Algorithmic Risks Report Netherlands

Report July 2023

Dutch Data Protection Authority | Department for the Coordination of Algorithmic Oversight

Periodic insight into the risks and effects of
the use of algorithms in the Netherlands

AUTORITEIT
PERSOONSGEGEVENS

Table of contents

Introduction

- Central topics
- Summary of the risk profile and recommendations

Go to chapter →

1. General trends

- Awareness of undesirable effects
- Risk profile
- The chilling effect

Go to chapter →

2. Algorithms in practice

- Law enforcement
- Payment transactions
- Municipal social facilities

Go to chapter →

3. Policies and regulations

- International
- National
- Formation of the DCA

Go to chapter →

AUTORITEIT
PERSOONSGEGEVENS

Introduction

This report concerns systems and applications using algorithms and Artificial Intelligence (AI) that can impact groups and individuals.

In essence, these AI systems automate actions and decisions previously carried out by people, or perform activities that were previously impossible without AI systems. In short, the report concerns algorithms and AI. This ranges from relatively simple applications, involving the functioning of a single static algorithm, to highly complex applications of ***machine learning*** and neural networks. For the purposes of our risk analysis, we do not distinguish between the precise definitions of algorithms or AI. In all cases, this report refers to 'algorithms' and 'systems and/or applications'. In principle, this then concerns algorithms and systems and/or applications which through their actions may affect groups and individuals. The Department for the Coordination of Algorithmic Oversight (referred to in this report by its Dutch initialism 'DCA') of the Dutch Data Protection Authority (referred to in this report by its Dutch initialism 'AP') contributes to improving responsible use of algorithms. This is achieved by monitoring and periodically reporting on risks for and effects on public values and fundamental rights in the development and use of algorithms and AI. These values and rights include non-discrimination, transparency and explainability, prevention of deceptive or misleading information, freedom of expression and equality of opportunity.

This report describes algorithmic risks in the Netherlands. Relevant algorithmic risks are those that may affect individuals, groups and individuals or society as a whole. And that subsequently may disrupt society. The DCA has drafted this Algorithmic Risks Report Netherlands (referred to hereinafter as 'this report') to make relevant stakeholders – private and public organisations, politicians, policymakers and the public – aware of these risks in a timely manner so that preventative action can be taken. The first section of this report outlines the most important recent developments in the use of algorithms and algorithmic risk management in the Netherlands. The second section covers a number of specific areas in more detail through relevant case studies. The third section focuses on policy developments and institutional frameworks.

This report does not contain any predictions. Using current knowledge and available information, the DCA aims to provide a compact and understandable overview of current risks and control challenges associated with the use of algorithms. Where possible, the DCA proposes policies that can mitigate risks. The analyses and recommendations presented in this report offer organisations and policymakers insights into how to reduce the probability of the development and use of algorithms having undesirable effects on fundamental rights, public values and fundamental freedoms. In addition, this report also provides a method to improve understanding of algorithms and strengthen dialogue about opportunities and risks of algorithms in society.

We welcome your comments on this report and any suggestions for improvements. Please email these to: dca@autoriteitpersoonsgegevens.nl.

Key points

- It is possible to develop and use algorithms responsibly, thereby providing social value. But the use of these algorithms also comes with risks that need to be managed.
- The importance of adequate risk management will continue to increase in the near future, as algorithms become more powerful and new uses and risks emerge.
- Dutch society needs to take additional steps to control downsides in the societal impact of using algorithms and AI. This calls for: (i) expedited establishment of legal transparency requirements; (ii) clear regulation; and (iii) enhancement of internal and external supervision. It also requires organisations to invest in training programs and allocate more human resources to manage algorithm risks.
- Organisations that doubt the adequacy of their risk management are advised to be cautious in their use of algorithms and AI.
- In those areas where algorithms and AI can have significant societal impact, it is recommended that organisations are accountable before, during and after their use. Case studies show that, irrespective of their societal function, also these organisations too often still view their algorithmic processes as a purely internal, organisational matter.
- The AP welcomes the introduction of an algorithm register for public organisations in The Netherlands. In addition, the AP sees scope for a risk-based mandatory requirement for public organisations to register their algorithms. A deadline should be introduced for the initial entry of high-risk algorithms. Government algorithms and the associated risks will need to be identified by the first half of 2024 to determine whether they should be classified as high risk under the AI Regulation. These algorithms will, accordingly, also have to be recorded in the algorithm register. The DCA will emphatically monitor and report on progress within the government. Private organisations with a significant social mission can be expected to take proactive steps towards greater transparency and explainability.

Summary of risk profile and recommendations

This report provides an initial insight into the risks and effects linked to the use of algorithms and AI in the Netherlands.

This report is a product of the new algorithmic oversight activities of the AP, carried out within the DCA. The DCA identifies and analyses risks and impacts of algorithms through all sectors and domains of society. The DCA focuses on algorithms whose risks and effects directly or indirectly affect groups and individuals, or all people in society. As a consequence, algorithms applied primarily as a safeguard or control component in products, systems or processes often fall outside the scope of the oversight envisaged by the DCA through the report. In principle, these algorithms can generally be expected to have no impact on public values and fundamental rights. As part of its monitoring task, the DCA will publish this report every six months. The first full report will be published at the end of 2023 and will partly be based on information passed on to us by other regulators concerning potential risks. This first report is also intended to develop the reporting methodology.

It is possible to responsibly develop and use algorithms that can affect groups and individuals.

It has to be recognized that this requires effort as it involves risks that must be actively managed, prior to, during and after their use. This requires not only procedures and safeguards in the development phase of algorithms, but also an ecosystem of risk management and accountability once algorithms are actively being used. If insufficient attention is paid to the responsible development and use of algorithms, the Netherlands may miss out on important, positive applications of algorithms.

The preliminary initial insight is that Dutch society needs to take additional steps to control algorithmic risks.

Decisive political decision making – even in the outgoing phase of the Rutte IV government – plays a big role in this. There is a significant increase in awareness of and attention to algorithm risks. At the same time and not withstanding positive exceptions, there is across the board still a lack of mature, fundamental, compulsory algorithm-specific tools to adequately monitor and manage high-risk algorithms.

The process of working towards a manageable and regulated use of algorithms requires political support at this very moment, as algorithmic systems and applications are continuing to evolve rapidly and are becoming more widely used. The AP therefore calls on the outgoing Dutch cabinet, the Senate and the House of Representatives to make progress or even to accelerate efforts on policy, implementation and internal-external oversight with regard to algorithms.

Where fundamental, compulsory frameworks are still lacking, both public and private organisations need to determine an appropriate approach for the management of algorithmic risks.

This should be in addition to and in conjunction with compliance with existing laws and regulations, such as the General Data Protection Regulation (GDPR). These legislative pieces already contain some important mandatory risk-management related elements. It is clear that the level of maturity of risk management varies by sector and organisation. Nevertheless, there is still insufficient clarity on developments and applications to provide a comprehensive overall picture. It is a focus point for the DCA to contribute to improvement of control together with sectoral regulators who play a leading role in realising uniform risk management focused on sectors. The DCA expects that the greatest challenges are related to those sectors, applications and innovations where there is no sectoral regulator overseeing the overall functioning of organisations.

Initiatives such as registers, product standardisation, assessment frameworks and audit techniques are often still in the pilot phase.

To some extent, this first report is another example of this. Legislators and regulators need to pioneer with the development of a comprehensive set of management measures for algorithmic systems that are already being used in all areas of society. It is therefore important that policymakers remain focused on creating tools and principles, such as registers, review frameworks and transparency, mandatory in the development and use of high-risk algorithms. In this respect, 'the best is the enemy of the good', that's to say a lot can be achieved with simple rules. Organisations must then provide sufficient financial and human capacity to quickly comply with such rules. Where necessary, they should receive support with this.

A catch-up effort in achieving control of high-risk 'traditional' algorithms is necessary...

Many organisations are at the early stages of becoming more transparent about the risky yet simple algorithms they use – in some cases already for many years. In particular, this concerns organisations that have not set up an institutionalised approach and clear accountability for checking possible bias and fairness of algorithms targeting society, groups and individuals. This applies both prior to and during the use of algorithms. It is in addition noted that, from the perspective of a coordination supervisory authority overseeing the entire landscape where algorithms are being used, it is difficult to observe and determine the precise overall state of risk management of algorithms. This blurred vision is linked to the lack of a structured overview of the manner in which various types of organisations in the various societal sectors have structured their risk management (or have not yet done so). In this regard, more insight into the mechanisms and procedures used by organisations that are 'good examples' can also be useful to organisations where risk management is still being developed.

...and the emergence of sophisticated and complex algorithms adds to that challenge.

These more complex algorithms, based for example on self-learning neural networks, are particularly challenging in terms of testing, transparency and evaluation. This increases the challenge for organisations to implement effective risk management at a rapid pace. This is all the more true because risks materialise in a different way through complex algorithms in comparison with simple algorithms. The control mechanisms therefore needs to be structured differently. The same applies to transparency, accountability and explainability. For complex algorithms, such as neural networks, this is a specific field of study that is still under development ***(Explainable AI)***.

Organisations should exercise caution when using algorithms until sufficient risk management measures are in place.

The AP warns against using new applications as long as there is no assurance that the use of these applications does not violate fundamental rights and public values. potential risks concerning the violation of fundamental rights and public values have not been identified. This applies to the use of all types of algorithms in organisations' processes where the outcome of these algorithms may affect society, groups or individuals. Examples are systems for facial recognition or applications for detecting fraud risks. Not only is the identification of risks for fundamental rigths and public values to some degrees already a legal requirement, for example under the GDPR, it also follows from the lessons organisations should have incorporated from the disruptive case-based reasoning of recent years. As a consequence of these high profiles cases, there is no widespread awareness of the need for active risk management to avoid negative consequences, such as discrimination, arbitrariness and misdirection. As a consequence, an organisation must be ready to deploy such risk management techniques at an organisation-wide level before it develops systems or applications with algorithms.

Organisations that are frontrunners in the development and use of algorithms should be especially aware of the effort required from them to safeguard public values and fundamental rights.

The algorithms used by these organisations tend to have more impact. As a result, if used incorrectly, trust in algorithms can be seriously undermined. While new algorithmic techniques are powerful, they also have the potential to be more disruptive. For example by being able to create a false reality, ***deepfakes*** being an example., The more these kinds of sophisticated systems are implemented, the harder it becomes for people – whether citizens, customers or employees – to be certain about what can and cannot be trusted.

The Netherlands is certainly not alone in this situation – international cooperation is crucial.

Globally, the same challenges are at play. The European Union has the opportunity to take the lead in shaping regulatory frameworks if a political agreement is reached on the AI Act by the end of this year. The AP observes that the draft legislation increasingly contains elements that contribute to the protecting fundamental rights and freedoms, such as the right not to be discriminated against and freedom of expression. This is a move in the right direction.

Although the AI Act is important, it will not be a panacea for the current challenges.

First, for manyof high-risk applications, the compliance mechanism will be based on the system producer's self-assessment. As such, there is no independent assurance in advance that a system entering the market will not infringe fundamental rights. In addition, once agreement is reached, the regulations will not be binding for several years. In anticipation of this, it is therefore necessary to consider ways to give substance to the key provisions at this time already. Given the global development and use of systems and applications, such as ***large language models*** (LLMs), regulations must also be transnational to benefit from innovation. It makes sense to pursue a European or global approach, in line with developments in the EU and recent agreements on the subject by the G7.

Adverse effects of existing algorithms are often under the radar; risks of new technologies are directly in the spotlight. .

In the absence of an overview of all algorithms in use, it is impossible to determine which algorithms currently in use pose the greatest concerns with regard to public values and fundamental rights. It has already been seen in well-known case studies, for example in the fields of law enforcement, payment transactions and social facilities, that there have been algorithmic applications that present a risk of discrimination, unfairness or lack of accountability, either regarding individuals or groups of people. These issues will require systematic reporting and assessment in order to manage this from the perspective of continuous risk assessment. Potential data protection problems in LLMs are currently being investigated in the EU and addressed where necessary on the basis of existing laws and regulations. For the AP, the framework for any investigation into systems and applications comprises the GDPR and the Dutch Police Data Act [Wet politiegegevens].

Societally significant organisations often view the implementation of algorithms in their core processes as an internal choice and an issue of implementation and operationalisation of their formal duties, resulting in limited or no accountability before, during, and after the deployment.

The Dutch police's Crime Anticipation System (CAS) is an example of an algorithmic system deployed throughout Dutch society. The technical information publicly known about this algorithm is limited and not updated. Dutch financial institutions use algorithms to assist them in their legal duty to monitor transactions as part of their anti-money laundering and sanctions checks. This type of algorithm carries the risk that it may contribute to unwanted discriminatory effects. Financial institutions therefore use technical evaluation tools to test their models for bias, but again there is a lack of transparency (systematic and public) regarding exactly which evaluation tools they use. More transparency and the complimentary understandable explanations can increase public trust and help improve the quality of these tools.

QUICKLY TO THIS SUBJECT

General trends →
Awareness of undesirable effects →
Risk profile →
The chilling effect →

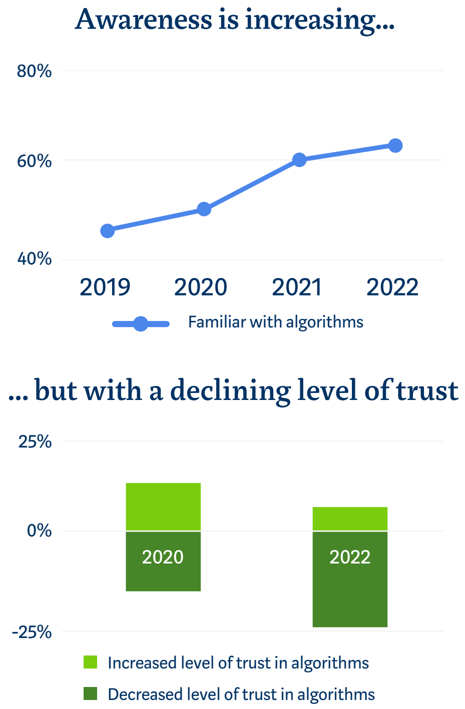
1. General trends

Awareness of undesirable effects

Awareness of the undesirable effects of using algorithms is increasing, while technology continues to develop at a rapid pace. The use of algorithms is increasing in all parts of society. And people are becoming increasingly aware of the effects of using algorithms. An example of a recent development is generative AI, for example in the form of LLMs. Many people in the Netherlands are now familiar with ChatGPT from OpenAI. By collecting, searching and analysing large amounts of text, a chatbot can formulate answers to questions or perform LLM tasks like the automated creation of presentations, plans, documents or letters. Applications are also emerging for images and video, which use an LLM via text-to-image to generate images or video based on text. These visual applications are also already finding their way into existing tools. For example, the use of LLMs creates opportunities to make processes more efficient. However, the use of this new kind of application of algorithms also raises growing and sometimes new societal concerns, for example, about the rapid spread of fake news, privacy breaches and copyright violations.

Confidence in algorithms is declining. Although awareness of algorithms has increased in recent years, the degree of confidence in algorithms has declined more than it has increased (see Figure 1). Studies conducted by Motivaction for KPMG show that absolute levels of confidence are also low. Only 20-30% of people have confidence in the use of algorithms, irrespective of whether the algorithms are those of government organisations, financial institutions, retail organisations, healthcare institutions, booking websites or tech companies. According to those people interviewed, by far the best way for organisations to increase this confidence is to be more transparent (around 40%) and to focus algorithms on improving society rather than detecting incidents (around 25%).

GRAPH 1: DUTCH CITIZENS AND ALGORITHMS



Awareness is increasing...

... but with a declining level of trust

SOURCE: KPMG (2022) – RESEARCH ON TRUST IN ALGORITHMS AMONG DUTCH CITIZEN

In recent months, several organisations and agencies have cautioned about the various effects of using-generative AI. Prominent developers and AI experts, fuelled in part by public and political pressure, have cautioned against its rapid development. They suggested a six-month development pause to focus on the impact on society. While these concerns are worth discussing, they can also distract from the actual issues facing society today, and from the instruments to develop and use algorithms responsibly.

In the Netherlands, the Minister of Education, Culture and Science immediately stopped the use of a fraud detection algorithm at the Education Executive Agency (DUO) at the end of June. This was in response to journalistic investigations into possible discriminatory effects of the algorithm. The decision to stop the use of this algorithm with immediate effect confirms the perception that awareness of algorithm risks is high, but also that internal risk identification and control within – in this case – government organisations is not yet sufficiently robust to be relied upon. Stable control of algorithm risks only occurs when organisations identify signals like this early on through internal evaluative risk assessment and control systems are periodically monitored and where necessary addressed by adjusting the use of the algorithm.

At the same time, the positive effects of AI should not be overlooked, in this report as well. Increasingly, applications that make a tangible contribution to individuals, groups and society are becoming apparent. Most obviously, these are applications in healthcare and for people with disabilities. But applications in industry, agriculture, infrastructure and civil society also show that innovation, if responsible, can definitely make a positive contribution to society.

Risk profile

To make complex algorithms manageable, additional effort is needed by society as a whole... With the lightning-fast developments and high stakes for society, all parties involved need to make additional efforts with regard to manageability. The range of applications and possibilities is growing daily, as is apparent in LLMs. These rapid developments, as well as regulations that are not always comprehensive or clear, put Dutch society in danger of risky algorithms entering the very fabric of society in an insufficiently controlled manner. Widespread use of system technology requires an appropriate framework, especially in the case of generative applications. Public values and fundamental rights must be protected and balanced with innovation.

...Meanwhile, we remain in a situation where even simple algorithms can do considerable damage. Simple systems and applications in which algorithms play only a limited role can also have profound effects on individuals, groups and society. Examples include the Dutch benefits scandal [Toeslagenaffaire] and fraud risk systems used by municipal authorities. These are often systems and applications based on historical data or implicit assumptions used to derive risk factors from that data. They do not actually identify fraud or other identifiable behaviour, but merely derive the risk of that behaviour from data about identified past behaviour. Unprecedented behaviour, spurious associations or unjustified assumptions often lead to a distorted picture of reality. There is a significant risk of discrimination, lack of transparency or unintended arbitrariness.

For many people, the use of algorithms remains hidden by a lack of transparency; they simply do not know that an algorithm is involved. In many cases, this lack of transparency conflicts with current laws and regulations. Non-compliance with current laws and regulations further widens the gap vis-à-vis future additional laws and regulations. For society, groups and especially individuals, the importance of transparency in systems and applications that can affect public values and fundamental rights is a crucial aspect. Without transparency, individuals and groups lose control, have limited ability to challenge decisions and it becomes practically impossible to contest outcomes and impacts. When there is a lack of transparency or an explanation about the use of an algorithm, manipulative applications can occur. Influencing behaviour (online or offline) through profiling and small or large decisions made by algorithms can have serious impacts, like in games of chance, for example, or even democratic processes.

Organisations at the forefront of adopting new technology and new algorithmic systems and applications should be aware of the extra effort required. To provide or use useful and responsible systems and applications in society, it is essential to manage the risks. In particular, new use of AI requires more from organisations than existing and known technologies. Risk mapping is a more difficult process, but it is necessary because many risks cannot be completely eliminated. This requires monitoring, control and a mature organisation, both prior to and during use. Such an organisation complies with applicable laws and regulations, has appropriate knowledge and skills, develops responsibly, and continues to monitor deployment. In this sense, even small or start-up companies can be mature organisations. Without organisations acting maturely and responsibly, society will be denied useful, valuable systems and applications. It is precisely to capitalise on the opportunities provided by algorithms that we need to invest now in managing risks, responsible development and robust organisations that are willing to develop within democratically established frameworks.

Increasing use and dependence on algorithms may go hand in hand with increasing market power for large tech companies. Many complex algorithms require huge computing power and huge amounts of data. The risk is that only a small group of tech companies have the computing and data power to develop the most sophisticated models. Fundamental rights, democracy and the rule of law also provide essential protection against concentrations of power and the misuse of this power. In the digital society, consolidations of power are not reserved for the State: private technology players also have very significant power.

The further emergence of algorithms could lead to fundamental shifts in our society. It is important to adequately educate citizens to make them not only digitally proficient, but also aware of how algorithms work and the risks they pose in society. Existing systems may also require a completely new approach due to the impact of algorithms, according to the International Monetary Fund (IMF), among others. For example, education (the labour market is changing), healthcare (different manner of diagnostics and different role of medical specialists) and government (relationship between taxation of capital and labour). Based on international AI standards, such as those of UNESCO, issues such as the impact of algorithms on people with disabilities and the climate impact should also be considered.

To manage risks arising from the use of AI, organisations should make efforts with regard to transparency, the dialogue with society and anticipation of new laws and regulations. Firstly, transparency in the development and deployment of systems and applications contributes to trust in innovation and understanding how systems and applications work in society. It also offers perspective for people who may be negatively affected or where there is a '***chilling effect***'. Existing laws and regulations for many sectors, areas and applications already provide guidelines concerning transparency. Compliance with these is therefore obviously very important. Secondly, organisations should be aware of potential risks and impacts to public values and fundamental rights when developing and using systems and applications. This requires a continuous dialogue with society about these risks and impacts, for example, by consulting stakeholders when developing and monitoring risks and impacts. Thirdly, the DCA strongly recommends incorporating the principles of impending laws and regulations into the development and deployment of systems and applications now. To take that important step forward, it may help to study or connect with the various key principles of international ***frameworks***. If at least some elements of new legislation or relevant frameworks are present in an organisation, this can contribute to responsible innovation, by following UNESCO's ten AI principles, for example. If organisations do not act in anticipation of future legislation, they will be unable to meet society's desire for innovation responsibly. This could cause society to miss out on socially relevant innovation, in key areas such as health, climate and democracy, for example.

The chilling effect

The chilling effect refers to the phenomenon that people adjust their behaviour when they feel their fundamental rights are affected, irrespective of whether this is really the case.

In a shopping street with visible camera surveillance, people behave differently because they are aware of the cameras. The very idea of being watched acts as a deterrent or disincentive to our behaviour. This effect applies equally well if the cameras are not functional.

A new form of the chilling effect occurs in the interaction between humans and algorithms: people may also possibly change their behaviour if they know or feel that they will be judged not by a human but by an algorithm. For instance, job seekers may discontinue a job application process if they become aware that an algorithm will evaluate them. The chilling effect then changes people's behaviour, without a direct violation of rights having taken place.

QUICKLY TO THIS SUBJECT

Algorithms in practice →
Law enforcement →
Payment transactions →
Municipal social facilities →

2. Algorithms in practice

Several types of algorithmic systems used in the Netherlands have recently been in the spotlight.

For example, in December 2022, the EU's Fundamental Rights Agency (FRA) published a report that detailed the risk of group discrimination by ***predictive policing*** (PP) algorithms. In addition, the National Coordinator against Discrimination and Racism (NCDR) said it received signals about discrimination by financial institutions in April 2023. Furthermore, in May 2023, the AP asked five municipal authorities for clarification on the use of the 'fraud scorecard', an algorithm to identify the risk of welfare fraud. Following these signals, this section offers a brief description of three types of systems for enforcement, payments and municipal social facilities. In this report, the DCA describes the algorithmic systems and application based on the overall algorithms coordinator task. This is not, therefore, a description from the perspective of the GDPR regulatory task.

Law enforcement

Bias and subsequent group discrimination are major risks of PP. Algorithms are widely used in law enforcement tasks in various ways, including for PP. A well-known example is the Crime Anticipation System (CAS) deployed by the Dutch police. This has been much discussed over the past year, for example in the Dutch House of Representatives, by the Netherlands Court of Audit [***Algemene Rekenkamer***], and by the FRA. Predictive policing is the use of algorithms to predict crimes by certain individuals or at certain locations and times. The CAS does the latter: it divides the Netherlands into 125-metre squares and predicts the probability of a crime occurring (being reported) in such a square. Worldwide, it is the only PP system operating on a national scale. The effectiveness of the CAS is subject to debate. Outside the Netherlands, questions have been raised about the use of PP systems. For example, the German Federal Constitutional Court [***Bundesverfassungsgericht***] labelled a PP system as unconstitutional in early 2023.

The Netherlands Court of Audit stated in 2022 that the police are not properly monitoring the CAS for bias and other risks. Due to bias, some people do not get sufficient protection and others are disproportionately surveilled. Bias can arise because the data used by algorithms give a distorted picture, with unintended discrimination as a possible consequence. The CAS does not use data from actively detected crime to reduce bias due to a positive feedback loop via its own actions. The CAS predicts which crimes will be reported primarily based on previous reports and public data from Statistics Netherlands [CBS], such as household composition, income figures and gender distribution. However, that does not prevent all bias. For example, willingness to report varies by neighbourhood and can again be influenced by the use of the system (feedback loop). The CAS also takes into account whether known suspects live in the neighbourhood, resulting in the possible inclusion of actively detected crime incidents. Public CBS data may be a proxy (misleading predictor) for certain ethnic groups. This could result in discriminatory police deployment.

Greater transparency is the first step towards better accountability. Proactive transparency to outside parties can help identify risks and understand the effectiveness or social benefit of using a PP system. For example, a first step is to update and periodically maintain a public record of the variables used in CAS and their calibration, following on from the one-time disclosure in 2021 based on a request under the Government Information (Public Access) Act [*Wob-verzoek*].

Also of interest is accountability for the assessment of the added value of algorithmic solutions. It is tempting for organisations, looking to improve processes to achieve certain goals (such as reducing crime), to fall prey to the idea that technological innovations are fundamentally neutral. And that when they are allowed to be used, they only offer opportunities. This is also known as '***technological chauvinism***'. By explicitly considering the risks of using algorithmic systems during decision-making, a more explicit assessment is created. The organisation can, or should also be accountable for this.

On the other hand, many enforcement algorithms have little risk, work well and have a positive impact. Using algorithms to automate routine tasks is common within enforcement organisations and has a long, successful history of use. Within the enforcement domain, it is therefore important to focus on the algorithms that specifically pose risks to public interests and fundamental rights.

Monitoring of payment transactions

Deploying algorithms to monitor payment transactions brings with it the responsibility of preventing group discrimination. Financial services algorithms are also being used to monitor payment transactions and potentially unwanted transactions more efficiently. In the financial sector, current developments are welcomed because of the legal obligation to thoroughly monitor Dutch payment transactions for illegal activities. For example, there are risks of money laundering, terrorist financing and more general fraud that banks should be alert to and actively look out for.

It is impossible to monitor each individual transaction due to the volume of digital payments. By applying pattern recognition algorithms based on data on unusual transaction patterns, ***modus operandi*** and risk indicators, similar unusual transactions can be quickly recognised. Suspected illegal transactions can be 'paused' by the system, which freezes the transaction. An employee of the bank then manually reviews whether the algorithm's estimate is justified. In the case of a '*false positive*', meaning a false classification as a potentially illegal transaction, the employee marks the report by the algorithm as erroneous. The transaction can then proceed. If the reporting of the algorithm may be justified, then further investigation can take place to determine whether the transaction indeed violates applicable legislation.

However, the use of algorithms for monitoring is not without risks. Where algorithms are deployed to recognise unusual patterns, there is also immediately the risk that bias can result in undesirable discriminatory effects. It is essential to identify potential risks to public values and fundamental rights in advance, as well as to identify and manage known and unforeseen risks during use. It is important to be aware that, as with other applications of technology in society, there could potentially be a chilling effect. This should also be taken into account when identifying and managing risks. Adequate risk management can prevent the use of algorithms with a high risk to public values and fundamental rights from actually affecting groups or individuals. The Dutch Central Bank [DNB] examines whether financial institutions have policies and procedures and take measures to manage the risk of discrimination.

Municipal social benefits facilities

Insufficient maturity levels at some municipal organisations stand in the way of responsible use of algorithms for key social facilities. Various media have reported on recurring issues with algorithm use for fraud prediction in the social facilities sector, among others. Even after the landmark ruling of February 2020 by the Court of The Hague on the Systemic Risk Identification (SyRI) anti-fraud system, municipal authorities throughout the Netherlands continue to use algorithms to detect fraud risks. These systems and applications vary in complexity, but in many cases can have a major impact.

Many of these systems use algorithms to estimate the probability of fraud by individual recipients or certain groups of recipients based on historical data. Past experience, perceptions of risk, as well as indicators with no known origin or value are used to identify individuals or groups as potential fraudsters. For example, a profession such as hairdressing may pose a higher risk of fraud than that of a lawyer. People with owner-occupied houses may be much less likely to be flagged as potential fraudsters than people living in rented houses or mobile homes.

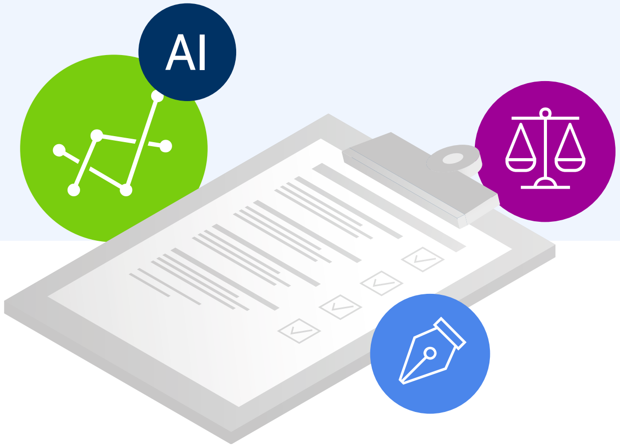
The use of these systems that estimate an individual's potential risk of fraud can impact the lives of individuals, families and entire groups in our society. Being flagged as a potential fraudster can cause people considerable emotional and financial damage. People are under suspicion from the outset and, due to the opacity of systems used, find it difficult to find out why they are classified as fraudsters and what they can do about it. The scale at which this happens can also translate this individual harm into substantial societal harm. This is evident from the recent controversy concerning social benefits and the SyRI system.

The risks of algorithms to predict fraud are high and their use therefore requires sufficient *checks and balances*. Municipal organisations must be set up to continuously monitor for and identify risks of such algorithm use.
The checks and balances required for this should apply to both the development and the adoption of such systems, since risks can arise and be noticed at both stages.

In both the development and adoption stages of systems to predict fraud, the usefulness of an algorithm must be assessed against the risks to fundamental rights and public values. Systems that predict fraud pose major risks to public values and fundamental rights. They should therefore be constantly assessed. Appropriate measures should be taken where necessary. Sometimes abandoning the use of these high-risk systems can be a legitimate outcome of an informed decision-making process, if it turns out that the benefits do not sufficiently outweigh the risks to public values and fundamental rights. Again, the aforementioned technological chauvinism must also be avoided.

QUICKLY TO THIS SUBJECT

Policies and regulations →
International →
National →
Formation of the DCA →



3. Policies and regulations

Creating new, complementary laws and regulations to manage algorithm risk is underway and is a global challenge.

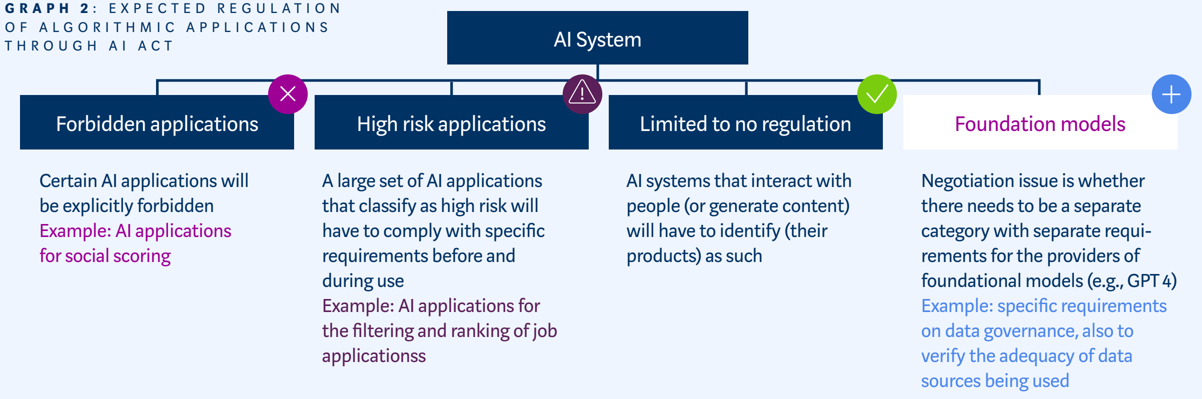
International AI frameworks, such as those of UNESCO and the OECD, provide guidance on the requirements to be set. The task now is to translate this into binding national and international legislation to create additional safeguards to complement relevant existing laws and regulations.

International

The AI Regulation will ban or more heavily regulate algorithmic systems and applications. Systems and applications that unduly restrict people's free choice and exploit or manipulate people will be banned. In addition, the AI Regulation regulates a number of 'high-risk' systems. Almost all specifically regulable applications of high-risk systems listed in the AI Regulation affect public values and fundamental rights, such as the right to data protection or equal treatment, for example, in the areas of biometrics, recruitment and selection and education, but also systems used in critical infrastructure or added to other systems, such as cars and emergency medical equipment. Such systems must meet additional conditions. These include requirements regarding transparency, human oversight and accuracy of data. It is important that developers continuously monitor the risks of algorithms and mitigate risks to security, fundamental rights and other public interests. The organisations that will use these systems in practice (users) will also have to follow rules.

They may be required to conduct a fundamental rights ***impact assessment***. In addition, the AI Regulation is likely to include additional rules on **general purpose AI** and ***foundation models***, including generative AI. See figure 2. Oversight of the AI Regulation is likely to be largely on a national basis. Talks on this have already started in the Netherlands.

GRAPH 2: EXPECTED REGULATION OF ALGORITHMIC APPLICATIONS THROUGH AI ACT



AI System

Forbidden applications

Certain AI applications will be explicitly forbidden
Example: AI applications for social scoring

High risk applications

A large set of AI applications that classify as high risk will have to comply with specific requirements before and during use
Example: AI applications for the filtering and ranking of job applicationss

Limited to no regulation

AI systems that interact with people (or generate content) will have to identify (their products) as such

Foundation models

Negotiation issue is whether there needs to be a separate category with separate requirements for the providers of foundational models (e.g., GPT 4)
Example: specific requirements on data governance, also to verify the adequacy of data sources being used

The AP welcomes the increasing focus on protecting fundamental rights and freedoms in the negotiations on the AI Regulation. Specifically, developers must ensure that they monitor and address risks, such as risks to data protection, discrimination, democracy and rule of law, etc. And regulators will have to check that this takes place properly. Negotiations regarding the AI Regulation are currently ongoing. It would be a significant outcome to secure a final political agreement around the end of 2023, and rapid implementation, in order to respond to the rapid pace of innovation with regulations that effectively protect public values and fundamental rights.

At the same time, we should not expect miracles from the AI Regulation, especially since regulatory compliance relies primarily on producers' own judgement when it comes to high-risk systems. Producers of high-risk systems for recruitment and selection, educational assessment, fraud detection, and also ***predictive policing***, for example, are not required to have a system tested externally before they market it. An own assessment of compliance with essential requirements and standards suffices. As a result, there is an explicit possibility that unsuitable high-risk algorithmic systems will enter the market and be used by private and public parties. Such a system disappears from the market only when the regulator determines that the AI system is faulty. However, damage may have been done by then.

The future AI convention will set the tone internationally for the regulation of AI. A Convention on Artificial Intelligence is also in the pipeline. A preliminary version (zero draft) of this convention was shared by the Council of Europe in January 2023. The convention builds on the European Convention on Human Rights and Convention 108 on automated processing of personal data. Participating countries must ensure, among other things, that fundamental rights are protected when government organisations use algorithms in their decision-making. These countries must also ensure that individual freedom, human dignity and autonomy, democratic processes, public health and the environment are protected when governments, but presumably also private parties, use algorithms. In addition, the convention contains principles on issues such as equality, privacy, transparency and monitoring. The convention further prescribes measures, such as for legal protection and human review. Because the convention will also apply outside the EU and involves countries such as the United States and Japan, it has the potential to become an international standard for the development and use of algorithms in both the public and private sectors.

National

At the national level in the Netherlands, there is a clear commitment to improving algorithm risk management, starting with the government. For instance, the relevant parties are pursuing human rights testing of algorithms and are preparing an implementation framework for algorithm use. The AP deems both of these aspirations to be goodsteps towards providing clarity to government organisations. It also helps that there is one clear assessment framework to which the government explicitly refers, the Human Rights and Algorithms Impact Assessment (IAMA). However, the AP does highlight: (i) a proportionate application of such review mechanisms and frameworks; and (ii) a risk-oriented phasing-in by organisations. It must be proportionate and manageable for organisations. Adhering to the IAMA is a careful, comprehensive but also time-consuming process. This is not proportional for every type of algorithm. It is emphatically proportionate for key algorithms that are part of socially critical processes associated with the public functions of government organisations. In the DCA's view, therefore, the first step is for organisations to classify their most important algorithms, apply a human rights test and follow implementation frameworks. Government organisations need to identify high-risk algorithms for this, which is also necessary in the lead-up to the AI Regulation. The final regulation is expected to come into force in early 2024.
he AI Regulation will become applicable after a period to be determined (2-3 years). Some government algorithms will be classified as high risk under the AI Regulation. A full overview of algorithms used and developed, with associated risks and classification, will need to be produced in the first half of 2024. This is necessary to prepare for obligations under the AI Regulation and to register these algorithms in the algorithm register. The DCA will emphatically monitor and follow progress within the government and report on it in subsequent reports.

Public-private partnerships can assist start-ups active in the algorithm field with careful product development and regulatory compliance. An example is the ELSA labs. These are a form of incubators, where ethical, legal and societal concerns are explicitly considered from the start when developing algorithmic systems. Start-ups are not burdened with transition challenges from legacy systems. Such labs provide an opportunity to develop a system on a small scale in a managed way and draw lessons from real-world experience, by using synthetic data, for example. This will allow for controlled innovation. The cooperation between government organisations, knowledge institutions and the private sector is very helpful in this regard.

The AP has been designated as the coordinating party with regard to algorithm supervision. The DCA was launched in early 2023 to fulfil this new task. Supervision in the Netherlands is organised thematically and by sector. Sectoral supervision involves a great deal of specific knowledge, but also the danger that the overall risks to more general public values and fundamental rights are neglected. Increasingly, executive boards, market regulators and state inspectorates are cooperating more frequently and more intensively in the supervision of algorithms, partly through working groups. As external regulators, the goal is to share knowledge and jointly control algorithms and the effects of algorithm use. In the second half of 2023, the DCA will compile an overview of how algorithm risks are viewed from different supervisory roles based on a survey conducted among supervisors.

Regulators and market participants are preparing for new regulations. A swift and unified approach is needed to align new regulations with existing ones, provide explanations and set up national supervision correctly and in line with other affiliated countries. For systems and applications in the private sector, it is important to be mindful of public values and fundamental rights in addition to existing regulations and standards, and to secure them in anticipation of – and in the spirit of – new regulations, such as the AI Regulation. The manner of implementation may vary by supervisory area and sector, and requires the same innovative approach from supervisory organisations as when deploying new technology.

The AP is positive about the Dutch national algorithm register and sees scope for a mandatory but risk-based acceleration of entries in the register. The Dutch national algorithm register is the basis for oversight and understanding of the algorithms used by government organisations. A positive aspect is that the register also specifically describes how and why an organisation applies the algorithm. By the end of June 2023, about 120 algorithms had been registered, mainly from some of the larger Dutch municipalities. Of concern is that the registry currently includes algorithms that carry limited risk, such as a system to automatically add document numbers to decisions. Initially, the focus should be on high-risk systems. In this context, classification can provisionally be based on self-assessment, using the provisional list of high-risk systems under the AI Regulation. The AP favours mandatory registration of high-risk systems in the Dutch national algorithm register. There should be a deadline for initial entry in the register. This should be a top priority, keeping in mind that the most important thing at first is to get the basic data on the riskiest algorithms recorded as soon as possible. This also then gives regulators an objective starting point from where they can engage with organisations.

Formation of the DCA

The AP is the coordinating body for the supervision of algorithms as of 2023. The AP is fulfilling this role with its new DCA department. This task was allocated to the AP in response to the desire to better protect public values and fundamental rights when developing and using algorithms.

The focus of the DCA is on improving the protection of public values and fundamental rights. Such as preventing discrimination and arbitrariness and promoting transparency. Furthermore, the DCA considers the fairness of algorithms and the prevention of deceptive or misleading information. The activities of the DCA are separate from the intensification of AP's supervision of algorithms processing personal data.

Risk identification is an important part of the DCA's activities. This particular activity is related to strengthening supervisory collaboration and promoting and facilitating joint standard setting nd ***guidance*** for organisations. The focus in risk identification is on identifying and analysing cross-sector and overarching risks and effects of algorithms, and on developments in policies and regulations.

The DCA is working on networks and structures to receive and discuss signals. In its coordinating role, the DCA aims to liaise with regulators, sector representatives, implementing organisations, interest groups, civil society, academia and specialist organisations and institutes. The DCA will construct these networks and set up reporting structures. International networks are also an important resource for highlighting internationally identified risks. By continuing to monitor signals over time, the DCA can identify trends. This report will provide a periodic overview of these risks and impacts, both in the context of the overarching part and in the part identifying and discussing specific examples of algorithm risks based on media, stakeholders, international monitoring and supervisory practice. The AP expressly does not discuss these specific examples in this report from the perspective of its role as GDPR regulator.

Cooperation with regulators is essential. Existing cooperation will be intensified, both nationally and internationally, and additional cooperation established where necessary.
The supervisory landscape in the Netherlands is extensive, and is characterised by its strong sectoral and thematic knowledge. This can be maximised through cooperation. To ensure effective oversight together, the DCA will begin the process of charting the landscape of algorithm oversight. This should clarify which regulators play a role, what tasks are required and where they are allocated, and it should also allow the opportunity to look for any possible gaps in supervision.

The DCA is also in the process of developing a perspective on algorithm risk. To support risk assessment, the DCA is working on a framework that can be used to examine the types of algorithm systems and control systems, and the manner in which algorithms are developed and used. The combination of these three elements determines the extent to which risks to values and fundamental rights, like the risk of discrimination, lack of fairness, lack of transparency and explainability, and deceptive or misleading information, can materialise.
The DCA expects to release an initial vision document for consultation in the second half of this year.

AUTORITEIT
PERSOONSGEGEVENS