INSTRUCTIONS FOR USE (IFU) FOR HIGH-RISK AI SYSTEMS UNDER THE EU AI ACT

working template

Knowledge Centre Data & Society

Artificial Intelligence Flanders

Knowledge Centre Data & Society

Citation: T. Gils and W. Ooms (Knowledge Centre Data & Society), “Instructions for use (IFU) for high-risk AI systems under the EU AI Act – working template”, October 2024.

Contact: thomas.gils@kuleuven.be or wannes.ooms@kuleuven.be

www.data-en-maatschappij.ai

ABOUT THE KNOWLEDGE CENTRE DATA & SOCIETY

The Knowledge Centre Data & Society (KCDS) is the central hub in Flanders for the legal, social and ethical aspects of data-driven and AI applications. The Knowledge Centre brings together knowledge and experience on this topic tailored to industry, policy, civil society and the general public. Specifically, our objectives include:

- **Disseminating information and knowledge** on the ethical, legal and social aspects of data-driven applications and AI. All publications are made publicly available and aim to create a positive and proactive effect between these innovations and our society.
- **Promoting structural initiatives** that strengthen vision development and valorise the social and economic opportunities of data-driven applications and AI among governments, industry and other social actors.
- **Stimulating public awareness** and debate on the benefits & drawbacks and the social, ethical and legal aspects of data-driven applications and AI, in all layers of society.
- **Building and supporting a network and learning environment** for stakeholders and strengthening collaboration between different policy levels and actors.
- Contributing to the development of legal frameworks and guidelines on the use and framing of AI and data-driven applications for policy makers, businesses, organisations and employees. Our policy prototyping project is one of the activities that we develop in order to achieve this objective.

Please visit our website for more information about the KCDS, our objectives and our offering.

Before using this template - This template was created based on the Knowledge Centre Data & Society's 2023 Policy Prototyping Project as well as on our own expertise and interpretation of the AI Act's obligations. As a result, the layout of these instructions of use and the grouping of the required information in this template may differ from what you find elsewhere. This template was drafted when the AI Act entered into force in August 2024. Before completing this template, we recommend checking if new guidelines have been published by authorities relating to article 13 of the AI Act and its obligations.

This template marks sections and information as either "required" or "recommended but optional". Information marked as "required" is either explicitly included in the AI Act or is, in our interpretation, necessary for the proper functioning of the instructions of use (e.g. including operating instructions for the system in this IFU) and to interpret and understand the information explicitly required in the AI Act. Information marked as "recommended but optional" is not required under the AI Act but has proven useful in our policy prototyping project.

GENERAL GUIDELINES

TARGET THE PRIMARY USERS: ensure that an IFU is clearly targeted at the envisaged, primary professional users. Explicate or describe who the primary user should or is expected to be. If the AI system allows for multiple types of users to be involved, consider explicitly indicating the target audience for separate parts of the IFU.

STRUCTURE: ensure that the IFU follows a logical structure so that the target audience can understand the document properly without having to jump back and forth between sections. Start from a logical table of contents. The table of contents used by this template is only a suggestion and may be altered in accordance with the needs of your use case.

LANGUAGE: strive to use simple, concrete and clear language, adapted to the target audience. Avoid overly generic wording or technical jargon (if possible). Consider drafting a glossary of terms if various, similar concepts need to be used. Provide translations to the local language.

DESIGN: use a clear and legible font and page layout. Consider using visual elements such as images, tables or graphs (if possible and/or desirable) to enhance understanding.

TABLE OF CONTENTS

Who we are (required)

[In this section, you must identify your company and provide contact details. If you are working with an authorised representative in the EU (because your company is not established in the EU), you must also provide the identity and contact details of your representative.]

Provider: (Company X)

Product/service: (Product/service name and/or description)

Address: xxxx

Email: contact@company.com

Phone:

Authorised representative (if relevant):

Product or Service Description

Description (recommended but optional)

[In this paragraph, describe the product or service that involves the use of a high-risk AI system and to which the instructions relate. If the AI system has multiple versions, specify the version being used and describe how it (and the instructions for use) relate(s) to previous versions. This can include its nature (what is it?) and intended purpose as well as who is/are the expected user(s) (i.e. the 'deployer' under the AI Act) and expected persons affected by the product or service. You can clarify in this paragraph which parts of the product or service use AI. You can also generally describe the expected method of use of the product or service, the basic underlying technical process (e.g. which input data is used and why, which model, AI-technique and/or software is involved…) and the expected outcome/output.]

[Product/service name] is a [product/service description] for [expected user]. [Product/service name] is meant to be used on [expected persons, processes or data affected by the AI system].

[Product/service name] works by [describe the workings of the AI system/product.

Intended purpose (required)

[Describe here in more detail the purpose(s) for which you intend the AI system to be used, including its context and conditions of use. Relevant questions that could guide you are: when and how do you intend the system to be used? What is the AI system meant to achieve? How or for what objective should(n't) the output be used?]

Benefits for users and others (recommended but optional)

[You can optionally add a section describing the (intended) benefits for users and affected persons of the AI system. This can complement the above intended purpose and direct the use of the AI system.]

System requirements (required)

[In this section, describe the minimum computational, hardware or other technical resources required for the AI system to function. System requirements can include compatible operating systems, technical specifications such as available storage, RAM, processing power, related devices, internet bandwidth, etc.]

Installation, operation instructions and use (required)

[This section should include minimum but detailed instructions on how to install the AI system (where relevant) as well as on how to use or operate the AI system. This can include, for example, the conditions in which the system should be used, the actions the user must perform for proper use, who or what (e.g. other devices, products, services, data, etc.) is affected by the use of the AI-system, user qualifications (e.g. professional knowledge or skills), requirements of the affected person (e.g. age, location, gender), timing requirements, environmental conditions and the broader functional setting, what the expected effects or outputs of the AI system are and how the results of the AI system should be further used or interpreted. If these installation or usage/operation instructions are very detailed or lengthy, consider using a separate, technical manual featuring the detailed installation and operation instructions while integrating a clear link to such document in the IFU.

This section can also describe the system's compatibility with other devices or AI systems used in the relevant sector.]

Interpretation of results (identifying anomalies) (required)

[This section must include information for the user to interpret the results of the AI system as well as to identify anomalies in the results.

This information can include an overview of the potential results of the system as well as the meaning of these results. In addition to this overview, the general method by which the AI system arrives at the result (based on the input data) can be described, this may include mention of the relevant variables and value allocation to those variables, as well as an explanation of the variables. The section can also provide steps or guidelines on how to act based on the results of the AI system. The explanation can include how reliable the result is and, if relevant, how the result could be verified in the event of doubt and what to do in the event of unclear results.]

Data specifications

[Generally, ensure that information regarding input, training, validation and/or testing data is sufficiently detailed. Designate, at least, (i) the data used/required and its relevance, (ii) the data source or other collection modalities and (iii) data quality characteristics.]

Input data specifications for the user (required)

[This section should specify the input data to be provided by the user. This may include the data format, the data type (images, measurements, attributes, etc.) as well as the needed data quality and the minimum data quantity. This can include the capability of the system to handle (or not handle) different levels of quantity or quality of input data. You can also specify the data categories required (i.e. that you have the required data points for the relevant variables). If applicable, this section can also include the (technical) means by or source from which the data must have been gathered to be useful (or instructions on how to collect the input data) as well as any foreseeable circumstances, changes or attributes of the input data that could hinder the operation of the system (e.g. persons in a picture may not wear glasses....).

For clarity, this section may also feature which data are not (to be) used in the working of the AI system which may be relevant for the user to address concerns (internal or external) related to bias of the AI system (e.g. if gender, age, etc are not relevant personal data for an AI system used in recruitment).]

Specifications regarding the training/validation/testing data and related data sources (required)

[This section should describe relevant information on the training, validation and testing data sets used, taking into account the purpose of the AI system. This may include the sources of the training data, whether by referring to where the sources were found or the persons/sources it was collected from. This may also include information on the safeguards in place for the data sources (e.g. that personal data was collected in accordance with applicable rules, that commercial data was appropriately licensed etc.). This section can also describe the reliability of the data.

In terms of the training data itself, this section can describe the type of training data (e.g. biometric, images, etc.) as well as the data labelling practices used to train the AI system and information pertaining to the data set such as when it was created, which data is

included or not, etc. In addition, you can provide the training methodologies, techniques and training data sets used and data cleaning methodologies (e.g. outliers detection).]

Technical capabilities and usage limitations (required)

Technical capabilities and usage limitations

[This section should describe the technical capabilities, characteristics, and limitations of the AI system.

This includes particularly the (technical) characteristics and capabilities relevant for the user to explain the output of the system. This may include for example the underlying technical process (e.g. which AI model, AI-technique and/or software is used (e.g. proprietary or third party origin), the applied training/validation/testing methods, design specifications (including the general logic of the AI system and the algorithms), key design choices including their rationale and underlying assumptions, main classification choices and the relevance of different parameters.). In addition, you can provide the general system architecture or a high-level overview of the algorithm and its various components.

Relevant usage limitations that may impact the accuracy or usefulness of the system's output should also be included. This section should also describe reasonably foreseeable misuse[1], or circumstances (known and foreseeable) related to the use of the AI system which may lead to risks to health, safety, and fundamental rights (e.g. circumstances for which the system is not suited). This can include using the system for an adjacent, but not intended use or using the system in an adversarial way and can include limitations related to the user of the AI system, for example if certain expertise is required to use the system or interpret the results.]

Expected performance regarding specific (groups of) persons

[In this section, if appropriate/relevant, describe (in plain language) if the system's performance is affected by specific persons or specific groups of persons on which the system is intended to be used. For example, if the system is primarily trained on Caucasian people, people in a certain age group etc.]

[1] This is misuse both from reasonably foreseeable human behaviour and interactions with other systems

Accuracy, robustness, cybersecurity and performance (required)

[In this section, you must describe the level of accuracy, robustness, and cybersecurity for which your system was tested and validated, and which users may expect. You must also provide any known, foreseeable circumstances that may impact these levels of accuracy, robustness, and cybersecurity.]

Accuracy

[The disclosed accuracy metrics should be specific to the type of AI system you are providing. In general, we recommend providing extensive and explicit information regarding the concrete levels of accuracy, performance and other relevant/related metrics (e.g. evaluation or bias/fairness metrics). If necessary, additional background information should be provided to contextualize the metrics and enable a correct interpretation of results.

Below we have listed some concrete metrics used in our own policy prototyping exercise:

For instance, you can use the accuracy/error rate of the system's performance. The accuracy metrics can detail the sensitivity and specificity of the system. You can also provide 'area under the curve' performance metrics (complemented by confidence intervals). You can provide information on a threshold-based analysis of the system's output (relating to its sensitivity and specificity compared to acceptable thresholds).

Furthermore, you can provide information on whether these levels are in line with standard market practice or not. You can additionally provide information on the methods and sample sizes used during the system's testing. You can provide additional details around the system's accuracy based on race, age, gender,... of the involved natural persons or groups of natural persons (assuming this affects the system's accuracy) or add information on other factors which may impact accuracy. This explanation may be more technical and detailed than the information provided above on the expected performance regarding specific persons.]

Robustness

[Describe whether measures are in place to mitigate foreseeable risks. These risks may be intrinsic to the system (e.g. limitations of the systems and how they may be corrected) or external to it (e.g. malicious actions, behaviour of the user or input data).

Include descriptions on how the robustness of the system was tested. Describe, for example, the testing setup and variations used to ensure robustness as well as any measures for reliability.]

Cybersecurity

[Descriptions of the level of cybersecurity can include, where relevant, the various cybersecurity measures taken (incl. if and how products, services and data are encrypted), and/or implemented cybersecurity standards as well as followed best practices. This description can take the form of an evaluation of the AI system's cybersecurity, focusing on AI-specific cybersecurity vulnerabilities, along with details about any unaddressed cybersecurity risks. Cybersecurity details can be provided both for the system and its associated device(s) (if applicable) and for the storage environment of any used data. Information on the presence of backup systems can also be provided. This description can include organisational measures taken to ensure cybersecurity.

The description can also include advice or best practices for users to maintain cybersecurity, such as ensuring the safety of their own devices, ensuring only authorised access, and always ensuring the software is up to date. Ensure that such information is sufficiently concrete and actionable, following a realistic allocation of responsibilities.]

Human Oversight (required)

[In this section, detail the human oversight measures that are in place, or should be put in place by the user to ensure that AI systems can effectively be overseen by their users in order to prevent risks related to health, safety or fundamental rights. We recommend to consider making an explicit distinction between organisational and technical oversight measures.]

Organisational measures

[Describe procedural measures that are recommended for the user to properly interpret and use the results of the AI system. This may include as appropriate, measures for the user to not over-rely on the system's result, circumstances in which the use and/or the results of the AI systems should not be used or should be overruled, checks that should be performed to ensure that the limitations of the AI system are taken into account, training or education that should be followed by the user of the AI system, when confirmation of the results is required and by whom, measures to report incidents with the AI systems back to you, and measures to be taken to monitor the performance of the system over time. It should be clear which of these measures are taken by you as the provider and which are expected to be taken by the user of the system.]

Technical measures

[In this section describe the technical measures put in place to facilitate the interpretation of the outputs of the high-risk AI systems by the users. This can include technical interventions to remind the user of how to interpret the system's results, pre-installed measures to prevent overreliance on the system, to stop the operation of the system or to disregard or override the system or its results. The system could include measures to assess the reliability of the result (e.g. a confidence score) or measures that prevent a result due to poor input data.]

Logs (required)

[This section should describe the ways in which the AI system allows the user to collect, store and interpret logs of the system's use. The logging capabilities can include monitoring the use of the AI system, as well as any risks to health, safety or fundamental rights.

This section should describe how and where the user can access the logs and their history and find guidance on how to read and interpret them. The section should describe how the logging capabilities allow the user to identify situations that may present a risk to the health, safety or fundamental rights of persons, or signify a substantial modification of the AI system. You should explain how logging allows the monitoring of the operation of the AI system and allow users to report to you and the required authorities if a risk (to health, safety or fundamental rights) or a serious incident presents itself. The section should also explain how the logging and user reports facilitate your post-market monitoring of the system. This can include how you as a provider use the logs provided by the system's use (or if you do not). This can include improving the system's performance, solving errors, etc.

Additionally, this section can describe which actions and data points are logged when the system is used (e.g., user interactions, errors and exceptions, performance, technical information...).

If your AI system enables remote biometric identification systems[2] (except AI systems intended to be used for biometric verification for identity confirmation), the logging must include at a minimum recordings of the periods of use of the system, the reference database against which input data has been checked by the system, the input data for which a match was found, the identification of the persons involved in the verification of the results. This section should explain these measures to users and allow them to report situations presenting risks and serious incidents to you.]

[2] Meaning under the AI Act: "an AI system for the purpose of identifying natural persons, without their active involvement, typically at a distance through the comparison of a person's biometric data with the biometric data contained in a reference database"

Lifetime, Maintenance, and Care

Expected lifetime (required)

[In this section, describe the expected lifetime of the AI system. This includes whether the duration of the lifetime is determined, how long you intend to maintain and update the system as well as intended updates which you are aware of. Alternatively, if you cannot determine the system's lifetime accurately, you can also provide information on which factors influence the lifetime (e.g. technological advancements, practical circumstances, etc.) and what the estimated lifetime is at the time of drafting the IFU.]

Foreseeable changes and evolution of the system (required)

[If, at the moment of the initial conformity assessment, you know that the system, or its performance, will undergo pre-determined changes, this section must detail those changes.

If the changes are not (yet) specific, you can generally describe their nature and the elements that determine them (e.g. updates to improve the system's performance, updates based on user feedback).]

Maintenance and care measures (required)

[In this section, describe the necessary maintenance and care measures for the AI system to function properly, including the frequency of these measures. This includes information about software updates.

You should clearly distinguish between the measures you take as a provider (e.g. updating and maintaining security of external storage) and measures that must be taken by the user (e.g. installing the latest updates or allowing updates to be auto installed by you).

If users need to perform routine checks (e.g. calibration) of the system, devices (e.g. included or needed hardware) and network, specify the frequency as well as the instructions for the check here. This should include steps the user must take if they encounter errors or concerns during the check (e.g. contacting you or instructions on how they can resolve the issues themselves). If users need specific training, expertise or tools to perform the maintenance and care, this should also be specified here. This can include measures to be taken by the user to monitor the performance of the system over time.

You can use this section to explain the need and benefits of updating and maintaining the AI system. In addition, you can also specify maintenance and updates of adjacent devices and systems to ensure the functioning of the AI system (e.g. requiring up-to-date web browsers, operating systems, etc.).

This section can include information on how users can provide feedback relating to the maintenance and care of the system.]

FAQ & Troubleshooting (recommended but optional)

[You can optionally add a FAQ and troubleshooting section to the IFU. This may help reduce the number of questions you receive from users and provide them with easy access to the solution for common problems.]

References (recommended but optional)

[You can optionally add a section in which you refer to other documents or policies of your organisation which may be relevant for the user. This can include for example, general terms and conditions of use, your privacy policy, reference materials, code of conduct, etc.

This section can also include standards or best practices with which the AI system complies.]

Acknowledgements

Knowledge Centre Data & Society

Thomas Gils
Research Associate
Centre for IT and IP Law/ KCDS

Lotte Cools
Research Associate
Centre for IT and IP Law/ KCDS

Wannes Ooms
Research Associate
Centre for IT and IP Law/ KCDS

Prof. dr. Jan De Bruyne
Co-director KCDS

Knowledge Centre
Data & Society