Moving forward responsibly

The AP's vision on generative AI

AP | bescherming in een digitale wereld

Consultation version May 2025

Table of contents

Consultation

The Autoriteit Persoonsgegevens (AP) (Dutch Data Protection Authority) invites you to respond to this document by email to **genai-loket@autoriteitpersoonsgegevens.nl**. The consultation is open until 27 June 2025. We will summarise the responses we receive in one document, without including any names of persons and organisations or contact details. We wil l publish the summary on the AP's website and use it to further improve this vision paper. The final version of this vision paper will follow later in 2025.

Management summary

1. **Generative AI is a high-impact technology that will touch upon many domains of society.** As a society, we are at the beginning of a social transformation. There are major opportunities to increase prosperity and well-being in domains such as healthcare, education, research and innovation in trade and industry. To seize these opportunities, it is crucial to deploy generative AI in such a way that it contributes to protecting and strengthening fundamental rights and public values. This includes a society-wide dialogue and consensus on how generative AI relates to our fundamental rights and public values.

2. **While generative AI has so far fallen short in lawfulness, the AP sees a possibility to work towards lawful development and deployment of generative AI.** A legal analysis by the AP has uncovered some irregularities and uncertainties with regard to the GDPR. Subsequently, the development and deployment of this technology can be organised in a way that prevents these irregularities and uncertainties. We discuss these preconditions and possibilities in the light of the GDPR in an article entitled 'GDPR Preconditions for generative AI'.

3. **At the application level – where generative AI is practically used by people and organisations – there are a number of preconditions for responsible deployment.** These applications must have a clear purpose and fulfil that purpose after a thorough risk assessment with appropriate (GDPR) safeguards. Other preconditions are, for example, sufficient AI literacy in the deployment and control of systems and data.

4. **On a societal level, the AP identifies a number of principles for embracing responsible generative AI.** These include European digital autonomy, societal resilience, democratic governance, a functioning market for responsible solutions and the ability to correct through the generative AI-chain. All of these principles can be promoted using societal tools such as investing in European AI model providers and driving transparency through regulation and public model cards. But also by working on auditing and risk management requirements for AI models and the AI systems based on them.

5. **In order to steer towards the responsible use of generative AI in society, a broad public and political debate is important.** Choices such as whether or not to prefer open-source models are now relevant. This requires the involvement of all layers of society: from education to developer, from citizen to director and from consumer to CEO.

6. **The desired future scenario for the AP entails one where effective regulation enables the development and use of responsible generative AI.** The AP will take concrete steps in the coming period to get there: we will set up a generative AI advice centre, organise periodic meetings and discuss the right forms of guidance on the subject. With this, we want to contribute to a shared guiding perspective for the role of generative AI in society. The AP contributes to further standardisation of the technology in an EU context.

1. Introduction

Generative AI has gone through a period of rapid development and societal adoption over the past two and a half years. This technology has become part of the digital fabric of our society. This brings with it a whole new set of opportunities and challenges: economic, technological, legal and social.

An impactful development such as the emergence of generative AI deserves the full attention of society. The AP aims for a future where digital technology is at the service of society and the social interest by, for example, contributing to a wide range of fundamental rights such as the right to education, the freedom to conduct a business, the right to social security and social assistance or the right to health protection. A carefully weighed and responsible embrace of generative AI will make it possible to take steps in all these areas.

At the same time, the technological functioning of generative AI makes fundamental rights worthy of additional safeguarding and protection. Think of the right to non-discrimination and the right to a fair trial, the right to property such as copyright or the right to the protection of personal data. In the area of data protection, the AP estimates that the vast majority of available foundation models fall short in terms of legality due to the 'scraping' of (special categories of) personal data. The European Data Protection Board (EDPB) found that the use of these foundation models by Dutch and European parties is not necessarily unlawful. The AP considers the lawful development and deployment of generative AI possible, but also sees challenges in areas other than data protection.

In this vision paper, the AP takes a forward-looking perspective on generative AI. What do we mean by it, what opportunities does the technology offer us, what risks does it entail and what needs to be done to embrace generative AI responsibly? We will take a high-level look at the protection of our fundamental rights, not only based on our role as supervisory authority for compliance with the GDPR and the Dutch Directive on data protection in the law enforcement sector (RGR), but also as the coordinating AI and algorithm supervisory authority and to contribute to the preparation of supervision of compliance with the EU AI Act.

This vision paper is primarily aimed at professionals who have to deal with generative AI in their work. By that we mean, in a broad sense, developers who make daily decisions about what generative AI can and cannot do, as well as administrators and managers who have to decide about the deployment and use of generative AI in their organisation.

With this vision paper, the AP aims to contribute to the public debate about generative AI and to outline what is needed for safe and responsible deployment. The AP realises that technological developments follow each other at lightning speed, and tomorrow's knowledge has already moved on from today's knowledge. This vision paper is, therefore, by no means a static analysis, but should be seen as an invitation to discuss this topic with each other. As a society we must shape the role therein for generative AI.

2. Generative AI: the state of the technology

Generative AI has become an integral part of our day-to-day lives. But what we exactly mean by generative AI can vary from conversation to conversation, sometimes causing confusion. This chapter explains how generative AI will be approached in this vision paper and outlines the latest technological developments.

What is generative AI?

Generative artificial intelligence (Generative AI) is a form of AI that is capable of generating new data. The most popular generative AI applications create texts and images that are all but impossible to identify as having been generated by AI. This vision paper is about AI models that are able to generate realistic data and about the systems and applications that these models are part of.

Generative AI models can generate and all kinds of different forms of output and can be guided therein. These models can serve as a foundation for many different specialist applications and can be used for all kinds of purposes. These types of models are often referred to as 'foundation models' or 'general-purpose AI models'. In our chosen approach, models with those denominators all fall under generative AI.

Beyond the scope of this vision paper lie many other forms of AI. AI technology has been in development for a long time and is, for example, widely used for classification. The AI is then only trained to recognise patterns, rather than generate them.

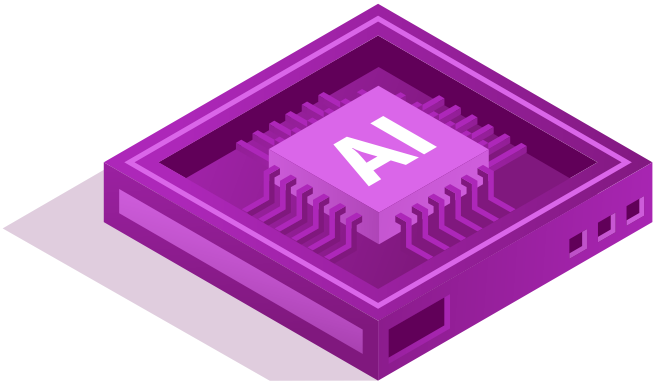
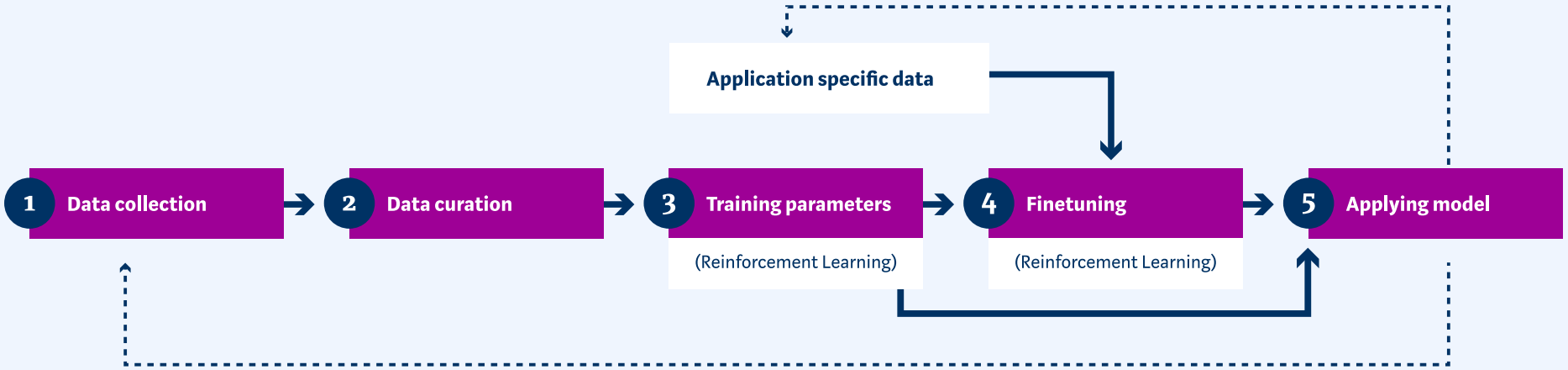


FIGURE 1 | SCHEMATIC REPRESENTATION OF GENERATIVE AI CHAIN



Simply put, training a generative AI model can be seen as a set of sequential steps. Because generated data is often collected again for training, there is a feedback loop.

1. **Step 1: Data collection.** Sample data is collected, often through scraping.
2. **Step 2: Data curation.** Unwanted examples (such as personal data or hateful content) are removed in a curation step.
3. **Step 3: Training parameters.** The parameters are trained on the basis of patterns in sample data, possibly through reinforcement learning.
4. **Step 4: Finetuning.** Finetuning adapts the model to a specific application or limitation.
5. **Step 5: Deployment of the model.** The outcome of these steps is a trained generative AI model that is deployed in an application.

Technological developments

The technology behind generative AI is still in full development. In the figure below, we give an overview of a number of important developments. This overview is not complete but gives an impression of developments within different layers of technology. We distinguish between developments in the models themselves (model layer), developments with regard to the systems in which the models are used (system layer), and the applications through which those systems make contact with the environment (application layer).

In all three layers of technology, we have seen many developments in recent years. Within the model layer, which ultimately forms the core of the technology, scientists and developers are experimenting with improvements in training algorithms and adjustments to model architecture.[1][2][3] The fact that many models – including advanced ones – are publicly available for download means that everyone around the world can build on this. Within the system layer, innovations contribute to further exploiting the model capabilities. For example, connecting a model to a database can cause a system to use memory, rather than just reacting to the last input[4][5]. In the application layer, we see a lot of innovation in the way the systems interact with the

FIGURE 2 | SCHEMATIC REPRESENTATION OF THE TECHNOLOGY

	Model layer	System layer	Application layer
Description	The trained model is a collection of files containing the model architectecture and the learned patterns.	The model is embedded and run within a system. The input and output are processed and possibly adapted.	The system interacts with the environment in the form of an application, such as an chat app.
Technological development	**Model size** Increasing number of parameters in the models	**RAG architecture** Database communication for explicit information needs	**Multimodal interaction** Vision and audio as a medium for interaction
	Multimodal Next to text, models now handle visual and audio data	**Chain-of-thought** Breaking inference into structured steps	**Embedding into tools** Interactive assistence into existing code, text or visual editors
	Mixture of experts Architecture stimulating build up of specific expertise	**Local deployment** Running systems on own hardware	**AI-agents** Autonomous execution of actions based on instructions. Such as generating code and subsequently running it
	Input shape and size Model allows for increased input size, including context	**Outputfilters** Filtering step to prevent unwanted outcomes	
	Reinforcement Learning Learning from human or reflective feedback	**Model Context Protocol** Approach to extend modelinput with context	

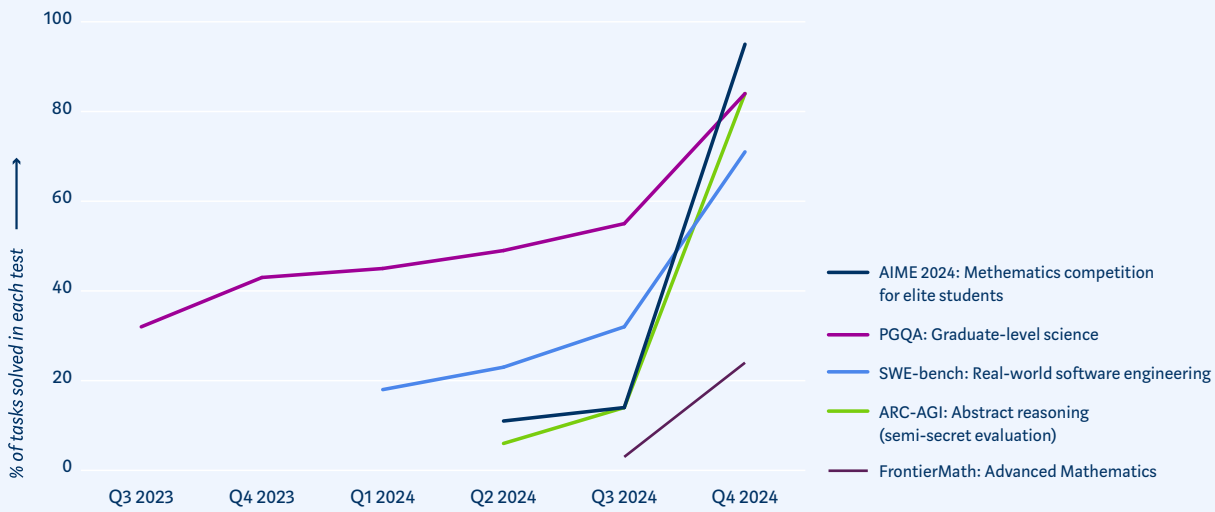
1. Mistral AI (December 2023). 'Mixtral of experts'
2. Meta (December 2024). 'Large Concept Models: Language Modeling in a Sentence Representation Space'
3. Deepseek (January 2025). 'DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning'
4. Cohere. 'Retrieval Augmented Generation (RAG)'
5. OpenAI (September 2024). 'Learning to reason with LLMs'

environment. For example, there is a lot of attention for autonomous AI Agents[6], i.e. applications that perform tasks autonomously, which we will go into in greater detail in Chapter 4 regarding trends.

Qualities of generative AI

As a result of the above developments, the quality of generative AI output has rapidly increased. To measure this, models and systems are subjected to all kinds of tests and benchmarks, for example with multiple-choice questions about biology and physics[7]. By 2024, progress in generative AI technology seemed to be slowing down as an ever-increasing amount of information on the Internet is low-quality output from generative AI. This prevents the collection of high-quality training data. Since December 2024, there has been an acceleration in the increase in quality. Reinforcement learning is an important part of this. This gives the model the freedom to develop response strategies, and an incentive to build on strategies that provide good answers. A model can learn in this way, for example, by looking back on its own answers and possibly also improving them[8].

FIGURE 3 | KEY BENCHMARKS SCORES OVER TIME



SOURCE GRAPH: INTERNATIONAL AI SAFETY REPORT (BENGIO ET AL., 2025)

The question is to what extent these developments will translate into AI models capable of many more tasks in the coming years. There is still great uncertainty about the further development of the qualities of this technology[9]. We can assume that, regardless of the progress in the model layer, many new applications will be added in the coming years on the basis of the existing models. Further progress in the model layer will, therefore, only reinforce this.

6. Anthropic (December 2024). 'Building effective agents'
7. Rein, D., Hou, B. L., Stickland, A. C., Petty, J., Pang, R. Y., Dirani, J., ... Bowman, S. R. (2024), 'GPQA: A Graduate-Level Google-Proof Q&A Benchmark.' First Conference on Language Modeling.
8. Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., ... & He, Y. (2025), 'Deepseek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning,' *arXiv preprint arXiv:2501.12948*.
9. Bengio, Y., Mindermann, S., Privitera, D., Besiroglu, T., Bommasani, R., Casper, S., ... & Zeng, Y. (2025), *International AI Safety Report*. arXiv preprint arXiv:2501.17805.

3. Deploying generative AI: market and applications

Generative AI is rapidly taking hold in both the private and public sectors. Dutch citizens are also increasingly using applications based on generative AI for personal purposes. The promise is great and expectations are high, leading organisations across the board to prepare for a further increase in use in the near future.

The market for generative AI has experienced explosive growth since the end of 2022. The launch of ChatGPT by Open AI ushered in a period of unprecedented rapid societal adoption, the end of which is not yet in sight. This has also led to fierce competition between technology companies worldwide, so that the development and deployment of generative AI has now become closely intertwined with geopolitical developments.

Many industries are now working and experimenting with generative AI. It is important to track the deployment of generative AI in order to properly assess the risks and safety issues involved. Models are usually extensively tested before being placed on the market, but their use in practice (and the applications that build on the models) can never be fully replicated – while the risks and incidents manifest themselves precisely there. It is, therefore, important that we get a good overview of this, so that we can learn what is needed for responsible use and how to mitigate the main risks.

Private sector

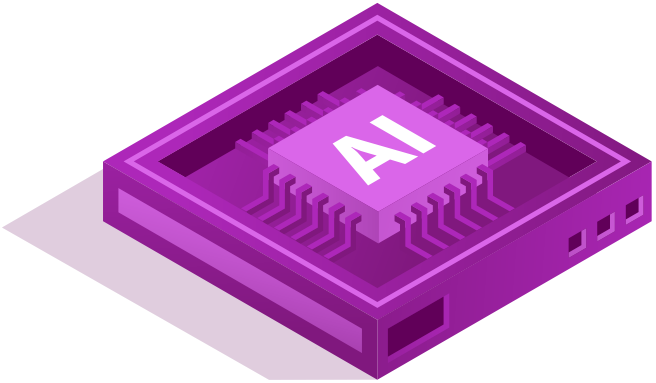
Within the private sector generative AI is deployed for all kinds of business processes. In McKinsey's *The State of AI 2024*, 65% of organisations surveyed say they regularly use generative AI: a doubling compared to the same study a year earlier.[10] Capgemini paints a similar picture in its research on the use of generative AI in different sectors, reporting that the use of generative AI has grown significantly in all domains in one year.[11] Organisations deploy generative AI across a wide variety of business processes, often with the goal of increasing efficiency and productivity – from IT to logistics and from HR to marketing. This illustrates the versatility of generative AI and also shows that people in very different positions in their work will have to deal with a form of generative AI.

Public sector

The public sector is experimenting with generative AI, but this is often still in its infancy. The latest quick scan by the Netherlands Organisation for Applied Scientific Research (TNO) from July 2024 shows that AI is increasingly being used in public services in all kinds of domains.[12] The share of generative AI was still low at the time of the research, but we can partly explain this due to the fact that many experiments were still in their infancy. The individual use of civil servants was also not included. TNO also points to the risk of 'shadow IT': technology or software used by individual employees within an organisation without formal approval from the IT department. The government-wide position (April 2025) on the use of generative AI within government

10. McKinsey (May 2024). 'The state of AI in early 2024: Gen AI adoption spikes and starts to generate value'
11. Capgemini Research Institute (2024). 'Generative AI in organizations in 2024'
12. TNO (June 2024). 'Increasing use of artificial intelligence by public authorities'

organisations is also important in this context, stating that the use of generative AI is encouraged, provided that it is agreed with suppliers and not under consumer conditions.[13] Based on this view, it is likely that the use of generative AI in the public sector will increase in the coming period.



Personal use

Regarding the personal use of generative AI, a generational gap is visible. More than one in three Dutch people now regularly use generative AI[14]. The differences between age groups are considerable. For example, more than 50% of respondents aged between 18 and 34 indicate that they use generative AI, while this is only 10% in the 65-to-75 age group. This generational difference is not strange. Today's young people have grown up with the internet and are, therefore, often more confident to experiment with new technologies in their daily lives[15]. People mainly use generative AI tools for searching and collecting information and generating ideas. A large majority of Dutch people believe that generative AI will make their work easier (77%) and more fun (77%) in the next two years[16].

Practical examples

In practice, the deployment of generative AI has many different forms of expression. The fictional examples below are based on existing applications and intended to give an impression of what is currently possible with generative AI in a different contexts, and how this impacts users and environment.

Accountancy

Tailored financial advice

Accounting firm Loukili & Laheij has taken significant steps since purchasing tax software based on generative AI. Standard reports and financial forecasts can now be generated in minutes. The auditors of Loukili & Laheij only need to validate these, so they can spend more time on personal contact and tailored financial advice. The software also supports them in this by making suggestions for possible solutions based on historical customer data and customer demand. For customers, this is of great added value. The software also proves to be effective in error detection.

At the same time, this also entails a risk for Loukili & Laheij, i.e. the risk of employees relying too much on the software and assessing output insufficiently critically, while critical assessment is particularly important in light of annually changing legislation and tax systems.

13. Dutch government (April 2025), 'The whole-of-government position for the deployment of generative AI,

14. KPMG (2024). 'Algorithm Trust Monitor 2024'
15. Algosoc (2024), "AI Opinion Monitor"
16. Deloitte (October 2024). 'Trust in generative AI: a European and Dutch perspective'

General practicioner

Chatbot as a doctor's assistant

The *ZorgModern* medical practice has opted for a chatbot as a doctor's assistant. Not only to save money, but also because good staff was hard to find. The chatbot is now the first point of contact for patients of the practice. Through an app, patients describe their complaints. The chatbot asks a few follow-up questions and then makes a summary that is sent to the doctor. The doctor assesses the message and sends a personal answer to the patient, either advising the patient on what to do or asking them to make an appointment.

Patients are generally satisfied with the short line of communication between them and the doctor. However, older patients feel less at home at this practice. Due to their lower digital skills, the app feels like an obstacle to good care.

Private use

Personal holiday advisor

Spain, Germany, France, or a long trip to Thailand or South Africa? Is it the right season for that? And are three weeks enough to see all the sights? These questions that you used to spend a few evenings googling can now be answered in a few minutes by a language model. And with the right prompts, you can generate a detailed travel plan that meets all your wishes and requirements in no time: from recommendations for hotels and restaurants to tips for nature parks and cultural highlights. Or information about local eating habits and emergency numbers. And even during your trip, the digital personal holiday advisor is always there for you.

Education

Digital reading coach

At De Ekster primary school, 8-year-old Noor is taught every week by a personalised digital reading coach. Noor has dyslexia and therefore has trouble keeping up in the regular reading comprehension classes. It was frustrating and demotivating. On the computer, she can concentrate better and she gets to read texts that are specially tailored to her. For example, a text about Disney, which she is a big fan of. The texts that her digital reading coach generates not only match her interests, but also contain specific words and spelling patterns that are tailored to her level. For Noor, reading has not only become more fun, but she also progresses faster.

In most of the above fictional cases, personal data will be processed. If personal data are processed, the GDPR applies and the controller must have a legal basis for processing under the GDPR. The above personal data may have been processed on the basis of consent. Having a legal basis is one of the conditions for using generative AI responsibly and lawfully. For a run-down of some of the GDPR preconditions, see the guidance document that the AP has drawn up thereon ("GDPR Preconditions for generative AI").

4. Trends in generative AI: emerging applications

Generative AI has the potential to provide a fundamental change in the way we work, communicate and gather information. From autonomous performers (AI agents) to search engines and new virtual social actors we interact with. Technology offers new opportunities in all these areas but also poses challenges and can directly and indirectly impact fundamental rights and values.

The impact of generative AI is seen in practice and is related to the form in which it is deployed by organisations and individuals. In this regard, there is still plenty of experimentation with new application possibilities. In the chapter "State of the Technology", we described the technological developments in the field of generative AI. These developments underlie the possibilities of application in practice. But in what ways does generative AI bring about change? That depends on how the technology is deployed. In this chapter we cover some of the trends in forms of application that we see with respect to generative AI.

Generative AI as an autonomous agent (*AI Agents*): the idea of allowing AI models to autonomously perform actions to automate tasks is not new, but generative AI creates new opportunities for this. First, prompts using human language can be better interpreted by large language models. Goals are divided into chunks and carried out step by step. In addition, these models can generate instructions for all kinds of different instruments, including to operate the mouse and keyboard of computers virtually[17]. Together, these improvements bring the field closer to autonomous agents, or AI agents. One AI developer claims that by the end of 2027, there will be models that can do almost anything most people can[18]. These autonomous agents can always be on and working in the background, making them virtually invisible. For companies, autonomous agents can offer a solution, but the deployment may also have an impact on employment, among other things.

Generative AI as a 24/7 personal assistant: the integration of generative AI into existing software and the combination of more efficient models with stronger mobile hardware and larger internet bandwidth will make it possible to use (local) generative AI as an assistant that is continuously on. If people want to make continuous notes during the day, the AI assistant should listen in continuously. As a result, it could pick up a lot of personal details, also from persons who have no awareness that the assistant is on or cannot exercise any control over it.

Generative AI as a social actor: the output generated by generative AI is very similar to something made by humans. As a result, you may feel that you have human contact when you interact with a virtual (chat)bot. This gives AI the opportunity to operate on a new level. Some experience it as a virtual friend or loved one in so-called 'companion apps', or as a digital therapist who helps with mental problems. Also, generative AI as a teacher can personalise a curriculum for a student or simply help to practise a job application, by stimulating a social setting. Finally, chatbots are increasingly being used to take over front office-like tasks, such as interacting with customers and making appointments.

Generative AI as a tool for optimisation and efficiency in organisations: one of the promises of generative AI is that it can take over routine cognitive tasks from employees. It can, for example, transcribe and summarise (online) meetings or answer customer queries about products. This could potentially increase productivity, but

17. Anthropic. 'Computer use (beta)',
18. ArsTechnica (January 2025). 'Anthropic chief says AI could surpass "almost all humans at almost everything" shortly after 2027'

it remains to be seen whether this is actually the case. Performing routine tasks can actually provide relief for an employee before going back to cognitively more demanding tasks that cannot yet be automated. The routine tasks thus contribute to overall productivity. If these disappear, it is not a given that more time for more demanding tasks also ensures more or better execution of these tasks[19].

Generative AI as a researcher: generative AI contributes to the creation, control and elaboration of ideas for research. This may result in an acceleration in research findings in many directions, for example in physics, mathematics and biology. Generative AI is especially efficient in detecting patterns and connections and can review and revise its own suggestions to a certain extent. This makes it ideal for academic disciplines that build on previous sources. At the same time, generative AI is not necessarily to be expected to cause a paradigm shift, where in fact it may be needed for major scientific breakthroughs.

Generative AI as a basis for image and sound: more and more online content is generated by generative AI models and applications. As these are getting better and better, this content is often difficult to distinguish from non-generated content[20]. A logical consequence of this is that people have less and less confidence in the authenticity of digital information. Also, offering publicly available models makes it relatively easy to exploit the technology and generate content that is misleading, without direct evidence that it is generated content. This is because all the 'new' content that is generated reproduces the sample data it has been trained to use. This means that existing biases contained in the sample data will also be reflected in the content generated by the AI. These prejudices are, consequently, further perpetuated in society.

Generative AI as a search engine: generative AI is often used as a source of knowledge. This is where it enters the domain of the more traditional search engine, which returns a list of relevant existing sources based on search terms. An important difference is that generative AI also processes the content from public sources, or has processed it during training. Based on this, new information is generated to answer the query. This makes it easy for the user to get information, but there is also a risk in that convenience. There is no guarantee that the information generated is correct, and the simple presentation makes it less intuitive for people to estimate or find out. Partly for this reason, we have recently started to see a development where generative AI search engines refer more explicitly to existing sources. At the same time, traditional search engines are also moving towards generative AI, with a generated summary of found information appearing at the top of the results.

19. TNO (January 2025). 'Working health and safety technology radar'
20. Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A., Krumhuber, E. G., & Dawel, A. (2023), 'AI hyperrealism: Why AI faces are perceived as more real than human ones.' Psychological Science, 34(12), 1390-1403.

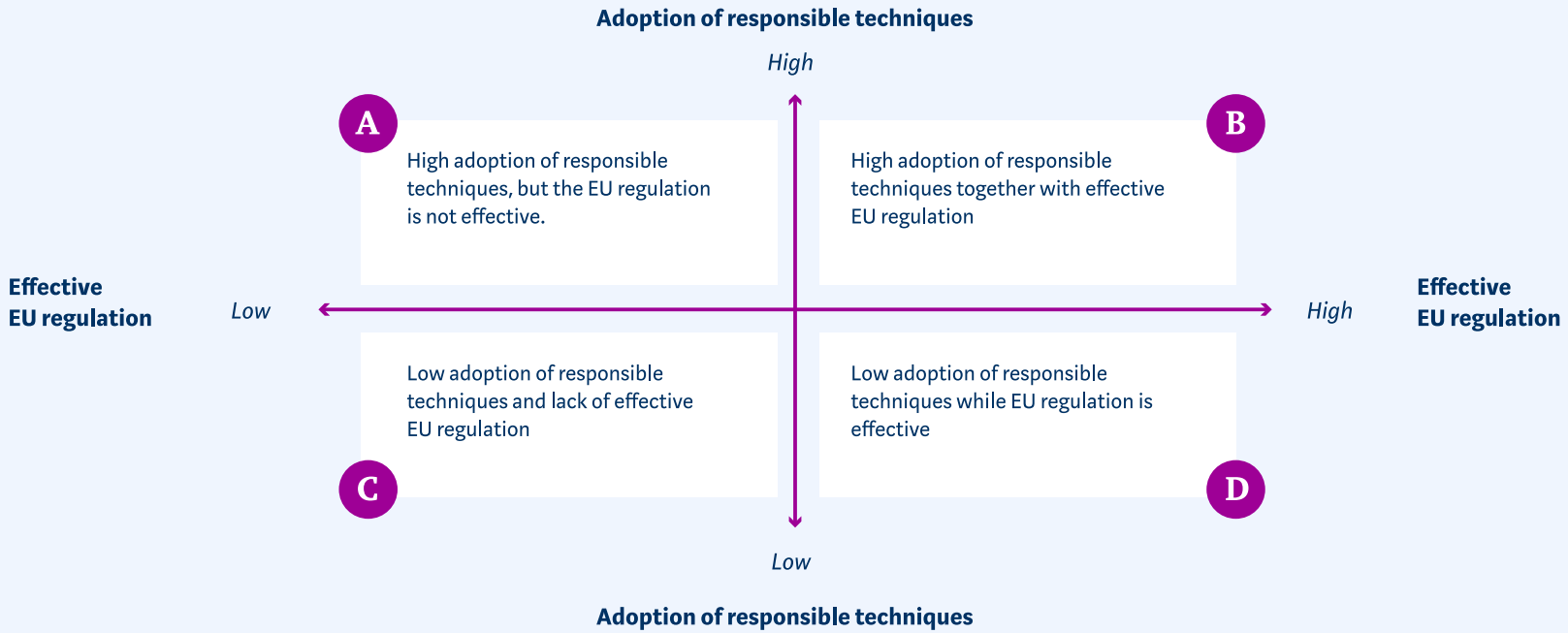
5. Future scenario's: generative AI in 2030

With a technology developing so rapidly, it is difficult to predict what the future will look like, even in the relatively short term. This chapter explores four future scenarios for how generative AI may be embedded in our society by 2030.

The scenarios are laid out along two axes of core regulatory and technological uncertainties. The first key uncertainty concerns the use of responsible technological solutions. Can researchers create technological solutions to mitigate the risks associated with generative AI? And do these solutions successfully move to the market? The second key uncertainty concerns the effectiveness of regulation in the European market. In the coming years, it must be shown whether the recently adopted European laws to manage (generative) AI are effective in practice, also in combination with existing legislation.

By combining the two axes, we arrive at four scenarios, each of which presents a different vision of the future. Below we outline the future for each scenario with the year 2030 as a horizon:

FIGURE 4 | FUTURE SCENARIOS OF GENERATIVE AI



A

Scenario A: 'Survival of the fittest'

Extensive use of responsible technological solutions, but ineffective regulation within the EU

There is a thriving market for responsible generative AI solutions. This is not so much due to legislation and regulations, but due to the continued demand from the market. This is the direct result of increased awareness and knowledge about the opportunities and risks of generative AI in society. After a few incidents in previous years, people have become aware of the importance of technology that supports democratic values and protects people's fundamental rights. The widespread use of responsible solutions has been a strong bottom-up movement, with the development of European alternatives stemming mainly from the idealism of some entrepreneurs and private financiers. It is mainly conscious citizens who benefit from this. At the same time, AI-related incidents continue to occur on a regular basis: the use of generative AI has increased rapidly, but lagging enforcement capacity has made citizens unprotected when their rights are violated by an AI system. Citizens depend on the benevolence of companies and cannot enforce the rights they have on paper. As a result, the responsible solutions increasingly turn out to be a sham: when things go wrong, injustice is hard to repair.

B

Scenario B: 'Values at work'

Extensive use of responsible technological solutions, with effective regulation within the EU

Companies, organisations and consumers are making full use of value-driven generative AI applications. The EU legislation that caused many eyebrows to frown at its introduction a few years ago has paid off. A thriving AI ecosystem has emerged, partly due to strong European investments. Artificial intelligence is an integral part of the curriculum of European primary and secondary schools. Interdisciplinary scientific research into AI has also received a boost, as a result of which the number of patent applications for AI-driven innovations has risen sharply in recent years. Generative AI applications with personal data can be used in a GDPR-compliant way because this data can be removed from models. Supervisory authorities play an important facilitating role in all this. Close collaboration at national and EU level, among other things, has created an internationally competitive level playing field, with a strong focus on harmonisation and constant exchange between supervisory authorities and the market. As a result, responsible generative AI applications have become the norm – and the best available products. Clear standards and frameworks also ensure that risks are identified and mitigated in a timely manner. There is a strong learning culture. Incidents are systematically used to improve systems and applications, increasing trust in generative AI in society and enabling citizens to exercise their rights.

C

Scenario C: 'Under the radar'
Little use of responsible technological solutions, with ineffective regulation within the EU

The regulation that aimed to make Europe a value-driven AI continent has overshot its goal. Legislation and regulations proved too complex in practice and public investment lagged behind, leaving many promising initiatives for the development of responsible generative AI stranded in the early stages of development. At the same time, the use of generative AI has not decreased. There are still many applications on the market for which it is not clear whether they are compliant due to conflicting laws and regulations, but which are frequently used due to the lack of effective control and enforcement. This leads to malfunctioning systems that often remain under the radar. Due to a lack of transparency and knowledge, citizens often do not know that there is unlawful treatment by an AI system, or fail to have erroneous personal data removed from models. Providers impose censorship restrictions on their language models, providing citizens with selective information on a wide range of topics. Regulators fail to protect citizens from this. Supervision is fragmented and the knowledge gap between policy makers, supervisory authorities and AI developers is huge. There is a lack of a common language, leaving citizens at the mercy of the vagaries of the market and low trust in generative AI. Incidents seem to affect particularly vulnerable groups in society. This is partly due to the increased gap between those with low and high digital literacy. Resistance among citizens is growing and calls for human contact are getting louder and louder.

D

Scenario D: 'Blindfolded conductor'
Little use of responsible technological solutions, with effective regulation within the EU

The mismatch between regulation and market developments creates a difficult European situation. The regulation of generative AI is effective in the sense that fundamental rights and public values are well protected. There are clear standards, citizens are informed in a clear way when a generative AI system is deployed and robust cooperation between supervisors provides a good overview of risks and the impact of incidents on citizens remains relatively limited. At the same time, there is little contact between regulators and market players. Laws and regulations have become increasingly detached from global technological and economic developments. The relatively small market that the EU represents for large market players means that they primarily focus their innovations on markets elsewhere. Companies and organisations are reluctant to use generative AI for fear of fines in the event of incidents. At the same time, sectoral markets where Europe was originally the market leader are losing their international position by limiting their innovation opportunities. The gap between the EU and other countries is growing, to the frustration of entrepreneurs and citizens who see what is possible abroad. The promise of digital strategic autonomy is proving difficult to live up to. Not only does this have economic consequences for Europe's position on the world stage, but it also puts an increasing strain on our acquired freedoms and fundamental rights.

Desired direction

Of the above future scenarios, the AP sees scenario B 'Values at work' as the most desirable scenario. In this scenario regulation is effective and we jointly ensure that value-driven innovation is achieved. The concluding chapter on responsible deployment further elaborates on this preferred future scenario, asking what it takes to get there and how the AP can contribute to it.

Box 1: Impact of public model access on risks

For the future development and deployment of generative AI, the open-source or closed-source availability of advanced models is an important and determining factor. More specifically, it is about whether the so-called model parameters become available, i.e. open-weight models. An open-weight model can be freely downloaded on the internet and used or adjusted with complete freedom. Hundreds of thousands of open-weight generative AI models are already available[21]. The following overview identifies a number of relevant implications in terms of risks arising from the disclosure or non-disclosure of the models' weights. This is a simplified view[22].

This comparison shows that making open-weight models available carries different risks than closed-weight models. Risks are added, while other risks are mitigated. Currently, both variants are in circulation. As a society, therefore, we will have to consider the risks of both variants.

According to the AP, this is the time to influence risks by encouraging or discouraging the disclosure of models. Determining the desired direction requires broad public discussion.

FIGURE 5 | COMPARISON OF OPEN-WEIGHT MODELS VS. CLOSED-WEIGHT MODELS

Risk category →	*Open-weight models*		*Closed-weight models*
Privacy and data protection	Personal data within the models is distributed	⚠ ↔ ✓	Provider can recall and correct models if needed
	Local deployment possible, less online traffic of data	✓ ↔ ⚠	Lots of input and output traffic between provider and users
Cybersecurity	GenAI cyber defense and attacks possible for anyone	↔	GenAI cyber defense and attacks on the basis of access
	Risk of hidden actions within models or accompanied malware	⚠ ↔ ✓	Provider controls the safety of the model
	Possibility for community based tests and controls	✓ ↔ ⚠	Online communication introduces vulnerabilities
Bias, stereotyping & discrimination	Comminuty can contribute to model alignment	✓ ↔ ⚠	Provider largely controls the models norms and values
	Many biased models availabel without a point of contact	⚠ ↔ ✓	Provider can be held liable for model bias
Usage with malicious intentions	Easier to generate misinformation at scale	⚠ ↔ ✓	Provider can limit access in case of misuse
	Powerfull AI is available and can do significant harm	⚠	
Autonomy and market concentration	More indepent, lowers threshold to enter market	✓ ↔ ⚠	Dependence on provider which grows with usage
	Power and control on applciation level	✓ ↔ ⚠	Providers obtain position of power over applications
	Anyone can drive innovations, more diverse and less orchestrated	↔	Providers determine the direction of innovation.

⚠ Hoog risico

✓ Laag risico

21. Hugging Face. 'Models'
22. This representation provides a first impression, but the impact on risks is incomplete and often more complicated. For example, there are services to run and fine-tune closed-weight models in a closed environment, creating hybrid solutions with a different risk picture.

Box 2: Generative AI and the GDPR

When generative AI models are trained on (untargeted) scraped data, the data sets for training those models often contain personal data. This means that the GDPR may apply from the first in the generative AI chain: data collection. But the GDPR may also apply later in the generative AI chain, for example when fine-tuning a foundation model or when implementing or using a generative AI application. The diagram below sets out a number of GDPR preconditions for the responsible development and use of generative AI applications that involve the processing of personal data. These preconditions have been further elaborated in the AP's guidance document "GDPR preconditions for generative AI".

From these preconditions, we can conclude that compliance with GDPR requirements still requires the necessary attention from controllers. It is plausible that special categories of personal data have been unlawfully scraped for the training of foundation models. It is also not a given that scraping of personal data is always permitted. Developers, fine-tuners and users of generative AI must recognise and consciously deal with the challenges surrounding the fulfilment of data subjects' rights.

This standard is the position of the AP at the time of publication. However, the AP closely follows the European standards regarding generative AI and actively contributes to this. This means that these preconditions can still be adjusted at a later stage.

FIGURE 6 | GDPR REQUIREMENTS FOR GENERATIVE AI

6. Legal framework and instruments

Generative AI is a technological development that cuts across all sectors and, therefore, does not fall under a single law. Different legal frameworks apply to the development and use of generative AI models and applications. Generative AI does not fall into a legal vacuum, but must conform to both technology-specific regulations and domain-specific legislation.

Depending on the sector in which the generative AI model is developed or deployed, different laws and regulations may apply. Consider, for example, financial legislation, consumer law, media legislation, healthcare legislation, administrative law and copyright protection.

New EU legislation

In addition to these existing laws and regulations, many new EU regulations have recently been developed to regulate the deployment and use of new technologies. This primarily concerns the entry into force of the GDPR (2016) to protect the personal data of citizens, but in addition, various laws have been adopted in recent years to, among other things, regulate digital platforms (DSA), counter market dominance by Big Tech companies (DMA), ensure cybersecurity (NIS2, CSR), ensure fair access to data and its protection (DA, DGA), and ensure that AI is developed and deployed in a responsible manner (AI Regulation). There are also several other legislative initiatives on the EU agenda that may also have an impact on generative AI, such as the Digital Fairness Act.

Fundamental rights of citizens

Furthermore, the development, deployment and use of generative AI also affects the fundamental rights of citizens[23]**.** On the one hand, fundamental rights are based on a vertical effect. That is, they protect citizens from government action. On the other hand, fundamental rights also have a horizontal effect. This means that citizens can invoke certain fundamental rights vis-à-vis other citizens (and businesses). Generative AI can contribute to the promotion of fundamental rights, but it can also pose risks to the protection of fundamental rights.

23. Charter of Fundamental Rights of the European Union.

Risks to fundamental rights

Generative AI models and applications can infringe fundamental rights of citizens if developers and users do not sufficiently respect these fundamental rights. Consider, for example, the training of personal data in a generative AI model or its processing in generative AI applications. If the protection of personal data is not sufficiently guaranteed, this could potentially constitute an infringement of the right to the protection of personal data. The development and use of generative AI models and applications may also affect the right to protection of intellectual property rights, for example when generating images or videos through generative AI

Strengthening fundamental rights

Generative AI can also strengthen certain fundamental rights. If generative AI applications bring about breakthrough changes in health care or education, this could potentially lead to improved accessibility for citizens within those sectors. In this way, generative AI indirectly improves fulfilment of the right to education and the right to healthcare.

Together with international (non-binding) policy agreements, these laws and regulations form the frameworks for the responsible development,

deployment and use of generative AI[24]**.** The figure below serves as a source of inspiration for control of the development, deployment and use of generative AI, without wanting to be exhaustive.

Lawfulness of data collection

With the launch of the first large generative AI applications, many questions have arisen about the legality of the data sets used to train large generative AI models. Data collection for these large generative AI models usually takes place via untargeted scraping. When scraping information from the internet, special categories of personal data can also be scraped. The GDPR puts a processing ban on these special categories of personal data, unless an exception can be invoked[25]. Because there is usually no direct relationship between the scraper and the data subject, it is often not possible to invoke consent as a ground for exception. It is possible that these special categories of personal data were manifestly made public by data subjects themselves[26]. However, it can sometimes also happen that someone else places special categories of personal data of a data subject on the internet. It cannot therefore be ruled out that, in limited cases, special categories of personal data may also have ended up unlawfully in generative AI models.

The AP takes into account that scraping these special categories of personal data may be secondary in nature to unlawfully placing them online. Furthermore, unlawfulness at an early stage of the generative AI chain does not always mean that the use of those models by other parties is also unlawful. However, we are aware of the risks involved in collecting special categories of personal data to train generative AI models. That is why the AP's focus in this vision paper is on the responsible use of generative AI in the Netherlands. Furthermore, the AP contributes to further standardisation of generative AI in an EU context. In this way, we contribute to the socially responsible creation of generative AI applications.

---

24. Examples of international agreements that affect the responsible use of generative AI are OECD Principles on Artificial Intelligence (2019), UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) and the EU Code of Practice on Disinformation (2022).
25. See Article 9 of the GDPR.
26. For a further explanation of the apparent disclosure in scraping, see the AP's Handbook on Scraping by Individuals and Private Organisations.

FIGURE 7 | INSTRUMENTS FOR RESPONSIBLE GENERATIVE AI

	Development phase	Deployment phase	Application phase
Risk management	• Impact-assessments (eg. DPIA, FRIA) (GDPR, AIA) • Security by design/secure development lifecycle (NIS2, ISO 27001) • Neutral data intermediairs (DGA)	• Impact-assessments (eg. DPIA, FRIA) • Data quality assessment (GDPR, AIA) • Conformity assesment (AIA)	• Impact-assessments (eg. DPIA, FRIA) • Transparency requirements (eg. deepfakes) (AIA)
Technical control measures	• Model card and technical documentation (AIA) • Data quality and interoperability requirements (DA) • Pseudonymization/anonymization (GDPR) • Fairness & bias detection	• Fairness & bias detection • Privacy Enhancing Technologies (eg. differential privacy) (GDPR) • Interoperability (DA) • Technical documentation (AIA)	• Fairness & bias detection • Monitoring & logging • Explainability & interpretability
Organisational control measures	• Privacy by design, Privacy by default (GDPR) • Appointing EU representative (AIA)	• Information security plan • AI-governance • Privacy by design, Privacy by default (GDPR)	• Training and awareness (AI literacy) (AIA) • Human intervention (GDPR) • Privacy by design, Privacy by default (GDPR)
Legal control measures	• Register of data mediation services (DGA) • Regulatory sandbox (AIA) • Prior consultation (GDPR) • Protection of intellectual property rights, trade secrets and confidential business information (AIA)	• Records of processing activities (GDPR) • Algorithmic audits • Independent audits (DSA)	• Incident reporting (eg. Data breach) (GDPR) • Post-market monitoring (AIA) • Points of contact/complaint mechanism (DSA, GDPR, AIA)

7. Values at work: towards responsible use of generative AI

The AP strives for a future in which the society enjoys the benefits of responsible forms of generative AI based on effective regulation (future scenario B). This chapter elaborates on how fundamental rights and values are protected in this future scenario and how this technology can be of service to society and social interest.

On the one hand this future scenario has the following overarching characteristics for the ecosystem of generative AI (at systemic level): European digital autonomy, social resilience, democratic governance, a functional market for responsible solutions and the ability to correct throughout the generative AI chain. On the other hand, individual AI applications in this future scenario show the following characteristics (at application level): open and transparent, the risks are identified, assessed and mitigated, clear purpose description, the systems and data are under controlled management and it is lawful.

In the previous chapters generative AI is described from a helicopter view. We discussed what the technology is (chapter 2), how it is deployed (chapter 3), and what trends are currently playing out (chapter 4). In addition, possible futures are described in the form of scenarios (chapter 5). In this chapter we provide details on how fundamental

FIGURE 8 | VALUES AT WORK: PERSPECTIVE FOR RESPONSIBLE APPLICATION

Characteristics of a situation in which fundamental rights and values are protected and the technology of generative AI is of service to society and the social interest

Systemic level

Value driven generative AI is an engine of innovation through:

- European digital autonomy
- Social resilience
- Democratic governance
- Functional market for responsible solutions
- Ability to correct throughout the generative AI chain

Application level

Applying generative AI responsibly is possible through characteristics:

- Open and transparent
- Risks identified, assessed and mitigated
- Clear purpose description
- Systems and data under controlled management
- Lawful

rights and values are protected in the desired direction ("values at work", scenario B), and how this technology can be of service to society and social interest.

We describe the layout of scenario B ("values at work") through characteristics. These characteristics enable steering. By realising them we create a societal foundation to move forward responsibly with generative AI, and utilise the potential benefits to welfare and well-being. First, the chapter focusses on characteristics at a (societal) systemic level and second the focus is at the application level. This vision paper concludes with actions that the AP is taking to contribute to the desired course.

Characteristics at systemic level: the role of generative AI in Society

In the future scenario "values at work" value driven generative AI forms an engine for innovation. A number of systemic characteristics contribute to this:

- **Systemic characteristic 1: European Digital Autonomy**

  Striving for this characteristic allows the EU to steer and intervene in the way this technology affects society. This takes into account that a strong dependency on an organisation can make it more difficult to correct it. Within the EU, activities are hold accountable to European rules, independent of pressure on regulation from outside the union. The Draghi report describes a degree of autonomy and competitiveness as the foundation for continuing to protect fundamental rights in the future[27].

  Possible instrument: stimulate the rise of European providers of generative AI.

- **Systemic characteristic 2: Social resilience**

  Striving for this characteristic makes that people throughout society know the basics of how generative AI works and, therefore, have realistic expectations of the technology. People can also look critically at its development and deployment[28]. This allows individuals to assess usage risks, but also to see how aspects of everyday life are affected by generative AI, such as our information provision

  Possible instruments: work on high AI literacy in society and periodic external audits of generative AI systems.

- **Systematic characteristic 3: Democratic governance**

  Striving for this characteristic ensures that, through the democratic process, the wishes and concerns of the general public are effectively incorporated into policy[29]. Broad social discussion underpins the direction, for example on the question of whether open-weight models are preferable to closed-weight models. Public contributions can also consist of tests or benchmarks for generative AI to contribute to controlled impact from various perspectives.

  Possible instruments: work on adequate expertise on AI system in parliament, low-threshold platforms for specialist benchmarks and stimulate societal debate on generative AI.

- **Systemic Characteristic 4: A well-functioning market for responsible solutions**

  Striving for this characteristic ensures limited dependence on one or a few parties. Instead, there is a market with multiple options throughout the generative AI chain, from training to applying generative AI. Regulation creates a level playing field where responsible solutions do not detract from quality. In addition, there is interoperability between applications: it is easy for users to switch between similar applications. This also means that the user data or the accumulated profile of users is available in such a way that a lock-in position is prevented.

  Possible instruments: Aim for a varied offering of available models, stimulate interoperability between systems and enable comparison platforms.

27. European Commission (September 2024). 'The Draghi report: In-depth analysis and recommendations'
28. AP (January 2025). 'Getting started with AI literacy'
29. AP (March 2025). 'Position paper AP democratische beheersing algoritmes en AI'

- **Systemic Characteristic 5: Ability to correct throughout the generative AI chain**

  Striving for this characteristic ensures that the embedding of generative AI models into digital systems is designed in such a way that vulnerabilities or illegalities can be corrected afterwards. If a model contains personal data or produces undesirable outcomes, each party in the generative AI chain has put in place a system to correct those personal data or undesirable outcomes.

  Possible instruments: correction methods such as machine unlearning, registration of model versions within high-risk applications.

Characteristics at application level

There are several characteristics of responsible use of generative AI at application level. This is about how an organisation approaches the deployment of individual generative AI systems in the desired future scenario:

- **Application characteristic 1: Open and transparent**

  Striving for this characteristic ensures that generative AI applications are clearly recognisable and lend themselves to further analysis where desired, for example by sharing (outcomes of) various assessment criteria on request. A high level of transparency is the basis for trust and enables the user to be resilient and to exercise their rights.

  Possible instruments: ensure AI interactions are clearly indicated to users, make model cards publicly available[30].

- **Application characteristic 2: Risks identified, assessed and mitigated**

  Striving for this characteristic ensures that a thorough risk assessment precedes deployment of generative AI applications. Legal and ethical considerations are made explicit and documented. The application does not contain unacceptable risks and is transparent about increased risks. Diversity of perspectives is essential in this analysis and balancing exercise.

  Possible instruments: develop risk management frameworks for AI models, monitor the risks of AI models in practice through random sampling.

- **Application characteristic 3: Clear purpose description**

  Striving for this characteristic ensures that the goal and usage of generative AI is formulated and documented as clear as possible. This is important because generative AI can be used inherently for many purposes. The application is set up to keep the use within the planned frameworks, which is monitored after commissioning. The employees involved in the deployment of the system are well aware of the functioning of the system and the possible risks and know how to recognise them in practice.

- **Application characteristic 4: Systems and data under controlled management**

  Striving for this characteristic ensures control over the environment where the generative AI system runs and the data is being processed. Reliance on third parties is limited for important processes. In addition, it is clear and transparent where the user data is located and to whom it is available.

  Possible instruments: Use open-weight models on own hardware, use private cloud solution.

- **Application characteristic 5: Lawfulness**

  Striving for this characteristic ensures the legal requirements from the relevant legislation, including the GDPR and the AI Act and any sectoral legislation, are met. This applies to the entire process, from the collection of data to the deployment of the application. Third-party models have an acceptable level of anonymity, as described by the EDPB[31].

30. Hugging Face. 'Model Cards'

31. European Data Protection Board (December 2024). 'Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models'

Actions by the AP

In the coming period, the AP will make an effort to contribute to the desired 'Values at work' vision of the future. We will do this on the one hand by paying extra attention to generative AI within our existing activities and on the other hand by starting a number of new activities that make a positive contribution to the responsible use of generative AI in our society.

As part of our regular work, the AP contributes to clear and realistic working methods for AI, including by actively writing standards and opinions. For example, the EDPB Opinion of December 2024 and the standards for high-risk systems in the AI Act. In addition, the AP is working on digital resilience, including by investing in the level of knowledge in society. We are, for example, providing guidance on AI literacy, publishing comprehensible information on our website and organising seminars. Risk identification is also an integral part of the work of the AP. For example, via the overview of GDPR risks for generative AI and the biannual risk reports on algorithms and AI. Finally, the AP is available for prior consultation and is setting up a sandbox for applications under the AI Act.

The AP takes a number of additional steps to make responsible progress with generative AI. Organisation can reach us through a new advice centre for generative AI. We will start by identifying the questions and challenges surrounding the responsible development and deployment of generative AI. To this end, we will organise a number of meetings and set up an online channel for questions and ideas about generative AI. The information we collect provides insight into, for example, compliance related questions and top-of-mind issues within organisations. This will be the basis for further periodic dialogue on the responsible development and deployment of generative AI. Additionally, this aids in prioritising the most urgent dilemmas first.

We will also work with concrete instruments and tools that contribute to protecting fundamental values in the development and deployment of generative AI. These include EU directives, efforts to encourage the use of methods to anonymise or remove personal data from models, guidelines for AI literacy and principles for transparency. And together with other regulators in the digital domain, we are also committed to jointly explaining normative standards, so that we create as much clarity as possible and thus legal certainty for organisations that want to work with generative AI. We are giving positive examples and use cases a platform to show and inspire what responsible generative AI can look like in practice. And of course, the AP will continue to monitor illegalities that occur in the playing field of generative AI.

This is how, together, we will make the responsible use of generative AI possible.

Annex 1. Glossary

The following is a brief (non-legal) definition of the terms and concepts used:

Concept	Description
AI *(Artificial Intelligence)*	Imitating human skills with a computer system, such as learning, planning, reasoning, anticipating and deciding independently without the intervention of human intelligence.
AI model	A software component containing knowledge to arrive at outcomes (such as predictions or classifications) based on input. Models learn patterns from sample data (sample input with a correct outcome) and process new input based on this to the best possible outcome.
AI system	A system designed to work with different levels of autonomy and using an AI model.
Generative AI	A form of AI that is capable of generating new data. The most popular applications create texts and images that are no longer identifiable as having been generated by AI. A user's input often indicates what to generate.
Machine learning	Methods to incorporate patterns from data into an AI model.
Neural network	A type of AI model where an abstraction of information can be stored in the form of parameters. These parameters are numbers that are chosen by training on sample data. A neural network determines an output based on an input and its parameters.
Deep learning	A method within machine learning. This method uses neural networks with many parameters.
The open-source model	An AI model that is publicly available. In that case, it is possible, for example, to download the decision criteria of a decision tree, or the parameters of a neural network.

Concept	Description
Closed-source model	An AI model that is only made available in a way where you get output back based on the input. The exact calculation of the decision is therefore not public.
The foundation model *(also called: General-purpose model)*	An AI model that has stored patterns in large neural networks based on a lot of data. The model can, therefore, be used in many different applications. In addition, the model can serve as a 'foundation' for models for specific purposes, with the foundation model becoming specialised.
Large language model (LLM)	An AI model that has stored patterns in large neural networks based on a lot of text data. The model can generate text and can be used, for example, in question-answer situations.