Ministry of Infrastructure
and Water Management

AI Impact Assessment

The tool for a responsible AI project

Version 2.0, December 2024. In collaboration with colleagues from the Ministry of Infrastructure and Water Management (I&W) at the Office of the CIO (CDIB), the Human Environment and Transport Inspectorate (ILT IDlab and analysis department) and the Directorate-General for Public Works and Water Management (RWS Datalab).

Please address any questions or comments to teamai@minienw.nl

Table of contents

The AI Impact Assessment helps in ensuring responsible AI by design

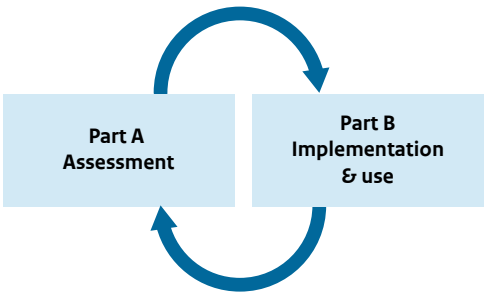
Artificial Intelligence (AI) offers opportunities, but also entails risk. It is important to have clarity with regard to the impact of an AI system before it is deployed in order to prevent unintended negative consequences. This helps to ensure that the opportunities that AI offers can be exploited to the full. In order to enable AI to be used responsibly, the ILT IDlab and analysis department, the RWS Datalab and Office of the CIO in the Ministry of I&W have developed the AI Impact Assessment (AIIA). The AIIA is intended to be used as a tool in supporting the thought process in order to increase the accountability, quality and **reproducibility** of AI deployment. The AIIA looks at obstacles in data collection, the AI system and algorithmics and takes account of relevant legislation and regulations. Once completed, an AIIA provides transparency with regard to the assessments made when deciding whether or not to use an **AI system**.

In this, we are assuming that AI is being applied in a *specific* context. If the AI system (or part of it) is used for a different purpose, it will be necessary to carry out another AIIA. Examples of this could include an image-recognition model for ships also being deployed for other vehicles.

The **AI Act** outlines a series of risk areas for AI systems, from unacceptable to minimal risk. The measures that need to be taken will ultimately depend on the level of risk. For this reason, you should always apply the risk levels outlined in the AI Act (see Appendix 1). For example, if the AI system moves from a low-risk application to a high-risk application, stricter requirements will be needed in order to comply with the AI Act.

Complete the AIIA with a multidisciplinary group
The AIIA needs to be completed by a multidisciplinary group from the organisation. This is because a different type of knowledge is needed in order to complete the different sections. The heading 'Who does what?' provides an overview of the roles that may be relevant for each chapter. Some questions will need to be answered by a data scientist and others by a legal expert.

The different parts of the AIIA
The AIIA is divided into two parts. Part A looks at the factors to be considered for using an AI system: what is the purpose and what effects are expected? This information is then used to assess the potential application of the AI system and any necessary measures. This ensures that the ethical discussion around the desirability of its application is demonstrable. Part B looks at the design, implementation and use of the AI system.



Although both parts are to be completed separately, there is a lot of crossover between them. When completing Part A, it will sometimes be necessary to get to grips with certain topics from Part B in order to reach an effective assessment. For example, the type of algorithm has an impact on sustainability and will therefore affect the assessment. Choices made during implementation can influence the decision as to whether or not to use the AI system. This may, for example, occur when an assumption regarding implementation in Part A turns out not to be right in Part B (e.g. because of the available data or use of a different IT infrastructure), as a result of which the assessment made in Part A is no longer correct. The table below shows the different potential applications of the AIIA.

Use the AIIA in all phases of your AI project

The AIIA can be used in every phase of development and procurement of an AI system. The AIIA is of most benefit if it is used from the start of an AI project. It is mandatory to have completed the AIIA when the system goes into production (even in the case of a pilot). The table below outlines the various possibilities. There is no pre-AIIA because the AIIA is compulsory for every AI system, irrespective of the risk based on the AI Act.

Quick-Scan AIIA	You use the AIIA Quick Scan to investigate whether an AI idea is feasible and desirable. The results will make it clear whether it is a good idea to procure or develop an AI system. When investigating the options, use the blue questions from Part A and use the blue questions from Part B for a more in-depth analysis if necessary This step can be used in order to make a go/no go decision.
Compiling a project plan	If you are compiling a project plan for the use of an AI system, it is mandatory to complete the AIIA. Use the AIIA in order to provide transparency with regard to the assessment of the AI system and account for the choices made. It is also an effective tool for discussion and to check that all relevant aspects have been taken into account. The guidelines entitled AI voor opdrachtgevers (AI for Commissioning Clients) provide practical tips for the development, implementation and management of an AI system.
During development	Choices made around implementation have an influence on the impact of an AI system (e.g. the data used or type of model). Use Part A to determine the impact and the choice to use an AI system. Use Part B to check whether relevant aspects have been taken into account in implementation.
AI system in production	The AIIA is mandatory at the moment when the AI system goes into production. Once in production, you should evaluate whether the AI project still meets the requirements. Check whether the area of application has been changed.

AIIA for impactful algorithms
The AIIA has been designed for use with AI systems, machine-based systems with a learning component. There are also simpler **algorithms**, which are basically a 'recipe' with predefined rules. You can also use the AIIA for **algorithms** of this kind, but it is not mandatory. The questions will help assess the impact of the algorithm and the implementation choices that are relevant in ensuring a responsible application.

Need help?
Do you feel that the AIIA is not the appropriate tool for an AI project or system you are considering? Perhaps you have other comments or questions about the AIIA? Is the completion of the AIIA raising questions? If so, contact the AI Team at I&W.

Version 2.0 accountability
The first version of the AIIA was developed by the ILT IDlab and analysis department, the RWS Datalab and the I&W Office of the CIO (CDIB). The AIIA was approved by the I&W Administrative Council on 4 July 2022. This version 2.0 (2024) has been adapted to add information about generative AI, and it also takes account of user experiences and information about the definitive version of the AI Act. The link with the

Fundamental Rights and Algorithm Impact Assessment (IAMA) has also been enhanced[1]. As a result, the AIIA now not only addresses the issue of how AI *can* be applied, but also whether it is *desirable* to deploy AI.

How to use this document

The following factors are important for completing the AIIA:

- The AIIA is mandatory at the moment when the AI system goes into production.
- The extent to which the AIIA is completed will depend on the expertise of the project leader and will also reflect the impact of the AI system. Unless otherwise indicated, a simple 'yes' or 'no' will not be sufficient as an answer to the questions.
- Blue questions are mandatory. They should always be completed.
- Green questions are intended to provide additional information. They should be completed if they are relevant.
- Words in bold are clickable terms that are defined in the Glossary of Terms included in the appendix.
- A fill-in template is available.

Please note that the Central Government Audit Service (*Auditdienst Rijk*) and the Netherlands Court of Audit (*Algemene Rekenkamer*) may check the AI system for correctness and security. A fully completed AIIA does not necessarily mean that the AI system is safe and secure. In order to ensure compliance with the AI Act (Regulation (EU) 2024/1689), Appendix 1 must be completed if your AI system is **high-risk**.

[1] This means that it is no longer necessary to complete both an AIIA and an IAMA, except in cases where additional assistance is needed in assessing fundamental rights. See Chapter 2.1.

Who does what?

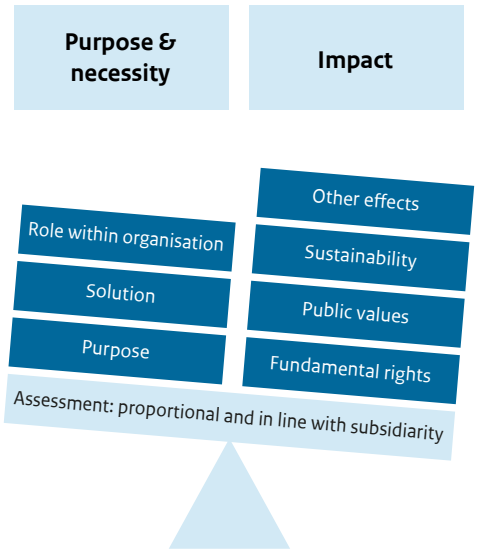
It is important for the AIIA to be completed by a multidisciplinary team because specific knowledge or expertise is necessary for the different sections. The table below includes a *suggestion* for the involvement of different roles for a specific chapter of the AIIA. Of course, it will depend on the scope and size of the project as to whether these roles are actually necessary or whether more should be involved than suggested in the table below.

	Ch 1. Purpose	Ch 2. Impact	Ch 3. Assessment	Ch 4. Technical robustness	Ch 5. Data governance	Ch 6. Risk management	Ch 7. Accountability
Stakeholders:		X					
CIO:	X	X	X	X	X	X	X
CISO:				X	X	X	X
Communication consultant:		X					X
Data scientists:	X	X	X	X	X	X	X
Data manager or source data owner:					X		
Domain expert:	X	X	X	X	X	X	
Privacy professional:	X	X	X	X	X		
Legal expert:		X	X		X		X
Commissioning client:	X	X	X	X	X	X	X
Other project team members:							
Project leader:	X	X	X	X	X	X	X
Strategic adviser on ethics:	X	X	X				

Part A: Assessment

Is the impact of the **AI system** proportionate to the intended objectives? This is a key question in the first part of the AIIA. Go through all of the questions and focus on the purposes, intended solution and expected effects. The proportionality of AI deployment is of particular importance. Perhaps there are other, less radical ways of achieving the objective?

If this is difficult to assess, for example because of a clash of different interests or because the AI system breaches fundamental rights, there are various tools available for conducting a structured conversation about ethical aspects. These include the IBDS data dialogue, DEDA, *Moreel beraad* (Moral Consultations) and many other methods. The I&W AI team offers internal workshops to support this conversation.



1 System purpose and necessity

These questions are about the purpose of the **AI system**, and its intended role within the organisation.

1.1 Purpose of the system

The 'application of AI' is not an aim in itself. An AI system is deployed in order to achieve a certain goal within the organisation, such as enabling the work to be done more efficiently or effectively. What problem needs to be solved? In this, look at the entire process in which AI plays a role. You can use the answers to these questions when completing the rest of the AIIA.

1. Provide a brief description of the intended **purpose** and intended **result** of the AI system (title, general description, definition of the problem, expected timeframe, location, target groups, the domain and operational process).
2. In which risk level in the AI Regulation does your AI system fall: **unacceptable, high or minimal risk**?

3. Where in the organisation (in which processes) is the AI system intended to be used?

In order to identify the risk level in the AI Act that applies to your system, we recommend doing the assessment in Appendix 1. This will give you greater clarity with regard to the legal obligations underlying your system.

If your system is in the '**unacceptable**' category, the AI system will not be permitted. In that case, it is not necessary to complete the rest of the AIIA.

If your AI system is in the '**high-risk**' category, you will also need to complete the remaining questions in Appendix 1. This will provide you with an overview of the additional obligations that high-risk systems need to meet.

For all other systems, categorised as low or minimal risk, the questions in the AIIA will be sufficient.

1.2 Intended solution

In this part, we look at the intended solution for the problem previously described, such as the AI technologies that will be applied and the data that will be used. Where necessary, also use the questions in Part B to help you to fully answer the questions below.

1 Provide a brief description of the intended **AI system** (technology, data and type of algorithm).

2 Why was this form of AI chosen (e.g. generative AI, linear regression or neural network)?

3 What alternatives were considered (e.g. no AI, less complex AI, different type of algorithm)?

1.3 Role within the organisation

Like any other IT system, an AI system has a commissioning client and a party with ultimate responsibility. Ownership is essential. In these questions, you determine the division of tasks in the development and use of the system. These roles are defined in the glossary of terms. Base your answers on these definitions.

1 Describe the division of tasks in setting up the AI system (such as the **developer**, **commissioning client**, **project leader**, IT **management organisations** and **person with ultimate responsibility**). If an external party is responsible for development: what contractual agreements are in place?

2 Who will be the **user** of the AI system, who are the **end users** who will work with the system and which **parties involved** will be impacted by the AI system?

3 Which stakeholders, people and/or groups have been consulted in the development of the **AI system**?

4 What feedback has been collected from teams or groups representing different backgrounds and experiences? And how was this feedback followed up?

1.4 Maintenance and administration

Like any other IT system, an AI system requires maintenance and administration. In the case of a proof of concept or pilot, it is also a good idea to investigate where and how the AI system will go into production. This enables the right choices to be made in advance and prevents, for example, an incompatible IT infrastructure or AI technologies that would result in an AI system that cannot be managed.

1 Describe the division of tasks for the administration and maintenance of the AI system (such as the **developer**, **commissioning client**, **project leader**, **management organisations** and **party with ultimate responsibility**). If an external party is responsible for developing the system: what contractual agreements are in place?

2 How are new laws and regulations that may be introduced or updated during the lifetime of the AI system taken into account?

3 Has the expertise required to manage the AI system been documented?

4 How are changes in the context of the AI system taken into account?

2 Impact

The questions in this section aim to identify the impact of the AI system's application in a specific context. Fundamental rights and sustainability are mentioned here explicitly, but there are also other effects which may have a positive or negative impact on broad prosperity. This refers to impact in the broadest sense of the word, such as the impact on specific target groups or on broad prosperity.

Please go through the chapter step by step because it has been arranged in a set sequence. For example, public values also relate to human rights and these are already covered in the 'Fundamental Rights' chapter. Part B also covers much of the impact.

2.1 Fundamental rights

Everyone with an interest in the operation of the **AI system** must be properly treated. This means that the (fundamental) rights of all **parties involved** must be safeguarded. When answering the questions, fundamental human rights are applicable and these are set down in the Dutch Constitution and the European Convention on Human Rights. The appendix to the Fundamental Rights and Algorithms Impact Assessment (IAMA) includes a list of fundamental rights clusters and this may potentially be useful in answering the questions below.

The AI system may result in a minor, medium or severe breach of fundamental rights. Extra careful consideration will need to be given to medium or severe breaches and additional measures may also be necessary.

Part B of the AIIA includes measures to protect a number of fundamental rights, such as personal data protection, the right to access information and a fair process. There may also be other fundamental rights that could be impacted by the use of the AI system.

Need help? Part 4 of the IAMA provides a step-by-step plan on fundamental rights that includes additional information on the subject[2].

1 What will the potential impact be on citizens' fundamental rights of using the **AI system**?

2 What legal basis underlies the use of the AI system and the intended decisions to be taken based on the AI system?

3 Which constitutional provisions may be applicable?

4 Which of these constitutional provisions may be breached in the event of improper implementation of the **AI system**?

[2] Questions 1, 2 and 3 of Chapter 4 of the IAMA correspond with the questions in Chapter 2.1, question 4 in the IAMA correspondence with Chapter 1.1 and questions 5, 6 and 7 can be used to make the assessment in Chapter 3.

2.2 Sustainability

Throughout its lifetime, an **AI system** creates an ecological footprint, through the consumption of energy, water and raw materials. For example, a continuous supply of water is required to run the **AI system**. An **AI system** also consumes energy. More specifically, it is important to ask whether the energy consumption is proportional in view of the problem definition and also how this energy was generated.

It is currently difficult to gauge the environmental impact of an AI system: companies are not, for example, transparent about energy consumption and research in this area is still in its early stages. For as long as there are no reliable figures available on impact, you should use this question mainly as a way of considering sustainable choices. It can sometimes be possible to reduce the environmental impact by using a different technology, infrastructure or model, and this often has the added advantage of making the model faster.

In general terms, it is possible to state that a 'simpler' model will be less energy intensive than a more extensive model such as a **large language model**. In addition, the size and type of input and output also have a major influence on energy consumption (e.g. text requires less computing power than images).

On the other hand, it is also possible to use an **AI system** to achieve environmental gains. This impact should be offset against the environmental costs of running the system, for example.

1 What will be the environmental impact of introducing the **AI system** (development, installation and use), and how will this be measured?

2 What measures have been taken to minimise the (negative) environmental impact of the AI system?

2.3 Other effects

This section looks at the effects or consequences that have not been mentioned earlier. Part B of the AIIA includes numerous effects, so it is important to look at these in order to be sure that everything has been included. In this chapter, it is also possible to mention issues that are of particular importance for this project in order to ensure they are given sufficient attention.

Below is a list of potential effects. The extent to which these points are relevant will depend on the scope of the project.

Other effects to consider:
- Public values: changing, context-related ideas about what we as a society perceive to be valuable. Further explanation and insight about public values can be found in the Ministry of the Interior and Kingdom Relations (BZK) toolbox for ethically responsible innovation3
- The organisation's mission and vision: for I&W, for example, this is working to achieve a better, clean, safe, sustainable and accessible Netherlands
- Short and long term
- Effects on the individual, organisation and society
- Positive and negative aspects of using the AI system

In this, you should not only consider the risks, but also the opportunities and positive effects of applying the AI system.

1 How does the AI system contribute to the organisation's mission?

2 In addition to the questions above, are there any other relevant effects (positive, negative, risks, for specific target groups, at different levels, broad prosperity) of the AI system that need to be taken into consideration?

[3] https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/nieuwe-technologieen-data-en-ethiek/publieke-waarden/toolbox-voor-ethisch-verantwoorde-innovatie/publieke-waarden-centraal/

3 Assessing whether or not to use the AI system

In this step, you determine whether it makes sense to deploy the AI system or not, by looking at the purpose, the solution and the impact of the AI system. Is it proportionate to apply the AI system? Does the intended application actually solve the problem? Or are other technologies also suitable? Clearly describe how this assessment has been made.

Provide clear reasons for your choices. Attempt to take everything carefully into account. Consider different groups of parties involved: no groups of people must be allowed to be unfairly disadvantaged. Of course, in some cases the impact will be so significant that further assessment is not necessary, for example if there is a breach of fundamental rights or it is not permitted according to the law and regulations.

Part B includes measures to ensure the responsible application of the AI system. It is possible that additional or specific measures may be necessary, for example within the process. You should mention these here.

1. Is the impact in proportion to the intended goals and are there other less radical ways of achieving these goals? In other words: is it **proportional** and in line with **subsidiarity** to deploy the system to achieve the stated goals?

2. Are there additional measures (e.g. as part of processes) that you could take to use the system responsibly?

Part B: Implementation and use of AI system

Part B of the AIIA looks at the design, implementation and use of the AI system. It includes more detail and in many cases requires greater knowledge about AI technology. Make sure that you involve this expertise, for example a data analyst, when completing and using this chapter.

4 Technical robustness

Whether or not an **AI system** works in the way for which it was intended is down to its technical **robustness**.

4.1 Bias

Bias means making assumptions about things, people or groups. There are two sides to it. On the one hand, it is necessary to draw conclusions in advance about the use of data in a new situation. When generalising, we always make assumptions. On the other hand, it is important to avoid unjustified distortion caused by forms of injustice and undesirable bias that may be at odds with human rights or regulations.

Bias can occur anywhere in the system. For example, it includes: **bias in the input**, **bias in the model** and **bias in the output**. When designing and using AI, it is important to take account of **data bias** and **design bias**. These types of bias are often caused by socio-economic assumptions in the data or made by a developer. The model can exacerbate these assumptions. This can mean that an AI system does not work effectively for all **parties involved** if you are not aware of the bias and correct for it.

At its core, this issue is all about awareness and integrity. It is impossible to work completely without bias. Bias can often last for decades and continue to go unrecognised, even when it should be. Take, for example, the language models (ChatGPT, Google Bing, Microsoft Co-pilot): these were probably trained using information on the internet and all of the assumptions this information has included for years. Therefore, rather than focusing on bias-free AI, we should aim to be as aware as possible of any potential discrimination.

Bias is closely related to human **diversity**, **equality** and **fairness**, but it is important to be aware that assumptions may also relate to non-human aspects, such as nature or the living environment. When considering how you aim to mitigate bias, it may also be relevant to draw a distinction between any **negative impact**, **no positive impact** and **positive impact** that the bias could have.

1 How will potentially undesirable **bias,** such as **bias in the input**, **bias in the model** and **bias in the output** of the **AI system** be taken into account?

Bias in the input (data)

2 Is the input (data) relevant and representative, taking account of the intended purpose (question 1 of 1.1) of the AI system?

3 In random sampling, have any subpopulations been protected if necessary?

4 Has the choice of input variables been substantiated and coordinated with the parties involved?

Bias in the model

5 What measures have been taken to prevent unfair or unjustified **bias** being created or exacerbated in an AI system?

6 Can the AI system be used by the intended **end users** (in other words irrespective of their characteristics, such as age, gender or capacity)?

Bias in the output (data)

7 Are there stop mechanisms, supervision mechanisms or monitoring mechanisms in place to prevent groups in society from being disproportionately affected by the negative implications of the AI system? Specifically for ILT: a distinction needs to be made here between *ondertoezichtstaanden* (supervised parties (OTS)) and the rest of society.

4.2 Accuracy

An accurate **AI system** performs effectively and is effective in its assessments. It is important to continuously measure the performance of an AI system (during both the development and production phases). The quality of the training data used is also important. Any AI system is a work in progress: it remains necessary to regularly test and retrain AI systems. It is desirable to have some way of quantifying the likelihood of the system making an incorrect assessment.

You can determine the **accuracy** of the system in advance by drawing up **acceptance criteria** for both the (training) data and the system. Examples of acceptance criteria might be a minimum amount of data or specific threshold values for the measurement system. There are numerous different types of measurement systems (often referred to as 'performance metrics' by data scientists) available for quantifying the quality of **models** such as an accuracy score, a precision score, and a recall or F1 score.

It is important to ensure that the measurement system and the acceptance criteria are properly geared to the data and intended purpose of the AI system[4]. Among other things, this must be coordinated with the findings from the risk analysis (see 'Risk management'), because new risks may emerge over the course of time when using an AI system. It is also important to ensure that the quality of the system is continuously monitored. During retraining or further development, you should re-evaluate the acceptance criteria and choice of measurement systems.

[4] The measurement system chosen must be suitable for the model and the data used to measure the quality. Take, for example, a system that has to label five words in every hundred words in a specific text. If the system labels 0 words in this text, the model has an accuracy of 95%. If you are determining the quality of the system based on the accuracy, the model therefore appears to be performing very well when in fact the recall is 0 and it is actually not performing well at all. For this reason, accuracy is not suitable for determining the performance of this model.

1 How will the continuous **accuracy** of the system be measured and safeguarded?

2 What **acceptance criteria** have been set up to measure the quality of the **input (data)** and **output (data)** of the **model**?

3 Are the **acceptance criteria** appropriate for the data and the purpose of the AI system?

4 How will the **output (data)** be regularly checked at random and continually monitored for correctness?

5 How will deviations in the output (data) relative to the acceptance criteria be analysed and corrected in a timely fashion?

6 What would the results be if alternative **models** were used?

4.3 Reliability

A **reliable** AI system produces consistent results in similar cases. When it comes to reliability, the key question is whether the individual **output (data)** can be reproduced using the same **model** and the same **input (data)**, the same settings and the same **parameters**. It is also important for the system to provide a reliable indication of how well the model will perform in new situations.

1. Is the **AI system reliable**?

2. What are the most important factors that influence the performance of the **AI system**?

3. Is a part of the (sub)dataset excluded from the model's learning process and only used to determine reliability or is the model's reliability calculated by means of cross-validation?

4. How has the (hyper)parameter tuning been substantiated and assessed?

4.4 Technical implementation

Technical implementation describes how the AI system is technically integrated within the organisation's ICT landscape. The specific hardware and software requirements of the AI system are documented to enable these to be taken into account in the roll-out and management of the system.

In addition, it is clear from the system architecture how the different software components interrelate. A carefully considered architecture reduces the operational risks involved in building a technological solution and creates a bridge between operational and technical requirements. In many cases, there will already be existing documentation and you should make sure you refer to this. You should also take a look at the Infrastructure and Water Management Enterprise Architecture (IWEA).

1. How has the AI system been implemented technically?

2. Has there been consideration of how the AI system fits into the existing technical and system infrastructure and have appropriate measures been taken for its roll-out (if applicable)?
3. Describe the system architecture (how do the software components interrelate)?
4. Have any specific hardware and software requirements been documented?
5. If the application is hosted externally, under what conditions is this happening?
6. How is access to the AI system and its components configured (think of the generic IT management measures)?
7. How can the AI system interact with other hardware or software (if applicable)?
8. How is the logging and monitoring configured?

4.5 Reproducibility

Reproducibility is about training, validation and testing. Reproducibility concerns such issues as registering the data used, developing the model, recording changes in the data, whether the same **input** (**data**) produces consistent results and whether there are certain situations or conditions that may affect the **output** (**data**).

Reproducibility is closely related to **traceability**. The main purpose of traceability is to ensure that the datasets and processes are properly documented. Data version management, the model and training the model are important aspects of this.

1. Is the **AI system reproducible** ? Has a process been set up to measure this?

2. Can **output** (**data**) obtained be reconstructed now or in the future (i.e. have previous versions of the **model**, datasets and conditions been saved by means of version management)?
3. Is it possible to reconstruct the model based on the given **parameters** and a fixed **seed**?
4. Can the broad outlines of the **AI system** be reproduced using the documentation?
5. How will the changes be documented during the system's lifetime?

4.6 Explainability

Technical **explainability** relates to the ability to understand both technical processes and human decisions related to them. It also needs to be clear which different design choices have been made and what the rationale is for using the **AI system**. See also: 'Accountability' for details about explainability, transparency and communication vis-à-vis users and other **parties involved**.

1. Is the **AI system** sufficiently **explainable** and interpretable for the **developers**?

1. In developing the AI system, how has account been taken of the model's explainability, for example for the users?

2. What technologies have been used to ensure that the AI system is explainable and why was this technology chosen?

5 Data governance

Data **governance** refers to the procedures in place regarding data with respect to access, ownership, usability, integrity and security. It also encompasses the quality of the data used.

Data governance includes privacy as well. Privacy is one of the fundamental human rights that could be compromised by AI. It is therefore important to ensure that there is effective data governance and protection of personal data, at least in accordance with the General Data Protection Regulation (**GDPR**) and the I&W privacy policy.

5.1 Data quality and integrity

Data quality is essential for the effective operation of an **AI system**. There are good reasons why the term 'garbage in = garbage out' exists. In addition, data collected may for example contain socially constructed **bias**, inaccuracies, errors and mistakes (see also 'Bias'). This needs to be addressed before these data are put to any further use.

To do this, you should use the FAIR principles (Findable, Accessible, Interoperable and Reusable) and the closely related FACT (Fair, Accurate, Confidential and Transparent) principles as inspiration. These principles are used within the GDPR. The datasets and working procedure will need to be tested and documented at every step: training, testing, roll-out phase and operational phase. This also applies to AI systems that have not been built in-house, but have been acquired.

Data quality is especially important if personal data will be used. According to the GDPR, it is essential for personal data to be correct and updated if necessary.[5] The data also need to be necessary for the purpose of the analysis. You should therefore also describe the purpose of the analysis and how you are dealing with the data and any errors in the datasets and outcomes.[6]

The Ministry of Justice & Security's 'Guidelines for the application of algorithms' ('*Richtlijnen voor het toepassen van algoritmen*')[7] make data quality explicit at a very operational level.

1. Which training data will be used as input for the algorithm and from which sources do the data originate?

2. How will the data quality be safeguarded?

[5] Article 5.1.(d), GDPR. See also Section 3 Rectification and erasure of data, GDPR.
[6] Richtlijnen voor het toepassen van algoritmen door overheden en publieksvoorlichting over data-analyses, JNV, p.20.
[7] https://open.overheid.nl/documenten/ronl-1411e45f-b822-49fa-9895-2d76e663787b/pdf

General

3. Is the data used necessary for the **AI system**?

4. How are you preventing unintended data duplications?

5. Is it possible to update the training and test data when the situation requires it? When will you decide to retrain, temporarily stop or further develop the AI system?[8]

6. Does the data meet the assumptions underlying the **model**?

7. How has the **input (data)** used in the AI system been collected and collated?

8. How will the data be labelled?

9. What factors (think of limitations in the method of collection, storage, etc.) affect the quality of the input (data)? And what can you do about that?

10. Has the input (data) been assessed for changes that occur during training, testing and evaluation? Also during the use of the algorithm over the course of time?

Output (data)

11. If the output (data) is used as new input, how will the output (data) be stored and checked for correctness and completeness?

12. How will you ensure that the output (data) is available in a timely fashion?

[8] Article 14.4.

5.2 Privacy and confidentiality

When designing an AI system, privacy legislation must be taken into account. Ultimately, it is easier to ensure this is done properly in advance rather than repairing it at a later stage. Of course, privacy must be safeguarded throughout the entire life cycle of the **AI system**.

When processing personal data, a **Data Protection Impact Assessment** (DPIA) pre-scan must be completed to determine whether it is necessary to complete the full DPIA.

In addition to personal data, other confidential information may be used that must not be made public. For example, this applies to the use of confidential information, such as classified information or trade secrets. These data must also be protected. The **AI Act** supplements the **GDPR**, by including additional rules for the use of (personal) data in AI systems. Confidential data must be sufficiently protected and secured (see Risk management).

1. What approaches are being adopted for personal data or confidential data?

Regarding personal data

2. Does the **AI system** work with personal data (is the GDPR applicable)? If so, please also complete the following questions. If not, proceed to 'Regarding confidential data'.
3. Is the processing of personal data proportional and in line with subsidiarity (use the assessment in Chapter 3 as a basis for this)?
4. Can the output of the AI system be tracked back directly or indirectly to individuals (is the GDPR applicable)?
5. Have officials been involved, such as the Chief Privacy Officer, Information Security Officer, Chief Information Officer, Privacy Officer, etc.?
6. How often is the quality of and necessity for processing personal data evaluated?

Regarding confidential data (i.e. not personal data)

7. Will confidential data be used or stored?
8. How will the security of this information be safeguarded?

6 Risk management

It is important to keep an eye out for risks. Unforeseen risks can lead an **AI system** to produce unreliable results. This can cause damage, as a result of poor performance of the AI system or because of **hacking attacks**.

6.1 Risk prevention

The development and **putting into service** of an **AI system** entails dangers that this AIIA aims to address as far as possible. However, unforeseen problems can still occur. It is important to determine how you will deal with these potential dangers. Mechanisms must be in place to manage risks and these mechanisms must have been tested. This involves such areas as preventing data poisoning, the scope of any management measures and the security of storage places for results. Risks can also occur after the AI system has been introduced. Check the risk management measures on a regular basis, at least once every three years or when major changes are made.

For high-risk AI systems, a risk management system is mandatory. You will find additional questions on this in Appendix 1. For other systems, a one-off risk analysis will be sufficient.

1. How has the system been tested for appropriate and targeted risk management measures?

6.2 Alternative procedure

It is advisable to have a plan in place in the event that problems arise with the **AI system**. This means that an alternative procedure must be available. Examples might include the option to revert from a machine-learning model back to a more limited rule-based **model**. You could even go a step further back and carry out the process manually. Of course, if it is acceptable for the system to be temporarily unavailable, having no plan at all is also an option.

The use of an AI system can cause certain human skills to deteriorate. The effect of the calculator on our mental arithmetic skills is an example of this. For this reason, any alternative procedure must not suddenly rely on this skill.

1. What will the plan be in the event of problems with the operation of the **AI system**?

2. What would be the impact of the system failing?

3. See the calculator example above. What equivalent effect could occur if the **AI system** is put into service and is this desirable?

6.3 Information security risks

As far as possible, an AI system must be built to be secure **by design**, with security being considered in the design phase. This is how information security risks, such as manipulation of the model, unauthorised access or **hacking attacks** can be managed. Draw up a list of foreseeable risks using the organisation's risk management process. This includes such areas as: mapping out the CIA triad; information classification levels; implementation of Government Information Security Baseline (BIO) measures; security tests; and, if the BIO security level does not suffice, possibly conducting an additional (technical) risk analysis.

Properly configured authorisations and a robust change process are essential in this. In addition, it is important to check whether errors and irregularities can be detected and technically accommodated. More practical guidance can be found in the Court of Audit's Algorithm Assessment Framework (*Toetsingskader Algoritmes van de Rekenkamer*)[9] or the ADR algorithm research framework (*Onderzoekskader algoritmes ADR*)[10]. The AIVD has a guide to the secure and safe development of AI systems.[11] In addition, the Open Worldwide Application Security Project (OWASP) has numerous resources about the safe and secure application of AI.

1. How are information security risks identified, reduced to an acceptable level and tested (from a technical perspective)?

2. How are unauthorised third parties prevented from taking advantage of vulnerabilities in the AI system?
3. What would the impact be of third parties having unauthorised access to the source code, data or results of the AI system?
4. Is it possible for people to take advantage of the fact that an AI system is being used instead of a human decision?
5. What is the system for recording who is using the AI system and for how long?
6. In addition to the standard I&W security measures, have additional measures been taken to secure the AI system?

[9] https://www.rekenkamer.nl/onderwerpen/algoritmes/toetsingskader
[10] https://www.rijksoverheid.nl/documenten/rapporten/2023/07/11/onderzoekskader-algoritmes-adr-2023
[11] https://www.aivd.nl/documenten/publicaties/2023/02/15/ai-systemen-ontwikkel-ze-veilig

7 Accountability

The Dutch central government must account for its actions within the organisation, to the House of Representatives and wider society. Although the technology is being applied increasingly often, there are also concerns about the use of AI. In order to provide accountability with regard to the use and results of AI systems, it is essential to set up a process for this purpose.

7.1 Transparency towards users

End users must be given an insight into the operation of an AI system (in addition to the explainability required for developers, see 'Technical Robustness'). This applies particularly to employees who use AI as part of their working process. They do not necessarily have to fully understand the AI system. The operation of an AI system and its limitations must however be clear in broad outline.

If a decision needs to be taken about the use of AI, as well as understanding its operation and limitations, it is essential to have a significant degree of influence over the decision.

1. In what way do you provide your end users with an insight into the operation and limitations of the AI system? And are these given sufficient attention for as long as they persist?
2. What role do people play in decisions based on the AI system's input ('humans in the loop') and how do you enable them to play this role?
3. How can the system be monitored and understood by everyone (human oversight)?

7.2 Communication to parties involved

This section is about two types of communication with the **end users**. Firstly, end users must be aware that they are dealing with the results of an **AI system** (and not a human, for example). Secondly, end users are entitled at all times to know how an **algorithm** determines the outcomes of an AI system. This also means that the purpose and limitations of the system must be communicated clearly, fairly and transparently. Both technological processes and the related human decisions need to be understandable, retrievable and corrected if necessary. This can be achieved, for example, by appointing a contact person with substantive knowledge about the AI system. Because of the self-learning nature of AI, it is not always possible to fully trace back the operation of the system. However, it must always be possible to provide an appropriate explanation of the process to end users.

In addition, it must be possible for citizens to retrieve information about the AI system or to invoke their rights pursuant to the GDPR. It must be possible for people to dispute the results of the AI system. This also means that the data, and the conditions in which the data is made available, must be stored (see Archiving).

1. To what extent are you transparent vis-à-vis different groups of parties involved about the AI systems and in what way?
2. Are mechanisms being set up to enable end users to make comments about the system (data, technology, target group, etc.)? And how and when are these validated (analysed and followed up on)?
3. Pursuant to the AI Act, does the system need to be included in the algorithm register and/or (for high-risk applications) in the EU database?

4 Are the **end user** of and **parties involved** in the AI system informed that the results are generated by an AI system and what this entails for them?

5 Has a manual been compiled?

6 What are the potential (psychological) side-effects, such as the risk of confusion, preference or cognitive fatigue in the **end user** of using the AI system?

7 In what way are different groups of **parties involved** (citizens, colleagues, managers, etc.) given an insight into the different aspects of the **AI system**? This includes such areas as data use, model or results.

8 How have you taken measures to achieve explainability specifically towards the **end user**?

9 Is the system sufficiently **transparent** to enable **deployers** to interpret the system's output (data) and use it appropriately?20

10 Have steps been taken to provide end users with training if necessary?

11 How are you ensuring that comments made by parties involved and end users are properly handled internally?

12 If a party involved wishes to lodge an objection,21 or submit a complaint about the AI system,22 is it clear what steps they should take? The same applies to lodging an appeal.

7.3 Verifiability

Verifiability refers to the way in which the data, the **model** and the results are evaluated. This process can be done by means of internal or external audits. Stricter requirements apply when using an AI system in riskier areas.

It is essential that there is insight into the sources, the system and the result. This is the **responsibility** of the **owner of the system**.

1 How will the **AI system** be verified and by whom?

2 In what way is accountability provided about the AI system?

3 Who provides the independent audit of the AI system? And in what way?

7.4 Archiving

Archiving is the storage of information to enable it to be reused in the future. Such reuse might include reconstructing the model (see 'Reproducibility'), or explaining how the system works to a new member of staff (see 'Explainability'), or to provide accountability to a **party involved** (see 'Accountability'). Archiving is also important in ensuring compliance with legislation and regulations. For example, a minimum retention period applies for high-risk applications of logs in an AI system and these are also described in the I&W **selection lists**.

Input (data)

1. How is the **input (data)** stored?
2. What is the retention period for the input (data)?

Model

3. How is the **model** stored?
4. How is version management arranged?

Output (data)

5. What is the retention period for the output (data)?

Glossary of Terms

Various terms are used in this document that may be defined differently in the literature. The following list clearly defines the terms used in this document.

Acceptance criteria	Conditions based on the intended purpose and agreed data that the **AI system** must meet. These conditions may concern the quantity of data, accuracy metrics for the **output** (**data**) or an independent output verification mechanism. Where possible, acceptance criteria must be made quantifiable, enabling these to be monitored by means of an appropriate measurement system. Good acceptance criteria are SMART and sufficiently differentiated in order to enable all aspects of the AI system to be effectively monitored.
Accuracy	Very precise or meticulous: refers to a system that is capable of making correct and
Accurate assessments	Expressed as a formula: TP+TN/(TP+TN+FP+FN). TP= True Positive, TN=True Negative, FP=False Positive, FN= False Negative. The higher the number of true results there are relative to false results, the higher the accuracy will be.
AI Act	European legislation that establishes rules for the development and use of AI systems.
AI system	A machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.
Algorithm	A set of rules and instructions that a computer automatically follows in making calculations in order to solve a problem or answer a question.
Area of application	Term used in AI legislation to refer to the context in which an AI system will be used. An example could be infrastructure.
Artificial intelligence	There is no set definition of AI. We apply the description of AI used by the Netherlands Court of Audit: "the ability [...] to correctly interpret external data, to learn from such data, and to use these learnings to achieve specific goals and tasks through flexible adaptation". Although it is not used in this document, we would also like to draw attention to the European Commission definition, which is as follows: AI refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.
Bias	Prejudice. Making assumptions about things, people or groups, in many cases not based on actual measurements.
Bias in the input	Data quality, consistency and integrity is an important precondition for an unbiased analysis.

Bias in the output	The way in which the **output** (**data**) is used can have an impact on people's lives. It is important to ensure that unjustified correlation does not result in causality.
Bias in the model	How correct are the **models**; to what extent do they correct for known flaws in the representativeness of the data? This may, for example, relate to what the **AI system** learns and what are considered to be undesirable learning effects.
BIO	Government Information Security Baseline (*Baseline Informatiebeveiliging Overheid*)[12], the framework of basic standards for information security within layers of government.
By design	Taking account of the relevant AI, privacy and security legislation in the design process. Examples include AI, privacy and security by design.
CIO	Chief Information Officer.
CISO	Chief Information Security Officer.
Commissioning client	A person or organisational division that commissions a contractor. This person is also ultimately responsible for completing an AIIA (together with the project leader).
Corruption	The misuse or exploitation of errors in the system or the exploitation of neutral system characteristics. 25 As distinct from **unintended corruption**.
Data bias	This refers to random samples that are not representative of the whole population.
Data pipeline	How the data moves from the field to the model; the process that the data goes through.
data subject	Natural person or organisation that has (or believes to have) an interest in the use or results of the system. A deliberate decision has been made not to use the word 'interested party', since this term is wider than the definition of 'interested party' used in Dutch administrative law. Examples might include citizens, supervised persons, but also the end users themselves.
Design bias	Problems in the technical design, including limitations in the computer tools, such as hardware and software.
Developer	An organisation or person that designs, develops and/or trains an **AI system**.
Diversity	This refers to the recognition of different types of 'subjects' in our analyses. In this, we try to prevent groups of relevant subjects being unjustly excluded from the development of an **AI system**, as a result of which the system does not cater to their needs.

[12] https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/cybersecurity/bio-en-ensia/baseline-informatiebeveiliging-overheid/

Domain expert	Someone who is very knowledgeable about the problem area in which the **AI system** is being built.
DPIA	Data Protection Impact Assessment. This tool is used to assess privacy risks involved in data processing.
End user	End users are the people who apply the **AI system** in practice within the organisation. This refers to natural persons. Who has their hands on the controls? Who within the organisation collects information from the AI system? Examples might be an inspector or a road traffic controller.
Entity	A position within an organisational department.
Equality	This refers to the notion that the same type of subject receives equal treatment.
Explainable	A statement of how input variables contribute to an output of the algorithm that must be explained.
Fairness	If not every subject is given equal treatment, it must be possible to explain this. In this, it is important that there is as comprehensive a picture as possible of the distinctive subject characteristics. This is both in order to demonstrate which characteristics actually play a role (and attribute a lower risk to party A than to party B) and which characteristics do not (as a result of which party A and B have a substantiated equal risk).
GDPR	General Data Protection Regulation. This privacy legislation regulates the care and protection of personal data by businesses and organisations.
Generative AI	A specific type of AI, in which algorithms are used to generate content. By means of a simple prompt, users can generate text, image, sound or computer code in an instant. The best-known example of this is ChatGPT.
Governance	The action or manner of governing, the code of conduct and supervision organisations. This relates to decisions that determine expectations, bestow power or verify performances. This consists either of a separate process or a specific part of management or leadership processes.
Hacking attack	Breaking into the **AI system**. With such consequences as pollution of data, unwanted leaks about an AI system (or its operation) or corruption of software or hardware.
High-risk AI	The AI Act lays down what high-risk AI is. These are often products closely related to fundamental rights and/or product safety. In all cases: an AI system that profiles individuals by processing personal data. In addition, all AI systems in the fields of: non-prohibited biometric data; critical infrastructure; education and vocational training; employment; access to and enjoyment of essential public and private services; law enforcement; migration, asylum and border control management; administration of justice and democratic processes.

Input (data)	The data processed for a predetermined purpose. In the context of an **AI system**, this may refer to raw data, such as observations from reality. In the context of the **model**, this normally refers to pre-processed data.
Interest group	Group of **stakeholders** to measure **diversity**. This may be a group of **end users** or a group of people impacted by the system.
Large Language Model	An LLM is generative AI that can generate text. It is trained on very large datasets and contains numerous parameters.
Limited-risk AI	The AI Act lays down what limited-risk AI is. AI that is geared to interact with humans, recognise emotions or produce manipulated images. Examples include spam filters, summarising texts, classifying aviation incident topics or AI systems that regulate office lighting.
Low-risk AI system	AI systems with a risk of manipulation or deception. These AI systems must be transparent and users must be informed about their interaction with the AI.
Management organisation	An organisation that sets up and optimises the **AI system's** application management.
Metadata	Data that describe the characteristics of other data. For example, whose data they are, or who sent them or when they were most recently changed.
Minimum-risk AI	Any AI system that is not prohibited or categorised as **high-risk AI** or **limited-risk AI**.
Model	A (simplified) mathematical representation of reality, which is used to process information. In an **AI system** the mathematical representation is often partially or completely learned according to an **algorithm**, as a result of which it is not possible to fully explain how the model reached its results, even for the **developers**.
Moral Consultations	Physical Environment Consultative Council (May 2021), *Moreel Beraad*.
Negative impact	Parties involved who experienced negative consequences as a result of the application of the **AI system**, for example because they are discriminated against based on a **bias** in the AI system.
No positive impact	**Parties involved** who do not by definition experience any negative impact from the use of the AI system, but remain, for example, in the same position as before. In this, there may be a danger that these parties involved do not experience the same 'positive impact' from the AI system as that experienced by other parties involved.
Output (data)	The data produced by the **AI system**. These are the results of the model.
Parameter	A variable within the **model**. If this variable is changed, this will also change the resulting size of the model or calculation.

Party with ultimate responsibility	A role within the organisation that bears **responsibility** for the **AI system**. For example, this refers to the responsibility to ensure that the right results are achieved for the AI system. This is usually the process owner.
Positive impact	**Parties involved** who experience favourable consequences as a result of the use of the **AI system**. This could be a minority group that is advantaged. This may entail the danger that this positive bias is excessively optimistic and therefore does not reflect reality. The drawback of this could also be a 'negative impact' for other parties involved.
Project leader	The person with ultimate responsibility for the project of which the **AI system** is a part. Together with the **commissioning client**, this person is also **ultimately responsible** for doing an AIIA.
Proportional	Proportionality is about ensuring a reasonable relationship between the purpose and the solution applied. Is the use of AI proportionate to the problem that is intended to be solved by the **algorithm**? The expected advantage must outweigh the risk that the AI entails.
Putting into service	The moment when an AI system is 'put into service' means the moment when it is used for the first time in a process. This is preceded by an external test or pilot. The AIIA must have been completed when the system is put into service.
Reliable	Having the characteristic of consistent behaviour and consistent results.
Reproducible	The ability to repeatedly achieve a similar result whenever a described procedure is executed.
Responsible	The actions of an **entity** can be traced back to that entity in a unique way and this entity is liable for these actions. Who is who? Please enter which persons have played a role in answering this AIIA.
Risk-free AI system	All AI that does not fall in the other categories. These systems are not covered by the AI Act.
Robustness	Developed using a preventive approach; behaving as predicted and described in advance. Avoiding unacceptable damage.
Seed	A 'seed' is the starting point of a random number generator. From this starting point, this generator always follows the same route to create new (pseudo-) random numbers. By documenting the seed, it is possible to repeat the route of (pseudo-) random numbers. This means that the seed is necessary to verify reconstruction of a **model** whenever the model makes any use of random numbers. The seed itself is also a number. There are no specific requirements for this number, and something 'recognisable' is often chosen (e.g. '123456', or '0, 42, 1234' or the date of birth of a **developer**).
Selection list	A list that describes how long the archive records must be retained, for example pursuant to the Netherlands Public Records Act (*Archiefwet*).

Stakeholder	Person or organisation that can influence a decision or activity, be influenced by it, or consider themselves to be influenced. A stakeholder can also be the owner of the data used, for example.
Traceability	When processes and results can be verified.
Transparent	When the operation and purposes of the **AI system** are clearly communicated and AI system results are **explainable**.
Type of algorithms	It is possible to use different technologies to create AI, such as neural networks, random forests or other forms of machine learning. However, less complex **algorithms**, such as business rules or decision trees, can also be used.
Unacceptable-risk AI system	AI systems that: apply subliminal, manipulative or deceptive techniques; make inappropriate use of vulnerabilities, categorise biometrically; apply social scoring; predict individuals' potential for criminality; create databases for facial recognition; infer emotions; recognise emotions remotely in real time.
Unintended corruption	Having an influence on the operation of the **AI system** without any malicious intent, for example by feeding in faulty input or pressing the wrong buttons. Unintended corruption falls under **reliability**. We draw a distinction between this and (intended) corruption.
User	According to the AI Act: "a [...] public authority, agency, or other body using an **AI system** under its own authority [...]". The user puts the system into use. This is never a natural person. Examples might include the ILT or RWS.

Appendix 1: Risk level assessment

The AI Act describes a number of risk areas of application for AI systems. The greater the risk, the more measures need to be taken. The risk level is determined by the area of application in which the AI system is deployed. The questions below help to determine the risk level of your AI system.

The classification of risk levels in the AI Act is absolute. This means that any other risks, such as those based on privacy legislation, do not count. For example, an AI system may have a significant impact in the area of privacy, but it may be in an area of application with minimum risk according to the AI Act. This means that it is not a high-risk AI system. The standard questions in the AIIA may however be of use in mitigating risks of this kind.

Definition of high-risk AI system (AI Act)

An AI system that profiles individuals by processing personal data. In addition, all AI systems in the fields of: non-prohibited (by the AI Act) biometric data; critical infrastructure; education and vocational training; employment; access to and enjoyment of essential public and private services; law enforcement; migration, asylum and border control management; administration of justice and democratic processes.

Still uncertain about which risk level applies to the AI system? If so, answer the following questions.

A) Assessment of unacceptable-risk

1. Does the AI system use subliminal messaging? Or are deliberately manipulative or misleading techniques used?
2. Does the AI system make use of manipulation/abuse of a vulnerable group, which could result in physical or psychological damage?
3. Does the AI system make use of ***social scoring***?[13]
4. Does the AI system aim to infer a person's emotions at work, outside of medical security considerations?
5. Does the AI system use biometric identification in a public space? Or biometric categorisation of people based on special personal data?

Have you answered no to all questions? If so, the AI system is probably not an unacceptable-risk application.

B) Assessment of high-risk

1. Does the AI system profile individuals?
2. Is the AI system a product or security component of a product within one of the following fields:

- **Machinery** (Directive 2006/42/EC)
- **Toys** (Directive 2009/48/EC)
- **Recreational crafts and watercraft** (Directive 2013/53/EU)
- **Lifts** (Directive 2014/33/EU)
- **Equipment and protective systems intended for use in a potentially explosive atmosphere**(Directive 2014/34/EU)
- **Radio equipment** (Directive 2014/53/EU)
- **Pressure equipment** (Directive 2014/68/EU)
- **Cableway installations** (Regulation (EU) 2016/424)
- **Personal protective equipment** (Regulation (EU) 2016/425)
- **Appliances burning gaseous fuels** (Regulation (EU) 2016/425)

[13] Social scoring: AI systems for the evaluation or classification of natural persons or groups of persons during a specific period based on their social behaviour, or known, derived or predicted personal personality characteristics (p. 51 AI Act)

- **Medical devices**(Regulation (EU) 2017/745)
- **In-vitro diagnostic medical devices** (Regulation (EU) 2017/746)

The Regulation also refers to a list of products that are also seen as high-risk AI applications. With the exception of Articles 6.1, 102 to 109 and 122, the high-risk obligations pursuant to the AI Act do not yet apply for these products. However, the requirements in the AI Act will be used at a later stage in shaping the specific product legislation that will apply to these products. It is not yet known when this will happen and it will vary according to the product. This applies to the following products and legislation:

- **(Protection of) civil aviation** (Regulation (EC) 300/2008 and Regulation (EU) 2018/1139)
- **Two- or three-wheel vehicles and quadricycles** (Regulation (EU) 168/2013)
- **Agricultural and forestry vehicles** (Regulation (EU) 167/2013)
- **Marine equipment** (Directive 2014/90/EU)
- **Interoperability of the railway system in the EU** (Directive (EU) 2016/797)
- **Motor vehicles and their trailers** (Directive (EU) 2018/858 and Directive (EU) 2019/2144)

3. Does the AI system use remote biometric identification of people?
4. Will the AI system be used as a security component in the management/operation of: critical digital infrastructure[14], road transport, the supply of water, gas, heating or electricity?
5. Does the AI system have an influence on people's recruitment and access to work?
6. Will the AI system be used in relation to the access and use of essential private and public services and payments?
7. Will the AI system be active in law enforcement?
8. Will the AI system be used in areas relating to migration, asylum and border control management?

Have you answered no to all questions? If so, the AI system is probably not a high-risk application.

C) Assessment of transparency obligation

1. Does it involve an AI system that interacts with people?
2. Can the AI system generate or manipulate artificial content (generative AI/GPAI)?
3. Can the AI system recognise emotion or categorise biometrically?
4. Does the system use deep-fake technologies?[15]

If you have answered 'yes' to any of the above questions, you should take the necessary transparency measures.[16] In general terms, this means: ensure that users know that they are dealing with an AI system.

[14] Critical digital infrastructure: internet exchange points, DSN service providers, top-level domain name registers, cloud computing, data centres.

[15] Image, audio or video material or text created or manipulated by AI that can incorrectly be seen by people as authentic and the truth.

[16] AI Act, Article 50.

Exceptions

If the system applies exclusively to one of the following areas, the standard questions in the AIIA will be sufficient. It will still be necessary for you to register the AI system (this can be done in the algorithm register) if it is a high-risk system that is covered by the exceptions.

Please note: this only applies if the AI system does not profile individuals. Otherwise, the system will **always** be a high-risk application.[17]

Exceptions, AI for:
1. Military purposes
2. Law enforcement and justice
3. Research into AI
4. Open source AI in the development phase (before entering the market)
5. Non-professional activities
6. The AI system does not fundamentally influence the outcome of the decision-making. Examples might include a system that carries out purely procedural tasks or serves as verification and improvement of a previous human activity (for example correcting language). Systems used to detect decision-making patterns are also included in this.[18]

N.B.: Many definitions relating to AI applications are still in the process of development. It may therefore be possible that, after completing these questions, it is not yet totally clear which category the AI system is in. You may wish to consider putting this question and others to the Dutch central government Regulatory AI Sandbox (more information).

[17] AI Act, Article 6.3.
[18] AI Act, Article 6.3.

Appendix 2: High-risk systems

Is the AI system a high-risk application (see Appendix 1)? Use the following additional questions to help you. This will enable you to apply the right safeguards to ensure that you use the AI system responsibly.

The AI Act describes a range of different parties, each of which has their own role or responsibility with regard to an AI system. This appendix includes the questions intended for the user (deployer) and developer (provider) of an AI system. A provider, for example, procures an AI system and provides it in their name. Alternatively, they make substantial changes to the system.[19]

Questions if you wish to use a high-risk AI system

According to the AI Act, high-risk systems must meet several requirements. The questions below will help you with this.

Effects
- Does the AI system constitute a significant risk to people's health, safety or fundamental rights?[20] Please provide reasons below as to why/why not.

Maintenance & management
- How will monitoring and operation of the system be safeguarded?[21]
- How will the logs produced by the AI system be retained for at least six months?[22]
- Have measures been taken to arrange human oversight by people with the necessary competence, training and authority?

Technical robustness
- Is the data (input) relevant and representative, taking account of the intended purpose (question 1 of 1.1) of the AI system?[23]
- Has technical documentation been provided for the AI system?[24]
- For the purposes of human oversight, is it possible to verify, interpret or possibly disregard the **output** (**data**)?[25]

Reliability
- Have the metrics for levels of accuracy been included in the user instructions?

Data governance
- Does the AI system's training data meet the following quality requirements?
- Relevant design choices for datasets.
- Providing transparency concerning the origin of data.
- Relevant processing activities, such as annotation, labelling, cleaning, updating, enhancement and aggregation.
- The setting of assumptions that the data must measure and represent.
- An assessment of the availability, quantity and suitability of the necessary datasets.
- An assessment of potential bias in the data with negative consequences for health, safety, fundamental rights and discrimination.

[19] AI Act, Article 25.
[20] AI Act, Article 6.2a.
[21] AI Act, Article 26.5.
[22] AI Act, Article 26.6.
[23] AI Act, Article 26.4.
[24] AI Act, Annex IV.
[25] AI Act, Article 14.4.

- Measures to detect, prevent and limit bias.
- Identifying and countering shortcomings that impede regulatory compliance.

Risk management
- How was the AI system tested for its intended purpose and to ensure that it meets the risk management requirements before being put into service?
- Has a risk management system been determined and documented? This will include the following steps:
  - A risk analysis of the AI system for health, safety or fundamental rights
  - An assessment and evaluation of the risks that may occur
  - An evaluation of new risks following market entry, based on the AI system's monitoring system[26]
  - The drawing up of risk management measures
- How will you safeguard potential collaboration with the supervisory authorities and other competent authorities?[27] This refers to such areas as contact persons, accessibility, etc.
- Is it likely that vulnerable groups (such as children) will have access to the AI system? In that case, the risk management systems will need to be especially strict.

N.B.: The AI office is developing a template for a questionnaire, partly using an automated tool, to provide deployers with the simplified way of meeting these obligations.

Communication
- How will you communicate about the putting into service of the high-risk AI system?[28]
- Has the AI system been registered in the EU database for high-risk systems?[29]
- Have user instructions been compiled? These must at least contain the following:[30]
  - The identity and contact details of the provider;
  - Characteristics, capacities and limitations (purpose);
  - Potential future changes;
  - Measures regarding human oversight;
  - The computational and hardware resources required, expected lifetime and any necessary maintenance and care measures (including the frequency);
  - A description of the mechanisms included in the AI system;
  - The levels of accuracy and the relevant accuracy metrics.

Generative AI
- Is there information and documentation to clarify the possibilities and limitation of the AI system?
- Is there policy with regard to how copyrights can be safeguarded by the AI system?
- Has a detailed summary of the content with which the AI was trained been provided?

Miscellaneous
- If use is being made of remote biometric identification, has permission been provided by a judicial authority?[31]

[26] The post-market monitoring system shall actively and systematically collect, document and analyse relevant data which may be provided by deployers or which may be collected through other sources on the performance of high-risk AI systems throughout their lifetime, and which allow the provider to evaluate the continuous compliance of AI systems with the requirements set out in Chapter III, Section 2.
[27] AI Act, Article 26.12.
[28] AI Act, Article 26.7.
[29] AI Act, Article 26.8 and Articles 49 & 71.
[30] AI Act, Article 13.2 and 13.3.
[31] AI Act, Article 26.10.

Questions for developers (providers) of high-risk AI systems

Developers (or to use the term in the AI Act: **providers**[32]) of a high-risk AI system must also answer the following questions.

- Has documentation been compiled to indicate compliance with requirements? See Annex IV of the AI Act for the requirements.
- How can technical documentation be found by users? Even if there is no involvement in the roll-out of your system?
- Has indication been provided that this is a high-risk system, and if so how?[33]
- Has a quality management system been set up?[34] You will also need to document this. The following minimum information will be required:
  - A strategy for regulatory compliance
  - Design procedures
  - Quality control procedures
  - An inspection procedure
  - An overview of the technical requirements and what is needed to enforce these
  - Data management procedures
  - A risk management system
  - A monitoring procedure
  - A procedure for reporting serious incidents
  - A communication strategy with the competent authorities
  - Systems and procedures for the registration of relevant documentation
  - Resource management, and security-of-supply related measures
  - An accountability framework
- How will the AI system logs be stored?[35]
- Is there a conformity assessment procedure and has this been recorded in an EU declaration of conformity?[36] How will you demonstrate that the requirements determined for the AI system have been met? These must be presented to a designated body and retained for at least ten years.
- Has the declaration of conformity been approved by the relevant supervisory authority?[37] This is necessary only if the high-risk AI system relates to a relevant high risk product (Appendix I).
- If the AI system is intended to comply with the AI Act, has CE Marking been provided, for example in the documentation?[38]
- If the system is no longer compliant with the AI Act, how will corrective measures be taken, such as removing this system from the market, deactivating or recalling it? How will users be informed about this?[39]
- Does the system comply with the European accessibility requirements?[40]

[32] '**Provider**': a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge.
[33] AI Act, Article 16.
[34] AI Act, Article 17.
[35] AI Act, Article 19.
[36] AI Act, Articles 43 & 47.
[37] AI Act, Article 16k.
[38] AI Act, Article 48.
[39] AI Act, Article 20.
[40] Directives (EU) 2016/2102 and (EU) 2019/882.

Appendix 3: Points to consider regarding generative AI

For the use of generative AI, such as large language models (LLMs), there are a number of important points to consider when completing the AIIA. This appendix focuses on the most important ways in which generative AI differs from other AI systems.

What is generative AI?
Generative AI is a type of AI in which algorithms are used to generate content. In its government-wide vision on generative AI[41], the Dutch Cabinet has stated that generative AI must serve the purpose of improving human well-being and autonomy, sustainability, prosperity, justice and security/safety. According to the vision, by applying responsible applications of generative AI, we are seizing the opportunities that this technology has to offer.

Under the bonnet, generative AI makes use of a neural network made up of billions of parameters. The output that is chosen is based on statistics and there is no underlying logic or knowledge about reality. The output varies and is difficult to reproduce or account for. This brings us immediately to the most important point to consider before completing the AIIA: you should only use generative AI if it is acceptable for the outcome not to be explainable or verifiable.

Provisional position for central government organisations: in principle not permitted
The provisional position on the use of generative AI in central government organisations currently sets strict requirements for the use of LLMS in central government: "Non-contracted generative AI applications, such as ChatGPT, Bard and Midjourney, do not generally comply demonstrably with the relevant privacy and copyright legislation. Because of this, their use by (or on behalf of) central government organisations is in principle not permitted in those cases where there is a risk of the law being broken unless the provider and the user demonstrably comply with relevant laws and regulations."

Points to consider when completing the AIIA
Below is a description of the most important ways in which generative AI differs from other AI systems. This will help when completing the AIIA. If any areas of the AIIA are not included below, this does not mean that they are not relevant, but that there are no specific points to consider relating to the fact that it is generative AI.

Purpose and necessity
- **Public values:** there are no specific points to consider for generative AI. The choice to use generative AI can often undermine transparency, explainability and sustainability.
- **Fundamental rights:** check whether the model has been trained and fine-tuned in an ethical way. This is the final step in which human feedback is used to improve the responses. As is the case with many other products, these people ('click workers') do not always have the best working conditions.[42]
- **Sustainability:** find out whether it is possible to achieve the same objectives using a less complex system or tool. Generative AI consumes huge amounts of energy both for training and using the model. Investigate whether energy-saving technologies can be applied, such as a smaller model, optimisation or a different software architecture.
- **Considerations:** when deciding whether or not to use generative AI, you must take into account the provisional decision on the use of generative AI in central government organisations (see introduction).

[41] https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2024Z00480&did=2024D01191
[42] AI draait op werk van miljoenen onzichtbare, slechtbetaalde mensen. Wie komt er voor ze op? - De Correspondent

Technical robustness

- **Bias:** always check the system's output for bias and assess what the risks of any bias might be. The data used to train a generative AI model often already contains bias, and the nature of the model (statistical language model) exacerbates this bias. Although software suppliers use human feedback and guardrails to remove the sharp edges, the bias remains in the model. In addition, other methods are available for reducing and/or detecting bias.[43]
- **Accuracy, reliability and reproducibility:** you should only use a generative AI model for purposes that do not require a high level of **accuracy**, reliability and reproducibility and/or you should take measures in setting up the system to reduce the likelihood of hallucinations. Do an assessment of the potential risks of an unreliable outcome. An LLM is a statistical language model with no knowledge of reality. The output is always different and not necessarily based on reality. A generative AI model also has the potential to 'hallucinate', generating responses that may sound logical and convincing, but are not true. Another challenging feature of generative AI is that it is difficult to measure the quality of the output.
- **Technical implementation:** In the case of an externally-hosted generative AI model, you should make contractual agreements about the use of the organisation's data (input). There is often a demand to use this as data to train the model further. There is potential for other (unauthorised) people to view these data (during use) if they make use of the right prompts. When procuring a system, the supplier must demonstrate that it meets all the requirements set. The Dutch Cabinet's position on generative AI expresses a preference for the use of open-source generative AI.
- **Explainability:** You should only use generative AI models if the explainability of the results is not a requirement. Do an assessment of the potential risks of this lack of explainability. Explain what the limitations are to any staff who have to work with the system. A generative AI is a highly complex system with billions of parameters, the operation of which is not explainable: it is a 'black box'. It is virtually impossible to demonstrate how the model reaches a response, although there are currently some initiatives to improve the transparency of this.[44]

Data governance

- **Data quality and integrity:** it is likely that the generative AI training data contains personal data and material protected by copyright. For most AI models, it is unknown which training data (and what the quality of it is) has been used in the model. The AI Act requires transparency. The GDPR does not allow personal data simply to be processed.
- **Privacy:** reach agreements with the supplier with regard to processing responsibility for the data entered (e.g. whether it will be used as training data). Draw up a proper processing agreement. If proper agreements are made, the likelihood of data breaches can be reduced.
- **Information security:** the OWASP has compiled a list of the ten most critical vulnerabilities, and therefore information security risks, of generative AI systems[45].

Accountability

- **Verifiability:** see 'explainability'.

[43] https://www.datacamp.com/blog/understanding-and-mitigating-bias-in-large-language-models-llms
[44] https://arxiv.org/abs/2309.01029
[45] OWASP Top 10: LLM & Generative AI Security Risks

This is a publication of:

Ministry of Infrastructure and Water Management

P.O. Box 20901

2500 EX The Hague

December 2024 | 73263