Ministry of Infrastructure
and Water Management

AI Impact Assessment

A tool to set up responsible AI projects

A collaborative production by colleagues from the Information Policy Directorate; the Human Environment and Transport Inspectorate (ILT; IDlab and Analysis Department), and the Directorate-General for Public Works and Water Management (Rijkswaterstaat /RWS Datalab).

Contents

Explanatory Notes

Artificial intelligence[1] can be employed to expedite processes or improve safety when performing such tasks as inspecting asphalt quality or law enforcement in the event of violations of maritime law. In addition to opening up opportunities, AI also entails dangers. In November 2020, the cio Council approved the development of a draft AI Impact Assessment (AIIA) in order to garner more attention to responsible AI. The ILT IDlab, the RWS Datalab, and the Information Policy Directorate of the Ministry of Infrastructure and Water Management embarked on this project and collectively developed this new version of the AIIA. The Administrative Council [*Bestuursraad*] endorsed the AIIA on 4 July 2022.

The AI Impact Assessment (AIIA) is utilised to facilitate discussions on AI systems. It explores obstacles in the data, the system, and the algorithms, taking account of applicable rules and regulations. The AIIA serves as an instrument for dialogue and for recording thought processes, thus enhancing, inter alia, accountability, quality, and **reproducibility**. The AIIA is expected to result in a comprehensible document which clearly manifests the considerations underlying the decision to use AI in a project.

The **commissioning client** bears primary **responsibility** for implementing or commissioning the AIIA. An AIIA **must** be drawn up for each **ai system**. However, each AIIA is expressly completed in a **proportionate** manner, befitting the impact and the risk profile of the application. Responsibility for proportionality is vested with the project leaders and the commissioning client.

AI can also be employed in research. In this scenario, attention needs to be focused on, e.g., issues such as false positives and false negatives, and on responsibility for and explainability of the results. Furthermore, AI can be used to generate hypotheses, which are subsequently elaborated using AI or other technologies. In short, it is also important for researchers to properly consider the pros and cons of AI, by reference to this AIIA. Evidently, irrelevant questions can be skipped, for example, if the system is not commissioned.

The AIIA must be completed – proportionately – before an AI system is put into operation. Subsequent regular adjustment of the AIIA is imperative, for example, if the AI system is being used for other purposes or if the AI system is modified. In addition, new risks may arise over the course of time. **Project leaders** / commissioning clients must monitor this periodically.

The efficient and responsible implementation of an AI project requires more than an AI Impact Assessment. Other helpful measures include the organisation of **moral consultations**, ensuring careful commissioning, and informing the stakeholders. Cf. the most recent guidelines entitled AI *voor opdrachtgevers* [AI for Commissioning clients], available from the Information Policy Directorate.

Is the AIIA insufficiently geared to an AI project or system that you are considering? Or do you have other comments or questions regarding the AIIA? Then please contact the CIO office at the Information Policy Directorate.

[1] This document uses the definition of AI as set down by the Netherlands Court of Audit [*Algemene Rekenkamer*]; see Definitions.

Guidelines for completion

Please take note of the following when completing the AIIA:

- The AIIA serves as an instrument for dialogue and verification.
- The AIIA must be adjusted if the **ai system** is going to be used for other purposes.
- **bold print** indicates clickable concepts, which are defined in Appendix "Definitions".
- The AIIA must be completed before an AI system is put into operation.
- The AIIA is mandatory, yet the extent to which it is completed falls under the expertise of the **project leader**.
- A simple "yes" or "no" does not suffice when answering the questions.
- The questions are coded and numbered. The initial letter refers to the chapter (e.g., T for Technological Robustness); the letter "o" is added if the question pertains to a green ● sub-question.

Questionnaire

The full AIIA comprises some 100 questions. An AIIA is mandatory when setting up or procuring AI systems, yet its completion is expressly proportionate: to be decided by the commissioning client and the project leaders themselves. We call this mandatory, yet flexible. This means, primarily, that you need to exercise common sense when reviewing the impact of your **ai system**. The umbrella questions in the blue ● boxes are mandatory for all forms of **artificial intelligence**; they help to facilitate a discussion on the advisability of the AI system. The sub-questions in the green ● boxes are intended to flesh out the general questions in more specific terms. As the relevance of the green questions is situation-dependent, they are not mandatory (i.e., they are flexible). Based on their own risk estimate, the **commissioning client** and the **project leader** may decide to complete such questions, nonetheless. Please note that the National Audit Service [*Auditdienst Rijk*] and the Netherlands Court of Audit [*Algemene Rekenkamer*] may check the system for correctness and security. Furthermore, a fully completed AIIA does not, by definition, entail that the AI is secure. Once the AI Act takes effect (COM/2021/206 final), completion of any questions marked with a red asterisk ✶ will also be mandatory for **high-risk ai**. In anticipation thereof, completion of such questions is already recommended for all AIIAs. All answers must be substantiated, i.e., a simple "yes" or "no" will never suffice as an answer to any question.

The Appendix entitled "Who Does What" contains an overview that may help to determine who needs to complete which question. For example, some questions are better completed by a data scientist, whereas others are more appropriately completed by a legal expert.

Introductory questions

The introductory questions pertain to the general aspects of the **ai system** that you are developing. These questions address the purpose of the system and the role that the system will have within the organisation. For example, the questions may deal with the technology employed or the person(s) who will be responsible.

Purpose of the system

These basic questions ask you to indicate the purpose for which you are developing a system. The answer provided to these questions is relevant to the rest of the AIIA.

i 1. Provide a brief description of the intended **ai system** (title, general outline, definition of the problem, the domain).

i 2. Why have you opted for these specific technologies? (In this respect, it is important for all the considerations, from robustness to human rights, impact on user and on end user, accountability, et cetera, to be included in the answer.)

i 3. What is the purpose of the AI system and what results is it intended to produce?

i 4. What goal is being linked to the AI system, as referred to in the Netherlands Court of Audit report *Aandacht voor Algoritmes* [Attention to Algorithms][2]? **Goal 1**, **Goal 2** or **Goal 3**?

Role within the organisation

Many of these questions could lead to a discussion. In addition to addressing the creation and details of an AI system, the questions also require thorough consideration of the overall impact that the AI system will have. As these are fundamental questions, you need to think them through carefully. Try and make the necessary differentiations. A positive impact on thousands of citizens, yet a negative impact on ten citizens, will not render an AI system immediately useless; however, appropriate customised solutions will need to be developed for the ten citizens affected by its negative impact.

These questions will also require you to set down the division of tasks within the development and operation of your system. These roles are defined in the Definitions. Base your answers on these definitions.

i 5. Which sections of the organisation intend to use the AI system, and what is the intended impact on the organisation?

i 6. Describe the division of tasks within the conception of the AI system (such as **developer**, **commissioning client**, **project leader**, **management organisations**, and **ultimately responsible party**).

i 7. Who will be the **user** of the AI system, who will be the **end users** working with the system, and which **involved parties** will the AI system impact?

[2] Netherlands Court of Audit (2021), *Aandacht voor Algoritmes* [Attention to Algorithms, in Dutch].

Fundamental rights & fairness

AI systems, like many technologies, can both promote and compromise basic rights. Considering the prime importance of protecting basic rights and the particular risks that the use of AI systems may entail in terms of injuring such basic rights, focusing separate attention on this issue is vital. This chapter is closely intertwined with the chapter on Data Governance, which deals with privacy. The right to privacy is a basic right; however, on account of its nature, the privacy issue will be discussed in a chapter of its own.

Basic rights

f 1. How may the use of the **ai system** impact basic civil rights?

f 2. Is the use of this system in the purview of achieving the intended goals **proportionate** and **subsidiary**? In other words: is its impact in proportion to the intended goals, and are there no other, less invasive methods to achieve such goals?

f 3. What legal basis underlies the use of the **ai system** and the intended decisions that will be taken based on the AI system?

People who have an interest in the operation of the **ai system** need to be treated properly. This means that the (fundamental) rights of all the **involved parties** must be safeguarded. Cf. the Fundamental Rights and Algorithms Impact Assessment[3] for a sound explanation of this issue. The scope of the questions posed in this AIIA is, in fact, insufficient to delineate this topic.

Basic human rights apply when answering the questions. These rights are laid down in the Constitution and in the European Convention on Human Rights.

fo 1. What constitutional provisions could be applicable?

fo 2. Which of these constitutional provisions could be infringed by an improper execution of the **ai system**? What measures are being taken to prevent this?

Bias

f 4. How are you taking account of potential unwanted **bias in the input**, **bias in the model**, and **bias in the output** of the **ai system**?[4] ✶

BIAS pertains to making assumptions regarding objects, people, or groups. There are two sides to this issue. On the one hand, projecting conclusions about data on a new situation is a necessity. The fact is that in generalisations, we always make assumptions. On the other hand, it is imperative to prevent any misinterpretation through forms of unjustified and unwanted bias that may violate human rights.

[3] Dutch Ministry of the Interior and Kingdom Relations (2021), Fundamental Rights and Algorithms Impact Assessment.
[4] AI Act, Article 14, Paragraph 4.

Bias may be contained within all the aspects of the system: **bias in the input**, **bias in the model**, and **bias in the output**. Several types of bias are relevant in the AI development and implementation stages, e.g., **data bias** and **design bias**. Such types of bias are often caused by socio-economic assumptions, as a result of which they may enhance such socio-economic assumptions. Failure to adjust for such forms of bias may hamper the proper functioning of AI systems for all the **involved parties**.

The core element of this theme is awareness and integrity. Operating without any bias whatsoever is impossible. Many manifestations of bias have existed for decades and will – unjustly – not be recognised as such. Therefore, rather than focusing on zero-bias AI, we need to pursue maximum awareness of potential discrimination. Furthermore, it is important to ask critical questions regarding the origin and content of data and regarding the operation of AI systems.

Bias is closely related to **diversity**, **equality**, and **fairness** among people. We need to be aware, however, of the fact that assumptions may also pertain to such non-human aspects as nature or the living environment. In addition, with respect to the manner in which you are planning to mitigate bias, it may be relevant to distinguish between a potentially **negative impact**, **no positive impact**, and a potentially **positive impact**.

A positive impact may be explained as follows. In statistics, bias refers to a systematic error or difference. This is not always wrong. The fact is that many such systematic errors are deliberately introduced in models, e.g., in the form of regularisation. This helps to reduce variance, even if it entails a (minor) systematic difference. Bias can also be used to a positive effect. For example, an AI system may be intentionally developed to eliminate or reduce discrimination, through the introduction of bias.

Bias in input (data)

fo 3. Is the **input data** representative for the issue on which a decision needs to be taken?

fo 4. Are sub-populations protected, if necessary, when random sampling?

fo 5. Has the selection of input variables been substantiated and coordinated with the **involved parties**?

Bias in the model

fo 6. What measures have been taken to prevent the creation or fostering of unjust or unfair **bias** in an **ai system**?

fo 7. Can the **ai system** be used by the intended **end users** (i.e., irrespective of their characteristics, such as age, gender, or capacity)?

Bias in the output (data)

fo 8. Are there stop mechanisms, supervision mechanisms or monitoring mechanisms in place in order to prevent societal groups from being disproportionately affected by the negative implications of the AI system? Specifically for the Human Environment and Transport Inspectorate ILT: a distinction needs to be made here between supervised parties (OTS) and the rest of society.

Stakeholder participation

f 5. Have all **stakeholders** been mapped by way of a stakeholder analysis, and has a stakeholder dialogue been initiated?

Several target groups are involved in **stakeholder** participation, in pursuit of **diversity**, non-discrimination, and justice. Achieving fairness in AI requires proper consideration of inclusion and diversity throughout the **ai system** life cycle. In this AIIA, this concept frequently also extends to **involved parties**.

Coordination of AI systems is essential in order to encourage AI system developers to look beyond their own wavelength and make them aware of implicit assumptions or consequences. Such coordination may extend to, e.g., your own team, the customer, the end user, involved parties, hands-on experts, domain experts such as universities, other government organisations, et cetera. Setting up moral consultations[5] is recommended.

fo 9. Which individuals and/or groups have been consulted in the development of the **ai system**?

fo 10. Are the stakeholders aware of the reason why specific input variables (that may include them) have been selected?

fo 11. What feedback has been collected from teams or groups representing different backgrounds and kinds of experience? And how has the feedback been followed up?

fo 12. How will the AI system be introduced among the Ministry of Infrastructure and Water Management staff?

fo 13. How will the AI system be introduced in society?

[5] Physical Environment Consultative Council (May 2021), *Moreel Beraad* [Moral Consultations; in Dutch].

Technological robustness

Technological **robustness** reflects whether an **ai system** fulfils its intended purpose.

Accuracy

t1. How is the continuous accuracy of the system measured and safeguarded?

In general, an **ai system** needs to perform well. In order to minimise the risk of incorrect assessments, continuous monitoring of an AI system's performance is important. This extends to monitoring the AI system in both the development stage and the production stage. The quality of the data used is of equal importance. Ergo, an AI system is a work in progress; regular testing and re-training of AI systems remains a necessity. Quantification of the risk of incorrect assessment nonetheless is advisable.

The **accuracy** of the system can be determined by drawing up **acceptance criteria** for both the data and the system. Compliance with such criteria can be monitored by means of a metric. Acceptance criteria could be, e.g., a quantity of data or certain threshold values of the measuring system. Many different types of measuring systems (often referred to as "performance metrics" by data scientists) are available to quantify the quality of **models**, such as an accuracy score, a precision score, and a recall or F1 score. Important in this respect is that the measuring system and the acceptance criteria are properly geared to the data and the intended purpose of the AI system.[6] This must be coordinated with, inter alia, the outcome of the risk analysis (cf. "Risk management"), as over the course of time, the use of an AI system may give rise to new or different risks. Furthermore, continuous monitoring of the quality of the system is important, as is re-evaluation of the acceptance criteria and choice of measuring systems, if need be, during the re-training or continued development stages.

to1. What **acceptance criteria** have been set to measure the quality of the **input (data)** and the **output (data)** of the **model**?

to2. Do the acceptance criteria chime with the data and the purpose of the AI system?

to3. What evaluation measuring systems (performance metrics) will you be using to safeguard compliance with the acceptance criteria and why?[7] ✶

to4. How is the **output (data)** (periodically) checked for correctness, at random and continuously?

to5. How are differences in the output (data) vis-à-vis the acceptance criteria analysed in a timely fashion and corrected?

to6. What would be the results of using alternative **models**?

[6] The measuring system selected must befit the model and the data that is used to measure the quality. Take, for example, a system intended to label 5 in 100 words of a particular text. If the system labels 0 words in this text, the accuracy of the model is 95%. Ergo, if the accuracy score is used to determine the quality of the system, the model appears to perform really well. However, the recall is 0, which means that its performance is not good at all. Accuracy is, therefore, not appropriate for measuring the performance of this model.

[7] AI Act Article 15, Paragraphs 1 and 2.

Reliability

t 2. Is the **ai system reliable**?

A **reliable** AI system produces consistent results in similar cases. The key question with respect to reliability 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. Furthermore, it is important for the system to provide a reliable indication of how well the model is going to perform in new situations.

to 7. What are the main factors affecting the performance of the **ai system**?

to 8. Is part of the (sub) data set excluded from the model's learning process and only being used to determine its reliability, or is the reliability of the model calculated based on cross validation?

to 9. How has the (hyper) parameter tuning been substantiated and how is it assessed?

Technical implementation

The **technical implementation** outlines how the AI system has been integrated within the IT landscape of the organisation from a technical perspective. The specific hardware and software requirements of the AI system have been documented in order to be taken into account in the roll-out and management of the system. In addition, the system architecture reflects the inter-relationship between the different software components. A well-considered architecture reduces the operational risks entailed in the construction of a technical solution and builds a bridge between business requirements and technical requirements.

t 3. How has the AI system been implemented in technical terms?

to 10. Have you considered how the AI system fits into the existing technical and system infrastructure, and have appropriate measures been taken for its roll-out (if applicable)?

to 11. What is the system architecture like (how are the software components inter-related)?

to 12. Have specific hardware and software requirements, if any, been documented?

Reproducibility

t 4. Is the **ai system reproducible**? Has a process been set up to measure this?

Reproducibility refers to the registration of, e.g., which data has been used, the model development process, whether changes in the data have been tracked, whether the same **input (data)** produces consistent results, and whether certain situations or conditions may affect the **output (data)**. Reproducibility is about training, validation, and testing.

Reproducibility is closely related to traceability. The main purpose of **traceability** is the proper documentation of the data sets and processes. Data version management, the algorithm, and the training play a key role in this respect.

to 13. Can you reconstruct, now or in the future, the **output (data)** produced (i.e., have previous versions of the **model**, data sets, and conditions been saved through version management)?

to 14. Can the model be reconstructed on the basis of the given **hyperparameters** and a fixed seed?

to 15. Can the outlines of the **ai system** be reproduced on the basis of documentation?

to 16. How are modifications during the lifespan of the system being documented?

Explainability

t 5. Is the **ai system** sufficiently **explainable** and interpretable to the **developers**?

Technical **explainability** refers to the ability to understand both technical processes and related human decisions. Furthermore, the design choices made must be clear, as must the rationale underlying the use of the **ai system**. Cf. "Accountability" for explainability vis-à-vis **involved parties**.

to 17. During he AI system development process, how have you considered the explainability of the model?

to 18. To what extent can the particular way in which the AI system operates be explained to an external AI expert (cf. "Explainability")?

to 19. Has the expertise required for maintenance of the AI system been documented?

Data governance

Data **governance** refers to the (administrative) procedures in place regarding data, such as access, ownership, usability, integrity, and security. It also extends to the quality of the data that is being used.

Data governance also covers privacy. Privacy is one of the fundamental human rights that could potentially be compromised by AI. Adequate data governance and the protection of personal data, in accordance with the General Data Protection Regulation (GDPR) is, therefore, crucial.

Essential elements in terms of AI systems include providing transparency regarding data protection and privacy risks; reducing such risks to an acceptable level; and ordering regular (technical) assessments of such risks (e.g., by conducting a penetration test). Achieving this will require completion of the organisation's risk management process relating to data protection and privacy, prior to implementation of the AI system. Products that need to be delivered in this process include: CIA triad (BIA); implementation of and testing for compliance with the Netherlands Government Information Security Baseline (BIO); DPIA (when processing personal data); security tests; and, if necessary, a plan for improvement.

Data quality and integrity

d 1. How is the quality of the data being safeguarded?[8] ✶

Data quality is essential for the operation of an **ai system**. Collected data may contain, e.g., socially constructed **bias**, inaccuracies, errors, and mistakes (cf. "Bias"). This needs to be addressed before this data is used any further. The data sets and the procedures must be tested and documented at every step: training, testing, roll-out phase, and operational phase. This also extends to AI systems that have been procured elsewhere, rather than have been created within the organisation. The Dutch Public Records Act[9] sets requirements regarding storage methods and data retention periods.

[8] AI Act, Article 10, Paragraphs 2 and 3.
[9] Public Records Act 1995 [*Archiefwet 1995*]; (https://wetten.overheid.nl/BWBR0007376/2020-01-01)

General

do 1. Is the data used necessary for the **ai system**?

do 2. How are unintended data duplications prevented?

do 3. Can training and test data be updated if required in a particular situation? When will you decide to re-train, suspend, or further develop the AI system?[10] ✶

Input (data)

do 4. Does the data meet the assumptions underlying the **model**?

do 5. How has the **input (data)** that is used in the AI system been collected and combined?

do 6. How is the data being labelled?

do 7. What factors impact the quality of the input (data)? And how can this be addressed?

do 8. Has the input (data) been assessed for changes that occur during the training, testing, and evaluation phases? And over time, during use of the algorithm?

Output (data)

do 9. If output (data) is used as new input for another model, how is the output (data) being stored (e.g., a feedback loop)?

do 10. How are you ensuring that the output (data) is available in a timely fashion?

[10] AI Act, Article 14, Paragraph 4.

Privacy and data protection

d 2. What procedures are in place regarding personal data or confidential information (as recorded in the relevant DPIA)?

Privacy and data protection must be safeguarded throughout the life cycle of the **ai system**. Electronically stored data on human behaviour may enable AI systems to derive age, gender, and political, religious, or sexual preferences. When using personal data, you need to ensure that it cannot be used for discriminatory purposes; cf. "Bias".

In addition to personal data, other confidential data may be used that should not be disclosed. This applies to, e.g., the use of such confidential information as classified information or trade secrets. Such data must also be properly protected. The AI Regulation sets out additional rules regarding the use of (personal) data in AI systems.

Regarding personal data

do 11. Does the **ai system** work with personal data[11] (i.e., does the GDPR apply)? If so, please complete the following questions as well. If not, please continue under "Regarding confidential data".

do 12. Can the output of the AI system be traced back to individuals (i.e., does the GDPR apply)? If so, please complete the following questions as well.

do 13. Have far-reaching protective measures been implemented to secure the personal data?[12] ✶

do 14. Have officials been involved, e.g., the Data Protection Officer, the Privacy Consultant, the Information Security Officer, the Chief Information Officer, et cetera?

do 15. How often is the quality and the necessity of processing personal data evaluated?

do 16. Has attention been paid to third party rights regarding dissemination of information on the AI system? ✶

Regarding confidential data (not being personal data)

do 17. Is confidential data being used or stored?

do 18. How is the security of this information safeguarded?

[11] Definition as per Dutch GDPR [AVG], Article 4, Paragraph 1.

[12] AI Act, Article 10, Paragraph 5.

Risk management

Monitoring potential risks is important. Unforeseen risks may result in an **ai system** producing unreliable results, which can cause damage. The prevention principle is in place to ensure that damage is minimised. Damage can be incurred as a result of the AI system malfunctioning or, e.g., as a result of external **hacking attacks**.

Risk management

r 1. How has the AI system been tested for appropriate risk management measures?[13] ✶

The development and **putting into operation** of an **ai system** entail dangers that are addressed by this AIIA to the maximum extent possible. However, unforeseen issues may nonetheless arise. It is important to determine how such potential dangers will be dealt with. This also means that mechanisms need to be put in place to manage risks, which mechanisms must be properly tested. Such mechanisms may pertain to, e.g., the prevention of data poisoning, the scope of countermeasures, and the security of outcome storage locations. In addition, account must be taken of the fact that new risks may arise following the introduction of the AI system. Ergo, the countermeasures must be checked on a regular basis.

ro 1. How has access to the AI system and its components been structured? (For example, generic IT control measures)

ro 2. How has the **ai system** been tested for its intended purpose, before it is put into operation?[14] ✶

ro 3. Is it probable that vulnerable groups (such as children) will have access to the AI system? In that case, the risk management measures must be tightened up.[15] ✶

ro 4. Apart from the standard security measures in place within the Ministry of Infrastructure and Water Management, have additional measures been taken to secure the AI system?

ro 5. How will the alternative plan be set in motion, in the event of problems with the **ai system**?

ro 6. Has the implementation been proven correct, e.g., by means of unit tests, integration tests, and end-to-end tests, if applicable?

ro 7. How can the AI system interact with other hardware or software (if applicable)?

[13] AI Act, Article 9, Paragraph 5.
[14] AI Act, Article 9, Paragraphs 6 and 7.
[15] AI Act, Article 9, Paragraph 8.

Alternative procedure

r 2. What will be the plan in the event of issues involving the operation of the **ai system**?

It is advisable to have a plan ready in the event of issues arising with the **ai system**. This means that an alternative procedure must be available for situations in which issues occur involving the operation of the system. Such a plan may involve the option of reverting from machine learning to a more limited rule-based **model**.

It is good to realise that human expertise develops differently from that of an AI system. Take, for example, the effect of calculators on our mental arithmetic skills. An alternative procedure must accommodate this. What would be the impact of an AI system generating erroneous results?

ro 8. What would be the impact of the AI system failing?

ro 9. Cf. the above example regarding the calculators. What would be a potential equivalent effect if the AI system is put into operation, and would this be desirable?

ro 10. Is the ai system immune to errors or irregularities involved in interaction with natural persons or other systems?[16] ✶

Hacking attacks and corruption

r 3. How are information security risks being identified, reduced to an acceptable level, and tested (from a technical perspective)?

Information security risks, such as **hacking attacks** and **corruption**, must be managed to the maximum extent possible. Foreseeable risks must be framed by identifying them via the organisation's risk management process. This comprises, e.g., 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. Other important measures are the detection and technical obviation of errors and irregularities.

ro 11. How are unauthorised third parties prevented from taking advantage of vulnerabilities of the **ai system**?[17] ✶

ro 12. What would be the impact of unauthorised third parties accessing the source code, data, or outcomes of the **ai system**?

ro 13. Is it possible for someone to take advantage of the fact that an **ai system** is used rather than a human decision?

ro 14. How are you registering who is using the **ai system** and for how long?[18] ✶

[16] AI Act, Article 15, Paragraphs 1 and 3.
[17] AI Act, Article 15, Paragraphs 1 and 4.
[18] AI Act, Article 12, Paragraph 1.

Accountability

Actions within the national government are accountable within the organisation, to the House of Representatives, and to society. Currently, considerable attention is being focused on AI. The technology is increasingly implemented within the national government; however, ethical considerations regarding the use of AI are a source of great concern. Consequently, appropriate mechanisms must be set up to warrant **accountability** for and the results produced by AI systems.

Communication

v 1. Are you transparent vis-à-vis involved parties and end users regarding the limitations and the operation of the AI system? And do such limitations continue to receive sufficient attention for as long as they continue to apply?

v 2. Are mechanisms being set up to enable end users to comment on the system (data, technology, target group, et cetera)? And how or when are such comments validated (analysed and monitored)?

This section discusses two forms of communication to **end users**. Firstly, end users must be informed that they will be dealing with the results of an **ai system**. Secondly, end users are entitled, at all times, to know how an **algorithm** is determining the outcomes of an AI system. This also means that the purpose and limitations of the AI system must be communicated clearly and squarely. Both technological processes and related human decisions must be comprehensible and retrievable, e.g., by appointing a contact person commanding substantive knowledge of the AI system. Considering the self-learning nature of AI, such information cannot always be traced back for the full 100%. However, as a minimum, it must be possible to provide end users with an appropriate explanation of the process.

In addition, with respect to any question within this AIIA, it is important for citizens to be able to retrieve information on the AI system. They must have the opportunity to dispute the results of the AI system. This also means that data and the conditions under which the data has been made available must be stored (see "Archiving").

Communication with the ai system

vo 1. Are the **end users** of and the **parties involved** in the AI system informed of the fact that the results are generated by an AI system and of what this entails for them?

vo 2. Have instructions been drawn up for end use? Such instructions must comprise, as a minimum, the following:[19] ✶

- The name and contact data of the provider;
- Features, capacities, and limitations;
- Potential future modifications;
- Human supervision;
- Expected life span.

vo 3. What are the potential (psychological) side effects, such as the risk of confusion, preference or cognitive fatigue of the **end user** when using the AI system?

Communication regarding the outcomes of the AI system

vo 4. To what extent can the reason why the AI system operates in a certain way be explained to an **involved party**?

vo 5. Is the system sufficiently **transparent** in order to enable **end users** to interpret and make appropriate use of the output (data) of the system?[20] ✶

vo 6. Have steps been taken to provide end users with in-house training, if necessary?

Communication relating to the AI system

vo 7. How are you ensuring the proper processing within the organisation of comments submitted by involved parties and end users?

vo 8. Are involved parties aware of the steps that they may take if they wish to lodge an objection[21] to or a complaint against a decision from the AI system[22]? The same extends to lodging appeals.[23] ✶

Verifiability

v 3. How is the **ai system** verified?

v 4. How has human verification and supervision been safeguarded?

[19] AI Act, Article 13, Paragraphs 2 and 3.
[20] AI Act, Article 13, Paragraph 1.
[21] Dutch General Administrative Law Act [AWB], Article 7:1.
[22] Dutch General Administrative Law Act [AWB], Article 9.
[23] Dutch General Administrative Law Act [AWB], Article 8:1.

Verifiability refers to the methods by which the data and **model** evaluation processes and results can be checked. This verification process, in the form of audits, can take place internally or externally. Stricter requirements will need to be set with respect to systems that operate in more critical fields.

Insight into the sources, the system, and the outcomes is essential. This **responsibility** will generally be vested with the **user**.

In order to make autonomous use of **ai systems**, the **end user** must have a sufficient understanding of the system or be able to retrieve its working. Furthermore, it is important for knowledge of the AI system to be easily transferable, in the event of the system being used by a new end user who has not been involved in its development. For that reason, AI systems must, wherever possible, be set up in consultation with the intended end user. Supervision can be realised by way of **governance** mechanisms.

vo 9. How are you taking account of new legislation and regulations that may come into effect during the life span of this AI system?

vo 10. How are you ensuring the possibility of independent verification of the AI system?

vo 11. How are you verifying and interpreting the correctness of the **input (data)**?

vo 12. How are you verifying and interpreting the correctness of the **model**?

vo 13. How are you verifying and interpreting the correctness of the **output (data)**?[24] ✶

Archiving

Archiving refers to the storage of information for future use or to be used for other purposes. Examples of such purposes are reconstruction of the model (see "Reproducibility"), explaining the construction of the system to new staff (see "Explainability"), and giving account to an **involved party** (see "Accountability").

Input (data)

vo 14. How is the input (data) stored?

vo 15. What is the retention period for the input (data)?

Model

vo 16. How is the **model** stored?

Output (data)

vo 17. Can users correctly interpret the output (data)?

vo 18. What is the retention period for the output (data)?

[24] AI Act, Article 14, Paragraph 4.

Climate adaptation

AI systems can be helpful when developing solutions to the most urgent societal concerns. It is important, however, for their use to be as ecological as possible. The environmental sustainability of the full **ai system** supply chain must be safeguarded.

On the other hand, some AI systems are used for the particular purpose of gaining environmental benefits. Such impact must be weighed against the environmental costs of, e.g., running the system.

It goes without saying that proportionality is a key consideration here: spending a great deal of time and energy on measuring the environmental impact of a system whose ecological footprint is very small would hardly be justifiable.

vo 19. Will the introduction of the **ai system** (development, installation, and use) impact the environment, and how will such impact be measured?

vo 20. How is the impact of the AI system weighed against the environmental costs of running the AI system?

vo 21. What measures have been taken to minimise the environmental impact of the AI system?

Appendices

Definitions

Literature defines many of the concepts used in this document in different ways. Below is a list of univocal definitions used in this document.

acceptance criteria	Conditions to be met by the **ai system**, geared to the intended purpose and data. Such conditions may pertain to, e.g., the quantity of data, an accuracy standard for the **output (data)**, or an independent output verification mechanism. Wherever possible, acceptance criteria must be rendered quantifiable, in order to enable monitoring using an appropriate measuring system. Proper acceptance criteria are SMART and sufficiently different from one another, in order to enable efficient monitoring of all the relevant aspects of the AI system.
accuracy	A system is considered accurate if it is capable of making correct, accurate assessments. In 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 versus false results, the higher the accuracy.
ai system	A system that has (in part) been developed through the application of self-learning algorithms (machine learning, statistics, or logics) on historical data, for the purpose of producing predictions or recommendations, or of independent decision-making.
algorithm	A "recipe" or finite sequence of mathematical instructions departing from a given initial state and leading to a pre-set goal. Such **algorithms** are usually implemented into a computer program.
algorithm types	Various technologies can be employed to build AI, such as neural networks, random forests, or other forms of machine learning. Less complex **algorithms**, such as business rules or decision trees, can also be used.
artificial intelligence	There is no univocal definition for AI. In this document, we use the Netherlands Court of Audit description: "The ability [.] to correctly interpret external data, to learn from such data, and to use such lessons for the realisation of specific goals and tasks through flexible adjustment". We would also like to point out the description set down by the European Commission, although this is not yet used in this document: "Artificial intelligence (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 regarding objects, people, or groups, in many cases not based on actual measurements.
bias in input	Quality, consistency, and integrity of data is a key prerequisite for unbiased analysis.

bias in output	The manner in which the **output (data)** is used may affect the lives of people. In this respect, it is important that unjust correlation will not lead to causality.
bias in the model	How correct are the **models**; to what extent do they adjust for known flaws in the representativeness of the data? This concept may also refer to, e.g., what the **ai system** is learning and what is considered unwanted learning effects.
cio	Chief Information Officer.
ciso	Chief Information Security Officer.
commissioning client	An individual or organisational unit commissioning a contractor. The commissioning client bears ultimate responsibility (together with the **project leader**) for drawing up an AIIA.
corruption	The misuse or exploitation of system errors, or the exploitation of apparently neutral system features[25], as distinct from **unintended corruption**.
data bias	Refers to random samples that are not representative of the entire population.
data pipeline	Indicates how the data is delivered from the field to the **model**; the data movement process.
design bias	Problems in the technical design, including limitations of computer tools such as hardware and software.
developer	An organisation or an individual designing, developing, and/or training an **ai system**.
diversity	This refers to the identification of different types of "subjects" in our analyses. In this respect, we attempt to prevent groups of relevant subjects from unjustly being left out in the development of an **ai system**, as a result of which they would not be accommodated by the system.
domain expert	Someone commanding a great deal of knowledge on the problem area in which the **ai system** is built.
end user	End users are the individuals using the **ai system** in actual practice, within the "**user**" organisation. End users are natural persons. Whose hands are on the controls? Who within the organisation is collecting information from the AI system? Examples of end users are inspectors or road traffic controllers.
entity	A position within an organisational department.
equality	This refers to the concept that every similar subject is given equal treatment.

[25] Example: third parties may – accidentally (unintended corruption) or intentionally (intended **corruption**) – use incorrect information to complete a form whose results are used for **input (data)** of the **model**, resulting in the **algorithm** potentially running on faulty data.

explainability	AI is considered explainable when it is possible to explain how input variables contribute to an output of an algorithm.
fairness	Differences in subject treatment need to be able to be explained. To this end, a comprehensive picture of distinctive subject characteristics is essential, in order to demonstrate which characteristics actually play a part (rating the risk for party A lower than the risk for party B) as well as which characteristics actually do not (thus substantiating an equal risk for both party A and party B).
goal no. 1	The Netherlands Court of Audit lists three potential goals of AI. Goal no. 1 focuses on the automation of simple human actions. Such types of **algorithms** are often prescriptive and execute actions automatically, without human intervention. They pose little risk of errors impacting citizens, considering their high technological **transparency** and simple areas of application.
goal no. 2	The Netherlands Court of Audit lists three potential goals of AI. Goal no. 2 focuses on the facilitation of operational management. Compared to **goal no. 1**, goal no. 2 often involves the use of more complex data. Many **algorithms** are predictive, without automatic decision-making. There is a limited risk of errors impacting citizens, as the algorithm only engages in preparatory "work".
goal no. 3	The Netherlands Court of Audit lists three potential goals of AI. Goal no. 3 focuses on (risk) predictions and does not involve automatic decision-making. The risk of errors impacting citizens is high. For example, results may violate the law, or feature (unwanted) deviations ensuing from hidden limitations in the **input (data)**. This will jeopardise explainability. In addition, it entails the risk that the recommendation generated by the **algorithm** will affect the eventual staff decision.
governance	The action or manner of governing, the behavioural code, and the supervision of organisations. Governance concerns decisions that determine expectations, bestow power, or verify performances. It involves either a separate process or a specific component of management or leadership processes.
hacking attack	Breaking into the **ai system**, resulting in, e.g., data pollution, unwanted leaking out of (the operation of) an AI system, or corruption of software or hardware.
high-risk ai	High-risk AI is defined in the AI Regulation and often refers to products that closely relate to fundamental rights and/or product safety. Examples include AI in aircraft, ships, railway systems, road traffic, aircraft navigation, and drinking water supply. In anticipation of the AI Regulation coming into effect, we need to approach AI in a responsible manner. This means that we need to be aware of any high risks involved in an AI system.
input (data)	The data that is used for an intended purpose. Within the context of an ai system, this may involve raw data, e.g., observations from reality. Within the context of the model, such data will, as a rule, involve pre-processed data.

interest group	Body of **stakeholders** to measure **diversity**. Can be either a group of **end users** or a group of individuals impacted by the system.
involved party	Natural person or organisation with an interest or a self-perceived interest in the use or the outcomes of the system. The term "interested party" is deliberately not used in this context, as the concept is wider in scope than the term "interested party" as defined in administrative law. Examples include citizens, supervised parties, and end users themselves.
limited-risk ai	Limited-risk AI is defined in the AI Regulation. AI geared to interaction with humans, recognising emotions, or producing manipulated images. Examples include spam filters, summarising texts, classifying aviation incident topics, or AI systems regulating office lighting.
management organisation	An organisation that sets up and optimises the **ai system** application management.
minimum-risk ai	Any AI system that is neither prohibited nor falls under the **high-risk ai** or **limited-risk ai** categories.
model	A (simplified) mathematical representation of reality, which is used to process information. In an **ai system**, the mathematical representation is frequently "learned", partially or in its entirety, according to an **algorithm**. Thus, even the **developers** will not be able to fully explain how the model arrives at its outcomes.
moral consultations	Physical Environment Consultative Council (May 2021), *Moreel Beraad* [Moral Consultations].
negative impact	Negative consequences experienced by **involved parties**, ensuing from the use of the **ai system**, e.g., discrimination as a result of a bias in the AI system.
no positive impact	**Involved parties** that do not, by definition, experience any **negative impact** from the use of the **ai system** but, for example, remain in the same situation as before. This may entail the risk of such involved parties not experiencing the same "**positive impact**" from the use of the AI system as experienced by other involved parties.
output (data)	The data produced by an **ai system**, i.e., the results of the **model**.
parameter	A variable within the **model**. Modification of this variable will also modify the resulting variable of the model or of the calculation.
positive impact	Positive consequences experienced by **involved parties**, ensuing from the use of the **ai system**. For example, preferential treatment of a minority group. This may entail the risk that the positive bias is too optimistic and, therefore, not factual. Furthermore, it may have the drawback of a "**negative impact**" on other involved parties.
project leader	The party bearing ultimate responsibility for the project that comprises the **ai system**. The project leader also bears **ultimate responsibility** (together with the **commissioning client**) for drawing up an AIIA.
proportionate	AI is an encroaching technology involving explainability issues. Is the use of AI proportionate to the problem to be solved using the **algorithm**? The expected advantage must exceed the risk entailed in AI.
putting into operation	The putting into operation of an **ai system** refers to the first time it is used outside the organisation. In practice, this also entails an external test or pilot. The AIIA must be completed before the system is put into operation.

reliable	Featuring consistent behaviour and producing consistent results.
reproducibility	The ability to generate similar results, time and again, when an outlined procedure is executed.
responsibility	Refers to the possibility of tracing back actions to an entity in a unique manner, which entity is accountable for the said actions.
robustness	A system is considered robust if it has been developed using a preventative approach and behaves as foreseen and pre-outlined, thus preventing unacceptable damage.
seed	A "seed" is the point of departure for a random number generator. This generator always takes the same "route", starting from this point, to create new (pseudo) random numbers. Documenting the "seed" enables repetition of the "route" of (pseudo) random numbers. This means that this seed is required to verify reconstruction of a **model**, when the model uses random numbers for any of its operations. The seed itself is also a number. No specific requirements are set for this number; ergo, in many cases, something "recognisable" is chosen (e.g., "123456", or "0, 42, 1234", or the **developer**'s date of birth).
stakeholder	Individual or organisation that is capable of influencing a decision or activity, can be influenced by such decision or activity, or considers itself influenced. A stakeholder may be, e.g., the owner of data that is used.
subsidiarity	AI is an encroaching technology, involving explainability issues. Can the problem be resolved using less far-reaching means?
traceability	Processes and results are considered traceable when they can be verified.
transparent	An **ai system** is considered transparent when its operation and goals are communicated clearly, and its results are **explainable**.
ultimately responsible party	A role within the organisation carrying **responsibility** for the **ai system**. This comprises, e.g., responsibility for achieving the proper results with the AI system.
unintended corruption	Influencing the operation of the **ai system** without any malicious intent, e.g., by feeding faulty input or pressing the wrong buttons. Unintended corruption falls under **reliability** and is distinct from (intended) **corruption**.
user	According to the AI Regulation: "A (…) public authority, agency, or other body using an **ai system** under its **authority** (…)". The system is put into use by the user, which is never a natural person. The user may be, e.g., the Human Environment and Transport Inspectorate (ILT) or the Directorate-General for Public Works and Water Management (Rijkswaterstaat / RWS).

Who is who

Please state which individuals have played a part in the completion of this AIIA.

interest group: .....

ciso or **cio**: .....

Communications consultant: .....

Data scientists: .....

Data manager or source data owner: .....

domain expert: .....

Data Protection Officer: .....

Legal expert: .....

commissioning client: .....

Other members of the project team: .....

project leader: .....

Strategic ethics consultant: .....

Who does what

	Ch 1	Ch 2	Ch 3	Ch 4	Ch 5	Ch 6
interest group:						
ciso or **cio**:					X	
Communications consultant:		X				X
Data scientists:		X	X	X	X	X
Data manager or source data owner:						X
domain expert:		X	X	X	X	X
Data Protection Officer:				X		
Legal expert:		X				
commissioning client:	X	X	X	X	X	X
Other members of the project team:						
project leader:	X	X	X	X	X	X
Strategic ethics consultant:						

A publication of the Dutch Ministry of Infrastructure and Water Management

Postbox 20901

2500 EX The Hague

The Netherlands

March 2023