Beyond the "Black Box":

Enabling Meaningful Transparency of Algorithmic Decision-Making Systems through Public Registers

by

Maya Murad

B.A. Economics (2016)

American University of Beirut

Submitted to the Integrated Design and Management Program
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering and Management

at the

Massachusetts Institute of Technology

June 2021

Signature of Author_____

Integrated Design and Management Program
May 14, 2021

Certified by_____

Leigh Hafrey
Thesis Supervisor
Senior Lecturer, MIT Sloan School of Management

Accepted by _____

Matthew S. Kressy
Director, Integrated Design & Management Program

Beyond the “Black Box”:

Enabling Meaningful Transparency of Algorithmic Decision-Making Systems through Public Registers

by

Maya Murad

Submitted to the Integrated Design and Management Program
in Partial Fulfillment of the Requirements for the Degree of Master of Science in Engineering and Management

May 14, 2021

ABSTRACT

Deployments of algorithmic decision-making systems (ADMs) by the public sector have been plagued with opacity. There is a baseline lack of visibility of the context and purpose of the ADM system as well as its potential risks to individuals and collective goods. In many cases, citizens are unaware of the very existence of algorithmic systems that they interact with or that help decide their access to benefits or influence policing. Moreover, disclosures concerning algorithmic systems often take place when their shortcomings (potential harms) are inadvertently exposed, often through the work of public interest groups.

Given the increasing adoption of algorithmic systems to automate decisions and services in the public sector, there is a need to operationalize transparency requirements to enable better accountability. While algorithmic transparency can take on many forms, this thesis mainly focuses on the role of public ADM registers in enabling meaningful transparency to the public. In the past year, at least five local governments have launched their very first ADM registers. Drawing upon these early experiences, relevant stakeholder interviews and specifically considering Amsterdam as a case study, we attempt to formalize the concept of a register as both a standardized and interpretable ADM disclosure mechanism, as well as a governance framework that enables coordination between a number of stakeholders to provide of transparency to the public. We also propose models through which public interest groups and civilians can be engaged in the creation, development and launch of public ADM systems through the governance of a register, and outline key benefits and limitations of such models.

Thesis Supervisor: Leigh Hafrey
Title: Senior Lecturer at MIT Sloan School of Management

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Algorithms are increasingly being embraced by the public sector as a means to automate decision making. The main perceived benefits by the government to adopt algorithmic decision making systems (ADMs) is to improve public sector operational efficiency, capture savings and enable more accurate decision making. Applications of ADMs in the public sector to date include prioritizing access to services, deciding benefit entitlements and enabling real time monitoring, risk flagging and predictive policing.

There is a growing body of literature documenting the failure points of implementations of ADMs in the public sector such as cases of bias and discrimination, privacy infringements among others, as discussed in Chapter 2. Deployments of ADM systems by the public sector have been plagued with opacity, with little to no information shared with the public or other regulatory bodies on the context, purpose, performance, human control and reliability of these systems. Many public authorities do not disclose altogether their use of ADM systems. Moreover, regulation related to the use and oversight of ADM systems is still in its infancy. These factors contribute to the general public's mistrust of ADMs and raises concerns about the accountability of public ADM systems.

Deploying ethical and reliable ADM systems requires adequate transparency mechanisms that enable accountability. The transparency impediment can be overcome through the provision of relevant disclosures related to the system and the development of an open governance model around it. The concept of deploying a public ADM register that documents all the algorithmic systems in use by the public sector and makes relevant disclosures to the general public has been gaining traction recently. In 2020, several cities launched their own ADM registers including Amsterdam, Helsinki, Nantes, Antibes and New York City. These registers are still quite novel and have not been assessed to date. Moreover, there is currently no comprehensively mandated disclosures regarding ADM use and all information shared is mainly self-reported by the public sector.

The main goal of this thesis is to explore the role of public registers in providing meaningful transparency for ADM systems deployed by the public sector. It also aims to formalize the concept of a public register as a standardized and interpretable disclosure mechanism that enhances the understanding of ADM systems and provides visibility on their impact and reach, as well as outline the need for an ADM governance mechanism that enables transparency. Finally, this work explores potential mechanisms through which civil society can be engaged in the ADM disclosure process to ensure it meets its goals.

In Chapter 2, we review existing literature on the need for algorithmic transparency as well as its provisioning in different regulatory contexts. We find that most existing regulation and governance models do not adequately provide for the protection of individual and collective harms resulting from ADM system deployment in the public sector. A baseline of transparency is the first step in terms of driving accountability and ethical ADM system deployment that minimizes harms and protects rights.

In Chapter 3, we explore different mechanisms through which algorithmic transparency has been operationalized to date. We argue the need for ADM registers to tie together different disclosure requirements and modalities to provide a comprehensive view of the system in a way that is interoperable. This section includes a draft of a comprehensive disclosures framework based on a summary of existing literature.

In Chapter 4, we provide an overview of existing ADM registers deployed to date as well as attempt to formalize the concept of an ADM framework. Leveraging the disclosure framework from Chapter 3, as well inputs from expert stakeholders in algorithmic transparency, policy, governance and advocacy, we attempt to formalize the concept, purpose, design and governance of a public register.

In Chapter 5, we consider Amsterdam's AI register as a case study. Deployed in September 2020 in partnership with the city of Helsinki, Amsterdam's register constitutes the first version of an ADM register launched by a public authority. What is interesting about this case study is that the register was launched despite no existing mandate requiring algorithmic transparency. We explore in this section the context that led to the creation of the register, the key design choices that ensued as well as its reception by its citizens.

In Chapter 6, we argue the need to consider civilian engagement in operationalizing algorithmic transparency. We specifically consider different possible engagement models of public interest groups through the governance of an ADM system register.

Finally in Chapter 7 and 8, we summarize key findings for public authorities and policy groups that are interested in operationalizing algorithmic transparency in a register and outline open points and recommended areas of future work.

CHAPTER 2

TRANPARENCY AS A BASELINE FOR ADM ACCOUNTABILITY

In this chapter, we explore the need for improved transparency mechanisms to enable the accountability of algorithmic decision making systems in the public sector.

Putting accountability back in the algorithm

The scope of this research focuses on algorithmic decision making systems deployed by the public sector. The choice of use of this terminology is deliberate and reflects the need for a transparency mechanism at the 'unit' level of a decision-making system that is enabled by algorithmic processes. An 'algorithm', in its simplest expression is "a set of rules that precisely define a sequence of operations" (Stone, 1971). This may include a wide range of operations with different levels of complexity that can include *rule-based* approaches as well as *machine learning* techniques. Such operations, as simple or as complex as they may be, proceduralize the decision-making process, introduce a level of automation in the system, and have implications on data collection, handling and storage. In this context, we are concerned by algorithms that enable decision making systems in the public sector. The system itself might have portions that are fully administered by human beings, or are under human oversight. We argue in our research that the algorithmic decision making system as a whole should be considered as the baseline unit for transparency provision. *The appendix includes a glossary of key terms used*.

Another terminology concern that is relevant to address is the colloquial interpretation of the term 'algorithm' that obscures the accountability of the decision making system the algorithm in question is a part of (Lum & Chowdhury, 2021). The term suggests to the public that the system is complex and devoid of human accountability. It is important to clarify that algorithms used by the public sector consist of decision making and/or support tools that were designed by humans and for whose impact and harms the commissioning authority should be accountable for.

ADM regulation to safeguard from individual and collective harms

There are different types of harms that can be caused by the poorly regulated deployment of ADM systems. These harms can manifest themselves at the individual or collective level and have different degrees of severity, as well as scale of impact (i.e. number of individuals affected). The table below includes a summary of potential harms discussed in literature.

[TABLE 1] SUMMARY OF POTENTIAL HARMS THAT CAN BE CAUSED BY ADM SYSTEMS

TYPE OF HARM	LIST OF HARMS	SOURCE
Individual	**[a] Bias and discrimination** Can manifest through the reproduction of existing undue biases and patterns of marginalization, inequality and discrimination already present in the data processed as well as the overarching design of the system.	(Alan Turing Institute, 2019; Bejtullahu-Michalopoulos & Florin, 2018)

	[b] Denial of autonomy, recourse, and rights Can occur when systems are automated with no clear outlet for objection, recourse or feedback.	(Alan Turing Institute, 2019; Bejtullahu-Michalopoulos & Florin, 2018)
	[c] Non-transparent, unexplainable or unjustifiable outcomes Certain types of algorithmic processes such as deep learning models are less interpretable than simple statistical methods, leading to outputs that are not explainable; this may be problematic when the model is not learning in a representative manner and reproducing biases.	(Alan Turing Institute, 2019; Bejtullahu-Michalopoulos & Florin, 2018)
	[d] Threats to data protection and privacy Can appear at different stages of the design and deployment of ADM systems, data can be collected, stored or handled in ways that violate data consent and privacy.	(Alan Turing Institute, 2019; Bejtullahu-Michalopoulos & Florin, 2018)
	[e] Unreliable, unsafe, or poor quality outcomes Can arise from poorly designed systems, or systems that perform well on training data but not during deployment due to data mismatch	(Alan Turing Institute, 2019; Bejtullahu-Michalopoulos & Florin, 2018)
Collective	**[f] Isolation and disintegration of social connection** Can manifest itself in the polarization of social dynamics and limiting exposure to other worldviews	(Alan Turing Institute, 2019)
	[j] Excessive surveillance and excessive social control Can take place through undue influence exerted by actors owning and deploying these types of systems.	(Bejtullahu-Michalopoulos & Florin, 2018)
	[k] Manipulation or malignant use Can appear in the form of interference with democratic processes or violation of human rights.	(Bejtullahu-Michalopoulos & Florin, 2018)

There are well-documented examples that illustrate the individual harms that can be caused by the poorly regulated deployment of ADM systems, as illustrated below. In these examples, the harms were discovered after the systems were deployed, often adversely impacting a number of individuals before being stopped or mitigated. The discovery of these harms triggers added scrutiny and often results in the uncovering of information that helps shed light onto the functioning of the algorithmic systems, as well as their performance and impact, that otherwise would have remained shielded from the general public.

Individual harms materialize generally quicker and more visibly than collective harms (Mittelstadt, 2017). Collective harms include harms experienced by entire segments of civil society and can adversely impact collective goods such as democratic institutions. It is important to note that collective harms are often overlooked, as they are less visible and less likely to trigger redress actions.

Examples of individual harms caused by ADM systems:

- **[Harms a, b]** In the context of the COVID-19 pandemic in the UK, secondary education examinations were canceled and an algorithm designed by the regulator Ofqual was assigned the task of determining students' final grades based on teacher predictions and with a goal to avoid grade inflation (Hughes, 2020). The algorithm's results were scrapped after public outcry as it was found to disproportionately downgrade students from poorer backgrounds, ignoring their individual performance (Lee, 2020)**.** It is also important to note that the algorithm in question is a simple statistical model that did perform as intended and was a product of human-led decisions (Burgess, 2020). This is an example of how the term 'algorithm' can obscure accountability.
- **[Harms a, b, e]** The COMPAS recidivism algorithm is a commercial tool created by Northpointe that was used in U.S. courts to assess the likelihood of a defendant becoming a recidivist. An

analysis of the tool's performance found that black defendants were far more likely than white defendants to be incorrectly flagged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk (Angwin et al., 2016).

- **[Harms e]** In Chicago, a data mining tool was developed to identify children at risk for serious injury or death. The algorithm mined existing data at the Department of Children and Family Services and assigned a score to children who were the subject of an abuse allegation. The algorithm was later found to be unreliable with an overwhelming number of flagged false-positive cases, as well as an increase in the number of children deaths going undetected. The DCFS decided to end the program two years after its launch (Marx, 2017).
- **[Harms d]** An NHS 'datastore' was proposed to be created in the UK to centralize patient data and leverage ADM capabilities through outsourcing to big tech providers (Gould et al., 2020). Civil society representatives (Foxglove and Open Democracy) pressured the government to provide transparency on the procurement contracts, which later on were revealed to grant intellectual property rights to selected big tech firms and allow them to train their models on a large pool of unreleased health-related data (Fitzgerald & Crider, 2020).

Recent failures and harms caused by ADMs raised the need for improved governance mechanisms which should encompass legal, ethical and procedural requirements and entail an array of different regulatory approaches (Bejtullahu-Michalopoulos & Florin, 2018). Regulators are therefore faced with the challenge of enabling the beneficial outcomes of ADM systems while minimizing the risk of potential harm.

Existing regulation is inadequate to drive ADM accountability

Accountability can be thought of "primarily as a legal and ethical obligation on an individual or organization to account for its activities, accept responsibility for them, and to disclose the results in a transparent manner" (Bejtullahu-Michalopoulos & Florin, 2018). In this context, accountability should translate into regulatory and legal tools whose goals are to enable ethical deployment of ADM systems and safeguard both individuals and collectives from their potential harm. Accountability mechanisms can take on a variety of forms that can be supply and/or demand-driven. *Supply-side accountability* reflects the measures the supplier of the algorithmic system implements to enable better accountability. In the context of the public sector, supply-side measures can take on the form of internal controls and protocols, the creation of designated institutions and oversight bodies, among others. *Demand-side accountability* reflects the many forms through which stakeholders outside the public sector can influence better accountability of the algorithmic systems in use. These may include investigative reporting, public perception risks, protests, among others.

Existing regulation to drive ADM supply-side accountability include:

- In the EU, the primary basis of regulation of data and data processing is the European General Data Protection Regulation (GDPR) (Radley-Gardner et al., 2016). GPDR's primary goal is to enable individual control over data and has provisions concerning the right of access to personal data and how it is processed, the right to be forgotten and data erasure and the right to object processing of personal data. A study by European Parliamentary Research Service has indicated that the GDPR is "not likely to be sufficient" to adequately safeguard the accountability of ADM systems (Koene et al., 2019). Articles 22 "right to an explanation" and 13-15 "rights to "meaningful information about the logic involved in automated decisions" and 35 "data protection impact assessment" have a narrow focus on personal data and are restricted in applicability.

- Beyond GDPR, France passed its "Law for a Digital Republic" (*LOI N° 2016-1321*) which introduces new obligations regarding administrative ADM and goes beyond the scope of 'automated processing' to include decisions made based on algorithmic processing. Administrations (i.e. the public sector) are required to inform persons affected by ADMs and must, upon request, communicate the rules based upon which the system functions in an intelligible manner. To date, few local administrations in France have complied with this law. The cities of Nantes and Antibes have both released in the past year an algorithmic system log to the general public *(further discussed in Chapter 4).*

- In the UK, two main regulatory tools help regulate data processing and ADMs, data protection impact assessments (DPIAs) and equality impact assessments. DPIAs disclose information relevant to the functioning of the data processing system, including data fields and sources, the system's function within broader administrative processes, the responsible officials, and the effects and legal basis for data processing. DPIAs are mandatory for systems flagged as high risk. Equality assessments are associated with the "2010 Equality Act", whose goal is to safeguard against the risk of discrimination (Equality and Human Rights Commission, 2010). Both of these instruments were found to have shortfalls in their direct mandate as well as lagged in terms of safeguarding against broader harms (Ada Lovelace Institute, 2020b)

- In the US, privacy impact assessments are carried out by the Department of Homeland Security to identify and mitigate risk relating to personally identifiable information (US Department of Homeland Security, 2009). It is important to note that, unlike the EU, privacy is not a constitutional right in the US. Privacy legislation differs widely, most of the laws enacted or being considered have an individual consumer focus such as the California Data Privacy Act (State of California Department of Justice, 2018). Existing legislation does not have provisions to protect against the broader class of individual harms ADMs can produce. In that regard, the "Algorithmic Accountability Act of 2019" was introduced to propose regulations on high risk ADM systems requiring them to undergo algorithmic impact assessments (Booker & Wyden, 2019). The bill did not undergo a vote.

ADM regulation is still in its infancy and requires significant work to be effectively put into practice. The starting point for most of the regulatory tools outlined above is data privacy and there is a need for a more consistent approach in terms of recognizing, assessing and safeguarding against both individual and collective harms. Transparency has been a critical enabler of data privacy mandates and we expect these stipulations to be considered at an algorithmic decision making system level.

In April 2021, the European commission released its proposal for an EU-wide "Artificial Intelligence Act" to enable the development of secure, ethical and trustworthy artificial intelligence within the single market (European Commission, 2021). The proposed legislation adopts a horizontal approach across the ADM application space and outlines a proportionate risk-based approach as well as codes of conduct for non-high-risk systems. This proposal's scope goes significantly beyond the existing legislation and constitutes a starting premise for comprehensive ADM accountability, as it defines frameworks of disclosures and oversight mechanisms. The basis of this proposal is a categorization of ADM systems into four potential risk levels: unacceptable, high, limited and minimal risk. Applications with unacceptable risk that are considered dangerous for EU citizen rights, such as social credit scoring systems or

manipulative uses of AI, will be banned. High risk systems include vaguely defined types of use cases that can pose a high level of risk on citizen rights. Transparency requirements are applicable to providers of high risk systems, whether in the public or private sector, and include the following provisions:

- Conformity assessment based on internal controls or third party assessment before the system is placed on the market; this assessment considers both the quality of the system's management as well as the technical documentation on the AI system itself including its risk assessment and monitoring plan;
- High level system disclosures to the general public by registering the system on an EU database;
- Informing end users when interacting with an AI system and providing them with proof of conformity.

The EU "Artificial Intelligence Act" is a promising development that is likely to be contested and iterated upon before coming into law. The proposed regulation is likely to catalyze increasing change in public and private sector approaches to algorithmic transparency and may potentially have global repercussions similar to how GDPR influenced other data privacy regulation.

The proposed EU "AI Act" is also criticized for its vagueness concerning the defined risk levels of ADM systems and the use cases outlined for high risk systems. One of the challenges to overcome is how to avoid misuse of the proposed risk screening model as only high risk systems have mandatory disclosures. The proposed regulation does not clearly define how the risk screening will be conducted. Joseph Foti, Chief Research Officer at the Open Government Partnership, argues that screening outcomes should be disclosed to the general public, regardless of whether the system is high or low impact. Furthermore, there should be additional considerations for non-conformity with risk screening and mandatory disclosures. Other criticisms include the lack of provisions for redress for those affected or harmed by AI systems as well as a lack of citizen engagement models in the proposal (Espinoza & Murgia, 2021).

Transparency as enabler of ADM accountability

Both transparency and answerability are required to safeguard the accountability of ADM systems (Krafft et al., 2020). Transparency ensures that the "factors that influence the decision of an algorithmic system should be [...] visible [and understood] to people employing or affected by the system (Kossow et al., 2021). In this context, transparency is not an end goal but an ideal sought to enable accountability. Transparency does not require ADM systems to be just or fair, but provides visibility for accountability mechanisms to be effective. Transparency can also enable demand-side accountability mechanisms.

Transparency can also be thought of at two levels: a first-order level focusing on the system itself, its functioning, design and implementation; and a second-order level focusing on the governance of the system (Kaminski, 2020). Both of these types of transparency are necessary in order to enable accountability and may face limitations when imposed in practice. ADM systems often have blurred accountability on which party is responsible for the systems intended and unintended goals (Krafft et al., 2020). Within a single ADM system, there can be distributed responsibility between different stakeholders involved in the system's commissioning, development, implementation and assessment. Information asymmetries can arise in these conditions and result in agency loss when overarching responsibilities and transparency requirements are not clear.

Having adequate and clear governance of an ADM system (second-order transparency) can enable better disclosures of the system's functioning and risk profile (first-order transparency), specifically in contexts where key components of the system are procured. Therefore it is critical to also consider ADM

governance models that enable clarity of accountability around system responsibilities and disclosures.

There are limitations to the level of algorithmic transparency that can be achieved, specifically in contexts of *intentional* or *intrinsic opacity* of an ADM system (Burrell, 2016). *Intentional opacity* exists in the context of protection of trade secrets and intellectual property regarding algorithmic systems. This type of opacity can be mitigated through regulation that outlines the disclosure obligations of companies who have trade secrecy concerns. Existing proposals include disclosing the technical information, such as source code and datasets used, to third party auditors, who are equipped to assess the system for conformity with existing standards and regulation, similar to how a financial audit would take place. *Intrinsic opacity* exists due to the nature of certain machine learning methods, such as deep learning, which are difficult to interpret (i.e. understand why a certain outcome was reached). A whole field of research called explainable artificial intelligence (XAI) has emerged in recent years to attempt to solve the interpretability problem.

In the next chapter, we will explore different mechanisms through which ADM transparency can be operationalized.

CHAPTER 3

OPERATIONALIZING ALGORITHMIC TRANSPARENCY

Transparency is required to enable ADM accountability. In this Chapter, we will explore how algorithmic transparency can be operationalized to achieve visibility of the system as whole and provide meaningful insights into its functioning and governance.

Limitations of existing algorithmic transparency mechanisms

Many countries have fragmented tools to enable ADM transparency and do not provide a comprehensive visibility of the system. Information pertinent to these systems is often scattered across multiple documents available to different subsets of stakeholders and are generally not available to the public. Furthermore, some of these protocols are not mandatory or being effectively enforced. Finally, there is a lack of standardized reporting requirements as well as safeguards to verify disclosures and enable compliance (Ada Lovelace Institute, 2020b).

[TABLE 2] REVIEW OF EXISTING ALGORITHMIC TRANSPARENCY MECHANISMS IN USE

MECHANISM	EXISTING IMPLEMENTATIONS	DISCUSSION
Assessments and evaluations	● Privacy impact assessments on handling of personal data (mandatory in EU, UK, US and Canada) ● Human rights impact assessments, go beyond privacy to consider nondiscrimination and other legal rights (mandatory in UK - Equality IA) ● Algorithmic impact assessments to determine the overall impact level of an ADM prior to implementation (mandatory in Canada) ● Third party ADM audits to evaluate system functioning, performance and posed risk (very limited implementation and regulatory mandate to date); audits can be commissioned by the system owner or performed without permission from the system owner	Privacy IAs are the most widely implemented type of assessment; however, they are limited in scope to handling personal data and are inadequate to consider all the harms that can be caused by ADMs. Human rights IAs and Algorithmic IAs offer a more comprehensive view of harm assessments; however, they have limited implementation. A number of public interest groups have called for mandatory comprehensive IAs (Access Now, 2020; AI Now, 2018; Alan Turing Institute, 2019). Post-deployment audits and evaluations can be a powerful tool to understand the impact of an algorithm. These audits can be performed via either *black-box* or *white-box* evaluation. A *black-box* evaluation considers only the inputs and outputs of the algorithmic system whereas a *white-box* evaluation also considers the source code of the algorithm. Black-box evaluations often can be performed without permission of the system owner to better understand the functioning of the ADM system, however it is not the most optimal type of evaluation (Lepri et al., 2018).
Procurement disclosures	● Government spending data can offer insights on ADM procurement ● Procurement audit trails, which can include procurement tender process, vendor assessment and vendor	Analysis of publicly disclosed government spending data to understand ADM procurement practices can be limited by data reporting quality and lays the burden of analysis on public interest groups. Open public procurements practices are gaining traction

	agreement (mandatory in UK, France, Canada)	and can be an important tool to understand procured ADM implementations. According to the Head of Ethics in a public sector organization, the procurement audit trail should ideally include an evaluation of the vendor and their track record in a similar spirit to human rights IAs.
Open sourcing	• Open sourcing of ADM training data and/or source code (limited implementations in the public sector)	Open sourcing has limited applicability in practice, specifically when considering datasets with personal identifiable data or when dealing with proprietary data and data processing mechanisms. Furthermore, most interviewed expert stakeholders that source code data offers very limited insights into the functioning of an ADM system.
Freedom of information / Public records requests	• Regulation enabling citizens / public interest groups to access public records under certain conditions (applicable in most countries)	Public record requests pose the burden on achieving transparency on civil society and public interest groups. A comprehensive public records study in the US showed that it failed to produce meaningful transparency due to poor record keeping from local government and contracting practice (Brauneis & Goodman, 2018).

A review of existing disclosure mechanisms suggests the need to both strengthen existing transparency mechanisms as well as provide comprehensive documentation on ADMs in a way that promotes meaningful transparency. Instead of privileging certain types of disclosures that look at standalone elements of the system, we should be "looking across" the system and consider them as "socio-technical systems" (Ananny & Crawford, 2018)

Disclosures to build a comprehensive view of ADM systems

It is important to differentiate between transparency relevant when a user interacts with an ADM system and one that enables a comprehensive view of the system as a whole. The scope of this work focuses on the latter.

Below is a summary of existing literature and recommendations on what disclosures should be included to give an exhaustive and complete view of an ADM system in a way that enables accountability. The basis of this disclosure framework is adapted from the "reviewable ADM framework" approach, which offers a systematic framework to practical transparency by breaking down the ADM process into stages (Cobbe et al., 2021). We distinguish between different types of disclosures:

- *Self-reported* information and ones *verified by third parties*;
- Information describing the *process* to create and manage the system and ones explaining and evaluating its *outcomes*.

[TABLE 3] SUMMARY OF EXISTING LITERATURE ON RELEVANT DISCLOSURES FOR ADM SYSTEMS

DISCLOSURE ELEMENT	TYPE	STAGE	INFORMATION TO INCLUDE	LITERATURE REFERENCE
System purpose	Process - self reported	Commissioning	Rationale behind the creation of the ADM system including the values, norms and legal basis behind its commissioning. The	(Centre for Data Ethics & Innovation, 2020; Cobbe et al., 2021;

			intended goals the system aims to achieve should also be disclosed here.	Etalab, 2021; Tow Center, 2015)
System accountability and governance	Process - self reported	Commissioning	Outline of the key stakeholders responsible for the commissioning, development, management, monitoring and review of the ADM system. The contact information of the system owner should be included as well.	(Etalab, 2021; Tow Center, 2015)
Procurement	Process - self reported	Commissioning	Key procurement agreements in place related to the ADM system's creation. The agreements may be required to be shared legally depending on local regulations. There are calls for the public sector to be transparent about sensitive data being collected, handled or trained on models by private sector companies.	(Cobbe et al., 2021)
Impact Assessment	Outcomes - self reported	Commissioning	Assessment of the potential implications and risks of the ADM system, which should It should include assessment of the following: compliance with existing regulation, concerns about bias and discrimination, privacy concerns and ethical issues raised. Some of these assessments may be legally required, for example Data Protection Impact Assessments are required under GDPR.	(Castellucia & Le Metayer, 2019) (Kaminski, 2020) (Cobbe et al., 2021)
Data use & handling	Process - self reported	Model Building	Overview of data used to train and test the ADM system, which should include information regarding the data collection mechanism, data handling, pre-processing as well as a privacy notice. Source data should be shared when possible. There are current efforts to standardize the documentation process around datasets, such as 'datasheets for datasets' (Gebru et al., 2020)	(Cobbe et al., 2021; Etalab, 2021; Tow Center, 2015)
Data quality assessment	Output - self reported	Model building	An assessment of the quality of data used based on its intended use cases, including considerations of data completeness, representation and accuracy among others. Existing frameworks for dataset quality assessment include the 'data nutrition project' (Chmielinski et al., 2020)	(Chmielinski et al., 2020; Tow Center, 2015)
System architecture	Process - self reported	Model building	Technical explanation of algorithm functioning including outlining statistical methods used to process data and build models. The source code should be shared if possible.	(Centre for Data Ethics & Innovation, 2020; Etalab, 2021; Tow Center, 2015)
System performance assessment	Outcomes - self reported	Deployment	Reporting accuracy and fairness metrics, ideally against benchmarks. Limitations may exists here if using black box models	(Tow Center, 2015)

			that have low interpretability	
Decision making process and human oversight	Process - self reported	Deployment	Outline of key use cases and explanation of decision making processes for each. Information provided should also clarify which steps in the process are automated and how and where human review and oversight takes place.	(Centre for Data Ethics & Innovation, 2020; Cobbe et al., 2021; Etalab, 2021)
System performance monitoring	Process - self reported	Deployment	Outline of the process to continuously monitor and evaluate the system's performance related to accuracy and fairness metrics.	(Centre for Data Ethics & Innovation, 2020)
Audit	Process & outcomes - performed third parties	Investigation	There are no commonly established standards for auditing algorithmic systems. Third-party audits are suggested to validate impact assessment, scrutinize critical self reported elements and ensure compliance with existing regulation.	(Ada Lovelace Institute, 2020a; Cobbe et al., 2021)

The above framework is a starting point to consider the types of information whose disclosure is necessary to provide *comprehensive* transparency of an ADM system. This framework does not consider who the information is shared with nor the sharing/displaying modality best practices. It is important to consider the target audience for a transparency mechanism to ensure understandability. In the case of disclosures to the general public, interpretability of the information provided should be a priority, and it is important to avoid creating a "false sense of transparency" by overwhelming the audience with documentation. A "spectrum of disclosures" may be required to provide the right level of disclosures to each target stakeholder (Kaminski, 2020). Another consideration to keep in mind regarding a comprehensive ADM disclosure framework is how these disclosures are applicable to different types of ADM systems, based on the intended scope or risk level for example. We saw how a proportionate risk-based approach towards regulating ADMs is being proposed in the EU with mandatory disclosures and assessments only required for systems qualified as high risk.

In the next chapter we will refine the disclosures' framework and propose preliminary recommendations on intended audience, modalities and types of ADMs to disclose via the governance of an ADM register to provide *meaningful* transparency. ADM registers have been surfacing recently as a tool to provide comprehensive disclosures and centralize information that typically is fragmented across different modalities and stakeholders.

CHAPTER 4

PUBLIC REGISTERS TO CREATE MEANINGFUL TRANSPARENCY ON ADM SYSTEMS

Public registers are an algorithmic transparency tool that can help provide comprehensive visibility of ADM systems. A handful of local governments have deployed initial versions of public ADM registers. In this chapter, we will provide an overview of existing register implementations and attempt to formalize the concept, purpose and disclosure modality of ADM registers based on the insights of expert stakeholders in the field.

What is an ADM register?

In its simplest form, an ADM register is a log of algorithmic decision making systems used by a public authority that have some level of direct impact on its citizens. In this form, the value of a register is to provide a baseline of visibility of the algorithmic systems a public authority uses, which is an improvement on the status quo, as in many contexts the very existence of these systems is opaque to the general public and even within public sector stakeholders.

The goal of ADM registers, as it is currently being discussed and implemented in the public sector, is to go beyond providing a baseline of visibility and achieve *meaningful* transparency on an ADM system level to concerned stakeholders. What is meant by providing meaningful transparency? The concept of meaningful transparency of an ADM system can be understood as providing the relevant information that can contribute to a good level of understanding of the system, in a format and modality that are intelligible to and interpretable by the target audience. Different stakeholders have different information priority needs. For example, a citizen affected by a given system would be interested in understanding how its algorithm impacts decision making, what its potential risks are and which pathways they can follow to claim redress, whereas an oversight body would be primarily interested in the technical performance of the system as well as its compliance with existing regulation.

The concept of a *public* ADM register is one that makes disclosures available to the general public following a standardized disclosure framework. All stakeholders interviewed in the context of this research agreed that the creation of public ADM registers should be encouraged, especially given that there is no mandatory oversight of ADMs. Therefore making relevant disclosures available to the general public can enable public interest groups to continue their historical role of investigating potential harms as regulatory oversight catches up. Public registers therefore enable demand-side accountability.

Emergence of first public ADM registers

The earliest calls for the creation of public registers in the EU are by two public interest groups, Access Now and AlgorithmWatch (2020). These calls came as a response to the European Commission's recommendations of expanding the use of artificial intelligence (European Commission, 2020), as an attempt to reconcile between the EU's AI ambition's and the reality of protection of human rights. Both these organizations call upon the creation of registers as mandatory disclosure mechanisms for ADM systems deployed by the public sector. Similar calls are also happening in the UK (Safak & Parker, 2020)

and the US (Reisman et al., 2018). In France the requirement to disclose ADM systems used in the public sector has been put into law in 2016 (*LOI N° 2016-1321*).

The first public ADM registers by the public sector have been implemented in the past year, these include:

- The cities of Amsterdam and Helsinki, which worked closely together to launch their respective registers in partnership with Saidot, a Finnish company that developed a platform that enables standardization of transparency requirements across an organization's ADM portfolio. Amsterdam and Helsinki's registers are not comprehensive of all of their ADM systems, however they do reflect a citizen-centered approach to creating transparency.The three respective parties launched a joint white paper outlining their collective experiences (Haataja et al., 2020). *We will expand on Amsterdam's experience in launching its register in the next chapter.*

- The cities of Antibes and Nantes-Metropole, which have launched their respective registers as mandated by the 2016 "Law for a Digital Republic". These represent the first attempts to comply with the law.

- New York City's compliance reporting which functions as a high level directory of algorithmic systems in use by the city's agencies. This reporting is mandatory as required per Executive Order 50 (Thamkittikasem, 2020).

[TABLE 4] OVERVIEW OF EXISTING PUBLIC ADM REGISTERS TO DATE

PUBLIC ADM REGISTER OWNER	YEAR LAUNCHED	NUMBER OF SYSTEMS LOGGED	INTERFACE TYPE	DISCLOSURES INCLUDED
City of Amsterdam[1]	Sept 2020	4	Interactive interface on a dedicated website with nested information	● System context and purpose ● System owner and contact info ● Procurement disclosure ● Datasets used ● Data processing and model overview ● Decision making process overview ● Nondiscrimination considerations ● Human oversight of system ● System risk level and risk management process ● Feedback processes
City of Helsinki[2]	Sept 2020	5	Interactive interface on a dedicated website with nested information	● System context and purpose ● System owner and contact info ● Procurement disclosure ● Datasets used ● Data processing and model overview ● Decision making process overview ● Nondiscrimination considerations ● Human oversight of system ● System risk level and risk

[1] https://algoritmeregister.amsterdam.nl/

[2] https://ai.hel.fi/en/ai-register/

				management methods ● Feedback processes
Nantes-Metropole[3]	Oct 2020	2	Interactive interface explaining service accompanied by downloadable documents explaining algorithms enabling the city's decision-making systems	● Service explanation ● Simulation of how service would work for citizen ● Technical specifications of algorithm and source code ● Decision making process overview
City of Antibes[4]	Jan 2021	7	Downloadable document of municipality's algorithm database	● System owner ● System context, purpose and legal basis ● Explanation of algorithm contribution to decision making process ● Datasets used and data processing ● Model technical functioning
New York City[5]	2020	16	Downloadable compliance report of city's use of algorithmic systems	● System owner ● System context and purpose ● Overall functioning description (high-level model architecture and decision making process overview)

Best practices for public ADM registers

Leveraging the existing literature on transparency mechanisms, learnings from early implementations of public registers as well as input from expert stakeholders from public interest groups, policy makers, government and technologists, we aim to formalize the concept of a public ADM register and offer recommendations.

When building a public ADM register, it is important to consider the following:

- Existing political and legal context
    - A number of factors can influence the implementation of a public register including the existing legal mandate and institutional mechanisms governing algorithmic systems, the adoption of open governance practices and the capacity of the public authority to carry out innovation projects.
    - These factors should be considered in order to establish compatibility and continuity with existing transparency practices. It is also important to consider forward-compatibility of newly adopted practices and ensure there is enough flexibility given that applications of ADM systems in government is a nascent field.

- The *unit* of disclosure
    - Throughout this work we have purposefully used the term *algorithmic decision making*

[3] https://data.nantesmetropole.fr/pages/algorithmes_nantes_metropole/

[4] https://www.antibes-juanlespins.com/images/pdf/Administration/20210129-Inventairev1desalgorithmes.pdf

[5] https://www1.nyc.gov/assets/ampo/downloads/pdf/AMPO-CY-2020-Agency-Compliance-Reporting.pdf

system, as opposed to *data*, *data processing*, *algorithm*, *statistical tool* or *AI*, to reflect the adequate unit of disclosure, as also argued in relevant literature (Bejtullahu-Michalopoulos & Florin, 2018; Castellucia & Le Metayer, 2019; European Commission, 2021; Safak & Parker, 2020). An ADM system can potentially include several algorithmic tools that enable a decision making process that has some level of impact on civil society; the key action here being the decision. Some aspects of this decision-making system can have human controlled inputs whereas others can be automated. It is important to provide transparency on the functioning of the system as whole and not just disclose the algorithmic enablers of the system. As discussed in Chapter 2, the governance of the system and the internal controls set on the algorithmic decision making system are critical to provide meaningful transparency.

- Should all ADM systems be disclosed? Systems can have different risks of harm and safety to individuals and collective goods. There are existing proposals for risk-proportionate approaches where disclosures are mandated for high risk systems only, such as the EU AI Act proposal (European Commission, 2021). This proposal, however, doesn't consider the mechanism through which screening occurs and does not include considerations about how to safeguard against noncompliance. We argue here that, in a context of lack of regulatory oversight, a register should include all ADM systems whose use impacts citizens directly (e.g. deciding benefits, allocations, violations, monitoring, etc) as well as those that have an impact on public goods (e.g. democratic institutions, environment, public financing), regardless of risk level. Lower-impact systems can have less stringent disclosure requirements; however, there should be a minimum level of disclosures for all systems. At the very least, low-risk systems should disclose the assessed risk level of the system as well as relevant information that can allow third parties to validate the risk assessment.

- The intended audience
  - A publicly available ADM register has a wide audience with varied information needs, engagement in public affairs and understanding of technology. Potential stakeholder groups include:
    - Civilians who have directly interacted with an ADM system and want to understand how the system impacts them, within this group we can diverse levels of technological understanding;
    - Public interest representatives, which can include researchers, investigative journalists and advocacy groups seeking to preserve the best interests of the general public, specific groups of society or public goods;
    - Other local government and policy officials.
  - It is important that the public authority implementing a public register considers the target groups the register is designed for and ensures the tool meets their information needs. It is likely that there will be tradeoffs between what information to include and how to display it in order to meet the requirements of different groups. *We will explore further in the next chapter how the city of Amsterdam approached these design choices.*

- The relevant disclosures
  - The ADM disclosures' framework included below outlines the recommended disclosures to include for each ADM system. The format has been adapted from the disclosures framework in Chapter 3 to include insights from existing register applications as well as feedback from expert practitioners. It is important to note that this framework clarifies 'what to disclose' but not 'how should the information be disclosed'. The modality of disclosure would largely depend on the intended audience and goals of the register.

    - Low impact systems should disclose at the very minimum the following sections: (1), (2), (3), (4), (5), (8), (11) and (12).
    - Most interviewed stakeholders agree that technical details such as source code and data are not most relevant or practical to share with the general public and can introduce a "false sense of transparency". A user interface study found that *white-box* explanations had little advantage over *black-box* explanations towards contributing to actual understanding of ADM systems (Cheng et al., 2019). *White-box* explanations show the inner workings of the algorithm whereas *black-box* explanations do not, they mainly show the outcome reached based on some inputted parameters. *White-box* explanations are more relevant for stakeholders interested in examining the system's technical functioning, such as certain public interest representatives or oversight groups. The City of Amsterdam's case study (in Chapter 5) illustrates the implementation of contracting requirements guaranteeing procedural transparency to the general public, while technical information is only required to be shared with third-party auditors.
    - Some good mechanisms to provide the general public required visibility into the technical aspects of the system are to include dummy data, data quality assessments as well as simulation environments of the system's functionality under different conditions.
    - In certain contexts, it may be necessary to not disclose certain technical information to the general public that would help users "game the system" of a certain policy and gain an unfair advantage such as claiming benefits (Lepri et al., 2018). This is a risk that should be considered against the disclosure framework provided for each ADM system.

- The disclosure modality
    - The value of a public register comes from the standardization of the disclosure mechanism and data interoperability. Each ADM system should be disclosed in a standardized way while considering the information needs of different target audiences, representing an interesting design challenge.
    - The display interface should have strong accessibility considerations, such as language to use, font size, etc. The interface should also be easy to navigate and locate the information relevant to each target audience. Usability tests are necessary in deploying registers to ensure the register is meeting its intended goals. Current best practices indicate that the needs of diverse stakeholder groups are best met using an interactive interface with layered information, where the baseline information accessible to all groups is displayed first and more technical documentation is embedded below.
    - Each section should include explanations in "plain language" that can be interpretable by individuals with little technical background. This is especially important for sections (1), (4), (8) and (11). Additional technical explanations and documentation should be included in each section to meet the information needs of other intended audiences.
    - Further recommendations on interface display and copy can be found in the GDPR's guidelines on transparency (Art 29 WP, 2016).

[TABLE 5] ADM REGISTER DISCLOSURES FRAMEWORK			
ADM DEVELOPMENT PHASES	**DISCLOSURE ELEMENTS**	**DISCLOSURE TYPE**	**INFORMATION TO INCLUDE**
			Legend: *(*) depends on local regulation,* *(**) not necessary to disclose to general public*
Commissioning	(1) System purpose	Process	• Context and motivation to create ADM system, this may include legal basis, policy, values and norms guiding its development • High level goals the system aims to deliver on • Key use cases and if relevant citizen groups impacted by system
	(2) System accountability / governance	Process	• List of key stakeholders involved in commissioning, design, development and management of ADM system • Ideally should list a *system owner* who is accountable for the entire system, as well as their contact information
	(3) Procurement procedures (*)	Process + Outcome	• List of vendors used and outline of services provided • Explanation of procurement process, vendor evaluation, ethical assessment and vendor agreement (*)
	(4) Impact assessments (*)	Process + Outcome	• Overall risk level of system and justification of assessment • May include assessments of compliance with local regulation, bias and discrimination, privacy, ethical concerns and other risks (*)
System building	(5) Data use & handling	Process	• Information on data collection mechanisms, processing, storage and privacy notice (*) for both training and deployment • Process to ensure data quality and bias mitigation
	(6) Data quality assessment	Outcome	• Explanation of types of data collected • Source data (**), or illustration of dummy data if source code data is not publicly shareable • Evaluation of adequacy of data used for specific use cases, including completeness, representation, accuracy concerns
	(7) System architecture	Process	• Technical explanation of model functioning, including data processing, model training and testing • Provide source code (**), or model testing environment if possible
Deployment	(8) Decision making process and human oversight	Process	• Schematic outline of decision making process for each intended use case, clarifying whether each step is automated and whether it is subject to human oversight

			• Explanation of extent of human oversight on overall system
	(9) System performance assessment	Process	• Accuracy, fairness and other relevant metrics on test and deployment contexts if possible • Should be accompanied by supporting text assessing performance and reliability of system
Post-launch	(10) System monitoring and review	Process	• Overview of key processes to monitor, review and evaluate process, and stakeholders involved in this process • Outline of key monitoring considerations, and escalation processes • Length or timeline of review cycles
	(11) Redress (*)	Process	• Outline of remedial process, including how redress requests are reviewed, assessed and remediated • Relevant contact information and steps to initiate remedial process (*)
	(12) Feedback loops	Process	• Process to provide feedback on ADM system and information included in register • Explanation of how feedback is reviewed and addressed
	(13) Audit performance	Outcome	• Audit reports performed by third parties to evaluate system performance and risks -- *currently no standard algorithmic audit framework*

Required governance to deploy a public ADM register

A public register is enabled by a governance mechanism that supports the documentation and disclosure process of the ADM systems used by a public authority, especially in a context where there is distributed responsibility for commissioning, building, deploying and monitoring the system between different stakeholders. Currently, ADM systems used by the public sector are largely deployed without the intention of making relevant public disclosures available. In such circumstances, a disclosure governance system would have to be set up post-ADM system deployment and the timeline to deploy a public register from the moment of internal buy-in is about six to nine months. Typically, a register is launched via a special projects unit that would report into a technology or innovation service of the public authority which would act as the project manager. The first key activity of the project manager is to get the buy-in of individual departments or units that have deployed ADM systems to participate in a register pilot. The buy-in process can be more difficult to achieve when there is no official mandate for disclosures to the public. The next key activities include establishing an internal ADM system knowledge base to understand how many ADM systems are being deployed by the public authority, the governance of each system and the existing documentation available. Many public authorities, such as the city of Amsterdam, operate in a siloed approach with little guarantee of consistency of approach between different agencies or services. Therefore, initial work is required to map existing disclosure documentation and identify gaps between the desired disclosures. Another important activity is to design the register based on its goals and target audiences. User research is an important tool to understand how best to design the register interface and display the disclosure information in a way that is adapted to the local context. The project manager

can choose to design the register in-house or outsource it to a vendor. Typically a few ADM systems would be selected to pilot in-house while the design of the register is being finalized. The learnings from this early development stage would then contribute to an easier onboarding of the remaining ADM systems.

Public authorities can benefit from adopting a proactive documentation approach, starting from the early stages of ADM system commissioning, regardless of whether there is an existing intention to create a public register, or not. Standardized documentation throughout the process can enable internal learnings and facilitate oversight efforts. The service owning the ADM system being commissioned can adopt documentation governance best practices to coordinate between the different stakeholders involved in the system. IBM's AI Factsheets 360 offers insight into how to coordinate between different stakeholders involved in the ADM system's lifecycle. These stakeholders include, among others, an overall system owner, policy representatives, data scientists, AI/ML practitioners, system engineers and potentially technical vendors. Each stakeholder has a limited purview of the set of facts related to the overall ADM system. A key part of the documentation governance is identifying the owner of each set of documentation as well its reviewer. A system owner, who has full responsibility towards the system achieving its goals while mitigating any potential risks, should take upon the role of reviewing the documentation in its entirety.

CHAPTER 5

CASE STUDY:
THE CITY OF AMSTERDAM'S AI REGISTER

In this chapter, we will consider the City of Amsterdam's public register as a case study where we review the register's development context, intended goals, design and development process and overall reception. To develop this case study, we interviewed a policy officer of the City of Amsterdam who was involved in the development and launch of the register, as well as accessed internal documentation on the register's design process and feedback data.

Creation context

In September of 2020, the cities of Amsterdam and Helsinki issued a joint press release announcing the launch of their respective AI registers, as the first cities in the world to do so. Both public authorities partnered together to make this effort possible as they were facing similar challenges relating to deploying algorithmic systems and shared a mission "to create as much understanding about algorithms as possible and be transparent about the way [they] are used" (City of Helsinki & City of Amsterdam, 2020).

The launch of the register in Amsterdam was possible due to a favorable political context. Since 2009, Amsterdam has launched its "Smart City" initiative as a platform to engage with citizens, government and business to pilot projects guiding the city's sustainable growth. This initiative marks the city's "commitment to move away from government bureaucracy" and promote active citizen participation in public life (JSG, 2018). Amsterdam's Chief Technology Office is an enabler of these efforts to innovate in policy and governance. Furthermore, a comprehensive audit performed by the Netherlands Court of Audit of the use of algorithms by the Dutch government found that "private citizens do not play a prominent role in the use of algorithms" and therefore recommended to "provide insight in algorithms for citizens and explain where and how they can obtain more information about algorithms" (Netherlands Court of Audit, 2021).

Public register goals

The City of Amsterdam's AI Register was created with three main goals. According to a relevant policy officer, one of the goals is to "demonstrate that it is possible to be transparent about the use of algorithmic systems in the public sector and encourage other governments to do the same [...] which is a goal that we believe we were able to achieve". The version of the register that is currently publicly available was described as an "MVP" (minimum viable product). The register is still under development and will likely go through iteration to be able to achieve its other intended goals.

Another goal behind the City of Amsterdam's adoption of a public register was to enable the recording of relevant documentation of all algorithmic systems used by the City. The public register does not currently document all systems used; only four are recorded at the moment. According to an involved policy officer, this goal has not been reached yet due to several internal reasons. First of all, there is no official mandate to disclose the City's ADMs via the governance of a public register. Given that each deployed system has its own governance, individual buy-in is required to onboard and document the system, and

the burden of work to be done lies with the existing system owners to implement, which take time to implement. More ADM systems should be added to the register in the future.

Finally, the third goal of the register is to promote the understanding of the algorithmic systems used by the city and create meaningful transparency for the citizens of Amsterdam. *The impact of the register on achieving this goal is still being reviewed; insights from preliminary feedback are included below.*

Relevant regulation

There is no current official mandate or regulation impacting the need for disclosures to the general public of ADM systems in the public sector for the City of Amsterdam. However, the first emerging procedures impacting transparency requirements of ADM systems have been recently outlined in the "Standard Clauses for Municipalities for Fair Use of Algorithmic Systems" which include conditions on contracting and procurement from external technological vendors (Township Amsterdam, n.d.). These procedures include language that clarifies the contractor rights with regards to data use and the quality of the system developed in terms of suitability, accuracy and compliance with existing laws.

One important distinction made is the difference between *technical* and *procedural* transparency. Technical transparency is defined as "the provision of information enabling the Municipality to understand the technical operation of the Algorithmic System" which may include the source code, technical specifications of the algorithm and development method. Procedural transparency is defined as "the provision of information on the purpose of the Algorithmic System and the process followed in the development and application of the Algorithmic System and the data used in that context", this includes providing an understanding of the choices and assumptions made in the development of the process, outlining how human intervention take place and identifying the system's risks and mitigation plan. As per the outlined standard clauses, the Municipality has a right to *procedural* transparency and also has the right to disclose it to the general public as it enables the municipality to provide explainability to its citizens and establish the quality and risks of the system and the controls used. *Technical* transparency is only required in the context of an audit with the information being shared directly with the auditor with no requirements to share with the Municipality. These procedures are meant to enable the continuation of the AI register and include systems that have been contracted from a third party.

Register design and development process

The cities of Amsterdam and Helsinki partnered with Saidot, a Finnish company that enables the standardization of transparency requirements across an organization's ADM portfolio. The rationale for the City of Amsterdam to partner with an external vendor was to accelerate the development process and leverage Saidot's existing transparency metadata model. Several offices within the city's government were engaged to develop implementation processes that would enable the creation of the register, including the Chief Technology and Chief Information offices.

The starting point to design the register, given that there were no references at the time, was to better understand the information needs of the citizens of Amsterdam. Consultative sessions were held with civilians and experts in the field of open data around the question of "What information do you need for the democratic control of algorithms?". The stakeholder groups were asked open-ended questions on the topic and were shown early prototypes of the register for feedback. This part of the design process helped define the stakeholders information requirements as well as improve the usability of the register interface. Key insights from this design phase include:

- The need for visual explanations in addition to simple text-based explanations specifically in outlining where human control takes place in the decision making system as well as explaining the algorithm functioning;
- The need for practical examples and illustrations, such as showing dummy data to illustrate the dataset in case there are privacy concerns, or presenting a testing environment to simulate how the algorithm operates under different conditions;
- The need for clarity of political accountability around the system and identifying the system owner;
- The need for clarity on the risk profile of the system and how the assessment was conducted.

A key design decision in the register development was defining a target audience. Saidot and the City of Amsterdam considered three different types of civilians:

- (1) Technical experts who understand the concept of AI, can read and write code and are interested in understanding how systems work;
- (2) Civilians with basic understanding of AI who come from diverse non-tech backgrounds and have some interest in technology and how it impacts their lives;
- (3) Civilians with no understanding of AI and have an indifferent or negative attitude towards technology.

The second group, civilians with a basic understanding of AI, were selected as the target audience to design the register for as this group includes journalists, policy researchers and other public interest representatives that historically play a role in providing explanations to the third group, civilians with no understanding of AI.

Another outcome from the design phase of the register was reaching an operational understanding of what providing meaningful transparency to citizens means, which is through creating a good understanding of the city's ADM systems and how it affects them. Given that the register needs to cater to individuals with varying levels of technical understanding, some tradeoffs where necessary to ensure accessibility.

It is important to note that although Amsterdam's register is publicized as an *AI* register, the *unit* of disclosure of the register is at an *algorithmic decision-making system* level as defined in the previous chapter.

Amsterdam's AI register 1.0

The City of Amsterdam's public AI register is hosted on a dedicated website[6] with both Dutch and English explanations. We include below excerpts from the Beta platform launched in September of 2020 and map the disclosures included to the framework from Chapter 4.

[6] https://algoritmeregister.amsterdam.nl/

Figure 1. *'City of Amsterdam Algorithm Register Beta' landing page (in English)*

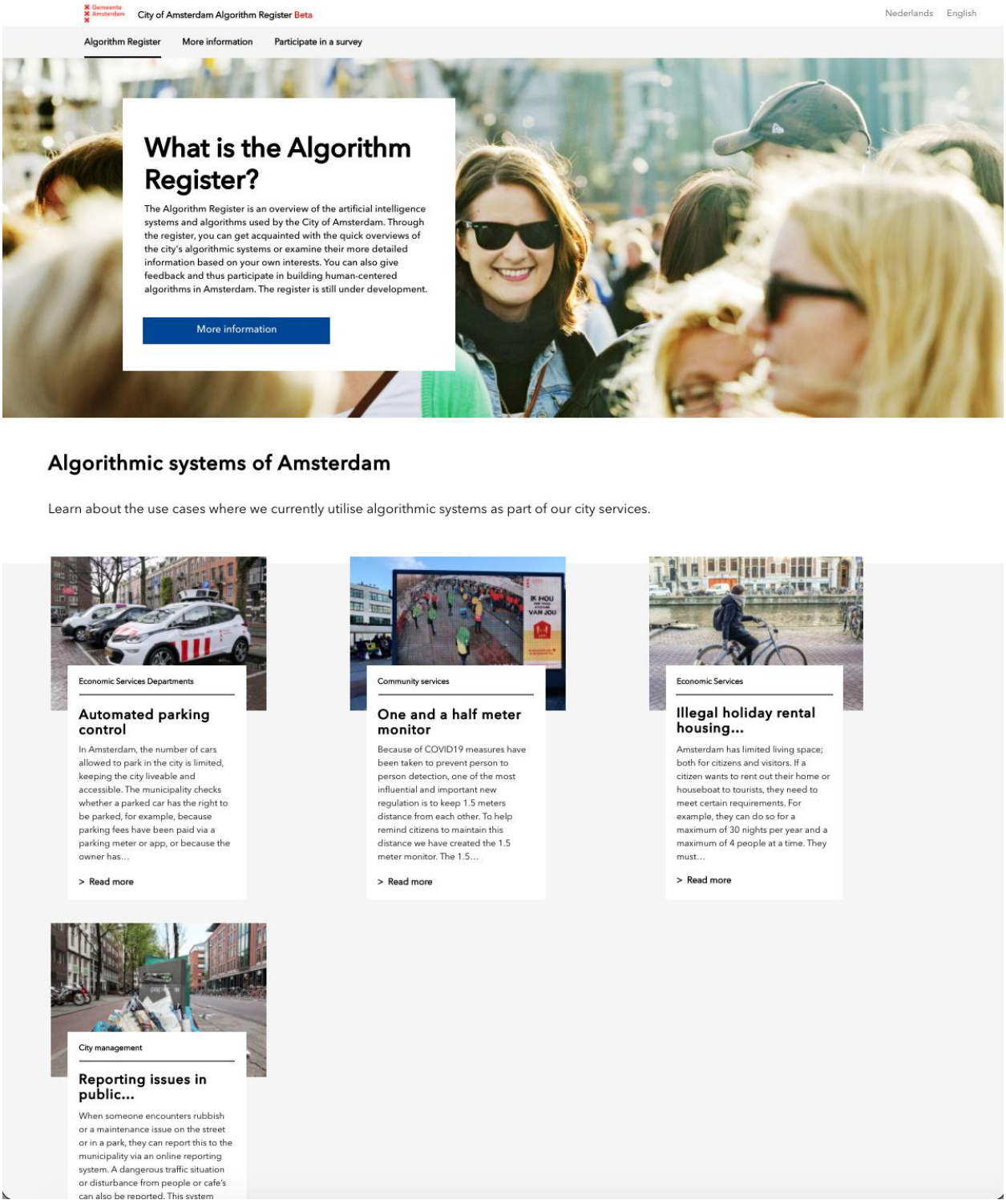


Figure 2 *'Automated parking control' ADM system overview (part 1/2) - high level overview and governance*

Economic Services Departments

Overview

Automated parking control

Tags Parking control | Image recognition | License plate reading

System purpose (1)

In Amsterdam, the number of cars allowed to park in the city is limited, keeping the city liveable and accessible. The municipality checks whether a parked car has the right to be parked, for example, because parking fees have been paid via a parking meter or app, or because the owner has a parking permit. Enforcement is done with the help of scan cars equpiped with camera's, automating the process of license plate identification and background checks with specific scanning equipment and AI-based identification service. The service is currently in use for more than 150,000 street parking spaces in the City of Amsterdam.

The service follows a three-step process. In the first step, the scan cars drive through the city and use object recognition software to scan and identify the license plates of surrounding cars. After the identification, the license plate number is checked against the National Parking Register to validate if the car has permission to park at a given location. Whenever no payment has been made for current parking, the case is sent to a human inspector for further processing. In the last step, the parking inspector assesses the scanned images to determine whether the license plate was recognized correctly and whether there is a special situation such as loading or unloading. The parking inspector may also verify the situation on-site. Whenever there is no valid reason found for non-paid parking, a parking ticket is issued.

Link to service

System governance (2)

Contact information

Department
Parking services

Contact team for inquiries

External vendors
Egis Parking Services B.V.

Contact email
parkeerdata@amsterdam.nl

Contact phone
+31 20 624 1111

Figure 3 *'Automated parking control' ADM system overview (part 2/2)* - *interactive user interface with layered sections*

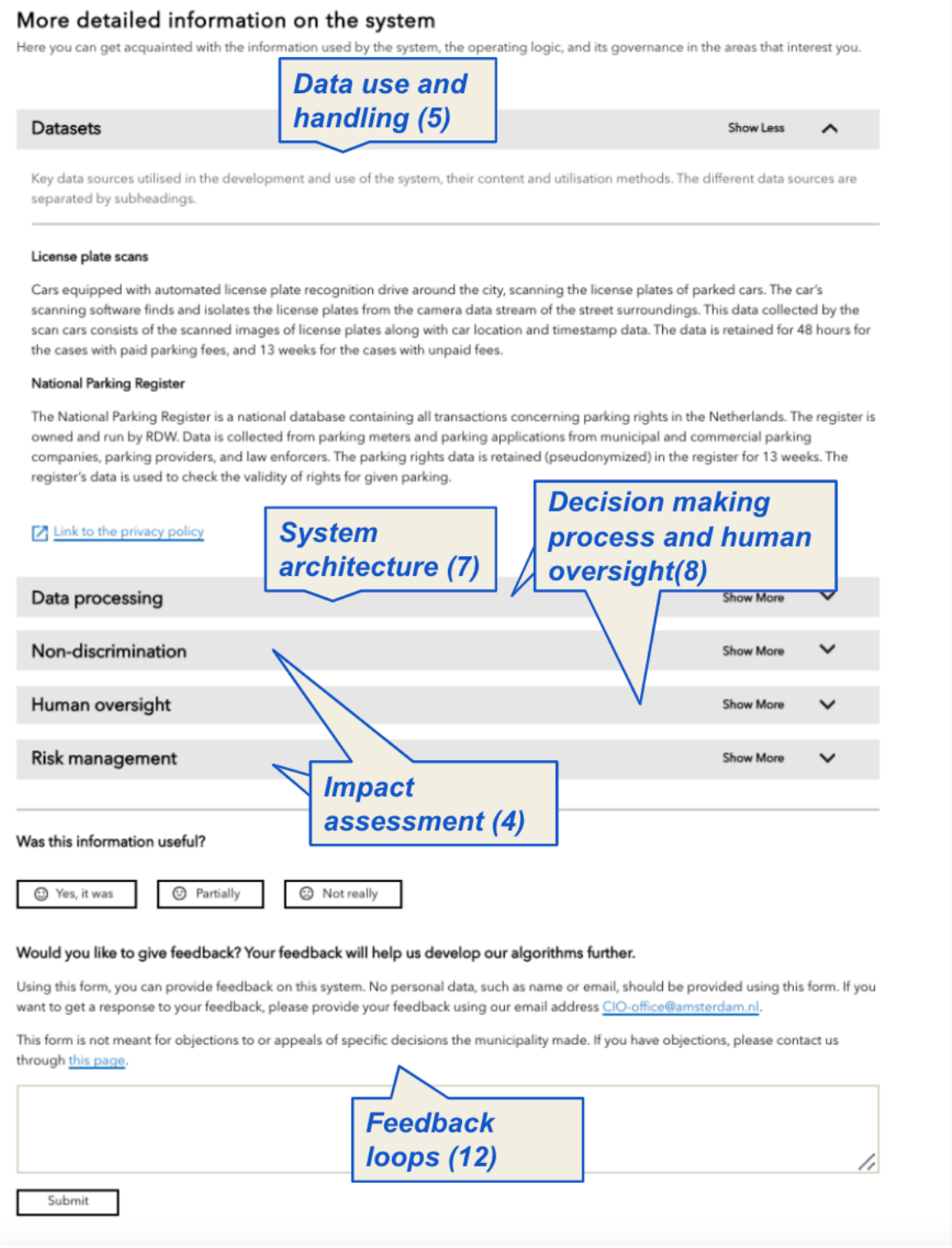
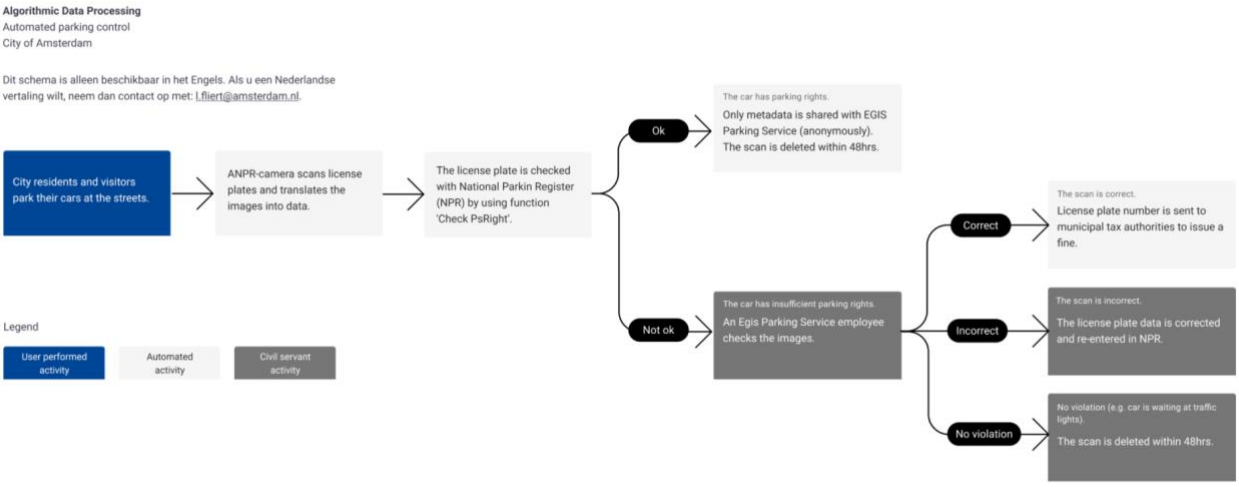


Figure 4 *'Automated parking control' data processing schematic outlining automated and human-controlled parts of the decision making process*



Initial reception and feedback

The City of Amsterdam has collected feedback on the register via the online survey provided in both English and Dutch on the register's portal. The survey has collected 25 responses in total as of April 2021. It is important to note that the results of this survey are not representative of the general population due to the small sample size as well as the strong technical background of the majority of respondents technical (24 respondents had at least a moderately good understanding of algorithms and artificial intelligence).

Key findings based on the initial feedback survey include:

- 68% of respondents have reported that the register contributed to some level of increased understanding of the City's use of algorithms, the rest reported no change in understanding. There's a recognition that the register is a "step in the right direction" and is setting a good precedent for transparency but improvements are needed.

- 50% of respondents did not find what they were looking for on the register website. The most common reasons mentioned is that the information displayed was "too high level" and that the algorithmic system they were looking for was not included in the register.

- In terms of perception of algorithm use by public authorities, 40% of respondents felt somewhat better after visiting the register, 24% felt worse and 32% had no change in perception. A number of respondents, including those that had an improved perception, have noted in their answers feeling "concerned" about the "fear of misuse" and being "worried about the use of AI without proper oversight and ethical guidelines" and are looking for practical safeguards against these concerns to be implemented and disclosed in the register.

Other civilian feedback has emerged on public online forums, although still predominantly coming from individuals with strong tech knowledge (Hacker News Forum, 2020). Some additional insights include:

- The burden of explainability when deploying less interpretable machine learning techniques can be a barrier for transparency and could potentially discourage public authorities from leveraging less interpretable techniques potentially at the expense of accuracy.

- Assessments of data quality and safeguards against biases should be disclosed as well.

- Self-reported risk profiling without disclosure of the assessment mechanism and the safeguards against risks at the system level does not provide adequate transparency. One example raised is Amsterdam's parking control systems, where the register cites there is no risk of nondiscrimination on the algorithm itself, without additional explanation of the assessment.

- There is a need to engage with the general public early on in the design process of ADM systems. "I wouldn't be surprised to see a future where all [machine learning] applications affecting citizens and the general public interest (excluding defense, military, etc.) will be obliged to be open-source and subject to public scrutiny."

CHAPTER 6

CIVIL SOCIETY ENGAGEMENT IN PUBLIC ADM REGISTERS

In this chapter we will explore mechanisms through which civil society can be engaged in the ADM lifecycle via the governance of a register. Public registers are designed to provide transparency to the general public and can potentially represent an opportunity to engage directly with civilians on public affairs involving ADM systems. We specifically single out the role of public interest representatives, which include research, advocacy and activism groups representing the best interest of civilians in the face of algorithmic systems' deployments. Existing algorithmic transparency mechanisms do not have clearly laid out plans to engage civil society in the process.

Why engage public interest representatives?

Public interest representatives have played a critical role in bringing visibility to problematic ADM implementations in the public sector and in helping pull the plug on harmful systems. For instance, all of the examples of ADM failures in the public sector cited in Chapter 2 have been brought to light by public interest representatives.

Public interest representatives also play a critical role in shaping up government best practices around ADM systems (Kuziemski & Misuraca, 2020). Given the current lack of oversight mechanisms of ADM systems and safeguards against noncompliance with proposed transparency requirements, we can still envision public interest representatives continuing to play a critical role in promoting meaningful transparency of ADM systems through providing unofficial oversight and demand-side accountability. In the context of the implementation of an ADM register, public interest groups can leverage the information disclosed to investigate ADM systems and potentially request more information in case of abnormal findings.

Public interest group engagement can also benefit public policy decisions relating to ADM systems specifically by leveraging feedback in the commissioning and design stages. This is currently a missed opportunity whenever public authorities do not disclose their intentions to leverage ADM systems or actively engage civil society in the process. The implementation of representative deliberative processes has been documented to lead to better policy outcomes, enhance public trust in public authority and help counteract polarization of public opinion on policy matters (OECD, 2020).

Potential limitations of public interest group engagement

One of the most important pitfalls concerning public interest group engagement in ADM transparency mechanisms is having these groups turn into "free auditors of the state" (Safak & Parker, 2020), where they perform the growing and labor-intensive groundwork of analyzing complex systems for the benefit of the public due to lacking formal oversight and accountability mechanisms. While many public interest groups have specific mandates regarding the safeguarding of the interests of civil society and/or minority groups (and their work may involve investigating ADM system implementations), the primary burden of accountability should lie with the local government. Furthermore, the labor burden on public interest groups can be eased with formalized channels to access documentation and communicate with ADM

system owners. Currently, accessing documentation constitutes a major pain point for public interest groups, even in contexts where there are "Freedom of Information" laws (Brauneis & Goodman, 2018; openDemocracy, 2021).

Another potential pitfall is to avoid the *tokenization* of the participation of public interest groups in the development of ADM systems by public authorities, wherein these systems can appear to gain legitimacy in the public perception thanks to their association with public interest groups. This can be especially problematic in situations where the recommendations of public interest groups were not included in the design process to begin with. There should be additional considerations on how to disclose the engagement of public interest representatives while avoiding such pitfalls.

Existing public engagement models

There are several existing models of civil society and public interest group engagement implemented by the public sector. Many of these modalities are adapted to the local context and differ based on the criticality of the policy considered. These models can be formal (instituted into existing governance and regulation) or informal. The most widely used models are public consultations, calls for submissions or responses to existing policy, roundtable discussions and surveys. There are also dedicated governance models for representative civilian engagement, as documented by the OECD, such as "citizen initiative review models", "citizen councils" and "observatories" (OECD, 2020). It is important to note that the output of these models can be collective (i.e. all concerned parties reach one set of decisions or recommendation) or fragmented (i.e. representing standalone opinions or recommendations).

Some relevant learnings from existing engagement models include:

- Participants of the engagement should be given a clear and well-defined task which is typically linked to a public problem;
- Participants should have access to relevant documentation and training if required to help inform their opinion/decision and their time commitment should be compensated;
- The commissioning authority's commitment to engage and respond to public interest representative inputs should be publicly disclosed along with relevant information about the procedures and timelines they will be following;
- When a formal engagement process is launched, the full process, feedback collected, outcomes and public authority response to the feedback should be disclosed.

Recommendations for engagement models in the context of an ADM register

We have established in Chapter 4 that an ADM register can also be thought of as a governance mechanism enabling technical documentation as well as multi-stakeholder engagement in the ADM system's lifecycle. Different types of stakeholders may be needed to be engaged at different points of the ADM development process. Some stakeholders are necessary to be engaged at an overall register level, to ensure that the register meets its goals as a transparency mechanism, and others at a particular ADM system level, to mitigate potential risks. There is a general consensus that such engagements should start at the commissioning stage to ensure that the right solution is being developed to address the policy problem.

Recommended engagement touchpoints are summarized below:

[TABLE 6] CIVIL SOCIETY ENGAGEMENT IN ADM LIFECYCLE			
Types of ADM systems	**Lifecycle stage**	**Potential engagement goals**	**Potential target groups and engagement mechanisms**
High risk as defined by the following factors (probability of harm * scale of harm * degree of harm)	Commissioning	Identifying the concerns of target groups and ensuring the solution commissioned is well suited to solve the policy problem	Detailed discussions with the groups affected by the system's implementation and involving relevant advocacy groups and intermediaries. Formal and informal feedback loops with policy and advocacy groups
	Model Building	Ensuring the algorithmic system is fit for deployment	Validating with research groups the accuracy and fairness results of the developed model, which may involve sharing privileged content on the model's functioning that would not be available to the general public. Involved PIG would be allowed to share their findings but not disclose source data and code.
	Deployment	Ensuring right checks and balances are in place	Reviewing processes in place with affected groups either directly or through intermediaries as well as involving advocacy groups
	Investigation	Mitigating post-launch risks and ensuring adequate system visibility to the public	Feedback surveys open to the general public, as well as potential audits to be performed by third parties which may be a PIG if a regulatory oversight body does not exist
Lower risk systems	Commissioning	Identifying the concerns of target groups and ensuring the solution commissioned is well suited to solve the policy problem	Policy groups via formal or informal channels
	Model Building	*Optional*	
	Deployment	*Optional*	
	Investigation	Evaluating the effectiveness of launched system and ensuring adequate system visibility to the public	Feedback surveys open to the general public, as well as potential audits to be performed by third parties which may be a PIG if a regulatory oversight body does not exist

One of the critical ways public interest representatives can be engaged is in the validation of system risk screening results, especially if there is no official oversight body performing that role. We make the argument for an established process to formally engage the public in the process of validating the risk level of the system prior to the system's launch. For this to be feasible, there should be a requirement from the public authority to disclose the commissioning of new ADM systems. There should also be a formal channel for select public interest groups to access privileged documentation and share informal feedback, which can be done via the governance of a register used early in the development phase.

Beyond the scope of a register, public interest representatives should actively be engaged in setting standards regarding algorithmic systems as well as initiating broader policy efforts to drive reliable, ethical and accountable systems.

As the use of ADM systems by public authorities is expected to increase, governments should invest in educating their citizens about algorithms in general and providing the means for representative segments of society to engage in the ADM development feedback process. A significant barrier right to achieving meaningful transparency is that engagement in public debate on ADM use mainly involves individuals or groups who have strong technological knowledge and may not be representative of the overall population, as seen in Amsterdam's case study.

CHAPTER 7

SUMMARY OF FINDINGS

A register is a tool public authorities can leverage to achieve meaningful transparency around the algorithmic decision making systems deployed to support existing policies and services. As calls are increasingly emerging to improve the oversight and accountability of mechanisms governing algorithmic systems, public authorities can establish a baseline for demand-side accountability by providing relevant documentation on the procedural and technical functioning of their systems through a public register. We learn from the City of Amsterdam's journey to release their first ADM register that it is possible to make the first steps towards algorithmic transparency even when there is no existing regulatory mandate. Several governments are considering implementing algorithmic transparency mechanisms such as registries, for which we make the following recommendations:

- Consider the existing regulatory context and public authority positioning with regards to ADM accountability and implement transparency mechanisms that allow for flexibility and forward compatibility;

- Adopt a proactive documentation approach, beginning with the ADM system's commissioning stage;

- Clarify transparency requirements from external vendors and third parties;

- Adopt a citizen-centric approach, and deeply considering the information needs of citizens and how to best address them;

- If following a risk-based approach to ADM screening, develop well-defined guidelines for ADM system risk screening and enforcing mandatory disclosure of assessments conducted for systems of all risk levels to avoid miscreening and altered incentives;

- Engage civil society throughout the ADM development lifecycle and embed engagement mechanisms in transparency modalities.

CHAPTER 8

OPEN POINTS AND FUTURE WORK

The intended scope of this thesis is to provide preliminary generalized findings on the best practices of implementing public registers with the purpose of achieving meaningful transparency of algorithmic systems deployed by public authorities. This work touches upon multiple disciplines and there are several pathways forward to refine the research, that include:

- Further refining the existing register disclosure framework, through stress-testing its adequacy under different scenarios and ADM risk profiles, ideally leveraging implemented examples.

- Developing guidelines for responsible ADM system development in the public sector that takes on a proactive approach towards system documentation and civil society engagement for specific governments and/or regulatory contexts.

- Exploring human-computer interaction factors that can improve the register design with a goal to enable better understandability of ADM systems affecting citizens. This may include exploring mechanisms that can enable the understanding of the functioning of an algorithmic system to audiences with low technical understanding as well as enabling demand-side accountability mechanisms.

- Defining regulatory mechanisms to safeguard the transparency and accountability of ADM systems deployed in the public sector. Localization of research efforts is important and we encourage considering case studies outside of European Union and North America.

- Exploring how to operationalize algorithmic transparency for high risk systems deployed by the *private* sector and potentially considering public registers as a viable tool. There are significant hurdles to this effect, which include trade secrecy concerns as well as enforcing compliance.

APPENDIX

Key definitions

- **Algorithm**
  An algorithm is "is a set of rules that precisely define a sequence of operations" (Stone, 1971). Algorithms are currently most commonly referred to in the context of computer science, as the "computational generation of knowledge or decisions" (Gillespie, 2016). Algorithms may include a wide range of operations with different levels of complexity that can include *rule-based* approaches as well as *machine learning* techniques.

- **Data processing**
  Data processing is a series of operations on data, especially by a computer, to retrieve, transform or classify information. Data processing is a pre-requisite for the implementation of AI models or other forms of statistical analysis on data. Algorithms may be used to automate data processing.

- **Artificial Intelligence (AI)**
  AI is used to describe the set of scientific disciplines and technologies that enable machines to carry out tasks generally associated with human intelligence. "Algorithms and AI are intrinsically related, but they are not the same. Every computer program is built using algorithms, but not every program is regarded as AI" (Innes & Beacon, 2021). Application of AI includes computer vision, speech recognition, pattern recognition and translation.

- **Machine Learning (ML)**
  ML is a subset of AI where that allows the machine to learn from data without explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data. Certain types of ML models are not interpretable, meaning we do not understand how the model is learning and reaching a certain outcome.

- **Algorithmic decision making systems (ADMs)**
  Algorithmic decision making systems refer to the use of algorithms to support decisions such as prioritization, classification, association, and filtering (Cheng et al., 2019). The introduction of algorithmic systems helps procedurialize and automate parts or the entire decision making process. An ADM system can be decomposed into a number of steps automated by an algorithmic system or carried out by a human based on inputs from an algorithmic system or other assessments.

- **ADM governance**
  ADM governance refers to the set of legal, ethical and professional behaviours or conventions that, taken together, guide the development and use of data and decision-making algorithms. This governance can entail an array of different regulatory approaches from international and state law, to collective self-regulation and civil initiatives. The main goal of ADM governance is to proactively minimize the risks associated with the deployment of ADMs. The governance of ADM systems involve technical and non-technical considerations that relate to the system's methodology, human input and control, the domain of application as well as the system's purpose (Bejtullahu-Michalopoulos & Florin, 2018).

- **ADM registers**
  In its simplest form, an ADM register is a log of algorithmic decision making systems used by an entity. ADM registers can also be thought of as "a standardised, searchable and archivable way to document the decisions and assumptions that were made in the process of developing, implementing, managing and ultimately dismantling an algorithm" (Haataja et al., 2020).

- **Public interest representatives**
  Public interests representatives include groups and individuals that represent civil society and/or specific groups in matters of public discourse and advocate for their rights.

Research Methodology

The main research techniques leveraged in this thesis are:

- Extensive literature review on algorithmic transparency, specifically looking into requirements to operationalize ADM transparency in the public sector;
- Expert interviews to understand transparency requirements from the viewpoint of different stakeholders and to validate the proposals made to operationalize ADM transparency;
- Case study on the implementation of an ADM register in Amsterdam through the review of internal documents documenting the design process and the feedback received the date, as well as access to a policy officer in Amsterdam involved in the development and implementation of the register.

List of expert stakeholders interviewed:

- *Anonymous*, Head of Data Ethics in a public sector organization
- *Anonymous*, Policy Officer working for the City of Amsterdam and involved in the development and implementation of the city's AI register
- *Anonymous*, Researcher on algorithmic accountability at an independent research group
- Joseph Foti, Chief Research Officer at the Open Government Partnership
- Meeri Haataja, CEO of Saidot

BIBLIOGRAPHY

Access Now. (2020, June 11). Trust and excellence—The EU is missing the mark again on AI and human rights. *Access Now*. https://www.accessnow.org/trust-and-excellence-the-eu-is-missing-the-mark-again-on-ai-and-human-rights/

Ada Lovelace Institute. (2020a). *Examining the Black Box: Tools for assessing algorithmic systems*. https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/

Ada Lovelace Institute. (2020b). *Transparency mechanisms for UK public-sector algorithmic decision-making systems*. https://www.adalovelaceinstitute.org/evidence-review/transparency-mechanisms-for-uk-public-sector-algorithmic-decision-making-systems/

AI Now. (2018). *AI Now Algorithmic impact assessments*. http://equinepromotion.com/site/ai-now-algorithmic-impact-assessment-164ef2

Alan Turing Institute. (2019). *Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector* (SSRN Scholarly Paper ID 3403301). Social Science Research Network. https://doi.org/10.2139/ssrn.3403301

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

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=1B8jKuq-H9G4ZEq4_95FZ7ZaZ9a3rKDs

Art 29 WP. (2016). *WP29 Guidelines on transparency under Regulation 2016/679*.

Bejtullahu-Michalopoulos, K., & Florin, M.-V. (2018, November 26). *The Governance of Decision-Making Algorithms*.

Booker, C., & Wyden, R. (2019). *Algorithmic Accountability Act of 2019*. 15.

Brauneis, R., & Goodman, E. P. (2018). Algorithmic Transparency for the Smart City. *TECHNOLOGY Vol., 20*, 74.

Burgess, M. (2020, August 20). The lessons we all must learn from the A-levels algorithm debacle. *Wired UK*. https://www.wired.co.uk/article/gcse-results-alevels-algorithm-explained

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

Castellucia, C., & Le Metayer, D. (2019). *Understanding algorithmic decision-making: Opportunities and challenges*. https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2019)624261

Centre for Data Ethics & Innovation. (2020). *Review into bias in algorithmic decision-making* (p. 151).

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. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300789

Chmielinski, K. S., Newman, S., Taylor, M., Joseph, J., Thomas, K., Yurkofsky, J., & Qiu, Y. C. (2020). *The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence*. 7.

City of Helsinki & City of Amsterdam. (2020). *Press release: Helsinki and Amsterdam first cities in the world to launch open AI register*. News Powered by Cision. https://news.cision.com/fi/city-of-helsinki/r/helsinki-and-amsterdam-first-cities-in-the-world-to-launch-open-ai-register,c3204076

Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems. *ArXiv:2102.04201 [Cs]*. http://arxiv.org/abs/2102.04201

Equality and Human Rights Commission. (2010). *Public sector equality duty FAQs | Equality and Human Rights Commission*. https://www.equalityhumanrights.com/en/public-sector-equality-duty-

scotland/public-sector-equality-duty-faqs

Espinoza, J., & Murgia, M. (2021, April 24). *Europe attempts to take leading role in regulating uses of AI*. https://www.ft.com/content/360faa3e-4110-4f38-b618-dd695deece90

Etalab. (2021). *Fiche pratique: L'inventaire des principaux traitements algorithmiques | guides.etalab.gouv.fr*. https://guides.etalab.gouv.fr/algorithmes/inventaire/#en-pratique-quelles-informations-dois-je-indiquer-dans-mon-inventaire

European Commission. (2020). *White Paper on Artificial Intelligence: A European approach to excellence and trust* [Text]. European Commission - European Commission. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European Commission. (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) | Shaping Europe's digital future*. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence

Fitzgerald, M., & Crider, C. (2020). *Under pressure, UK government releases NHS COVID data deals with big tech*. OpenDemocracy. https://www.opendemocracy.net/en/ournhs/under-pressure-uk-government-releases-nhs-covid-data-deals-big-tech/

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2020). Datasheets for Datasets. *ArXiv:1803.09010 [Cs]*. http://arxiv.org/abs/1803.09010

Gillespie, T. (2016). *Algorithm*.

Gould, M., Joshi, D. I., & Tang, M. (2020). *The power of data in a pandemic—Technology in the NHS*. https://healthtech.blog.gov.uk/2020/03/28/the-power-of-data-in-a-pandemic/

Haataja, M., van de Fliert, L., & Rautio, P. (2020). *Public AI Registers: Realising AI transparency and civic participation in government use of AI*. Saidot. https://uploads-ssl.webflow.com/5c8abedb10ed656ecfb65fd9/5f6f334b49d5444079726a79_AI%20Registers%20-

%20White%20paper%201.0.pdf

Hacker News Forum. (2020). *City of Amsterdam's Algorithm Register | Hacker News*. Y Combinator Hacker News. https://news.ycombinator.com/item?id=24623639

Hughes, D. (2020, August 17). *How the A-level grade algorithm works, and what Ofqual's system meant for results this year*. Inews.Co.Uk. https://inews.co.uk/news/education/a-level-algorithm-what-ofqual-grades-how-work-results-2020-explained-581250

IBM Research. (n.d.). *AI FactSheets 360*. AI Factsheets 360. Retrieved May 2, 2021, from aifs360.mybluemix.net/governance

Innes, K., & Beacon, R. (2021). *Government by Algorithm: The Myths, Challenges and Opportunities*. Institute for Global Change. https://institute.global/policy/government-algorithm-myths-challenges-and-opportunities

JSG. (2018). Amsterdam Smart City (ASC). *HBS Digital Innovation and Transformation*. https://digital.hbs.edu/platform-digit/submission/amsterdam-smart-city-asc/

Kaminski, M. E. (2020). *Understanding Transparency in Algorithmic Accountability* (SSRN Scholarly Paper ID 3622657). Social Science Research Network. https://papers.ssrn.com/abstract=3622657

Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). *A governance framework for algorithmic accountability and transparency.* Publications Office. https://data.europa.eu/doi/10.2861/59990

Kossow, N., Windwehr, S., & Jenkins, M. (2021). *Algorithmic transparency and accountability* (p. 24).

Krafft, T. D., Zweig, K. A., & König, P. D. (2020). How to regulate algorithmic decision-making: A framework of regulatory requirements for different applications. *Regulation & Governance*, *n/a*(n/a). https://doi.org/10.1111/rego.12369

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

Lee, G. (2020, August 14). *FactCheck: Did England exam system favour private schools?* Channel 4 News. https://www.channel4.com/news/factcheck/factcheck-did-england-exam-system-favour-private-schools

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, Transparent, and Accountable Algorithmic Decision-making Processes. *Philosophy & Technology*, *31*(4), 611–627. https://doi.org/10.1007/s13347-017-0279-x

LOI n° 2016-1321. (n.d.). Retrieved April 30, 2021, from https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000031589829/

Lum, K., & Chowdhury, R. (2021). *What is an "algorithm"? It depends whom you ask*. MIT Technology Review. https://www.technologyreview.com/2021/02/26/1020007/what-is-an-algorithm/

Marx, D. J., Gary. (2017). *Data mining program designed to predict child abuse proves unreliable, DCFS says*. Chicagotribune.Com. https://www.chicagotribune.com/investigations/ct-dcfs-eckerd-met-20171206-story.html

Mittelstadt, B. (2017). From Individual to Group Privacy in Big Data Analytics. *Philosophy & Technology*, *30*(4), 475–494. https://doi.org/10.1007/s13347-017-0253-7

Netherlands Court of Audit. (2021, January 26). *Understanding algorithms—Report—Netherlands Court of Audit* [Rapport]. Algemene Rekenkamer. https://english.rekenkamer.nl/publications/reports/2021/01/26/understanding-algorithms

OECD. (2020). *Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave | OECD iLibrary*. ../../../../gov-2020-630-en/index.html

openDemocracy. (2021). *Freedom of Information*. OpenDemocracy. https://www.opendemocracy.net/en/freedom-of-information/

Radley-Gardner, O., Beale, H., & Zimmermann, R. (Eds.). (2016). *Fundamental Texts On European Private Law*. Hart Publishing. https://doi.org/10.5040/9781782258674

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). *Algorithmic Impact Assessments: A*

Practical Framework For Public Agency Accountability (pp. 3–30). AI Now. https://ainowinstitute.org/aiareport2018.pdf

Safak, C., & Parker, I. (2020). *Meaningful transparency and (in)visible algorithms*. https://www.adalovelaceinstitute.org/blog/meaningful-transparency-and-invisible-algorithms/

State of California Department of Justice. (2018, October 15). *California Consumer Privacy Act (CCPA)*. State of California - Department of Justice - Office of the Attorney General. https://oag.ca.gov/privacy/ccpa

Stone, H. S. (1971). *Introduction to Computer Organization and Data Structures*. McGraw-Hill, Inc.

Thamkittikasem, J. (2020). *Implementing Executive Order 50 (2019): Summary of Agency Compliance Reporting* (p. 18). NYC Algorithms Management and Policy Officer.

Tow Center. (2015). Towards a Standard for Algorithmic Transparency in the Media | by Tow Center | Tow Center | Medium. *Tow Center*. https://medium.com/tow-center/towards-a-standard-for-algorithmic-transparency-in-the-media-81c7b68c3391

Township Amsterdam. (n.d.). *Standard Clauses for Municipalities for Fair Use of Algorithmic Applications*. https://assets.amsterdam.nl/publish/pages/934668/standard_clauses_for_municipalities_web.pdf

US Department of Homeland Security. (2009, July 6). *Privacy Impact Assessments*. Department of Homeland Security. https://www.dhs.gov/privacy-impact-assessments