Gemeente Amsterdam

Helsinki

Saidot

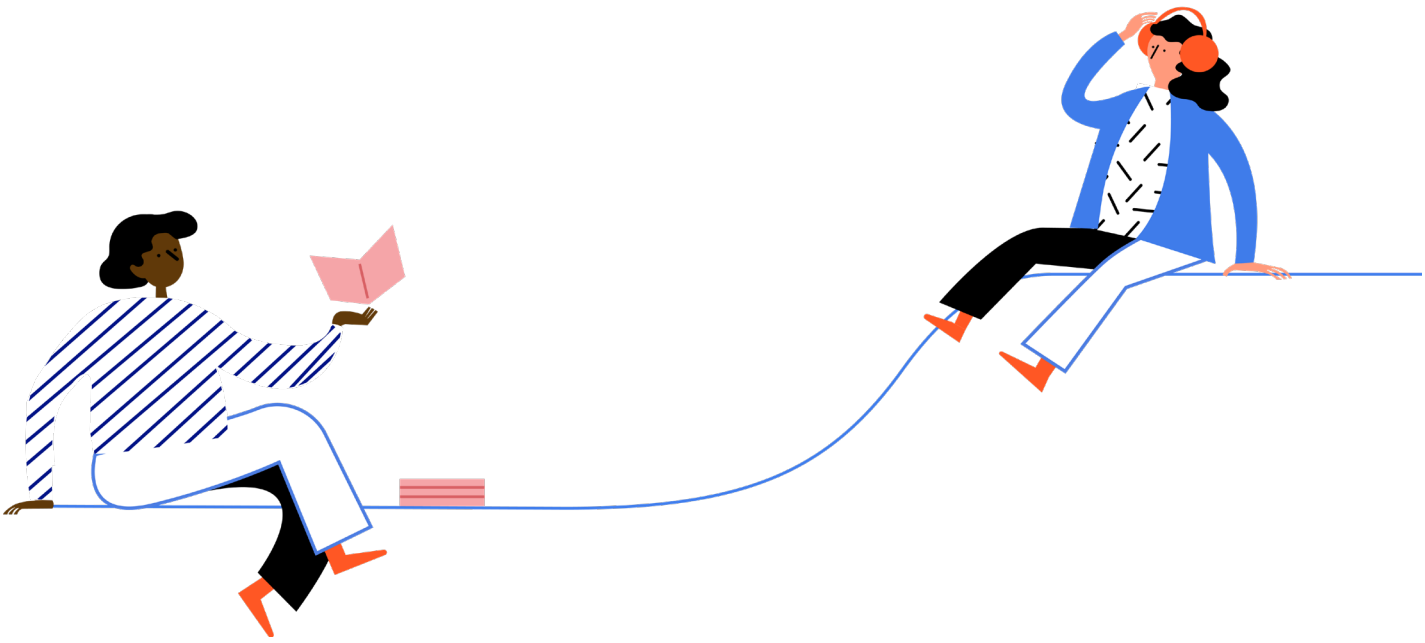
September 2020

Public AI Registers

Realising AI transparency and civic participation in government use of AI

Whitepaper written by **Meeri Haataja**, **Linda van de Fliert** and **Pasi Rautio**

Contents



Version 1.0

September 2020

Executive summary

Algorithms and artificial intelligence (AI) have become an integral part of our daily lives. Governments use algorithms to improve the accessibility and experience of their services, to optimise processes, or to manage public space, for example. In the cities of Amsterdam and Helsinki alone, more than thirty AI projects are taking place at the moment. However, regular citizens, each one of us, still have limited access to understand, participate in and debate about these developments.

Democratic governments are open and allow the public to voice their opinions on government activities and the use of public authority. We believe this should apply not only to human actions but also to the ways we automate those actions with the help of algorithms. Every citizen should have access to understandable and up-to-date information about how algorithms affect their lives and the grounds on which algorithmic decisions are made. It is no wonder transparency is referred to as the most cited principle for trustworthy AI[1,2].

We think it is time to align the actions of our governments to meet this expectation. For this, we introduce the concept of a public AI register as a means for transparency and civic participation in government use of AI. In this paper, we share our work on AI registers and suggest the concept for broader adaptation.

About the register

The AI register is 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. With this, transparency, and when applicable, explainability, can be given for public debate, independent auditors, and individuals citizens. For civil society, it is a window into the artificial intelligence systems used by a government organisation. Ultimately, we hope it will become a catalyst for meaningful democratic participation and a platform for fostering mutual trust.

Building on the suggestions raised in various AI ethics and human-rights related research, policy papers and existing legal frameworks, we suggest a documentation framework for AI registers that is ready for any government organisation to take into use. With this, we hope not only to bring down the barriers of moving into action but also, to welcome feedback for making it even better.

We suggest government organisations to ensure meaningful transparency via AI registers into the following aspects of all their algorithmic systems: purpose and impacts, accountability, datasets, data processing, non-discrimination, human oversight, risks and their mitigation measures, and explainability. Producing this information, in collaboration with AI technology suppliers and other partners, and publishing it for civil participation, should become a standard procedure across government AI projects.

We welcome our readers to explore how it looks in practice by visiting our public AI registers: **City of Amsterdam Algorithm Register** and **City of Helsinki AI Register**. Both registers are enabled by Saidot's AI transparency platform. We hope they will inspire active civic participation through our AI registers, helping us further align our AI with the values of the people we serve.

Introduction

The government can only function when there is trust between the government and people. Whether it is local government or agencies operating on a state level, citizens have to be able to trust that the decisions and actions will drive the best of the society and ultimately, its people.

Algorithms and AI have become an inseparable part of our daily lives from optimising delivery routes to facial recognition used by law enforcement. Governments use algorithms for various causes: to improve the accessibility and experience of the city's services, to optimise processes and services, or to manage public space. Every quarter new applications are deployed while public technologists find new use cases where algorithms can be helpful for society. Algorithms are also starting to have an impact on the relationship between government and people. This change process, if anything, requires trust. In fact, without trust, we as government organisations are not even qualified to use algorithms and the urban data we may have.

In democratic governments, trust is secured through many constitutional safeguards. Good administration allows every citizen to understand the grounds of its decisions, and to have a possibility to contest such decisions. We must remind ourselves of this reality also in the context of increasing algorithmic impact in our societies. We believe such expectation of openness provides a good ground for trust also in societies, where automation plays an ever-increasing role.

Algorithms and AI, when used by a government, should be seen as nothing more than automation of human actions and decisions and therefore need to be held accountable according to the same standard as human actions. They must be fair, lawful and efficient, and verifiably adhering to the constitution and other relevant laws. Like we today hold people accountable over government decisions, so is accountability essential also concerning algorithms. Therefore, our algorithms must be made transparent and open to democratic control.

However, algorithms and AI are often complex technical systems, for which one may need extensive technological know-how to understand how they work. And as they are getting more technologically advanced, they become even less self-explanatory. In this reality, to make the trust equation work, we must seek for standardised ways for transparency and prioritise understandability. This is what the AI register is all about.

Transparency as the first principle for trustworthy AI

The primary purpose of transparency in the context of AI is to allow stakeholders to understand how the system works, how its decisions were done ('explainability') and to contest its behaviours ('accountability'). Transparency not only enables accountability for people directly impacted but also allows the people responsible, independent auditors and civil society activists to evaluate the workings of the system. In the public domain, transparency is a means for protecting democracy and everyone's right to influence society's development and their living conditions.

Transparency is often referred to as the most cited principle for trustworthy AI. In the **Principled Artificial Intelligence study conducted by Berkman Klein Center for Internet and Society at the Harvard University**, transparency and explainability principles were present in 94% of documents researched in the dataset covering thirty-six prominent AI principles documents published around the world[1]. The Principled Artificial Intelligence study is a great resource when looking for well-established frameworks or public debate around public control on AI.

Other specific resources to be mindful of:

- **The General Data Protection Regulation (GDPR)** provides European governments ground rules for transparency while positioning transparent information as one of the fundamental rights of a data subject (Article 12)[3].
- **The EU Commission's High-Level Expert Group on Artificial Intelligence** defines transparency through traceability, explainability, and communication in their **Ethics Guidelines for Trustworthy AI**[4]. Their **Assessment List for Trustworthy AI (ALTAI)** is another great resource for teams implementing AI[5].
- **The Council of Europe's report Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights** is a helpful resource for human rights impact assessments of algorithms6. It suggests public disclosure of information on algorithmic systems as one key measure to protect human rights in artificial intelligence.
- **National regulations concerning the information openness of government activities** should be referred to as the baseline for openness and good practice on information management in government[7,8]. One should also refer to possible ongoing national regulative work on automated decision-making in government.

Concept of an AI register

Although an AI register is a novel concept, we're not the first ones to talk about it. Over the last year, similar suggestions have been brought into discussion also by civil society organisations. Most recently, AlgorithmWatch and Access Now suggested the similar concept in their responses to the consultation of EU Commission High-Level Expert Group on AI during summer 2020[9,10]. We believe, however, this is the first time when such a concept has been implemented by governmental organisations.

Willingness to support public control over AI and enable this through transparent communication can realise only if we have the right information to support such transparency. In reality, many algorithms in use in government are lacking such information or at best; they're only made available to the development teams themselves. Therefore, the first step is to document what we're doing; what decisions we are making, what assumptions are those decisions built on, what risks do we see, and how do we plan to mitigate them.

AI register is a window into the artificial intelligence systems used by a government organisation. The register provides information which helps anyone to get acquainted with the overviews of government AI systems or to examine their more detailed information based on stakeholder interests. It is also a channel to provide feedback on those systems and participate in building trustworthy AI with the government.

A register enables government organisations to comply with citizen requests on the workings of algorithmic systems. It helps document (in a standardised, searchable and archivable way) the decisions and assumptions that were made in the process of developing, implementing, managing and ultimately dismantling an algorithm. With this documentation, transparency, and when applicable, explainability, can be given when asked.

However, we believe the benefits of a register don't end there. A register, when designed based on user-friendliness and their information needs, can also be a tool to proactively share transparency on the algorithms that an organisation uses, or even plans to use. It is this aspect of an AI register as a public resource that deserves care and attention if we want to bring government transparency a step forward.

See our summary on the primary purposes of an AI register from the perspectives of the two main user groups.

Before continuing to the suggested contents of an AI register, we want to pause to remind that such creation of transparent information is only the first step. An AI register is not a solution in itself to the harmful effects of algorithms, incomplete datasets or biased developers, but we believe it is a needed foundation to tackle these problems. Only with the right information, and with a requirement for government AI teams to address such important aspects in their development process, we're able to involve those people, who can help us make our AI trustworthy.

How can an AI register help citizens and civil servants in government organisations?

For citizens, it is a medium for civil participation and influencing how algorithms impact their living conditions. While governments process a lot of data of their residents, algorithms in public services must respect the same principles of responsibility, transparency, and security as all the other government services. Everyone should have access to understandable and up-to-date information about how algorithms affect their lives. And should be able to voice their opinion on it. For journalists, scientists, politicians, civil society organisations and activists, AI register can become an invaluable help while ensuring public control over societal impacts of AI.

For civil servants, an AI register can be a tool for managing AI governance in the development and operations processes. It will help teams create and maintain documentation systematically across all their projects, regardless of the technologies and partners involved. Civil servants and their vendors will find it helpful in guiding what kind of transparency is needed and how to provide this information understandably. It is also not a checklist, but rather the home of the process and metadata information produced by following a responsible and participatory AI development process.

Suggestion for the contents

Next, we will summarise what we see as the necessary and sufficient contents of an AI register. This approach is a synthesis of our literature review on AI ethics and transparency papers and a series of expert discussions to validate the contents. During this process, our focus has been to form a practical concept for defining what information is necessary to allow different stakeholders to understand and contest how the algorithms work.

Overview

The overview of your algorithmic system is the most important document and, arguably, the most difficult one to produce. The description should be written in clear and understandable language that is approachable by everyone regardless of their technical background or previous experience with algorithms.

Based on our experience of writing and testing such overview descriptions and their understandability, we suggest focusing on answering the following questions in the overview of your algorithm:

- **What is the main goal or societal benefit of the process this system contributes to?**
  (e.g. Parking control is needed because the number of cars allowed to park in the city is limited, keeping the city liveable and accessible)
- **How is it being used, and in which use cases?**
- **Who are the impacted people and the expected impacts?**
- **How does the system work at a high level?**

Accountability

The accountability section helps you assign the responsibilities for your system. Accountable people make sure the system is developed and maintained responsibly over its entire lifecycle. We should not only look at the internal actors but also the external partners with a key role in either development or maintenance of the system.

We suggest to have in place at least the following details in the accountability section:

- **Who are the responsible organisations and departments?**
  If there is a separation between the person ultimately accountable (i.e. business owner) and the person responsible for the technical development (i.e. technical owner), make it explicit.
- **How can the responsible parties be contacted (contact person, email, and phone)?**
- **Which external partners and suppliers are involved?**

Besides, we recommend assigning a responsible manager, technical owner, and possible external auditors for every system in the register.

Datasets

The datasets section summarises what information your algorithmic system is trained with and what data it processes when in use. Data is the bloodline of all algorithmic systems, and thus knowing the data is essential in understanding how algorithmic systems work. While algorithms used by a government organisation can utilise a wide array of datasets from open data to sensitive data collected in public services processes, describing a dataset can be a weary process. We encourage to start documenting a big picture of all datasets in use, and then, adding details to each dataset and their sources.

We suggest to add at least the following information:

- **What is the name of your dataset?** Name your dataset by reflecting the type and the origin of your data.
- **Description of the data represented in your dataset.** Focus on revealing the origins of the data and how your algorithm uses it. If possible, additional information about the data quality and any considerations you had about it.

We recommend assigning each dataset with further details on the source, the license, the version, and the provenance of the dataset. When possible, adding links to open datasets, anonymised samples, and a list of features will facilitate the traceability of your system.

Data processing

The data processing section addresses the ways your system uses automated data processing while creating its outcomes. It should be put in place with the intent for traceability that enables repeating the processing and resulting in the same outcomes.

- **What is the model architecture of your system?** The description summarises the key building blocks of the system while processing the input data into model outcomes.
- **How is the performance of the system measured and interpreted?**

Besides, the documentation should cover information on the key features (the measurable properties of a phenomenon observed), and potential parameters used in the model. When possible, projects should consider publishing the source code or, at a minimum, having it accessible when needed for independent auditing.

Non-discrimination

The non-discrimination section should provide focused information on the ways to evaluate the realisation of equality in the system context and the measures used to promote the realisation of equality. According to anti-discrimination laws, no one may be discriminated against based on age, origin, nationality, language, religion, belief, opinion, political activity, trade union activity, family relationships, state of health, disability, sexual orientation and/or other personal characteristics, and/or proxies to these attributes[11,12].

We hope the following questions can be useful when describing how you have addressed the realisation of equality and non-discrimination in your system context:

- **How is unfair bias interpreted in the system context, and what are the measures put in place to test for it?**
- **Is the system designed and implemented considering the accessibility of the service for people with disabilities?**
- **How were the people impacted by the system involved in the design and development of the system?**

Human oversight

The human oversight section should provide information on the ways humans are in the loop or in control, ensuring that your system does not undermine human autonomy or cause other adverse effects. Depending on the level of autonomy of your system, this can be realised in many different ways. We recommend familiarising with the EU Commission's Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment for further details5.

An AI register should contain a meaningful description of the following:

- **Description of the capability and support for human intervention in the system design and development, decision cycles, and in the monitoring of the system's operation.**
- **Description of the necessary competencies required for successfully performing the function and the training provided for gaining such skills and competencies.**

Risks

The risks section summarises the known risks associated with your system, and the methods used for risk mitigation. Risks can relate to both material and immaterial harm or damages caused by taking the system into use. In future, AI registers can provide an essential means for participatory identification of potential risks.

- **Description of the key tradeoffs between risks and benefits.** What were the key tradeoffs your organisation had to make while balancing the risks and the benefits during the development, implementation and managing the algorithmic system, and how did you come to such decisions?
- **Description of the risk level and risk management methods.** Focus on the risks producing legal or other significant effects, and the ones posing a risk of injury, death, or significant material or immaterial harm.
- **List of identified risks and the measures put in place to mitigate these risks.** Consider categorising your risks per type, scope, severity, and probability.
- **Privacy impact assessments, human rights impacts assessments and any other risk assessment methods applied.**

Explainability

For any system having significant social impacts on individuals and a high level of automation, such as any automated decision-making systems, explainability on the reasons that led to the specific outcome should be enabled. Such explainability can be provided in either AI registers or via other means.

References

In addition to the presented sections and the mandatory information, we suggest the AI registers contain references to further documentation regarding the privacy policies, information and system governance models, supplier contracts, and possible audit reports.

Our experiences and the path forward

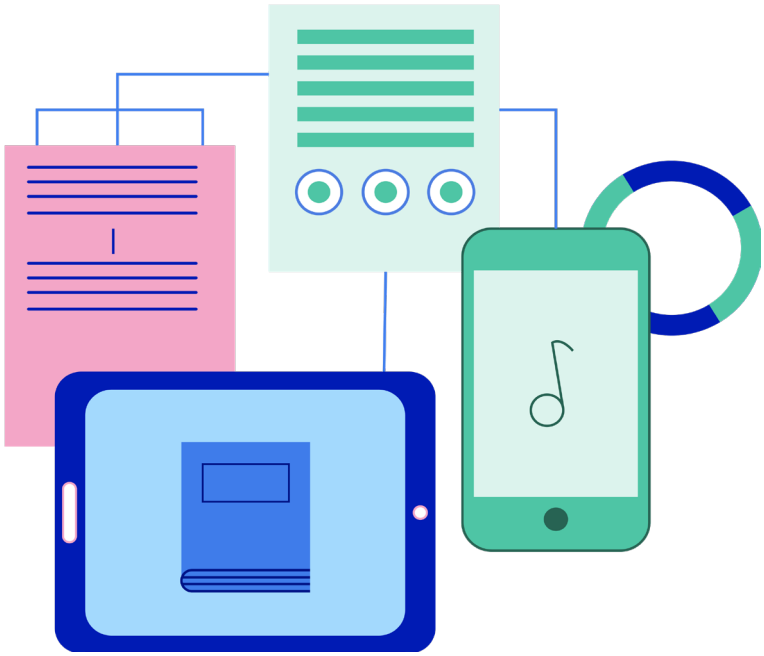
Having implemented the first versions of our AI registers in Amsterdam and Helsinki, we're also aware that this is only the beginning of our journey. Over the course of preparing these registers, we've come across topics, which will require further learning and continued development to make AI registers a success. We will summarise such key themes next.

- **Integrating AI registers into organisation's AI development processes** is critical in turning AI registers not only into a means of communication but also a key tool for helping AI teams in the good governance of their AI applications/ solutions/technology development. As an example, the assignment of accountabilities, or ways to deal with external suppliers, are such themes where current processes may need to be challenged while working with an AI register. We believe identifying such areas is also proof that what we're doing can have a significant influence on making our AI development processes more responsible.

- **Finding the right balance for the level of information** to be provided via AI registers will require persistence in collecting feedback from both external and internal users, and iterating based on that. We recognise there are a lot of varying levels of information needs, and also, potential bottlenecks for how to enable provisioning of such information in its most detailed level. For example, IP related concerns may need to be addressed via renewed supplier contracts, where larger transparency is expected. Starting from the level of detail relevant for the wide public, we expect to go in more detail when our teams progress in using the AI register as a governance tool for algorithms.

- **Using AI registers as a means for early citizen feedback for government AI projects** is something we regard as a potentially very useful application. For now, the systems which are published in our AI registers are at a minimum in the pilot phase, but most of them, already in use. However, we believe the AI register could be beneficial even earlier in the process to gain citizen participation and feedback. In the future, for example, systems could be searched from the AI register by their development stage.

- **Facilitating information sharing between government organisations** can potentially make the introduction of new trustworthy algorithmic systems more efficient. For example, finding potential suppliers for chatbot development will become easier when government organisations can visit each other's AI registers. We believe there can be some significant potential benefits also in collaborative learning on identifying human-rights related risks and finding good ways to address those, via greater visibility into AI portfolios of different government entities.



Conclusions

In this paper, we've presented a practical concept of an AI register as a means for realising transparency and public control for government AI. While developing the first versions of the Amsterdam and Helsinki AI registers, we've felt we're creating a concept and a solution of wider relevance. In the spirit of open innovation, we wanted to write this white paper to allow other public technologists to familiarise with our work and the rationales driving our decisions.

We're living unprecedented times where trust is becoming a scarcity in our societies. In November 2019, the cities of Amsterdam and Helsinki agreed to create ways to utilise the Internet in a more humane and democratic manner. As part of this initiative, cities aimed to develop shared terminology and ways to explain the use of artificial intelligence to the city residents. This is what has now materialised into the algorithm and AI registers.

Finally, we'd like to leave our readers with a thought from the City of Helsinki's Chief Digital Officer Mikko Rusama: "Without trust, there is no use for AI." We hope what we've presented here can be an essential step for further nurturing this trust in the cities of Helsinki and Amsterdam, and potentially, a new normal for broader government openness of algorithmic activities.

Contacts

City of Amsterdam
Linda van de Fliert
Innovation Officer
l.fliert@amsterdam.nl

City of Helsinki
Pasi Rautio
Project Manager, Data, AI and Robotics
pasi.rautio@hel.fi

Saidot Ltd
Meeri Haataja
CEO & Co-Founder
meeri@saidot.ai

References

1. Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika (2020): Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center Research Publication No. 2020-1.

2. Mozilla Foundation (2020): Creating Trustworthy AI - a Mozilla whitepaper on challenges and opportunities in the AI era (version 0.9).

3. European Parliament and Council of European Union Regulation (EU) 2016/679 (2016): General Data Protection Regulation (GDPR)

4. EU High-Level Expert Group on Artificial Intelligence (2019): Ethics Guidelines for Trustworthy AI.

5. EU High-Level Expert Group on Artificial Intelligence (2020): Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment.

6. Council of Europe (2019): Unboxing Artificial Intelligence: 10 steps to protect Human Rights

7. The Act on the Openness of Government Activities of Finland (Julkisuuslaki)

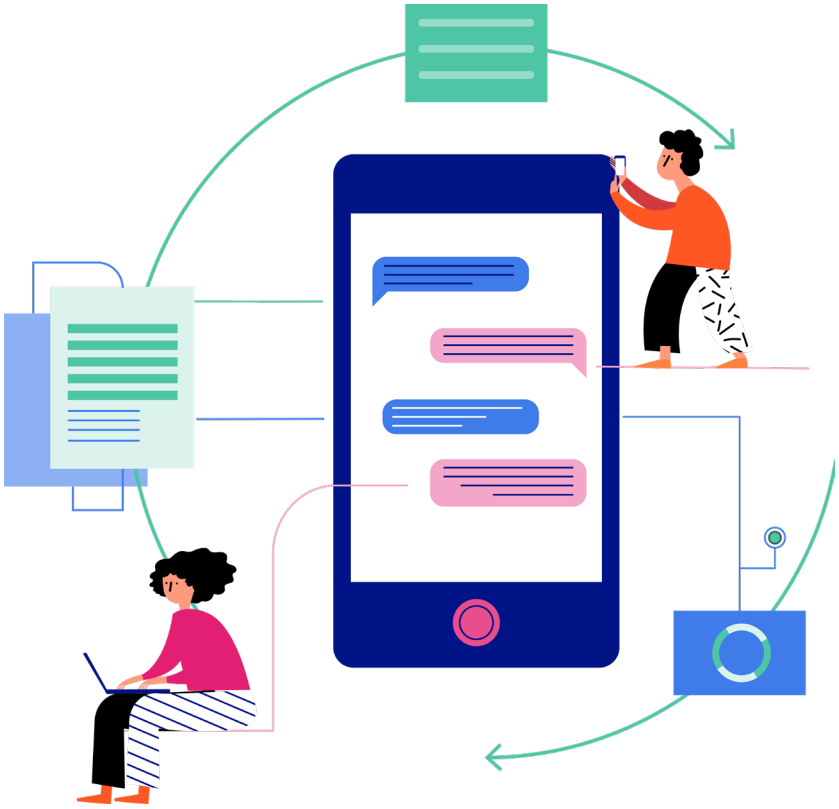
8. The Government Information Act of The Netherlands (Wet openbaarheid van bestuur)

9. Algorithm Watch response to the European Commission's consultation on AI (2020)

10. Access Now's submission to the Consultation on the "White Paper on Artificial Intelligence - a European approach to excellence and trust" (2020)

11. The Non-Discrimination Act (21/2004) of Finland (Yhdenvertaisuuslaki)

12. The General Equal Treatment Act of The Netherlands (Algemene wet gelijke behandeling)



Gemeente Amsterdam

Helsinki

Saidot