Ministry of the Interior and Kingdom Relations

Responsible Use of Generative AI

Government-Wide Guide

Working on the foundations of the Netherlands and of our Kingdom.

Table of contents

Introduction

Generative AI can be considered a powerful extension of human analytical and creative abilities. It has extensive capabilities with which societal issues can be addressed and can increase labour productivity. At the same time, certain rights may be under pressure and there are significant risks involved. For example, generative AI can reinforce discriminatory dynamics and socioeconomic inequalities.[1]

The Dutch government has a leading role when it comes to using emerging technologies such as generative AI responsibly and safely. As such, government organisations must have the right toolbox to use this technology responsibly and lawfully in all aspects of their work.

One of the criteria here is compliance with the requirements of the European AI Act (hereinafter AI Act) and other relevant regulations. To ensure compliance with laws and regulations, legal issues must be carefully considered when using generative AI. Once it is clear that generative AI can contribute positively to achieving organisational goals, it must be determined whether the intended use is possible, legal, and responsible. Digital ethics is fundamental to this process. In addition to legal and ethical considerations, consideration must be given to security, with specific attention to privacy and cybersecurity. Finally, responsible use consists largely of assigning appropriate roles and responsibilities.

Reading Guide

Chapter 1 explains the delineation of this guide through the positioning and scope of the document. Chapter 2 explains what generative AI is and how the technology is being used in government. Chapter 3 discusses the opportunities presented by generative AI, as well as low-level application areas in government organisations. Chapter 4 presents a roadmap with recommendations for organisations that wish to start using generative AI. It contains references to chapters 5, 6, 7 and 8, which provide information on, respectively, the legal, ethical, security, and governance aspects of generative AI. Chapter 9 contains a list of concrete recommendations that relevant government employees can communicate to end users. The guide concludes with a glossary and an appendix that succinctly explains how generative AI systems are created.

1 Government-Wide Vision on Generative AI

1. Positioning and scope

1.1 Scope

The guide delves deeper into the technological, organisational, ethical and legal conditions that help a government organisation use generative AI responsibly. This guide covers the entire palette of working with generative AI: from experimentation to implementation and monitoring.

This document is only intended to provide guidance and is not binding. The following falls outside the scope of this document:

- The establishment or provision of new requirements or conditions relating to the use of generative AI;
- The manner in which the regulatory bodies fulfil their role of overseeing the use of generative AI applications;
- The provision of guidance on the procurement, development and use of AI in a broader sense. This paper only addresses generative AI;
- The interpretation of standards or integration of existing frameworks.

1.2 Related documents

- Government-wide Vision on Generative AI of the Netherlands: This government-wide vision emphasises the importance of generative AI that serves to enhance human well-being, autonomy, sustainability, prosperity, justice and security.
- The Government-Wide position on the Use of Generative AI (published March 2025).
- The Algorithm Framework: provides an overview of existing laws and regulations that may apply to algorithms.
- The Algorithm Register: a publicly available national publication of algorithms used by the government with accompanying explanations.

This guide is consistent with the Government-Wide Vision on Generative AI and the revised position for government organisations on the responsible use of generative AI. When applicable, this guide may also be used alongside existing laws and regulations that may play a role in the use of generative AI and, particularly, the European AI Act. The reason for this is twofold. Firstly, generative AI is currently already widely used within government organisations[2], while only a limited number of European and national standards have been established and various provisions of the AI regulation are not yet fully in force. This document serves as an immediately deployable guide to provide government organisations with the right tools to use generative AI responsibly. Secondly, this guide offers concrete recommendations for government organisations and end users. In this context, end users refers to employees of government organisations.

1.3 Organisation-specific version of the guide

Government organisations are encouraged to publish their own version of the guide:

- The recommendations for government employees which they can translate into their own organization-specific recommendations can be found in Chapter 9, Recommendations.
- Government organisations may adopt additional, stricter rules if, for example, they process a lot of confidential or sensitive information.
- If an organisation creates its own version, the title must include the name of the organisation.

1.4 Status and management of the document

This document does not require any central monitoring or control mechanisms. However, it is recommended that this document is translated in a decentralised manner with

2 See reports: Focus on AI in Central Government | Report | Netherlands Court of Audit and 3 Rapid Assessment of the Impact of Generative AI on Government Personnel. | Report | Rijksoverheid.nl

appropriate control mechanisms. Organisations can contact their own or local CIO and CDO offices, for example, for help refining the guide.

As developments in generative AI are continuing at a rapid pace this guide is considered a living document. As such, it is periodically supplemented and updated with newly acquired knowledge from pilots and living labs and requirements under new laws and regulations.

2. What is **generative AI** and how is it used in government?

2.1 What is generative AI?

Generative AI is a form of artificial intelligence that uses algorithms to generate content at a user's request. With a simple natural language command—a *prompt*—users can have AI generate content such as text, image, sound, video or computer code at the press of a button. Generative AI is also an example 'general purpose AI' (GPAI). According to the AI Act, this is powerful AI that can be adapted and used for multiple applications and domains.

Generative AI technology is trained on large amounts of data from which the system can learn to recognise patterns and structures. Based on such patterns, the trained model can predict the most logical next element in, say, a sentence, picture or piece of music. A more detailed explanation of how a generative AI system is created is included in the appendix of this guide. One of the most recognisable applications of generative AI for the general public is the AI chatbot. Well-known examples include ***ChatGPT***, ***CoPilot*** and ***Gemini***. These are chatbots based on ***large language models*** (LLM), which specialise in natural language processing with a focus on text processing and prediction.

Generative AI models can be offered in several ways. Providers can choose to make the source code and any other components—such as training data, performance *metrics* and controls—accessible (*open source*) so that the model can be reused by third parties. The Allen Institute for AI's *Open Language Models* (OLM) are one example of this. When providers keep the source code and training data protected, it is considered a *closed source* model (*proprietary model*). *GPT-4*[3] is an example of a *closed source* generative AI model.

Generative AI can also be used in 'agentic AI'. Agentic AI refers to autonomous AI systems that can independently pursue goals and take actions without continuous human intervention. In addition to using generative AI to make decisions, such systems can create new content, such as text or images, to achieve their goals. Without adequate monitoring, AI systems risk making inappropriate decisions that are not in line with government policy or ethics. This problem will only increase as the number of agents increases, with the additional risk of conflicts between agents.

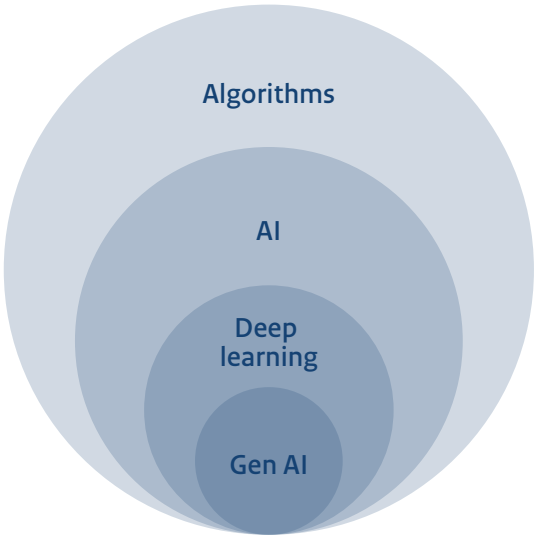

Figure 1: relationship between generative AI and broader playing field

3 For more examples of open and closed source models, see this table from Radboud University: https://opening-up-chatgpt.github.io

2.2 Generative AI in government

Generative AI offers the government opportunities to improve processes, optimise services to citizens, and increase government performance in general. Various government organisations are using generative AI or conducting pilots. To gain a better understanding of certain generative AI *use cases* in government, and to get a better idea of the possibilities for the responsible use of generative AI, an AI community has been established in which government employees can share insights and experiences. This community can be found at: https://aienalgoritmes.pleio.nl/.

3. Opportunities presented by generative AI

This chapter describes ways in which the Dutch government can capitalise on the opportunities presented by generative AI.

3.1 Renewed objectives

Generative AI can contribute to achieving government organisations' goals, as also described in the Government-Wide Vision on Generative AI of the Netherlands[4] and the Rapid Assessment of the Impact of Generative AI on Government Personnel[5] report. Generative AI enables the government to innovate in two ways:

- Firstly, the use of generative AI may enable the government to achieve its current goals more efficiently and effectively, e.g., by making certain processes faster and easier.
- Secondly, using generative AI, the government even may be able to adjust, renew or elevate its goals as the technology provides new insights.

3.2 Areas of application

To capitalise on the opportunities presented by generative AI, it is prudent to classify these opportunities in terms of added value, risk and security. This will create a selection of applications that can benefit many government employees and pose little risk. These are applications for 'everyday use'. They have a relatively low impact, which makes them more accessible to use after analysis and approval by the responsible party (director) within the organisation. These may include the following categories:

- Generating new ideas. Generative AI can help with brainstorming by suggesting new viewpoints to solve problems.
- Automation of repetitive tasks. Generative AI can help with automatic form filling or—like traditional AI—with analysing bulk data.
- Finding information. Generative AI can help find and locate public or internal information faster. An example of this is sAmmie, the virtual colleague used by the province of Noord-Brabant. A tool can also provide suggestions in response to a question, such as requests under the Open Government Act.
- Supporting (policy)officers . Generative AI provides a user-friendly interface to easily extract relevant and prioritised answers from complex information concerning policy and other topics. For example, it can suggest answers to questions such as, "What exact regulations should my new AI system take into account?" Or, "What rules in hospitality legislation are conflicting, and in what areas?"
- Greater involvement in society. Generative AI can scan and analyse public sources to support the identification of emerging sentiments, trends and themes.
- Software development. Generative AI can help developers write and check code, e.g., by making suggestions or detecting errors. Generative AI can also generate automated test cases.
- Text editing. Generative AI can help translate, simplify or improve certain texts. It can also help rewrite text for different audiences and create corresponding FAQs.

4 Government-Wide Vision on Generative AI | Report | Rijksoverheid.nl
5 Rapid Assessment of the Impact of Generative AI on Government Personnel | Report | Rijksoverheid.nl

- Image editing. Generative AI can help generate and edit images, charts or infographics for internal communications.[6]
- Scenario studies and forecasts. Generative AI can create realistic scenarios and offer employees advice on how to prepare for them. These scenarios may include workplace simulations or trends in society or industry. This application can be useful in areas such as employee training or education.
- Publishing information. The use of Generative AI can aid in compliance with legislation or increasing citizen confidence by more easily achieving active disclosure of documents and responding to requests under the Open Government Act within days.
- Aiding in scientific research. Scientists can use generative AI to make research more efficient and creative and to process growing amounts of data, but they also have the opportunity to study or develop generative AI themselves. Generic frameworks are inadequate because scientific research has specific goals and takes place in a specific context. The responsible development and application of generative AI must be done in line with the principles of research integrity and government-wide policy. To address this more specifically, national knowledge institutes and planning agencies (united in the RKI network) are jointly developing a framework for the use and continued development of generative AI in the contexts of scientific research.

Access to complete, recent and high-quality data is crucial for the generative AI tool to work properly. Even then, it is worth remembering that the current state of the technology sometimes leaves much to be desired. In addition, certain applications may greatly interfere with the primary process or involve the use of sensitive data. Risk assessments are essential for such applications[7]. They can also be integrated in contracted systems to prevent the potential dissemination of confidential data. These include applications relating to interaction with citizens, selection techniques and support with decision making and implementation. Finally, it is important to always include some form of human intervention when using generative AI. This keeps the official in control when using generative AI.

6 See also: The national government's visual identity and imagery - Beeldkompas.
7 For more information see the Government-Wide Position on the Use of Generative AI.

4. **Getting started** with generative AI

This chapter provides a general, practical roadmap to help use generative AI sensibly and responsibly.

4.1 Define the goal and application areas

Generative AI is a means to an end, just like any other technology. Putting the goals first helps prevent getting hung up on technological innovation without context. The starting point is always the challenge or problem, not the solution. Generative AI is a powerful addition to the repertoire of potential solutions to problems or further improvements to already well-running processes.

Challenges include demonstrable improvement to public services, increased productivity, reduced workload or reduced project costs. In addition, certain *use cases* for generative AI must be avoided at all times.

4.2 Involve the right people with the right skills

To use AI technology responsibly, it is prudent to involve a diverse group of employees. In any case, the following areas of expertise are needed when building or acquiring generative AI:

- managers who understand the context and impact;
- lawyers, ethicists, and privacy and cybersecurity officers who can help develop and use the solution responsibly and securely;
- data specialists with insight into the quality and availability of needed data;
- software developers, UI/UX designers and architects who know how to build or integrate solutions;
- purchasers with experience of everything involved when purchasing software with AI.

Based on Article 4 of the AI Act, organisations developing or using AI systems must ensure that their staff and other stakeholders have sufficient knowledge of AI ('AI literacy'). They should consider the technical knowledge, experience and training of users, the context in which the AI systems are used, and the individuals or groups for whom the AI systems are intended.

Consider establishing an ethics committee within your organisation. This should be a multidisciplinary, diverse group in which there is sufficient knowledge about the various forms of AI.[8]

Working with generative AI requires varying degrees of knowledge and skills depending on the role involved. These stakeholders can be categorised as follows (this list is not exhaustive):

1. General: any official who is not familiar with generative AI but will have access to it is recommended to undergo basic training to better understand the technology and its risks. This could include training such as the offerings by RADIO.
2. Policy and operations: officials using generative AI for information retrieval and document generation. It is recommended that this group can always fall back on internal or external training and handouts to use the tools responsibly.
3. Developers: officials with advanced digital skills working on the development or implementation of generative AI solutions. Learning opportunities can focus on the technical and ethical aspects and implementation challenges associated with generative AI.
4. Data analysts: officials who collect, organise, analyse and visualise data. Training may cover the use of generative AI to facilitate automated data analysis, synthesis of complex information and generation of predictive models.
5. Decision makers: administrators responsible for creating a government culture ready for the use of generative AI.

8 In that context, see for example: ipo-whitepaper-verkenning-chatgpt.pdf.

It is recommended that they have access to information that helps them understand sound and responsible use of generative AI and its potential impact on organisational culture, governance, ethics and strategy.

6. Compliance officers: staff focused on compliance with legal frameworks related to the use of generative AI. They are expected to have sufficient knowledge of the legal, ethical and organisational measures necessary to use generative AI responsibly.

4.3 Create an AI governance structure

Effective AI governance within an organisation helps ensure the proper implementation of AI models and helps identify, evaluate and manage risks. The governance structure for generative AI may fit well with existing AI or data governance. A good AI governance structure may consist of:

- An AI strategy or policy: adhere to the government-wide position and translate this guidance into an internal version. Submit this policy for review by information officers within your organisation, such as the CPO, CDO, CISO,CIO, FG and/or AI Officer.
  - Establish a policy or strategy that provides information on the ethical, legal, and security aspects of generative AI.
  - The policy or guidelines must include recommendations for end users to ensure that they use the generative AI application as safely and effectively as possible. Prevent employees from using the generative AI application for other purposes, and agree on whether or not to log prompt history for transparency and traceability. Ensure that all end users within the organisation know what the system may and may not be used for.
  - Establish a 'fall-back' scenario in your strategy for when generative AI may or can no longer be used.
  - When generative AI is used to interact with citizens, the citizen should be explicitly notified that they are interacting with generative AI. Citizens must always have the option to speak with a real person if questions or ambiguities arise.
- An AI steering committee and accountability structure:
  - Assign responsibilities within the organisation. For example, appoint an officer responsible for coordinating the intersection between cybersecurity and the use of generative AI. This person maintains close contact with parties such as the CISO (if present). Also appoint an officer responsible for privacy and generative AI, who works with the chief privacy officer (CPO) and/or data protection officer (DPO). Appoint a support role that employees can contact with questions about generative AI.
  - Ensure that you are not discriminating by developing or using the generative AI. Establish a protocol if discrimination occurs. There is an existing discrimination protocol that can help with this.
  - Establish whether you can have a sustainability impact analysis conducted and minimise your emissions. Options include fine-tuning pre-trained systems and favouring energy-efficient data servers.
  - Systematically identify the security risks of generative AI applications. The OWASP Risk Rating Methodology can help with this. If necessary, seek advice from an in-house or external cybersecurity specialist.
  - Create a structure that enables accountability for errors of the generative AI application. Ensure that errors or discrimination can be reported.

- A documentation strategy: for overview and transparency outside the organisation.
  - Be transparent about the use of generative AI applications and show why the application is used and by whom. This must be documented, preferably through the Algorithm Register. For internal documentation, summarise the added value of the generative AI application, e.g., by what percent an outcome improves, by how much time is saved or better utilised, or by evaluating performance regularly with varying prompts or versions of models.
  - Maintain an up-to-date record of all AI applications used within your organisation (e.g., through publication in the Algorithm Register). Include a detailed description of the system; the Publication Standard can be used for this. At a minimum, describe the public purpose and target audience. It should be clear whether the interaction is aimed at internal or external users.
  - Document considerations about data access and use. For example, describe whether the generative AI model will be trained further at any point, and with what data. Describe whether this data is stored, who has access, and how its integrity is ensured.
  - Collect conducted risk assessments such as DPIAs and algorithm impact assessments in a central location. Make sure these can be easily retrieved and serve as references for others to avoid duplication.

4.4 Conducting risk analyses

Once the goal has been considered, the right people are involved, and an outline of a deployed generative AI governance structure has been created, the next step is to conduct a risk analysis (see the position on generative AI). This should be done before purchasing a generative AI system or during in-house development.

- When developing generative AI applications in-house, ensure that privacy, security and human rights risks are considered in the design. Consult resources such as the Algorithm Framework for guidance on this topic.
- Conduct the necessary risk analyses. An overview of various tools is provided in the Algorithm Framework. Such tools include a Fundamental Rights and Algorithms Impact Assessment (IAMA) or AI Impact Assessment (AIIA), but may also be a proprietary tool. The purpose of such an analysis is to spark interdisciplinary discussion among various roles in an organisation about the use of and risks associated with an algorithm. A Data Protection Impact Assessment (DPIA) must be performed if required under the GDPR. A pre-DPIA scan is always required if personal data is processed[9]. This applies to all generative AI applications, both those used on a project basis and generative AI applications used systematically by employees in their daily work. The outcomes of these assessments should guide policy, as described in the Government-Wide Position on the Use of Generative AI. These assessments are especially important when using generative AI in high-risk applications. The AI Act determines the risk classification of an AI system according to its use and deployment. If the system is used in a high-risk application, see Chapter 5.1 for more information on the AI Act.

The use of AI systems under consumer conditions[10] currently poses unacceptable risks. This form of licensing may only be used for private use or for exploration or experimentation with new AI systems in a controlled, non-production environment, provided no sensitive or confidential data is used. This restriction may be lifted once providers meet the legal and regulatory requirements. Discussions are underway to resolve this government-wide.

4.5 Generative AI procurement or development

Following completion of the risk analyses, the application can be procured or commissioned. Generative AI can be used and integrated in a variety of ways. Each method has unique advantages and security risks that affect how the technology should be handled. The most common applications of generative AI are:

1. Public generative AI applications and web services. These are off-the-shelf AI tools accessible to the general public via the internet, such as ChatGPT.
2. Embedded generative AI applications. This involves integrating AI functionality into existing software or devices, such as AI-enabled word processors. One example is Github Copilot, an application that assists in code development by providing suggestions.
3. Public generative AI APIs. These interfaces enable developers to integrate AI functionality into their own applications without having to build the underlying technology.
4. Local development. This involves creating and training AI models on local systems. This offers more control but also requires more resources.
5. Cloud solutions. This approach uses cloud infrastructure to host AI models, providing scalability and flexibility.

Observe the following steps when procuring or building generative AI applications:

- Learn from other government organisations that have implemented generative AI solutions.
- Have a good plan to increase employee AI literacy in the application. Consult employees who will be working with the application early on in the process.
- Consult commercial peers from the start when procuring generative AI applications. Pay attention to supplier management to ensure compliance with legal frameworks. The Algorithm Framework includes information on procurement conditions.
- Examine all options to avoid v*endor lock-in.* Explore the offerings of local or European generative AI service providers. These are often more likely to meet national and European security standards than non-European service providers. This also enhances Dutch and European digital sovereignty.
- Various models differ in their abilities to deal with the Dutch language. Use Benchmarks such as OpenGPT-X to

9 During the DPIA process, the Chief Privacy Officer (CPO) must be consulted before the results of the assessment are presented to directors. The Data Protection Officer (DPO) must also be consulted.

10 These are conditions set by the software vendor for using the software for personal use. This is in contrast to commercial or organisations' terms and conditions that include specific terms to protect the organisation.

help evaluate performance. European initiatives, such as the EuroLLM and OpenGPT-X Teuken models, pay extra attention to European languages.

- Establish clear r*equirements*. Include a problem statement, highlight your data strategy and make arrangements for data access, use, storage and logging. Enforce clarity about how the service is monitored by the provider and ensure control over your own data.
- An open source application or model should be used, if possible. Open source models offer more insight in aspects such as transparency. However, transparency must never impact the security of generative AI models. Explore the options of of*fline a*nd on-*premise g*enerative AI models. Of*fline and/or on-premise* AI models are generally not connected to the Internet. This keeps the generative AI deployment within the organisation and processes data using a local server. The quality of models that can be used offline (often op*en source)* is continuously improving and already frequently approaches the performance of well-known, large online models. However, achieving this performance requires significant investment, partly due to the hardware required.
- Check the data settings of the generative AI system and disable data sharing for external optimisation of the model. This prevents the data entered by end users from being used to train the model, which, in turn, prevents the disclosure of personal data which may violate legislation such as the GDPR.
- When selecting LLMs, consider the capabilities of various F*oundation Models*. The Holistic Evaluation of Langue Models by the Stanford Center for Research on Foundation Models (CRFM) provides a good comparison of models using several criteria.

5. **Legal** aspects

This chapter discusses the key regulatory and legal frameworks relevant to the use of generative AI.

5.1 AI Act

The AI Act provides an important legislative framework for the development and use of AI in the European Union (EU). The AI Act divides AI systems into several risk categories. The rules may be more or less onerous, depending on the category into which an AI system falls. In particular, high-risk applications require additional considerations when government organisations want to use them. The regulation, which is being implemented in stages, also sets requirements for generative AI (in the AI Act, this falls under 'general purpose AI, GPAI').

In practice, government entities will mostly use GPAI models and systems. These can be purchased from a vendor and used for domain-specific applications. It is always important to consider whether you, as a government entity, are a provider and/or deployer under the AI regulation, as this determines what requirements you must meet.

5.2 Privacy legislation

In the development stage, generative AI models are often trained by '*scraping*' large amounts of data from the internet. This is likely to involve processing various forms of personal data without consent. Because many generative AI models currently cannot guarantee that no personal data has been processed, they are unlikely to comply with privacy legislation[11] such as the General Data Protection Regulation (GDPR) and the GDPR Implementation Act. No court ruling has been made on this matter in an EU context at the time of writing. In addition, generative AI models used by the police and judiciary must comply with the Police Data Act and the Judicial Data and Criminal Records Act. This list is not exhaustive; the applicable laws may differ for each context.

It is also important for the user to be wary of data input when using generative AI. If personal data is entered into the AI system (with a legal basis), this data may be stored by the developers of the model to be used for training purposes. This is standard for non-contracted services such as *ChatGPT* and can cause data breaches. Using in-house *open source* applications or self-developed generative AI models prevents this, but poses technical barriers. When purchasing generative AI applications or integrating generative AI models into existing software (e.g., MS *Copilot*), it is important to have an understanding of how user data is used, stored and shared, and to make agreements with vendors about this.

5.3 Copyright

As it is not always possible to trace the origin of the data used to train generative AI models, it is highly likely that the training data contains copyrighted material such as works of art, books, photographs and publications. Employees must never assume that permission has been obtained for all training data used for a generative AI model. Many providers are not currently transparent about the use of copyrighted works in the training process. The AI Act changes this by imposing transparency requirements on providers of large GPAI models (Article 53 of the AI Act).

This requires providers to establish an internal policy for copyright compliance and disclose a detailed summary of the training content used. This helps rights holders better monitor compliance with text and data mining requirements.

As such, training generative AI with works of literature, science or art may sometimes be allowed under the copyright exception for text and data mining. Even then, to avoid infringing on the creators' copyright, the output of generative AI must not be too similar to the work used to train the AI. Rights holders can take legal action if necessary.

5.4 Procurement

Government organisations use public procurement to purchase software with generative AI components. Public procurement is subject to both general and specific conditions. These may be off-the-shelf solutions, an existing

11 DPA: scraping is almost always illegal | Data Protection Authority

solution with modifications, or a new solution developed specifically for the government. When purchasing high-risk AI applications, the principal is likely to take a more active role in working with the vendor. For example, the principal must specify the legal and ethical limits of the ultimate operation of the AI system. When purchasing a turnkey solution, the vendor must be able to demonstrate that the AI application meets all requirements. In addition to legal requirements, aspects such as ethics, sustainability, and sovereignty should be considered. For more information on procurement, see the Algorithm Framework and the General Government Terms and Conditions for IT Contracts (ARBIT). As part of the Dutch Digitalisation Strategy, we will deliberate on a collaborative approach to the procurement of digitalisation solutions, including generative AI, to strengthen the Dutch government's market position and benefit from increased efficiency.

5.5 Other laws and regulations

Other laws that an organisation must adhere to when using generative AI include the General Administrative Law Act, the Open Government Act, the Public Records Act and the Digital Government Act. These laws apply to IT systems and, by extension, to AI applications. The requirements from these laws are compiled in the Algorithm Framework.

6. **Ethical** aspects

This chapter outlines the key ethical values and themes (critical reflection on what is desirable) that come into play when using generative AI. In particular, this chapter focuses on the ethical issues directly involved in the use of generative AI, rather than broader societal issues such as the long-term impact of generative AI on the labour market. For information on the broader considerations, please refer to the Government-Wide Vision on Generative AI.

6.1 Bias and discrimination

Generative AI models—or the data used to train them—may contain biases or prejudices. This is called *bias* . This may stem from the data collection process, such as *selection bias*. This means that limited training data creates an inaccurate reflection of reality, resulting in blind spots. However, the primary risk of generative AI models involves the adoption discriminatory or undesirable patterns from the training data. To counter this, when developing generative AI models, training data should be critically examined and mitigating measures should be taken where possible (also called 'guard rails'). For models without guard rails, clear agreements should be made with end users regarding the use of outcomes. There is also a risk of *groupthink*: a generative AI model may arrive at consensus without considering or showing alternative viewpoints. It is important that users consider output critically and remain open to non-conventional thought.

Under Dutch and European law, *discrimination* on the basis of gender, race, religion or any other ground is forbidden. The AI Act sets requirements for providers of AI systems to prevent discrimination[12]. These requirements also apply to developers of generative AI models. However, some degree of bias is often a given in AI models based on real data.

We urge government organisations to carefully consider which anti-discrimination requirements they impose. To that end, the Non-Discrimination by Design guideline can serve as a reference. Outputs from generative AI systems whose use may affect human rights must be tested for unwanted bias promptly and regularly—not only during the development of the model but also after it has been put into use. It is also prudent for government organisations to examine the risks if it is determined that a generative AI model cannot meet these anti-discrimination requirements.

6.2 Transparency

It is important that the government is transparent about the development and use of a generative AI model to allow the verification of ethical training and use of the model. For example, when training a model, consideration should be given to human rights and the risk of underpaying workers. Transparency can be promoted in the model itself by being open about the datasets used to train the model and providing clear documentation of the system, including information about the techniques used, source code and parameters. Transparency rules from the GDPR also apply when personal data is processed. For purchased models, this should be included in the terms of purchase. When fine-tuning or developing generative AI models in-house, this should be logged in technical documentation.

In addition, it is important that the government is transparent about the use of generative AI when it affects decision-making. Individuals whose data is processed or who are in direct contact with the AI system through interaction with an *AI chatbot*, for example, must be notified, as also mandated under the AI Act. This can be done through the inclusion of a privacy statement about the collection and use of data[13], for example.

When using a generative AI system to create a government product, we recommend reporting the system as an information source and have the prompts and outputs from the

12 These requirements and associated measures will also be included in the Algorithm Framework: Algorithm Framework - Algorithm Framework [draft] (minbzk.github.io)

13 The requirements of a privacy statement have been established by the Data Protection Authority: Right to information | Data Protection Authority

system available to share with stakeholders and for evaluation, monitoring and inspection purposes.

6.3 Trustworthiness

Government organisations have a responsibility to provide citizens with reliable information. The use of generative AI in this process could have negative consequences for the government's credibility and reputation. A generative AI system does not weigh the reliability of information when answering a prompt, which has the potential to damage the trust of citizens. This risk still exists if there is transparent communication about the use of generative AI applications, as people may view the information more critically. The use of generative AI in communication with citizens requires thorough consideration of the risks. Some risks can be overcome by a disclaimer indicating that the output of the generative AI system may contain errors, or by limiting responses to references to official publications. Having a human element providing input as part of the generative AI model ('human-in-the-loop') also contributes to the reliability of the output.

The use of AI for content creation—such as the generation of photos, videos and *voice-overs*—may conflict with existing policy frameworks, such as that visual communication must always be a realistic representation of Dutch society and that image manipulation must be used with restraint.[14] As such, we strongly advise against the use of AI-generated imagery in both internal and external communications for the time being.

6.4 Explainability

Explainability refers to being able to explain how an algorithm works so that people understand its operation. Generative AI models are '*blackbox*' models. This means that the model's outputs cannot be properly reasoned, where a simpler algorithm such as a simple decision tree can be fully explained. The billions of computations required for each output do not allow human understanding of the generative AI model at the lowest level. This is one of the reasons why generative AI applications cannot be used for direct decision-making, only in support thereof. One way to partially explain how a generative AI model works is to provide explanations of the AI model's capabilities, limitations, and training data so that users know what the model can and cannot do well. Transparency is of even greater importance when using generative AI applications.

We recommend that government organisations carefully consider whether using a *blackbox* model is the best choice, or whether another model or algorithm can provide the same solution. It is also prudent to examine the risks if the outcomes of a generative AI system cannot be explained.

6.5 Quality

Although AI-generated content often looks credible at first glance, generative AI models do not always provide factually correct answers. This is because generative AI is optimised for plausibility, not accuracy. Generating false outputs is called 'confabulating', more commonly known as 'hallucinating'. This involves a generative AI model presenting a fabrication as fact. Both the quality of the training data of a generative AI model and the quality of the prompt greatly influence the quality of the output. The outputs of generative AI models may provide an incomplete picture or be outdated due to incomplete or outdated training data, for example.

6.6 Sustainability

Training and using generative AI models requires a lot of computing power, hardware and data, depending on the type of model. The required computing power and data servers have an impact on the climate. In particular, the energy consumption, CO2 emissions and water consumption of AI models can be high; for example, generating one kWh of electricity requires an average of 0.93 litres of water (primarily by cooling systems). Producing the hardware also requires scarce resources. There are currently no standardised methods to measure the environmental impact of generative AI, and estimates vary widely. As such, many AI service providers do not have clear figures on the emissions of their models.

New techniques such as *quantum computing*, more efficient architectures or the optimisation of existing architectures could make models more economical. At the same time, we cannot assume that technological greening will be enough. Organisations will have to make choices to make the digital economy sustainable. As such, we recommend that organisations consider the reason for selecting a generative AI model, and whether a more sustainable alternative can provide the same solution.

14 See, for example: https://www.beeldkompas.nl/kennisbank/rijkshuisstijl-beeld

7. **Security** aspects

This chapter provides insight into the potential security risks that may arise from the use of generative AI.

7.1 Protection of and from AI: Cyber attacks

While generative AI technology offers numerous cybersecurity opportunities on the one hand, the technology also poses security risks. Generative AI models can be attacked or used to carry out an attack. A project is underway within the central government to develop actions and measures to manage cyber risks when using AI.

Generative AI systems that lack robust security mechanisms can be abused by malicious actors. One example is a *prompt injection attack*, which involves manipulating a generative AI system through a particular prompt or instruction. The attackers' aim is to coerce the model into producing malicious output, such as releasing confidential information, generating manipulative content or spreading disinformation. To prevent such attacks, generative AI systems must contain sufficient security mechanisms and must be regularly tested for potential vulnerabilities.[15] [16]

Generative AI models can also be used to create *deepfakes* and personalised *phishing* emails or *voice* messages more efficiently and on a larger scale.

7.2 Privacy and data breaches

As described in chapter 3, the training of generative AI models can lead to violation of privacy rights. Generative AI models are trained on vast amounts of data, which may contain personal data, or personal information may be shared as output.[17] Entering confidential information into a publicly available generative AI system can also lead to data breaches, which poses security risks. For more information see the EDPS publication for advice on information security and privacy when using generative AI. Furthermore, even non-sensitive information can become sensitive when entered into an AI system, as the government only uses a limited number of IP addresses, and public AI systems can combine input data to compile a profile.

7.3 Information security

One basic framework of standards applies to all levels of government: the Government Information Security Baseline. This framework sets information security standards and requires organisations to conduct a GISB *quick scan* and risk analysis. The results must be discussed with the party responsible for the identified risks. The Chief Information Security Officer (CISO) can also be consulted for advice.

15 Develop AI systems securely | Press release | General Intelligence and Security Service

16 To inform developers and organisations of the potential security risks involved in the use of generative AI technology, the OWASP has compiled a list of the ten most critical vulnerabilities in generative AI systems: OWASP Top 10 for Large Language Model Applications | OWASP Foundation

17 Source: Rathenau Techscan

8. Governance

As described in the previous chapters, the use of generative AI technology can pose various regulatory, ethical and safety risks. Effective AI governance within an organisation helps identify, evaluate and manage these risks. This chapter does not provide a blueprint for the structure of AI governance within organisations. However, it does briefly outline the issues a government organisation can consider in the governance of generative AI.

8.1 Accountability

It is important to consider accountability when using generative AI. Accountability involves reporting to the appropriate bodies about how generative AI was used, what risks were taken, what decisions were made, what outcomes were achieved and what costs were incurred. Control mechanisms are implemented to facilitate proper and appropriate accountability. In addition to reports, internal and external audits may be conducted.

8.2 RASCI

A RASCI model can help to clarify roles and responsibilities when using generative AI. This model helps establish *responsibility* for the use of generative AI, *main accountability*, *support*, which role is *consulted*, and which role is *informed*.

8.3 Whistleblowing, objections and appeals

Given the sensitivity and potential major impact that generative AI applications can have on citizens, businesses and employees, it is important to establish and link the appropriate processes[18] for whistleblowers and for objections and appeals; see also Directive 2019/1937. There must be clear procedures in place for employees to anonymously file an internal or external report about negative effects of the use of generative AI. It must be clear to citizens, businesses and employees how they can object to and appeal decisions directly or indirectly resulting from the use of generative AI applications.

18 Processes can be linked to a data breach procedure or incident process within the organisation, for example.

9. Recommendations for end users

This chapter describes some practical recommendations for government organisations to communicate to their employees to promote the responsible use of generative AI. The recommendations are aimed at CIO offices to enable them to create guidelines for end users. CIO offices can translate these recommendations to organisation-specific contexts or focus on a particular generative AI application. The recommendations are categorised according to the topics from the previous chapters.

9.1 Legal aspects

- Note that the outputs of generative AI may be based on copyrighted information. Do not accept the outputs of the generative AI application blindly but use them for support or as a starting point. Do not use the results of a generative AI model if you suspect copyright infringement. When in doubt, contact an internal or external copyright specialist.
- Processing personal data in a generative AI system without a legal basis is unlawful under privacy legislation such as the GDPR. Never enter sensitive information or personal data into an AI model without a well-founded legal basis. Input information may intentionally or unintentionally reappear in the outcomes of the model.

9.2 Ethical aspects

- Before choosing to use generative AI, always consider carefully what you want to achieve and whether this technology is the most appropriate tool to achieve that. Never use generative AI to outsource skills you do not possess or understand yourself.
- Only use generative AI as a tool, not as a replacement. Do not use generated content verbatim. Do not humanise this technology; formulate your own answer using the output.
- Evaluate generative AI outputs critically to identify possible biases.
- When using generated content in communications to citizens, make sure the text or image is captioned or watermarked to make it clear that it was created by an AI model.
- The use of generative AI may be unavoidable in some cases, such as when using search engines or automatically translated websites. Handle this with integrity. In other words, be aware that these results are based on generative AI and check whether the results can be used verbatim or whether additions or adjustments are necessary. With the increasing presence of AI in systems, it is important to be aware of whether the systems you use are using this technology and how.

9.3 Tips for writing effective and ethically safe prompts

- Experiment with different instructions and ask follow-up and clarification questions. This can help determine whether the model produces a substantively correct answer. Keep in mind that LLMs do not have a real memory within a session. They mimic memory by incorporating some or all of the content of the previous prompts and answers into the new prompt behind the scenes. To be sure they provide the clarification you seek, you should include the texts most important to you in your prompt as context.
- Avoid entering biased or suggestive questions or assignments. When you ask a generative AI model a suggestive question, there is a high chance you will receive an affirmative answer. Always remain critical in reviewing answers; if necessary, check them with a colleague or consult a reputable source.
- Be specific and precise through customary language and sufficient detail. Avoid the use of jargon if it is not necessary for answering, and divide a complex task into multiple, simpler sub-tasks. Validate the tasks and sub-tasks independently and in context with the results of other tasks.

- By default, generative AI models have no context. It is important to provide plenty of background information in your instructions, and use examples whenever possible. If you are writing an instruction with a specific purpose, state the purpose (and, potentially, the target audience).
- Include rules in your prompt. Clearly state what you expect from the generative AI model. For example, specify how long the answer should be and in what style and format the answer should be given. For text generation, for example, this can be done in a step-by-step explanation, table or list. For image generation, you can impart a particular style or colour effect, and for sound generation, a genre or tone-of-voice. It can sometimes help to assign the model a specific role in your instruction.
- To increase the explainability of output from generative AI systems, you can apply 'chain-of-thought prompting'. This is a method in which instructions are formulated incrementally to obtain better results from a language model. The user then asks the model to provide a rationale for the output.
- Get help from experienced colleagues or experts and delve further into the possibilities and limitations of this new technology. For example, take educational training at RADIO or a prompt engineering course.

10. Glossary

Algorithm Impact Assessment
An algorithm impact assessment is a tool for making informed decisions when using algorithms and artificial intelligence. Examples of algorithm impact assessments include the Fundamental Rights and Algorithms Impact Assessment (FRAIA) and the AI Impact Assessment (AIIA).

Artificial Intelligence (AI)
A technology which, for explicit or implicit purposes, determines how it generates outputs based on received inputs (such as data). Outputs include predictions, content, recommendations and decisions. The technology can learn, reason and perform tasks in ways that would normally require human intelligence.

AI Act
The European AI Act is one of the world's first comprehensive laws specifically for AI. The AI Act establishes frameworks and requirements for governments and companies regarding the development and use of AI systems.

Chief Information Officer (CIO)
The Chief Information Officer is responsible for developing and coordinating the organisation's information and digitalisation policy, as well as overseeing the development and management of its information systems in accordance with that policy.

Data Protection Impact Assessment (DPIA)
A DPIA is a tool to identify the privacy risks of a data processing operation in advance, enabling an organisation to take mitigating measures.

End users
Employees using generative AI in their work.

Data Protection Officer (DPO)
The Data Protection Officer oversees the implementation of and compliance with privacy laws within an organisation.

Generative AI
A type of AI that uses complex algorithms to generate content such as text, images, computer code or videos.

High-risk AI
The AI Act defines high-risk AI applications as those that pose a high risk to the health and safety or fundamental rights of individuals. To be allowed on the European market, such applications must meet several obligations.

Large Language model (LLM)
A specialised type of generative AI model trained on large amounts of text to understand existing content and generate textual content.

AI 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. An AI model is trained on a task, such as categorising documents or, in the case of an LLM, generating sentences similar to the training data. *GPT-4* is an example of an AI model.

AI 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. Examples of AI systems include *ChatGPT, Google Gemini and Copilot*.

Training
The process of teaching an AI model to recognise patterns in data.

11. **Appendix:** How is a generative AI system created?

To gain a thorough understanding of a generative AI system, it is important to understand the stages involved in its creation. The development of generative AI systems can be divided into several stages. The Rathenau Instituut distinguishes the following 5 stages (see Figure 2).[19]

1. Data collection: in the first stage, large amounts of data are collected to train the model, such as scientific articles, books, photos and videos. This data is often publicly available online.
2. Data preparation: in the second stage, the data is filtered and cleaned. This involves anonymising personal data and removing duplicates, for example.
3. Training: this stage is also called pre-training. During this stage, the model learns to recognise patterns in the data. This process requires considerable computing power and is performed on specialised *hardware.*
4. Optimisation: at this stage, the model is further refined and adjusted. The model can be fine-tuned with specialised (organisation-specific) knowledge and may be trained to produce socially acceptable answers. This involves special techniques such as *Reinforcement Learning from Human Feedbackedback*[20] (RLHF),
5. Deployment: in the final stage, the model is made available to users and put to use in practice. The model can be duplicated and then accessed through a consumer interface (e.g., as a chatbot or image generator) by tens of thousands of users simultaneously. The applications of generative AI are growing rapidly. Two noteworthy trends are: *Retrieval Augmented Generation* (RAG), where the AI system has access to a database of documents it can search and include in its output. Ag*entic AI also* gives the AI system the ability to perform independent actions, such as creating files. Note that new applications of generative AI involve their own (potentially still unknown) risks.

19 Generative AI | Rathenau Instituut

20 which involves incorporating human feedback into the training process of AI algorithms to guide or improve the AI algorithm's learning. This human feedback may enable the algorithm to learn at a faster and more effective pace. The aim is often to use human expertise to steer AI algorithms in a desired direction.

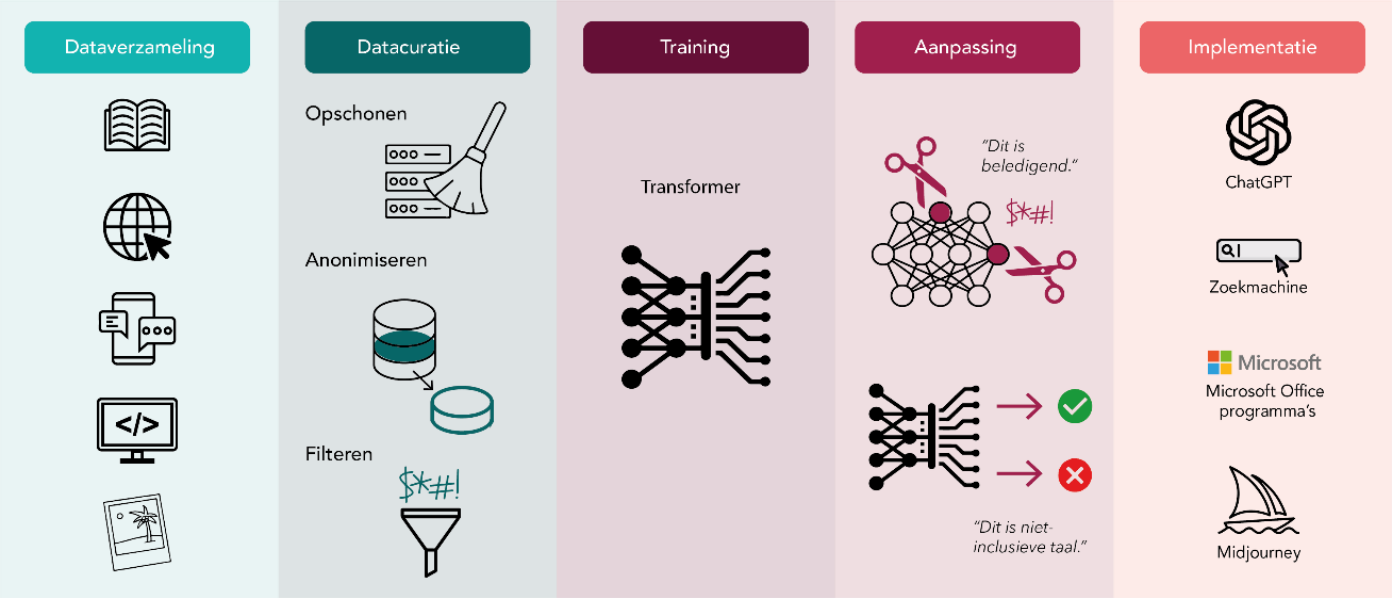

Figure 2: development stages of a generative AI system (source: Rathenau Instituut)

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
P.O. Box 20011
2500 EA The Hague

Date
January 2025