Beyond transparency: harnessing algorithm registries for effective algorithm governance

The Dutch government's creation of a registry for all their algorithms is a positive first step towards increasing public control. But the future will decide whether this is good enough, and if organizations need to take further steps beyond transparency to manage algorithm risks.

We argue that an algorithm registry can also provide a foundation for managing algorithms risks internally. On top of public transparency, there are more functions that organizations should consider when implementing an algorithm registry. Collaboration and knowledge management, risk assessments and general governance are as well functionalities that help organizations to gain more (internal) control over their algorithms. It's up to each organization to determine the best approach for them. But it goes without saying that with the right measures in place, algorithm registries help to increase public trust in algorithms and internally assure that they are used ethically and responsibly.

Frank van Praat
is a director and leads the Trusted Analytics team at KPMG.

Ylja Remmits
is a senior consultant in the Trusted Analytics team at KPMG.


TWORTHY
AI

INTRODUCTION

On February 15th 2023, Dutch State Secretary for Digitalization Van Huffelen made a bold commitment with potentially far-reaching implications. During a debate about the usage of algorithms and data ethics within the Dutch government, she promised that by the end of 2023, information on all government algorithms would be publicly available in the recently launched algorithm registry of the Dutch government ([Over]). This ambitious promise poses a significant challenge. The broad and complex nature of algorithms and their widespread use makes it difficult to obtain a complete overview of all algorithms used in the public sector. In fact, we observe that many organizations – also in the private sector – struggle to create and maintain an inventory of their own algorithms to start with.

There is an increasing demand from society and in the parliament ([Klav21], [Dass22]) for greater control over algorithms. In this light Van Huffelen's ambition is logical. However, it is open to debate whether the added transparency provided by the Dutch registry will actually effectively mitigate the risks inherent in the usage of algorithms by the public sector. We believe that a public algorithm registry is not enough to enable public oversight or to minimize the potential devastating impact of algorithms and AI. We merely believe that a complete and comprehensive overview of algorithms could be a great start for end-to-end algorithm governance. In this article, we argue that a registry's true value lies in its use as a tool for governance and risk management.

We believe that a public algorithm registry is not enough to enable public oversight or to minimize the potential devastating impact of algorithms and AI

TRANSPARENCY, PUBLIC OVERSIGHT AND INTERNAL CONTROL

The Algorithm registry of the Dutch Government ([Over]) was presented in December 2022. It was explicitly presented as a first step and it contains information on over a hundred algorithms from twelve different governmental organizations such as municipalities, provinces and agencies. Registration of all algorithms used in the public sector in this registry will become mandatory in the following years. In a letter to the parliament, Van Huffelen explains that citizens should be able to trust that algorithms adhere to public values, law and standards and that their effects can be explained. The registry gives citizens, interest groups, and the media access to general information about the algorithms used by the central government. The presented information in the registry empowers its readers to analyze the algorithms and pose relevant questions. However, transparency alone is only a small contribution towards the goals as stated by Van Huffelen.

The true value of transparency lies in the actions taken by stakeholders in response to the information they receive. For instance, citizens can obtain information about specific algorithms by contacting the relevant public sector organization, and (special) interest groups can challenge the way in which algorithms are deployed, or how specific data is used. Although the registry facilitates these actions, the current set up relies on action and effective challenging by third parties in the prevention of algorithmic risks and errors in the public sector. Relying solely on public scrutiny as a means of ensuring algorithmic accountability is a cumbersome and time-consuming process that demands a lot of effort from external stakeholders. Organizations must also take proactive measures to ensure that their use of algorithms is aligned with ethical and legal considerations.

To proactively manage the risks associated with use of algorithms in (governmental) organizations, it is crucial that organizations are internally in control over these systems themselves. With the introduction of legislation such as the AI Act this is not only the responsible thing to do, but also required by law. This involves obtaining an overview of the algorithms that are being used, to conduct a thorough risk analysis of each one in order to identify potential issues and prevent irreversible mistakes. In order to create and maintain this overview, a comprehensive registry would be a fitting tool. So, rather than being a transparency tool, the registry should be used by organizations as a means of assessing and managing risks. By also utilizing the registry for internal control purposes, organizations are forced to keep it up to date. The registry serves as an internal control

General transparency versus individual explainability

From the viewpoint of an individual citizen or customer, the value of an algorithm registry which provides general transparency is questionable. For them, explainability, which answers questions like "Why is my request denied?" or "Why do I have to provide more detailed information?" is more important. It is imperative for organizations to provide individuals with clear and concise information about decisions that impact them, whether they engage with an algorithmic system or a human one. In doing so, individuals can better understand how the system operates and any potential implications. On top of that, it is also crucial that the algorithm can provide meaningful feedback about how its output was derived. This feedback can be critical to ensuring that individuals – citizens and customers, but also users in the organization – have a comprehensive understanding of the decision-making process and can assess the fairness and accuracy of the decision that has been made. Individual citizens benefit more from information that is proactively shared and pertains to them, rather than general information that can only be found through search efforts on one of the thousands of government websites ([AmAL23]) or hidden in a corner of the customer care pages.

system, allowing organizations to remain in control of the deployment and risk management of their algorithms. While some form of oversight is required, it need not necessarily be public in nature. For example, a designated internal or external supervisor can approach their supervision in a structured risk-based way and monitor the system of internal control for each organization individually. The Dutch Data Protection Authority (Autoriteit Persoonsgegevens) started coordinating these efforts ([AP22]). In the case of general public supervision, it may be expected that certain high-risk algorithms or organizations may skip public oversight entirely because an overall structural approach to enforce compliance is missing.

FUNCTIONAL CHOICES FOR ALGORITHM REGISTRIES

In the previous section, the algorithm registry as a means to establish internal control was introduced. However, internal control (more comprehensively stated: governance) is a broad concept and can manifest in different ways depending on the type and risk appetite of organizations. It is important to note that solely having a registry is inadequate for governance purposes and that supplementary measures are often necessary. Generally, the following core functionalities can be distinguished for a registry:

Core function 1: Public transparency

Public transparency encompasses two main themes: 1) the provision of transparency and 2) accountability reporting. The former involves providing information on the algorithms being used, how they are being used, and how they impact citizens, businesses, or organizations. Aside from oversight, this also supports demystification[1]. Examples of algorithms in practice might help to give a more realistic image of the risks and issues that are already at stake which require genuine public debate. The latter involves being able to report to stakeholders on the ethical choices made, the technology involved, and the extent to which standards and regulations are being met.

For public control and trust, information on the use of an algorithm should not only describe what the algorithm does in factual terms, but also provide insight into its impact on citizens. It should answer the question, "How does this algorithm affect my life or that of my target audience?". Research shows that citizens primarily prefer information on privacy, human control over the algorithm, and the reasons for using the algorithm ([Ding21]). Organizations need to avoid technical or organization-specific jargon, such as acronyms or process names, and instead strive to use language that is accessible to a broad audience. The disclosure of such information is relatively general, and the reader needs to interpret whether this information applies to their specific situation. As such, a registry alone is insufficient to provide satisfactory answers to individual questions. In situations where citizens expect or require information about the impact of algorithms on decisions affecting their lives, explainability becomes a critical factor.

1 Demystification is one of the five keys tasks when embedding AI in society according to [Shei21].

Core function 2: Collaboration and knowledge management

An algorithm registry is able to serve as a valuable resource to enhance knowledge and expertise in the use of algorithms in organizations. By allowing searches for specific information about the inner workings and technologies of algorithms, a registry can provide a knowledge repository for developers to share new techniques and foster innovation. Furthermore, promoting knowledge-sharing can accelerate the adoption of algorithms within organizations in general.

The information that fulfills this core function is more substantive, comprehensive and detailed than the information that is provided for transparency and accountability purposes. The audience for this function typically consists of data engineers, -analysts and -scientists who are more likely to possess a greater level of familiarity with technical language. As such, the avoidance of jargon is less critical in this core function. The additional information that is added to the registry aimed specifically at knowledge-sharing and collaboration is not required to be included in the part of the registry that is available to the general public.

For this function, it is important that the registry is easily searchable for example on the grounds of an organization-specific taxonomy, and that a comprehensive overview of all algorithms in a specific category can be easily obtained.

Core function 3: Integrated risk assessment

Core function 3 of an algorithm registry is to provide organizations with a comprehensive view of the risks associated with algorithm use. By facilitating risk assessments and identifying measures to mitigate the risks, a registry can help organizations to better understand the potential risks and to take proactive steps to mitigate them.

To fully realize the benefits of this function, organizations must develop a methodology for classifying the algorithms used in their operations based on their risk levels. This could involve a system of classification such as low, medium, or high risk or based on factors such as the level of complexity, autonomy and impact of a particular algorithm. Depending on the classification, specific measures to mitigate risks may be required or recommended. The algorithm registry can support the risk identification and mitigation process by maintaining a standard risk and control measure catalog that connects directly to associated risk levels.

In order to effectively implement the integral risk assessment function of the algorithm registry, organizations must have access to up-to-date information about the algorithms in use, as well as the ability to monitor their use and assess their impact on decision-making. The registry must be regularly updated and maintained to enable organizations to assess the combined risks posed by multiple algorithms. In some cases, the use of multiple algorithms may increase the overall level of risk associated with a particular decision. The registry can play a pivotal role in enabling organizations to take a holistic view of their algorithm use and to identify and manage any potential risks that may arise.

Banks already have experience with this for their quantitative financial models. They use a model inventory that serves as a central registry to support so-called Model Risk Management (MRM). With MRM, banks keep an eye on the risks of models, track possible shortcomings and specific dependencies, and ensure that internal reviews (validations) are carried out ([KPMG19]).

Core function 4: Algorithm governance

Algorithm governance refers to the policies, procedures, and controls (core function 3) that an organization puts in place to manage the lifecycle of its algorithms. As algorithms become increasingly prevalent and critical to the functioning of organizations, there is a growing need for effective governance to ensure that they are developed, implemented, and used in a trustworthy manner.

This core function plays a crucial role in algorithm governance by establishing ownership, responsibility, and accountability for algorithms. By doing so, an organization can ensure that there is a clear understanding of who is responsible for the development, implementation, and use of each algorithm. This information can be used to make informed decisions about which algorithms to develop, how to deploy them, and how to monitor their performance.

In addition to providing insight into an organization's portfolio of algorithms and their ownership, the core function also facilitates active management of algorithm performance and added value. By continuously monitoring an algorithm's performance, an organization can identify potential issues and take corrective action before they become serious problems. This can help to improve the effectiveness and efficiency of algorithms, as well as enhance their overall value to the organization.

This can also support compliance, similar to how a *record of processing activities* – "verwerkingsregister" in Dutch – is used. The GDPR mandates organizations to maintain a comprehensive record of all processing activities under its responsibility. This register allows a full overview of what data is processed, for what purpose. The registry is a tool that supports compliance and a tool through which compliance with key aspects of the GDPR can be demonstrated.

SCOPING THE REGISTRY: DECISIONS ON WIDTH AND DEPTH

In scoping for algorithm registries, there are two aspects to consider, namely "width" and "depth". Width refers to the range of algorithms that are included in the registry (scope), while depth refers to the level of detail captured for each algorithm.

Width

Deciding which algorithms to include in a registry is challenging, given the broadness and variance of definitions of algorithms and AI. A socio-technical approach to algorithms, where the interplay between the technology behind algorithms (complexity), the processes in which they operate and their impact on society (impact), and the level of human oversight (autonomy) is of crucial importance, instead of the code of the algorithm.

- *Impact*. The impact of an algorithm on individuals or groups is measured based on the extent to which it influences various outcomes. This impact can be minimal, such as when the algorithm only affects internal financial reporting. However, the impact can be more significant when for example the results of the algorithm are used as input for policy development. The impact is highest when an algorithm is used in processes with a direct impact on the rights and obligations or decisions about citizens or businesses, or when the results of an algorithm have a significant impact on physical safety.
- *Autonomy*. The degree of meaningful human control and supervision over the algorithm. Non-autonomous algorithms are controlled by humans, and the results are valued by humans. In the case of autonomous algorithms, there are automatic results and consequences without an effective 'human in the loop' making decisions.
- *Complexity*. The complexity of the technology used. The simplest algorithms are rule-based algorithms that are a direct translation of existing regulations or policies. More advanced algorithms are based on machine learning or a complex composition of other algorithms.

In this perspective, it is worth noting that not all algorithms need to be registered. Organizations may choose additional criteria based on the above dimensions to limit the scope of their algorithm registry regarding their specific needs and goals. For example, the EU's AI Act only requires high-risk AI systems to be included in the proposed European database. Through a scoping exercise, organizations can define which algorithms should be included in the registry and what level of control is applicable. The next step is to determine what exactly is registered at what point during an algorithm (development) lifecycle, which we refer to as the depth of the system.

Depth

Next to the width decision to be made, the depth of the information per algorithm is as important to detail out. Three important factors are to be considered.

Firstly, the desired depth of information is closely related to the purpose of the registry and the recipient of the information. For example, if the registry is only aimed at providing public transparency, it probably does not contain the right information to be able to check the substantive functioning of an algorithm. Conversely, information aimed at risk management is likely to be incomprehensible to the average citizen, who is not familiar with the technology and jargon used. For knowledge sharing or (internal) validation of algorithms, it can go a step further. For example, if a data scientist wants to delve into a specific technology for peer review, the algorithm developer will have to provide all desired information via the registry.

Secondly, the quality of information is also crucial. While a comprehensive description of the algorithm may include many technical details, if the quality of that information is poor or lacking in substance, it may not provide meaningful insights into the algorithm's performance or effectiveness. For instance, the algorithm registry of the Dutch government regularly lacks in-depth and insightful information on specific algorithmic applications. For example, under the "proportionality" section of a license plate recognition algorithm, the only information provided is "No, this is an addition to manual enforcement." Such a description fails to provide any insights into how the algorithm's proportionality was determined.

Finally, practical considerations related to the feasibility and resources required for data collection and preparation should also be taken into account when determining the level of detail for an algorithm registry. To fill the algorithm registry with meaningful information, input is needed from various experts, and the content must be

aligned with various parties. It can be expected that it will take several days per algorithm to collect and enrich the information about an algorithm. In addition, the timing of inputting information into the registry should also be considered. For impactful use cases, it may be useful to keep the information in the registry up to date tracking the process of development, while in other cases, registration afterwards may be sufficient.

TRANSPARENCY ALONE IS NOT ENOUGH TO ENABLE PUBLIC OVERSIGHT ON ALGORITHMS

A registry of algorithms solely for the purpose of public control would be a missed opportunity. We argue that it is essential that responsible use of algorithms not only becomes a public responsibility but is also anchored internally in algorithm governance within organizations. An algorithm registry can be a powerful tool to assist in achieving this goal, when designed as such.

The Algorithm registry of the Dutch Government is a great start to inventory the use of algorithms in the Dutch governments. However, in its current form it is not enough to serve the Dutch government's ambitions. To truly build a registry that adds value, the government should decide what core (internal) functionalities the registry should have. Chosen functionalities in turn direct choices on which algorithms should be included in the registry (the width) and what information should be included (the depth). Explicit design choices guided by clear goals, ensure that the algorithm registry is not a mandatory one-off exercise, but a valuable tool for ongoing governance.

References

[AmAL23] De Avondshow met Arjen Lubach (2023, January 31). *Waarom heeft de overheid zoveel websites | De Avondshow met Arjen Lubach (S3)* [Video]. YouTube. Retrieved from: https://www.youtube.com/watch?v=CRmSlxj1mLM

[AP22] Autoriteit Persoonsgegevens (2022). *Contouren Algoritmetoezicht AP Naar Tweede Kamer*. Retrieved from https://autoriteitpersoonsgegevens.nl/nl/nieuws/contouren-algoritmetoezicht-ap-naar-tweede-kamer

[Dass22] Dassen (2022, October 28). *Kamerstuk 35 925 VII, nr. 26* [Motie]. Retrieved from: https://zoek.officielebekendmakingen.nl/kst-35925-VII-26.html

[Ding21] Dingemans, E., Bijster, F., Smulders, M., & Van Dalen, B. (2021). *Informatiebehoeften van burgers over de inzet van algoritmes door overheden.* Het PON & Telos.

[Klav21] Klaver (2021, January 19). *Kamerstuk 35 510, nr. 16* [Motie]. Retrieved from: https://zoek.officielebekendmakingen.nl/kst-35510-16.html

[KPMG19] KPMG (2019). *Model Risk Management toolkit.* KPMG Netherlands.

[Over] Overheid.nl (n.d.) *Het Algoritmeregister van de Nederlandse overheid.* Retrieved February 15, 2023, from: https://algoritmes.overheid.nl

[Shei21] Sheikh, H., Prins, C., & Schrijvers, E. (2021). *Mission AI: The New System Technology.* WRR.

About the authors

Frank van Praat is a director at KPMG where he leads the Trusted Analytics team. He has extensive knowledge of Emerging Technology Risk Management, Advanced Analytics Governance and the human side of AI.

Ylja Remmits is a senior consultant in the Trusted Analytics team at KPMG. Ylja has broad experience helping both governments and private organizations implement policy, processes and governance on algorithms and AI.