AI & Algorithmic Risks Report Netherlands

Report winter 2023/2024

Dutch Data Protection Authority | Department For The Coordination Of Algorithmic Oversight (DCA)

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

AUTORITEIT
PERSOONSGEGEVENS

Contents

Introduction

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

In their core, these AI systems automate actions and decisions which were previously done by humans, or which were not possible before. Simply put: it is about algorithms and AI. This extends from relatively simple applications, in which a single algorithm functions, to complex applications of ***machine learning*** or ***neural networks***. The risk analysis in this report makes no distinction. 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') monitors the possible effects of the use of algorithms and AI on public values and fundamental rights. The DCA will periodically report on this in the AI & Algorithmic Risks Report Netherlands (ARR). In doing so, the AP contributes to more responsible use of algorithms.

The AI & Algorithmic Risks Report Netherlands (ARR) describes risks and trends and developments in these risks. Relevant algorithmic risks are those that may affect individuals, groups and individuals or society and could subsequently disrupt the latter. The DCA has drafted this AI & 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. There are two comments that must be made when describing trends and developments in the risks. First, it can be noted that the use of algorithms and AI not only poses risks but can also make positive contributions to society, including to strengthen public values and fundamental rights. However, given the role of the AP in algorithm monitoring, the emphasis is on mitigating and controlling said risks. Secondly, this periodic report is focused on trends and developments in algorithmic and AI risks, which means that certain elements are emphasized, in addition to the structural risks that are present.

This report does not contain any predictions. Using current knowledge and available information, the AP aims to provide a compact and understandable overview of the current risks and control challenges associated with the use of algrithms. Where possible, the AP 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.

In this report. The first chapter of this report outlines the most recent overarching developments for the Netherlands in the use of algorithms and their risk management. The second chapter describes the developments and challenges of generative AI and ***foundation models***. The third and fourth chapters are thematic and discuss algorithms and AI in the workplace and in education. Finally, the fifth chapter focuses on policy development and institutional frameworks at national and international level.

Algorithmic supervision under construction

The ARR remains a work in progress and can contain errors. The Netherlands aims to be a global leader in working on careful control of algorithms and AI, so that their deployment is beneficial for people and society. The start of coordinating algorithm monitoring at the AP in 2023 and the periodic and systematic analyses in this ARR are an example of this. The DCA started in January 2023 and is still under construction. The first edition of the RAN, which was published in July, discussed the work of the DCA in more depth.

This is the second edition of the RAN, which will be published every six months. The content in this edition is based on the knowledge that is being obtained through the AP's supervisory network. This includes analyses by agencies, its own observations and discussions with more than 100 relevant national and international organizations. However, the developments are moving quickly and the full picture is still incomplete on many fronts. With this in mind, the AP is still trying to provide the best possible complete picture of current risks and developments in measures to control and mitigate risks. And also to link policy recommendations to risks and measures in a constructive manner. However, errors or omissions in this ARR are possible.

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

Key messages

1. The overarching risk profile mandates a call for action.

The cautious estimate is that the risks of algorithms and AI have increased over the past year. An important reason for this increase is that new, generative AI systems are rapidly coming onto the market and are increasingly being used in organizations. Although improved risk management is being worked on, these improvements can not seem to keep up with the pace of innovation. This is partly because legal frameworks and product standards are still in development. Adequate risk management is important to improve trust in this technology and unlock opportunities to innovate. This also applies to the use of new technologies when risks are still unidentified and citizens may encounter potentially negative effects more often.

2. There is a lack of knowledge about where and when risks and negative effects can occur.

Only a few regulators and authorities in the Netherlands are currently receiving signals or complains from citizens about incidents with algorithms and AI within their field of expertise. This may be due to a lack of transparency about the use of algorithms and AI and a lack of reporting obligations for organizations. At the same time, an increasing number of AI incidents are getting attention around the world, partly due to the emergence of generative AI. The Dutch algorithm register for government organizations is important for increasing the transparency of impactful algorithms that are being used. The AP observes a significant increase in the number of registered algorithms in the second half of 2023. It is important that government organizations increase their registrations in this register to promote transparency. And that the proposed mandatory registration of high-risk government applications of algorithms and AI results in a concrete proposal soon.

3. Preparing the Netherlands for the future through a master plan for algorithms and AI.

New applications of algorithms and AI attract a lot of attention. The use of these new applications will take further flight in the coming months and years. A more coherent effort is needed to get a grip on the use and risks of algorithms and AI in order to reap their benefits as a society. This involves more than just the establishment of supervision. There are many actions possible and needed, it is important to have a comprehensive approach regarding the economic, safety, social and fundamental rights perspectives. The AP recommends the development of a Dutch "master plan" for algorithms and AI, which focuses on five mail elements: (1) human control, (2) secure applications and systems, (3) organizations ***in control***, (4) national ecosystem & national infrastructure and (5) international standards and cooperation. This master plan, with similarities to the famous Dutch Delta plan, must be seen in the context of the European market and the open Dutch economy. Chapter 1 discusses this recommendation in more detail.

4. The versatility of generative AI requires appropriate supervision.

The adoption of generative AI in the Netherlands is widespread and fast. When using generative AI, questions arise about the legality of these systems and applications. New systemic risks and user risks are also areas of great concern. It is important to continue working on the concretization of the European supervision of ***foundation models*** (which are the basis for applications of generative AI) and the supervision of organizations that develop these models. The AI Act provides a basis for regulation of these models and organizations. Chapter 2 discusses this in more detail.

5. More and more workers are being managed by algorithms.

From bus drivers to service technicians to health care workers: algorithms and AI play an increasingly important role in the management of labor. This increase is visible both in platform work and in traditional jobs, and this can be beneficial for workers, if for example it increases their physical safety. However, it can also have negative consequences for work pressure and personal autonomy of workers, employment rights of workers and in the assessment of work performance. Currently, European legislation for algorithmic management in platform work is being discussed. The AP sees opportunities and the need for an increase in transparency and the contestability of the role of algorithms in labor management. Chapter 3 discusses this in more detail.

6. Learning instruments and materials for students are increasingly influenced by algorithms and AI.

Even more algorithms and AI are being used in education: from individual adaptive learning systems in primary school to analytical tools in higher education to promote the progress of students through their educational programs. However, there are many reasons why educational profiling

or prediction does not fit well with a student's situation. Careful embedding of AI into the educational approach and an awareness of the limitations of these applications among teachers and school boards is therefore crucial. The AP advises educational institutions to include AI deployment and AI management processes in the design of the ICT strategy, with sufficient support from internal or external expertise. In addition, AI knowledge among teachers also needs to be increased. Chapter 4 discusses this in more detail.

7. The foundation for the oversight of algorithms and AI is almost ready, it is now time for further construction work.

A political agreement at the European level on the AI Act was reached on December 9th. Development on the level of detail will follow in the coming months to allow the AI Act to enter into force by mid-2024, followed by a transitional period before the rules will be applied. This is an important step that will contribute to the responsible use and control of AI. Future experience will have to determine whether regulation will provide an adequate basis for safe use of algorithms and AI. In the Netherlands, supervisory authorities are jointly preparing for the corresponding supervisory task under the AI Act. However, control of algorithms and AI requires more than just product-level supervision.
It is important to also work on internal supervision and external control at an organizational level. An investment is also needed in cooperation between authorities so that citizens can submit complaints or there are signals to detect and address algorithmic abuses. Therefore, based on its task of coordinating algorithm, the AP aspires to facilitate these developments. Chapter 5 discusses this in more detail.

1. Overarching developments

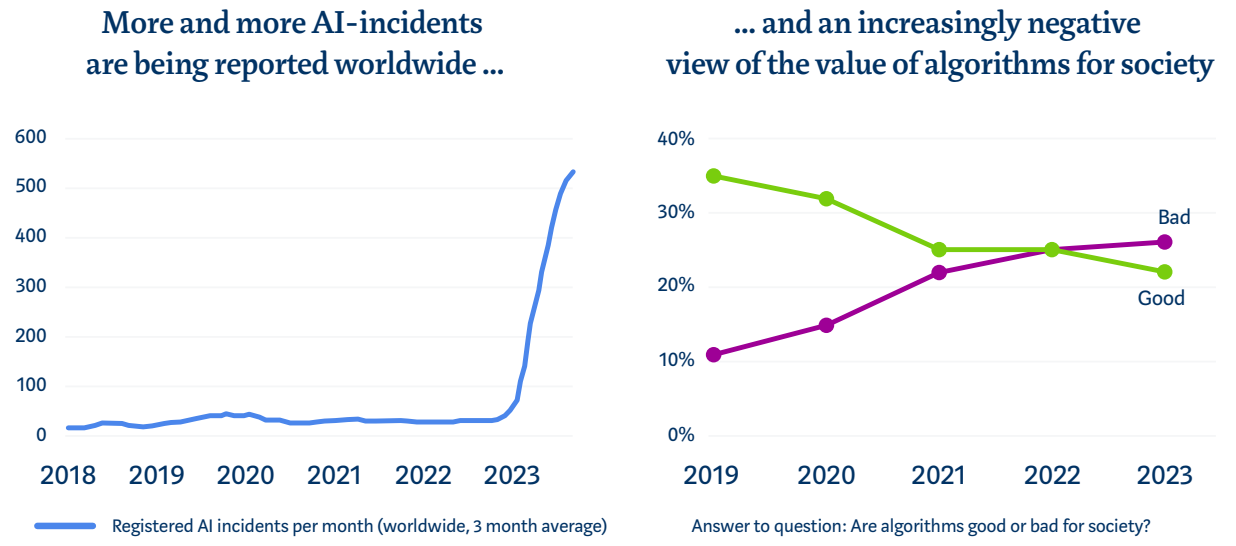
Main points

Risks from algorithms and AI remain present in the Netherlands and are even increasing, but steps are being taken to address these risks structurally. Developments are continuing rapidly. Where language models broke through for the general public a year ago, these models are now more sophisticated and can be applied in even more ways. At the same time, through this attention and broad application, we see that not only known, but also unknown risks, are materializing.

The OECD AI Incidents Monitor shows a significant increase. This OECD monitor is about global incidents described in news articles[1]. Graph 1 shows the strong growth of incidents – a ten-fold increase (961 %) compared to last year. The increasing number of known AI incidents shows the various risks to public values and human rights. So many more incidents have probably come to light in 2023 because generative AI is increasingly being deployed and this has consequently received a lot of attention in 2023. It also shows that many risks are as yet unseen or are poorly predicted and these incidents are probably the tip of the iceberg.

Dutch citizens have become increasingly negative about algorithms in recent years. In 2023, for the first time, more Dutch people think that algorithms for society are bad (26 %) rather than good (22 %) (Graph 1). This structural decrease goes hand in hand with increasing awareness of algorithms, from around 45 % in 2019 to more than 70 % in 2023. A study by KMPG, conducted by Motivaction, shows these results. It fits in with a broader picture in which 22 % of Dutch people have decreased trust in algorithms in 2023.

GRAPH 1: AI INCIDENTS AND PUBLIC IMAGE OF ALGORITHMS



SOURCE: OECD AI INCIDENTS MONITOR (AIM) AND KPMG (2023) - ALGORITME VERTROUWENSMONITOR

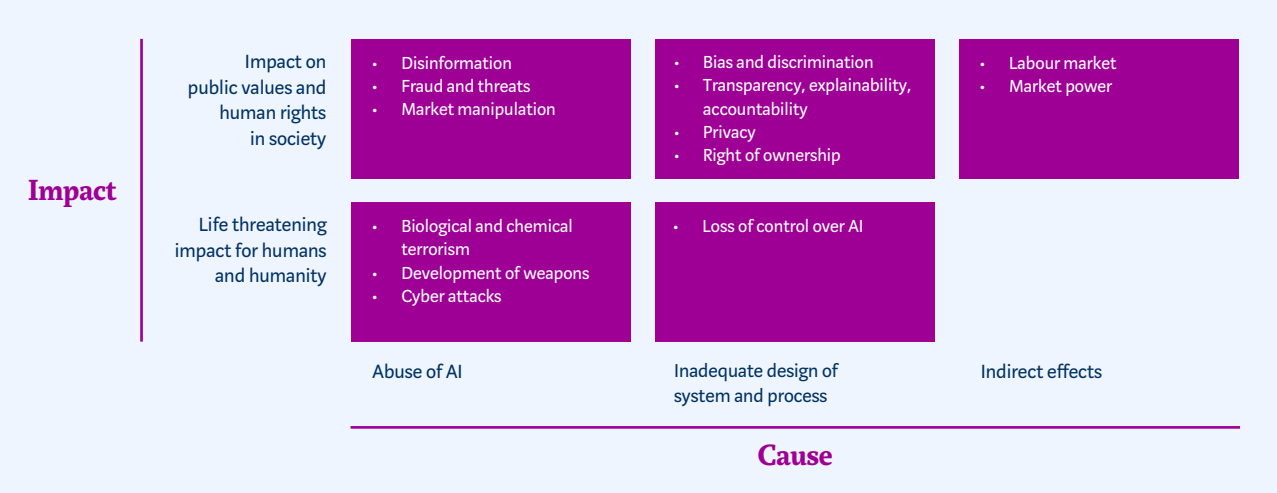
Whereas in 2022, 21 % of Dutch people still had a positive attitude towards the use of algorithms within implementing organizations, this decreased to 10 % in 2023. Additionally, a similar decline can be observed for banks and insurers. In the other hand a positive attitude towards the use of algorithms in retail organizations increases (up to 29 %).

In 2023, the structural and cross-organisational control of algorithms started at national and international level. Due to the rapid development of algorithms and, for example, language models, the awareness is that mastery cannot be left behind and is structurally necessary. Both at the level of individual systems and applications, at organisational level, and at transcending national and international level. Two examples illustrate how this control slowly begins to take shape: the start and growth of the algorithm register in the Netherlands and the entry into force of the European Digital Services Act (DSA). The algorithm register offers public organisations the opportunity to register impactful algorithms and provide the public insight. This provides transparency for citizens and interested experts, journalists and politicians. It also forces public organizations to gain an insight into their own systems, applications and processes. Therefore, this directly contributes to the control of algorithms. The DSA regulates large digital platforms, especially in the field of social media, and forces these providers to address risks and provide greater transparency. From these actions we see a start in making the risks of using AI and algorithms manageable. More registration, transparency and monitoring will probably also lead to increased awareness of incidents, where they have not been revealed to date.

Identifying and managing the risks of algorithms and AI gains wider international attention, helping to achieve common understanding, language and tools. Within the G7, since May 2023, the ***Hiroshima AI Process*** has been working on international principles and a voluntary code of conduct for developers that complement existing regional initiatives such as the European AI promotion. In October 2023, the United Nations created an ***AI Advisory Body*** to advise on an international and inclusive governance of algorithms and AI. These initiatives come in addition to existing global cooperation within UNESCO and the OECD.

An *AI Safety Summit* in the UK focused on the risks of groundbreaking AI systems and will be followed up in the coming year. The idea behind the AI Safety Summit is that the risks of groundbreaking AI systems can have an impact on public values and fundamental rights (transparency, explanation, honesty, non-discrimination, etc.) as well as threatening the intrinsic safety of people through unforeseen consequences. In the final statement, 29 countries, including the Netherlands, point to risks in the field of cybersecurity, biotechnology and disinformation that can be catastrophic in the end. It is noted that these risks may arise both from an accident and from deliberate misuse of these groundbreaking AI technologies. The conversation that took place during the safety summit emphasizes that one type of risk does not exclude the existence of another type of risk. Given that the negative risks to public values and fundamental rights are already materializing, these are also acute. In order to talk about the different types of risks of algorithms and AI, it helps to have a risk matrix. The DCA provides a first step to this, whereby risks can be distinguished by two axes: cause (abuse of AI, direct unmanaged and uncontrolled effects, and indirect effects) and impact (on fundamental rights and values of people and society, life-threatening impact on people and humanity), see Graph 2.

GRAPH 2: ALGORITHMS & AI- CONCEPTUAL RISKS MATRIX



In the United States, the White House has taken concrete action that will affect the use of algorithms and AI within the federal government in the coming years. The White House pays extensive attention to the control of algorithms and AI in a recently published executive order[2]. President Biden talks about AI as one of the most important technologies of our time: "*The President has been clear that we must seize the opportunities AI presents while managing its risks.*"

Algorithm cases

Recent case studies show that possible undesirable effects of algorithms are not always easy to assess. The Netherlands Institute for Human Rights [College voor de Rechten van de Mens] has published information in recent months on two cases in which an algorithm is central and which have been brought to the attention of the Institute. In the two cases, anti-cheat software at the VU University Amsterdam and the dating app[4] Breeze, the Netherlands Institute for Human Rights delivered a judgment. A consequence of both judgments is that they provide relevant insights into the control of algorithms. In the Breeze case, it is important to note that the company itself has proactively asked for a judgement from the Institute on how it intends to correct its algorithm and remove possible discrimination. This proactive step is welcomed by the AP because it can minimise negative impact at an early stage. According to the Institute, Breeze must adapt the algorithm used to prevent discrimination, by introducing people with dark skin tone and of non-Dutch origin just as often to other users as people of light skin colour and Dutch descent. In the anti-cheat-software case[5], a student felt discriminated by the system that was used for taking online exams at VU Amsterdam. She experienced problems because, from her point of view, the software did not recognise her face because of her dark skin tone. The Institute's final judgment was that there was no demonstrable discrimination in the particular case of the applicant. The university showed that the student had not experienced anymore problems during her exams than the other students and that the problems were not caused by her skin colour. The Institute however stresses that the use of such systems and/or applications in other cases can lead to discrimination. It also shows why it is important for those involved to know that when they come into contact with AI or algorithms. So that they can contest any outcome or ask for accountability for the deployment of such systems.

Problems are also detected abroad and the impact of unmanaged algorithm risks is felt. Two recent examples stand out: the "Robodebt scandal" in Australia and the anti-fraud approach by the Department for Work and Pensions in the UK. Both cases show that the risks and effects that have occurred explicitly in the Netherlands also occur elsewhere.

The Australian *Robodebt* scandal shows the consequences of automated decision-making without human supervision and without a safety net for model errors. The Robodebt scandal shows similarities to the Dutch surcharges affair in terms of impact generated. In this case, uncontrolled algorithmic decision-making has had a major impact on people's lives. In the Robodebt scandal, income figures from annual tax returns of people with benefits were compared with the biweekly declaration of income with the benefit agency and this was fully automated. If the figures from the two data streams did not match, an algorithm calculated that someone had received too much in benefits and that they would have to repay. Then entirely automatically, without the authorisation of an official, a letter containing the amount due was sent to the 'debtor'. However, the model did not take into account fluctuations in income, such as for seasonal workers. These groups with a fluctuating income were thus wrongly imposed huge recovery fines. These recovery fines have led to major personal problems for victims with a very large impact on this group and their families.

The focus on the use of an anti-fraud algorithm by the Department for Work and Pensions (DWP) in the United Kingdom shows the importance of adequate control to avoid ambiguity about fairness. The DWP has been strongly committed to the use of algorithms to combat benefit fraud. During the same period, the benefits of a large number of Bulgarian women living in the UK were discontinued. This does not necessarily mean that there is discrimination but the problem is that the DWP itself does not know. The DWP indicated that it was limited to the testing for unfair outcomes of the algorithms used and that the results of its own fairness tests are not clear. This indicates a lack of proper control measures and therefore irresponsible algorithm use.

In 2023, visible auditing and evaluation of the use of algorithms took place in large municipalities. Not only by municipalities themselves but also by controlling organisations that are close to municipalities. For example, the Amsterdam Court of Audit [Rekenkamer Metropoolregio Amsterdam] recently published the study 'Algorithms'[6]. The Court concludes that, in the municipality of Amsterdam, the management framework and practice still pay little attention to three points: (1) the fairness of algorithms, (2) the privacy protection of citizens and (3) the openness about the development and application of algorithms. These concerns are probably not only relevant for the municipality of

Amsterdam but also for other municipalities and similar public and private organisations. In line with this, Algorithm Audit has also provided advice in the report 'Risk profiling re-examine social assistance benefits'[7]. This advisory report states that algorithmic risk profiling can only be used under strict conditions for the process of selecting citizens with social assistance benefits for re-examination. The report indicates that a combined use of different selection methods is desirable to break tunnel vision and feedback loops. Additionally, one condition is that support algorithms for the selection of citizens must be explained for re-examination. On the other hand, complex training methods cannot comply with this.

It is positive that the search is being carried out for connecting points, building blocks and tools to control the use of algorithms in municipal processes. At present, many audits and evaluations will show that control as a whole is still insufficient. These outcomes should be used as basis in order to achieve concrete and specific improvements in control elements. It is important to prioritise those areas that have the most impact on citizens.

In 2023, more than 40 % of Dutch regulators investigated the impact and use of algorithms and AI. A survey was carried out by the AP among 24 Dutch regulators, to gain more insight into the role of algorithms and AI in the areas where supervision is carried out. Interestingly, the investigation by supervisors reveals algorithm risks. For example, last June, the Dutch Authority for Consumers & Markets addressed[8] web shops that use misleading ***countdowns***. On some websites, in addition to online offers, you can sometimes see a ***countdown timer*** and if the time is up, the offer would no longer be valid. Unfortunately, the timer puts pressure on consumers to make purchases faster. However, research showed that the offers on certain web shops continued to exist after the end of this countdown.

Due to a lack of complaints and reports, few cases come to the attention of a supervisor. The National Institute for Human Rights [College voor de Rechten van de Mens] is one of the regulators that has received reports of possible discrimination in which an algorithm is involved. A recent opinion on the use of anti-cheat software can be found in this report in the section 'Casuistic algorithms'. This case clearly shows that the demonstration of discrimination in the use of algorithms remains a complex matter. Supervisors need specific knowledge of both the sector, the impact and the technology used. Chapter 5 deals with the results of the survey.

Risk management in generative AI

Generative AI puts existing frameworks to the test. Generative AI systems or models produce or manipulate material such as images, audio or text. These technologies have been one of the most eye-catching developments in algorithms and AI in 2023.

Organisations that want to develop or deploy generative AI can partly rely on existing regulations, management measures and standards. However, concrete or additional measures, ***frameworks*** – or plugins – are needed to use these new technologies responsibly. Frameworks that are tested at multiple levels are, for example, organisational and legal frameworks, but also thinking frameworks and technical frameworks or standards. A first step is to comply with current frameworks such as the GDPR.

Organisations must already work on sufficient knowledge and a mature organisation. While regulations, management measures and standards are still in development, organisations can already take steps to achieve an adequate form of risk management for the deployment of generative AI.

- **Algorithmic literacy within organizations must be at a sufficient level.** This means that everyone involved in the development, deployment and use of algorithms must have sufficient knowledge of the technology, deployment and risks, including the final users and the people who make decisions in an organisation. In managerial positions, there must be knowledge about how an AI model is created, in order to determine whether or not it is appropriate to use it for a particular purpose. This is all the more important because ***foundation models*** are designed for versatility.

- **Also, an organisation must be sufficiently mature and have the right people, knowledge and 'checks and balances' in-house.** This also means that the advice of strategic advisors and opinions of citizens or consumers should find resonance in an organisation.

Control of generative AI applications is still in its infancy. Existing control tools do not fit well with the risks of generative AI. It is partly due to this reason, that the AI Act has been extended to include a specific supervisory regime for foundation models (generally purpose AI models), which will need to be more concrete in the coming period. In the development of these instruments, the AP sees room for at least four points to be taken into account. Until these control measures are sufficiently developed and embedded, caution is needed in the development and deployment of generative AI.

- **Customisation is needed in the form of an impact assessment for generative AI that matches existing and future impact assessments (DPIA, IAMA, FRIA).** Allows organisations to identify, mitigate and account for the potential risks of the application and concrete deployment of generative AI in advance. There is no impact assessment that specifically takes into account certain characteristic use and concentration risks that apply to applications and concrete deployment of generative AI.

- **Auditing standards for algorithms need to be further complemented and conceived for the auditing of applications and concrete deployment of generative AI.** When these systems and applications are deployed then risks, effects and measures should be monitored periodically. For example, through audits and this requires auditing standards. Currently, these standards are in development for algorithms. These audit standards can, in principle, also be applied to the concrete use of generative AI but must be focused on the specific characteristics.

- **Transparency of generative AI systems and applications is an important issue that needs to be made more concrete.** A lot of research is being done on technical solutions and guidelines for transparency. However, the current versions of well-known systems and applications are often limited in transparency about the origin and use of data, impact, feedback loops and other elements that are important for responsible deployment and control.

- **Developers must carry out a risk analysis to identify and mitigate the reasonably foreseeable effects of the models they have developed in different applications.** When generative AI is applied in specific sectors, products or applications, a number of risks are foreseeable for the developer. Developers can reasonably be expected to be familiar with this and take mitigating measures. Furthermore, developers need to be transparent to organisations that will use their models. In this way, organisations can adjust their procurement, control and measures accordingly. A 'manual' to the generative model can help an end user to determine whether the model is suitable and whether there are any dangers in the deployment of the model for different purposes.

Inspiration for mitigating algorithm risks can also come from other areas. A proven method for mitigating algorithm risks in the financial sector is the use of circuit breakers. These automatically shut down processes – in this case the trading on financial exchanges – if the risks exceed predetermined limits. This approach, an emergency brake after

exceeding a threshold, can also be used in other areas to mitigate known risks with not always known effects. The DSA supervision of very large online platforms also provides inspiration for the approach and supervision of very large organisations that develop and/or deploy algorithmic systems or applications. Chapter 5 provides more information about the DSA.

Transparency as a basis

The AP is positive about the growing number of registrations in the national algorithm register. Since the first edition of the ARR was released in July 2023, dozens of registrations have been added to the algorithm register of algorithms used in the public sector. In total, around 190 algorithms were registered in November 2023 (see Graph 3). However, this is only a small part of actual high-risk algorithms. In order to ensure that all government organisations are transparent about the use of impactful algorithms, the AP calls for quickly clarification of the envisaged obligation to register such high-risk algorithms in the public sector. As the upcoming AI Act entails registration obligations for AI system providers, it is obvious that a proposal for complete or partial registration obligation in the national algorithm register follows quickly. The DCA also recommends prioritising algorithms with evidently high risks for publication in the Algorithm Register. Registered algorithms are still often medium-risk or low-risk algorithms. The designation of high-risk systems under the AI Act could be instrumental for that prioritisation.

The DSA requires very large platforms and search engines to be transparent about their algorithmic risks. Providers such as Instagram, TikTok and Google Search must map, address and publish systemic risks of their services under the DSA. The DSA regulates digital services and the corresponding rules will apply to regulated organisations in the EU by mid-February 2024. These include risks such as spreading disinformation and influencing elections, but also negative effects on fundamental rights, such as the right to privacy, freedom of expression and non-discrimination. This includes the risks of using, for example, recommendation algorithms or algorithms to block online content. Digital service providers should also regularly disclose how they moderate illegal content or disinformation and explain the algorithms they use for that purpose. In fact, the DSA forces large platforms and search engines to control the risks of the algorithms they use and new supervision is being organised.

Algorithm distortion

The use of algorithms in existing processes can lead to a radical change in character. Algorithms are never neutral, i.e. they can change the context of their efforts. We call this algorithm distortion and it can occur in different ways. Inadvertently, or in the form of a conscious adaptation to the environment in order to allow an algorithm to function.

Intended algorithm distortion is often due to algorithms needing data to function. An illustration of this could be bus services where travellers have to sign up for a ride. Travelers indicate by phone or via an app what time they want to get to which stop and where they want to go. These notifications provide data for an algorithm that determines the most efficient route for the bus at short intervals: stops without registration are skipped because it saves fuel and

GRAPH 3: INCREASING USE OF DUTCH NATIONAL ALGORITHM REGISTER

**) from march 31st until October 31st 2023 **) Reference data: November 6th, 2023. Number of organizations with one of more algorithm registered per category (for example: 2 provinces have jointly registered 9 algorithms).*

time. However, passengers cannot simply stand at the stop and take the next bus. Because they have to register, such a bus service shifts somewhat from being public transport to a taxi service. This is an algorithm distortion.

Unintended algorithm distortion occurs when algorithms accelerate one process, while related processes are not accelerate. For example, scan cars with image recognition can check the streets of a city faster for parking violations than a parking officer. However, if notification of offenders does not accelerate, a violation – before anyone gets the chance to rectify that violation – has been detected and punished more often than in a similar process without algorithmic use. As a result, the relationship between helpfulness and punishment in enforcement practice becomes very different. This is also an example of algorithm distortion.

A good awareness of the function of processes within an organisation and their context gives insight into algorithm distortion, prior to the use of algorithms. Bus services that see 'predictability' and 'accessibility' of public transport as important functions may opt for an old-fashioned timetable instead of a route algorithm that skips empty stops. Parking organisations that realise that checking and collecting fines are connected to offenders know that more intensive checks require and regulate adjustments in those adjacent processes.

Control measures in practice

Some control measures are easier said than done: two major operationalisation challenges concern the completion of bias and fairness testing and the completion of meaningful human intervention. Both topics receive a lot of attention and are central to many frameworks and discussions about management measures. However, it is difficult to make these subjects concrete. However, this is required of organisations when applying open standards.

Preventing bias and unwanted discrimination is essential for the deployment of reliable and sustainable application of algorithms and AI. In its publication on non-discrimination in algorithms the Rathenau Institute notes that – in the context of algorithm testing – there is uncertainty about the precise meaning of discrimination (operationalisation of fairness metrics) and the use of algorithmic systems by definition poses the risk of indirect discrimination. A first step therefore, is to open a discussion on what level of risk (political or organisational) is considered acceptable when using learning algorithms. A second step is an answer to the question of which fairness metric – for example, the degree of deviation in the percentage of ***false positives*** among different groups – is suitable for which types of algorithmic processes. In line with this, the governmental Audit Office (Auditdienst Rijk) recently revised its ***research framework Algorithms***[9]. In this context, seven management measures are provided for the bias and discrimination sub-area. These measures are formulated in such a way that they are or can become verifiable.

Meaningful human intervention requires a lot from an organisation. Humans cannot become "stamping machines". When parts of processes are automated through the use of algorithms, this often affects the entire process. Due to the decoupling of components, the interaction between man and machine is of great importance. This requires careful organisation of processes, including human intervention for impactful decisions. This is even mandatory under certain circumstances under the General Data Protection Regulation (GDPR), such as automated decision-making. How this human intervention can be made meaningful and concrete often seems to be a journey of discovery at a given moment. At the core, meaningful human intervention will at least mean that humans can make an independent judgment (based on expert judgement) and have the expertise to control the machine, but can also do the process manually without the intervention of an algorithm. As a result, people remain in control of the entire process and retain knowledge of the decision, but also about possible errors or undesirable effects. This provides a guarantee for the popular term "the human measure". Organisations must organise their processes and workforce in such a way that this meaningful human intervention is possible in the process. And that there are employees with sufficient knowledge and powers to be able to do this. Workers who have to check automated decisions under pressure with insufficient expertise run the risk of becoming a 'human stamping machine', thus eliminating meaningful human intervention.

The first algorithm officers are appointed

The AP has called on organisations to improve their grip on algorithms on their own initiative. A key message in the first ARR (summer 2023) was that in many cases the control of algorithms is not yet at a sufficient level, although this is becoming increasingly important. A related observation was that organisations insufficiently evaluate the existing use of algorithms and also reflect too little on this.

Some organisations are now working on appointing an internal regulator of algorithms or an official algorithms. The AP welcomes the fact that organisations are taking responsibility in this way.

Different organisations shape such a function in different ways. There is no legal obligation to describe the tasks of such a function. Some organisations create a new position with a separate appointment while other organisations extend the tasks of the Data Protection Officer (DPO). The protected role that the FG has within an organisation can also serve as an inspiration for how this new function can take shape. The officer will then have to have additional knowledge and skills in order to be able to fill the role properly.

How these new functions take place must be shown in practice. In any case, there are a number of ways in which an official can give the algorithms or internal algorithm supervisor merit. First, by collecting information within the organisation, creating an overview of all algorithm usage that takes place and keeping track of that overview. Secondly, by setting up a group in which all organizational parts that use algorithms are represented and all relevant expertise is brought together. So that people within the organisation can learn from each other and there is a shared vision of the algorithm use within the organisation. Thirdly, by translating such a vision into a clear policy, with clear rules for the entire organisation. However, it is important that the staff member can rely on a firm and clear mandate of the summit of the organisation and to anchor the position in terms of responsibilities, powers and resources that are available.

The DCA will continue to work on the operationalisation of this supervisory and control issue. The DCA wants to further explore the origin and possible role of the algorithms officer – or whatever name is given to it. That is why the DCA will organise a meeting in 2024 for people who are algorithm officers or perform a similar function.

GRAPH 4

Masterplan algorithms & AI: Ambition for 2030

Human control

Goal People have sufficient knowledge to safely use algorithms and AI, and are sufficiently protected against the risks of algorithms and AI.

Indicators
- ↑ Knowledge within Dutch society
- ↑ Trust in algorithms and AI

Actions
- Education of young, working-age and elderly people
- Accessibility of complaints offices and clear follow-up
- Algorithmic transparency

Secure applications and systems

Goal All impactful applications and systems that use algorithms and AI are safe, also with regard to fundamental rights and public values.

Indicators
- ↑ Number of registered applications and systems
- ↓ Number of reported incidents

Actions
- Private and/or public-private support initiatives that contribute and stimulate
- Investments in sufficient supervisory capacity

Organisations in control

Goal Organisations are fully – and at all stages – in control of the use of algorithms and AI in their processes and the consequences of their application.

Indicators
- ↑ Number of impact assessments and evaluations performed
- ↑ Period audit reports show organisational improvement

Actions
- Clear action plans and guidance for all economic sectors
- Supervisory framework on organisational risk management

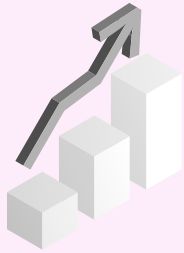
National ecosystem and infrastructure

Goal Algorithms and AI contribute in a safe way to Dutch welfare, wellness and stability.

Indicators
- ↑ Dutch performance on an international "broad index" of algorithm/AI maturity and leadership

Actions
- Innovation agenda that contributes to these goals along the full spectrum
- Build-up of knowledge through scientific research and AI centres of expertise

Monitoring and control



Substantial yearly improvements to reach ambition in 2030.

International standards and cooperation

Goal Addressing global interconnectedness in AI-systems through global standard setting and supervision; alignment of EU AI's Act with European Charter.

Indicators
- ↑ Global stability of AI systems
- ↑ Effective European cooperation on AI topics
- ↑ Dutch influence on AI regulation

Actions
- Focus on safeguarding public values and fundamental rights in AI regulation
- Contribute to international knowledge sharing and cooperation

Master plan for the control and use of algorithms and AI

Inspired by previous major societal challenges, a strategic master plan, in light of the Dutch delta plan, has to be considered for controlling and using algorithms and AI in the Netherlands. More information on the historical meaning of the Delta plan can be found in the additional box in this chapter. The Scientific Council for Government Policy (WRR) has previously pointed out that embedding a system technology such as AI requires a long-term interaction of society and technology to guide the implementation of such a technology. In particular, governance and infrastructure are important components of such a plan. With initiatives like the 2022 Work Agenda Value-driven Digitalization [Werkagenda Waardengedreven Digitaliseren] and the 2019 Strategic Action Plan for Artificial Intelligence [Strategisch Actieplan voor Artificiële Intelligentie], the Dutch government has already taken some first steps. The Dutch government should take the next step through enriching and bringing these initiatives together and combine them with the latest developments and regulations, such as generative AI and the upcoming AI Act. The master plan that will be created from these efforts should be seen as a concretization of specific steps and a vision for Dutch society, while facing the challenges and changes that algorithms and AI are bringing to the Netherlands.

A master plan with targets for 2030 can offer a vision and a horizon. As a coordinating supervisor of algorithms and AI, the AP has some initial ideas. Such a master plan is explicitly about more than all the organization of both internal and external supervision. The master plan must also be seen in the context of the European market and the open Dutch economy. The AP is happy to give all policy makers, politicians, social partners, supervisors and civil society the space to further this idea. Graph 4 provides inspiration for such a master plan based on the AP's first observations as the coordinating algorithms and AI supervisor and the policy messages that are brought together in the RAN.

The master plan for control and use of algorithms and AI can serve as a compass for a wide range of stakeholders and provide clarity, realism and direction. In a process of transition, not everything can happen at once. By drawing up a structured, tiered and targeted plan, divided into five core pillars, it can be ensured that citizens, businesses and the government work together towards a society in which algorithms and AI are responsibly embedded and beneficial. So that not only prosperity, well-being and stability are enhanced, but also fundamental rights and public values are well protected.

The central objective could be to ensure human control in an era where dependence on algorithms and AI can increase. It is therefore important to raise awareness about risks and strengthen education about algorithms and AI. In this way, users retain control and understanding about how the outcome of algorithms and AI affect decisions and processes. This does not only support safe and trustworthy AI systems, but also promotes the development of safe and reliable applications.

For organisations, it can provide a framework that assists them in being "in control" when using algorithms and AI. The master plan aims to build capacity and provide clear guidelines so that organisations of all sizes know what is expected of them. In this area, inspiration can be drawn from recent initiatives within the U.S. federal government, where government organisations are required to set up a governance structure for AI, in order to develop an AI strategy for each organisation and to devise a transition plan for how the organisation will reach a higher level of control in the coming years. In addition progress will be actively monitored in the coming years. In order to support organisational transitions, it is important that there is also a national ecosystem and infrastructure for increasing AI knowledge. A recent initiative from the Dutch government to build its own open language model (GPT-NL) is an interesting example of being in control.

International cooperation and the development of standards can also contribute to the master plan. By actively contributing to international discussions and cooperation, the Netherlands can be of influence to European and global AI regulations. This will strengthen the position of the Netherlands as an innovative and responsible society in the field of algorithms and AI.

Structural investments in algorithmic literacy are needed in order to be able to deal with algorithms as a society. Dutch society is one of the most digitized societies in the world. Over the past twenty years, strong investments have been made in digital literacy to cope with digitalization as a society and to build a strong digital knowledge-based economy. This is once again necessary for the safe development and deployment of algorithms in society as almost everyone comes into contact with algorithms now or in the near future. To manage this, knowledge of algorithms and their risks and effects is essential. This does not mean that everyone should have the same knowledge. For instance, teachers or doctors need to know how to assess or use an algorithm. Workers need to know what the use of an algorithm means

for them and how to defend themselves against unwanted effects. Directors and boards of organisations must have adequate knowledge to be able to oversee and assess the risks, impacts and opportunities for control before taking a decision on the deployment. Algorithmic literacy is an important basis for dealing with the frameworks and regulations that are currently being worked on as a society.

Delta plan - historical meaning

The Netherlands is known for its ingenious physical infrastructure that protects the parts of the Nederland's that sit/reside below sea level. In 1953 the southwest of the Netherlands was hit by disastrous floods. After these floods, the Delta plan was introduced by the Dutch government to comprehensively protect the country from flooding for years to come. Massive physical infrastructure projects started and various innovative flood defences, locks and dams have, since then, been built.

The current situation, in which the Netherlands needs to be protected from oncoming risks of AI, warrants a novel, new type of Delta plan. This time not a physical infrastructure to protect the citizens from the water, but a digital infrastructure that protects citizens and society from the risks that come with the deployment of algorithms and AI.

QUICKLY TO THIS PART

Risks of use of generative AI →
Systemic risks of foundation models →
Disinformation →
Supervision of foundation models and generative AI →

2. Generative AI & foundation models

The adoption of generative AI in the Netherlands is happening widespread and fast. With generative AI, it is possible to create textual or audiovisual output that closely matches a specific request or prompt. The rise of generative AI is strongly linked to developments in the field of foundation models, which form the basis for these types of techniques. The application possibilities for generative AI are very diverse and the technology can be integrated into existing software products. This adds an additional layer of complexity for regulation and supervision. In addition, the emergence of generative AI introduces new user risks and systemic risks. The shape the regulation and supervision of foundation models and generative AI is currently under discussion globally, with a focus on the largest models. The AI Act provides the world's first AI rulebook with European rules and European supervision. It is important and urgent that steps are also taken in monitoring the use of generative AI by organizations. This includes responsible human-machine interaction and how organizations are transparent about their use of generative AI.

About half of all Dutch people are familiar with generative AI but how they perceive this technique differs.[10] The number of Dutch people who actually use these types of applications is lower. The perception and image of generative AI in the Netherlands is mixed. For example, it seems that Dutch consumers are more skeptical about the use of these techniques than consumers in other countries.[11] Conversely, other research suggests that two-thirds of Dutch consumers find it useful to obtain medical advice through generative AI, and 7 out of 10 Dutch consumers trust the text and output that has been created with generative AI.[12] Publications from consulting firms indicate that 80 % of IT leaders within Dutch organizations expect generative AI in an organization to play a role in supporting efficiency and scalability, but at the same time 65 % is concerned about ethical considerations in the use these types of systems.[13]

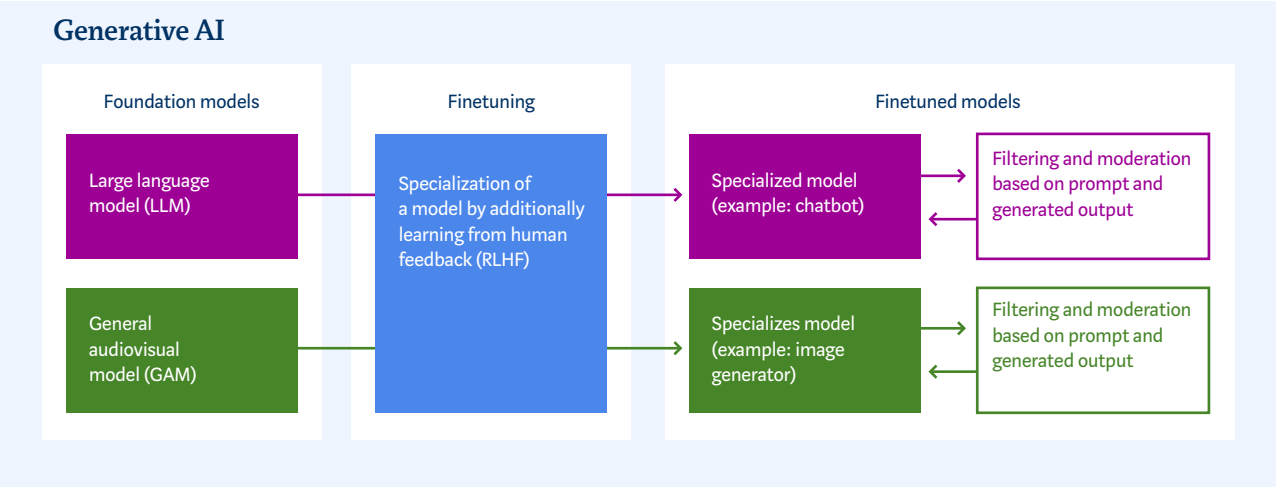
Increasingly powerful foundation models and applications lead to the increase in the use of generative AI. In a simplified explanation, a foundation model is a system that is trained with ***deep learning*** on large amounts of text and/or audiovisual material. Examples of foundation models are

language models – also known as ***large language models*** (LLMs) – such as Claude 2 (developed by Antrophic), GPT4 (OpenAI), Llama-2 (Meta) and Palm 2 (Google). The usability of foundation models has increased sharply in recent years due to three developments that reinforce each other: scientific breakthroughs, even more data to train these models and greater computing power. A significant example of one of these scientific breakthroughs is in the architecture of transformer models. This makes it possible to process large amounts of text more efficiently, leading to significant advances in language understanding and language creation.[14] In the context of the AI Act, foundation models are also referred to as ***general purpose AI models*** (GPAI).

Foundation models need to be adapted for usable AI applications. For example, to pursue certain behaviors, developers use finetuning that is emphatically based on human feedback. Model trainers then compare the suitability of different outputs – for example, two generated answers to a question – depending on the intended purpose of the specific application for which training is being conducted. This feedback makes the model output match the desired outcome. An example is a chatbot that is programmed for customer service use. The human feedback training process is used to ensure that the chatbot communicates with a certain tone, politeness and level of detail. Another example is a specialized language model for writing public friendly weather forecasts based on the model results of a technical weather system and this requires additional training.
For applications with a great potential for user interaction, such as chatbots, filtering and moderation has to especially take place. For specific applications, it may also be necessary to train the foundation model with additional data, for example articles from scientific medical journals. Graph 5 provides a simplified representation of the way in which generative AI works, noting that developments are moving rapidly.

With the emergence and increasing use of generative AI applications, there is a greater need to enforce existing legal rules and develop new policies for emerging risks. An OECD study in G7 jurisdictions shows that all countries are concerned about the possibility of generative AI being widely used for disinformation and manipulation (see Graph 6). The use of generative AI by malicious actors is a risk. However, the legal tension between generative AI applications and, for example, privacy rights and intellectual property rights is also seen as a concern by G7 member states. These are risks that are often directly linked to the data on which foundation models are trained (e.g. through ***scraping***) and the output that is generated. For instance, a generated song with the voice of a pop artist. At present, the regulatory framework (or lack thereof) for generative AI and foundation models does not sufficiently address these potential risks.

GRAPH 5: AI-CHATBOTS THAT ARE BEING USED FOR GENERATIVE PURPOSES ARE A SPECIALIZATION OF LARGE LANGUAGE MODELS



Explanation: *This figure offers a highly simplified representation of how generative AI works. The basis for generative AI lies in a foundation model. This is a very large model that has been trained using algorithms to recognize connections and patterns in a huge amount of text and/or audiovisual data. Practical use of these foundation models requires additional specialization, which can be provided by training the model via human feedback on elements that are, or are not, desirable for the purpose in which the model must specialize (for example: answering chat questions in a friendly form in a first person perspective). This specialized model is offered to users after the training and feedback phase. Depending on its purpose, the model is instructed by its creator to impose restrictions on itself. For example the refusal to answering medical questions or supressing an answer that contains swear words.*

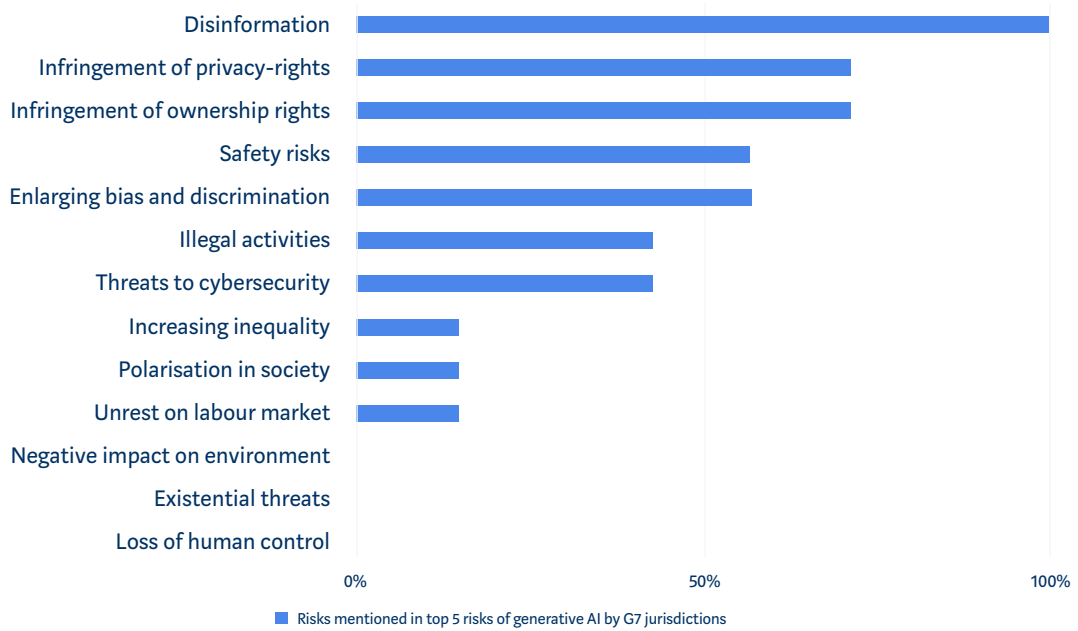
At the end of October 2023, the G7 indicated that jurisdictions should continue to work on specific regulations for AI. In the meantime, AI-developers are encouraged to adhere to general principles, for example for testing and monitoring abuse, reporting incidents and providing information on their models.

This chapter does not address the legality and security risks in the use of individual generative AI models and AI applications. Legality and security risks can be prohibitive in practice. The AP has previously expressed concerns about the risks to the protection of personal data. These and other risks have prompted the AP to call on organizations to critically assess how they develop and use generative AI applications.[15]

Risks of use of generative AI

Generative AI systems bring new usage risks, stemming from the characteristics of the current foundation models. These risks are for example related to the generative AI system's sensitivity to instructions from the user (***prompting***) and the degree of randomness of the output, which can sometimes be misleading. Often, these systems do not give the user any insight into the margins of uncertainty, or the alternative outputs and the sources. Limited transparency makes it more difficult to assess and value the generated output. Another important risk is the existence of bias in training data which can lead to distorted outcomes. When individuals and organizations are insufficiently aware of these shortcomings, or do not take into account how they are being handled in the output, this increases the risk of erroneous conclusions and actions. Furthermore, this may lead to discrimination and arbitrariness in the actions of a person or organization. These risks can be avoided – at least partly – by a conscious and informed approach to generative AI systems. A user should at least be aware of the specific limitations of a specific generative AI system (and also develop experience with prompting) and how this affects the output of a system.

GRAPH 6: G7 SEES ABUSE OF GENERATIVE AI FOR DISINFORMATION AND MANIPULATION AS BIGGEST RISK



SOURCE: OECD (2023), "TOWARDS A G7 COMMON UNDERSTANDING ON GENERATIVE AI"

Users are confronted with faulty output from generative AI, this is a risk for users and can occur in different forms. These faulty outcomes are sometimes referred to as "hallucinations", but this is an undesirable definition because it attributes a mystical and human trait to errors inherent in the stochastic character of language models, where outcomes are determined by plausibility and arbitrariness. Faulty, erroneous or false output may be perceived as plausible, but the language model itself cannot determine the fact that it predicts the random probability and plausibility of the output. The current use of foundation models is still relatively new. As a result, many individuals and organizations that use generative AI do not know where these errors come from and in what forms they can occur. A recent scientific research paper distinguishes at least two types of errors in generative AI from a user perspective: factual illusion and hopeful imaginations (***silver linings***).

- **A factual illusion occurs from the perspective of the user when a language model makes a misleading mistake in response to a factual correct prompt.** This occurs, for example, when a language model responds to the prompt 'the first football match in the

Netherlands' that 'the first organized football match in the Netherlands took place on 14 December 1889 between the teams of [...] RAP (Reaction and Relaxation after Labour) and VVHV'. This is incorrect, because the first football match already took place in 1865 and between other teams. The crux here is that the user must be able to recognize this error.

- **A hopeful imaginative answer occurs when a language model makes a misleading mistake in response to a factually incorrect prompt.** For example, the answer "Stronger Netherlands and early goal" in response to the question "What explains the loss of West Germany in the 1974 World Cup final against the Netherlands?". Here too, the crux is that the user must be able to recognize this error.

The faulty or erroneous output of generative AI can be further divided into categories such as numerical errors, misinterpretation of abbreviations, assigning fictitious quotes to individuals and organizations, and mentioning non-existent characters or entities. For the erroneous but plausible output generated by these models, the likelihood of incorrect but plausible output increases when a prompt is very different from the training data, when a prompt provides room for interpretation and how directing a prompt is.

Another risk in the use of these models is the role of randomness and arbitrariness in the output. This is inherent to the functioning of the current foundation models. These models, which do not have a fixed knowledge base such as specific databases, generate output based on probabilities. And in their applications, the output of these models can vary with their degree of plausibility as this is necessary for flexible and realistic application. It is sometimes difficult for users to identify the degree of arbitrariness in the output. An example is that if a language model is asked ten times about the most famous Dutch footballer, the answer is 'Johan Cruijff' nine times and once 'Johan Cruyff'. Graph 7 gives another example. The degree of variability in output is, in principle, a model parameter that can be adjusted. However, due to user-friendliness, this variation is not available in some popular tools for generative AI. The variability in the output of generative AI emphasizes the need for users to be aware of this and to interpret the results with a degree of carefulness and skepticism. In situations where precision is essential, the use of generative AI is not appropriate.

As with all applications of algorithms and AI, users must also take bias into account that can develop in the training phase of a model. The fact that AI systems can be biased is now widely known. An important cause of bias is training data that are insufficiently representative of today's society or as is desired for its application. In relation to generative AI, this emphatically highlights image models, which for example are very one-sided i.e. doctors disproportionately visualized as men and nurses disproportionately often as women. In addition, AI models for image creation often present different variants of output to a user, making the bias more prominent and visible. In language models, of course, the same bias is present, but it is less visible. For example, because it has been suppressed via finetuning.

GRAPH 7: GENERATIVE AI AND APPARENT ARBITRARINESS

Each output of a generative AI-model is a new combination of apparent randomness and probability based on patterns identified in the data during previous training.

Explanation: *anwered generated by ChatGPT in Dutch on November 9th 2023. The answers are generated in a new session based on the same prompt. The language model is sensitive for exact word choices in the prompt. When the words "what" was changed in "mention", different restaurants were named in the answers.*

The use of generative AI risks emphasize the need for careful human-machine interaction... In traditional algorithmic processes, AI functions autonomously, with possible human intervention to assess the outcomes. Generative AI, on the other hand, often requires active input from users and this creates a different dynamic between humans and AI-applications. Consequentially, it is essential that users understand how an AI-model operates and what its limitations it has. A first rule of thumb may be that a generative AI model cannot be used as a search engine, fact database or calculator – although there is a nuance in place if the model uses suitable plug-ins and its use and corresponding steps and logic is transparent and verifiable in the output. A second rule of thumb may be that a user must be prepared for faulty, erroneous or false output, as well as arbitrariness and bias in the output.

...also because too much dependence on generative AI, risks the loss of valuable human insights and creativity. Generative AI models offer efficiency and new perspectives. However, this does not automatically mean that generative AI should always be the first step in the process of conceptualizing and creation. An approach where human thinking is the first and foremost element in the process ensures the presence of human expertise and nuance that AI simply cannot replicate. Generative AI can then be helpful for refinements (e.g. writing out a letter), checking and rewriting (missing something or can it be clearer?) or variation (options).

Systemic risks of foundation models

Scaling up foundation models introduces the risk of concentration of power and influence. In recent years, the amount of data on which foundation models are trained has increased. They are currently the biggest models that show the most impressive skills in language and image creation. Training these models takes increasing amounts of data, computing power and time. The high initial investment costs create barriers to entry that only big tech companies seem to be able to overcome. Another consequence is the chance that some foundation models will become dominant, which limits the ecosystem. The dependence on a small number of foundation models entails concentration risk (***single point of failure***). The model errors (bias), security issues (cybersecurity) and organizational risks (governance and probability of failure) that can be associated with a foundation model can have effects through all services and users who rely on such a model.

Scientists and regulators also point to the risk of homogenization and herd behaviour in AI models, which can arise from foundation models. Homogenization means that, at the moment, almost all powerful foundation models are derived from (the technique behind) a small number of models, which were developed a few years ago. Already in 2021, the Center for Research on Foundation Models (Stanford University) mentioned the risk that all these models are vulnerable to the same risks and, for example, sensitive to the same type of bias.[16] Furthermore, according to the institute, this increasing homogenization is having a dangerous effect. From an ethical and safety point of view, reducing the common risks should therefore be the central challenge in the further development of these models.

Continuing this line of thinking, the US Securities and Exchange Commission (SEC) has pointed out the risk for market participants in the financial sector who rely on the same (generative) AI models with herd behaviour. This can simultaneously lead to monoculture and greater network interconnectivity in the financial system, which leads to systemic risks.[17]

Homogeneity of AI models may lead to consistent disadvantages for certain groups which can only be measured when we look at the ecosystem as a whole. Recent research has also revealed this in more traditional machine learning models. Such systematic disadvantages for certain groups can occur, for example, in AI for recruitment and selection for dermatological research. Research also shows that bias reduction within an individual system usually benefits people who are already correctly assessed by other systems.[18]

The increasing use of generative AI can also have an impact on the performance of foundation models (*model collapse*) through feedback effects over time. This risk arises when foundation models are trained on synthetic data that are the result of previous foundation models. The more the output of generative AI finds its way into the real world, the harder it becomes to distinguish these generated texts or images from authentic texts or images.[19] By training on generated data, a foundation model distorts itself, just as a photo of a photo is of lower quality than the original. To maintain the performance of foundation models, the need for identifiable authentic data increases.

Disinformation

Generative AI models are very suitable for image generation. This will therefore require a different approach to identifying the reliability of images. These generated images differ in quality and equality of 'real' images, such as news photos, but the rapid development of these models improves the quality and likeness. As a result, images that were judged to be 'real', but were actually generated have gotten a lot of attention this past year. These generative models are increasingly accessible, not only for the generation of images but also the generation of sounds and voices. People are more and more at risk of identifying generated images as authentic 'real' images. Which can attribute to, for example, the spread of disinformation and misinformation and the associated threat to the functioning of our democracy. In order to cope with such risks, the AI Act will require a form of 'watermarking' or labelling of generated images in certain cases. A number of developers of generative AI models have also indicated that they want to integrate these watermarks or labels in their models. Due to the rapidly growing number of applications of AI models, increasing numbers of users can generate images themselves.
But these models are also adapted for specific use, so that the responsibility for including watermarks or labels may have to be for the end users. Where people now trust the images they see, a whole set of safeguards and proper education is needed to prevent people from losing confidence in the reliability of images and from causing damage to society.

Verifying the origin of images is becoming increasingly important. There are already initiatives to do this and an initiative that has taken a different approach is the Content Authenticity Initiative. This is a coalition of media companies, technology companies and NGOs, who have jointly developed technical standards that, among other things, enable cryptography to verify the origin of images. A number of companies are using the C2PA (Coalition for Content Provenance and Authenticity) standard, which has already been launched in some cameras in late 2023. This enables two important guarantees for the reliability of images. Firstly, that organizations can justify the origin of images and organize their processes accordingly. Secondly, it becomes possible for anyone who 'consumes' images to develop their assessment and knowledge of the reliability of images. Trusting your senses is no longer always possible, this may shift to trusting the source and verifiability of image. Being able to check the origin and reliability of images that have a high social value or impact is a valuable contribution to controlling algorithms and the effects they can have on society.

Supervision of foundation models and generative AI

In Europe, the AI Act will introduce a supervisory regime for foundation models, also known as general purpose AI (GPAI) models. The AP welcomes the supervisory powers provided to a new European AI Office within the European Commission. This also creates an agile structure to respond to further developments in foundation models and their capabilities in the future. The AP will work with risk alerts and will advise on the further development of rules for GPAI models to contribute to this regulatory structure.

The AI Act will provide for a tiered regime for foundation models. Some general obligations will apply to all GPAI models. For GPAI models with systemic risk, additional and specific requirements will apply. Providers of high-risk AI systems that build on GPAI models will have to comply with the regular provisions of the Regulation and will have to, for example, check their AI systems for compliance.

All providers of foundation models will be subject to obligations that help mitigate risks further down the AI chain. Providers of GPAI models should (1) document their model and (2) provide information to 'downstream' providers of AI systems. Furthermore, providers may need to develop a policy to respect copyrights and summarize the data that is used to train the model. Also, providers of models that can be used to generate content, such as images and text, should include a digital watermark in the generated content.

The AI Act sets out an additional regime for foundation models with systemic risk (systemically important GPAI models). According to the AI Act, such a model exists if it can have a significant impact on the internal market. This has to be assessed on the basis of technical possibilities or the significance of its impact. To determine this, quantitative criteria such as the large amount of computing power for the training of the model, the number of parameters in the model and the number of business users that use the

model can be considered. A foundation model is in any case systemically relevant from a certain threshold of computing power that is required for training, with 10^{25} ***floating point operations per second*** (FLOPs) as the probable lower limit. Preliminary estimates circulating in the media are that most GPAI models currently do not meet the limit, with ***GPT4*** (OpenAI) and ***Gemini*** (Google) as possible exceptions.

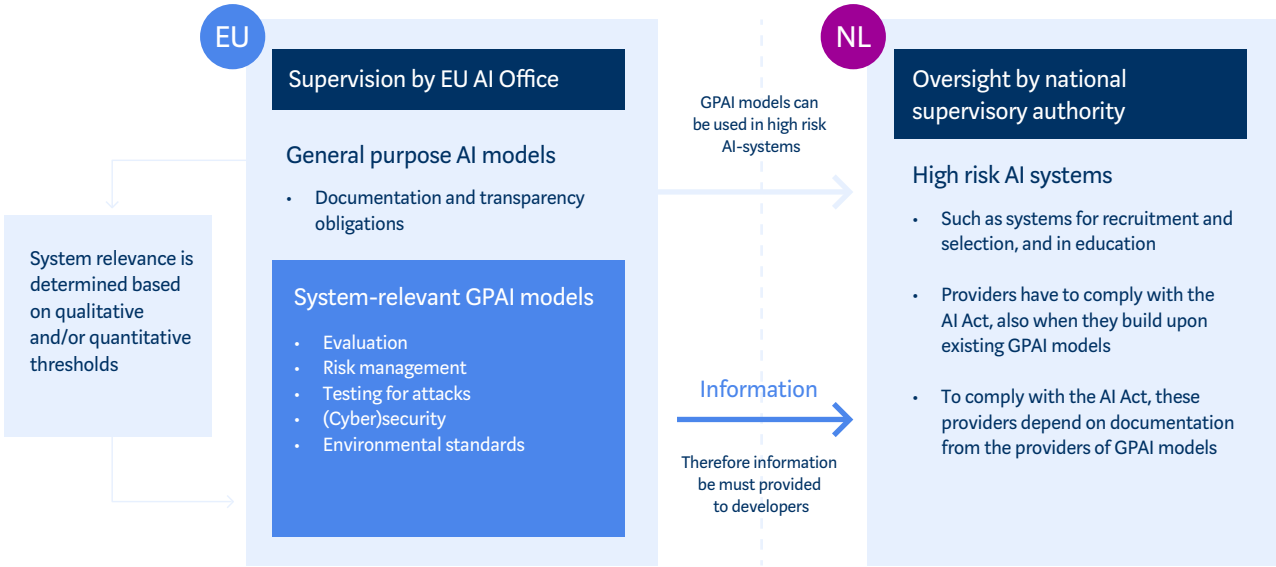
Providers of systemically relevant foundation models are subject to stricter requirements. The additional requirements that providers of these models must meet are that they (1) evaluate their models (2) identify and address potential systemic risks of their models, (3) test their models for attacks, (4) report incidents, (5) take security measures both in the field of cybersecurity and for the physical infrastructure of the model and (6) to meet certain environmental requirements.

A new European AI Office will oversee these foundation models. This AI Office will be able to enforce the rules on foundation models in the AI Act. The AI Board plays a role through their involvement in the further development of standards for foundation models (see further chapter 5). In addition, the AI Office can also designate systemically relevant foundation models based on qualitative criteria. Graph 8 provides a schematic representation of the intended regulation of GPAI models in the AI Regulation.

In addition to the AI Act, there is a need for frameworks for the safe use of generative AI by organizations, the AP wants to contribute to this. Because the human-machine interaction in generative AI is different from other algorithms and AI, there is a need for clarity about how organizations can responsibly embed generative AI into their processes. This can build on producers' responsibility under the AI Act to provide users with, simply put, a guide for safe use but it cannot be ruled out that an organization-wide perspective is also needed. For example, a specific elaboration for generative AI that fits in with the overarching international standards for AI risk management. The AP will contribute to the discussion about how this should take shape with national and international partners in 2024.

New frameworks for foundation models are also being developed outside Europe. The President of the United States recently introduced a supervisory regime for the large foundation models through an ***executive order***, due to their potential threat to national security. The idea behind these measures is that foundation models can at least in theory also be abused by malicious actors and it is therefore important to have a certain level of control over them. The US approach demands that from a certain model size – which it is likely is not as yet reached – developers of these foundation models must provide the federal government with information about these models on a continuous basis. In doing so, they should report on the training, development and production of new foundation models and protective measures, both physical and cybersecurity-oriented measures. This includes transparency on the ownership and control of model weights. Further technical development of these measures will follow in the coming period.

GRAPH 8: REGULATION OF GENERAL PURPOSE AI MODELS



3. Algorithms and AI in the workplace

Algorithms and AI are increasingly being deployed in many areas and thus become intertwined with many aspects of our society. Many algorithms are used for various purposes and may not always be easily perceivable. One of the purposes that algorithms are deployed for is the management of labour forces. Such deployment of algorithms in the workplace can increase efficiency, but can also lead to undesirable effects for employees.

Labour market development

The emergence of algorithms and AI affects many labour sectors in different ways. In many labour sectors, major changes are expected through the implementation and use of algorithms and AI.[20] According to the Dutch Central Planning Bureau (CPB), the development of new technologies, including AI, ensures both the disappearance of existing jobs and the introduction of new jobs.[21] The effects of AI on the work floor are only gradually being noticed and are therefore only partially mapped. Clear risk analyses are essential to create effective regulation for the deployment of AI in the workplace, therefore, further risk analysis is imperative.

The platform work sector is a sector that develops rapidly, partially due to the use of AI-systems. Platform work goes hand in hand with the use of algorithms. This sector increased by 65 % from 2015 to 2019.[22] Currently, there are around 100.000 people working through a platform in the Netherland.

Algorithms play an increasingly important role in work that is not done through platforms. For example, algorithms are increasingly used in the management of staff, but also in the recruitment and selection of new staff.[23] The use of algorithms for recruitment and selection involves many risks, while the correct functioning of these systems and applications is rarely proven.

Algorithms and AI in the workplace

More than 75% of companies in the Netherlands will use AI applications in the workplace over the next five years[24]. In addition, 40 % of large companies are already using AI systems in the workplace.[25] This leads to automation of the l management of labour, the so-called 'algorithmic management'. The ***International Labour Organisation*** describes six levels of automation of such algorithmic management in a recent publication, signifying its importance.[26] Although work and production optimisation models have been used for a long time, organisations often succeed in increasing employee efficiency through using data and AI-systems. The deployment of AI systems in the workplace will be regulated by the AI Act as far as recruitment, selection and decisions are made about people who work.

On the international level, the effects of algorithms and AI in the workplace are placed high on the legislative agenda. The challenges of AI systems in the workplace are mentioned as a point of focus for the coming years by the European Commission.[27] The United States government will also conduct research into the effects of algorithms and AI on the workplace.[28] In the beginning of 2024, these first reports are expected to be published, which will in turn inspire policy discussions. Mapping the risks of algorithms and AI in the workplace is therefore an essential first step in a legislative process. In cases where risks are already well defined and mapped, legislative processes have been set in motion. The European Platform Work Directive is an example of this two-step process.

The use of algorithms and AI to promote labour productivity is standard practice in some sectors. For example, in the delivery sector, where labour productivity and proper planning are of great importance. But algorithmic management systems are also used in other employment relationships to increase efficiency and enable employees to do more work in less time. Algorithms have a large ***span of control*** and make it possible to control employees on a larger scale when compared to a human manager.

Work monitoring by algorithms can ensure that people are approached as a labor commodity. Some employees need to increasingly rely on an algorithmic management system in their daily work. However, this introduces risks that can be caused by the use of algorithms. Employees can start to feel continuously watched and analysed by an algorithmic system. Employees may also dread bad days with reduced productivity, since this will not pass unnoticed by the watchful algorithmic system. Through algorithmic management, the overall relationship of authority in the workplace shifts from

human-to-human to algorithm-to-human. The greater the role of algorithms in labour management, the more important the role of a manager who can bring a human measure.

Risks of algorithmic management systems

Algorithms and AI are used in specific professions and therefore also have sector-specific risks, in addition to the more general risks. A system that manages a group of warehouse employees can be less useful for a fast food restaurant than for example for an online store. The risks may also play out differently, given the different context in which the system is deployed. The use of algorithmic management systems therefore requires specific knowledge of the particularities and size of an organisation. Using algorithms to manage employees can therefore work differently across different sectors and companies. Getting a grip on the risks of using an algorithmic management system requires customisation.

Algorithmic systems may increase employee workload. This increase in workload can cause employees to burn out faster than they would under human management systems if there is insufficient attention for personal circumstances. Moreover, algorithmic systems may act as inhibitor between signals from employees to employer. Therefore, it can be more difficult for these workers to discuss labour pressures with their supervisors.

Algorithms and AI do not have a perfect insight into the physical world. This can lead to inefficiencies, burdens and frustration among employees. The labour pace of algorithmically managed work may also be unreasonably low. As seems to be the case with several delivery services in the Netherlands. If parcels have to be delivered within a limited period of time, delivery providers may have to wait before they can start delivery in the next delivery window. This can mean that delivery drivers have to wait in front of a house and cannot deliver their parcel yet, because the system does not provide clarity as to which package for which house, has entered into the delivery period.

The use of algorithmic systems may lead to behavioural changes in employees. For example, in call centres, algorithmic systems are used that distribute incoming telephone traffic between the available workers and at the same time monitor employees. Another example is that there are systems that analyse the emotion in a call center employee's voice and point the employee to a lack of audible empathy. Algorithmic management systems can therefore nudge or even encourage employees to change their behavior to work more productively or more effectively.

Algorithms and AI may affect employee autonomy. Human managers mostly allow employees to indicate their wishes to perform processes in a different manner than the algorithm prescribes. In the case of algorithmic management systems, the processes and coordination of labour may be more rigid. In part this is because acquired knowledge of employees does not quickly translate into algorithmic management systems. Therefore, the ability to respond to future challenges and use of experience and individual qualities may decrease when using an algorithmic management system. Employees may decreasingly be in charge of their own work when algorithms increasingly make decisions in the work space.

Negative impacts on public values and fundamental rights. The effects of the deployment of algorithmic systems in the workplace may also have an impact on public values and fundamental rights. For example, on the right to privacy but also on the right to equal treatment. Moreover, algorithmic management systems may affect the extent to which and how employees can receive information. Algorithmic management systems often require the collection of a lot of data and gain insights from this, which are not accessible to employees. This unequal information position can lead to a weaker position of employees.

Platform work

Platform workers face greater risks when compared to others through algorithmic management. Platform workers often lack the means of opposition in their relation to the platform providers. Moreover, there are fewer human managers operating on the 'work floor'. As a result, the risks of unreasonable workload, diminished autonomy and safety and behavioural change for platform workers through nudging are potentially even greater. Providing proper protection

to these platform workers is thus an essential part of a sustainable structuring of the labour market in which algorithms and AI will play an ever-increasing role. The European Platform Work Directive, on which a political agreement has been reached on December 13th, will be important for this.

Risks deploying algorithms and AI for platform workers

In the case of platform work, protection from against risks is difficult. Platform workers often lacks the social structure that can speed up resistance or push back against the provider or platform. If the nature of the work involves platform workers having little or no contact with colleagues, then platform workers also share limited experiences or hardships. This makes it difficult to discover common problems together and to propose targeted solutions for this together with a platform.

Assignments that platform workers are offered via a platform, cannot always be refused without consequences. The refusal of assignments that yield little, take a long time or involve additional work can have an impact on the offer of future assignments and therefore on future income. Platform workers who accept more assignments will be more likely to be eligible for assignments with good merits or pleasant work. Platform workers are therefore very much dependent on the algorithms that are being used by platforms.

Algorithmic rating systems may cause behavioural change among platform workers. ***Rating systems*** are very popular with providers of platform work. In addition, good ratings have a positive effect on many aspects of the work performed by a platform worker. However, platform workers also rely on having good ratings for getting work.
The ratings are taken into account by algorithms when distributing assignments among the available platform workers. The relationships between platform workers and customers are designed to ensure that platform employees aspire to receive the highest possible score.

Future regulation

The Platform Work Directive will introduce requirements for the use of algorithmic management by digital labour platforms. The Directive will clarify the employment status of people working through a platform. But the directive will also set requirements in terms of transparency and limits on algorithmic management by platform providers.

Pending legislation, organisations can work on their own to ensure that algorithmic management systems run properly. Sufficient organisational measures are of great importance here. Managing workforce with algorithms and AI – if arranged in a responsible way – can bring all kinds of benefits for both employees and employers.[30]
The introduction of new technologies has changed the labour market over and over again.[31] Algorithms and AI will not be an exception, but their effects on employees need to be closely monitored to mitigate risks. Organisations should be set up in such a way that algorithms and AI in the workplace take into account the quality of jobs and working conditions and that they do not only serve efficiency purposes but also mitigate the risks accordingly.

Interview: Training algorithms and AI

This interview was conducted in November 2023. The interviewee is a platform worker in the Netherlands who is training AI models, like the LLM's that received a lot of attention in 2023. An estimated tens of thousands of people worldwide are working as platform workers to assess and improve the outcomes of algorithms and AI models.

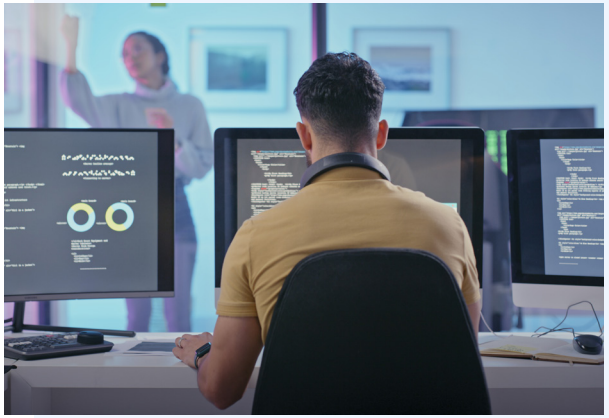
Can you describe your work as an AI trainer?
"I work for a number of online providers, largely training algorithms on accuracy and relevance, for example when it concerns search results. Recently, rewriting and assessing answers from AI models has been added. I judge them on the accuracy and quality of the answers. And I check if the model doesn't pretend to be a person. But also the creation of prompts and the rewriting of answers from the AI model are among the assignments I do."

What made you consider to become an AI trainer?
"This type of work fits me well because all the work is done on online platforms. It gives me freedom, it can be combined with other creative activities and does not depend on your place of residence."

Are there any disadvantages to having this freedom? "Yes, this type of work does not provide me with a regular income, and the supply of assignments varies greatly. There are no guarantees or rights for me as a platform worker, nothing to fall back on, such as a contract. You can lose your income from one day to the next. Also, fees can suddenly be lowered. Payments that are not made – which does occurs – are difficult to get without a human point of contact."

What exactly does training algorithms involve?
"I work on projects with thousands of people worldwide who improve these models and systems and provide answers, texts and reviews. The fact that increasing volumes of people from all over the world are working on these AI models is causing the language quality in these models to deteriorate. This is due to continuous pressure on people's qualifications and the quality that people have to deliver, without having any rights to fall back on."

Do algorithms influence your work? "Algorithmic assessment of the quality of your assignment or the time you spent on your assignments can sometimes cause you to lose access to projects. For example when you finish an assignment too quickly. Or if the quality of your assignment is judged as not being sufficient. However, it is often not clear what the rules or guidelines are. So in some projects you have to comply with rules that are not known or clear. This makes you feel interchangeable as a worker.

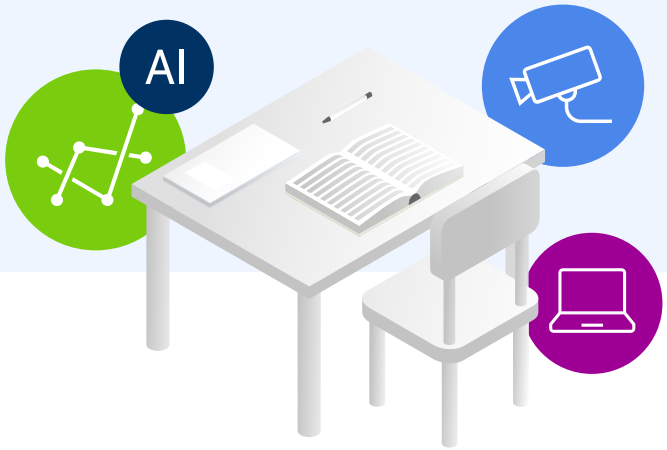


Especially when systems assess whether you are entitled to your income that month. But there are also projects or providers where there is an abundance of information on the basis of which you have to work on assignments."

Did this work affect your trust in algorithms?
"I didn't immediately gain more or less confidence in algorithms. But this does help me to value algorithms and AI. As a platform worker, you can see the quality of these AI models. This helps to relativize the hype. At the same time, you also see the blind faith that people put in the models while using them. This is evident from the interactions between people and the models, such as the prompts that people write and which I see. I'm worried about how much confidence people put in models that essentially still contain so much human work."

QUICKLY TO THIS PART

Adaptive learning → | Learning analytics → | Generative AI → | Policy approach →



4. Algorithms and AI in education

The education sector (primary, secondary, vocational and higher education) increasingly makes use of algorithms and AI. For example, educational institutions use adaptive learning systems, which offer the teaching materials automatically and adapted to the individual learner. Systems with 'learning analytics' give new insights based on all kinds of data, with the aim of improving the flow of students through the education system or the quality of education. At its core, these systems often provide profiling and prediction of pupils or students. There are all kinds of possible reasons why profiling or prediction does not fit well with the situation of individual students. Careful embedding of algorithms and AI into education and knowledge of their limitations is crucial, as it concerns children and young adults. Pupils and students use a lot of generative AI and they do this to outsource some of their assignments. That challenges education to incorporate the use of generative AI in a good way into the goals that are pursued.

Such a thing has previously been done successfully with the rise of the internet and Wikipedia. The AI Act will regulate a number of AI-systems for education as high-risk applications that require compliance with product standards, but these will mainly apply to developers and will not always be sufficiently concrete. The DCA therefore advises to set up policy strategies and control processes for algorithms and AI within educational institutions, under the guidance of the existing external support organisations. In addition, the education sector should collectively agree on clear product standards with developers of AI-systems. Increasing AI-knowledge among teachers is also a point of focus, to ensure careful embedding and control of algorithms in education.

Education is currently undergoing an intensification of digitalisation and algorithm use. In the years before the coronavirus pandemic, the education sector had already increasingly shifted towards digital teaching resources and learning environments. The pandemic has intensified this digitalisation in education.[32] This makes it easier to collect data on the performance of students. the presence of more data often means that it is easier to use algorithms. For the current intensification of digitalisation and the use of algorithms in education, not only more data, but also more financial resources are available. The National Growth Fund [***Nationaal Groeifonds***] has made 80 million euros available to the National Education Lab AI [***Nationaal Onderwijslab AI***],[33] which conducts research into the opportunities and risks of AI in education.[34] Many schools have also used part of the National Education Programme's [Nationaal Programma Onderwijs] budget to purchase digital learning resources.[35]

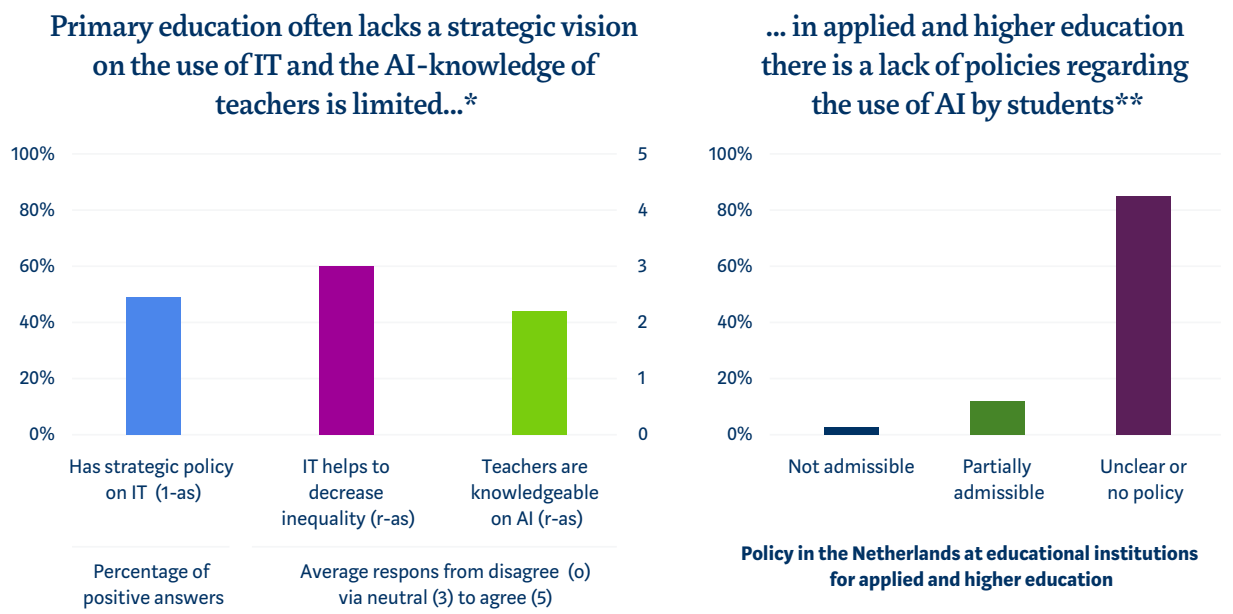
In order to use algorithms responsibly, the education sector needs to be able to make a better estimation on what the sector can and wants to achieve with the use of digital resources. For example, figures from the Kohnstamm Institute[36] and Npuls magazine[37] (2023, see Graph 9) show that half of primary schools do not have a strategic ICT policy. In addition, more than half of teachers have no knowledge of AI. Also, only a few universities and vocational schools [***MBO's***] have policies for AI-use by students. However, most students at universities already use Large Language Models. In the case of Erasmus University, this amount to even more than 90 percent.[38] These figures underline the need for vision and policy for digitalisation in education.

Two well-known algorithmic applications in education are adaptive learning systems (AL) and learning analytics (LA). There are many more applications of algorithms in education, from applications that help teachers compile lessons in secondary education to algorithms that help in the scoring of students' work. AL and LA, however, are the best-known examples of commercially available systems that are used by educational institutions and show the risks of algorithm use in education. AL is mainly used in primary education, LA primarily in higher education. These are the subsectors that use algorithms the most, the vocational and secondary education sectors currently use fewer algorithms.

Adaptive learning

Adaptive learning systems select practice material for individual students based on how well a student has made previous exercises. More than half of primary school pupils practice with education materials in such a system on a daily basis. Greatly simplified AL works as follows: a student does three tasks in a row quickly and correctly. The AL-system captures this and interprets it as a sign that the student is at a higher level. On this basis, the system then selects three tasks at a higher level of comprehension. A student who makes the tasks slower or incorrectly is presented with a lower level of difficulty. The system therefore profiles the

GRAPH 9: EDUCATIONAL INSTITUTIONS DO NOT ALWAYS HAVE POLICIES ON THE USE OF IT AND AI-APPLICATION



)SOURCE**: KOHNSTAMM INSTITUUT (2023). *)SOURCE**: NPULS(2023)

student and adjusts the profile after each assignment. Based on that profile, it predicts which assignment offers the pupil the best chance of progress and selects these. Many AL applications use an algorithm that was once developed to rank chess players. The fact that AL is mainly used in primary education is due to the fact that the clearly delineated skills pupils learn in primary education lend themselves well to the application of AL. In addition, primary education forms a large market, as each school works towards the same learning goals.[39] Primary education is therefore more attractive for providers of algorithm applications than higher education, which is more differentiated when it comes to teaching materials.

Teachers use dashboards of adaptive learning systems, which provide real-time insights into the interaction between the system and the student. Dashboards are digital environments in which the progress of individual students is shown, in the shorter or longer term. Dashboards are found in several parts of the education sector and are not exclusively linked to AL-systems. In secondary education, however, they often are. A dashboard linked to an AL-system helps teachers to quickly adjust the learning process of pupils. The dashboard shows data with which the teacher can see which pupil is progressing well at that moment and who needs attention. There are dashboards that mainly show data, but there are also dashboards that analyse and interpret the data as well. For example, by visually categorising pupils or even proposing a certain intervention by a teacher.

Despite the intention of providing tailor-made education to each child, adaptive learning systems can lead to inequality of opportunity. The models at the base of AL-systems are often less complex than reality in the classroom. This means that students who structurally study easily receive exercises that encourage them to develop, or that keep them working at their level. The system has more difficulty with identifying the needs of pupils who do not yet master the educational material. The system does not take into account differences, other than the varying success ratios between students, such as different housing backgrounds and concentration spans, whereas such differences can explain the sometimes differing needs of children with similar numbers of correct and wrong answers. If a system has difficulty recognising such underlying differences between pupils, some students will be misprofiled and the system therefore selects less fitting tasks for these students.

In order to ensure equal opportunities in the use of adaptive learning systems, teachers and adaptive learning systems should better complement each other. If students are improperly profiled by an AL-system, the system often presents them with the same tasks, while they have different needs. To provide these students with what they need, teachers should assess what is needed based on their dashboard and their own perception. This requires some knowledge about the operation of the AL-system and the difference between model and practice. We call this "algorithmic literacy." and this is not always present. Furthermore, if they are algorithmically literate, teachers should be able to actually use their skills in supporting pupils. Schools therefore cannot make unlimited use of AL systems, in order to free up teachers for other tasks.

School boards, teachers and developers of AL-systems share the responsibility to reduce the risk of inequality of opportunity due to AL systems and to realise the opportunities for personalised education. Teachers should be able to intervene when AL-systems make inaccurate, inappropriate or incorrect choices for their pupils. They need "algorithmic literacy". School boards should shape their policies in such a way that teachers can acquire and use the necessary algorithmic literacy, for example through training. Furthermore, it is to be recommended that school boards should avoid irresponsible purchases of AL-systems, by determining which implementation options the school has before making the purchase. In addition, it is also important to establish a clear organisation of implementation and use with a clear division of responsibility. Research and pilots are needed to better understand the interaction between students, AL-systems and teachers. This should result in better interaction between AL-systems and teachers in order to create synergy between humans and machines. For this, developers, boards and teachers have to take action. When it comes to this, support organisations with expertise in this field, such as Kennisnet and SURF, are indispensable.

Learning analytics

Learning analytics (LA) is the use of data to gain insight into the progress of pupils and students and the quality of education. LA is divided into two purposes: firstly, the guidance of students and improving the flow of students through the education system, and secondly, improving education. An adaptive learning system that shows teachers real-time data about students therefore essentially also uses LA. However, the term 'learning analytics' usually refers to gaining insights

from data over longer periods of time, focusing on structural interventions such as adjustments to the curriculum or determining more precisely what hinders students' progress. LA-systems are not yet used massively. However, higher education institutions in particular are actively exploring application possibilities. SURF assists them in doing so.

LA is considered to be promising[40] and is available, but educational institutions generally have no vision of or policy for the deployment of LA-applications. Large learning management systems, such as Canvas and BlackBoard, offer integrated LA. For example, the University of Utrecht has already set up an LA-team.[41] However, many educational institutions do not yet know what exactly they can and want to do with an LA application in a way that students, teachers and board are satisfied with it. Collecting data without a clear purpose can lead to data being used in irresponsible ways. In addition, this is not allowed when it concerns personal data. It is therefore advisable for educational institutions not to deploy learning analytics on a large scale before a clear vision for its use has been developed.

Ill-considered use of LA can have a major impact on individual academic careers and privacy. The use of LA to improve progress of individual students requires profiling and the processing of personal data. That always means an infringement on their privacy. The use of LA without a clear purpose and clear frameworks significantly increases the likelihood of unlawful or disproportionate privacy infringements. The use of LA in decision-making on academic careers also requires a well thought out plan. Using data and profiles without clarity about what the data means and what conclusions can be drawn from it increases the risk of unjust treatment of students.

The education sector must develop clear, safe *use cases* for the adoption of LA, in which the interests of students are paramount. The higher education sector is working, in collaboration with among others SURF, on knowledge and expertise about LA. This is a cautious approach that is to be encouraged. The interests of students should always be paramount in the development and use of LA. Therefore, it is recommended that educational institutions actively involve student populations in developing final use cases. The first experiences from higher education already show that students are open to this and that their involvement raises the use cases to a higher level by making them more useful and meaningful for all concerned.

Generative AI

Generative AI is now becoming ubiquitous and requires a clear policy from educational institutions, similar to that for the use of internet resources. A small survey from Erasmus University Rotterdam shows that 92 percent of students use ChatGPT for different purposes.[42] In secondary education, students use generative AI for homework assignments. However, institutions in vocational and higher education still have little policy on what is and is not allowed with regards to generative AI. The evident risks of generative AI to the education sector are the unauthorised use by pupils and students, such as plagiarism, and misinformation through faulty but plausible output. In addition, the question is whether and how generative AI can be used lawfully. Partially because of this legality issue, it is currently wise for educational institutions themselves to be reluctant to use generative AI. Consequently, the sector can focus on mastering the use of generative AI by students. This has previously been done successfully with regards to the use of internet resources, such as Wikipedia.

Education cannot eliminate generative AI and thus can better teach pupils and students how to properly use it. It is foreseeable that generative AI, just like Google and Wikipedia, will be used daily and ubiquitously. A consideration for education is to follow a two-track plan. On the one hand, pupils and students need to learn to deal with generative AI by using it. This means that education needs testing and teaching methods that guide pupils and students with controlled use of generative AI. This must be done in a way that adequately takes into account the legality questions surrounding foundation models and derived applications. On the other hand, pupils and students need to be able to assess generative AI by being able to also do what generative AI can do.

Policy approach

The AI Act imposes requirements for profiling and evaluating algorithmic applications in education in certain cases. The education sector has to complement these requirements. This means that the systems themselves must meet strict requirements in order to be admitted to the European market. However, there is some time to go until the AI Act will be applied. Moreover, the requirements of the AI Act are general and not specific to education, and do not apply to all actors in the chain. For responsible use of applications of algorithms, partners in the education sector therefore also need to decide which requirements systems must meet and pose those requirements to developers.

Educational institutions should prepare their organisation for integrating algorithm applications into education prior to their purchase. Upcoming and new regulations such as the AI Act do not consider yet how users actually embed algorithmic applications in their organisations. Proper embedding is of great importance to mitigating risks. Teachers must be sufficiently algorithmically literate to interpret the outcomes of systems they work with. Educational institutions should clearly delineate responsibilities and think in advance about the evaluation of algorithmic applications. Furthermore, it is important that educational institutions identify possible negative effects of algorithmic applications and determine in advance how they can be covered.

QUICKLY TO THIS PART

AI Act → | AI standards → | Digital Services Act → | Supervision and control in the Netherlands → | Insight into the supervision of algorithms and AI in the Netherlands → | Regulatory sandboxes →

5. Policy and regulations

New, more robust policies and regulations for algorithms and AI are emerging. New legislation has entered into force or will do so in the foreseeable future. The negotiations of the AI Act have been concluded with a political agreement and the AI Act can enter into force in 2024. This is an important step that will contribute to the control and control of AI systems. The Digital Services Act (DSA) is an important European regulation that has recently entered into force and requires very large platforms and search engines to address algorithmic risks. On December 13th, a political agreement has also been reached in the negotiations on the Platform Work Directive, which will provide rules on algorithmic decision-making and monitoring of people who are working via a platform. This new legislation greatly contributes to protection and guidance for the management of algorithmic risks. However, even though these regulations are closely related, coherence between these regulations is still a challenge. A lack of coherence can create ambiguity about rules, supervision and therefore also about the management of algorithmic risks.

AI Act

There is a political agreement on the AI Act. Public, private and supervisory organizations should therefore prepare for the implementation in order to comply with the upcoming AI Act. This regulation is likely to apply in mid-2026 at the earliest. The ban on certain AI systems is likely to apply after 6 months (end of 2024); the provisions on generative AI models and governance are likely to apply after 1 year (mid 2025). Now that there is a political agreement, it seems that the structure of the AI Act is not substantially different from what is described in the previous ARR (July 2023). The AI Act will therefore, in particular, explicitly regulate the development and market entrance of AI systems. However, a number of new elements have been added which are likely to have an impact on the effectiveness of the regulation in managing and controlling algorithm risks, which will be explained in the following paragraphs. With the start of this new phase, it is important to emphasize once again that the AI Act will only succeed if governments, private parties and supervisors work towards clear and adequate standards, a consistent interpretation of regulations and an effective structure for supervision and cooperation in the field of AI.

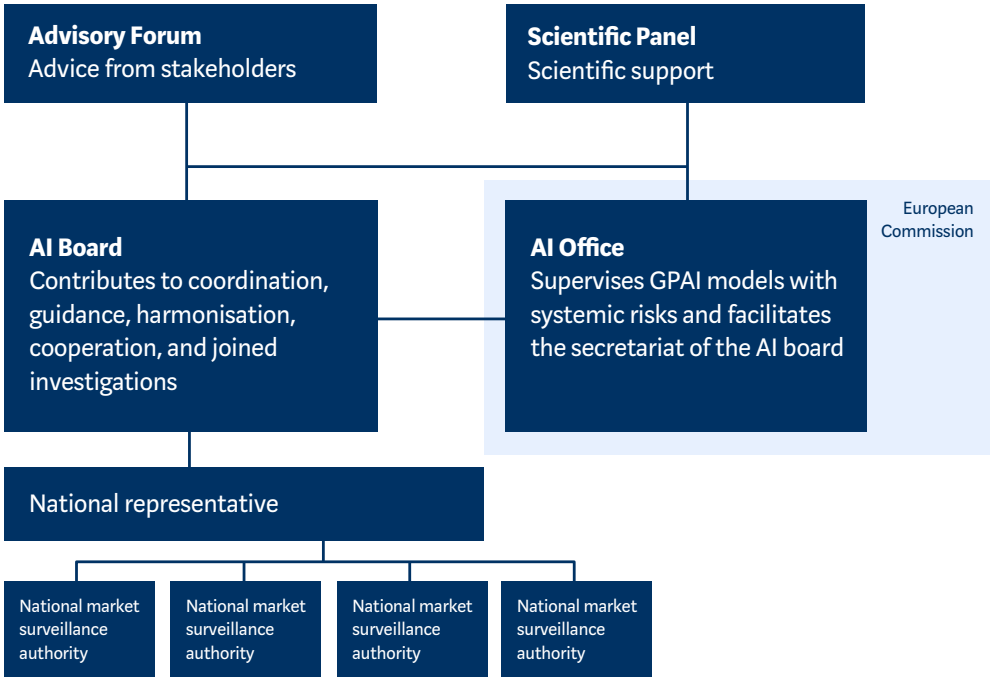
The *"carve-out"* for high risk systems limited to preparatory or supporting tasks has made the regulatory framework of the AI Act more ambiguous. In previous texts of the draft AI Act, certain systems were classified as 'high risk', such as systems intended to make decisions on the recruitment and selection of job candidates. Such AI systems can only be placed on the market if they meet the requirements of the AI Act. However, the regulation does not apply to AI systems with preparatory or supporting tasks ('carve-out'). Especially for AI systems that are intended to perform a preparatory task or that have an additional role in human assessment, for example for the confirmation of a decision. This carve-out is not without risks because people tend to rely too much on preparatory assessments of AI systems, a principle known as "overreliance." Even if an AI system has only a preparatory or supportive task, the risks of discrimination can materialize, even though people are involved in the process of decision-making. In addition, there is a risk that providers incorrectly place their systems on the market as 'preparatory' or 'supporting' systems. It is then up to the market surveillance authority to start enforcement.

The AI Act gives rights to people who are impacted by AI systems. People will have the right to file complaints about AI systems. They also have the right to receive explanations of decisions that are based on the use of high-risk AI systems, such as personnel recruitment and selection systems or for education purposes. Organisations using an AI system will also have to assess the potentially negative effects of the system on fundamental rights in a ***fundamental rights impact assessment*** (FRIA) and take measures to prevent it.

Governmental *use* of high-risk AI systems should also be registered in a European database. As a result, the European AI register will be a mix of a product register (which products have a CE marking and are authorized in the European internal market) and a usage register that may in some aspects be similar to the algorithm register that has been developed by the Dutch government.

European problems call for European solutions. AI systems can have a societal impact in all European member states. Therefore, it should be possible to act decisively under the AI Act if needed. To achieve this, part of the supervision will be set up at European level. This supervisory structure consists of several components. First of all, an ***"AI Board"*** will be established which will consist of national representatives and will have an advisory and coordinating role. For example, the Board can provide opinions and recommendations on which AI systems should be identified as high risk for the standards and the development of the AI landscape in Europe. In addition, there will be an ***Advisory Forum*** and a ***Scientific Panel***, where the latter also has a role in supervising general purpose AI. The supervision of general purpose AI will be a task of the "AI Office" that will be part of the European Commission. Graph 10 gives a schematic representation of the intended governance structure.

GRAPH 10: EUROPEAN GOVERNANCE FOR AI ACT SUPERVISION (SIMPLIFIED)



The AI Act will provide for a tiered regime for so-called general purpose AI models (foundation models). Horizontal obligations will apply to all general purpose AI models. Additional requirements will apply when these models present a systemic risk. Chapter 2 deals with foundation models and their supervision.

AI standards

Standards ("*standards*") will address and specify the requirements of the AI Act. The AI Act will provide general rules ("essential requirements") that developers must comply with when developing their AI system. These include risk management, the quality of datasets, transparency, human oversight and cyber security. Standardization organizations CEN and CENELEC are already working on standards for developers at the request of the European Commission.[43] The NEN, the organization that manages standards in the Netherlands, participates in these organizations and represents the interests of the Dutch market and participants. Developers are not obligated to adhere to the standards. But if they do, it is assumed that their AI systems comply with the essential requirements of the AI Act. In practice, the standards will therefore play a major role in demonstrating compliance and assessing conformity.

The development of AI standards should be done in a balanced and transparent manner, taking into account all interests and the protection of fundamental rights. The Dutch government has a special responsibility in this regard. Standards are flexible, provide clarity and are co-produced by organizations with a lot of technological knowledge. At the same time, the development of standards is largely out of sight of society and risk the improper representation of some interests. It is therefore important that not only companies that develop AI are represented in organizations such as NEN, but also public institutions and organizations that stand for the protection of fundamental rights and the interests of citizens. Where public interests are at stake, public institutions must provide insight into their own commitment and input into these norms. This promotes the legitimacy of government action and gives more insight into and knowledge on this otherwise private and closed process. The standards themselves must also be publicly available. The European Commission should closely monitor a fair balance of interests and the protection of fundamental rights as soon as the Commission harmonizes and approves standards. Finally, market surveillance operators should check whether a harmonized standard in a specific case in practice actually respects the requirements of the AI Act, such as requirements on transparency or the prevention of discrimination.

Digital Services Act

In addition to the AI Act, there are also other European (sector-specific) regulations that impose requirements on the deployment of algorithms and AI. An example of this is the Digital Services Act (DSA). The DSA regulates internet services such as hosting services, social media services, online marketplaces and search engines, in order to protect the users of online platforms and holding these platforms responsible for the safety for users of their services. Some of the provisions in the DSA that target very large platforms and very large search engines already apply. Other elements of the DSA will apply in February 2024.

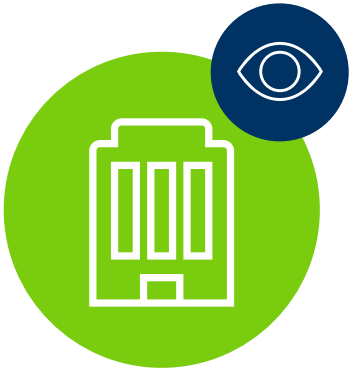
The DSA sets out rules on how digital service providers should act against illegal content. For example, large online platforms should act against disinformation and misinformation campaigns. Especially if the dissemination of such material poses a risk to the electoral process. A concern here is the increasing role of generative AI to produce huge amounts of content on a very large scale, cheap and fast, which can then be distributed through digital service providers. Generative AI can also be used for plausible or realistic looking misinformation and disinformation. At the same time, digital service providers can also use AI systems and algorithms to identify and sometimes remove generated content. The fight against misinformation and disinformation becomes, in a way, an arms race of AI systems.

The discovery and removal of illegal content and misinformation and disinformation with the help of AI systems also creates risks. Unjustified removal of content by an algorithm creates tension with certain fundamental rights, such as freedom of expression and equal treatment. There is a chance that deleted content is not illegal at all or cannot be labeled as disinformation or misinformation. In doing so, digital service providers may play a role in making sure that online content is removed. The lack of human control is also problematic. From the 2023 Transparency Report of X (formerly: Twitter) can be inferred that in the Content Moderation Team, only one employee is able to speak the Dutch language.[44] Unjust removal of legitimate content with algorithms and AI is therefore a risk that needs to be taken seriously.

The European Commission is the supervisory authority for very large platforms, search engines, and the special requirements that apply to these service providers. These service providers have a particular responsibility in the DSA to take action against risks posed by their services. Several platforms, such as X, TikTok and Meta, have already received requests for information on the dissemination of unwanted content on these platforms under the European Commission's DSA. The Commission is therefore taking a first step in keeping very large platforms in check with the DSA in order to take action against the dissemination of disinformation through recommendation algorithms on their platforms. The DSA is therefore a valuable tool to understand and mitigate specific risks of algorithm and AI use. The DSA also enables societies to gain more control over the spread of disinformation and misinformation driven by algorithms.

Supervision and control in the Netherlands

Algorithms and AI are a system technology and become intertwined by all parts of society, including oversight and supervision. It is expected that almost all supervisory and oversight authorities, agencies and inspectorates will come into contact with algorithms and AI when exercising supervision or even in the performance of their tasks. Algorithms are always used in a specific context. Therefore, sectoral knowledge in oversight and supervision is important.
On the other hand, it is necessary to avoid fragmented oversight and supervision. Cooperation is crucial to organizing well-functioning and coherent oversight and supervision of the development and deployment of algorithms and AI.



The Netherlands has a large number of oversight and supervisory organizations to oversee and supervise the development and deployment of algorithms and AI. These organizations oversee and supervise algorithms and AI, monitor their operation and legality, or deal with complaints about public and private organizations that use algorithms and AI. State institutes, market surveillance authorities and inspections carry out a large part of these tasks. In addition, the National Ombudsman and local ombudsmen play an important role. There are also other public and private organizations that have a monitoring function, for example for specific certifications or quality supervision. Finally, citizens and organizations can also go to court to enforce their rights.

Monitoring of algorithms and AI takes place on the basis of a variety of existing and new frameworks. Oversight and supervisory organizations base their activities on different legal frameworks and differences in independence or positioning. International treaties, European regulations and directives, as well as national legislation, form a solid basis for a large part of the oversight and supervisory organizations. These legal frameworks do not often contain specific provisions that focus purely on algorithms and AI or their development and deployment. In addition to legislation, oversight and supervision can be complemented and strengthened by norms and standards. On the one hand, this sometimes offers more flexibility in changing circumstances. On the other hand, these norms and standards offer less guidance for intervention and sanctioning in case of violations.

Multi-layered surveillance is necessary for the application of system technology. Oversight and supervision and are not always shaped as a direct, one-on-one interaction between an oversight or supervisory organization and a supervised organization (the 'supervised'). Where the number of supervised is limited, direct oversight or supervision of an organization's products, services or processes can take place. However, in the case of a large number of supervised, it is often no longer possible to directly monitor or consistently carry out checks. In this case, it can be more effective to use layered supervision. For example, internal supervision in an organization can be organized, for example as is done by appointing a Data Protection Officer (DPO). Such an internal supervisor can identify problems and give advice at an early stage. And as a strategic advisor, they can protect an organization and citizens or consumers from violations or unwanted risks and effects. Another example of a "supervisory layer" is to certify routine procedures or standard products and services and have them supervised by a designated third-party oversight body, such as an auditor.

Algorithms and AI require an internal structure of supervision and control that enables responsible technological innovation. Algorithms and AI are being used in an increasing number of sectors. This makes direct monitoring of all development and deployment from an oversight of supervisory organization impossible. In the case of algorithms

and AI, layered supervision is more appropriate, because the risks, the relationships between actors and their interests vary greatly and depend on the specific applications and context. From the workplace to the board, attention should therefore be paid to identifying, reducing and controlling risks in the development and deployment of algorithms and AI. This task involves introducing internal positions such as an algorithm officer, strengthening knowledge and expertise in the organization, or appointing an AI ***governance board***. The emergence of initiatives for AI governance is a positive development for being in control of Algorithms and AO.

Citizens do not always know which oversight or supervisory organization they can turn to. For citizens, oversight and supervisory organizations are an important safeguard against infringements, violations, risks or undesirable effects and defend themselves against them. Are there signs that standards are being violated? Or is a citizen confronted with risks or undesirable effects of algorithms? This person can then report this to a number of complaint desks of supervisors or ombudspersons. A problem here is that due to the considerable number of oversight and supervisory organizations in the Netherlands, there is uncertainty among many citizens about which organization they can turn to. Appealing to specific organizations that are well-known like the Ombudsman is usually the only option for people to take action if they have a complaint or want to report something. The Dutch government provides no comprehensive overview of all oversight and supervisory organizations where citizens could effectively seek assistance. This lack of information can lead to reduced willingness to report and can weaken trust in government and institutions. This is all the more true if there is a possible misconduct on the part of a public organization. A guiding overview of oversight and supervisory organizations, with associated complaint desks and mandates, contributes to ensuring access to oversight and supervision and the ability to take corrective action. Being able to defend against abuses and illegalities in the use of algorithms and AI is also an obligation from a human rights perspective.

It is important to invest in the cooperation between complaint desks to detect algorithmic abuses. Citizens with complaints about algorithms and AI should be able to go to all relevant complaint desks without obstruction. While citizens are not yet complaining in large numbers about the negative effects of algorithms and AI experienced by them, they should have the opportunity to do so. The fact that the number of complaints is limited, is often due to a lack of transparency about the use of algorithms. As a result, citizens are not aware that there is an algorithm present in a specific process. It may also be unclear to citizens that the cause of a problem is actually in an algorithmic system. In 2024, the DCA will work with relevant organizations that have a complaints desk to strengthen the cooperation between them. The goal is to recognize the presence of an algorithm in an early stage of a complaint and to be able to provide citizens with the assistance that is needed.

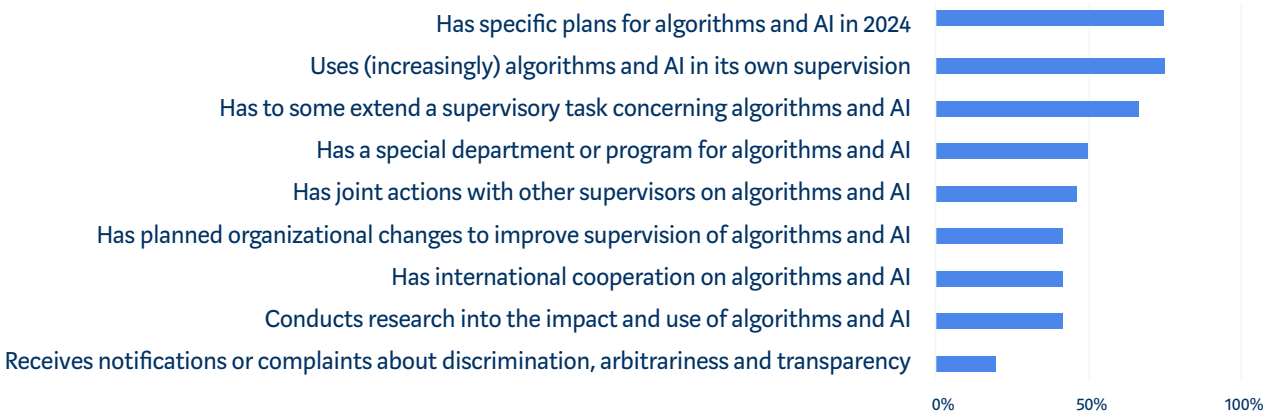
Insight into the oversight and supervision of algorithms and AI in the Netherlands

In order to achieve effective oversight and supervision of algorithms and AI, it is of great importance to gain insight into how oversight and supervision of algorithms is currently been done in the Netherlands. In the summer of 2023, a survey on algorithm and AI risks and observations was sent to 33 oversight and supervisory organizations in the Netherlands. Of these, 24 oversight and supervisory authorities i completed the survey. The results provide insight into the landscape of the oversight and supervision of algorithms and AI in the Netherlands. The results of this survey show that a majority of respondents are aware of the risks and opportunities of algorithms and AI, but that it should also be noted that it differs per respondent on how much algorithms and AI are part of their day-to-day work. Graph 11 gives an overview of the survey results.

Two-thirds of the respondents to this survey supervise algorithms and AI to some extent. The impact of the use of algorithms and AI systems and/or applications varies by sector and therefore also per supervisor. Two-thirds of respondents say they supervise algorithms to some extent and a smaller number structurally oversees the use of algorithms. Other respondents are either mainly extra alert to potential risks or the group of respondents that do not supervise algorithms, which indicate that it is currently either not present or not present enough in their sector. Or that their expertise in this field is lacking.

More than 40% of the respondents to this survey are investigating the impact and use of algorithms and AI. This ranges from explorations, provision of targeted information and investigations into potential algorithm-related abuses to violations in which an algorithm plays a direct role. These investigations result in guidance for organizations and citizens, clarification and explanation of standards, advice, warnings and sanctions.

GRAPH 11: HOW ARE ALGORITHMS AND AI REPRESENTED IN THE WORK OF SUPERVISORY AUTHORITIES AND INSPECTORATES IN THE NETHERLANDS? RESULTS OF A SURVEY AMONG DUTCH SUPERVISORY AUTHORITIES AND INSPECTORATES



The survey was completed by Dutch supervisory authorities and inspectorates in mid-2023. The AP has received a response from 24 organisations.

Four respondents to this survey received concrete complaints or signals from citizens about the use of algorithms and AI. One reason for this may be that citizens do not know that there is an algorithm in play. For this reason, improving transparency in the use of algorithms is one of the core themes of the DCA. In addition, the accessibility and cooperation between respondents and complaint desks are important. The DCA will facilitate this development in 2024.

This survey stressed the need for mutual cooperation. The increase in the use of algorithms in requires more expertise and collaboration. Almost half of the respondents have contact with other respondents about systems and/or applications in the field of work. And 40% of the respondents sometimes have contact with international partners on this issue, often in existing cooperation structures. One of the pillars of the DCA is to further strengthen the cooperation between respondents in this area. For example, the AP organizes workshops and knowledge sessions will evaluate the needs that are present on the basis of this survey.

The majority of respondents have ambitions and plans to do more with algorithms in 2024. A number of respondents indicate that they want to start with research pilots to further explore the opportunities and risks of using algorithms. There is also a strong focus on new legislation, such as the AI Act and its impact on the work of respondents. In response, employees are increasingly offered training and courses to stay up to date on the developments of algorithms and AI. Some of the respondents are also recruiting new people for their work on algorithms and AI.

Respondents also use algorithms and AI for their own work. Three quarters of the respondents in this survey are engaged in expanding the use of algorithms to be able to work more in a more targeted and effective way. For example, systematic checks where less human capacity is needed or the detection of risks and deviations in a sector. Specific trends in sectors can also be collected more quickly. This shows that the world of oversight and supervision is not that different from the rest of society. Of course, respondents must also invest in responsible deployment and adequate control of algorithms. Transparency is part of this, for example by registering algorithms in the Dutch national algorithm register.

Regulatory sandboxes

Regulatory sandboxes can be used for the responsible and safe development of algorithms and AI. These are environments in which new methods or technologies can be tried under the watchful eye of a supervisor. Depending on its purpose, a regulatory sandbox can take different forms. In such a sandbox, a supervisor could for example be able to intervene quickly if risks manifest themselves. A function of the sandbox can be to develop algorithms within the existing laws and regulations, and provide guidance, and also direction to parties on how to comply with the rules. The purpose of a sandbox is to ensure that innovation takes place within the legal frameworks. And in doing so, for example, identifying and answering difficult legal questions, promoting compliance and lowering the threshold to market access. The role of oversight and supervision is mainly to give direction to developers. Knowledge gained can then be shared with the market, so that there is more clarity for all

relevant market participants about how to comply with new regulations. In addition, laws, regulations, guidelines and policies can be tailored to new developments and insights gained in a sandbox.

In the Netherlands, a regulatory sandbox for the AI Act is being prepared. Under the AI Act, member states will have to offer a sandbox. In this, market surveillance authorities guide and advise developers of AI systems, so that they can make their systems compliant. In the Netherlands, a group of market authorities and inspectorates, together with the Ministry of Economic Affairs, are exploring how they can set up the Dutch sandbox for the AI Act.

[1] OECD. (2023). OECD AI Incidents Monitor (AIM). OECD.AI Policy Observatory. https://oecd.ai/en/incidents

[2] Young, S. (2023). Proposed memorandum fort the heads of executive departments and agencies. [Memorandum]. Executive Office of the President – Office of Management and Budget. https://www.whitehouse.gov/wp-content/uploads/2023/11/AI-in-Government-Memo-draft-for-public-review.pdf

[3] The College of Human Rights. (17 November 2023). 2023-111. https://oordelen.mensenrechten.nl/oordeel/2023-111

[4] The College of Human Rights. (1 August 2023). 2023-82. https://oordelen.mensenrechten.nl/oordeel/2023-82

[5] The College of Human Rights. (17 October 2023). Student not discriminated against by exam software Proctorio, but VU should have handled the complaint more carefully [news message]. https://www.mensenrechten.nl/actueel/nieuws/2023/10/17/student-niet-gediscrimineerd-door-tentamensoftware-proctorio-maar-vu-had-de-klacht-zorgvuldiger-moeten-behandelen

[6] Court of Auditors Amsterdam Metropolitan Area. (2023) Algorithms https://www.rekenkamer.amsterdam.nl/onderzoek/algoritmen/

[7] Algorithm Audit. (2023) Advisory document Risk profiling re-examine social assistance (AA:2023:02:A) https://algorithmaudit.eu/nl/algoprudence/

[8] Consumer and Market Authority. (27 June 2023). ACM appeals to webshops that use misleading countdown timers [news release]. https://www.acm.nl/nl/publicaties/acm-spreekt-webshops-aan-die-misleidende-count-downtimers-gebruiken

[9] Audit Service of Rijk. (2023). Research framework Algorithms. https://open.overheid.nl/documenten/61b54381-d331-40ed-8fce-b2883b195f25/file

[10] Deloitte. (2023). Digital Consumer Trends Report 2023. https://www2.deloitte.com/nl/nl/pages/technologie-media-telecom/articles/digital-consumer-trends-23-report.html

[11] Consultancy.nl. (8 November 2023). Dutch consumers are not so positive about generative AI [news message]. https://www.consultancy.nl/nieuws/49601/nederlandse-consument-staat-niet-zo-positief-tegenover-generatieve-ai

[12] Capgemini Research Institute. (2023). Why consumers love generative AI. https://prod.ucwe.capgemini.com/wp-content/uploads/2023/05/Final-Web-Version-Report-Creative-Gen-AI.pdf

[13] Salesforce. (24 July 2023). New Salesforce research reveals increasing pressure on IT teams and a prominent role for generative AI. Emerce. https://www.emerce.nl/wire/nieuw-salesforceonderzoek-onthult-toenemende-druk-itteams-prominente-rol-generatieve-ai

[14] Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J.Q., Demszky, D., Liang, P. (2021). On the Opportunities and Risks of Foundation Models. arXiv. https://doi.org/10.48550/arXiv.2108.07258

[15] Personal Data Authority. (2023). Blog post: concerns about generative AI. https://www.autoriteitpersoonsgegevens.nl/actueel/blogpost-zorgen-om-generatieve-ai

[16] Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J.Q., Demszky, D., Liang, P. (2021). On the Opportunities and Risks of Foundation Models. arXiv. https://doi.org/10.48550/arXiv.2108.07258

[17] Gensler, G. (17 July 2023). "Isaac Newton to AI" Remarks before the National Press Club U.S. Securities and Exchange Commission. https://www.sec.gov/news/speech/gensler-isaac-newton-ai-remarks-07-17-2023

[18] Toups, C., Bommasani, R., Creel, K.A., Bana, S.H., Jurafsky, D., Liang, P. (2023). Ecosystem-level Analysis of Deployed Machine Learning Reveals Homogeneous Outcomes. arXiv. https://doi.org/10.48550/arXiv.2307.05862

[19] Competition & Markets Authority. (2023). AI Foundation Models: Initial Report. https://www.gov.uk/government/publications/ai-foundation-models-initial-report

[20] OECD. (2023). Employment Outlook 2023: Artificial intelligence and jobs. https://oecd.org/employment-outlook/2023/

[21] Boy, E., van den Berge, W., Goos, M., Kerkemezos, Y. (2023). Technology, the labour market and the role of policy. Central Planning Bureau. https://www.cpb.nl/sites/default/files/omnidownload/CPB-Publicatie-Technologie-de-arbeidsmarkt-en-de-rol-van-beleid.pdf

[22] Klijs, B., Smits, W. (2023). Employment. In Monitor online platforms 2022 (hfst.4). Central Statistical Office. https://www.cbs.nl/nl-nl/longread/rapportages/2023/monitor-online-platformen-2022/4-werkgelegenheid

[23] Vervliet, T. (2022). Digital recruitment & selection: algorithm use in employers' recruitment & selection process and awareness of risks of exclusion and discrimination (Report 2022-62). SEO Economic Research. https://publicaties.mensenrechten.nl/publicatie/546c77bc-4db8-407e-ae52-41ea26510c68

[24] https://ec.europa.eu/commission/presscorner/detail/en/speech_23_5366

[25] Riso, S., Adăscăliței, D., Rodriguez Contreras, R. (2023). Ethical digitalisation at work: From theory to practice. Publications Office of the European Union. https://www.eurofound.europa.eu/en/publications/2023/ethical-digitalisation-work-theory-practice

[26] Baiocco, S., Fernandez-Macías, E., Rani, U., Pesole, A. (2022). The Algorithmic Management of work and its implications in different contexts. International Labour Organisation, European Commission. https://www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_849220.pdf

[27] von der Leyen, U. (13 September 2023). 2023 State of the Union Address by President von der Leyen [Speech]. https://ec.europa.eu/commission/presscorner/detail/en/speech_23_4426

[28] Harris, L.A., Jaikaran, C. (2023). Highlights of the 2023 Executive Order on Artificial Intelligence for Congress (R47843). Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R47843

[29] Smith, C. (10 February 2022). Using AI-Based Tools to Build Empathy in the Contact Center [Blog]. Genesys. https://www.genesys.com/blog/post/using-ai-based-tools-to-build-empathy-in-the-contact-center

[30] OECD. (2023). Employment Outlook 2023: Artificial intelligence and jobs. https://oecd.org/employment-outlook/2023/

[31] Boy, E., van den Berge, W., Goos, M., Kerkemezos, Y. (2023). Technology, the labour market and the role of policy. Central Planning Bureau. https://www.cpb.nl/sites/default/files/omnidownload/CPB-Publicatie-Technologie-de-arbeidsmarkt-en-de-rol-van-beleid.pdf

[32] Inspection of Education. (z.d.). Measures taken by so-schools to adapt education during the coronavirus crisis. https://www.onderwijsinspectie.nl/onderwerpen/corona-onderzoeken/gevolgen-van-16-maanden-corona-voor-het-onderwijs/gevolgen-speciaal-onderwijs/maatregel-en-scholen-voor-aanpassen-van-onderwijs

[33] National Growth Fund. (s.d.) National Education Lab. https://www.nationaalgroeifonds.nl/over-zicht-lopende-projecten/thema-onderwijs/nationaal-onderwijslab

[34] Radboud University. (z.d.). About the National Education Lab AI. https://www.ru.nl/nolai/over-nolai

[35] Moerland, S., Elings, M. (21 October 2021). Extra money for primary education mainly used for attracting teachers. NOS News. https://nos.nl/artikel/2402556-extra-geld-voor-basisonderwijs-vooral-gebruikt-voor-aantrekken-leerkracht

[36] Karssen, M., Krepel, A., Stronkhorst, E., Lourens, J.M.P., Bruck, S., Van Kessel, M. & Saab, N. (2023). Monitor digitization primary and secondary education (report 1102). Kohnstamm Instituut. https://kohnstamminstituut.nl/rapport/monitor-digitalisering-primair-en-voortgezet-onderwijs/

[37] Scharwächter, V. (2023). ChatGPT: Overview of AI guidelines for each educational institution. In Smarter Education with AI, September 2023, 62-63. NPuls. https://npuls.nl/actueel/npuls-presenteert-het-magazine-slimmer-onderwijs-met-ai/

[38] Stokke, L., Gallagher, E. (29 March 2023). ChatGPT is here, deal with it. UKrant. https://ukrant.nl/magazine/chatgpt-is-er-dus-deal-ermee-studenten-willen-regels-voor-ai-gebruik-blijkt-uit-enquete-ukrant/#:~:text=Toch%20lijken%20studenten%20nog%20zo,complete%20opdrachten%20hebben%20laten%20maken

[39] The Education Council. (2022). Use of Intelligent Technology https://www.onderwijsraad.nl/publicaties/adviezen/2022/09/28/inzet-van-intelligente-technologie

[40] Dijkstra, P. (14 June 2022). Experimental environment learning analytics: get started with study data with an eye for privacy. Surf Communities. https://communities.surf.nl/learning-analytics/artikel/experimenteeromgeving-learning-analytics-aan-de-slag-met-studiedata-met

[41] Utrecht University. (z.d.). La in education. https://www.uu.nl/onderwijs/learning-analytics/la-in-het-onderwijs

[42] Smaling, E., (16 November 2023). Almost everyone uses ChatGPT, but plaguing students goes too far. Erasmus Magazine. https://www.erasmusmagazine.nl/2023/11/16/bijna-iedereen-gebruikt-chatgpt-maar-plagieren-gaat-studenten-te-ver/#:~:text=Van%20de%20154%20deelnemers%20aan,te%20zetten%20tijdens%20hun%20opleiding

[43] Commission IMPLEMENTING DECISION on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence, C(2023)3215. (22 May 2023). Directorate-General for Communications Networks, Content and Technology. https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=nl

[44] X. (2023). DSA Transparency Report. https://transparency.twitter.com/dsa-transparency-report.html

AUTORITEIT
PERSOONSGEGEVENS