The government-wide vision on
Generative AI
of the Netherlands

GENERATIVE AI

Table of contents

1 Introduction

Artificial intelligence (AI), as a systems technology, has a major impact on all domains and sectors of our society. It affects all government policies.[1] The increasing prevalence of generative AI applications has led to a greater integration of AI into the daily lives of many individuals in the Netherlands, both in personal and professional contexts.

Generative AI is not only used by professionals such as data analysts, advertisers and journalists. It is also being used to write a poem or to create a customised weekly menu. Generative AI can be used in many ways.

Generative AI is a form of AI that uses algorithms to generate content.[2] Users can generate text, images, sound, or computer code with a simple prompt. Since the launch of ChatGPT in late November 2022, the number of users rose to over 100 million worldwide within two months. Around 180 million users were active by the end of 2023. It has been estimated that over one and a half million people in the Netherlands are currently utilising generative AI.[3] Apart from ChatGPT, there are a number of other generative AI applications that are publicly available, such as Midjourney, DALL-E, and Google Bard. This involves the use of so-called large language models (LLMs).[4]

Generative AI can be considered a powerful extension of human analytical and creative abilities. When coupled with related technologies, it has great potential to address societal and scientific issues. It can also increase labour productivity. While there are opportunities, it is important to acknowledge that there are also significant risks involved. The achievement of various public values and fundamental rights may be at risk. For instance, generative AI can be used to create and disseminate misinformation; it can reinforce discriminatory dynamics; and it can reinforce socio-economic inequalities.

The government has an important responsibility to steer the development, application and embedding of generative AI in the right direction. The use of generative AI has become widespread due to its scalability, rapid development, and low threshold for usage. It is therefore necessary to formulate a vision for this technology and to link concrete actions to it.

---

1.   *See also the 2021 report Opgave AI by the Scientific Council for Government Policy (Dutch: Wetenschappelijke Raad voor het Regeringsbeleid, WRR).*
2.   *A type of AI that uses complex algorithms to generate new content such as text, images, computer code or videos. The best-known example of this is the ChatGPT chatbot.*
3.   *https://autoriteitpersoonsgegevens.nl/actueel/ap-vraagt-om-opheldering-over-chatgpt*
4.   *An LLM is a type of AI that is trained on large amounts of text to understand existing content and generate new content. It is considered a specialised form of generative AI.*

The government intends to explore **the potential of genera-tive AI to improve human wellbeing and promote autonomy, sustainability, prosperity, justice, and security**. By focusing specifically on responsible applications of generative AI, we are seizing the opportunities presented by this technology. We do this for all sectors. By taking a responsible and open approach to the application of generative AI, society as a whole can benefit. The aim is to create a strong AI ecosystem in the Netherlands and the EU, in which many responsible generative AIs can be innovated. The government intends to create a framework **for the responsible development and use of generative AI**, while preserving our digital, open strategic autonomy.

The government recognises that the impact of generative AI - and AI in general - depends on an interplay of technological, economic, institutional and societal factors. The government therefore stresses the importance of continuing to monitor and analyse the developments and impact of generative AI. In order to ensure the proper social embedding of generative AI, it is essential to take technological developments into account at an early stage and to adopt a learning and evaluation approach. We are committed to doing this together with all stakeholders in the Netherlands, but also in an international context.

The government recognises the added value of generative AI, provided it is developed and applied responsibly. The AI Act is an important legal framework for this. Through responsible experimentation and a learning approach, we can use generative AI innovatively in the Netherlands while exploring its potential. The government collaborates closely with companies that have the necessary knowledge and skills. The government, in collaboration with the industry, intends to examine the potential advantages of generative AI in addressing societal issues, such as the energy transition. The government

will intensify its dialogue with the business community on this matter.

With this government-wide vision, the government emphasises the importance of taking action on the matter, both in the short and long term. In doing so, this vision aligns with the ambitions of the Value-Driven Digitalisation Work Agenda.[5] In particular, the government is focusing on the impact of this new digital technology on society. This is closely related to the Digital Economy Strategy, in particular in terms of creating the right framework conditions for well-functioning digital markets and services, stimulating digital innovation, and strengthening cybersecurity.[6] The presentation of this vision also fulfils the Dekker-Abdulaziz and Rajkowski motion, which your House of Representatives of the Netherlands adopted by a large majority in April 2023.[7]

This vision is the result of numerous meetings and discussions in different fields and sectors such as healthcare, the labour market, education and public administration. These discussions will continue even after the publication of this vision. This involved actively seeking collaboration with departments, implementing organisations, subnational governments, knowledge and higher education institutions, developers and citizens.[8]

a Delineation

This vision focuses specifically on generative AI. Unlike task-specific AI systems, such as those used for facial recognition on a smartphone, generative AI is capable of creating content on its own. In addition, some generative AI systems (including systems with underlying large language models (LLMs)) can be used to perform a wide range of tasks. Furthermore, online tools such as ChatGPT and integration with search engines or applications such as Microsoft Office make

generative AI available to a wider audience. The development of generative AI is far from stagnant. The next generation of generative AI systems will likely be able to handle multiple modalities simultaneously and be much more capable than the products currently on the market.

b Reading guide

This vision first addresses the question of what technology is involved, and then outlines the expected short- and long-term technological developments. It then considers the (social) impact of generative AI. This is followed by an outline of existing policies and regulations, which will provide the framework within which the government's vision for generative AI will be presented. The national, European and international contexts are discussed here. In order to ensure that citizens and businesses in the Netherlands and Europe can reap the full benefits of this technology, while being protected from its excesses, the government sets out four principles in this report. These principles are linked to actions to realise this vision over the coming years.

5. See line 2.3 of the Work Agenda 'Anticiperen op nieuwe digitale technologie': https://www.rijksoverheid.nl/documenten/rapporten/2022/11/04/bijlage-1-werkagenda-waardengedreven-digitaliseren
6. https://www.rijksoverheid.nl/documenten/kamerstukken/2022/11/18/strategie-digitale-economie
7. Amended motion by members of the House of Representatives Dekker-Abdulaziz and Rajkowski of 4 April 2023
'Integrale visie op nieuwe AI-producten' (Parliamentary papers 2022/23 26 643, no 1003).
8. Appendix 1 shows exactly how this open approach was designed and the lessons learned.

2 Generative AI

This chapter explores what generative AI means from a technological standpoint within this vision, and how it will (potentially) evolve in the coming years. Finally, it discusses the playing field for this technology.

a What is generative AI?

Generative AI is a form of AI[1] capable of generating content such as text, audio, images, computer code and videos. The distinction between content created by generative AI and content created by humans is not always immediately apparent to humans.

One of the most recognisable applications of generative AI are AI chatbots. These digital assistants can communicate via text in a way that closely resembles human interaction. Well-known examples are ChatGPT and Google Bard, both AI chatbots using LLMs. The strength and growing success of these models lies in their versatility, from writing (computer) code to playing board games.

Other generative AI systems can generate images or audio such as OpenAI's DALL-E 3 and Google's MusicLM. AI-generated images of people, i.e. not real people, are becoming increasingly common in advertisements and websites. Generative AI-generated audio will be particularly prevalent from 2023 onwards. Examples include generating audio based on the music of existing artists or deployment in healthcare. It allows people with ALS, for example, to continue to communicate in their own voice

The creation of generative AI models consists of three phases: pre-training, fine-tuning and deployment. In the pre-training phase, the model is fed with large amounts of data (such as text, audio or images) from different sources. The model is capable of recognising patterns in the data during pre-training. This task requires a considerable amount of computational resources and is performed on specialised hardware. The pre-training phase is followed by the fine-tuning phase. In this phase, the model is trained to follow the user's instructions, any expertise is added, and the model can be trained to give socially acceptable responses. This involves special techniques, such as Reinforcement Learning from Human Feedback (RLHF).[2] The fine-tuning phase is followed by the deployment phase, in which the model is made available to users.[3] The model can be duplicated and then used by tens of thousands of users simultaneously via a consumer interface. More infor-

---

1. We follow the OECD's recently revised definition of "AI system" (2023): a machine-based system that derives, for explicit or implicit purposes, from the inputs it receives, how to generate outputs such as predictions, content, recommendations or decisions that can affect physical or virtual environments. Different AI systems vary in their degree of autonomy and adaptability after their implementation/deployment.
2. In the case of RLHF, human feedback is incorporated into the training process of AI algorithms to guide or improve the AI algorithm's learning. It is suggested that this feedback could potentially aid the algorithm in learning at a faster and more effective pace. The aim is often to use human expertise to steer AI algorithms in a particular desired direction.
3. Generative AI attracts a diverse range of users, varying in expertise and objectives.

mation on the technical development process of generative AI can be found in Appendix 2.

b Trends

Five trends can be identified in the developments around generative AI:

1. *Models are becoming more skilled and applicable.* This includes both honing existing skills and developing new ones. For example, current models can assist a user with programming tasks, whereas the previous generation of models were barely capable of doing so.

2. *Models are increasingly being fitted with 'guard rails'. Guardrails* are safety measures that govern the interaction between an AI model and a user, and which can be used as a basis for monitoring. However, there is still a long way to go in this area. For example, current models regularly produce incorrect results ('hallucinations'). And in some cases, security measures and ethical frameworks can also be circumvented relatively easily.

3. *Models are becoming multimodal.* While AI models could initially only handle text, audio or images, new AI systems have been developed in the past year that can handle these forms of content simultaneously.

4. *Models are becoming (more) independent.* New AI systems are able to autonomously connect digital tools and then use them independently in a sequence. Data collection and task planning and execution can also be done independently.

5. *Models are becoming more cost-effective.* While larger AI models have become much more capable, they also require much more computing power. For this reason, active efforts are being made to create more compact, affordable and faster models without significant loss of performance.

Monitoring developments

We publish an annual Generative AI Monitor to (continue to) track the development and use of generative AI for and by governments.

c Speed of development

Even after becoming available to the general public in late 2022, developments in generative AI will continue at a rapid pace. The computing power used to train generative AI models is increasing by a factor of four every year. In addition, the algorithmic efficiency of AI models is increasing by a factor of 2.5 per year.[4] This stacked exponential growth has led to much more proficient generative AI systems in recent years. With a high degree of autonomy, they are now able to automate complex processes, perform sophisticated data analysis or show a user how to repair their bike based on a photo. Generative AI systems can also be connected to the internet and given instructions to perform all sorts of actions, such as booking an airline ticket. Improving AI capabilities relies heavily on scaling. Developers can train a better model simply by using more AI chips and more data. This increase in scale is expected to continue in the coming years. Generative AI models will thus become much more proficient.

The emergence of increasingly capable generative AI systems has raised the question of whether this means we are heading towards *artificial general intelligence* (AGI). AGI refers to technology that exhibits intelligence across a wide range of domains, performing at or above human levels with these capabilities.[5] To date, there is no scientific consensus that we could speak of AGI.[6] What is certain, however, is that major geopolitical powers and tech companies are investing significant sums in the development of advanced AI systems. This has led to competition between countries and companies claiming to be developing AGI. EU and Dutch companies are not actively involved at the moment because they do not have the resources to compete with these players

4. *See also*: AI Trends – Epoch (epochai.org)
5. *Taecharungroj, V. (2023). "What Can ChatGPT Do?" Analysing Early Reactions to the Innovative AI Chatbot on Twitter. Big Data and Cognitive Computing, 7(1), 35.*
6. *Tredinnick, L., & Laybats, C. (2023). The dangers of generative artificial intelligence. Business Information Review.*

d Playing field

It is important to note that the development of generative AI models often requires a substantial investment in computing infrastructure. This has pushed the development of generative AI towards commercial parties. AI labs in the US, with the support of their cloud providers, are currently leading the way in the development of cutting-edge generative AI models. Mainly because they have the computing power, talent and data needed to train and develop generative AI. The current dynamic of winner-takes-all may reinforce the dominance of these companies. As a result, there is a growing dependence of European organisations and citizens on a limited number of generative AI developers.[7] The high investment required to purchase sufficient computing power and the lack of attractive business models make it almost impossible for Dutch organisations to enter this dynamic. In other member states, some companies are already training new generative AI models, but even these players lag behind major US and Chinese compet-

itors. To gain a foothold in this market, European cooperation is necessary.

Not all AI labs market their models in the same way. Most companies only distribute their models via an API or a consumer product. Other companies, such as Meta, deliberately choose to make the parameters of their AI models public. This allows users to fine-tune the model, making it more flexible to use. A disadvantage of this approach is that any security measures - e.g. against racism or illegal use - can easily be removed from the model. The development of generative AI models relies heavily on a concentrated hardware chain. More than 75% of all state-of-the-art AI chips are designed by US-based NVIDIA and manufactured at TSMC in Taiwan. This uses lithography machines, almost all of which are made in the Netherlands. The Netherlands is home to a strong semiconductor ecosystem. This gives it a unique position in the AI development chain.

The establishment an AI advisory council at the highest level

The establishment of an AI Advisory Council, (or Rapid Response Team AI), is being considered at the highest level to provide the government with short- and long-term advice.

High Performance Computing

The Netherlands participates in the EuroHPC partnership under Horizon Europe in the field of high-performance computing (HPC), enabling Dutch companies and knowledge institutions to participate in European projects on HPC and quantum computing. HPC makes a significant contribution to complex problems by enabling complex calculations to be performed at high speed.

---

7. *See also Agenda DOSA:* Agenda Digitale Open Strategische Autonomie | Report | Rijksoverheid.nl

3 The impact of generative AI

The development of generative AI, as outlined in Chapter 2, will intensify in the coming years and is likely to have a major impact on people, society, work and the economy. Below, the expected impact on (Dutch) society is divided into opportunities and possibilities on the one hand and risks and challenges on the other.

The impact of generative AI will have to manifest itself further. This could take years. The ultimate impact of generative AI depends on an interplay of technological, economic, institutional and societal factors. The government therefore emphasises the importance of monitoring and analysing the developments and consequences of generative AI (see also under 'actions' in Chapter 5). This chapter discusses the expected opportunities and risks, based on initial scientific findings and expert predictions.

The impact of generative AI can be both positive and negative. For example, while generative AI offers opportunities when it comes to generating information, it can also lead to disinformation and inimitability. So the risks mentioned below are mostly 'the other side of the coin'. Whether a given capability of generative AI ultimately manifests as an opportunity or a risk depends on the specific development, the application of the technology and the intentions or expertise of the user. It is therefore important that generative AI is properly guided and supervised. Chapters 4 and 5 discuss the details of this. This chapter aims to provide an overview of the opportunities and risks posed by generative AI in different social spheres and sectors. Some of the impacts mentioned arise not so much from generative AI models *themselves*, but from the use and social acceptance (or non-acceptance) of the technology.

a Opportunities and possibilities

Generative AI opens up a wide range of possibilities, such as creating conversational transcripts, composing music, synthesising images, and discovering and designing new molecules and materials. In addition, there is still plenty of research being done on the exact opportunities offered by generative AI and how to make the most of them. It is clear that, compared to smaller and more specialised AI models, a new generation of generative AI models can be used as 'base models' in different domains for a variety of general purposes. Generative AI models can therefore be used in numerous sectors

and domains to optimise **processes, automate** and assist in tasks such as collecting, summarising and elaborating (large amounts of) information or writing computer code. Generative AI has the potential to improve efficiency, reduce costs, assist in decision-making, enhance service delivery, and offer innovative solutions.

Generative AI can perform tasks in various roles, such as **production, learning** and **problem solving**, or a combination of these. This presents opportunities for individuals, businesses, government and society as a whole. Given the potential impact of generative AI models and systems on society and the economy, the government is committed to promoting responsible experimentation and use across various sectors and domains.

Generative AI as a production tool

Generative AI creates opportunities for producing all kinds of digital content, such as answering questions, summarising texts, creating videos and writing. This is already affecting the daily lives of individual citizens. Many use AI chatbots like ChatGPT for all kinds of tasks in their (personal) lives, such as for coming up with a recipe, personal sports training schedules or for generating poems and cover letters.

For businesses and organisations, generative AI as a production tool also offers **promising opportunities** for the efficiency and quality of all kinds of **business processes.** For example, generative AI is already being used to support and speed up administrative processes, assist customer service, write computer code and automate industrial processes.

In various industrial processes, generative AI can quickly produce a large number of design alternatives. In the **manufac-** turing industry, for example, the design process in mechanical engineering can be significantly accelerated and the maintenance of complex machines can be carried out more easily and more cheaply. In addition, generative AI models can help make proactive decisions and reduce costs associated with overproduction or inventory shortages by simulating different production scenarios based on predicted customer demand.

In the **cultural sector**, generative AI is used to support the creative process, for example in scriptwriting by film scriptwriters or in generating descriptions of artworks. Generative AI is also used to create simulations for training purposes, working out future scenarios, or producing a virtual representation of a product or process (known as *digital twins*). For the **music industry** generative AI can act as a catalyst for creativity and innovation. In this area, it is able to generate new compositions, creating new melodies, harmonies and rhythms that can inspire or support musicians in their work.

Due to the capacity of generative AI to produce computer code through the use of 'prompts', this technology presents numerous possibilities within the **ICT industry**. Generative AI has great potential to improve software and application programming and support IT professionals in their work. This can significantly speed up the development process and leave IT professionals with more time for other tasks.

Generative AI can lead to overall **productivity growth.**[1] Historically, productivity growth has led to an increase in **material prosperity**, better health and more leisure time. According to economists, generative AI is predicted to have a productivity effect[2] not only for large companies but also for small and medium-sized enterprises (SMEs) due to its low-threshold

applications.[3] Generative AI can also make certain tasks (such as performing financial analysis and legal processes) available to *in-house*[4] staff of SMEs in a cost-effective way.[5] Productivity and wealth growth, as well as the creation of **new jobs** resulting from the deployment of generative AI, are potential outcomes.[6] A direct effect is the emergence of new professional groups (such as information specialists and ICT practitioners) with the skills to deploy generative AI applications.[7] There will also be a demand for workers with the necessary competences to responsibly deploy AI to support their work. An indirect effect on employment may arise from the potential growth of disposable income in the economy. AI-induced productivity growth can lead to income growth, which in turn can increase demand for goods and services and result in employment growth. Most economists do not anticipate a long-term

Innovation labs with SMEs

Innovation Labs will be launched from AiNed in 2024. InnovationLabs are partnerships between public and private entities that aim to develop AI innovations, with a particular focus on supporting SMEs, start-ups, and scale-ups.

1. https://www.technologyreview.com/2021/06/10/1026008/the-coming-productivity-boom/
2. https://www.kentclarkcenter.org/surveys/ai-and-productivity-growth/
3. Mills, K. (2019). How AI could help small business. Harvard Business Review.
4. For example, through Artificial Intelligence as a Service (AIaaS).
5. OECD (3 February 2021) the Digital Transformation of SMEs. Chapter 5. Artificial Intelligence, changing landscape for SMEs.
6. https://www.wsj.com/tech/ai/the-new-jobs-for-humans-in-the-ai-era-db7d8acd
7. OECD (2023). OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market.

decline in the total number of jobs in the economy due to AI-driven automation.[8] However, there may be distributional issues, which are discussed as a potential 'challenge or risk' in section 3b. It is important to note that specific preconditions must be met for productivity growth, including knowledge building in organisations.

As a production tool, generative AI can have a positive impact on the nature of work performed by humans. For instance, the implementation of AI can automate mundane tasks, such as transcribing meetings, transcribing audio, or responding to common inquiries, freeing up employees' time. There are indications that AI technology is assisting workers with less knowledge and experience to keep pace with their more experienced counterparts.[9] This increases the sense of professional autonomy and competence. The factors mentioned above may increase the (perceived) **quality of work**. The opposite can also occur, as explained in more detail below under the section titled 'challenges'.

Government can also benefit from generative AI as a production tool. It provides the public sector with opportunities to enhance processes, improve overall government functioning, and **optimise services to citizens**. For instance, by improving communication with inhabitants. Generative AI also has the potential to make government information more accessible to everyone by providing language level adjustments. In this way, technology can contribute to clear and inclusive communication with citizens. Generative AI has the potential to enhance the efficiency of legal and administrative processes by automating forms, as seen in 'Legal Tech', while still allowing for customisation. This is conditional on the technology being used ethically and properly regulated.

Generative AI can play a role in data-driven policy making and evaluation by rapidly analysing large amounts of data, generating training materials, and simulating policy scenarios. It also offers opportunities for internal knowledge development.

Generative AI as a learning tool

Generative AI models can be used to analyse huge amounts of data quickly. Generative AI can therefore be used, for example, to explain complex texts, interpret key topics and draw conclusions. Users can therefore also use generative AI as a learning tool.

We see this role, for example, in the field of **language and translation**, Generative AI models are capable of translating large amounts of text with high accuracy.[10] The technology can be used to translate content, making relevant information available to a wider audience. This can be helpful for language learning and for translating websites or educational materials, including those related to government.

Generative AI can therefore play an important role as a **learning tool** in **education.** For instance, this technology can assist students in generating summaries, clarifying learning material, and composing practice questions. Generative AI models can create personalised feedback, recommendations, and interventions by analysing users' learning patterns to tailor teaching to personal learning needs.[11] Teachers can use generative AI to design teaching methods or improve teaching materials. Furthermore, generative AI enables the prediction of future student performance based on past student data, which can help identify students who may require additional support.[12]

The function of a learning tool is exemplified by the use of digital search engines, which are utilised by the majority of

Generative AI pilots at and with public sector bodies

Pilots within the public sector are being carried out to test how generative AI can be used responsibly and safely, for example in the area of proactive service delivery.

individuals in their daily lives. Generative AI has already been integrated into search engines by companies like Google and Microsoft, significantly improving their functionality.

Finally, generative AI applications like ChatGPT are capable of explaining concepts in a variety of ways. The tool can be utilised for academic or professional purposes, as well as for everyday topics such as explaining the operation of a fuse box or how to save energy. Due to the interactive nature of applications like ChatGPT, users can request customised or more detailed explanations.

Generative AI as a problem solver

Generative AI can be a valuable tool for **problem solving**. This is evident in the **scientific domain**, particularly in the development of new drugs. The drug development process is often long, complex and costly. Generative AI has shown promising

8.   *In the long term, it can be argued that the creation of new jobs and displacement of old ones balance each other out as supply and demand reach equilibrium through price adjustments. See: David H. Author (2015). 'Why Are There Still So Many Jobs? The History and Future of Workplace Automation.' Journal of Economic Perspectives 29(3): pp. 3-30.*
9.   *Brynjolfsson, Li & Raymond (2023). 'Generative AI at Work.' National Bureau of Economic Research. Working paper no. 31161.*
10. *Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of AI, 7(1), 52-62.*
11. *Abunaseer, H. The Use of Generative AI in Education: Applications, and Impact. Technology and the Curriculum: Summer 2023.*
12. *Wang, T., Lund, B. D., Marengo, A., Pagano, A., Mannuru, N. R., Teel, Z. A., & Pange, J. (2023). Exploring the Potential Impact of Artificial Intelligence (AI) on International Students in Higher Education: Generative AI, Chatbots, Analytics, and International Student Success. Applied Sciences, 13(11), 6716.*

results in speeding up and improving the drug development process.[13] This allows for time-saving and potential cost reduction. These benefits are not limited to drug development and extend to other areas of science.

Generative AI also offers opportunities for developing and improving **materials**. Research into new materials for batteries or microchips, for example, can take months, if not years. Generative AI models can generate new chemicals, molecules, and materials much faster than humans, contributing to a more efficient process.[14]

The role of problem solver is already being deployed in **healthcare**. For example, generative AI is being experimented with as a consultant in cancer treatment.[15]There are opportunities to analyse data within clinical trials to predict which patients will benefit from a new treatment. In this way, generative AI may eventually play a role in reducing late or incorrect diagnoses. In addition, generative AI in healthcare can be used to perform repetitive and administrative tasks, such as summarising patient conversations and completing patient records, among others. Medical professionals can dedicate more time to substantive work. This approach can alleviate the burden on healthcare systems and enhance the quality of care.

Generative AI can contribute as a problem solver to address major societal issues. Despite concerns about the energy consumption of generative AI technology, this is offset by the potential contributions that generative AI can make to the sustainability transition. For example, generative AI can be used to analyse natural ecosystems or predict climate trends[16]. In addition, there are already generative AI applications that allow maritime companies to monitor their emissions or generate operational strategies for sustainable industries.[17] The role of (generative) AI as a problem solver for societal challenges can also be seen in the **military domain**. Examples include modelling and simulation (*wargaming*)[18] and deployment in operational-tactical planning through accessible big data analytics.[19] There are also opportunities in the **cybersecurity domain**. AI applications, for example, allow organisations to automatically detect attacks via detected anomalies in their network. According to the Cyber Security Council (CSR) in Autumn 2023, Generative AI can generate analytics automatically, enabling actions to be taken based on the data.[20]

b Challenges and risks

Generative AI presents both opportunities and risks, with the latter often arising from the former's potential applications. Below, we distinguish between the impacts on individual citizens, market design, labour and income, and society as a whole.

Impact on individual citizens

There are risks associated with using generative AI. The initial challenge is that **discriminatory dynamics** can be amplified due to existing **bias** (bias or selectivity embedded in training data and model parameters[21]).[22] This bias may be reinforced by the fact that widely used AI models from major developers are created by a select group of individuals with often one-sided perspectives. Bias has negative consequences for the social recognition and representation of individuals who use or are influenced by generative AI. **Equal treatment and non-discrimination** are thus under pressure. The lack of transparency, explainability, and complexity of AI models can conceal bias and discriminatory effects for extended periods.

A second challenge concerns the possible violations of **rights** on **privacy**, **data protection** and **copyright and related rights.** As a result, training data, mostly obtained through large-scale (web)scraping[23] from public sources on the internet or other digital sources, may contain (special) personal data.[24] There is often a lack of transparency on what data is used and how. The content generated may be inaccurate, outdated, incorrect, inappropriate, or offensive and may also take on

13. *Bilodeau, C., Jin, W., Jaakkola, T., Barzilay, R., & Jensen, K. F. (2022). Generative models for molecular discovery: Recent advances and challenges. Wiley Interdisciplinary Reviews: Computational Molecular Science, 12(5), e1608.*
14. *Liu, Y., Yang, Z., Yu, Z., Liu, Z., Liu, D., Lin, H., ... & Shi, S. (2023). Generative artificial intelligence and its applications in materials science: Current situation and future perspectives. Journal of Materiomics.*
15. *Sorin, V., Klang, E., Sklair-Levy, M., Cohen, I., Zippel, D. B., Balint Lahat, N., ... & Barash, Y. (2023). Large language model (ChatGPT) as a support tool for breast tumour board. NPJ Breast Cancer, 9(1), 44.*
16. https://www.abnamro.nl/nl/media/rapport-generatieve-ai-pakt-rol-in-de-duurzaamheidstransitie-december-2023_tcm16-216530.pdf
17. https://bearing.ai/
18. https://magazines.defensie.nl/defensiekrant/2019/23/06_wargaming_23
19. https://open.overheid.nl/documenten/d49f42ca-181b-4e2f-9986-b412de40f2f5/file
20. *See also:* https://www.cybersecurityraad.nl/actueel/nieuws/2023/12/22/csr-brief-over-ai-en-cybersecurity
21. ChatGPT Replicates Gender Bias in Recommendation Letters | Scientific American
22. Humans Absorb Bias from AI--And Keep It after They Stop Using the Algorithm - Scientific American
23. *Webscraping is the use of software to extract information from web pages for subsequent analysis.*
24. *See also: Roundtable of G7 Data Protection and Privacy Authorities Statement on Generative AI (21 June 2023), online via:* Roundtable of G7 Data Protection and Privacy Authorities Statement on Generative AI -Personal Information Protection Commission-(ppc.go.jp).

Opportunities and possibilities

Challenges and risks

Generative AI as a
Production tool

Efficiency and quality business processes

Creative process

IT

Material prosperity

New Jobs

Quality of work

Government functioning

Legal tech

Generative AI as a
Learning tool

Language and translation

Education

Search engines

Interactive support

Generative AI as a
Problem solver

Scientific domain
Drug development

Materials like batteries

Healthcare

Societal issues, sustainability

Military domain

Cybersecurity

Impact on individual citizens

Bias/discrimination

Privacy, data protection, user autonomy

Cognitive development
Social development

Copyright, neighbouring and database rights, portrait rights

Dependence and market power

Increasing dependence on US Tech companies
Strategic dependencies

Concentration of power
Entry barriers

Labour and labour market

Employment, Income distribution Wage decline

Quality of work

Income distribution
Job security
Polarisation labour market

Impact on society

Superstar firms
Increasing social and economic inequality

Substantial energy demand
Climate change

Degradation of information ecosystem
Mis- and disinformation

Uncertain reliability automation bias

Military security
Systemic security

Misuse and abuse
Hate speech

a life of its own.[25] According to the Rathenau Institute, it has been observed that generative AI has the potential to extract highly personal information, such as an individual's mood or thoughts, from their interactions with the system.[26] This issue raises concerns about the potential for unwanted control and manipulation through the use of hyper-personalised content and *dark patterns*, which may exploit our desires and unconscious cognitive processes. This could be a potential curtailment of **user autonomy**.[27]

Encouraging the development of language models for languages such as Frisian and Papiamento

We actively encourage the development and improvement of (open and public) language models trained on languages such as Frisian, Papiamento or sign language.

Another challenge is the impact of generic AI systems on the **cognitive development** of people using these systems.[28] Concerns have been expressed about the loss of people's knowledge and skills,[29] particularly in the areas of creativity, critical reflection, and understanding. With the increasing prevalence of generative AI, there is a risk that individuals may lose cognitive skills. Also, **social development** may be negatively affected if generative AI systems increasingly replace (intimate) human interactions.[30]

Finally, when scraping to train a generative AI model, it may be necessary to consider **copyright, neighbouring and database rights, as well as other materials and databases**. It is important to note that the output of generative AI may infringe both the **portrait right** and the aforementioned rights.[31] For example, several lawsuits have already been filed in the US regarding copyright and generative AI.[32] Chapter 4 outlines the current and future legal frameworks, including privacy and data protection legislation, the Constitution, copyright law, and the upcoming European AI Act.

Dependence and market power
There is an **increasing dependence** on a limited number of technology companies. The Netherlands predominantly use generative AI models and services from a limited number of **US tech companies**. These companies have significant amounts

Responsible generative AI tools through the 'Rijks AI-validatieteam' (Government AI Validation Team)

The Government AI Validation Team develops (publicly available) guardrails and tools for generative AI models. To gain further knowledge and experience in AI validation, , the government has established an AI Validation Team . The team is currently exploring the measurability of risks and opportunities associated with generative AI, among other topics. The team consists of software engineers who will work together with policymakers to develop practical tools for validating (generative) AI.

25. *Appendix 'Handelingen II' 2022/23, nr. 3381.*
26. *Rathenau Institute (2023) Generatieve AI: p. 21.*
27. https://www.wired.com/story/ai-chatbots-can-guess-your-personal-information/
28. *Rathenau Institute (2023), Generative AI: pp. 24-25.*
29. *J. Pitt (2023), "ChatSh*t and Other Conversations (That We Should Be Having, But Mostly Are Not), IEEE Technology and Society Magazine, vol. 42, no. 3, pp. 7-13.*
30. *Danaher, J. (2019) The rise of the robots and the crisis of moral patience. AI & Society 34, 129–136.*
Rathenau Institute (2023). Generative AI: p. 25.
31. *See also:* https://open.overheid.nl/documenten/dpc-c82f1b6b5ce7c6826069b7b8579835360ab041ea/pdf
32. https://www.nytimes.com/2023/08/21/arts/design/copyright-ai-artwork.htm

of data, computing power, and development capacity.[33] An increasing share of the global market for generative AI models also lies outside Europe, **mainly in China**. Therefore, the companies involved are in a better position than European companies to develop generative AI models, which will largely determine the direction in which the technology develops. Given the importance of generative AI for the innovation strength and long-term earning capacity of the Netherlands, this may lead to **strategic dependencies**.[34]

The development of generative AI reinforces the **concentration of power** in digital markets, thereby increasing the risk of power abuse. Economies of scale are gaining importance, particularly in relation to the role of data in the development of generative AI. This reinforces the *winner-takes-all-* dynamic.[35] This is especially true for the market of models used as the technological basis for applications of generative AI.[36] In this market, a select group of tech companies are in competition with each other. There is also integration of generative AI with various (existing) services (ecosystem formation). For other companies, this **increases** the **entry barriers** to compete on this market too. Productive market forces can be hindered when innovative new entrants are given less of a chance and developers of specific applications become trapped. For instance, this could result in **unfair trade practices**, increased prices for accessing and using AI applications and infrastructure, and limited options for consumers.[37]

The development of generative AI depends on a handful of large companies. This can result in unequal access to technology and disproportionate opportunities to take advantage of it. Smaller companies, educational institutions, teachers, students, and socio-economically vulnerable groups may be disadvantaged, which can increase social inequalities within a society as well as between societies worldwide. Finally, market power is often the prelude to exerting social and political influence.[38]

Labour and labour market

Regarding work, there are concerns about the impact of generative AI on **employment**, **the quality of work** and **income distribution.**

Most economists do not anticipate a decrease in overall **employment** due to generative AI. Yet the impact on employment can be uneven and disruptive in the short term. The low availability and scalability of generative AI will significantly influence the speed of implementation of this automation. Professions and sectors that are susceptible to (generative) AI, such as creative work, data analysis, legal work, and office support[39], may experience changes in tasks and potential job displacement. As retraining and job searching take time, this may lead to a short-term **rise** in **unemployment**. If jobs disappear quickly, this may also lead to a **temporary wage decline**.[40] In the medium term, it can cause long-term unemployment among workers who are unable or unwilling to transfer to new jobs, leading to a detachment from the labour market. This phenomenon is commonly referred to as '*scarring*'.

Generative AI also affects the **quality of work**. Although it presents opportunities to enhance work quality, there is also a risk of tasks becoming more entrenched as they take over complex

tasks in the workplace. This can limit employees' autonomy and put pressure on the human dimension of professional relationships.[41]

33. https://www.economist.com/business/2023/09/18/could-openai-be-the-next-tech-giant
34. Agenda Digitale Open Strategische Autonomie | Report | Rijksoverheid.nl
35. *Market dynamics where one or a few firms are so dominant that competition is almost impossible. The market may move towards a situation of (quasi-)monopoly, also known as tipping.*
36. *The threats and opportunities for competition differ by market level. Instead, for developers of applications of generative AI, there are especially many opportunities, for example in the form of alternative business models and new forms of competition within existing market structures.*
37. https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns
38. *Rathenau Institute (2023), Generative AI: pp. 30-31.*
39. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#business-and-society
40. *Greenhouse, S. (8 February 2023) US experts warn AI likely to kill off jobs – and widen wealth inequality. The Guardian.*
41. Summary - Artificial intelligence for worker management: an overview | Safety and health at work EU-OSHA (europa.eu)

There is also the challenge of income **distribution and job security** in the longer term. Accelerated and easily accessible automation by generative AI may disproportionately affect workers who lack the skills required in new jobs. Individuals with sufficient access and skills, who perform tasks that significantly contribute to productivity, have the potential to become a select group of 'superstar employees'.[42] On the other hand, workers whose skills are automated may actually lose productivity. As a result, generative AI can contribute to **polarisation in the labour market** and from rising income inequality[43], as was the case with previous waves of automation.[44] It is still uncertain which groups will be affected (relatively more) by these inequality effects. Unlike previous technological disruptions, generative AI can have a significant impact on relatively highly skilled and high-paying jobs, both positively and negatively. Some of the tasks associated with these functions (which are cognitive yet routine in nature) are at high risk of being automated. On the other hand, highly skilled workers are expected to benefit most from the deployment of generative AI.[45] The final effects of inequality are related to the use of generative AI in the workplace, whether as a supplement or a substitute, the distribution of tasks among functions and workers, and the quality and effectiveness of retraining.[46]

Impact on society
Challenges for society resulting from the use of generative AI arise on several fronts. First of all, generative AI has a potentially disruptive impact on the social and societal domain. The wide use of this technology may lead to only a limited number of people and companies (the so-called 'superstar firms'[47]) benefiting from it, which may lead to **increasing social and economic inequality.**

SER examines AI and the labour market

The Social and Economic Council (SER) is mapping the impact of AI (including generative AI) on labour productivity, quantity and quality of work.

Another significant societal challenge concerns the **substantial amount of energy** needed to train and operate generative AI models. If this energy does not come from renewable sources, generative AI may have an undesirable effect on **climate change**. Even when using renewable resources, generative AI can contribute adversely to climate change by consuming energy resources that would otherwise be used by existing sectors. Currently, the climate impact of generative AI models is still relatively limited. According to the Massachusetts Institute of Technology (MIT), training OpenAI's GPT-3 resulted in approximately 500 tonnes of CO_2 emissions.[48] This is comparable to a thousand cars driving a thousand kilometres.[49] New, larger models require substantially more energy for both training and use, which could increase emissions to tens, possibly hundreds of megatons of CO_2.[50] Besides training, the use of generative AI systems also brings climate impacts. A study shows that for a chat conversation of 20 to 50 answers, about 50 milligrams of cooling water is needed. This is equivalent to a bottle of water per session.[51]

A third challenge concerns the **degradation of our information ecosystem.** Generative AI has already demonstrated that faster and larger scale can contribute to **the creation and dissemination of mis- and disinformation.**[52] It has been possible to generate deepfakes for some years. Generative AI tools such as Midjourney, Synthesia, and D-ID have greatly increased the scale, simplicity, and realism of this process.[53] Several campaigns have been identified in which public sector bodies and other entities have disseminated misinformation using AI-generated news broadcasts that are almost indistinguishable from reality. Research indicates that individuals struggle to differentiate between real and synthetic faces and tend to have greater confidence in artificial faces.[54] The undermining effect of misinformation is amplified as (generative) AI systems also start interacting and prioritising content in *newsfeeds* and timelines on social media. Misinformation can be spread more

42. *Benzell, S. & Brynjolfsson, E. (2019). 'Digital Abundance and Scarce Genius: Implications for Wages, Interest Rates, and Growth.' National Bureau of Economic Research.*
43. AI and the Labor Market - Clark Center Forum (kentclarkcenter.org)
44. *Acemoglu, Koster & Özgen (2023). 'Robots and Workers: Evidence from the Netherlands.' National Bureau of Economic Research, working paper no. 31009.*
45. *Pizzinelli et al. (2023). 'Labor Market Exposure to AI: Cross-Country Differences and Distributional Implications.' International Monetary Fund Working Paper No. 2023/216.*
46. https://www.technologyreview.com/2023/03/25/1070275/chatgpt-revolutionize-economy-decide-what-looks-like
47. *Autor, D. et al. (2020). The Fall of the Labor Share and the Rise of Superstar Firms, The Quarterly Journal of Economics.*
48. We're getting a better idea of AI's true carbon footprint | MIT Technology Review
49. https://www.tudelft.nl/stories/articles/duurzame-kunstmatige-intelligentie-van-chatgpt-naar-groene-ai
50. *De Vries, A. (2023). The growing energy footprint of artificial intelligence, Joule (2023).*
51. *Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI less "thirsty". Uncovering and addressing the secret water footprint of AI models (arXiv:2304.03271). See also: Rathenau Institute (2023), Generative AI.*
52. *Bontridder & Poullet (2021). The Role of Artificial Intelligence in Disinformation. Data & Policy 3(E32).*
53. https://www.rijksoverheid.nl/documenten/kamerstukken/2023/06/16/tk-beleidsreactie-op-de-wodc-onderzoeken-naar-de-regulering-van-deepfakes-en-immersieve-technologieen
54. https://www.pnas.org/doi/10.1073/pnas.2120481119

Research into more sustainable use of generative AI

We examine the sustainability aspect in the development and use of generative AI (by the government) and, where possible, take measures to reduce negative impacts.

effectively and have a greater impact when allowed to circulate in this way.[55] For a democracy to function effectively, well-informed citizens and a shared understanding of reality are crucial prerequisites, as is widespread support for democratic institutions. Independent and high-quality news and information are crucial in this context. Generative AI applications potentially compromise this.[56] So far, traditional methods of fact-checking, informing and educating users, and detection tools appear to be less effective for content generated by generative AI.[57] This means that even in democratic societies, the spread of misinformation can destabilise or undermine social debate and core democratic processes such as elections.[58]

Also, the **uncertain reliability** of many generative AI models and applications negatively affects the quality of our information ecosystem. The models are based on probability and lack understanding. There are many examples of generative AI programmes unintentionally generating false information. For example, chatbots like ChatGPT regularly refer to non-existent (scientific) sources and 'make up' data presented as facts.[59] This so-called 'hallucination' carries great risks when it comes to truth-telling, especially as many people are prone to *automation bias* in which they place too much trust in the results of automated systems.[60]

As the opportunities for manipulation and misinformation increase, it is vital to understand how different public sector bodies use and regulate these technologies. The use of generative AI in authoritarian regimes enables unprecedented control over information, suppression of dissent, and increased citizen surveillance, resulting in significant human rights implications such as privacy violations and restrictions on freedom of expression.[61] This poses significant challenges to fundamental freedoms and human rights worldwide.

A fourth type of societal challenge concerns **military security**, especially in terms of the effects that generative AI will have on the international security domain. **Systemic security risks** may arise from the reinforcement of existing inequalities through the deployment of generative AI, rapid and large-scale changes in the labour market, or shifts in economic and military relations as a result of the deployment of advanced generative AI with potential implications for geopolitical relations.

---

55. *This was the case, for example, when the founder of the open-source research organisation Bellingcat, Eliot Higgins, had posted pictures of Donald Trump's arrest on Twitter in March 2023. Although Higgins had mentioned that he had created the photos with the generative AI tool Midjourney, the photos were shared thousands of times, partly through news channels:* https://www.nytimes.com/2023/04/08/business/media/ai-generated-images.html

56. *'Dat zijn toch gewoon ál onze artikelen?' – De Groene Amsterdammer.*

57. *OECD (2023). As language models and generative AI take the world by storm, the OECD is tracking the policy implications - OECD.AI*

58. *Parliamentary Paper 2023-2024, 35165 No. 46.*

59. *Beutel, Geerits & Kielstein (2023). Artificial Hallucination: GPT on LSD? Critical Care 27(148).*
     Hallucinations Could Blunt ChatGPT's Success - IEEE Spectrum.

60. *Goddard et al. (2012), Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators, Journal of the American Medical Informatics Association 19(1): 121-127.*

61. The Repressive Power of Artificial Intelligence | Freedom House

A challenge arises from the intentional **misuse of generative AI models**. As these models become more proficient, the potential for risky **abuse** increases (due to increased computing power, availability and capability).[62] For instance, generative AI models can be used to identify vulnerabilities in computer code and execute large-scale cyber attacks autonomously, without human intervention. Generative AI could also help malicious actors create new viruses.[63] As indicated earlier, the technology could also be used to create and distribute criminal content online on a large scale, including threats or *hate speech*. Online, threats against politicians, administrators, journalists, columnists, and scientists are increasing. This poses a risk of reducing willingness to perform important functions in a **democratic rule of law** and hindering free democratic debate.[64]

The sixth societal challenge pertains to the risk of incidents. As generative AI models become more proficient, they will be used more often in complex, societal processes. This increases the likelihood and impact of **incidents**, for example if AI models generate incorrect results in crucial processes. Another possibility is that AI systems will more or less autonomously pursue goals in ways that cause harm.[65] The (still) inscrutable (black box) nature of generative AI models makes it difficult to avoid such accidents. However, it is also possible that generative AI learns to make significantly fewer mistakes than humans, so it may no longer be desirable for such tasks to be performed by humans.

---

62. *Generative AI can already be abused - for example, an AI system called 'WormGPT' is circulating on the dark web that can automatically generate personalised phishing emails. There is also the risk of data poisoning:* Forcing Generative Models to Degenerate Ones: The Power of Data Poisoning Attacks for NeurIPS 2023 | IBM Research

63. *The necessary biotechnology already exists. However, generative AI could make this technology accessible to a larger number of malicious actors by providing knowledge or advice on planning and executing bioterrorist attacks. See:* Anthropic \ Frontier Threats Red Teaming for AI Safety.

64. *"Koester de Democratie! Een dringende oproep om de democratische rechtsorde weer voor iedereen te laten werken." Final report 'Adviescommissie Versterken Weerbaarheid en Democratische Rechtsorde', 2-11-2023.*

65. https://www.newscientist.com/article/mg25834382-000-what-is-the-ai-alignment-problem-and-how-can-it-be-solved/

Vision on Generative AI

Why

Society can harness the full potential of generative AI if the government actively contributes to the safe and equitable development and use of generative AI that serves human welfare, autonomy, sustainability, and increases our prosperity.

Ambition

The Netherlands aspires to be a front-runner within Europe in the application and regulation of safe and just generative AI and promotes a strong AI ecosystem in the Netherlands and the EU, in which responsible generative AI can thrive.

Learning approach

The rapid developments justify an iterative and learning approach. A coordinated approach is necessary to proactively address this development. Active collaboration is essential for developing coherent and effective policies and also for engaging in a broad social dialogue with a diverse set of stakeholders.

Action lines

A — Collaborate

B — Closely monitoring all developments

C — Shaping and applying laws and regulations

D — Increase knowledge and skills

E — Innovating with generative AI

F — Strong and clear supervision and enforcement

4 Vision on generative AI

This chapter presents the government-wide vision of generative AI. The starting point here is a values-driven approach, in line with the Value-Driven Digitalisation Work Agenda,[1] the 'Agenda Coalities voor de Digitale Samenleving',[2] the Digital Economy Strategy[3] and the EU's coordinated plan on AI.[4]

This approach is the basis of four central principles that guide the development, application and embedding of generative AI in our society. As will be explained in more detail, the government aims for generative AI that is secure and equitable and contributes to human wellbeing, sustainability and prosperity. With its values-driven approach, the Netherlands has the opportunity to play a leading role in Europe and globally. Realising this vision requires new actions from the Dutch government. Chapter 5 provides an overview of this.

Prior to creating the vision, existing policies and applicable laws and regulations applicable to (generative) AI were examined. This is not a one-off activity. Due to the rapid development of generative AI, it is crucial to monitor the public values that are under pressure, the challenges and opportunities presented by generative AI (as discussed in the previous chapter), and the extent to which policies are responding to these issues. The first part of this chapter is devoted to this legislative and policy analysis. The second part discusses the four guiding principles for the government-wide vision of generative AI. Chapter 5 describes the actions the government is taking to make this vision a reality.

a Current laws, regulations and policies

This section discusses the existing policies, laws and regulations applicable to both the development and use of generative AI. It also discusses the European AI Act. Finally, international developments in generative AI are highlighted.

i AI-policy

The Dutch government has been working on policies related to AI for some time now. The policies that were partly drafted for more traditional or 'narrow' AI are largely appropriate for the challenges and opportunities presented by generative AI. However, the widespread availability of generative AI, its scale, and the rapid pace at which it is developing call for a future-proof

vision and associated actions to address the increased risks or seize the opportunities.

AI has a high priority in the Dutch policy for the digital society. It is a crucial technology that we aim to adopt alongside the leading European innovators.[5] Since 2019, the Strategic Action Plan for AI (SAPAI) has been in place, which aims to capitalise on AI's opportunities and safeguard public interests in AI. The Value-Driven Digitalisation Work Agenda lists policy priorities around protecting public values in AI. In the field of AI, this includes making the application of algorithms fair and transparent. To achieve this goal, it is important to ensure that digital participation is accessible to everyone, the digital world is reliable, and individuals have control over their digital lives. The Digital Economy Strategy addresses, among other things, capitalising on opportunities and streamlining the AI market.

1. https://www.rijksoverheid.nl/documenten/rapporten/2022/11/04/bijlage-1-werkagenda-waardengedreven-digitaliseren
2. https://open.overheid.nl/documenten/10c88500-cdb5-4815-bd00-c91515242ea3/fil
3. open.overheid.nl/documenten/ronl-c6a34954523bef54ca41011f629b77b7b611045f/pdf
4. https://digital-strategy.ec.europa.eu/nl/library/coordinated-plan-artificial-intelligence-2021-review
5. See also: https://open.overheid.nl/documenten/ronl-e14cdcee-690c-4995-9870-fa41413190d6f/pdf

As mentioned in the July 2023 letter 'Regulating Algorithms⁶', the government is taking steps to regulate AI. These are key initiatives in this regard:

- Focusing on **responsible AI applications**. Through the **Dutch AI Coalition (NL-AIC)**, government, industry, educational and research institutions and civil society organisations are working together on socially responsible AI applications. Among others, through labs in which scientists, entrepreneurs and public institutions - the so-called ELSA labs - **research** the **ethical, legal and social aspects of AI.** The NL-AIC has developed an **AI course** that is available to everyone free of charge.
- **The AiNed programme** is a multi-year public-private programme within the National Growth Fund. The programme aims to position the Netherlands as a leading country in AI and contribute to economic recovery and growth, as well as the structural strengthening of the country's economic base. Additionally, it aims to promote a people-oriented and responsible application of AI. Through AiNed, investments in recent years have included attracting exceptional AI talent and increasing the number of Dutch parties able to participate in AI research and innovation projects with European cooperation.
- It is important that the government facilitates a support structure that manages the **development of AI for education**. With funding from the National Growth Fund, the Ministries of Economic Affairs and Climate Policy and Education, Culture and Science are investing substantially in **the National Education Lab AI (NOLAI) for a period of ten years.** Teachers, scientists, and companies collaborate to responsibly develop and evaluate advanced digital innovations, such as AI, in primary education.[7] For instance, AI is being utilised to develop a centralised teacher dashboard and to offer customised assistance to students in their learning.
- In addition, the National Growth Fund programme Npuls is developing a **national AI point and AI vision for Secondary vocational education (MBO), higher vocational education (HBO) and university education (WO)**. The aim is to prepare the sectors for the transformation of education and to help shape these changes in collaboration with partners and institutions.
- Major investments have already been made in the field of **secure AI**. The Innovation Centre for Artificial Intelligence (ICAI) is conducting extensive research and experimentation in collaboration with industry, government, and the knowledge sector. The ROBUST programme has a total budget of 87 million. For example, ICAI has launched AI labs in various Dutch cities in recent years. These labs facilitate collaboration between public sector bodies, companies, and the scientific community.

ii Laws and regulations in the Netherlands and the EU

Different legal frameworks apply to the development and application of generative AI. The following discusses how the development of generative AI relates to fundamental rights with a particular focus on the European AI Act.

Fundamental rights
The use and development of generative AI could potentially affect the realisation of fundamental rights. The fundamental rights affected by generative AI are the prohibition of discrimination and the right to privacy and data protection.

The prohibition of discrimination and the principle of equality are included in Article 1 of the Constitution. The prohibition of discrimination is enshrined in Article 14 of the European Convention on Human Rights (ECHR) and Article 26 of the International Covenant on Civil and Political Rights (ICCPR). As previously noted in chapter 3b, generative AI can inadvertently perpetuate bias or discrimination. Bias can enter generative AI systems in various ways, including through developer and training data. As only a handful of AI developers shape generative AI and applications, (unconscious) bias may inadvertently be introduced into the models. In addition, training data may contain bias and enter the model, spreading and amplifying this bias widely.[8]

The use and development of generative AI may raise concerns regarding the protection of privacy and data. The right to privacy is set out in Article 10 of the Constitution. In addition, the right to privacy is protected by Article 8 of the European Convention on Human Rights (*Right to respect private and family life*) and Article 17 of the International Covenant on Civil and Political Rights.

---

6.   *Parliamentary paper 2022-2023, 26 643 no 1056.*
7.   https://www.rijksoverheid.nl/documenten/kamerstukken/2023/07/06/visiebrief-digitalisering-in-het-funderend-onderwijs
8.   *Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems, Volume 3, 2023.*

Privacy and data protection law

Under Article 8(1) of the Charter of Fundamental Rights of the European Union and Article 16(1) of the Treaty on the Functioning of the European Union, individuals have the right to the protection of their personal data. The main general rules for processing personal data are contained in the General Data Protection Regulation (GDPR) and, additionally for the Netherlands, the GDPR Implementation Act (Uitvoeringswet AVG, UAVG). These rules apply to both public and private parties falling within the scope of this legislation. Competent authorities with law enforcement tasks are subject to different rules for the protection of personal data. This is the Police Data Act (Wpg) and the Judicial and Criminal Records Act (Wjsg). Data protection law sets standards for the proper and lawful processing of personal and police data. These standards include general principles and bases, rules on transparency, personal data security, and data subjects' rights, such as the right to access, rectify, and delete their data. In principle, the processing of special categories of personal data, such as data revealing ethnic origin, biometric data for the purpose of uniquely identifying a person, or data concerning health, is prohibited unless strict conditions laid down by law are met. To avoid a serious risk of circumvention, the right to data protection is independent of technologies used, such as generative AI. The Telecommunications Act (Telecommunicatiewet, Tw) is supplementary to the GDPR as lex specialis. As a general rule, internet user consent is required by the Tw when confidential communications are interfered with or when cookies or similar techniques are used.

Monitoring compliance with personal data protection legislation is in the hands of an independent regulator. In the Netherlands, the Dutch Data Protection Authority (AP) has the authority to do so. The AP has the authority and ability to investigate whether parties are adhering to their obligations under personal data protection laws and take corrective action accordingly. It is therefore not the task of the government to monitor the legality of data processing in the private sector in specific cases.

The European Data Protection Board (EDPB) is a collaborative and independent body consisting of all national privacy regulators from the EU and the European Data Protection Supervisor (EDPS). As generative AI is a cross-border phenomenon that requires a harmonised approach, the EDPB has established a ChatGPT task force to promote cooperation and information exchange on possible enforcement measures.[9] The Netherlands Authority for Consumers and Markets (ACM) is the regulator of compliance with the Tw.

Copyright

Copyright law includes regulations for safeguarding literary, scientific, or artistic works from unauthorised disclosure and reproduction. The Copyright Act contains a restriction on the reproduction right for the purpose of text and data mining[10] that can be used to train AI. Anyone can invoke that restriction but only when lawful access to the works has been obtained. This restriction does not apply when rightholders have explicitly reserved the right to make copies for text and data mining purposes. In the case of works made available online, reservations should be made by machine-readable means, such as in the metadata where the conditions for use of a website are indicated. If a reservation is made correctly, prior permission from the rights holders is required again to make the necessary copies for text and data mining. Permission to use may be subject to conditions, such as payment of a fee and proper attribution including the creator's name. Similar rules apply to training generative AI models with copies of neighbouring and database-protected performances.

An output of generative AI as such is not eligible for copyright protection. After all, there is no creation of the human mind.

If there is cooperation between a human and an AI system, the output may be eligible for copyright protection under circumstances. This requires creative choices in developing the idea in the prompt. The human influence on the final result is probably not enough to constitute a work in the sense of copyright law. However, any creative choices added later to the final work that is made public may alter this. There is a possibility that the output may violate copyright, related rights, database rights, or portrait rights.

It is up to the courts, and ultimately the European Court of Justice, to determine whether current generative AI models can successfully invoke the text and data mining exception. At this stage, the court has not been given the opportunity to rule on the matter. Outside the EU, however, several lawsuits have already been filed centring on whether training (generative) AI models constitutes copyright infringement.[11] There is a realistic possibility of copyright infringement when developing generative AI tools due to the lack of legitimate access to works and/or failure to respect reservations made.[12] Therefore, it is crucial to maintain ongoing monitoring in the upcoming years and, if required, to strengthen policies concerning this matter in the European context.

---

9.   https://edpb.europa.eu/system/files/2023-09/20230919-20plenagenda_public.pdf
10. *An automated analysis technique aimed at decomposing text and data in digital form to generate information such as patterns, trends and interrelationships.*
11. *Rathenau Institute (2023), Generative AI.*
12. *See opinion of the Landsadvocaat (2023) on use of generative AI tools:* https://open.overheid.nl/documenten/16d72572-da6b-422c-8cf8-cdc95a523093/file

Protection of trade secrets

Innovative companies are increasingly exposed to practices aimed at unlawfully obtaining trade secrets, such as misappropriation, unauthorised copying, economic espionage or breach of confidentiality requirements, both within and outside the EU. As defined in the Trade Secrets Protection Act, trade secrets are know-how and business information that is valuable because it is secret and is intended to remain secret. The holder has also taken measures to keep them secret. However, developments such as generic AI carry increased risk. For example, in addition to what is mentioned above in relation to copyright, in the training phase of generative AI, the model may be fed with data that is classified as a trade secret. As companies' investment in intellectual capital affects their ability to innovate and compete, great care should be taken to ensure that the application of generic AI does not undermine this. If it does, it can have a negative impact on profitability and the willingness to innovate further.

Competition and market regulation

Effective market forces are a prerequisite for providing Dutch businesses with sufficient choice at a fair price and are an incentive for innovation. Past experience of power concentration and dependence in other technology markets has shown that the benefits of some technologies remain too much in the hands of a few large technology companies, with little flow through to entrepreneurs and consumers. This could ultimately hamper productivity growth in the economy as a whole. A similar dynamic can be seen in the field of generative AI, for example in the position of model developers. Competition law can be used by regulators to tackle anti-competitive behaviour (such as abuse of a dominant position) and to prevent undue barriers to entry.

In addition, the European Digital Markets Act (DMA) contains specific rules for the largest online platforms, so-called gatekeepers. These are platforms that businesses and consumers can no longer ignore. Many gatekeepers in the AI industry have gained significant market power, which they can leverage in other technology markets like cloud computing, resulting in spillover effects. The rules in the DMA aim to increase the contestability of the position of gatekeepers and reduce their dependence on them. For instance, the DMA includes obligations related to interoperability and data, as well as prohibitions aimed at preventing the transfer of market power. The DMA provides opportunities to apply rules to AI markets. For instance, some AI applications may already fall under the DMA's scope. The DMA allows for flexibility to expand the scope and add more obligations if needed, based on market research

European AI Act

The AI Act is considered to be the primary legislative framework for the development and use of AI in the European Union. Requirements have been established for generative AI, and a system for monitoring these requirements has been put in place. On 8 December 2023, a preliminary political agreement was reached in Brussels concerning the AI Act. The Act will come into force once it has been approved by all EU Member States and the full European Parliament. The aim of this European act is to facilitate the integration of secure AI systems into the internal market, while also protecting public health and fundamental rights. In order to achieve this goal, certain criteria will be established for AI systems, taking into account the level of risk they may pose. While some AI practices are prohibited, other AI systems are subject to high requirements due to their high-risk scope, such as for recruitment and selection or law enforcement. The act will have direct effect in the Netherlands and the regulation will automatically become part of Dutch law. Part of this, such as the supervision of prohibited and high-risk AI applications, will be further established by a Dutch law. The AI Act covers Generative AI and the powerful AI models that often underpin it, which can be used for a wide range of applications. These models are also known as general-purpose AI (GPAI) models. For example, it is mandated that all GPAI models adhere to transparency standards. This guarantees that businesses utilising these models for particular AI applications are provided with the essential technical documentation to conform with the AI Act. The most powerful GPAI models with systemic risks will be subject to additional obligations in terms of risk management, serious incident monitoring, and conducting model reviews. These obligations will be implemented through codes of practice that the European Commission will develop in collaboration with industry, science, civil society, and other stakeholders. The European Commission will establish a regulatory body, known as the AI Office, to enforce the new regulations for GPAI models.

Generative AI systems, such as chatbots and systems that generate images and videos, must meet additional transparency requirements under the AI Act. Providers of generative AI systems must ensure that it is clear to people when they are interacting with AI or when content has been generated by AI.

This approach aligns with the Netherlands' commitment to the AI Act. The government deems it appropriate to impose obligations on companies and public sector bodies to safeguard individuals from potential risks posed by AI applications. The development and use of generative AI must be safe to gain the trust of society and the market and to fully exploit the potential of this technology.

The European AI Act has been under negotiation between EU member states and with the European Parliament since 21 April 2021. After approval by EU member states and the European Parliament, the law will come into effect. Based on the political agreement reached in December 2023, the Netherlands' authorities and businesses will have between half a year and two years to ensure that AI systems being developed, purchased, and used comply with the requirements of the AI Act. This timeline depends on the level of risk involved; for instance, some AI practices may be prohibited as early as 6 months. For high-risk AI applications, there are timelines of 24 and 36 months, during which the implementation law will be developed in consultation with stakeholders and presented to parliament. The requirements will apply to GPAI models, which encompass most large AI models that generate content, after a period of 12 months. European supervision will be established within this timeframe.

iii International developments

Generative AI is an international phenomenon. Considering the impact of this technology on the global population and on geopolitical and international relations, it is crucial for the Netherlands to take an active role on the international stage. Several key developments in (generative) AI are outlined below.

- **The CAI (Committee on AI) of the Council of Europe** is developing an AI treaty aimed at regulating AI systems in line with the Council's human rights, democracy and rule of law standards. The European Commission negotiates on behalf of the EU and works closely with EU member states. The treaty is scheduled to be completed by April 2024.
- **The Global Partnership on AI (GPAI)** is an initiative of France and Canada to promote cross-border cooperation between experts working on responsible AI.
- **The OECD AI Expert Group (AIGO)** works on the implementation of OECD AI principles, research in a variety of fields and exchange of best practices on AI in the OECD and other countries. The Netherlands is a member of this group, collaborating with other OECD countries to develop responsible and ethical AI technologies.
- **The G7 Hiroshima AI Process** focuses on establishing international guidelines for organisations developing advanced AI systems, and aims to promote safe, secure and reliable AI worldwide. The non-exhaustive list of guiding principles is discussed and developed as a living document to build on the OECD's existing AI principles in response to recent developments in advanced AI systems.
- The **EU-US Trade and Technology Council** was established at the EU-US summit on 15 June 2021 in Brussels. Its purpose is to serve as a forum for the US and the EU to coordinate major global trade, economic, and technology issues and to strengthen transatlantic trade and economic relations. Within this council, the EU and the US collaborate to develop dependable AI technology, among other objectives.[13]
- **The UN High Level Advisory Body on AI**, established by the UN Secretary General, is tasked with formulating recommendations for international AI governance structure. This advisory board consists of 39 experts, published an interim report[14] in December 2023 and will present a final opinion in September 2024 at the 'Summit for the Future'.[15]
- **UNESCO** is also active in the field of AI and is also a partner in the Global Partnership on AI (GPAI). In 2021, a recommendation on AI ethics was adopted, partly at the initiative of the Netherlands. This was followed by policy papers on ChatGPT and AI Foundational Models in relation to the recommendation in 2023. The Netherlands aims to maintain focus on implementing this recommendation, emphasising the connection between AI, ethics, and human rights.

On the initiative of the Biden administration, on 21 July 2023, leading US AI companies signed the **"Voluntary Commitments on AI "**, a set of voluntary principles that emphasise security, trust and transparency in AI development. In parallel, the US established the AI Contact Group, a cross-regional group of 21 countries, including the Netherlands. In addition, President Biden (US) issued the first presidential *executive order* on AI on 30 October 2023. The executive order aims to promote safe, secure, and reliable AI innovation. It focuses on current and future AI developments.[16]

The Netherlands initiated the first international dialogue on responsible use of AI in the military domain in February 2023 by organising the **REAIM summit** in The Hague. More than 50 countries have signed a Call to Action, which is a significant accomplishment. The Call to Action is viewed by the government as a crucial initial step in engaging as many countries and stakeholders as possible in setting the international agenda on this topic. It serves as a broad foundation for further discussions on international frameworks for the responsible application of AI in the military domain. The Netherlands is co-hosting the next edition of REAIM. It will take place in South Korea in 2024.

During REAIM, the **Global Commission on Responsible AI in the Military Domain** was launched. The commission aims to provide concrete and policy-relevant advice on the governance of military AI within two years, which will be important for talks and negotiations between states, likely in a UN context.

The Netherlands is an active contributor to NATO's Data and Artificial Intelligence Review Board (**DARB**). This board is developing a certification standard for artificial intelligence in the military domain. The standard should ensure that companies and institutions within the alliance act in accordance with international law and NATO's norms and values. Within the DARB, there is a special focus on the opportunities that Generative AI offers for military applications, subject to responsible deployment as set out in the standard.

---

13. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/stronger-europe-world/eu-us-trade-and-technology-council_en
14. https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf
15. https://www.un.org/techenvoy/ai-advisory-body
16. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

b Four principles for generative AI

The preceding section shows that the government is actively involved in regulating and promoting AI, including generative AI, at national, European, and global levels. The rules and initiatives that result will aid society in responsibly harnessing generative AI. At the same time, it is important to consider the *possibility* that current policies may not sufficiently address certain risks or protect necessary conditions for taking advantage of opportunities.[17] Given the expected impact of generative AI, policy gaps could have major implications for people, the economy and society. This could lead to the identification of potential policy actions and the establishment of necessary preconditions. This requires the government to have a **proactive and open-minded approach, as well as vision and courage**. To effectively address both opportunities and challenges, policymakers must be agile in their decision-making. Here, open cooperation between government, industry and science is important in order to identify signals early and make adjustments. An ecosystem approach to AI responds to the rapid technological developments around generative AI that have implications for society, the market and citizens.

The government-wide vision for generative AI is based on four value-driven principles that are in line with the Value-Driven Digitalisation Work Agenda[18], the Digital Economy Strategy and the 'Agenda Coalities voor de Digitale Samenleving' (Agenda Coalitions for the Digital Society). The Dutch government is committed to the **safe** and **equitable** development and use of generative AI that **serves** human welfare, autonomy, **sustainability, and increases** our prosperity. These four principles articulate the government's objectives (which are also aligned with the Sustainable Development Goals (SDGs[19])) in the development, use and embedding of generative AI. They also pro-

vide insight into how generative AI intervenes in the different values central to the principles.

The (intended) policy actions contribute to the implementation of several principles and are presented separately in Chapter 5 for ease of reference. However, these policy actions alone will not be sufficient. The need for any new actions or policies should be assessed frequently in the coming years using a learning and iterative approach. Subnational governments and implementing organisations will continue to be actively involved in this process, particularly with regard to feasibility.

1 Principle 1: Generative AI is developed and applied in a safe way

The Dutch government is committed to the safe development and use of generative AI systems. More specifically, this means actively contributing to the mitigation of **abuse**, **incidents** and **systemic security risks** of and by generative AI models. This is done at national, European, and international levels because many potential security risks do not respect national borders and cannot be remedied by unilateral efforts of the Dutch government.

Mitigation of misuse of generative AI systems
Chapter 3 describes how generative AI models can be abused by malicious actors if they lack robust security mechanisms. The most recent generation of AI models already includes

more 'guardrails' to prevent misuse. However, it is still possible to 'jailbreak' models.[20] Open-source - models may be particularly vulnerable to this.[21] Generative AI models can already be misused to create disinformation and phishing emails [22], manipulative content and deepfakes . Future generations of generative AI models with increased skills could enable even more far-reaching misuse, such as automated cyberattacks, or assist in the synthesis of dangerous viruses or chemicals. Misuse of generative AI models disregards national borders, as explained in Chapter 2. For this reason, the Dutch government is committed in the European context to **limiting opportunities for abuse,** and to encouraging techniques that reduce the potential for abuse. We invest in national resilience against abuse, particularly in the cyber domain. This is demonstrated through our action point 'Increase knowledge and skills'.

Mitigation of incidents by using generative AI systems
After all, the government encourages the development of intrinsically safe generative AI models- **use** of **generative AI-** models **should not lead to (large-scale) incidents**. To this end, developers should prevent systems from generating erroneous or risky information. This is no easy task: despite considerable efforts by AI developers to reduce so-called *hallucination* , text models still regularly generate inaccurate or even dangerous output with great veracity. Modern generative AI models also have a significant black box nature. This means that we cannot predict when a model will behave in an undesirable or unreliable way, nor can we verify that the objectives we give a model have been correctly embedded in the model. This lack of interpretability and explainability becomes more problematic when future models are used to take actions or make decisions independently. The government advocates for safety and interpretation rules for generative AI models through international

---

17. The Rathenau Institute cautions that current and future policy frameworks may not be up to the challenges posed by generative AI. See: Rathenau Institute (2023). Generative AI: p. 33.
18. Everyone can participate in the digital age, trust the digital world, and have control over their digital lives.
19. See also: https://www.rijksoverheid.nl/onderwerpen/ontwikkelingssamenwerking/internationale-afspraken-ontwikkelingssamenwerking
20. Within the context of LLMs, jailbreaking refers to designing prompts with the intention of exploiting model biases in order to generate output that is inconsistent with the purpose of the model. For example, the model will answer questions that would not normally be answered by the model.
21. Abdelnabi, S., Greshake, K., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023, November). Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection. In Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security (pp. 79-90). See also Appendix 2 for more information regarding open-source generative AI models.
22. This is also pointed out by the Cyber Security Council (2023). See also: https://www.cybersecurityraad.nl/actueel/nieuws/2023/12/22/csr-brief-over-ai-en-cybersecurity

organisations such as the OECD and the UN. Additionally, the government encourages research into responsible and transparent AI models. The government acknowledges that open-source models can provide solutions for transparency and explainability. It encourages the use of open source by applying the 'open source unless' principle in procurement and development.[23] However, it is important to note that transparency should not compromise the safety of generative AI models.

Mitigation of systemic safety risks

The widespread adoption of generative AI should not exacerbate social inequality, instability, or disrupt vital processes. **Systemic security risks** could arise from the reinforcement of existing inequalities through the deployment of generative AI, rapid and large-scale changes in the labour market, or shifts in economic and military relations resulting from the deployment of advanced generative AI. Systemic security risks are difficult to prevent through targeted actions due to their diffuse nature. The government is committed to **researching and monitoring** how generative AI can lead to wider, unintended social change. A proactive approach is essential in this case. Unintended systemic effects of generative AI should not simply "befall us", as has happened with social media. Systemic security risks often have an international component. The government is therefore actively pursuing cooperation with like-minded countries when it comes to mitigating such risks.

2 Principle 2: Generative AI is developed and applied equitably

The government aims to ensure that the development, use, and impact of generative AI align with principles of equity.

Lawful development and deployment

Generative AI should be developed and deployed legitimately. Herein, we distinguish several elements specific to generative AI. First, it must be clear who **bears responsibility** for the proper functioning of AI models and who is responsible under what conditions for any harmful or undesirable outcomes. Because of the black box nature of AI models, as well as the complexity of the socio-technological structure of AI systems, there is room for confusion of responsibility.[24] For example, who is responsible for harmful or illegal deployment or output of a particular generative AI model? The model developer, the application developer, the organisation implementing the tool or the user?[25]

Second, the development and application of generative AI must respect already **existing laws and regulations** such as the **Constitution, privacy and data protection law** and **copyright law**. As explained earlier (in Chapter 3), the ways in which generative AI models 'harvest' and process data may infringe on these rights. Conformity to existing regulations requires translation of existing legal frameworks to the context of generative AI. Where there are ambiguities or policy gaps, or where regulations are no longer considered 'fit for purpose', consideration should be given to strengthening existing frameworks. In addition, it remains important to ensure that supervisors have

the knowledge, capacity and resources to carry out their tasks effectively now and in the future. The importance of analysing and (where necessary) adjusting regulatory frameworks and capabilities of supervisors is also repeatedly stressed by the Rathenau Institute.[26]

Opportunity equality

The government recognises that the development and deployment of generative AI may put equality of opportunity under pressure. The possible cause of this is twofold. A first potential factor is **unequal access** to generative AI applications for residents (due to income differences). A second potential factor is a **digital skills gap.** Both factors can have a significant influence on **social and economic opportunities**. The government aims to ensure that everyone in our society has the means and skills to benefit from the opportunities offered by generative AI applications. This requires investment in low-threshold accessibility and technological citizenship.

Non-discrimination

As described, many AI systems suffer from **bias**, **selectivity** and **stereotypical** thinking embedded in underlying data and model parameters. In this way, generative AI can act as a catalyst for discriminatory dynamics, for example when it comes to recruitment policies in organisations and companies.[27] The government considers this unacceptable in the light of the principles of non-discrimination. The government endorses the importance of developing and deploying methods to mitigate bias and discrimination. Several methods are under development, such as data curation, 'constitutional AI' (where an AI model is automatically trained to give answers that fit constitutional principles)[28] or 'democratic AI', where a representative selection of people are involved in the development of an AI model or application. The government is also encour-

23. For an overview of the government's efforts on open-source software, see Parliamentary Papers II 2022/2023, 26 643, no 1057.
24. Novelli et al. (2023). 'Accountability in Artificial Intelligence: What It Is and How It Works.' AI & Society.
25. This issue is the subject of a lively legal-ethical academic debate. See, for example: Zech (2021). 'Liability for AI: Public Policy Considerations'. ERA Forum, 22: pp. 147-158 & Hacker (2023). 'The European AI Liability Directives – Critique of a Half-Hearted Approach and Lessons for the Future'. Computer Law & Security Review 51.
26. Rathenau Institute (2023), Generative AI: chapter 4.
27. https://www.forbes.com/sites/forbeshumanresourcescouncil/2021/10/14/understanding-bias-in-ai-enabled-hiring/?sh=3f9734837b96
28. For an example of constitutional AI, see: https://www.anthropic.com/index/collective-constitutional-ai-aligning-a-language-model-with-public-input

aging extensive testing of generative AI models to verify bias and discriminatory outcomes. This should be done periodically, as models can deteriorate over time.

Transparency and explainability
The government endorses the importance of transparency and explainability of (generative) AI models. The black box nature[29] of generative AI models hinders a basic understanding of how AI models work. This damages **procedural justice**: without insight and explanation, the fairness of processes and procedures driven by generative AI cannot be verified. Moreover, the lack of transparency hinders the **correction of biases** in the underlying data and model parameters. The government is committed to increasing transparency and understanding of the training data and operation of AI models, tailored to the needs of the context in which a model is applied. The transparency obligations that will apply to AI systems under the AI Act offer a solution to this. As a government, we are also setting a good example ourselves by, among other things, imposing conditions on the origin of (training) data when procuring generative AI systems. Moreover, it should be possible for scientists to look under the bonnet of AI models. Users can benefit from clear and concise *model cards* - a kind of leaflet with technical details and possible limitations of the AI model in question.[30] Such a leaflet can also help (end) users determine whether a model is suitable, and whether any dangers exist in its deployment. The AP also signals this in its second report on AI & Algorithm Risks Netherlands.[31] Open-source models can also help in some cases when it comes to transparency.[32]

Generative AI for greater equality
Generative AI should help promote equality and bridge (socio-economic) gaps, both within and between countries. This is in line with several of the Sustainable Development Goals (SDGs).[33] This requires, firstly, monitoring the impact of automation on wage and employment trends, especially where **increasing income and wealth inequality**. AI may cause wages for different jobs to diverge more, as was the case with previous automation.[34] Secondly, policies and initiatives that promote economic equality, such as education and retraining programmes, social safety nets, and inclusive AI development that ensures a more balanced distribution of opportunities, can help counter economic inequality and reduce the digital divide. It is important to ensure that our social safety net is well-equipped for the socio-economic transitions that generative AI is expected to trigger in the mid to long term. Finally, equity requires fair conditions for all parties involved in model development and training. It is important to note that human 'labellers' who assist in improving AI models should receive fair compensation and working conditions. Unfortunately, this is not always the case.[35]

3 Principle 3: Generative AI that serves human welfare and safeguards human autonomy

The deployment and development of generative AI should serve human welfare. According to the World Health Organisation (WHO) definition, wellbeing is a positive state of physical, mental and social wellbeing.[36] According to WHO, wellbeing includes both quality of life and the feeling of (being able to make) a meaningful contribution to the world.

Health
Generative AI applications that contribute to **physical and mental health**, such as making accurate and efficient diagnoses, improving care and other public welfare services, and aiding medical research, should be encouraged. At the same time, the government acknowledges the importance of preserving **commonality** and **humanity** in our living environment, for the purpose of social cohesion and mental wellbeing. It may be undesirable to replace certain forms of human contact with AI-driven processes. Automation of contact can potentially foster loneliness and social alienation, or impair our social capabilities.[37] All this requires a robust public debate about the desired role of generative AI in society, considering the opportunities and drawbacks, the possibilities and possible limits that should be placed on the deployment of generative AI. We organise this in broad social coalitions.[38]

---

29. *Black box models: An (AI) model for which there is a lack of understanding of how the model's prediction was made and what the basis of the model is.*
30. *See:* Model Cards (huggingface.co)
31. *Dutch Data Protection Authority (2023). Report AI & algorithm risks Netherlands (RAN) - autumn 2023.*
32. *See also Appendix 2 for more explanation on open source and generative AI and associated challenges.*
33. *Such as fighting all forms of poverty (SDG 1) and reducing inequality (SDG 10).*
34. *The reverse can also happen: scenarios are conceivable in which the average worker becomes more productive through generative AI deployed in a supportive (rather than displacing) way. This is described under 'opportunities' in chapter 3.*
35. https://ssir.org/articles/entry/ai-workers-mechanical-turk
36. *See: WHO (2021). Health Promotion Glossary of Terms 2021. URL:* Health Promotion Glossary of Terms 2021 (who.int)
37. *Turkle, S. (2015). Reclaiming Conversation: The Power of Talk in a Digital Age. New York: Penguin Press. &*
38. *An example is the social coalition 'Over Informatie Gesproken'. (Talking About Information).*

Personal and professional autonomy
The government also considers it important that the deployment of generative AI promotes human **self-development** and not at the expense of **personal autonomy,** both privately and in the workplace. As outlined in Chapter 3, generative AI applications have the potential to limit users' autonomy. Direct and indirect methods can compromise personal autonomy. Direct methods include deception, while indirect methods include 'dark patterns' or microtargeting of (dis)information. The answer may lie in a combination of digital resilience and effective enforcement of the Digital Services Act's ban on dark patterns and certain forms of profiling. In the workplace, the deployment of generative AI may compromise **professional autonomy** by increasing the intransparency of processes through AI automation. Far-reaching automation of work processes must be avoided as it can erode the humanity and meaning associated with work. This requires that workers maintain control over their work content and maintain social working relationships, as advocated by the WRR.[39] It is essential to anticipate the labour market of the future and to ensure that people have the skills to maintain their job content in the context of new technologies. Social partners could play a significant role in integrating generative AI into companies and organisations through participation councils and collective bargaining agreements.[40]

4 Principle 4: Generative AI contributes to sustainability and prosperity

The Dutch government is committed to **sustainable** generative AI that contributes to our **prosperity**. Generative AI is being used to promote sustainable economic growth, reduce labour shortages, and lead to innovative new products and solutions. It can also help address societal issues such as climate change.

Level playing field and productivity growth
A prerequisite for this is ensuring **healthy competition** between AI developers, to promote both market accessibility and models as well as competitive prices. This requires effective enforcement of European and national competition rules, as well as the Digital Markets Act (DMA). Competition authorities require the necessary tools, expertise, and capacity to promptly intervene when necessary to prevent anti-competitive behaviour. Other preconditions for successful implementation of AI within organisations include having the necessary knowledge, availability of skilled workers, sufficient organisational support, and an adequate digital infrastructure.

When generative AI is used responsibly, it offers all kinds of opportunities to enhance **productivity** in the workplace. Even now, *coding assistants* can help software engineers write higher-quality code in the same amount of time, and language models help with writing or editing. As generative AI systems become even more skilled in the future, these types of systems will be able to take over time-consuming tasks in more and more ways, leaving more time for core tasks. The Dutch government is committed to the responsible adoption of (generative) AI models that assist with human tasks. Where possible and desirable, these models can also fully automate tasks. In doing so, we are also considering the adoption of generative AI applications in public sectors, such as healthcare. Deploy-ing generative AI can help organisations reduce their human resource shortages and contribute to economic growth. Productivity increases may also lead to job creation as demand for goods and services rises with higher incomes.

Responsible and innovative generative AI deployment
The deployment of generative AI systems can not only speed up existing tasks, it can also contribute to the **innovative** capability of the Netherlands. Our vision is for Dutch companies and organisations to be at the forefront of applying generative AI models to create innovative products and business models. Generative AI models can also be developed and applied to accelerate scientific research and development. As a government, we promote responsible generative AI through investments, public-private partnerships, and collaboration with knowledge institutions. This approach allows us to take advantage of the opportunities presented by generative AI in addressing societal problems.

Not only Dutch companies, but also Dutch consumers should benefit from generative AI. To this end, we want consumers to have access to a wide range of responsibly generative AI-driven **products and services**. The Netherlands and the EU are promoting the responsible development of generative AI systems in an international context. This includes making useful generative AI applications developed abroad readily available in the Netherlands. European regulations, including the AI Act, establish the necessary conditions for deploying (generative) AI in a manner that promotes safety, health, and fundamental rights. Securing European values and stimulating innovation can thus go hand in hand.

39. Het betere werk. De nieuwe maatschappelijke opdracht | Report | WRR
40. Workers could be the ones to regulate AI | Financial Times (ft.com)

The path of education and science

Generative AI can also improve education for both children and working adults seeking to upskill or retrain. Generative AI can be utilised to create new teaching materials or personalise teaching methods. Furthermore, if Dutch education takes advantage of such opportunities, generative AI can contribute to higher-quality education and our future **earning power**.

In science, generative AI can make a significant contribution to solving complex problems, especially with specific data sets, such as medical images and texts, protein structures or mathematical problems.[41] As a result, this technology can play an important role in **driving innovation**.

Sustainability

An important point is that the development and deployment of generative **AI should not have an unwanted impact** on our climate **and ecosystems.** This means that, as a government, we will not use technology if it causes great harm to people and the planet. In line with the Sustainable Development Goals, our goals include **sustainable innovation** and combating climate change.[42] Sustainability should be promoted by focusing on energy-efficient training processes and delivery, with priority given to the use of renewable energy sources. At the same time, the application of generative AI can be used to help mitigate climate change. For instance, generative AI models can help optimise energy use, or support scientific research into clean energy sources.

---

41. *Rathenau Institute (2023), Generative AI: p. 14.*
42. *See in particular SDG 9 (sustainable industry, innovation and infrastructure) and SDG 13 (combating climate change).*

5 Actions

In order to achieve a responsible embedding of generative AI in Dutch society, (concrete) actions, grouped into six action points, are listed below.[1] Some of these are already ongoing actions, while others are new.[2] A number of potentially new actions are also being explored or studied.

The measures will ensure that society can fully benefit from the opportunities and mitigate the risks of generative AI. Some actions intervene in the technology itself, others in the companies that produce the technology, and still others in the social context in which it is used.

Some actions contribute to the achievement of all the principles listed above, while others may contribute to several, or a single, principle. A number of potentially new policy actions are being explored or researched for which there is not yet funding. The outcome will also be submitted to the next government.[3]

Generative AI affects the whole of society. Dealing with the opportunities and challenges of generative AI is a task that must be approached **collectively**, based on a **learning and experimenting approach.** That means continuously conducting a **broad social dialogue** in the Netherlands and seeking **international cooperation**, within the EU and globally. This requires being well **informed about current developments** of generative AI and being aware of the socio-economic and sustainability implications. This will enable us to anticipate socio-economic and digital changes and assess the sustainability of implementing generative AI. In doing so, we emphasise responsible application of generative AI, so that the entire society benefits from the potential this technology has to offer. As the Netherlands, we aim to be a global leader with a strongly values-driven approach.

Staying informed is not enough, we will actively engage in **increasing knowledge and skills among public sector bodies, organisations and citizens**. The government sets a good example by encouraging **innovation** and investing in talent within the **Dutch and European AI ecosystem**. At the same time, it is important to increase **knowledge and skills** about generative AI. Education is crucial in this regard, which is why the government provides support to educational institutions to ensure they are adequately equipped to respond to technological advancements. This provides pupils, students and teachers with up-to-date knowledge and skills. This provides a solid foundation for the targeted deployment and evaluation of generative AI within appropriate ethical frameworks. Given the social impact of the subject matter, it is essential to have **future-proof laws and regulations**, legal protection, and strong and clear **supervision and enforcement**.

---

1. 1. Collaboration; 2. Closely monitoring all developments; 3. Drafting and applying laws and regulations; 4. Increase knowledge and skills; 5. Innovating with generative AI; 6. Strong and clear monitoring and enforcement.
2. *These actions are covered financially.*
3. *These actions are clearly marked as exploratory or investigative.*

a Collaborate

Discuss and debate
Public trust in new technology and the government's responsible role in society are crucial for a well-functioning digital economy. This necessitates not only a distinct role for the government but also the promotion of awareness and discussion among residents regarding the potential effects of generative AI and digital citizenship. Societal dialogues and debates, such as the Dutch AI-Parade and ECP's[4] guidance ethics approach, contribute to discussions about the role of generative AI in society and how to preserve the commonality and humanity of our environment. This conversation is also important in striking a balance between harnessing the societal and economic potential of AI and dealing with challenges facing generative AI.

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
Extending an ongoing social dialogue with residents, workers, trade unions and entrepreneurs on the impact and role of generative AI on their lives and society as a whole. The Dutch AI-Parade, organised by the NL-AI Coalition, has an important role to play in this, possibly through (further) expansion into education and the Legal Parade.	Principles 2. Justice and 3. Wellbeing and autonomy	Ongoing	Ministry of the Interior and Kingdom Relations, Ministry of Education, Culture and Science, Ministry of Economic Affairs and Climate Policy

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
Exploring the possibility of appointing a specific quartermaster or organisation to actively stimulate and coordinate various initiatives (both existing and new) specifically aimed at social dialogue on responsible deployment of generative AI.	All principles	2024	Ministry of the Interior and Kingdom Relations
In 2024, the Rathenau Institute organised dialogue sessions that are ideal for social debate on the impact of generative AI and its role in society and the economy.	Principles 2. Justice and 3. Wellbeing and autonomy	2024	Ministry of the Interior and Kingdom Relations
Promoting awareness and skills among citizens to protect their online privacy, particularly the data that can be used to train generative AI models.	Principles 2. Justice and 3. Wellbeing and autonomy	Ongoing	Ministry of Justice and Safety

4.   https://ecp.nl/project/aanpak-begeleidingsethiek/

Action summary	Fulfils principle(s)	Timeline	Owner
Encouraging initiatives that train generative AI models based on democratic input can promote participation in government. For example, through the 'Rijks AI-validatieteam' (Government AI Validation Team) (see also action 'Innovating with generative AI').	Principles 2. Justice	2025-2028	Ministry of the Interior and Kingdom Relations

Intergovernmental cooperation
Intergovernmental and inter-departmental cooperation is vital, as generative AI has profound implications for society as a whole. Addressing opportunities and challenges related to citizens and entrepreneurs is a joint responsibility of all levels of government, regardless of the distance between them. Generative AI systems can be used by decentralised authorities to develop solutions that are tailored to local or regional needs.

Intergovernmental cooperation is necessary to ensure consistency in the guidelines and standards to be applied. Decentralised authorities can reinforce national policies with insights from local practice.

A coordinated approach is necessary to proactively address this development as a nation. Public sector bodies, as well as implementing organisations, business, and civil society organisations, must collaborate to take action. This collaboration is essential not only for developing coherent and effective policies but also for engaging in a broad social dialogue. In doing so, it is also of added value to explore how and to what extent generative AI can play a role in improving government information management. The same applies to dealing with the legacy issues that several parties in government have to contend with.

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
Establishing an intergovernmental triage point for data sharing is recommended, especially in light of generative AI. The purpose of triage is to facilitate data sharing and gain insights into bottlenecks. This information can then be shared with others.	Principles 2. Justice and 3. Wellbeing and autonomy	2023-2024	Ministry of the Interior and Kingdom Relations in cooperation with Ministry of Social Affairs and Employment, Ministry of Justice and Security, Ministry of Health, Welfare and Sport, FIN and subnational governments

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
Pilot testing of responsible generative AI applications in concrete (proactive) government services.	Principles 3. Wellbeing and autonomy and 4. Sustainability and prosperity	2024-2026	Ministry of the Interior and Kingdom Relations in cooperation with subnational governments and implementing organisations

Action summary	Fulfils principle(s)	Timeline	Owner
Exploring how generative AI can be deployed in legal and administrative processes ('Legal Tech').	Principles 3. Wellbeing and autonomy and 4. Sustainability and prosperity	2024 and beyond	Ministry of the Interior and Kingdom Relations in cooperation with subnational governments and implementing organisations
Using generative AI to analyse large datasets for policy making and evaluation. This allows the government to better respond to societal needs and test the effectiveness of current measures.	Principles 4. Sustainability and prosperity	2024 and beyond	Ministry of Economic Affairs and Climate Policy in cooperation with subnational governments and implementing organisations
We are exploring the added value of responsible generative AI in promoting transparency and improving our information management as a (central) government. As well as the potential role of generative AI in solving legacy problems.[5]	Principles 2. Justice and 3. Wellbeing and autonomy	2024 and beyond	Ministry of Economic Affairs and Climate Policy (in cooperation with subnational governments and implementing organisations)

European and international cooperation
Generative AI is a cross-border phenomenon, a (geo)political issue with far-reaching implications for the international order. International cooperation is necessary, particularly with like-minded countries, on aspects such as human rights, personal security, and international security. Talks on the international governance of AI are expected to intensify in the coming months and years.

The Netherlands has valuable insights and networks to contribute actively to the international governance of AI due to its active role in cyber governance and standard setting, non-proliferation and disarmament, and its leading Research & Development position. The Netherlands will participate in this with a strong values-driven approach.
The US and China are global players when it comes to AI capabilities. The EU generally performs well, although it falls short in terms of fostering a productive (generative) AI ecosystem for companies. The Netherlands has a robust scientific foundation for AI research and, within the EU, is among the countries with specialisations that make them significant global players.[6] In the field of generative AI, the United States is the leading player, followed by China at a distance. The versatility of generative AI offers strategic choices for digital open strategic autonomy. These choices include strengthening the political-economic foundation, mitigating risky strategic dependencies, and enhancing the EU's geopolitical capacity to act. In October 2023, the government published the Agenda Digitale Open Strategische Autonomie (DOSA), which places a high priority on AI as a policy area.[7]

Scale is essential in generative AI: half a system delivers much less than half the value. In a competitive market where the best systems are seeing an increase in users and investment, success is determined by scale, which includes data, computing power, and investment. Europe provides an opportunity to achieve this economy of scale. Initiatives are also being developed at the EU level to enhance the knowledge and innovation position on generative AI. One such initiative is the Alliance for Language Technologies EDIC (ALT-EDIC).[8] This initiative is coordinated by France and includes several member states. The government is currently considering whether to join this initiative. In addition to cooperation, the objectives of this EDIC are to preserve linguistic and cultural diversity in Europe, to strengthen technological leadership and strategic autonomy, to respect European norms and values, and to raise awareness.

5. *See also activities on information governance and legacy approaches in the 'Geactualiseerde Werkagenda' (2024): https://open.overheid.nl/documenten/f3d07837-02d3-4523-84a7-65dad72ddad5/file*
6. *See the WRR report 'Opgave AI' (2021) 'De nieuwe systeemtechnologie', for a detailed explanation.*
7. *https://open.overheid.nl/documenten/5cb9749c-7efa-40db-9328-5da7fa5fcb7c/file*
8. *An EDIC is a legal framework in the EU that assists member states in establishing and executing multi-country projects. ALT-EDIC is one of the EDICs under preparation. See also: ALT-EDIC (europa.eu)*

The discussion regarding generative AI is also active in other EU networks. For instance, the AI Data Robotics Association (ADRA) has established a generative AI task force. Its purpose is to identify opportunities for Europe in generative AI and provide advice to the European Commission regarding investments from Horizon Europe, Digital Europe, and other similar instruments. The AI Alliance Forum, an AI platform aimed at entrepreneurs and policymakers to jointly define the way forward for European AI innovation, has urged the Commission to commit to multimodal AI.[9]

The development of the next generation of powerful AI models will require a significant investment in AI infrastructure to meet the increasing need for computing power and AI chips. This is a prerequisite to fulfil the Netherlands' ambition to be a forerunner in (responsible) generative AI and to be competitive in this field.[10]

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
The Netherlands is a participant in the EuroHPC partnership under Horizon Europe for high-performance computing (HPC). This allows Dutch companies and knowledge institutions to take part in European projects on HPC and quantum computing.	Principles 4. Sustainability and prosperity	2023	Ministry of Economic Affairs and Climate Policy
Netherlands is dedicated to enhancing the international rule of law. The government is committed to ensuring that secure AI deployment, driven by values, becomes the norm worldwide. Part of this involves contributing to international organisations that are working to establish guidelines for AI, as is currently happening within the UN. Pre-deployment audits of advanced models should be advocated, along with the establishment of rules for making the model weights of large models available . This provides an opportunity to address issues that the Netherlands cannot regulate unilaterally, such as the development of unsafe AI outside the country or the working conditions under which model fine-tuning often occurs.	Principles 1. Safety and 3. Wellbeing and autonomy	Ongoing	Ministry of the Foreign Affairs

9. *Multimodal AI is a field of AI that involves processing and interpreting information from multiple sensory modalities, such as images, text, audio and video. More information:* The AI4Media Strategic Research Agenda on AI for the Media Industry | Futurium (europa. eu)
10. open.overheid.nl/documenten/5cb9749c-7efa-40db-9328-5da7fa5fcb7c/file

Action summary	Fulfils principle(s)	Timeline	Owner
The Netherlands aims to establish global standards for the military use of generative AI, building on the progress made during the REAIM Summit in The Hague.[11]	Principles 1. Safety and 3. Wellbeing and autonomy	Ongoing	Ministry of Foreign Affairs in cooperation with Ministry of Defence
The government is implementing the DOSA Agenda, in which AI is one of the identified themes where there is a strategic dependency.[12]	All principles	2023-2024	Ministry of Economic Affairs and Climate Policy (in cooperation with other ministries)
A new round of risk assessments on critical technologies, including AI, has been launched at European level as part of the European Economic Security Strategy. These oversee the prevention of knowledge leakage and the management of risks to innovation capacity related to technology security. A questionnaire on this subject was completed at national level by the Ministry of Economic Affairs and Climate Policy. The Commission anticipates having initial results available in Q1.[13]	Principles 1. Safety	Ongoing	Ministry of Economic Affairs and Climate Policy

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
The Ministry of Economic Affairs and Climate Policy is currently exploring the possibility of joining the Alliance for Languages Technologies European Digital Infrastructure Consortium (ALT-EDIC). The ALT-EDIC brings existing open-source LLMs for use by industry players and SMEs, with a focus on mitigating bias.[14] In addition, the ALT-EDIC serves as a fund to encourage new LLMs and foundation models.	Principles 2. Justice and 4. Sustainability and prosperity	2023-2024	Ministry of Economic Affairs and Climate Policy

11. *Among other things, there will be a 'Global Commission AI' to promote mutual awareness worldwide, clarify what is meant by AI in the military domain, and determine how to achieve its responsible development, production, and application. See also:* https://www.rijksover-heid.nl/ministeries/ministerie-van-buitenlandse-zaken/nieuws/2023/02/16/call-to-action-verantwoord-gebruik-ai-in-het-militaire-domein.
12. https://open.overheid.nl/documenten/5cb9749c-7efa-40db-9328-5da7fa5fcb7c/file *(see also p. 26/29).*
13. *Parliamentary paper 22112-3826, no. 3826.*
14. *Open-source LLMs are transparent, which benefits the insightfulness of the models involved. This development approach can also promote ethical data curation to prevent bias, as well as privacy and copyright violations.*

b Closely monitoring all developments

It is very important to closely monitor developments in generative AI in general. There will also be a focus on specific issues: the impact on employment, democracy, sustainability and the climate. In doing so, the government will continue with the open approach used in the preparation of this vision.[15] This means that policy remains in close contact with subnational governments, implementing organisations, knowledge institutes, commercial parties, stakeholder organisations, employers, employees and citizens.

Monitoring developments around (generative) AI

The social impact of generative AI is expected to become more apparent in the future. Therefore, it is crucial to establish a system to monitor its long-term implications properly. This applies to both technological developments and the effects of generative AI on various sectors of our society and economy.[16] In this context, the WRR also highlights the importance of involving civil society in monitoring and keeping track of consequences of further embedding (generative) AI in Dutch society.

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
Encouraging the invitation of expertise from abroad (including outside the EU) to contribute to research programmes related to (generative) AI.	Principles 2. Justice and 4. Sustainability and prosperity	Ongoing	Ministry of Education, Culture and Science
In January 2023, the AP established the new Algorithms Coordination Directorate (DCA) to enhance coordinating supervision of algorithms. This includes proactively identifying and analysing, and gathering and sharing knowledge about, cross-sectoral and overarching risks and impacts of the development and use of algorithms, including generative AI.	Principles 1. Safety, 2. Justice and 3. Wellbeing and autonomy	2023-2027	Dutch Data Protection Authority
Working with industry to identify where barriers to innovation exist and where government can contribute, including through the Top Sector ICT within the Mission-Driven Innovation Policy 2024-2027.	Principles 4. Sustainability and prosperity	Ongoing (in 2024 first update progress)	Ministry of Economic Affairs and Climate Policy

---

15. *See Appendix 1 for more information on the open-ended approach used to arrive at this vision.*
16. *This is also pointed out by the Rathenau Institute in its scan on generative AI (2023).*

Action summary	Fulfils principle(s)	Timeline	Owner
The government is exploring the possibilities of a top-level AI advisory body (or Rapid Response Team AI (RRT-AI)).[17] This working/expert group can advise the government on key developments in (generative) AI.	All principles	2023-2024	Ministry of the Interior and Kingdom Relations, Ministry of Economic Affairs and Climate Policy and Ministry of Education, Culture and Science
There will be a government-wide inventory and monitoring of initiatives, developments and uses of generative AI by governments and (semi-) public organisations, for which we will develop intergovernmental indicators. This includes a focus on both opportunities and risks. The results will also be regularly shared with your House of Representatives of the Netherlands.[18]	All principles	2023-2025	Ministry of Economic Affairs and Climate Policy (in cooperation with subnational governments and implementing organisations)
As a government, it is crucial to stay informed about the current and future capabilities of AI systems, including generative AI. Therefore, we keep ourselves updated on the latest developments in this field. In this process, we closely examine opportunities to leverage existing (international) research and where we can complement the current state of the art, such as Epoch AI or Metaculus research[19], organisations and methodologies with a good track record in predicting technological breakthroughs and developments.	All principles	Ongoing	Ministry of the Interior and Kingdom Relations and Ministry of Economic Affairs and Climate Policy

Steering socio-economic transitions in the field of work and income
As stated in Chapter 3, the use of generative AI applications in professional and industrial contexts may cause shifts in the labour market and income, and affect job quality. Matching the skills of the workforce to the labour market of the future is essential. The transitions require consideration of the financial sustainability and adequacy of the current social security system.

---

17. *In this context, it is also worth noting the motion put forward by members Dekker-Abdulaziz and Rajkowski, which was adopted in October 2023:* https://www.tweedekamer.nl/kamerstukken/moties/detail?id=2023Z17682&did=2023D42892
18. *To achieve this, meetings are being organised with subnational governments, among others.*
19. *See also:* Exploring Metaculus's AI Track Record

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
The Social and Economic Council (SER) will assess the impact of AI, including generative AI, on labour productivity, as well as the quantity and quality of work.	Principles 2. Justice and 3. Wellbeing and autonomy	2024-2026	Ministry of the Interior and Kingdom Relations, Ministry of Economic Affairs and Climate Policy, Ministry of Education, Culture and Science and Ministry of Social Affairs and Employment

Sustainable development and use (generative) AI
The climate challenges are extensive. Generative AI has the potential to make a positive contribution to this. Currently, there are no standardised methods to measure the climate impact of (generative) AI. Currently, the development and deployment of generative AI creates a larger ecological foot- print (see also Chapter 3). As a government, it is important to consider sustainability when deploying generative AI. This means refraining from deploying the technology if it proves to be harmful to people and the planet. Therefore, it is important to also consider developments that promote the sustainability of generative AI. This could involve implementing more ener- gy-efficient training processes and relying on publicly available and accountable LLMs to reduce the frequency of the training process. Sustainable generative AI involves exploring oppor- tunities for how it can contribute to addressing climate chal- lenges through innovation and research.

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
Further examine the sustainability aspect in the development and use of generative AI (by the government) and take measures to reduce negative impacts where possible.	Principles 3. Wellbeing and autonomy and 4. Sustainability and prosperity	2024	Ministry of the Interior and Kingdom Relations
Encourage further research on how generative AI can positively contribute to various climate challenges in government.	Principles 3. Wellbeing and autonomy and 4. Sustainability and prosperity	2024-2025	Ministry of the Interior and Kingdom Relations

Democracy and (generative) AI

Generative AI poses several risks to democracy and our democratic rule of law. The accelerated production and dissemination of disinformation or criminal content, such as threats and hate speech, is of particular concern. Therefore, it is crucial to continue monitoring how generative AI affects democracy and what policy adjustments are necessary to renew and protect it. The government has previously indicated that it should review the policy focus on disinformation in light of new technologies.[20] At the same time, we should also utilise generative AI to enhance democracy. One example is to encourage research into language models for Frisian and Papiamento. This will make communication with and for citizens more understandable and inclusive. This statement contributes to the efficient delivery of services and promotes a more accessible and inclusive government.

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
The government-wide strategy progress letter addresses the effective tackling of disinformation and whether the policy focus should be adjusted in view of new technologies.	Principles 1. Safety, 2. Justice and 3. Wellbeing and autonomy	2024	Ministry of the Interior and Kingdom Relations

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
Encourage further research on how generative AI can contribute positively to democracy, for example in the area of civic participation.	Principles 2. Justice and 3. Wellbeing and autonomy	2024-2025	Ministry of the Interior and Kingdom Relations
Encouraging the development and improvement of (open and public) language models trained on languages such as, for example, Frisian, Papiamento and sign language.	Principles 2. Justice and 3. Wellbeing and autonomy	2024-2025	Ministry of the Interior and Kingdom Relations
Exploring how generative AI can support to clarify and make communication with citizens more inclusive.	Principles 3. Wellbeing and autonomy and 4. Sustainability and prosperity	Ongoing	Ministry of Economic Affairs and Climate Policy in cooperation with subnational governments

20. *Parliamentary paper II, 2023-2024, 35 165, no 64.*

c Shaping and applying laws and regulations

Laws and regulations are crucial to promote confidence in generative AI. Fortunately, generative AI operates within a legal framework. As mentioned in Chapter 4, generative AI needs to comply with legal frameworks. To ensure future-proof standards, they are often formulated in a broad and open manner. This can create uncertainty regarding the interpretation of laws and regulations. There is a role not only for regulators and judges, but also for central government working with subnational governments. With rapidly developing technologies, such as generative AI, it is crucial that regulations remain technology-neutral and adaptable.

Clear rules should be established for developers and providers of generative AI to mitigate or address any societal risks and challenges associated with the further growth of generative AI. In a digital society, citizens and businesses require legal certainty and must be able to rely on the government to establish appropriate frameworks and rules, as well as provide clear oversight of them. This should address both how existing legal instruments, as well as upcoming laws and regulations, are able to create a playing field in which generative AI can be developed and used in a safe, fair, lawful and transparent way. Clear national, European, and international frameworks can accelerate innovation and facilitate the development of new solutions. The European AI Act is the basis for this. That is why we are working hard on the government-wide implementation of the AI Act in 2024; parties must be well prepared for this by providing good information and guidance to stakeholders, among other things.

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
Implementation of the European AI Act, including • Appoint supervisors via implementation act • Consultation and parliamentary consideration of implementation act • Establishing European supervision • Educating (subnational) governments, industry and other stakeholders • Encourage guidance and standard setting including through regulators • Setting up Dutch regulatory sandbox • Developing European standards • Adoption of European implementing acts	All principles	2024-2027	Ministry of Economic Affairs and Climate Policy and Ministry of the Interior and Kingdom Relations in cooperation with Ministry of Justice and Security, Ministry of Education, Culture and Science, Ministry of Social Affairs and Employment, Ministry of Infrastructure and Water Management, Ministry of Health, Welfare and Sport, Ministry of Agriculture, Nature and Food Quality, Ministry of Foreign Affairs, Ministry of Finance, and Ministry of Defence
The government is actively participating in the negotiations of the AI Treaty in CAI (Committee on AI) of the Council of Europe.	Principles 1. Safety and 2. Justice	2023-2024	Ministry of the Interior and Kingdom Relations (in cooperation with JenV)

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
Ongoing monitoring (in a national and European context) whether the current legal framework (e.g. copyright and GDPR) is sufficient. By means of further research and surveillance signals.	All principles	Ongoing (via the Work Agenda)	Government-wide

d Increase knowledge and skills

The government is actively promoting the acquisition of knowledge and skills. This allows us to take full advantage of the opportunities provided by generative AI. Adequate commitment to knowledge and skills also benefits one's grasp of technology, promotes social equity, and encourages participation. As a government, we will set a positive example. At the same time, the government recognises the significance of enhancing generative AI knowledge and skills throughout society. This is achieved by supporting education to enable an adequate response to technological developments. In addition, the government is fully committed to developing the digital knowledge and skills necessary to responsibly handle generative AI.[21]

Government leading by example

The Dutch government has a commendable role in the responsible and safe development, procurement, and deployment of generative AI. Officials with the appropriate skills and expertise are required to responsibly procure, develop, or deploy this technology in various aspects of their work. Therefore, it is crucial for the government that employees possess the appropriate knowledge and skills, acquired through training, and operate within the correct frameworks. This will provide employees with the tools to use the technology responsibly and legitimately in their own work, the right conditions, frameworks and policies to procure the technology, and an awareness of the challenges and opportunities of generative AI.

It is important to consider that the implementation of public services can have an impact on residents' trust in the government, either positively or negatively. The Dutch government must therefore exemplify its role in an integrated manner.

As a government, we recognise the importance of innovation and experimentation in harnessing generative AI for public values. All generative AI applications must comply with relevant laws and regulations.[22] To determine whether a specific form of generative AI deployment is feasible, a risk analysis should be conducted for each unique case before use. These are a Data Protection Impact Assessment (DPIA) and an algorithm impact assessment (such as an Impact Assessment Fundamental Rights and Algorithms (IAMA)), which identifies risks and mitigation measures. The results of this should be submitted to the (departmental) Chief Information Officer and the Data Protection Officer for advice before deployment. The above points apply when using or (re)developing an open-source generative AI application. In the context of the 'Wet open overheid' (Open Government Act) (Woo) and encouraging transparency, the policy guideline "open (source), unless" applies.[23] Non-contracted generative AI applications[24] generally do not demonstrably comply with applicable privacy and copyright laws. Therefore, its use by (or on behalf of) central government organisations is not permitted where there is a risk of breach of the law, unless the provider and user can demonstrate compliance with applicable laws and regulations. Contracted generative AI applications should also comply with the General Government Terms and Conditions for IT Contracts 2022 and departmental procurement conditions (if they prevail). When utilising a generative AI application, it is crucial that employees receive proper guidance on how to use this technology responsibly. This can be done through training or guidelines for responsible use.

The increasing prevalence of generative AI in society holds great promise for individuals and organisations. Based on ethical conversation, further risk analyses and risk categorisation in line with the future AI Act, the (im)possibilities of the deployment of generative AI by central government organisations in coming years will be determined. This position does not categorically ban the technology, but rather enforces applicable laws and regulations. The use is not prohibited, but rather regulated. This will sustain and encourage experimentation with the technology.

---

21. *In this context, it is also worth noting the 'I-strategie Rijk' (2021-2025) and the emphasis on 'I-vakmanschap':* I-strategie Rijk I-strategie Rijk 2021-2025 - Digitale Overheid.
22. *See:* Parliamentary letter on preliminary position for central government organisations when using generative AI | Parliamentary paper | Rijksoverheid.nl
23. *The government acknowledges that open-source models can provide solutions for transparency and explainability. It encourages the use of open source by applying the 'open source, unless' principle in procurement and development. It should be noted, however, that transparency should not come at the expense of the safety of generative AI models (see also principle 1).*
24. *For instance, publicly available forms of generative AI developed by large technology companies (often offered online).*

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
A guide is currently being developed to provide direction to government organisations, including central government, when deploying generative AI. This guide outlines the possibilities and limitations of using generative AI in government. We are working together to transpose this guide into the context of decentralised authorities.[25]	Principles 1. Safety and 2. Justice	2023-2024	Ministry of the Interior and Kingdom Relations (in cooperation with subnational governments)
The General Intelligence and Security Service analyses and produces publications for government officials and society on risks and opportunities for increasing resilience (including generative AI).	Principles 1. Safety	Ongoing	General Intelligence and Security Service (AIVD)
The National Cyber Security Centre (NCSC) and General Intelligence and Security Service are committed to increasing technical knowledge on the subject of AI and resilience against cyber threats in this new domain and resilience against unwanted use of LLMs.[26] During the process, organisations are informed about urgent developments and provided with practical tools for the safe development of AI.[27]	Principles 1. Safety and 4. Sustainability and prosperity	2024	National Cyber Security Centre and General Intelligence and Security Service

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
To reduce dependence on temporary deployment, it is recommended that the (central) government invests in an in-house knowledge base and expertise on generative AI. In doing so, we explore the added value of a generative AI knowledge hub within the (central) government.[28]	Principles 2. Justice and 3. Wellbeing and autonomy	2024-2025	Ministry of the Interior and Kingdom Relations
Officials working on (generative) AI learn to recognise and explore moral ethical issues on this topic. To achieve this objective, it is possible to actively seek cooperation with 'programma Dialoog & Ethiek' (Dialogue and Ethics programme).	Principles 1. Safety and 2. Justice	2024-2025	Ministry of the Interior and Kingdom Relations

25. *Prior to the guide, if necessary, a risk analysis will be carried out on specific use case(s) through a DPIA and IAMA.*
26. AI: 'Cruciaal moment in de geschiedenis of een hype?' | Expert blogs | Nationaal Cyber Security Centrum (ncsc.nl)
27. AI-systemen: ontwikkel ze veilig | Publication | AIVD
28. *See also:* https://generatieveai.pleio.nl/

Action summary	Fulfils principle(s)	Timeline	Owner
Through RADIO ('Rijksacademie voor Digitalisering en Informatisering Overheid', National Academy for Digitalisation and Informatisation Government), engage in knowledge sharing on the (im)possibilities of secure use of generative AI.	Principles 1. Safety and 2. Justice	2024	Ministry of the Interior and Kingdom Relations
Invest in the knowledge and skills of civil service professionals and elected representatives at all levels of government through courses and workshops. Facilitate intergovernmental knowledge sharing on opportunities for the safe use of generative AI by exchanging practical experience and knowledge.	All principles	Ongoing	Ministry of the Interior and Kingdom Relations (in cooperation with subnational governments)
Drafting, developing or refining (intergovernmental) procurement conditions with a view to generative AI.	All principles	2024-2025	Ministry of the Interior and Kingdom Relations (in cooperation with subnational governments)

Investing in talent and computing power

As previously mentioned, generative AI is a cross-border phenomenon. The Dutch government considers it important, especially in the European context, to stimulate an ecosystem for (generative) AI through public-private cooperation and investment in this ecosystem, as well as to invest in (open) public alternative generative AI (see also action 'Innovating with generative AI'). Therefore, we investigate the required investment in large-scale scientific and technological infra-structure, such as supercomputers and computing power, at both national and EU levels to remain competitive in the field of LLMs and other forms of generative AI. The explicit focus is on developing and retaining AI talent to create generative AI that aligns with European standards and values. This also adds value to Europe's digital open strategic autonomy.

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
Invest in adequate computer infrastructure, including through SURF, to carry out appealing scientific projects in the Netherlands and the EU.	Principles 4. Sustainability and prosperity	Ongoing	Government-wide
The Netherlands is committed to investing in innovative projects and research in the field of safe and responsible AI. This includes efforts to enhance the interpretability and transparency of AI models.	Principles 1. Safety and 4. Sustainability and prosperity	Ongoing	Government-wide

Increasing knowledge and skills (in education

Deploying generative AI requires different skills. Both in using generative AI tools, as well as in assessing the content generated by this technology. This requires a greater dedication to media literacy for various target groups, with particular attention to awareness and to assessing the reliability of (generated) content. The increasing prevalence of generative AI has made skills such as critical and structured thinking more important in education. Generative AI can rapidly produce significant quantities of text, imagery, audio and computer code. Therefore, it is crucial to evaluate and assess this content objectively and with care.

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
The National Education Lab AI (NOLAI) conducts research into the pedagogical, societal and social consequences of generative AI.[29]	Principles 2. Justice and 3. Wellbeing and autonomy	2023-2024	Ministry of Education, Culture and Science and Ministry of Economic Affairs and Climate Policy
Promoting the responsible use of (generative) AI applications for societal challenges at higher vocational education knowledge institutions through 'AI in Action'.[30]	Principles 2. Justice and 4. Sustainability and prosperity	N/A	Ministry of the Interior and Kingdom Relations and Ministry of Education, Culture and Science in cooperation with universities of applied sciences
The government is committed to ensuring a permanent place for digital literacy in the national curriculum for primary and secondary education.[31]	Principles 2. Justice and 3. Wellbeing and autonomy	2023 and beyond	Ministry of Education, Culture and Science and Ministry of the Interior and Kingdom Relations

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
Enhancing digital skills[32] and digital awareness of people in the Netherlands so that they can consciously, critically and actively engage with AI w.[33]	Principles 2. Justice and 3. Wellbeing and autonomy	Ongoing	Ministry of the Interior and Kingdom Relations and Ministry of Education, Culture and Science

29. *See also:* https://www.rijksoverheid.nl/documenten/kamerstukken/2023/07/06/visiebrief-digitalisering-in-het-funderend-onderwijs
30. *See also: Agenda 'Coalities voor de digitale samenleving':* https://open.overheid.nl/documenten/1oc88500-cdb5-4815-bd00-c915a5242ea3/file
31. *In this context, see also the 'Expertisepunt digitale geletterdheid' (Digital Literacy Expertise Centre) launched in autumn 2023.*
32. https://www.rijksoverheid.nl/documenten/rapporten/2023/06/19/digitale-vaardigheden-van-nederlanders
33. *See also line 1.1 of the Value-Driven Digitalisation Work Agenda: Enhancing digital skills and knowledge.*

e Innovating with generative AI

With a government that aims to take the lead, comes a government that explores safe and responsible generative AI. This approach can help to reduce dependencies and identify potential risks and opportunities in specific applications. The aim is to be able to take full advantage of the economic, scientific and other societal opportunities that generative AI offers us in a responsible way. It is important for public organisations to collaborate with Dutch industry to enhance the knowledge and innovation base for the development and deployment of generative AI.

Due to the potential impact of concentrated development of powerful generative AI, it is crucial to create an environment in the Netherlands that encourages experimentation, testing, and scaling up of reliable and transparent generative AI models and tools. This could include validation or bias detection. This highlights the significance of high-quality datasets, particularly those in Dutch, as a crucial foundation for generative AI models. To gain further knowledge and experience in the validation of AI[34], , the government has established a 'Rijks AI-validatieteam' (Government AI Validation Team). One of the team's aims is to make the risks and opportunities of generative AI measurable. The team comprises software engineers who will collaborate with policymakers to create practical tools for validating (generative) AI.

The Netherlands has a great example of responsible innovation with generative AI in the realisation of GPT-NL.[35] TNO, NFI, and SURF, non-profit organisations, will collaborate to develop a language model that aligns with Dutch and European values, ensuring transparent, fair, and verifiable use of AI while respecting data ownership. Additionally, a connection to the national supercomputer at SURF will be established. The aim of GPT-NL is to decrease reliance on commercial entities and offer a responsible and transparent alternative to them. GPT-NL is a virtual facility that partners can access to contribute data and knowledge towards the development of language models and applications. The facility is open to partners in various fields, including security, health, education, and government services.

---

34. *Parliamentary papers II 2022/23, 26643, no 1056*
35. *In November 2023, €13.5 million was committed for the development of GPT-NL through 'Faciliteiten Toegepast Onderzoek' of the Netherlands Enterprise Agency (RVO) and the Ministry of Economic Affairs and Climate Policy.*

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
Encouraging the development of (open) Dutch and European LLMs in line with public values. GPT-NL is a starting point for this.[36]	All principles	2023-2026	Government-wide
The Open State Foundation (OSF) is currently developing an LLM programme on Dutch open government information, which includes publicly available parliamentary documents and speeches. This project is being funded by a grant from the Ministry of the Interior and Kingdom Relations. One of its aims is to identify the opportunities and risks of current language models for democracy.	Principles 2. Justice en 4. Sustainability and prosperity	2023-2024	OSF (in collaboration with the Ministry of the Interior and Kingdom Relations)

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
Exploring the establishment of a secure and usable public national AI (testing) facility for responsible (generative) AI.	All principles	2024-2025	Government-wide
Enabling the responsible use of generative AI in a safe government environment. One of the ways we do this is through various pilot projects.	Principles 1. Safety, 2. Justice and 4. Sustainability and prosperity	Ongoing	Ministry of the Interior and Kingdom Relations (in cooperation with subnational governments)
In 2024, AiNed will launch InnovationLabs. InnovationLabs are partnerships between public and private entities that aim to develop AI innovations, with a particular focus on supporting SMEs, start-ups, and scale-ups. The AiNed InnovationLabs aim to accelerate the introduction of AI innovations to the market by bringing together expertise in the field of (generative) AI from knowledge institutions and (deep) tech companies, while also promoting knowledge sharing.[37]	Principles 4. Sustainability and prosperity	2024	Ministry of Economic Affairs and Climate Policy

36. https://www.tno.nl/nl/newsroom/2023/11/nederland-start-bouw-gpt-nl-eigen-ai/
37. Announcement AiNed InnovatieLabs (2024 Stichting AiNed call) - AiNed

Action summary	Fulfils principle(s)	Timeline	Owner
Committing to an AiNed call focusing, among other things, on ELSA (Ethical, Legal and Societal) aspects of the AI Act for existing and rapidly developing AI technologies such as generative AI.[38]	All principles	2024	Ministry of Economic Affairs and Climate Policy
A 'Rijks AI-validatieteam' (Government AI Validation Team) facilitates publicly available benchmarking and tooling (such as bias detection, based on democratic input, for instance) to provide guardrails for responsible generative AI in the Netherlands.	Principles 1. Safety, 2. Justice and 4. Sustainability and prosperity	Ongoing (from 2024)	Ministry of the Interior and Kingdom Relations (in cooperation with subnational governments)
Including ethical frameworks, and possibly tools, around the responsible use of generative AI in the further development of the Implementation Framework for Algorithms (IKA)[39]. This includes support for developers and users in the implementation of the AI Act in provisions touching on generative AI.	Principles 2. Justice and 4. Sustainability and prosperity	2024-2026	Ministry of the Interior and Kingdom Relations (in cooperation with subnational governments)
Encouraging and/or researching disclosure methods for transparency about the provenance and veracity of AI-generated content. 'Water-marking' and 'earmarking' AI content using cryptography are examples of techniques that can be used.	Principles 1. Safety, 2. Justice and 3. Wellbeing and autonomy	2024-2026	Ministry of the Interior and Kingdom Relations

38. Announcement AiNed ELSA Labs (2024 NWO call) - AiNed
39. https://www.rijksoverheid.nl/documenten/rapporten/2023/06/30/implementatiekader-verantwoorde-inzet-van-algoritmen

f Strong and clear supervision and enforcement

Developers, policymakers, and regulators at both European and national levels should remain vigilant for any unintended consequences that may arise from generative AI in the future. A proactive approach is crucial in this context, with regulators and the public sector providing clear frameworks from the outset to guide the development of generative AI and prevent the emergence of unwanted (generative) AI.

Sectoral regulators play a crucial role in ensuring effective control over (generative) AI, in accordance with relevant laws, regulations, and public values. The Algorithms Coordination Directorate (DCA) of the Dutch Data Protection Authority (AP), established in 2023, has coordinating responsibilities in these areas.[40]

To ensure effective supervision, supervisors must acquire knowledge and information during development and implementation to monitor progress and make necessary adjustments. Effective collaboration is crucial, for instance, via the Digital Supervisors Cooperation Platform (SDT) or the Inspection Council. The ability to intervene in a timely and effective manner when violations and undesirable effects occur is a collaborative effort between regulators, judicial bodies, politics, and society. Openness is crucial to allow for analysis and monitoring of relevant developments in generative AI by science, journalism, citizens, and politics.

With the increasing number of generative AI applications in the coming years, it is essential to continually assess whether regulators have the knowledge and skills, capacity and resources to perform their duties effectively now and in the future.[41] This could also consider using regulators' practical knowledge of (generative) AI for legislative advice. This aligns well with a learning approach, which involves monitoring in the future to ensure that relevant laws and regulations remain effective and protected in light of advancements in generative AI.

Actions in progress

Action summary	Fulfils principle(s)	Timeline	Owner
Implementing regulatory oversight of the AI Act	All principles	2024-2027	Ministry of Economic Affairs and Climate Policy and Ministry of the Interior and Kingdom Relations in cooperation with Ministry of Justice and Security, Ministry of Education, Culture and Science, Ministry of Social Affairs and Employment, Ministry of Infrastructure and Water Management, Ministry of Health, Welfare and Sport, Ministry of Agriculture, Nature and Food Quality, Ministry of Foreign Affairs, Ministry of Finance, and Ministry of Defence
The implementation of regulatory sandboxes from the AI Act in the Netherlands in conjunction with regulators.	Principles 1. Safety, 2. Justice and 4. Sustainability and prosperity	2023-2024	Ministry of Economic Affairs and Climate Policy (in cooperation with other departments, regulators)

40. *See also:* https://www.rijksoverheid.nl/documenten/kamerstukken/2022/12/22/kamerbrief-over-inrichtingsnota-algoritmetoezichthouder
41. *Rathenau Institute (2023), Generative AI: p. 38.*

Action summary	Fulfils principle(s)	Timeline	Owner
Promoting joint guidance and explanations (from regulators) and creating overview in existing and new legal frameworks (such as the AI Act) in the field of algorithms and (generative) AI.	Principles 4: Sustainability and prosperity	Ongoing (from 2023)	Dutch Data Protection Authority
The government is pushing Europe to include AI applications within the scope of the Digital Markets Act and contribute to effective enforcement.	Principles 4. Sustainability and prosperity	Ongoing	Ministry of Economic Affairs and Climate Policy

New actions

Action summary	Fulfils principle(s)	Timeline	Owner
Continued commitment to legislative advice from regulators on legal frameworks for generative AI.	All principles	Prior to the implementation of laws and regulations and continuously thereafter	Government-wide

6 Follow-up and conclusion

Generative AI is increasingly enhancing people's analytical and creative abilities. Generative AI is a component of wider digitalisation and traditional AI advancements. Generative AI is distinguished from traditional forms of AI by its scale, rapid development, and widespread availability.

The influence of technology is anticipated to become more significant in the future. The impact of generative AI is already becoming apparent based on scientific findings and expert predictions. Therefore, it is essential to continue to monitor and analyse the developments and consequences of generative AI.

Generative AI has both positive and negative effects. The success of the technology relies heavily on its development, application, and integration. Adequate governance of generative AI is therefore important. In line with wider digitalisation policies, such as the Value-Driven Digitalisation Work Agenda, this vision takes a value-driven approach. The government aims for generative AI applications and related technologies to enhance human wellbeing and autonomy, sustainability, prosperity, justice, and security. Its ambition is to lead in this area in Europe and globally with Europe. The government

aims to establish a framework for the development and use of responsible generative AI that is independent of commercial or geopolitical power blocks.

To achieve this vision, specific policy actions have been outlined. The progress will be reported to your House of Representatives of the Netherlands by the end of 2024. Attention will be given to the necessity of any new actions or policies, also in view of the new government. Rapid developments call for an iterative and learning approach.

To achieve the principles set out in this vision, concrete actions will be monitored over the coming years. These actions focus on cooperation, close monitoring of all developments, design and application of laws and regulations, increasing knowledge and skills, innovation with generative AI, and strong and clear supervision (and enforcement). The success of these actions depends on the further development of a functional (generative) AI ecosystem in the Netherlands and Europe. This way, our country's role as one of the European leaders in safe and fair (generative) AI can develop, and people in the Netherlands can actually benefit from this technology.

Appendix 1: Vision process approach

The approach taken to arrive at the government-wide vision of generative AI is detailed below.

a   Open approach
A wide range of stakeholders contributed to arrive at a government-wide vision for generative AI. The government acknowledges and appreciates the contributions made by individuals and organisations towards achieving this vision in recent months.

As of May 2023, various measures have been implemented to establish an inclusive and widely accepted government-wide vision of generative AI, tested across multiple sectors and domains within Dutch society. The approach involved gathering input from multiple experts through various sessions, sharing results with the public intermittently, and testing outcomes through an online Pleio community and other channels.

- As of May 2023, various measures have been implemented to establish an inclusive and widely accepted government-wide vision of generative AI, tested across multiple sectors and domains within Dutch society. The approach involved gathering input from multiple experts through various sessions, sharing results with the public intermittently, and testing outcomes through an online Pleio community and other channels.
- Several (sector) sessions took place between June and November 2023. These sessions were held in various sectors, including public administration, mobility, healthcare, employment, and the economy. During these sessions, the vision was developed by gathering input and ensuring that all necessary aspects were included. These talks will continue in 2024.

- In autumn 2023, ECP organised meetings with the media, higher education, healthcare, and police.[1] These sessions focused on the ethical dilemmas surrounding the application of generative AI.
- In cooperation with the Netherlands AI Coalition (NL-AIC), a number of sessions with residents of the Netherlands have been organised. The aim of these sessions was to raise awareness about generative AI technology and to gather public opinion on its impact on society and the role of government in its implementation.

- During the vision process, outcomes and insights were shared multiple times, for instance, through Pleio.[2]
- To guarantee that the vision encompasses the widest possible outlook on generative AI, a working group on generative AI was established. The group includes representatives from various ministries, provinces (IPO), and municipalities (VNG).
- In addition, a sounding board group was established with members representing different social perspectives.[3]

b   Catshuis session
A Catshuis session on generative AI was held for members of the government on 6 September 2023. In it, the question of how the Netherlands can position itself as a country and a responsible testing ground in the field of generative AI was considered, with much attention to both the ethical aspects and the opportunities this technology offers. The Catshuis session highlighted that the Netherlands can make a difference in the responsible use of AI in the military domain, following the REAIM Summit initiative.

c   Techscan Rathenau
The Rathenau Institute was asked to analyse generative AI using the 'techscan method' due to its independence, knowledge, and expertise at the intersection of policy and digitalisation.[4] This report was published in December 2023. Its threefold purpose was to: (1) Identify opportunities and risks of generative AI's social impact from a public-values perspective at an early stage. (2) Evaluate existing policies to capitalise on opportunities and address risks. (3) Identify potential courses of action.

This tech scan will also fulfil the commitment made by the State Secretary for Ministry of the Interior and Kingdom Relations to member Van Weerdenburg (PVV) on 22 March 2023 during the Committee debate on 'Digital infrastructure and the economy', to present a study on the impact of AI on society after the summer.

---

1. https://begeleidingsethiek.nl/cases/
2. https://generatieveai.pleio.nl/
3. *The sounding board group included representatives from: Bits of Freedom, CIO-Platform, IPO, FNV, NL-AIC, Politie, SER, VNG and VNO-NCW.*
4. *Rathenau Institute report on generative AI (2023):* https://www.rathenau.nl/sites/default/files/2023-12/Scan_Generatieve_AI_Rathenau_Instituut.pdf

Appendix 2: How is generative AI created?

The creation of generative AI models can be divided into three phases: pre-training, fine-tuning and deployment. These stages are described below.

The **pre-training phase** is a crucial step in training generative AI models. During this phase, the model receives training data from both public and closed sources. The availability of data is essential as it enables models to analyse and categorise a broad range of concepts, language structures, contextual nuances, and representations of the world. Not only text, but also audio, video and images can serve as data sources, and this data can be combined in a single model.

During the pre-training phase, the model optimises its parameters to detect accurate correlations and patterns in the training data. Generative AI models can have as many as a trillion parameters, requiring trillions of iterations to reach a desired value. The large size of the process from input to output reduces its transparency and contributes to the models' black box nature. The process of training AI models requires a significant amount of computing power and is frequently the bottleneck. Therefore, generative AI models are trained on specialised hardware. Rapid developments in hardware in recent years have made it possible for more and more parties to train larger, more complex AI models. It is important to note that the pre-training phase in this type of model requires minimal human intervention. This enabled rapid scaling.

Fine-tuning phase: The pre-training phase produces a basic model that is not yet suitable for widespread use. The basic model is refined by f inetuning, where the model learns to follow user instructions. Fine-tuning can also be employed to incorporate specialist knowledge or specific values and norms into the model. A recent trend in fine-tuning involves the implementation of Reinforcement Learning from Human Feedback. This method is complex. It evaluates (generative) AI models based on the helpfulness, fairness, and safety of their outputs, rather than their predictive abilities. Here, the outcome is assessed and labelled by people. Developers have now presented methods where AI models themselves assess output integrally for ethical aspects, such as Constitutional AI.[1]

To enhance the quality and quantity of training material, it is logical to consider the use of synthetic data after exhausting available open and closed data sources. Synthetic data is created by combining various data sources and utilising generative AI. This has an implicit accelerating effect. Thanks to faster and more advanced AI, it is now possible to generate high-quality training and fine-tuning data. High-quality AI results in higher-quality training material, which in turn leads to improvement of the AI system. The fine-tuning phase results in a model suitable for use.

Application phase: During the application phase, the model is made available to users. The model is reusable and can be duplicated, enabling companies with significant computing power to support tens of thousands to millions of users simultaneously. Whereas in the training phase it takes months to train a model, once applied, a model can provide answers within seconds.[2]

A model developer may decide to make the model **open source** after the application phase. This means that the source code of the model (and sometimes other components) is published so that it can be viewed, analysed, reused and built upon with custom modifications (*fine-tuning*). Models can be improved in this way, but they can also be made worse because, for example, there is no longer any control over further fine-tuning of the model. Making the code and parameters of generative AI models available does not change the fact that they remain black box models whose capabilities are not directly traceable. In addition, the amount of information actually disclosed about the model may vary from provider to provider. As such, open source does not necessarily equate to greater transparency, security or moral decency.

---

1. *Constitutional AI aims to develop generative AI that aligns with human values by automatically testing the outcomes of an AI model against principles that can also be democratically drawn up.*
2. *One example is Microsoft 365 Copilot.*

Appendix 3: Glossary of terms government-wide vision of generative AI

1. *Algorithm*
   A set of rules and instructions that a computer executes.

2. *Artificial general intelligence (AGI)*
   A technology that possesses intelligence across a wide range of domains, including reasoning, planning and learning, and performs at or above human levels in these skills.

3. *Artificial intelligence (AI) chatbot*
   Digital chatbots that can communicate via text and images in a way that closely resembles human interaction. ChatGPT is currently one of the most well-known AI chatbots.

4. *Artificial intelligence (AI) system*
   A machine-based system that deduces how to generate outputs, such as predictions, content, recommendations, or decisions that may affect physical or virtual environments, from the inputs it receives, for explicit or implicit purposes. Different AI systems vary in their degree of autonomy and adaptability after their deployment (OECD 2023).

5. *Special categories of personal data*
   Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, or the processing of genetic data, biometric data for the purpose of uniquely identifying a person, health data, data relating to a person's sexual behaviour or sexual orientation.

6. *Black box model*
   An (AI) model for which there is a lack of understanding of how the model's prediction came about and what the basis for the model formed is.

7. *Dark patterns*
   Interfaces, especially in online user interfaces, that can impair consumer autonomy, decision-making and choice. They often mislead, coerce or manipulate consumers, which can result in direct or indirect harm. Measuring this harm, however, can be difficult.

8. *Deepfake*
   A photograph, video or audio created with technology that shows or hears a person doing or saying something they did not actually do or say.

9. *Disinformation*
   Disinformation is the deliberate, often covert, dissemination of misleading information, with the purpose of damaging public debate, democratic processes, the knowledge economy or public health.

10. *Fine-tuning*
    During the 'fine-tuning' of an AI model, an already trained model is adapted to a specific task or dataset. Rather than training a model 'from scratch', an existing model is used.

11. *Foundation model*
    A foundation model is a basic machine learning model that serves as a foundation for further specialised models. A large language model (LLM) is a type of foundation model. An example of a foundation model is GPT-4, the foundation model for ChatGPT.

12. *Generative AI*
    A form of AI that uses complex algorithms to generate new content such as text, images, computer code or videos. ChatGPT is the best-known example of this.

13. *Hallucination*
    Information generated by a 'large language model' (LLM) that is factually incorrect. The answers generated by the model are not based on the given inputs or actual information from the training data.
    Hallucinations can be caused by various factors, including insufficient information in the training data, lack of context, or inconsistent and erroneous information in the training data.

14. *Jailbreaking*
    Designing prompts to intentionally exploit model biases in order to generate output that is inconsistent with the model's intended purpose. As an example, the model will answer questions that would not normally be answered by the model.

15. *Large language model (LLM)*
    A specialised type of AI model trained on large amounts of text to understand existing content and generate content.

16. *Machine learning*
A subfield of artificial intelligence that enables computers to learn from data. A machine learning algorithm learns from examples and experiences to discover patterns and rules in data.

17. *Model*
An AI model is the result of training an algorithm on data. An algorithm is a set of instructions, and a model is the specific output generated by following those instructions using certain data. GPT-4 is an example of an AI model, in this case a *large language* model.

18. *Model collapse*
A phenomenon that can occur when LLMs are trained with 'contaminated' data; a database containing AI-generated data. Model collapse occurs when a language model generates inaccurate and repetitive output due to corrupted data.

19. *Model parameters*
Customisable settings in AI models that decide how an LLM generates output. Model parameters affect the quality, diversity and creativity of the output. Model parameters are created in a variety of ways, including mathematical calculations and human intervention.

20. *Open source*
Open source publishing of a generative AI model means that the source code of the model (and sometimes other components) is published for viewing, analysing, reusing and building on with own modifications (*fine-tuning*).

21. *Personal data*
Any information about an identified or identifiable natural person ('the data subject'). An 'identifiable' refers to a natural person who can be identified, directly or indi-

rectly, through an identifier such as a name, identification number, location data, online identifier, or through one or more elements that characterise the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person.

22. *Pre-training*
During pre-training, an AI model is provided with significant amounts of training data, including text, audio, images, and video, from various sources. The model is capable of recognising patterns in the data during pre-training. This task requires a considerable amount of computational resources and is performed on specialised hardware.

23. *Reinforcement Learning from Human Feedback (RLHF)*
In the case of RLHF, human feedback is incorporated into the training process of AI algorithms to guide or improve the AI algorithm's learning. It is suggested that this feedback could potentially aid the algorithm in learning at a faster and more effective pace. The objective is frequently to utilise human expertise to guide AI algorithms towards a specific desired outcome.

24. *System*
An AI system comprises not only the model but also the entire infrastructure surrounding it. This includes the hardware, software, data processing, input and output interfaces, and all the components needed to run the model effectively. ChatGPT is an example of an AI system.

25. *Task-specific AI ('narrow AI')*
AI that is programmed for a specific task is known as narrow or task-specific AI, as opposed to generative AI, which can be used for a wide range of tasks.

26. *Training*
The process of learning an algorithm to recognise patterns in data.

27. *Transparency*
A model is considered transparent when it is clear which formulas, operations, and values are used to generate its output. A transparent algorithm is the opposite of a black box algorithm.

28. *Explainability*
A model is considered explainable when it is possible to understand and explain why the model produces specific outputs. Explainability ensures that a human can understand why a model behaves in a certain way without requiring knowledge of the model's formulas, actions, and values. Hence, a model that is explainable does not necessarily mean that it is transparent, and a transparent model does not necessarily mean that it is explainable.

29. *Guard rails*
Restrictions, guidelines or safeguards put in place to ensure that the use of LLMs remains within ethical and responsible limits.

30. *Webscraping*
The extraction of information from web pages for subsequent analysis using software.

Colophon