Third Draft of the General-Purpose AI Code of Practice

COMMITMENTS BY PROVIDERS OF GENERAL-PURPOSE AI MODELS

TRANSPARENCY SECTION

Introductory note by the Chair and Vice-Chair of the Transparency Section

The Transparency section of the Code of Practice describes three Measures which Signatories commit to taking to comply with their transparency obligations under Article 53(1)(a) and (b) and the corresponding Annexes XI and XII AI Act.

In this third draft, to streamline fulfilment of the commitments contained in Measure I.1.1 and facilitate Signatories' compliance, we have included a user-friendly Model Documentation Form which allows Signatories to easily document the necessary information in a single place.

The Form clearly indicates for each item whether it is intended for downstream providers, the AI Office or national competent authorities. Whilst information intended for downstream providers should be made available to them proactively, information intended for the AI Office or national competent authorities is only to be made available following a request from the AI Office, either *ex officio* or based on a request to the AI Office from national competent authorities. Such requests will state the legal basis and purpose of the request and will concern only items from the Form strictly necessary for the AI Office to fulfil its tasks under the AI Act at the time of the request, or for national competent authorities to exercise their supervisory tasks under the AI Act at the time of the request, in particular to assess compliance of high-risk AI systems built on general-purpose AI models where the provider of the system is different from the provider of the model.

Finally, in accordance with Article 78 AI Act, the recipients of any of the information contained in the Model Documentation Form are obliged to respect the confidentiality of the information obtained, in particular intellectual property rights and confidential business information or trade secrets, and to put in place adequate and effective cybersecurity measures to protect the security and confidentiality of the information obtained.

Nuria Oliver
Working Group 1 Co-Chair

Rishi Bommasani
Working Group 1 Vice-Chair

RECITALS FOR TRANSPARENCY SECTION

Whereas:

a) The Signatories recognise the particular role and responsibility of providers of general-purpose AI models along the AI value chain, as the models they provide may form the basis for a range of downstream systems, often provided by downstream providers that need significant understanding of the models and their capabilities, both to enable the integration of such models into their products and to fulfil their obligations under the AI Act (see Recital 101 AI Act).
b) The Signatories recognise that in the case of a modification or fine-tuning of a model, the obligations for providers should be limited to that modification or fine-tuning to safeguard proportionality (see Recital 109 AI Act).

Commitment I.1. Documentation

LEGAL TEXT: Articles 53(1)(a), 53(1)(b), 53(2), 53(7), and Annexes XI and XII AI Act

In order to fulfil the obligations in Article 53(1), points (a) and (b) AI Act, Signatories commit to drawing up and keeping up-to-date model documentation in accordance with Measure I.1.1, providing relevant information to providers of AI systems who intend to integrate the general-purpose AI model into their AI systems (downstream providers hereafter), and to the AI Office upon request (possibly on behalf of national competent authorities upon request to the AI Office when this is strictly necessary for the exercise of their supervisory tasks under the AI Act, in particular to assess compliance of high-risk AI systems built on general-purpose AI models where the provider of the system is different from the provider of the model[1]), in accordance with Measure I.1.2, and ensuring quality, security, and integrity of the documented information in accordance with Measure I.1.3. These Measures do not apply to providers of open-source AI models satisfying the conditions specified in Article 53(2) AI Act, unless the models are general-purpose AI models with systemic risk.

Measure I.1.1. Drawing up and keeping up-to-date model documentation

Signatories, when placing a general-purpose AI model on the market, commit to having prepared a document entitled "Information and Documentation about the General-Purpose AI Model" (hereafter Model Documentation) containing all the information referred to in the Model Documentation Form below.

Signatories commit to reporting the information requested in the Computational Resources and Energy Consumption sections in consistency with any delegated act adopted in accordance with Article 53(5) AI Act to detail measurement and calculation methodologies with a view to allowing for comparable and verifiable documentation.

In case of relevant changes in the information contained in the Model Documentation, Signatories commit to updating the Model Documentation to reflect the new information while keeping previous versions of the Model Documentation for a period ending 10 years after the model has been withdrawn from the market.

[1] See Article 75(1) and (3) AI Act and Article 88(2) AI Act.

Measure I.1.2. Providing relevant information

Signatories, when placing a general-purpose AI model on the market, commit to publicly disclosing via their website, or via another means if they do not have a website, contact information for the AI Office and downstream providers to request access to the relevant information contained in the Model Documentation.

Upon a request from the AI Office pursuant to Articles 91 or 75(3) AI Act for one or more elements of the Model Documentation that are strictly necessary for the AI Office to fulfil its tasks under the AI Act or for national competent authorities to exercise their supervisory tasks under the AI Act, in particular to assess compliance of high-risk AI systems built on general-purpose AI models where the provider of the system is different from the provider of the model[2], Signatories commit to providing the relevant elements contained in the most up-to-date Model Documentation, or otherwise the necessary additional information, subject to the confidentiality safeguards and conditions provided for under Articles 53(7) and 78 AI Act.

Signatories commit to providing to downstream providers the information contained in the most up-to-date Model Documentation and intended for downstream providers, subject to the confidentiality safeguards and conditions provided for under Articles 53(7) and 78 AI Act. Furthermore, subject to the same confidentiality safeguards and conditions, Signatories commit to providing additional information necessary to enable downstream providers to have a good understanding of the capabilities and limitations of the general-purpose AI model and to comply with their obligations pursuant to the AI Act.

Signatories commit to taking all the above-described actions in a timely manner.

Signatories are encouraged to consider whether the documented information can be disclosed, in whole or in part, to the public to promote public transparency. Some of this information may also be required in a summarised form as part of the public summary for training content that providers must make publicly available under Article 53(1), point (d) AI Act to be determined in a template to be provided by the AI Office.

Measure I.1.3. Ensuring quality, integrity, and security of information

Signatories commit to ensuring that the documented information is controlled for quality and integrity, retained as evidence of compliance with obligations of the AI Act, and protected from unintended alterations. In the context of drawing-up, updating, and controlling the quality and security of the information and records, Signatories are encouraged to follow the established protocols and technical standards.

Model Documentation Form

Below is a static, non-editable version of the Model Documentation Form. In this version, the input fields cannot be filled in, and the button at the bottom of the form—intended to generate a version of the form containing only information intended for downstream providers—is non-functional. In the final draft of the Code, this Form will be fully interactive and editable.

[2] See Article 75(1) and (3) and Article 88(2) AI Act

Model Documentation Form

This Form includes all the information to be documented as part of Measure 1.1. Crosses on the right indicate whether the information documented is intended for the AI Office (AIO), national competent authorities (NCAs) or downstream providers (DPs), namely providers of AI systems who intend to integrate the general-purpose AI model into their AI systems. Whilst information intended for DPs should be made available to them proactively, information intended for the AIO or NCAs is only to be made available following a request from the AIO, either ex officio or based on a request to the AIO from NCAs. Such requests will state the legal basis and purpose of the request and will concern only items from the Form strictly necessary for the AIO to fulfil its tasks under the AI Act at the time of the request, or for NCAs to exercise their supervisory tasks under the AI Act at the time of the request, in particular to assess compliance of high-risk AI systems built on general-purpose AI models where the provider of the system is different from the provider of the model.

Any elements of information from the Model Documentation Form shared with the AIO, NCAs or DPs shall be treated in accordance with the confidentiality obligations and trade secret protections set out in Article 78.

Date the document was last updated: **Document version number:**

General information

		AIO	NCAs	DPs
Legal name for the model provider:		☒	☒	☒
Model family:	The identifier, if any, for the collection of models (e.g. Llama).	☒	☒	☒
Versioned model name:	The unique identifier for the model (e.g. Llama 3.1-405B).	☒	☒	☒
Model authenticity:	Evidence that establishes the provenance and authenticity of the model (e.g. a secure hash if binaries are distributed, the URL endpoint in the case of a service), where available.	☒	☒	☐
Release date:	Date when the model was first released through any distribution channel.	☒	☒	☒
Union market release:	Date when the model was placed on the Union market.	☒	☒	☒
Model dependencies:	The list of other general-purpose AI models that the model builds upon, if any (e.g. the list for llama-3.1-nemotron-70b would be [llama-3.1] and the list for llama-3.1 would be empty). For each listed model dependency, please include a copy or link to the associated Model Documentation or indicate that the Model Documentation is not accessible.	☒	☒	☒

Model properties

		AIO	NCAs	DPs
Model architecture:	A general description of the model architecture, e.g. a transformer architecture. [Recommended 20 words]	☒	☒	☒
	If the model is a general-purpose AI model with systemic risk, provide a detailed description of the model architecture, specifying where it departs from standard architectures where applicable. If the model is not a general-purpose AI model with systemic risk, write 'N/A'.	☒	☐	☐
Design specification of the model:	A general description of the key design choices of the model, including rationale and assumptions made, to provide basic understanding into how the model was designed. [Recommended 100 words]	☒	☒	☐
Input modalities: *For each selected modality, please include maximum input size or write 'N/A' if not defined.*	☐ Text; ☐ Images; ☐ Audio; ☐ Video; If any other please specify:	☒	☒	☒
	Maximum size: (Text); Maximum size: (Images); Maximum size: (Audio); Maximum size: (Video); Maximum size: (other)	☐	☐	☒
Output modalities: *For each selected modality, please include maximum output size or write 'N/A' if not defined.*	☐ Text; ☐ Images; ☐ Audio; ☐ Video; If any other please specify:	☒	☒	☒
	Maximum size: (Text); Maximum size: (Images); Maximum size: (Audio); Maximum size: (Video); Maximum size: (other)	☐	☐	☒
Total model size:	The total number of parameters of the model, recorded with at least two significant figures, e.g 7.3*10^10 parameters.	☒	☐	☐
Select the range that the total number of parameters belongs to.	☐ 1-500M; ☐ 500M-5B; ☐ 5B-15B; ☐ 15B-50B; ☐ 50B-100B; ☐ 100B-500B; ☐ 500B-1T; ☐ >1T	☐	☒	☒

Methods of distribution and licenses

		AIO	NCAs	DPs
Distribution channels:	A list of every distribution channel (e.g. enterprise or subscription-based access through existing software suites or enterprise-specific solutions; public or subscription-based access through an API; public or proprietary access through integrated development environments, device-specific applications or firmware, open-source repositories) where the model can be accessed by external parties to the knowledge of the model provider. For each listed distribution channel, please include a link to information about how the model can be accessed where available and the level of model access (e.g. weights-level access, black-box access) via the channel.	☒	☒	☒
License:	A link to model license(s) (otherwise attach a copy to this document) or indicate that none exists.	☒	☒	☐
	The type or category of license(s) under which the model could be made available to downstream providers.	☐	☐	☒
	A list of additional assets (e.g. training data, data processing code, model training code, model inference code, model evaluation code), if any, that are made available with a description of how each can be accessed and what licenses, if any, relate to their use.	☒	☒	☐

Use

		AIO	NCAs	DPs
Acceptable Use Policy:	Provide a link to the acceptable use policy applicable (or attach a copy to this document) or indicate that none exists.	☒	☒	☒
Intended uses:	A description of either (i) the uses that are intended by the provider (e.g. productivity enhancement, translation, creative content generation, data analysis, data visualisation, programming assistance, scheduling, customer support, variety of natural language tasks, etc..) or (ii) the uses that are restricted and/or prohibited by the provider (beyond those prohibited by EU or international law, including Article 5 AI Act), in both cases as specified in the information supplied by the provider in the instructions for use, terms and conditions, promotional or sales materials and statements, as well as in the technical documentation. If specifying (i) or (ii) is incompatible with the nature of the license under which the model is provided, then 'N/A' can be entered. [Recommended 200 words]	☒	☒	☒
Type and nature of AI systems in which the general-purpose AI model can be integrated:	A list or description of either (i) the type and nature of AI systems into which the general-purpose AI model can be integrated or (ii) the type and nature of AI systems into which the general-purpose AI model should not be integrated. Examples may include autonomous systems, conversational assistants, decision support systems, creative AI systems, predictive systems, cybersecurity, surveillance, or human-AI collaboration. [Recommended up to 300 words]	☒	☒	☒
Technical means for model integration:	A general description of the technical means (e.g. instructions for use, infrastructure, tools) required for the general-purpose AI model to be integrated into AI systems. [Recommended 100 words]	☐	☐	☒
Required hardware:	A description of any hardware, including the version, required to use the model where applicable. If not applicable (e.g. model offered via an API), 'N/A' should be entered. [Recommended 100 words]	☐	☐	☒
Required software:	A description of any software, including the version, required to use the model where applicable. If not applicable, 'N/A' should be entered. [Recommended 100 words]	☐	☐	☒

Training process

		AIO	NCAs	DPs
Design specification of the training process:	A general description of the main steps or stages involved in the training process, including training methodologies and techniques, the key design choices, assumptions made and what the model is designed to optimise for. For example, “the model is initialized with randomly selected weights and optimised using gradient-based optimization via the Adam optimizer in two steps. First, the model is trained to predict the next word on a large pretraining corpus using the cross-entropy loss, passing over the data for a single epoch. Second, the model is post-trained on a dataset of human preferences for 10 epochs to align the model with human values and make it more useful in responding to user prompts“. [Recommended 200 words]	☒	☒	☐
Relevance of different parameters:	The relevance of different parameters, as applicable.	☒	☒	☐
Decision rationale:	A description of how and why key design choices were made in model training. [Recommended 200 words]	☒	☒	☐

Information on the data used for training, testing, and validation

AIO NCAs DPs

Training data type/modality:
Select all that apply.

☐ Text ☐ Images ☐ Audio ☐ Video If any other please specify:

Training data provenance:
Select all that apply.

☐ Web crawling
☐ Private data licensed by or on behalf of rights holders, or acquired from other third parties
☐ User data
☐ Publicly available datasets
☐ Data annotation or creation potentially through relationships with third parties
☐ Data collected through other means
☐ Synthetically generated data (when created directly by the provider or on behalf of the provider)
If any other please specify:

How data was obtained and selected:
A description of the methods used to obtain and select data, including methods and resources used to annotate data, and models and methods used to generate synthetic data where applicable. [Recommended 300 words]

Number of data points:
The size (in number of data points) of the training, testing, and validation data respectively, together with the definition of the unit of data points (e.g. tokens or documents, images, hours of video or frames,...), recorded with at least two significant figures (e.g. 1.5x10^13 tokens).

Scope and main characteristics:
A general description of the scope and main characteristics of the training data, such as domain (e.g. healthcare, science, law,...), geography (e.g. global, restricted to a certain region,...), language, modality coverage, where applicable. In the case that previously acquired data was used, a description of how the model provider acquired the rights to the data and which products and services were involved if the data corresponds to user data from products and services. [Recommended 200 words]

Data curation methodologies:
General description of the data processing involved in transforming the acquired data into training data for the model, e.g. cleaning (e.g. filtering out irrelevant content such as ads), normalisation (e.g. tokenizing), augmentation (e.g. back-translation). [Recommended 300 words]

Measures to detect unsuitability of data sources (harmful data):
A description of any methods implemented in data acquisition or processing, if any, to address illegal or harmful content in the training data, including, but not limited to, child sexual abuse material (CSAM) and non-consensual intimate imagery (NCII). [Recommended 300 words]

Measures to detect unsuitability of data sources (personal data):
A description of any methods implemented in data acquisition or processing, if any, to address the prevalence of personal data in the training data, where relevant and applicable. [Recommended 200 words]

Measures to detect identifiable biases:
A description of any methods implemented in data acquisition or processing, if any, to address the prevalence of identifiable biases in the training data. [Recommended 200 words]

Computational resources

AIO NCAs DPs

Training time:
A description of what period is being measured along with the duration in wall clock days (e.g. 9x10^1 days), recorded with at least one significant figure.

The duration in hardware days (e.g. 4*10^5 Nvidia A100 days and 2x10^5 Nvidia H100 days) for the period described above, recorded with at least one significant figure.

Amount of computation used for training:
Measured or estimated amount of computation used for training, reported in computational operations and recorded with at least two significant figures (e.g. 2.4x10^25 floating point operations).

Measurement methodology:
A description of the methodology used to measure or estimate the amount of computation used for training.

Energy consumption

AIO NCAs DPs

		AIO	NCAs	DPs
Amount of energy used for training:	Measured or estimated amount of energy used for training, reported in Megawatt-hours and recorded with at least two significant figures (e.g. 1.0x10^2 MWh).	☒	☒	☐
Measurement methodology:	A description of the methodology used to measure or estimate the amount of energy used for training. If the amount of energy used for training cannot be estimated due to the lack of critical information from a compute or hardware provider, the provider should disclose the type of information they lack. [Recommended 100 words]	☒	☒	☐
Benchmarked amount of computation used for inference:	Benchmarked amount of computation used for inference costs, reported in floating point operations, recorded with at least two significant figures (e.g. 5.1x10^17 floating point operations).	☒	☒	☐
Measurement methodology:	A description of a computational task (e.g. generating 100000 tokens) and the hardware (e.g. 64 Nvidia A100s) used to measure or estimate the amount of computation used for inference.	☒	☒	☐

Additional information to be provided by providers of general-purpose AI models with systemic risk

AIO NCAs DPs

		AIO	NCAs	DPs
Evaluation:	A detailed description of the evaluation strategies that are not already included in the Model Report, including evaluation criteria, metrics, evaluation results and the methodology used for the identification of limitations based on available public evaluation protocols and tools or otherwise of other evaluation methodologies.	☒	☒	☐
Adversarial testing:	Where applicable, a detailed description of the measures put in place for the purpose of conducting internal and/or external adversarial testing (e.g. red teaming) unless they are already included in the Model Report.	☒	☒	☐
Model adaptations:	Where applicable, a detailed description of the measures put in place for the purpose of conducting model adaptations, including alignment and fine-tuning, unless they are already included in the Model Report.	☒	☒	☐
System architecture:	Where applicable, a detailed description of the system architecture explaining how software components build or feed into each other and integrate into the overall processing.	☒	☒	☐

GENERATE FORM FOR DOWNSTREAM PROVIDERS

If this pdf document is opened in Acrobat Reader, clicking the button to the left will generate a pdf document containing only the subset of the information entered into this form that is aimed at providers who intend to integrate the general-purpose AI model into their AI systems.