Frontrunner model for responsible AI governance in the public sector

The Dutch perspective

Popa, Diana Mariana

DOI
10.1007/s43681-024-00596-2

Publication date
2024

Document Version
Final published version

Published in
AI and Ethics

AI and Ethics
https://doi.org/10.1007/s43681-024-00596-2

ORIGINAL RESEARCH

Frontrunner model for responsible AI governance in the public sector: the Dutch perspective

Diana Mariana Popa[1]

Received: 31 July 2024 / Accepted: 27 September 2024

Abstract
Across the European Union, considerable discrepancies can be observed regarding the current state of AI adoption in the public sector and the complexity of functioning AI governance structures. This can be attributed to diverse levels of digitalisation, AI maturity and governance styles across EU member states. In the field of AI implementation and AI governance models in the public sector the frontrunner is the Netherlands, scoring first in the Global Index on Responsible AI. Analysing this example of good practices in terms of AI governance, with a focus on the delegation acceptance perspective, is of relevance for the state of art on AI governance within the EU. The article looks into the structure of the public Dutch Algorithm Register which currently contains over 400 entries, the AI framework for the public sector, supervisory structures in place and risks management approaches, addressing the importance of values in the development and deployment of AI systems and algorithms. The article demonstrates how in the case of AI also, early adaptors shape future behaviours, thus carrying a burden of responsibility when developing and deploying key enabling technologies in line with the core values.

Keywords AI Act · Algorithm register · Risk management · Governance · Values

1 Introduction

AI systems are used by public authorities in their daily activities, and often come into the media attention when the technology use has gone wrong, such as in cases where AI is used for police predicting, the child benefits scandal in the Netherlands that was characterized as a digital catastrophe [1] and the recent case of the use of AI applications for handling of Visa requests by the Dutch Ministry of Foreign affairs (Tweede Kammer Kemerstuck/debaat 20,240,516). This has a negative impact on technology acceptance and on subsequent delegation acceptance of AI in the public sector. Governments and public authorities must therefore proactively engage with citizens in the process of AI deployment, in order to timely build trust in both the technology and the institution deploying it. In most cases of AI usage by public authorities, the individuals (citizens) do not interact directly with the AI system themselves but receive a service through it. For political accountability of the governing act, transparency practices (regarding the use of AI in this case) giving insight into the public decision making process are proactively put into place in order to further avoid such situations, and also retroactively, as lessons learned and course correcting measures. For informing the public about how decisions are made and how citizens data is used, public authorities actively implement transparency measures through dissemination campaigns and digital (open) instruments. With the coming into force of the Artificial Intelligence Act (European [2], European Parliament [3]) in the EU, public entities have been simultaneously working on shaping internal structures for regulating and governing national AI landscapes and developed different open instruments for giving citizens insight into what and how algorithms are being used in the public administration act, in line with principles of political legitimacy and responsible AI governance.

The article further discusses two such instruments developed in the Netherlands: the Algorithm Register[4]– a national level register where public authorities publish the algorithms that they use and the Algorithm Framework (*AlgoritmeKader*) [5]– the actionable overarching

✉ Diana Mariana Popa
d.m.popa@tudelft.nl

[1] Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, The Netherlands

Published online: 16 October 2024

framework on AI governance and implementation. Earlier initiatives of algorithm registers are reported in the literature at municipality level, in the case of Amsterdam and Helsinki [6]. This is not accidental, since the Netherlands and Finland have a compatible approach to data, with Finland's approach reflecting the EU vision of human-centred digitalisation, focusing on the importance of data usage, sharing and protection [7]. Examples of transparency exercises in the Netherlands are abundant, given the Values-driven Digitalisation Work Agenda of the Dutch Government [8]. Examples include the Open Government Law [9], digital tools such as the online platform giving citizens insight into how different public organisations make use of their data such as Gegevensbijbesluiten. Additionally, the Values-driven Digitalisation Work Agenda of the Dutch Government [8] explicitly addresses the fact that everyone must be able to trust the digital world and must be in control of their digital lives, listing as actions in this scope: the development and use of online platforms where citizens can interact with the public authorities, the launch of the ELSA labs [10], where participants can collaborate to produce algorithms in line with human rights and public values and the development of AI labs around public tasks and "AI for good". Public policies putting the citizen at the centre of the digital governing act are numerous [11]. Different public entities within the national AI oversight system such as the Dutch Data Protection Authority, the Netherlands Court of Audit, Ministry of Interior are conducting investigations on the best way to involve citizens in the decision making process of AI development and deployment [12]. This data focused model of governance cascades down to all public entities, who implement the responsible digitalisation governance model by putting the citizen at the centre. In addition to these, deployers are conducting their own research studies targeted at specific population groups. Targeted research includes the use of AI by law enforcement, justice and security system, social benefits management and more. Numerous instruments for transparency and responsible deployment of AI system have been developed by the government and put at the disposal of both public authorities and public in general. Examples include the Toolbox for Responsible Innovation [13], the Impact Assessment Human Rights and Algoritmes developed by the Ministry of Interior [14], the responsible deployment of algorithms implementation framework listing 14 instruments that can be used by public authorities to deploy algorithms in a responsible manner [15, 16]. An example of inclusive AI design is in the case of the Scheveningen Living Lab, where in the development of the project for AI for crowd control an "ethics table" was organised, and different stakeholders came together: citizens, representatives of the municipality and the police, to discuss the impact of the AI deployed in the public space. This strategy of stakeholder engagement is implemented in various AI centred projects.

The rest of article is structured as follows: in the first chapter, the technical socio-cultural narrative of the technological development is set out, by identifying the main lenses of analysis, including: the deconstruction of trust as a key element for AI acceptance and its reconstruction in elements that are spread throughout the entire development—deployment—and—oversight AI ecosystem; the delegation acceptance perspective and social adoption perspective; the influence of the cultural stance. In the second chapter, AI governance is placed in relation to the digitalisation landscape: overall digitalisation level and digital literacy. The third chapter analyses in detail the two instruments—the Algorithm framework and the Algorithm register, identifying strengths and weaknesses, making an analysis of selected algorithms for illustration purposes and adding references to other forms of responsible digitalisation strategies. The forth chapter analyses two instruments in relation with the transparency mechanisms of Responsible AI governance. Section 5 brings in the risk management approach, identifying sources of risk throughout the ecosystem and product lifecycle and proposes mitigation measures. Finally, the conclusion section puts forward the responsibilities that early adopters have, alignment strategies and identifies future directions of research.

2 Theoretical frames

To analyse the role that these instruments have on AI acceptance, the article makes reference to three lenses identified as the main theoretical clusters in previous literature [17]: the user centric perspective, the delegation acceptance perspective and the societal adoption acceptance perspective. From these, the user centric perspective, reflecting individual usage of an AI system, is less relevant for the case of AI system deployed by public authorities in the exercise of the public administration act, unless referring to the way the public officer interacts with the system intermediating the service (agent based approach). The user centric approach could also be considered as having an indirect spillover effect in cases where the citizen, as end user of an AI application in a commercial or private setting, has had either a positive or negative experience with the technology and that result influences their perception of the usage of AI by public authorities, either with an impact on the self or in relation with the impact the technology has when deployed at societal level. However, the two scenarios of AI usage are very different, and this spillover effect therefore distorts perception that should be constructed separately, given the

distinct contexts in which the AI "consumption" or deployment takes place.

The delegation acceptance perspective is the one most relevant for the present analysis, divided between willingness to delegate and actual delegation, seeing the AI as an agent providing a service. In this approach, of relevance is not only trust in AI, but also in the deployer of the technology, namely the public authorities and government, that use the AI in the interest of the citizen (end-user of the result).

A multidirectional trust ecosystem is therefore constructed:

- the trust the user has in the AI as technology, therefore influenced by:
    - general levels of trust in technology,
    - previous positive or negative experience during interactions with technology (spillover effect);
    - overall level of understanding of how the specific technology functions,
- the trust the user has in the competence, intent and benevolence of the authorities deploying the technology influenced by:
    - both the overall reputation of the institution and any scandals or mishaps connected to technology and AI use.
- The trust citizens have in the government and the state of democracy– the functioning of the power division principles and active and impactful citizen participation in the political and social debate.

These factors come together to form a propensity to trust or to distrust novel technologies and their deployers, impacting acceptance levels. The societal AI acceptance perspective looks into the impact of AI on a macro level (societal level), and is influenced by the overall risk disposition towards technology and by technological and digital maturity and literacy. In this regards, the Netherlands has a low uncertainty avoidance index—53 on the scale developed by Hofstede [18], meaning that at country level risks taking is seen as acceptable and therefore rather inclined to taking risks when using new technologies. Seen from a different angle, this tendency puts that much more responsibility on public authorities deploying AI systems and on supervisory market authorities that have to have both AI products and regulating frameworks vetted before deployment and keep monitoring implementation. It makes therefore sense that the Netherlands has a leading role in the field of AI regulation and implementation. The style of AI governance should also be connected to the style of government in a certain country and state of democracy, namely in the case of the Netherlands: high level of citizen involvement in social and political debates and actions, healthy system of the checks and balances, high overall trust in the government and of culture characterized by a low power index and low uncertainty avoidance [18], values streaming from the polder model focusing on co-habitation, collaboration and consensus and high digital and AI literacy.

Trust is a dynamic relationship, with overspilling effects of breaches of trust. When choosing the values they wish to prioritize, public entities have to consider reputational risks, answering to both the political spectrum and to the citizen for their actions and decisions. What is also specific about AI usage in the public sector (the delegation perspective) in contrast to the user centred approach (most often analysed for commercial purpose AI consumption) is that impacted individuals most likely do not have a choice about the use of the AI for the service that they are soliciting (there is no opt-out possibility), or have no awareness, control or influence over the AI system using their data, fact that also reflects the power imbalance between citizens and the state [17]. It is also because of this lack of alternative that transparency is needed, in order to build trust in both technology and public institutions. A study on citizens evaluation of algorithms in Germany [19] showed that the personal importance of the area of application of the algorithm is important in the evaluation of the citizen towards that algorithm. Proactive transparency campaigns facilitate building of trust in the intent and benevolence of public authorities, and should include a careful balance of information quantity and complexity to avoid information fatigue on the one side and risk increase through disclosure on the other side. Transparency and explainability are therefore predictors of AI acceptance [17], and both predictors are addressed in the two instruments here under analysis, the centralised Register for Algorithms, and the actionable Algorithm Framework.

3 The digital landscape

Part of the digital economic security strategy of the EU [7, 20] AI is regulated by several pieces of legislation, the most important of which is the Artificial Intelligence Act (AIA). The AIA has brought substantial legislative challenges, both at European level during political negotiations of the framework and at member state level during preparations for implementation and entry into force. The speed with which the act is being translated into workable structures of responsibility is influenced by hard factors found at national level such as digital infrastructure, digital maturity and literacy, different governance models and available resources.

The level of digitalisation, technological and digital sovereignty is very diverse across the EU Member States, with differences as large as double between leading and lagging member states, for example in the Digital Economy and Society Index [21]. In the Netherlands, AI is part of the National technology strategy, in the category of digital and information technologies [22]. The Netherlands scores high on digitalisation and competitiveness, and on the use of responsible AI, ranking third in the Digital Economy and Society Index [21] second in the IMD World digital competitiveness ranking [23], and first in the Global Index on Responsible AI (2024). The Global Index on Responsible AI looks into three thematic dimensions Responsible AI capacities, Human Rights and AI and Responsible AI Governance across three pillars: government frameworks, government actions and non-state actors [24, 25] and of the three, the Netherlands scores highest for the pillar "government actions". The Responsible AI index lists 19 public initiatives in the Netherlands for Responsible IA governance, Responsible AI capacities and Human Rights and AI, with the Algorithm Register being a dimension of Responsible AI governance and a binding framework [24]. Of note here is the fact that Responsible AI Governance doesn't equal Responsible AI, but rather a reflection of the governing structures in place. This can also be a result of overregulating, which can slow down responsiveness to market dynamics. Deregulation of clutters legislation is one of the 2024 challenges for the Netherlands according to the IMD World Competitiveness index. Despite this top ranking, the way algorithms are used by public authorities is under constant public scrutiny, the topic being on the official agendas of government, NGOs and civil rights lobby organisations. Arguably, this healthy system of checks and balances and open scrutiny of the governing act has also contributed to making the Netherlands a sector leader: within the EU, the Netherlands ranks 4th on the use of AI applications by enterprises (EIT [26]) and in the same time "aims to be a global leader in working on careful control of algorithms and AI" (AP, 2023: 4). This ambition has been concretized by the lead position the Netherlands has in the Responsible AI global index, through (among the elements named above) the application of transparency principles into low threshold instruments and reports available to the general public and to different public and civic actors that might have an interest into to the way AI applications and algorithms are used in actions impacting individuals. The right balance of information quantity and information complexity captured in such instruments is therefore a matter of importance of public authorities [27] and one that also has to be considered within the national context. Also, the fact that the development and accessibility of these instruments in the Netherlands proceed the development and availability of their EU level equivalents is again a marker for the lead position the country has in the digital and AI landscape.

4 The algorithm register and algorithm framework

In alignment with the formal AIA requirement of publishing high risk systems in a public database and, not to lesser extent, in line with the democratic principle of transparency of the governing processes, in the Netherlands the publication of high risk, impactful and also general purpose algorithms used by public authorities (in their role of deployers) is done in a publicly available online register: algoritmes.overheid.nl. The registry can be seen as a precursor for the EU Database for high-risk AI systems, which is currently in development, and in line with the latter's requirements, as defined in Chapter VIII, art. 71, point 131 and Annex VIII of the AIA. Following the prescriptive characteristics mentioned in AI Act, the Dutch algorithm register is publicly accessible, user friendly, keyword searchable, and gives the user an overview into the content of the algorithm. In the same vein, the European level database is meant to facilitate the congruent efforts of both the Commission and Member States regarding AI. This EU database for AI high risk systems will also be publicly accessible, with exceptions at EU level regarding high-risk systems in the area of law enforcement, migration, border control placed in a closed circuit databased and with high-risk AI systems in the area of critical infrastructure only registered at national level.

Up to now, in the Dutch Algorithm register there are 488 entries (as of September '24), from 330 authorities at local and central level, with activities ranging from administration of the public spaces, taxes and social benefits, to official documents management and applications used by law enforcement. The information presented in the register is grouped in three main categories: general information (theme, begin data of use, contact information of deployer), responsible use (purpose and impact on individuals of companies, considerations on the needed usage, risk management approach and risk mitigation measures, type of human oversight, legal basis and links to relevant legislation including also sector legislation, impact tests used and links to the results of the tests– Data Protection Impact Assessments, Fundamental Rights Impact Assesments) and processing (including type of data processed, technical details and name of the developer and link to its website). These details enable process-based transparency Buijsman [27]. In the register, public organizations (but not only) choose which algorithms they publish, with this being enforced from 2025 onwards in the case of high-risk algorithms. In addition to these high-risk cases, such as the ones determining the risk profile for

additional controls, impactful algorithms are also included. Other algorithms that should be registered are the ones that come under media attention or are quoted by research, and the ones that are part of longitudinal experiments with a social impact [15]. The compulsory element regarding publication will be adapted in time, with exceptions applying for algorithms used in defence, law enforcement investigations, border management, intelligence work, in line with the registration exemption mentioned in Art 49 of the AIA.

In parallel with the register that government organisations will have to fill in, the Ministry of Interior is also working on a framework for the legal and ethical measures accompanying the use of the register (Auditdienst [28]). This is a complementary open instrument at the disposal of the public and mainly of the public administrators having to implement and oversee AI systems: the Algorithm Framework (*Algoritmekader).* Currently also in a development phase, it transposed both content of the main and additional legislative texts regarding AI into actionable pieces of information, which is perhaps the most challenging aspect of AI governance at the moment. The framework addresses the lifecycle of an AI system, its components, the rights streaming from the legislation, the different instruments used when making assessments about the risks and cross-cutting themes such as governance, data quality, privacy and data protection, technical robustness and safety. If the information contained in the algorithm register is rather process oriented, the information from the algorithm framework adds the governance layer.

To further ease usability and findability, the Algorithm Framework also lists possible organisational or technical roles that interact with the AI application at a certain moment in the products' lifecycle, such as acquisition advisor, data engineer, privacy professional, developer, legal officers, data scientists, ethics advisor, etc. These different organisational roles also have different levels of oversight responsibilities, even if not defined as such. Of note here is the fact that as AI adoption develops, new roles will be included in organisations, such as the one of Responsible for Algoritmes.

Despite the mapping exercise, the complexity of the information it contains makes the Algorithm Framework mainly an instrument dedicated to public entities (ministries, municipalities etc.), public servants or specialists working within the area of AI and data protection, and is of less interest to the end user, certainly when compared with the algorithm register. Given the matrix structure of the included components of the framework, including but not limited to: cross-cutting stakeholder roles of the system, rights of impacted individuals, life stage of the product, relevant legislation etc. the instrument is also a hands on way of facilitating the work of the public officers (in the broad sense) during project start, project documentation, decision making and so on. Since translating the multitude pieces of legislation is an intensively time consuming exercise, the development of the framework at central government level eases the work at regional/ local level. Based on an envisioned uptake of the use of both the algorithm register (not only in the case of compulsory high risk systems) and the framework as daily work instrument, what now is being a bottom up daily practice of policy making and execution at local level will move towards a (more) uniform practice when it comes to the governance and use of AI systems with consistent sector or regional patterns starting to form. The intersection of lifecycle stages, actions and responsibilities and roles makes it possible for different actors involved in the development and deployment of algorithms to quickly find the necessary documentation that has to be implemented in a AI or algorithm project.

In line with the digitalisation strategy of the Netherlands and because it is a subject pertaining to many aspects of social life across different public bodies, both the algorithm register and the algorithm framework require considerable resources, both financial and human resources, with a so called digital triangle for digitalisation (Ministry for the Interior as the coordinating entity, ministry of Justice and Security and the Ministry for Economic affairs and Climate) and specialists from different ministries, other public organisations using AI applications, and also researchers come together in a co-creation exercise to work on the public interface of the algoritmekader, to find and test solutions, present progress, relevant case studies and good practices to other target groups. Not the least, there are public releases of updated versions of the instruments, where all interested parties can attend and have their say. This last approach is in line with the overlap between the delegation acceptance and the societal adoption acceptance of AI, with citizens being able to shape the adoption of technology in line with their own interests (Koning 2024) Examples of public consultations of citizens at national level include the periodic "Demo release" of the Algorithm framework and the Algorithm register where everyone can participate online. Examples at local level are citizen consultation initiatives by different municipalities on the use of data and AI in "digital cities" and public initiatives to address inclusiveness of underrepresented groups [29]. Of note is that the local level initiatives feed in the national level instruments, by way of registration of individual algorithms. Involvement of citizens from the beginning of the development phase enables long term social and political acceptance of AI projects implemented in processes that affect society and the public space. From a 2023 report on the target groups, possible users of the Algorithm register are first of all citizens, then public authorities deploying the system, journalists,

supervisors, social organisations, researchers, politic representatives, private organisations, international organisations Ministerie van Binnenlandse Zaken en Koninkrijksrelaties [15, 16].

5 Values centred responsible AI governance

As reflected by the delegation perspective, there is a certain degree of passivity on the side of the individual impacted by the process facilitated by the AI system, in the sense that the individuals affected by the system are not the ones directly interacting with it, but the government agencies [17]. Through the availability of the register, transparency is brought into how the government uses algorithms in the decision making process affecting citizens, it serves to empower citizens and demystify the use of algorithms and AI applications [15, 30, 31] and improve algorithm literacy. Information into how algorithms are deployed, the type and effectiveness of human oversight [27], the checks and balances in place, raises trust within the public regarding the way their data is used during the decision making process, makes it possible for individuals to consent to a decision-making procedure that involves AI [27] and contributes to building public trust in both the new technology and the thoroughness of the public bodies. The existence of human oversight also facilitates acceptance of delegation of the task to the AI system. Additionally, in cases of litigation against decision facilitated by the use of algorithms, the understanding by the citizen of how the used algorithm worked is of importance for the building of his defence in court. This possibility is connected or limited by the AI literacy levels of the affected individual, as understanding of how an AI system was deployed is needed in order to allow for contestation of its use, literacy that arguably is difficult to achieve given that even in specialist circles, the "black box" effect of AI cannot be fully explained.

Transparency is needed for explainability, and both transparency and explainability are principles of responsible AI governance. Given that the explainability of AI is its most challenging aspect, facilitating understanding through clear and sufficient information helps build trust in the system itself and in the intentions of the deployer. With the highest score for Government Actions and third highest score for Human rights and AI in the 2024 Responsible AI report [24], the involvement of the public in the shaping of AI includes access to remedy and redress, transparency and explainability, and public participation and awareness campaigns developed by research institutes, universities and civil societies [25].

The use of AI in the public sector has as benefits a faster decision on a certain request from the individual and a facilitation of repetitive processes for the public institutions. As AI applications take over repetitive tasks, civil servants have more time to focus on core functions, with added value to the public service offered. Taking into account the organisational boundaries of a process, there are internal processes or activities such as writing of policy documents, writing of political programmes, transcription of debates and meetings, identifying patterns of organisational mobility or role specific tenure. External processes, or the so called public facing AI [32] include advice support tools ranging from questions from citizens regarding public services such as rights to subsidies, identifying alignment with political party ideology for determining voting choice. As more and more services are either taken over or processed this way, the AI system becomes a "digital colleague", with impact on the work patterns and behaviour of workers themselves, and also with concrete consequences regarding the results it uses. Examples of later are the determination of subsidies or discounts for certain services that are seen as legally binding in count cases when contested by the deployer, given the fact that the decision to use the system was a conscious choice, most likely with a risk mitigation plan proceeding deployment. Public organisations and commercial entities have to legally answer for this "digital employee", as they do for human employees.

The different level of AI knowledge of different deployers is another point to be considered when designing the instrument and considering the accessibility levels, as this level is quite different between central and local public entities. Training in both the legal frameworks and the practical use of the AI system and AI data readiness is therefore necessary, deployer obligation also included in the AIA (26): "deployers should ensure that the persons assigned to implement the instructions for use and human oversight as set out in this Regulation have the necessary competence, in particular an adequate level of AI literacy, training and authority to properly fulfil those tasks". Some deployers include in the algorithm register information about the training that staff has to follow before using a certain algorithm, such as is the case of algorithm processing requests for asylum, [33, 34]).

As technology is not neutral, but infused with values from as early as the design phase, algorithms and AI systems are also not value neutral and their use is embedded within a specific social—cultural context [19]. Moreover, in the case of AI systems and algorithms, they are also influenced by means of the values embedded in the training datasets on which the AI system is built. In a sociotechnical perspective, Kudina and van der Poel (2024) also make the case that the AI system should be developed while having the socio- cultural setting in which it will be deployed in mind, and as such impregnated with the (compatible) values

of that system. In a bilateral influencing relationship, values also shape technological use through the ways in which both authorities and the public use technology, what norms are enforced in a given ecosystem and how, in what measure individuals can exercise their rights in relation to the decisions taken and how often they do so etc.

Even if in the case of AI actual alignment with public values is not required by the legislation, it is a practice in line with democratic values that increases transparency of the act of governing and public decision making and it is a reflection of the governing style of a certain nation state, such as is the case of the Netherlands. The weight that this alignment receives in a certain context is influenced by the level of democratic index (citizen participation in the governing act) and as such, an argument for either developing AI systems with/in short supply chains and/or testing at the moment of acquisition for this alignment, especially in the case of long supply chains which might include several vulnerability points or in the case of suppliers who are based in divergent value systems. Admittingly, one should not fall in the cultural determinism pitfall when considering the weight that values have within the design of the system, as real life use often supersedes the lab setting expectations in which the technology was developed.

6 Risk management approaches

In the case of AI, a unique combination regarding risk management is observed: seen that AI is deployed as a product to be deployed on the unique market, it reflects the product risk management approach, and if it uses personal data for training purposes, or when deployed it uses personal data in a manner which could result in a decision with possible high impact on the individuals, the human rights framework also applies. The AI Act is a risk centred legislation, with risk coming up over 700 times in the text of the act, with high risk AI systems defined in Annex 3 and necessity for having a risk management system in place defined in Article 9. In practice however, accurately classifying a system as high risk, medium risk or general purpose is not that straightforward and organisations have own methodologies based on which the classification is documented. One proposed solution is to use risk thresholds. It is expected than that clearer guidelines and interpretation of how the risk classification applies to different case come from central levels of oversight, either national or European. The Act itself attributes both deployers and developers responsibilities regarding risk management. Deployers have to take special care when using high-risk system, such as conducting DPIAs and FRIAs, taking responsibility but not intervening on the system itself because that would make them a developer, but through additional measures– targeted communication campaigns, internal complain mechanism, additional staff for FRIAs. This does not normally happen in product safety legislation.

When it comes to AI systems, risk involves several aspects, including risks streaming from the technical provisions of the systems, to the negative consequences automated decisions might have on the individual. Given the broad use of AI systems, the risk levels (perceived as negative consequences or harm for the individual) are also varied, observation that is also reflected when analysing the type of algorithms now in the public register. This is reflected by the need to have sectorial approaches instead of blanket-regulations for all AI systems [2], something that is also addressed by the fact the AI product as such is also regulated by the sector legislations in which the systems are deployed.

In the public sector, the use of AI applications also bring risks together with benefits, with low risk applications of automating calculus to high-risk applications that have a high impact on the individual, for example when determining rights to certain services or processing biometric data. Of the currently included algorithms in the Dutch register, only 24 are high risk ones, which indicates that public authorities have a rather safe approach to publishing their data. As the Dutch Data Protection Authority reports [35], the high risk algorithms currently registered in the Algorithm register are only a part of the algorithms used by public authorities, with the year 2025 in mind as the year for making registration compulsory for all public entities. These existing high risk algorithms include systems that process biometric data, used for example in the process of identity and document confirmation. Seen that different public organisation use similar algorithms, the registry contains separate entries of each of these organisations, and comparing the information filled in by each one, differences in the argumentation towards risk management become visible. One such algorithm is the automatic ID document check and face recognition system [33], used in order to help public officers confirm real time the identity of the person coming at the desk for requesting registration in the public records registry or requesting a new ID, confirm the authenticity of the identity document issued by foreign entities and safely saving the data. Conformation of identity is based on biometrics processing of the face of the individual. The purpose of the system is to prevent look-alike fraud. Being classified as a deep learning, high risk system, the four eyes principle is applied and advice from foreign experts is sometimes necessary for document authenticity verification. In order to prevent automatic decision making, the final decision is always tested and approved by the public officer (human oversight) [33, 34, 36]) and in case of doubt, the Police or Customs officials

are asked for investigation. The system generates identification match and similarity matrix data.

Another high risk algorithm that caught the attention of the public officials, the media and NGOs because of the risk of profiling is the one deployed for the processing of Visa requests (Algoritme register, 2024, b.): Informatie Ondersteund Beslissen—Kort Verblijf (Schengen) Visum (KVV) and used as support process for the decision on the visa requests. The system uses profiling, based on data from the previous 5 years of other asylum requests and includes data on migration, and the data of the requester regarding nationality, gender, age group, marital status, occupation, previous result of visa requests, purpose of travel etc. [37]. Based on the imputed information a risk score is determined. As a Canadian ICT company was brought in as external consulting regarding the impact of the algorithm, which raised questions inside the House of Representatives related to the matter of data and technological autonomy. All of these details lead to the case coming under scrutiny also regarding responsible use of AI systems in the public sector. As in the case of the child benefits scandal, such public criticism can lead to algorithm aversion with overspilling effects, eroding public trust in both technology and public institutions. This case of the visa processing algorithm can be analysed from both the delegation acceptance perspective, and the societal acceptance perspective. Based on the public information, the person requesting a visa can understand how their (sensitive) personal data is processed and used in the AI supported decision making process. Also, the broader public receives information on how public authorities defend the public interest (based on what grounds individuals are awarded a visa for entering the Schengen space) therefore facilitating accountability.

Other AI applications of possible concern are those of smart camera deployment in the public space, when they can be used for emotional recognition. These are also frequently set up in shops and supermarkets in order to monitor behaviour, that could fall under the AIA category of emotional recognition, a prohibited category. At the moment there is no centralised overview of these cameras and the question is in what measure are both public authorities and the public aware of the way they are being deployed.

For the purpose of an in depth analysis at national level of the risks and mitigation measures in place when using AI and algorithms in public decision making, the singular use of the algorithm register seems restrictive, requiring additional requests for information, for example based on the Open Government law (Wet Openbaare Overheid) named above. Via this provision, internal decision making processes regarding the release, deployment and risk management of algorithms can be accessed. Offering clear and sufficient information beforehand (transparency) is therefore not enough, as accountability through explainability is also needed for additional inquiries and these factors are mentioned in many studies looking at delegation in relation to AI acceptance [17].

The way each algorithm's risk description and mitigation measures are logged in the register differs from case to case, also as a result of no strict instruction as to how the open fields should be filled in by the representative of the institution. The risk management description ranges from an empty field, to a short sentence stating that risk is minimum [due to the intervention in the last step of a public administrator], to a rich description of the impacts and measures taken to mitigate the risks, notably in the case of algorithms used by the Tax and Customs administration for distributing child benefits, in the latter case perhaps also as result of the media attention and rolling court cases following the child benefits affaire. This certainly is the most well-known case of use of AI and algorithms gone wrong, with grave consequences for the impacted individuals, not only financially but also leading to cases where minors were taken out of the family home as a result. This is an example of what high impact means and also of how algorithms influence the way individuals are classified in different groups such as "fraudster".

Data Protection Impact Assessments (DPIAs) and Fundamental Rights Impact Assessment (FRIAs) are also therefore necessary to be conducted before deployment, and they are named in the register as risk mitigation measures in cases where these were implemented, although the content itself is not detailed, in line with the compulsory element for high-risk AI system. This point touches upon the importance of values as they are embedded both within the design phase of the AI product itself and the legislative and operational framework regulating it. Just as the General Data Protection Regulation (GDPR) reflects the values of the EU, having at its hart the protection of fundamental rights and freedoms, through the AI Act specific harmful practices that come against EU values are prohibited. These prohibited practices are described in Article 5. In addition to these there are also high-risk systems, defined in Annex III. The AI Act represents therefore a hybrid approach between normal product safety and human rights legislation. For a product safety is paramount and the addition of human rights is somewhat contradictory to focusing on safety, requiring a balancing act between safety principles and human rights. All three AI Act, Dutch Algorithm Framework and Algorithm Register reflect the values of privacy (data minimization, risk management, Data Protection Impact Assessment, Impact Assessment human rights and Algorithms), justice and democracy (protecting vulnerable groups, civic rights, limiting state intervention in the private sphere, offering transparency into how personal data is handled). If the technical

aspects and values reflecting respect for privacy, democracy and justice are embedded within the different levels of legislation, addressing the ethical aspects proves more challenging since data ethics is always a political matter. From the ground level perspective, many discussions in the different structures working on the implementation of the Algorithm framework and the register in the Netherlands come back to issues regarding ethics, as public administrators face the challenge of embedding ethical principles into the was algorithms are used in everyday processes within their organisations. This is why many Dutch public institutions that implement AI systems (for example municipalities, ministries that fall either under different of under several legislative frameworks, such as the Ministry of Justice) have their own ethics committees for examining the ethical impact, while research institutions are perhaps more accustomed with having research proposals peer-reviewed by an ethics committee. This point is currently intensely debated among specialists working on the implementation frameworks of AI systems, as fluid and sometimes conflicting values are hard to implement into an overarching actionable framework, and rather require case by case analysis, which is time and resource consuming. Steaming from the product approach of the single market, standards are easier to develop and implement, while ethical principles need framing, contextualization and negotiation, as conflicting values at social level such as the security– privacy debate, industry practice sets the requirements for products and operationalize requirements. One should also reflect on the weight that the human rights component receives when designing AI application itself in line with legislative requirements in the EU, as the European approach is quite different than the ones of US or China [26].

As mentioned in the previous section, third party supply chains bring in their own risk factors, if the product is not build in-house but provided by an external supplier and its connected supply chain. In this regard, point 22 of the AI Act gives the case where an operator established in the EU can contract services to an operator outside the EU for an activity that uses high-risk AI systems. In the latter case, supply chain risks have to be considered, including: access to (other) systems or data through build-in backdoors, data that might be feed back to the supplier, data that has been purposely manipulated. Safety measures should be embedded from the design phase with data quality assurance in mind, human oversight, ethical analysis, human rights impact analysis, and also "hard" technical security measures.

One last point to be considered regarding the transparency principle in relation to the risk posture is that it can also be counterintuitive for public authorities to publish more information about high-risk systems than minimum necessary, as this could lead to vulnerability disclosure. The question is then when does information sharing, or transparency become a risk? The more information is presented in the register, the more the digitalisation strategy at national level becomes public. What could a threat actor derive based on the information presented in the algorithm register? The risk is mitigated in the case of high risk systems in the area law enforcement, border control and intelligence work, and critical infrastructure through the exemption of the obligations of the AIA in these areas. But as the digital footprint of direct AI usage and AI explainability increases, states should consider the potential danger that the increasing volume of information brings.

7 Conclusion

The article has showcased the state of the art regarding AI governance in the Netherlands, in line with the requirements of the AI act and also as a reflection of the digital, political, and social landscape that has an impact on the AI acceptance. At the moment, there are elements for improvement for both instruments analysed in this article, but nevertheless they supersede in maturity level the existing aggregating alternatives at EU level. As both the Register and the Framework are resources still in development, they do present a number of limitations: the search options are limited to the top layer of information, not reflecting the text within the descriptive fields, thus limiting research directions. The fact that both resources are only in Dutch limits their accessibility, which is regrettable considering the lead role of the Netherlands on this topic. Also, the fact that there is quite a lot of freedom regarding the information that is filled in, makes the content across the same field quite diverse. Regarding these points of improvement, an optimum threshold of information quality and quantity has to be considered in line with the transparency and explainability principles, so that both public entities filling in the information as well as final users of the register can find its use as easy and therefore useful as possible. As mentioned previously, the type of disclosed information can have an impact on the organisations risk posture, as inside information on the way decision support algorithms work could be used to manipulate the system (for example users requesting certain financial benefits). Certainly with the expected future compulsory character of filling in the register by all public authorities, future content analysis will be extremely important at both national and European level. It can be foreseen that through 1. the accumulation of information and 2. the awareness of the general public of the existence of both the algorithm register and framework and their potential use as a reliable source of information, the actual worth of the instruments will be proven, through future research and published case studies. AI literacy,

requested also by the AI act for human oversight, remains a key aspect to be addressed, since it entails different learning outcomes for different user groups: developers, deployers and the general public and requires a combination of technical, legal and ethical elements.

Transparency instruments facilitate the build-up of trust in the way AI systems were developed and in the way they are deployed by public authorities. Paradoxically, when trust exists, transparency is less important, as engagement with or reliance on the AI system will be done based on trust rather that the ability to control or verify the data and mechanism behind it. The Dutch perspective on AI governance is an example of good practices, but one that fits the context into which it is developed and is being deployed, and should not be transposed without previous scrutiny. Conversely, with the role of early adopter and frontrunner in the development and implementation of AI transparency instruments comes the responsibility of influencing current behaviour and shaping future behaviour, since the frontrunner will become an example to be analysed currently and retrospectively, writing lessons of success or of failure. As different member states develop their own AI governance models, future research should look into comparative analyses of the efficiency of alternative models, in relation to the context of implementation, connecting socio-technical aspects. As the Dutch AI governance and transparency instruments preceded in development and implementation their EU equivalents and also other national initiatives, proactively sharing good practices and lessons learned from technology deployment gone wrong can benefit the European market as a pull factor, since mature and tested initiatives can be easier taken up and implemented as standard product by others, saving time and resources.

Funding No funding was received to assist with the preparation of this manuscript. Part of the research conducted in preparation of this article was undertaken in the DigiNEB project.

Data availability The manuscript has no associated data.

Declarations

Competing interests The author has no competing interests to declare that are relevant to the content of this article.

References

1. Hirvonen, H.: Just accountability structures—a way to promote the safe use of automated decision-making in the public sector. AI & Soc. **39**, 155–167 (2024). https://doi.org/10.1007/s00146-023-01731-z
2. European Commission (2021). Proposal for a Regulation of the European Parliament and of the Council Laying down harmonised rules or artificial intelligence (Artificial Intelligence Act) and Amending certain union legislative acts. Brussel, 21.4.2021
3. European Parliament (2024). AI Act. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf
4. Algoritmekader: https://minbzk.github.io/Algoritmekader/
5. Algoritmeregister: https://algoritmes.overheid.nl/nl
6. Floridi, L.: Artificial intelligence as a public service: learning from Amsterdam and Helsinki. Philos. Technol. **33**, 541–546 (2020). https://doi.org/10.1007/s13347-020-00434-3
7. Okano-Heijmans, M., et al.: Clingendael. Strengthening digital economic security in Europe (2023).
8. Dutch Government (2024). Digitalisation Work Agenda. https://www.nldigitalgovernment.nl/overview/digitalisation-policy/value-driven-digitalisation-work-agenda/
9. WOO (2022). Wet Open Overheid. https://www.rijksoverheid.nl/onderwerpen/wet-open-overheid-woo
10. ELSA labs. Available at: https://nlaic.com/category/elsa-labs/
11. BLP. BL Planbureau voor de Leefomgeving (2023). *Betrokken burgers—Onmisbaar voor een toekomstbestendige leefomgeving.* Den Haag, 2023 PBL-publicatienummer: 4957.
12. Bernasco et al. (2024). Artificiële intelligentie, justitie en veiligheid. 1/24/ Wetenschappelijk Onderzoek- en Datacentrum. Ministerie van Justitie en Veiligheid en Boom juridisch.
13. Digital Government (2024). Toolbox for Responsible Innovation. https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/nieuwe-technologieen-data-en-ethiek/publieke-waarden/toolbox-voor-ethisch-verantwoorde-innovatie/
14. Straatman et al.: IAMA in actie Lessons learned van 15 IAMA-trajecten bij Nederlandse overheidsorganisaties. Rijks ICT Gilde (2024). https://open.overheid.nl/documenten/47e00e94-be86-4071-8e6e-e3da2a537771/file
15. ICTU, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2023). Doelgroepenanalyse Algoritmeregister. *Kwalitatief onderzoek naar hoe (potentiële) gebruikers het Algoritmeregister ervaren.* Available here. (consulted April 2024).
16. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2023b). Implementatiekader'Verantwoorde inzet van algoritmen'. https://open.overheid.nl/documenten/9b7b55fd-1762-499b-b089-2b7132c12402/file
17. Koenig, P.D.: Attitudes toward artificial intelligence: combining three theoretical perspectives on technology acceptance. AI & Soc. (2024). https://doi.org/10.1007/s00146-024-01987-z
18. Hofstede, G.: Cultures and Organizations: Software of the Mind: Intercultural Cooperation and Its Importance for Survival, McGraw-Hill Publishing Co (1991).
19. Wenzelburger, G.; König, P.; Felfeli, J.; Achtziger, A.: Algorithms in the public sector (2021). https://doi.org/10.1111/padm.12901
20. European Commission (2023). Joint communication to the European Parliament, The European Council and the Council on "European Economic Security Strategy".
21. Digital Economy and Society Index (2024). https://digital-decade-desi.digital-strategy.ec.europa.eu/datasets/desi/charts/desi-indicators?period=desi_2023&indicator=desi_bbspeed_10

00&breakdown=total&unit=pc_lines&country=AT,BE,BG,HR,CY,CZ,DK,EE,EU,FI,FR,DE,EL,HU,IE,IT,LV,LT,LU,MT,NL,PL,PT,RO,SK,SI,ES,SE
22. Bree, T. et al. (2023). TNO. Herijking sleuteltechnologieen Rapport.
23. IMD (2024). World digital competitiveness ranking. https://www.imd.org/centers/wcc/world-competitiveness-center/rankings/world-digital-competitiveness-ranking/
24. Adams, R., Adeleke, F., Florido, A., de Magalhaes Santos, L.G., Grossman, N., Junck, L., Stone, K.: Global Index on Responsible AI 2024, 1st edn. Global Center on AI Governance, South Africa (2024)
25. Adams, R., Adeleke, F., Florido, A., de Magalhaes Santos, L.G., Grossman, N., Junck, L., Stone, K.: Global Index on Responsible AI 2024 (1st Edition). South Africa: Global Center on AI Governance. Evidence Explorer. Public participation and awareness. The Netherlands (2024a). https://www.global-index.ai/thematic-areas-Public-Participation-and-Awareness.
26. EIT Digital (2024). Generative AI: Europe's quest for regulation and industry leadership. ISBN 978-91-87253-71-3
27. Buijsman, S.: Transparency for AI systems: a value-based approach. Ethics Inf. Technol. **26**, 34 (2024). https://doi.org/10.1007/s10676-024-09770-w
28. Auditdienst Rijk (2023). Bekendheid, toepasbaarheid en toegevoegde waarde handreiking 'non-discriminatie by design', 2023–0000154060
29. Goelzer, L. (2024). Inclusive participation for inclusive AI. Open research Amsterdam. Inclusive participation for inclusive AI - openresearch.amsterdam
30. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2023a). Handreiking Algoritmeregister Aan de slag met het Algoritmeregister. Accessed June 2024 at: https://algoritmes.pleio.nl/attachment/entity/f1a35292-7ea6-4e47-93fa-b3358e9ab2e0
31. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2023c). Doelgroepenanalyse Algoritmeregister Kwalitatief onderzoek naar hoe (potentiële) gebruikers het Algoritmeregister ervaren.
32. Sicular, S.; Krensky, P.; Judah, S.: Artificial intelligence requires an extended governance framework. Gartner. ID G00805856 (2024).
33. Algorithm register (2024b). Informatie Ondersteund Beslissen—Kort Verblijf (Schengen) Visum (KVV). Available at: https://algoritmes.overheid.nl/nl/algoritme/informatie-ondersteund-beslissen-kort-verblijf-schengen-visum-kvv-ministerie-van-buitenlandse-zaken/94596537#technischeWerking
34. Algorithm register (2024c) Case Matcher: Een zoek- en vindfunctie bij asielaanvragen. https://algoritmes.overheid.nl/nl/algoritme/case-matcher-een-zoek-en-vindfunctie-bij-asielaanvragen-ind/36268942#verantwoordGebruik
35. Dutch Data Protection Authority. (2023). Periodic insight into the risks and effects of the use of AI & algorithms in the Netherlands. Report winter 2023/2024.
36. Algorithm register (2024a). Geautomatiseerde documentcheck en gezichtsvergelijker. https://algoritmes.overheid.nl/nl/algoritme/geautomatiseerde-documentcheck-en-gezichtsvergelijker-gemeente-shertogenbosch/38325188#algemeneInformatie
37. Ministerie van Buitenlandse Zaken (2023). Informatie Ondersteund Beslissen Schengen Visum Kort Verblijf (KVV). Den Haag.