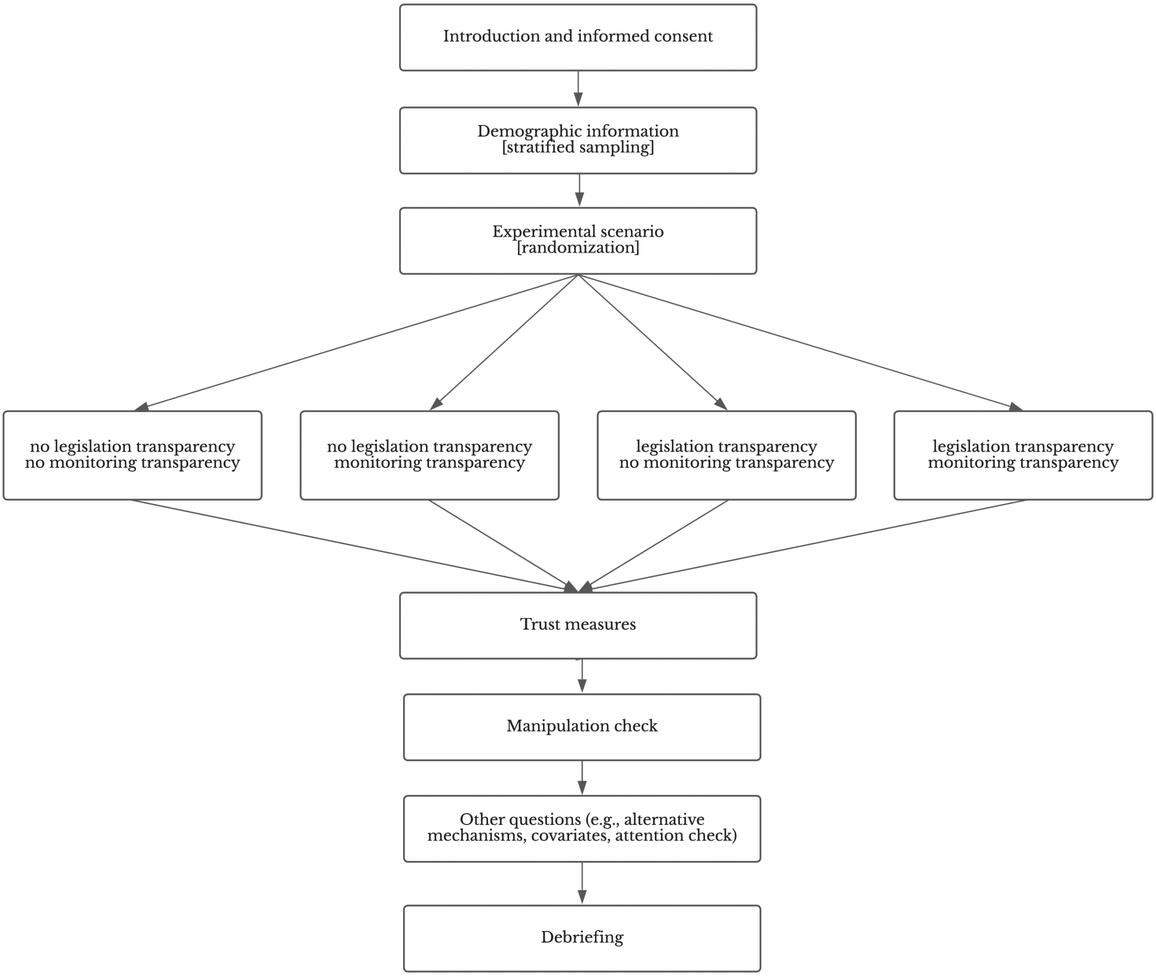
Figure 4.3 Coding tree: Implications


Implications
Negative
Positive
Regulatory and oversight challenges
Voluntary nature leads to incompleteness
No incentive to keep it up-to-date (checklist mentality)
Collective action problem
Transparency issues
Deceptive transparency
Inconsistencies in registrations hinder meaningful comparisons between algorithms or organizations
Conceptual challenges
Lack of clear goal AR
Lack of clear definition of (high-risk) algorithm
Adapting instead of deactivating algorithms
Resource challenges
Large financial and time investment
User challenges
Mismatch between ARs and intended users due to type of language used and visibility of ARs
Lack of connection between individual decisions to citizens and ARs
External (outside public organization)
Awareness within society about algorithms used by public organizations
Starting point for investigation
Internal (inside public organization)
Awareness within public organization about algorithms
Organizational learning
Disciplining effect

Appendices Chapter 5

Appendix 5A: Flowchart of experimental design

Figure 5.1 Flowchart experimental design


Introduction and informed consent
Demographic information [stratified sampling]
Experimental scenario [randomization]
no legislation transparency no monitoring transparency
no legislation transparency monitoring transparency
legislation transparency no monitoring transparency
legislation transparency monitoring transparency
Trust measures
Manipulation check
Other questions (e.g., alternative mechanisms, covariates, attention check)
Debriefing

Note. All items are detailed in the preregistration at Open Science Framework.

Appendix 5B: Experimental vignettes

Imagine the following situation:
Normally, you seldom see police in your street, but you have noticed that in the past week, there has been a police car driving through your street every day. Additionally, you noticed that your new neighbor, who recently immigrated to the Netherlands, was briefly stopped and questioned today on the street by two police officers as he was walking around the neighborhood. The officers wanted to know what he was doing and if he lived in the neighborhood.

There have been no burglaries in your neighborhood in the past months, and you find it strange that your neighbor was questioned by the officers on the street.

You decide to call the police station to ask about this, and you receive the explanation below.

- What is happening? A **computer program (algorithm)** is used to predict when and in which neighborhoods the risk of burglaries is highest.
- How does this program work? The computer program uses **personal information about neighborhood residents**, such as education level and income, to make these predictions. There are extra police patrols in your neighborhood this week because the program indicated an increased risk of burglaries.
- [MANIPULATION LEGISLATION TRANSPARENCY]
- [MANIPULATION MONITORING TRANSPARENCY]

[no legislation / no monitoring]

- Is this legal? The police did **not provide information** about whether there are any **laws** in place to control how the police use the computer program and personal information.
- Is anyone monitoring how police are using this program? The police did **not provide information** about who is **monitoring** their use of the computer program and personal information.

[no legislation / yes monitoring]

- Is this legal? The police did **not provide information** about whether there are any **laws** in place to control how the police use the computer program and personal information.
- Is anyone monitoring how police are using this program? The police indicated that an independent government agency - the **Data Protection Authority** - monitors their use of the computer program and personal information.

[yes legislation / no monitoring]

- Is this legal? The police indicated that the **Police Data Act** applies to this computer program. This law prescribes when and for which tasks personal information may be used.
- Is anyone monitoring how police are using this program? The police did **not provide information** about who is **monitoring** their use of the computer program and personal information.

[yes legislation / yes monitoring]

- Is this legal? The police indicated that the **Police Data Act** applies to this computer program. This law prescribes when and for which tasks personal information may be used.
- Is anyone monitoring how police are using this program? The police indicated that an independent government agency - the **Data Protection Authority** - monitors their use of the computer program and personal information.

Appendix 5C: Balance checks

Table 5.4 Balance checks

Variables	**Population** **(*N* = 17.811.291)**	**Sample** **(*n* = 877)**	**χ^2** **(*p*-value)**
Gender			
Females	50.25%	51.20%	0.28 (.598)
Males	49,75%	48.69%	0.36 (.552)
Non-binary	/	.11%	/
Age (in years)			
18-39	35,4%	32.04%	4.18 (.041)*
40-64	40,2%	42.30%	1.53 (.217)
65 and higher	24,4%	25.66%	0.68 (.409)
Education			
Low and Middle	59%	59.86%	0.24 (.627)
High	41%	40.14%	0.24 (.627)

Note: χ^2 follows from Fisher Exact probability test of the difference between the population and sample proportions.
* $p < .05$

The results of the chi-squared test revealed a slight imbalance across treatment groups in terms of age. We extended our main analysis by controlling for age and found no substantive differences in our results. Therefore, we present the findings without a control variable in the main body of the text.

Appendix 5D: Outcome variable "citizen trust in the use of predictive algorithms by the police"

The scale operationalizes trust using three dimensions (competence, benevolence and integrity), with 3 items for each dimension, resulting in 9 items in total, measured on a 5-point Likert scale (strongly disagree, disagree, neutral, agree, and strongly agree).

Competence

I think that, when it concerns the use of predictive algorithms …

[Competence 1]	The police are capable.
[Competence 2]	The police are experts.
[Competence 3]	The police carry out their duty very well.

Benevolence

I think that, when it concerns the use of predictive algorithms …

[Benevolence 1]	If citizens need help, the police will do their best to help them.
[Benevolence 2]	The police act in the interests of citizens.
[Benevolence 3]	The police are genuinely interested in the well-being of citizens.

Integrity

I think that, when it concerns the use of predictive algorithms …

[Integrity 1]	The police approach citizens in a sincere way.
[Integrity 2]	The police are sincere.
[Integrity 3]	The police are honest.

Appendix 5E: Confirmatory factor analysis description

Figure 5.2 shows the model, which consists of three common factors (competence, benevolence and integrity) that explain nine observed variables. See Table 5.5 for the correlation matrix of these observed variables. We consider the model fit to be satisfactory if RMSEA < .10 (Browne & Cudeck, 1992) and CFI > .90 (Bentler & Bonett, 1980), indicating respectively exact to mediocre fit and good fit. To scale the common factors, we used Unit Loading Identification. The first indicator of each common factor was constrained to 1. The model fit was satisfactory, RMSEA = .046, RMSEA 90% CI [.033, .059], CFI = .993. Therefore, the model was accepted. The common factors of the trust dimensions correlated strongly since all correlations were above 0.70, see Table 5.6. All observed variables were significantly predicted by the (theoretically expected) corresponding common factor. Additionally, all non-restricted factor loadings were approximately equal, see Table 5.7.

Figure 5.2 Common factor model

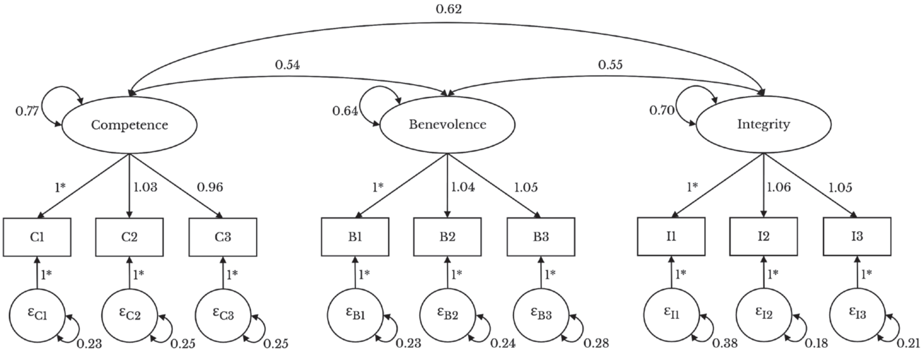


Table 5.5 Correlation matrix for trust items

Measure	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1. Competence 1	1.00								
2. Competence 2	0.78	1.00							
3. Competence 3	0.74	0.75	1.00						
4. Benevolence 1	0.59	0.55	0.57	1.00					
5. Benevolence 2	0.59	0.56	0.59	0.74	1.00				
6. Benevolence 3	0.58	0.56	0.60	0.73	0.73	1.00			
7. Integrity 1	0.58	0.62	0.64	0.54	0.59	0.55	1.00		
8. Integrity 2	0.67	0.65	0.67	0.65	0.64	0.62	0.72	1.00	
9. Integrity 3	0.64	0.62	0.65	0.63	0.64	0.60	0.72	0.80	1.00

Table 5.6 Estimated (co)variance

Variable(s)	(Co)Variance	Correlation	SE	*p*
Common factor covariances				
Competence ~ Benevolence	0.54	0.77	0.03	.000*
Competence ~ Integrity	0.62	0.84	0.04	.000*
Benevolence ~ Integrity	0.55	0.82	0.04	.000*
Variances				
Competence	0.77		0.05	.000*
Benevolence	0.64		0.04	.000*
Integrity	0.70		0.05	.000*
Competence 1	0.23		0.02	.000*
Competence 2	0.24		0.02	.000*
Competence 3	0.25		0.02	.000*
Benevolence 1	0.23		0.02	.000*
Benevolence 2	0.24		0.02	.000*
Benevolence 3	0.28		0.02	.000*
Integrity 1	0.38		0.02	.000*
Integrity 2	0.18		0.01	.000*
Integrity 3	0.21		0.01	.000*

* *p* < .05

Table 5.7 Factor loadings

Indicator	Factor loading	Standardized factor loading	SE	*p*
Competence				
Competence 1	1.00	0.88	-	-
Competence 2	1.03	0.88	0.03	.000*
Competence 3	0.96	0.86	0.03	.000*
Benevolence				
Benevolence 1	1.00	0.86	-	-
Benevolence 2	1.04	0.86	0.03	.000*
Benevolence 3	1.05	0.85	0.03	.000*
Integrity				
Integrity 1	1.00	0.81	-	-
Integrity 2	1.06	0.90	0.03	.000*
Integrity 3	1.05	0.89	0.03	.000*

* *p* < .05

Appendix 5F: Manipulation checks

We asked participants about whether they thought they received information about monitoring and we asked them about whether they received information about legislation regarding the police's use of algorithms. The results for the manipulation checks are shown in Tables 5.8 and 5.9.

Monitoring transparency

Please choose which statement applies to the scenario you just read:

□ The police indicated that an independent government agency –the Data Protection Authority– monitors the computer program and police's use of personal information.

□ The police did **not** provide information about who is monitoring their use of the computer program.

Table 5.8 Descriptive results for manipulation check monitoring transparency condition

	Answer chosen manipulation check		
Experimental monitoring transparency condition	**No monitoring transparency**	**Monitoring transparency**	***n***
No monitoring transparency	310	123	433
Monitoring transparency	111	333	444
Total	421	456	877

The manipulation check showed the experimental monitoring transparency condition to be successful. 73.32% of participants passed the manipulation check for monitoring transparency. The chi-squared test confirmed that participants who received information about monitoring were significantly more likely to indicate that they did, $\chi^2 = 188.8$, $p < .001$.

Legislating transparency

Please choose which statement applies to the scenario you just read:

□ The police indicated that there is a law in place –the Police Data Act– that controls how the police use the computer program.

□ The police did **not** provide information about laws that control how the police use the computer program.

Table 5.9 Descriptive results for manipulation check legislation transparency condition

	Answer chosen manipulation check		
Experimental legislation transparency condition	**No legislation transparency**	**Legislation transparency**	***n***
No legislation transparency	310	135	445
Legislation transparency	119	313	432
Total	429	448	877

The manipulation check showed the experimental legislation transparency condition to be successful. 71.04% of participants passed the manipulation check for legislation transparency. The chi-squared test confirmed that participants who received information about legislation were significantly more likely to indicate that they did, $\chi^2 = 153.92$, $p < .001$.

Co-author statements

Co-author statement for Utrecht School of Governance dissertations

Universiteit Utrecht

Contribution

This co-author statement regards the following contribution:

Explanations increase citizen trust in police algorithmic recommender systems: Findings from two experimental tests

Published in the following journal, volume or other outlet

Nieuwenhuizen, E. N., Meijer, A. J., Bex, F. J., & Grimmelikhuijsen, S. G. (2024). Explanations increase citizen trust in police algorithmic recommender systems: Findings from two experimental tests. *Public Performance & Management Review, 48*(3), 590–625. https://doi.org/10.1080/15309576.2024.2443140

The contribution is part of the following dissertation:

Algorithmic Transparency in Government – Esther Nieuwenhuizen

Author roles and contributions

Contributor Roles Taxonomy (CRediT) (source: https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement; see the source for an explanation of concepts)

	Esther Nieuwenhuizen	Albert Meijer	Floris Bex	Stephan Grimmelikhuijsen
Conceptualization	V	V	V	V
Research design	V	V	V	V
Privacy and ethics approval	V			V
Data collection	V			
Data analysis	V			
Data curation	V			
Writing - original draft	V	V	V	V
Writing - review & editing	V	V	V	V

Additional remarks:

n.a.

Signatures authors, including signature PhD candidate

Date	Name	Signature
11-9-2024	Esther Nieuwenhuizen	
13-9-2024	Albert Meijer	
11-9-2024	Floris Bex	
13-9-2024	Stephan Grimmelikhuijsen	

Co-author statement for Utrecht School of Governance dissertations

Universiteit Utrecht

Contribution

This co-author statement regards the following contribution:

Does institutional transparency affect citizen trust in predictive policing? Evidence from a survey-experiment in the Netherlands

Published in the following journal, volume or other outlet

Under review

The contribution is part of the following dissertation:

Algorithmic Transparency in Government – Esther Nieuwenhuizen

Author roles and contributions

Contributor Roles Taxonomy (CRediT) (source: https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement; see the source for an explanation of concepts)

	Esther Nieuwenhuizen	**Vishal Trehan**	**Gregory Porumbescu**
Conceptualization	V		
Research design	V	V	V
Privacy and ethics approval	V		
Data collection	V		
Data analysis	V		
Data curation	V		
Writing - original draft	V		
Writing - review & editing	V	V	V

Additional remarks:

n.a.

Signatures authors, including signature PhD candidate

Date	Name	Signature
11-09-2024	Esther Nieuwenhuizen	
12-09-2024	Vishal Trehan	Trehan
12-09-2024	Gregory Porumbescu	

Summary in English

Algorithms are playing an increasingly prominent role in many public service domains, ranging from healthcare to criminal justice. It is important that the use of algorithms is conducted in a responsible and trustworthy manner to mitigate concerns (e.g., biased, unfair or untransparent decision making) and make use of the value (e.g., efficiency, productivity or safety) that algorithms can offer to the public sector. Transparency is frequently mentioned as an important factor in building, maintaining and strengthening citizen trust in the use of algorithms by governments.

Still, there is ongoing debate about what transparency should entail in practice and how it actually influences trust. This dissertation therefore seeks to clarify how transparency about algorithm use by public sector organizations influences citizen trust. The following research question (RQ) is central to this dissertation:

RQ: *How does transparency affect citizen trust in algorithm use by government organizations?*

Answering this main research question provides an opportunity to address four shortcomings in the literature, leading to the formulation of four underlying sub-questions. These shortcomings, the corresponding sub-questions, and the insights gained are briefly outlined below.

Throughout the dissertation, the police serve as a recurring context to explore broader phenomena related to algorithm use within the public sector. Specifically, I examine two existing police algorithms: the Intelligent Crime Reporting Tool (Chapters 2 and 3) and the Crime Anticipation System (Chapter 5). One exception to this police-centered focus is Chapter 4, which examines algorithm registers. This chapter centers on transparency regarding the organizational embedding of algorithms across the broader public sector. Throughout all chapters, I discuss the broader relevance of my findings for public sector organizations. As such, this dissertation offers insights not only for the police but also for other public organizations.

Multi-level perspective on algorithmic transparency and trust

The literature on algorithmic transparency and trust shows that most studies have investigated this relationship in a rather narrow manner. They often focus on transparency of the algorithmic application itself, whereas general transparency literature suggests that transparency also plays a role at higher-order levels, such as at organizational and institutional levels. This includes transparency regarding organizational policies and institutional rules governing the appropriate implementation and use of algorithms in government. How this broader understanding of transparency can be applied to enhance trust in algorithmic governance remains unexplored in the literature, leading to the first sub-question:

RQ1: *How can we understand the relationship between trust and algorithmic transparency at the micro, meso and macro levels?*

Chapter 2 addresses this question by conceptualizing both algorithmic transparency and trust in algorithmic governance at the micro, meso, and macro levels. It then discusses how algorithmic transparency can contribute to building, maintaining, and strengthening citizens' trust in algorithmic governance. This results in a conceptual framework that can be used to empirically examine the relationship between transparency and trust in algorithmic governance. This is a multi-level framework that focuses on micro (specific algorithms), meso (organizational) and macro (institutional) levels of transparency, and on how these are linked to citizen trust. This chapter shows that to build, maintain and strengthen trust in algorithmic governance, it is not only necessary for algorithms themselves to be transparent (micro), but also the way they are used by government organizations (meso) and the regulations governing their use (macro) require openness.

Micro-level

Existing research offers contrasting insights into which types of explanations most effectively strengthen citizen trust in algorithmic outcomes, such as the recommendations of algorithmic recommender systems. Some studies argue that providing justifications for individual algorithmic decisions is crucial for generating trust in these decisions, while others show that providing information about the algorithmic decision procedure is most effective in strengthening trust in an algorithmic outcome. Altogether, the type of explanation that is most effective in strengthening citizen trust has only been investigated to a limited extent and has remained contested, leading to the following sub-question:

RQ2: *What are the effects of explanations on citizen trust in algorithmic recommendations?*

Chapter 3 addresses this question by testing the effects of various explanations provided by algorithmic recommender systems on citizen trust in two survey-experiments. The first experiment builds upon the concepts of procedural, rationale and combined explanations; the second experiment focuses on directive explanations. While nuances exist, the findings show that explaining algorithmic recommendations—in any form—strengthens trusting beliefs, trusting intentions, and trust-related behavior in citizens who receive digital public service deliveries. This may suggest that trust in algorithmic recommendations increases when citizens see that governments make an effort to provide an explanation, regardless of the nature of this explanation.

Meso-level

The scholarly debate about the new phenomenon of algorithm registers shows that the value of these registers is questioned: some scholars claim that governments only disclose politically non-sensitive algorithms and provide information with limited usefulness for accountability purposes. In contrast, a more optimistic view is also put forward: registers could help organizations to be transparent to the public, which increases public trust in algorithm use (Floridi, 2020). However, the academic debate on this topic is still in its early stages, making it difficult to understand what algorithm registers are and what impact they might have, resulting in the following sub-question:

RQ3: *What is the nature of algorithm registers, and what are their implications?*

To address this question, **Chapter 4** provides a theoretical understanding and an empirical mapping of the different factors that shape algorithm registers and their implications. Experiences from the Netherlands, a leading country in adopting public algorithm registers, show that a great diversity in algorithm registers exists. These experiences also show that many difficult choices must be made about what an algorithm register is, what objectives it tries to reach and how it should be designed. To create a useful and valuable algorithm register, these choices must be aligned. Furthermore, the findings show that the reality is nuanced regarding the implications of these registers. While public organizations indeed selectively disclose information and registers are currently not found useful by societal watchdogs and oversight authorities, there are also dynamics that could contribute to a more responsible use of algorithms by public organizations. For instance, the process of registering algorithms forces organizations to critically evaluate their algorithmic processes, potentially creating a disciplining effect. The chapter argues that algorithm registers are currently not a meaningful tool for transparency, but a meaningful box-ticking exercise for public organizations.

Macro-level

Previous research shows that algorithmic transparency is viewed as a potential solution to increase trust in algorithmic policing practices. Transparency can help address important concerns that influence trust in predictive policing. These concerns include, for instance, the use of biased data and the lack of accountability for algorithmic decisions. By integrating algorithms into organizations in such a way that outsiders can observe how these algorithms work and perform, these issues could be mitigated. While empirical studies mainly focus on providing transparency about how specific algorithms function, several scholars argue that transparency about the context in which algorithms operate is equally important. This highlights the importance of a broader approach to transparency: one that encompasses not only the functioning of the algorithms themselves, but also the institutional context (such as laws and regulations) in which they are applied. However, there is a lack of empirical research on providing transparency about the institutional context in which algorithms are embedded, leading to the following sub-question:

RQ4: *To what extent does institutional transparency affect citizen trust in predictive policing?*

Chapter 5 addresses this question by investigating whether two dimensions of institutional transparency, transparency about legislation and external monitoring, can strengthen trust in predictive policing. In doing so, the chapter provides a theoretical and methodological framework for conceptualizing and investigating institutional transparency in the context of algorithm use by governments. The findings in Chapter 5 show that disclosing information about how compliance with legislation on the use of predictive algorithms by the police is monitored by an external oversight body significantly increases citizen trust. However, providing information about the legislation itself does not have the same effect. This highlights the importance of transparency about external oversight in safeguarding citizen trust in the police's use of predictive algorithms.

Table 11.1 presents the main findings from Chapters 2-5 regarding the influence of transparency on citizen trust in the use of algorithms by governments.

Table 11.1 Main findings regarding the influence of transparency on citizen trust in the use of algorithms by governments

Level	Type of transparency	Definition	Main findings
Micro	Algorithmic transparency	Information about how an algorithm functions	▪ Explanations about how an algorithm functions and/or how it comes to its output significantly increase citizen trust in the algorithm.
Meso	Organizational transparency	Information about the organizational embedding of algorithms	▪ Partial or selective disclosure of information about the organizational embedding of algorithms can negatively impact trust in the use of algorithms by governments. ▪ The requirement to provide organizational transparency can foster a disciplining effect, as organizations have to critically asses their algorithmic decision-making practices.
Macro	Institutional transparency	Information about the institutional embedding of algorithms	▪ Information about the rules and regulations regarding an organization's use of algorithms does *not* increase citizen trust. ▪ Information about the monitoring of an organizations' compliance to rules and regulations regarding their algorithm use significantly increases citizen trust.

Conclusions

By combining the main conclusions of the different sub studies, I can formulate an answer to the central research question of this dissertation:

***RQ**: How does transparency affect citizen trust in algorithm use by government organizations?*

The findings in this dissertation reveal two mechanisms through which algorithmic transparency can foster citizen trust: a communicative (direct) route and a disciplining (indirect) route. Both routes have been discussed in the transparency literature before, but this dissertation shows how they coexist in the context of algorithmic transparency and interact with each other at different levels. This illustrates that the relationship between algorithmic transparency and trust occurs not only within individual levels but also through an interconnectedness across the micro, meso and macro levels.

The first route concerns the *communicative function* of transparency. This involves what organizations communicate externally about their algorithms: explaining how they work and produce their outputs, and how they are embedded within organizational and institutional frameworks. However, an overemphasis on the communicative function of transparency carries the risk of it being used manipulatively, by sharing only information that fosters trust. On the other hand, complete disclosure risks causing information overload: more transparency does not always lead to more trust.

The second route involves the *disciplining function* of transparency. The requirement to provide transparency encourages public organizations to engage in critical reflections, which can lead to improved algorithmic decision-making processes (a disciplining effect) and to better substantiated algorithmic decisions. By considering in advance how to communicate the algorithm and its integration within organizational and institutional contexts, these processes become more carefully designed in preparation for potential public exposure.

In the short term, the communicative function of transparency could *directly* foster trust by informing citizens and other relevant stakeholders about how the algorithms function, and how they are organizationally and institutionally embedded. In the long term, the disciplinary function of transparency could *indirectly* strengthen trust by encouraging public organizations to continually evaluate and refine their algorithmic practices. This dual focus, on both effective communication and rigorous internal reflection, can contribute to trustworthy use of algorithms by government organizations.

Finally, the findings in this dissertation demonstrate that the relationship between transparency and trust occurs not only within individual levels but also through an *interconnectedness* across the micro, meso and macro levels. Transparency at the macro level can, for example, influence trust at the meso level. Distinguishing between micro, meso and macro levels is relevant, as it offers both theoretical and practical insights into how transparency can be operationalized in theory and implemented in practice.

Scientific contributions

In academic and societal debates, there seems to be an overly optimistic view that making government algorithms transparent will automatically lead to trust. This belief is rooted in the broader discourse on transparency, where advocates of algorithmic transparency assume that once the inner workings of algorithms are made visible and understandable, public skepticism and mistrust will decrease or even disappear. This dissertation challenges and refines the naive, optimistic view, offering a more nuanced understanding of algorithmic transparency, by investigating how transparency works across three levels: algorithmic, organizational and institutional. These insights contribute to two significant bodies of literature: the literature on government transparency and the literature on algorithmization.

Contributions to the transparency literature

This dissertation contributes to the transparency literature by addressing *how* transparency can be provided and *what* it achieves. First, it demonstrates that providing meaningful transparency, especially about algorithmic decision-making, is complex and requires efforts at micro, meso, and macro levels. The findings from the different studies in this dissertation provide insights into how to design transparency initiatives that enhance trust in algorithm use by public sector organizations. Second, the dissertation shows that transparency does not automatically lead to trust. Trust depends on the type and amount of information provided. For instance, transparency that appears selective or insincere (e.g., "openwashing") can erode trust, as can excessive or overly complex information. These findings challenge the assumption that more transparency is always better and highlight the importance of tailored transparency strategies.

Contributions to the algorithmization literature

This dissertation also contributes to the algorithmization literature. First, it emphasizes the danger of transparency turning into a superficial box-ticking exercise. Continuous monitoring and regular updates are necessary to prevent superficial transparency and ensure accountability. Second, the dissertation proposes procedural transparency as a practical solution when full disclosure is constrained by confidentiality, such as in certain policing practices. By revealing the steps in decision-making (without disclosing sensitive outcomes) organizations can foster trust while protecting operational practices. Still, this approach must be applied cautiously to avoid being misused as a blanket justification for secrecy.

Recommendations for practice

This dissertation offers three recommendations for public sector organizations to enhance citizen trust in use of algorithms through transparency practices.

Recommendation 1: Embed algorithmic transparency "by design"

Algorithms are here to stay in public service delivery, making the trustworthy use of algorithms essential. However, not all forms of transparency build trust, and effective transparency is often difficult to achieve. On the *micro level*, public organizations should embed transparency into their algorithm design ("transparency-by-design") and ensure that their algorithmic decisions are clearly explained. This can, for example, be done through a rationale explanation, in which the reasoning behind a specific decision (such as the weighing of certain characteristics or individual circumstances) is explained.

Recommendation 2: Tailor transparency towards a specific audience

At the *meso level*, the dissertation calls for targeted transparency strategies tailored to the needs of different audiences. A one-size-fits-all approach can lead to mistrust if target group-specific expectations are not met. Transparency should empower stakeholders to understand and challenge algorithmic decisions. Politicians also have a role in using available transparency tools, like algorithm registers, to hold organizations accountable. This supports the disciplining function of transparency.

Recommendation 3: Strengthen and clarify oversight of transparency standards

On the *macro level*, continuous and proactive oversight is crucial. Regulators must define clear transparency standards to prevent checkbox compliance and ensure meaningful disclosure. Oversight should not be limited to algorithms alone but also encompass the organizational and institutional embedding of algorithms. Oversight at the organizational level can, for instance, include monitoring who is involved in the development of an algorithm or the organization's policies regarding algorithms. Oversight of the institutional embedding of algorithms involves, for example, monitoring compliance with laws and regulations concerning the (use of) algorithms by public organizations. Finally, public organizations must raise citizen awareness of oversight mechanisms. When citizens are aware that government use of algorithms is being actively monitored, it helps build their trust.

Samenvatting in het Nederlands

Algoritmes spelen een steeds prominentere rol in verschillende domeinen van publieke dienstverlening, variërend van de gezondheidszorg tot het strafrecht. Het is belangrijk dat het gebruik van algoritmes op een verantwoorde en betrouwbare manier plaatsvindt, om zorgen hierover (zoals bevooroordeelde, oneerlijke of ondoorzichtige besluitvorming) weg te nemen, en om de waarde die algoritmes kunnen bieden aan de publieke sector (zoals efficiëntie, productiviteit en veiligheid) te kunnen benutten. Transparantie wordt vaak genoemd als een belangrijke factor voor het opbouwen, behouden en vergroten van het vertrouwen van burgers in het gebruik van algoritmes door overheden.

Desalniettemin bestaat er een voortdurend debat over wat transparantie in de praktijk precies zou moeten inhouden en hoe dit daadwerkelijk het vertrouwen beïnvloedt. Deze dissertatie heeft daarom als doel om te verduidelijken hoe transparantie over het gebruik van algoritmes door organisaties in de publieke sector het vertrouwen van burgers beïnvloedt. De volgende onderzoeksvraag (OV) staat centraal in deze dissertatie:

OV: *Hoe beïnvloedt transparantie het vertrouwen van burgers in het gebruik van algoritmes door overheidsorganisaties?*

Het beantwoorden van deze hoofdvraag biedt de mogelijkheid om vier tekortkomingen in de literatuur te adresseren, wat leidt tot de formulering van vier onderliggende deelvragen. Deze tekortkomingen, de daaruit voortvloeiende deelvragen en de verkregen inzichten worden hieronder kort toegelicht.

In deze dissertatie dient de politieorganisatie als terugkerende context om bredere fenomenen rond het gebruik van algoritmes binnen de publieke sector te verkennen. Specifiek onderzoek ik twee bestaande politie-algoritmes: de Slimme Keuzehulp (hoofdstukken 2 en 3) en het Criminaliteits Anticipatie Systeem (hoofdstuk 5). Een uitzondering op deze politiefocus is hoofdstuk 4, dat zich richt op algoritmeregisters. Dit hoofdstuk draait om transparantie over de organisatorische inbedding van algoritmes binnen de bredere publieke sector. In alle hoofdstukken bespreek ik de bredere relevantie van mijn bevindingen voor organisaties in de publieke sector. Deze dissertatie biedt dan ook inzichten die niet alleen waardevol zijn voor de politie, maar ook voor andere publieke organisaties.

Multi-level perspectief op algoritmische transparantie en vertrouwen

De literatuur over algoritmische transparantie en vertrouwen laat zien dat de meeste studies deze relatie op vrij beperkte wijze hebben onderzocht. Ze richten zich vaak op de transparantie van de algoritmische toepassing zelf, terwijl bredere transparantieliteratuur suggereert dat transparantie ook een rol speelt op hogere niveaus, zoals het organisatorische en institutionele niveau. Dit omvat transparantie over organisatorisch beleid en institutionele regels die de juiste implementatie en het gebruik van algoritmes binnen de overheid sturen. Hoe deze bredere opvatting van transparantie kan worden toegepast om het vertrouwen in algoritmisch bestuur te versterken, blijft in de literatuur grotendeels onderbelicht. Dit leidt tot de eerste deelvraag:

OV1: *Hoe kunnen we de relatie tussen vertrouwen en algoritmische transparantie begrijpen op micro-, meso- en macroniveau?*

Hoofdstuk 2 gaat in op deze vraag door zowel algoritmische transparantie als vertrouwen in algoritmisch bestuur op micro-, meso- en macroniveau te conceptualiseren. Vervolgens wordt besproken hoe algoritmische transparantie bij zou kunnen dragen aan het opbouwen, behouden en versterken van het vertrouwen van burgers in algoritmisch bestuur. Dit resulteert in een conceptueel kader dat gebruikt kan worden om de relatie tussen transparantie en vertrouwen in algoritmisch bestuur empirisch te onderzoeken. Dit kader richt zich op transparantie op micro- (specifieke algoritmes), meso- (organisatorisch) en macro- (institutioneel) niveau, en beschrijft hoe deze niveaus samenhangen met het vertrouwen van burgers. Dit hoofdstuk laat zien dat het opbouwen, behouden en versterken van vertrouwen in algoritmisch bestuur niet alleen vraagt om transparantie over algoritmes zelf (micro), maar ook om transparantie over de manier waarop deze door overheidsorganisaties worden ingezet (meso) en over de regelgeving rondom het gebruik van algoritmes door deze organisaties (macro).

Microniveau

Bestaand onderzoek biedt tegenstrijdige inzichten over welke typen uitleg het meest effectief zijn in het versterken van het vertrouwen van burgers in algoritmische uitkomsten, zoals adviezen van algoritmische aanbevelingssystemen. Sommige onderzoeken stellen dat het geven van rechtvaardigingen voor individuele algoritmische beslissingen van belang is voor het creëren van vertrouwen in die beslissingen. Andere onderzoeken laten daarentegen zien dat het geven van informatie over de algoritmische besluitvormingsprocedure het meest effectief is om het vertrouwen in een algoritmische uitkomst te versterken. Al met al is er slechts in beperkte mate onderzoek gedaan naar welk type uitleg het meest effectief is in het vergroten van het vertrouwen van burgers, en over deze effectiviteit bestaat nog altijd discussie. Dit leidt tot de volgende deelvraag:

OV2: *Wat zijn de effecten van uitleg op het vertrouwen van burgers in algoritmische aanbevelingen?*

Hoofdstuk 3 behandelt deze vraag door in twee survey-experimenten de effecten te testen van verschillende vormen van uitleg over algoritmische aanbevelingen op het vertrouwen van burgers. Het eerste experiment bouwt voort op eerder onderzoek naar procedurele, rationele en gecombineerde uitleg; het tweede experiment richt zich op directieve uitleg. Hoewel er nuances zijn, laten de bevindingen zien dat het geven van uitleg over algoritmische aanbevelingen (in welke vorm dan ook) de vertrouwende houding (overtuigingen en intenties) en het op vertrouwen gebaseerde gedrag van burgers die digitale dienstverlening ontvangen, versterkt. Dit suggereert dat het vertrouwen in algoritmische aanbevelingen toeneemt wanneer burgers zien dat overheden moeite doen om een uitleg te geven, ongeacht de aard van deze uitleg.

Mesoniveau

In het wetenschappelijke debat bestaat onenigheid over de waarde van het nieuwe fenomeen algoritmeregisters. Sommige onderzoekers stellen dat overheden alleen politiek niet-gevoelige algoritmes openbaar maken en informatie verstrekken die beperkt bruikbaar is voor verantwoordingsdoeleinden. Daartegenover staat een meer optimistische visie: registers zouden, volgens andere onderzoekers, organisaties kunnen helpen transparant te zijn naar het publiek, wat het vertrouwen van burgers in het gebruik van algoritmes kan vergroten. Het academische debat over dit onderwerp bevindt zich echter nog in een vroeg stadium, waardoor het moeilijk is te begrijpen wat algoritmeregisters precies zijn en welke impact ze kunnen hebben. Dit leidt tot de volgende deelvraag:

OV3: *Wat is de aard van algoritmeregisters en wat zijn de implicaties ervan?*

Om deze vraag te beantwoorden, biedt **hoofdstuk 4** een theoretisch en empirisch overzicht van de verschillende kenmerken van algoritmeregisters en van hun implicaties. Ervaringen uit Nederland, een voorloper in het invoeren van publieke algoritmeregisters, tonen aan dat er een grote diversiteit aan algoritmeregisters bestaat. Ook laten deze ervaringen zien dat er veel lastige keuzes moeten worden gemaakt over wat een algoritmeregister precies is, welke doelstellingen het nastreeft en hoe het ontworpen wordt. Om een nuttig en waardevol algoritmeregister te creëren, moeten deze keuzes op elkaar worden afgestemd. Daarnaast laten de bevindingen zien dat de gevolgen van deze registers niet eenduidig zijn. Hoewel publieke organisaties inderdaad selectief informatie openbaar maken en registers momenteel door maatschappelijke waakhonden en toezichthouders niet als nuttig worden ervaren, zijn er ook dynamieken die kunnen bijdragen aan een verantwoordelijker gebruik van algoritmes door publieke organisaties. Zo dwingt het proces van het registreren van algoritmes organisaties om hun algoritmische processen kritisch te evalueren, wat mogelijk een disciplinerend effect heeft. In dit hoofdstuk wordt geconcludeerd dat algoritmeregisters momenteel geen betekenisvol instrument voor transparantie zijn, maar voor publieke organisaties vooral een betekenisvolle invuloefening.

Macroniveau

Eerder onderzoek laat zien dat algoritmische transparantie wordt gezien als een potentiële oplossing om het vertrouwen in algoritmische politiepraktijken te vergroten. Transparantie kan namelijk helpen bij het aanpakken van belangrijke zorgen die het vertrouwen in voorspellende politiezorg *(predictive policing)* beïnvloeden. Deze zorgen betreffen bijvoorbeeld het gebruik van bevooroordeelde data en het gebrek aan verantwoording over algoritmische beslissingen. Als buitenstaanders kunnen meekijken hoe deze algoritmes werken en presteren, kunnen deze problemen worden verminderd. Terwijl empirische studies zich voornamelijk richten op het bieden van transparantie over hoe specifieke algoritmes functioneren, beargumenteren verschillende onderzoekers dat transparantie over de context waarin algoritmes opereren even belangrijk is. Dit benadrukt het belang van een bredere benadering van transparantie, die niet alleen de werking van de algoritmes zelf omvat, maar ook institutionele context (zoals wet- en regelgeving) waarin zij worden toegepast. Het ontbreekt echter aan empirisch onderzoek naar transparantie over de institutionele context van algoritmes, wat leidt tot de volgende deelvraag:

OV4: *In hoeverre beïnvloedt institutionele transparantie het vertrouwen van burgers in voorspellende politiezorg?*

Hoofdstuk 5 adresseert deze vraag door te onderzoeken of twee dimensies van institutionele transparantie, namelijk transparantie over wetgeving en externe monitoring, het vertrouwen in voorspellende politiezorg kunnen versterken. Hiermee biedt het hoofdstuk een theoretisch en methodologisch kader voor het conceptualiseren en onderzoeken van institutionele transparantie over het gebruik van algoritmes door overheden. De bevindingen laten zien dat het openbaar maken van hoe de naleving van wetgeving rond het gebruik van voorspellende algoritmes door de politie wordt gemonitord door een externe toezichthouder het vertrouwen van burgers significant vergroot. Het verstrekken van informatie over die wetgeving zelf heeft dat effect echter niet. Dit onderstreept het belang van transparantie over externe controle bij het waarborgen van het vertrouwen van burgers in het politiegebruik van voorspellende algoritmes.

Tabel 12.1 presenteert de belangrijkste bevindingen uit Hoofdstukken 2–5 met betrekking tot de invloed van transparantie op het vertrouwen van burgers in het gebruik van algoritmen door overheden.

Tabel 12.1 Belangrijkste bevindingen over de invloed van transparantie op het vertrouwen van burgers in het gebruik van algoritmes door overheden

Niveau	**Type transparantie**	**Definitie**	**Belangrijkste bevindingen**
Micro	Algorithmische transparantie	Informatie over de werking van een algoritme	▪ Uitleg over hoe een algoritme werkt en/of waarom het tot zijn resultaat komt, verhoogt het vertrouwen van burgers in het algoritme significant.
Meso	Organisatorische transparantie	Informatie over de organisatorische inbedding van algoritmen	▪ Gedeeltelijke of selectieve openbaarmaking van informatie over de organisatorische inbedding van algoritmes kan het vertrouwen in het gebruik van algoritmes door overheden negatief beïnvloeden. ▪ De verplichting tot organisatorische transparantie kan een disciplinerend effect hebben, omdat organisaties hun eigen algoritmische besluitvormingspraktijken kritisch moeten beoordelen.
Macro	Institutionele transparantie	Informatie over de institutionele inbedding van algoritmen	▪ Informatie over de wet- en regelgeving met betrekking tot het gebruik van algoritmes door een organisatie verhoogt *niet* het vertrouwen van burgers. ▪ Informatie over de controle op de naleving van deze wet- en regelgeving door een organisatie verhoogt het vertrouwen van burgers significant.

Conclusies

Door de belangrijkste conclusies van de verschillende deelonderzoeken te combineren, kan ik een antwoord formuleren op de centrale onderzoeksvraag van deze dissertatie:

***OV:** Hoe beïnvloedt transparantie het vertrouwen van burgers in het gebruik van algoritmes door overheidsorganisaties?*

De bevindingen in deze dissertatie laten twee mechanismen zien waardoor algoritmische transparantie het vertrouwen van burgers kan bevorderen: een communicatieve (directe) route en een disciplinerende (indirecte) route. Beide routes zijn eerder besproken in de transparantieliteratuur, maar deze dissertatie laat zien hoe ze naast elkaar bestaan binnen de context van algoritmische transparantie en elkaar op verschillende niveaus beïnvloeden. Dit illustreert dat de relatie tussen algoritmische transparantie en vertrouwen niet alleen binnen individuele niveaus plaatsvindt, maar ook via een onderlinge verbondenheid tussen micro-, meso- en macroniveaus.

De eerste route betreft de *communicatieve functie* van transparantie. Dit gaat over wat organisaties extern communiceren over hun algoritmes: uitleggen hoe ze werken en hun output genereren, en hoe ze zijn ingebed binnen organisatorische en institutionele kaders. Een te grote nadruk op de communicatieve functie van transparantie brengt echter het risico met zich mee dat het manipulatief wordt ingezet, door alleen informatie te delen die vertrouwen bevordert. Aan de andere kant kan volledige openbaarmaking leiden tot een informatie *overload*: meer transparantie leidt niet altijd tot meer vertrouwen.

De tweede route betreft de *disciplinerende functie* van transparantie. De verplichting om transparant te zijn stimuleert publieke organisaties tot kritische reflectie, wat kan leiden tot verbeterde algoritmische besluitvormingsprocessen (een disciplinerend effect) en beter onderbouwde algoritmische beslissingen. Door vooraf na te denken over hoe het algoritme en de integratie ervan binnen organisatorische en institutionele contexten gecommuniceerd worden, worden deze processen zorgvuldiger ontworpen ter voorbereiding op mogelijke publieke openbaarmaking.

Op de korte termijn kan de communicatieve functie van transparantie *direct* vertrouwen bevorderen door burgers en andere relevante belanghebbenden te informeren over hoe algoritmes functioneren en hoe ze organisatorisch en institutioneel zijn ingebed. Op de lange termijn kan de disciplinerende functie van transparantie het vertrouwen *indirect* versterken door publieke organisaties aan te moedigen hun algoritmische praktijken voortdurend te evalueren en te verbeteren. Deze dubbele focus, zowel op effectieve communicatie als op zorgvuldige interne reflectie, kan bijdragen aan het vertrouwen van burgers in het gebruik van algoritmes door overheidsorganisaties.

Tot slot laten de bevindingen in deze dissertatie zien dat de relatie tussen transparantie en vertrouwen niet alleen binnen afzonderlijke niveaus plaatsvindt, maar ook via *onderlinge verbindingen* tussen het micro-, meso- en macroniveau. Transparantie op macroniveau kan bijvoorbeeld invloed hebben op vertrouwen op mesoniveau. Het onderscheid tussen micro-, meso- en macroniveaus is relevant omdat het zowel theoretische als praktische inzichten biedt in hoe transparantie in theorie kan worden geoperationaliseerd en in de praktijk kan worden geïmplementeerd.

Wetenschappelijke bijdragen

In academische en maatschappelijke debatten lijkt er een te optimistische opvatting te bestaan dat het transparant maken van overheidsalgoritmes automatisch tot vertrouwen leidt. Deze overtuiging is geworteld in het bredere discours over transparantie, waarbij voorstanders van algoritmische transparantie ervan uitgaan dat zodra de werking van algoritmes zichtbaar en begrijpelijk wordt gemaakt, publieke scepsis en wantrouwen zullen afnemen of zelfs verdwijnen. Dit proefschrift daagt deze naïeve, optimistische visie uit en biedt een genuanceerder begrip van algoritmische transparantie. Hierbij wordt onderzocht hoe transparantie werkt op drie niveaus: algoritmisch, organisatorisch en institutioneel. Deze inzichten dragen bij aan twee belangrijke wetenschappelijke literatuurstromingen: de literatuur over overheidstransparantie en de literatuur over algoritmisering.

Bijdragen aan de transparantieliteratuur

Dit proefschrift levert een bijdrage aan de transparantieliteratuur door te onderzoeken hoe transparantie kan worden geboden en wat het oplevert. Ten eerste toont het aan dat het bieden van betekenisvolle transparantie, vooral over algoritmische besluitvorming, complex is en inspanningen vereist op micro-, meso- en macroniveau. De bevindingen uit de verschillende studies in dit proefschrift geven inzicht in hoe transparantie-initiatieven kunnen worden ontworpen die het vertrouwen in het gebruik van algoritmes door publieke organisaties vergroten. Ten tweede laat het proefschrift zien dat transparantie niet automatisch leidt tot vertrouwen. Vertrouwen hangt af van het type en de hoeveelheid verstrekte informatie. Zo kan informatie die selectief of niet oprecht lijkt (bijvoorbeeld *"openwashing"*) het vertrouwen ondermijnen, net als te veel of te complexe informatie. Deze bevindingen dagen de veronderstelling uit dat meer transparantie altijd beter is en benadrukken het belang van maatwerk bij transparantiestrategieën.

Bijdragen aan de literatuur over algoritmisering

Daarnaast draagt dit proefschrift bij aan de literatuur over algoritmisering. Ten eerste benadrukt het proefschrift het risico dat transparantie over algoritmen (bijvoorbeeld via een register) een oppervlakkige invuloefening wordt. Continue monitoring en regelmatige updates zijn noodzakelijk om schijntransparantie te voorkomen en daadwerkelijke verantwoording te waarborgen. Ten tweede stelt het proefschrift procedurele transparantie voor als een praktische oplossing wanneer volledige openbaarmaking door vertrouwelijkheid wordt beperkt, zoals bij bepaalde politiepraktijken. Door de stappen in de besluitvorming te openbaren (zonder hierbij gevoelige uitkomsten prijs te geven) kunnen organisaties vertrouwen bevorderen en tegelijkertijd operationele processen beschermen. Deze aanpak moet echter met de nodige voorzichtigheid worden toegepast om te voorkomen dat deze wordt misbruikt als algemene rechtvaardiging voor geheimhouding.

Praktische aanbevelingen

Dit proefschrift doet drie aanbevelingen voor overheidsorganisaties om het vertrouwen van burgers in het gebruik van algoritmes via transparantie te bevorderen.

Aanbeveling 1: Integreer algoritmische transparantie in de ontwerpfase

Algoritmes zijn niet meer weg te denken uit de publieke dienstverlening, waardoor betrouwbaar gebruik ervan belangrijk is. Niet alle vormen van transparantie leiden echter tot vertrouwen, en effectieve transparantie is vaak moeilijk te realiseren. Op *microniveau* zouden overheidsorganisaties transparantie moeten inbouwen in het ontwerp van algoritmes *("transparency-by-design")* en ervoor moeten zorgen dat hun algoritmische beslissingen duidelijk worden uitgelegd. Dit kan bijvoorbeeld door middel van een inhoudelijke uitleg, waarin de redenering achter een specifieke beslissing, zoals de afweging van bepaalde kenmerken of individuele omstandigheden, wordt uitgelegd.

Aanbeveling 2: Stem transparantie af op een specifieke doelgroep

Op *mesoniveau* roept het proefschrift op tot gerichte transparantiestrategieën die zijn afgestemd op de behoeften van verschillende doelgroepen. Een *one-size-fits-all* aanpak kan wantrouwen opwekken als doelgroep-specifieke verwachtingen niet worden waargemaakt. Transparantie moet belanghebbenden in staat stellen algoritmische beslissingen te begrijpen en aan te vechten. Ook politici hebben een rol in het benutten van beschikbare transparantietools, zoals algoritmeregisters, om organisaties ter verantwoording te roepen. Dit ondersteunt de disciplinerende functie van transparantie.

Aanbeveling 3: Versterk en verduidelijk het toezicht op transparantiestandaarden

Op *macroniveau* is continu en proactief toezicht cruciaal. Toezichthouders moeten heldere transparantiestandaarden formuleren om een afvinkmentaliteit te voorkomen en zinvolle openbaarmaking te waarborgen. Toezicht moet zich niet beperken tot algoritmes, maar ook de organisatorische en institutionele inbedding van algoritmen omvatten. Bij toezicht op organisatorische inbedding kan gedacht worden aan controle op wie betrokken zijn bij de ontwikkeling van een algoritme of op het organisatiebeleid omtrent algoritmen. Toezicht op de institutionele inbedding van algoritmen betreft het monitoren van naleving van wet- en regelgeving omtrent het (gebruik van) algoritmen door publieke organisaties. Tot slot moeten overheidsorganisaties het bewustzijn van burgers over toezichtmechanismen vergroten. Wanneer burgers weten dat het gebruik van algoritmes door de overheid actief wordt gecontroleerd, draagt dit bij aan hun vertrouwen.

Dankwoord

Tijdens mijn promotietraject zijn er veel mensen geweest die een bijdrage hebben geleverd aan dit proefschrift. Ik ben heel dankbaar voor alle steun die ik heb gekregen, zowel op professioneel als persoonlijk vlak. Dit proefschrift had ik nooit kunnen schrijven zonder de hulp van velen, die ik hieronder persoonlijk wil bedanken.

Allereerst wil ik graag mijn begeleiders bedanken: Stephan, Albert en Floris.

Stephan, als mijn dagelijks begeleider heb jij een hele belangrijke rol gespeeld in mijn promotietraject. Wat ik ontzettend heb gewaardeerd is jouw manier van feedback geven. Niet alleen bracht je de knelpunten in mijn teksten op een constructieve manier naar voren, maar je gaf ook altijd suggesties hoe ik hiermee om kon gaan. Hierdoor bleef ik na het bespreken van mijn teksten niet verward of ontmoedigd achter. Op een pragmatische manier wist je voor elke situatie een oplossingsrichting te formuleren. En ook niet te vergeten: je was altijd goed voorbereid, las mijn stukken grondig door en maakte tijd voor me als ik ergens tegenaan liep. Je was heel benaderbaar en betrokken, en inhoudelijk een echte expert op het gebied van de theorieën en methoden die ik gebruikte (waar ik dankbaar gebruik van heb gemaakt). Een groot deel van mijn kennis over het opzetten en uitvoeren van experimenten heb ik aan jou te danken.

Albert, jij bent in staat om heel snel de rode draad of de grote lijnen te zien. Zelfs in mijn teksten die uit flarden, bulletpoints of vage ideeën bestonden, wist jij altijd precies tot de kern te komen. Tijdens het denk- en schrijfproces raakte ik soms verstrikt in mijn eigen onderzoeksrichting, maar jij hielp me om afstand te nemen en mijn werk te verbinden met bredere discussies in de wetenschappelijke literatuur. Ik heb veel van je geleerd, niet alleen op vakinhoudelijk gebied, maar ook bijvoorbeeld over hoe je je opstelt tegenover respondenten en hoe je je door het reviewproces navigeert. Ook ben ik dankbaar dat ik altijd bij je binnen kon lopen en dat je tijd voor me maakte om me te helpen met allerlei vragen op het gebied van onderzoek doen.

Floris, jouw scherpe oog voor details was van grote waarde tijdens mijn promotietraject. Jij zorgde ervoor dat er geen inconsistenties zaten in mijn overwegingen, redeneringen, of zelfs in woordkeuze. Daarnaast ben ik je erg dankbaar dat je mij al tijdens mijn master hebt geïntroduceerd in het Nationaal Politielab AI. Daarmee heb je de basis gelegd voor wat later een heel bijzonder promotietraject is geworden. Tijdens het promotieonderzoek was jouw inhoudelijke expertise op het gebied van algoritmen en de politie heel waardevol, en heb ik veel van je geleerd hierover. Bovendien zorgde je altijd voor een fijne en informele sfeer, zowel tijdens de ALGOPOL bijeenkomsten, als bij de bijeenkomsten en borrels van het Politielab.

Het Nationaal Politielab AI was een plek waar ik samen kon komen met onderzoekers uit diverse disciplines om onze studies naar het verantwoord gebruik van AI bij de politie te delen en te bespreken. Mijn dank gaat uit naar alle leden van het lab voor de inspirerende gesprekken, inzichten en samenwerking. Ook wil ik de rest van het ALGOPOL-team bedanken: Isabelle, Carlos, Merijn, José en Mirko. Het was een hele prettige ervaring om in dit team van onderzoekers te werken aan mijn promotieonderzoek. Isabelle, het was fijn om samen dit project te doorlopen, en al onze stappen en verwonderingen met elkaar te delen. Dankjewel voor je originele invalshoeken en ideeën! Verder wil ik graag de commissie bedanken voor de tijd en moeite de ze hebben gestoken in het lezen van dit proefschrift en hun deelname aan de verdediging.

A huge thanks to my colleagues from Rutgers University. First of all, thank you, Steve. Your daily walk-ins to my office were one of the highlights of my time at Rutgers. I really appreciated all the (life) advice you gave me. Thanks to you, I was able to navigate both the university and the public transportation systems in New Jersey and New York (which can be trickier than it looks!). Thank you for all the treats, the trip to the best hotdog place, and the tours of the neighborhood we lived in. Furthermore, I would like to thank the PhD students who gave me such a warm welcome and introduced me to SPAA. Thank you Ying, Jinah, Kayla, Monica and Mauricio (also for the fun hiking trip)! Finally, I want to thank my co-authors, Greg and Vishal. Without the two of you, I would not have come to Rutgers in the first place. Thank you for taking the time to work with me on a part of this dissertation!

Ook ben ik dankbaar voor de collega's bij USBO die me afgelopen jaren hebben geholpen op allerlei verschillende manieren. Esther, Odile en Lianne, bedankt voor jullie hulp bij al mijn vragen. Marcel, bedankt voor de lekkere broodjes en de gezelligheid! Bedankt Erna voor je (onderwijs)adviezen; ik heb veel van je geleerd over het begeleiden van studenten. En bedankt Marij, jij was niet alleen mijn BKO-tutor, maar ook een informele mentor bij wie ik altijd terecht kon als ik ergens tegenaan liep. Dankjewel voor je luisterend oor, scherpe vragen en waardevolle feedback tijdens mijn onderwijstraject en daarbuiten.

Daarnaast wil ik graag mijn kamergenoten bedanken. Door de jaren heen zijn er heel veel verschillende collega's in 0.11 geweest die een belangrijke rol hebben gespeeld in mijn promotietraject: Isabelle, Carlos, Roos, Jessy, Lauren, Sandra, Linde, Leonie en Resie. Jullie zorgden voor een fijne plek waar ik de hoogte- en dieptepunten, maar vooral ook de dagelijkse ditjes en datjes van mijn promotietraject kon delen.

Ik heb ontzettend veel gehad aan de steun, feedback en gezelligheid van andere AIO's. Toen ik tijdens corona begon aan dit traject waren andere promovendi de enige collega's die ik fysiek zag. Wij kregen gelukkig een uitzondering om toch op kantoor te werken. Het was bijzonder om deze tijd samen mee te maken. Bedankt Iris, Rosan, Michelle, Laura, Carlos, Isa, Henrico, Lars, Sheeling, Gaby en Lukas. Ik heb genoten van onze borrels, congresbezoeken, reizen (in Arizona!), Secret Santa activiteiten, wandelingen, AIO-platforms, en meer. Ook wil ik alle promovendi bedanken die daarna kwamen en mijn tijd op USBO bijzonder maakten. Zonder volledig te zijn, noem ik graag: Krista, Emma, Olaf, Rick, Lennart, Kees en Arjan.

Iris en Rosan, jullie wil ik nog extra in het zonnetje zetten. Rosan, wij begonnen tegelijk aan onze PhD's tijdens corona. Het feit dat alle vergaderingen en dataverzameling digitaal waren, leverde soms frustraties bij ons op en maakte ons lotgenoten. We gingen vaak lekker fietsen na werk om onze hoofden leeg te maken. Ook aten we geregeld samen, waarbij jij de heerlijkste gerechten kookte (met de Indonesische rijsttafel als hoogtepunt!). Dankjewel voor je nuchterheid (en tegelijk ook passie), de leuke schrijfweekenden en alle fijne gesprekken!

Iris, onze promotietrajecten liepen ook bijna parallel. Dit zorgde ervoor dat we elkaar goed begrepen, en elkaar konden helpen en motiveren tijdens elke fase van de PhD. Bijzonder belangrijk voor mij waren de dagelijkse wandelingen naar Vlaamsch Broodhuys, waar we (te dure) koffie haalden. Ik heb het even teruggezocht, en aan deze dagelijkse koffies (en taartjes wanneer er iets te vieren was) heb ik 494 euro uitgegeven! Dit was het helemaal waard: elke dag keek ik weer uit naar deze lekkere koffie en vooral onze wandeling daarna. Ook haalde ik veel motivatie uit onze biebdagen, waarbij we allebei hard aan de bak gingen met schrijven en onszelf trakteerden op een broodje Mario voor lunch. Bedankt dat je altijd voor me klaarstond de afgelopen jaren!

Naast de opbrengst van dit boek ben ik blij dat het promotietraject me zo'n mooie vriendschap met jullie heeft opgeleverd. In de afgelopen jaren heb ik zoveel kleine en grote momenten van mijn PhD met jullie gedeeld. Maar dat niet alleen, we deelden (vooral) ook lief en leed met elkaar. Ik had van te voren niet gedacht dat ik op meerdere vakanties zou gaan met collega's (waaronder kamperen op het erf bij Iris' ouders). Ik ben heel blij met jullie!

Zingen met Kwaya is een hele fijne uitlaatklep geweest de afgelopen jaren. Bedankt Hannah, Kila, Myrthe, Simone, Jaimy, en alle andere koorleden voor alle gezelligheid! Kila, naast vriendin was je ook een rolmodel voor mij. Ik heb veel van je geleerd over de academische wereld en hoe je je daarbinnen navigeert. Bedankt daarvoor! En Hannah, ik heb met heel veel plezier met jou de Valentijnoptredens geregeld; dat vond ik altijd het meest bijzondere optreden (waar ik elk jaar wel wat traantjes heb gelaten). Naast het regelen van optredens werden we ook erg close buiten koor om. Bedankt voor al je lieve steun tijdens mijn PhD. Als ik moe thuis kwam van een congres bood jij aan om een bakje eten langs te brengen of boodschappen voor me te doen. En toen ik koor miste omdat ik nog een les aan het voorbereiden was in de avond, appte je me om te vragen of je iets voor me kon doen (en als ik niet reageerde, appte je zelfs Emiel om te checken of alles goed met me ging). En niet te vergeten: bedankt voor het helpen met mijn drie (!) verhuizingen tijdens mijn PhD. Je bent een topper!

Tijdens het promotietraject zijn er heel veel vrienden die, elk op hun eigen manier, een belangrijke rol hebben gespeeld in mijn leven. Allereerst Mirron, met wie ik meer dan de helft van mijn promotietraject heb samengewoond (en zelfs 7 jaar in totaal). Dankjewel voor alles Mir: onze koffietjes in de zon voor de deur, de fietstochten, avondwandelingen, uitgebreide diners in de tuin tijdens corona, saunadagjes, series bingen, en meer. Tijdens corona werden we praktisch collega's, omdat we zoveel met elkaar deelden dat we helemaal op de hoogte waren van elkaars werk en we elkaar als sparringpartners gebruikten. Ik kijk met zo veel plezier terug op onze tijd als huisgenoten. Ik ben dan ook ontzettend blij dat je samen met Rick in het appartementencomplex naast ons komt wonen!

Thanks Jessie and Josh for making my PhD-time more fun during our trips together in Switzerland, Taiwan and the Netherlands. Jessie, I am very thankful that we met each other in Florida 8 years ago! Our friendship means a lot me. And thanks Josh for taking me along to your university in Zurich and introducing me to your colleagues. I really enjoyed our conversations. Celebrating the end of my PhD period with you two in Taiwan was amazing!

Ook wil ik een paar vriendengroepen bedanken. Mijn (roei)ploeg: onze wekelijkse etentjes op de maandagavond waren altijd een hele fijne start van de week. Bedankt voor al het lekkere eten, jullie gezelligheid, advies, nuchterheid, de leuke verhalen, kerstdiners, ploegvakanties, borrels en veel meer. Mijn vriendinnen van de middelbare school: ik denk niet dat ik hier had gestaan als ik niet door jullie was omringd op het CGU. Jullie nieuwsgierigheid werkt aanstekelijk. Julia en Anouk, ik ben ontzettend blij met onze vriendschap vanaf de brugklas. Heel veel dank voor jullie steun de afgelopen jaren!

Mijn studievrienden: bedankt voor jullie humor en interesse. Jullie waren áltijd geïnteresseerd in waar ik mee bezig was en benaderden elke situatie met een flinke dosis humor. Door jullie ervaring met USBO en eigen promotietrajecten begrepen jullie als geen ander waar ik mee bezig was. Ik kijk elke keer uit naar onze eetclub. Laura, ik wil jou nog extra bedanken voor al je steun tijdens mijn promotietraject. Ik heb heel veel aan je gehad. Ik kijk met veel plezier terug op alle leuke dingen die we afgelopen jaren gedaan en gezien hebben (diners, filmavonden, Oerol, oud en nieuw, Koningsdag, etc.). En ik vond onze biebdagen, waarin we beide aan ons promotieonderzoek werkten, erg fijn. Dankjewel voor alles!

En bedankt lieve Anne en Lars, en Eline. Ook al zien we elkaar niet veel, als we elkaar zien is het altijd fijn. Ik voel me gezegend met zoveel lieve vrienden!

En natuurlijk heel veel dank aan mijn lieve familie. Bedankt mama, papa, Marleen, Thymo, Renée en Martin, voor al jullie steun. Jullie hebben het hele proces van dichtbij meegemaakt en meegeleefd met alle hoogte- en dieptepunten. Mam en pap, ik was nooit aan een promotietraject begonnen zonder jullie. Jullie betrokkenheid, mama's nieuwsgierigheid, en papa's leergierigheid vormden een inspirerend voorbeeld dat goed van pas kwam tijdens dit traject. Tegelijkertijd deden de relativerende gesprekken met mama me erg goed, zeker op momenten dat ik het even niet meer zag zitten ("stop er maar gerust mee"). Marleen, ik ben ontzettend blij dat ik je altijd kon appen of bellen voor advies. Je zorgde daarnaast regelmatig voor de nodige afleiding met leuke verhalen en gezellige uitjes. Renée, jouw ervaringen bij en kennis over de politie waren waardevol om de politie als organisatie beter te begrijpen. Dankjewel dat jullie er altijd voor mij waren! Ook wil ik mijn lieve schoonfamilie bedanken: Edward, Renata en Dyana. Bedankt voor de warme manier waarop jullie mij in de familie hebben opgenomen en voor jullie interesse in mijn werk.

Tot slot, lieve Emiel, DANKJEWEL! Je was tijdens mijn hele promotietraject mijn persoonlijke "hype man". Bij elke gebeurtenis, hoe klein ook, benadrukte je steeds weer hoe trots je op me was en hoe goed ik het had gedaan. Je zorgde ervoor dat we alle belangrijke mijlpalen vierden. Ook stond je altijd voor me klaar als het minder ging. Ik kon mijn ei bij je kwijt, je luisterde naar me, kookte voor me en vrolijkte me op.

Een absoluut hoogtepunt van de afgelopen jaren was onze tijd in New York. Wat was het bijzonder om (voor het eerst!) samen te wonen op zo'n fantastische plek. Ik ben heel dankbaar dat we daar samen naartoe konden gaan, allebei een leuke werkplek hadden en zoveel mooie dingen hebben gedaan en gezien.

Bedankt voor je onvoorwaardelijke liefde, steun, enthousiasme en humor. Ik geniet van elk moment samen met jou!

About the author

Esther Nieuwenhuizen (1996) studied Governance and Organizational Science *(Bestuurs- en Organisatiewetenschap)* at Utrecht University, where she obtained her bachelor's degree in 2018. She completed a master's degree in Public Management at Utrecht University in 2020. During her master's program, her interest in the transparent use of algorithms by public organizations grew. For her master's thesis, Esther studied the relationship between explanations and citizen trust in algorithmic recommendations by the police. She received the H.A. Brasz thesis award for the best public administration thesis of the Netherlands.

After graduating, she remained at the Utrecht University School of Governance to start her PhD research. She was awarded both a Fulbright Scholarship and a Visiting Scholarship (Utrecht University) for a four-month research visit at Rutgers University-Newark in the United States. Alongside her research, she took several PhD courses at the Netherlands Institute of Governance. Furthermore, Esther has invested considerable time in teaching and developing her teaching skills, which resulted in her obtaining a university teaching qualification *(basiskwalificatie onderwijs)*. Moreover, she served as a member of the departmental research council. As of June 2025, Esther works as a Senior Police Inspector at the Inspectorate of Justice and Security.