ブリュッセル効果への対応 [第３回]
日本企業はEU-AI法にどう備えるべきか

2025年3月19日(水) 12:00-13:00

【プログラム】

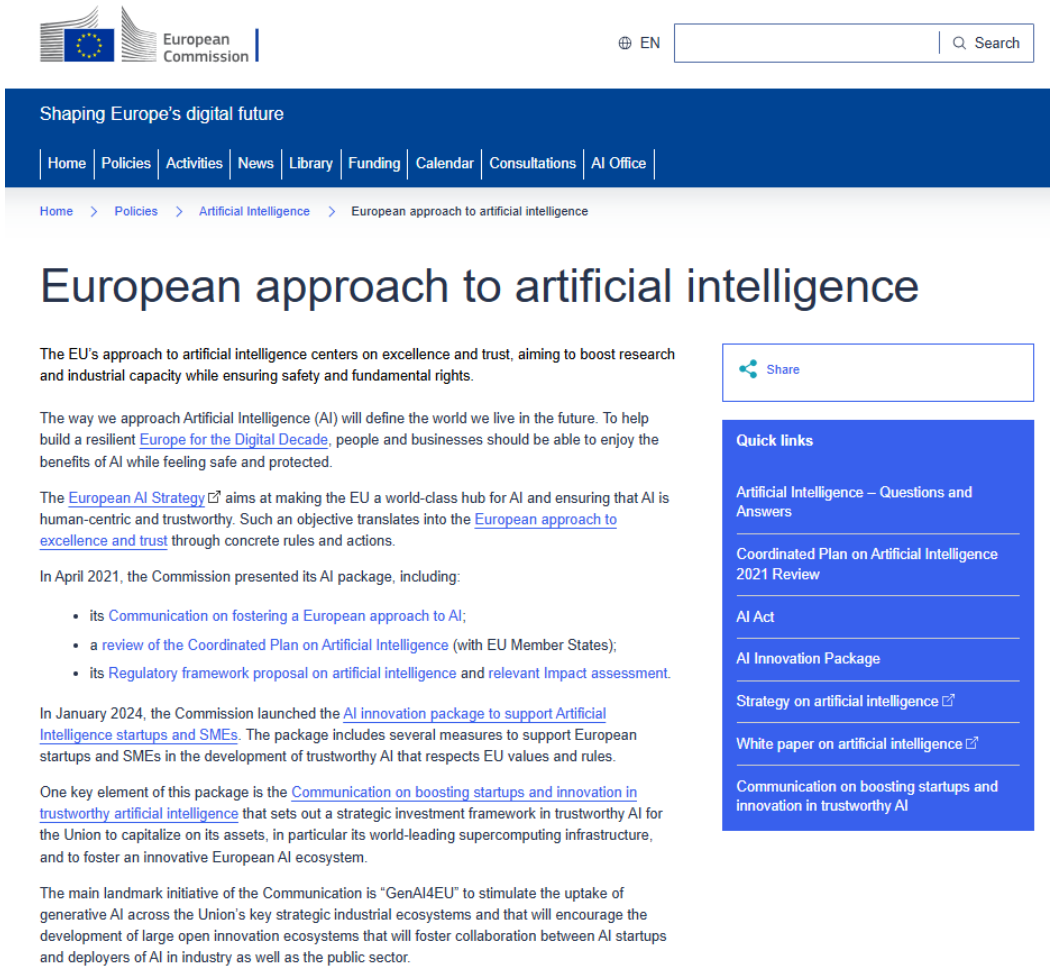
12:00-12:05	開会挨拶：飯田陽一（総務省）
12:05-12:30	CoP第三ドラフトの概要と日本企業が留意すべき対応の要点 ： 工藤郁子（大阪大学）、実積寿也（中央大学）、根本宗記（NTT）
12:30-12:55	パネルディスカッションとQ&A パネリスト： 工藤郁子（大阪大学） 実積寿也（中央大学） 根本宗記（NTT） 村上明子（日本AIセーフティ・インスティテュート） 司会：江間有沙（東京大学東京カレッジ）
12:55-13:00	閉会挨拶：村上明子

UTokyo | TOKYO COLLEGE

東京大学未来ビジョン研究センター
Institute for Future Initiatives
The University of Tokyo

AI Center UTokyo

EU-AI法の今後を追いかけるために



European Commission

EN

Search

Shaping Europe's digital future

Home | Policies | Activities | News | Library | Funding | Calendar | Consultations | AI Office

Home > Policies > Artificial Intelligence > European approach to artificial intelligence

European approach to artificial intelligence

The EU's approach to artificial intelligence centers on excellence and trust, aiming to boost research and industrial capacity while ensuring safety and fundamental rights.

The way we approach Artificial Intelligence (AI) will define the world we live in the future. To help build a resilient Europe for the Digital Decade, people and businesses should be able to enjoy the benefits of AI while feeling safe and protected.

The European AI Strategy aims at making the EU a world-class hub for AI and ensuring that AI is human-centric and trustworthy. Such an objective translates into the European approach to excellence and trust through concrete rules and actions.

In April 2021, the Commission presented its AI package, including:

- its Communication on fostering a European approach to AI;
- a review of the Coordinated Plan on Artificial Intelligence (with EU Member States);
- its Regulatory framework proposal on artificial intelligence and relevant Impact assessment.

In January 2024, the Commission launched the AI innovation package to support Artificial Intelligence startups and SMEs. The package includes several measures to support European startups and SMEs in the development of trustworthy AI that respects EU values and rules.

One key element of this package is the Communication on boosting startups and innovation in trustworthy artificial intelligence that sets out a strategic investment framework in trustworthy AI for the Union to capitalize on its assets, in particular its world-leading supercomputing infrastructure, and to foster an innovative European AI ecosystem.

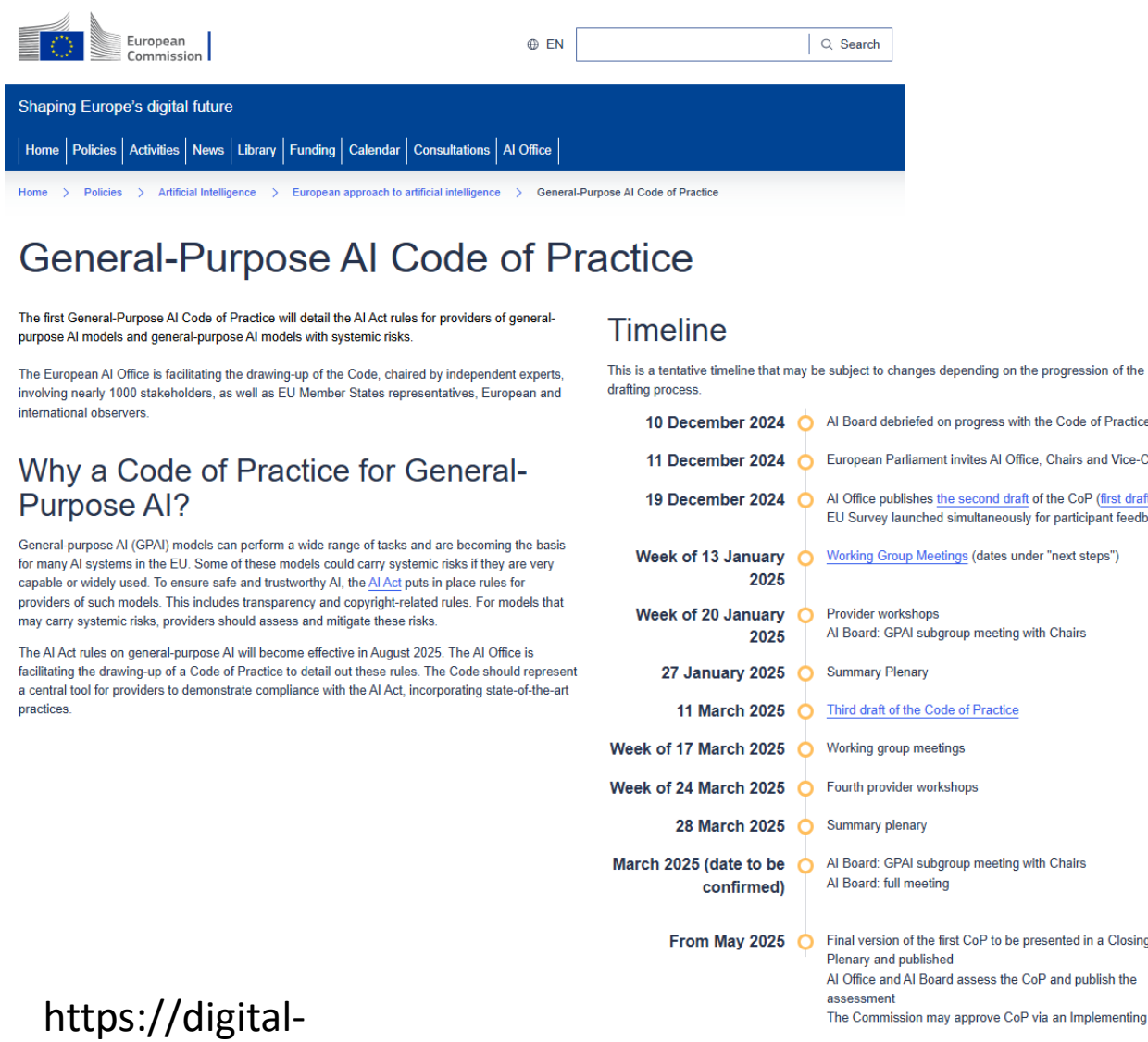
The main landmark initiative of the Communication is "GenAI4EU" to stimulate the uptake of generative AI across the Union's key strategic industrial ecosystems and that will encourage the development of large open innovation ecosystems that will foster collaboration between AI startups and deployers of AI in industry as well as the public sector.

Share

Quick links

- Artificial Intelligence – Questions and Answers
- Coordinated Plan on Artificial Intelligence 2021 Review
- AI Act
- AI Innovation Package
- Strategy on artificial intelligence
- White paper on artificial intelligence
- Communication on boosting startups and innovation in trustworthy AI

https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence



European Commission

EN

Search

Shaping Europe's digital future

Home | Policies | Activities | News | Library | Funding | Calendar | Consultations | AI Office

Home > Policies > Artificial Intelligence > European approach to artificial intelligence > General-Purpose AI Code of Practice

General-Purpose AI Code of Practice

The first General-Purpose AI Code of Practice will detail the AI Act rules for providers of general-purpose AI models and general-purpose AI models with systemic risks.

The European AI Office is facilitating the drawing-up of the Code, chaired by independent experts, involving nearly 1000 stakeholders, as well as EU Member States representatives, European and international observers.

Why a Code of Practice for General-Purpose AI?

General-purpose AI (GPAI) models can perform a wide range of tasks and are becoming the basis for many AI systems in the EU. Some of these models could carry systemic risks if they are very capable or widely used. To ensure safe and trustworthy AI, the AI Act puts in place rules for providers of such models. This includes transparency and copyright-related rules. For models that may carry systemic risks, providers should assess and mitigate these risks.

The AI Act rules on general-purpose AI will become effective in August 2025. The AI Office is facilitating the drawing-up of a Code of Practice to detail out these rules. The Code should represent a central tool for providers to demonstrate compliance with the AI Act, incorporating state-of-the-art practices.

Timeline

This is a tentative timeline that may be subject to changes depending on the progression of the drafting process.

10 December 2024	AI Board debriefed on progress with the Code of Practice (CoP)
11 December 2024	European Parliament invites AI Office, Chairs and Vice-Chairs
19 December 2024	AI Office publishes the second draft of the CoP (first draft) EU Survey launched simultaneously for participant feedback
Week of 13 January 2025	Working Group Meetings (dates under "next steps")
Week of 20 January 2025	Provider workshops AI Board: GPAI subgroup meeting with Chairs
27 January 2025	Summary Plenary
11 March 2025	Third draft of the Code of Practice
Week of 17 March 2025	Working group meetings
Week of 24 March 2025	Fourth provider workshops
28 March 2025	Summary plenary
March 2025 (date to be confirmed)	AI Board: GPAI subgroup meeting with Chairs AI Board: full meeting
From May 2025	Final version of the first CoP to be presented in a Closing Plenary and published AI Office and AI Board assess the CoP and publish the assessment The Commission may approve CoP via an Implementing Act

https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice

大阪大学

ブリュッセル効果への対応：日本企業はEU-AI法にどう備えるべきか3

EU AI法とCoP第3草案、そして規制の簡素化？

大阪大学 社会技術共創研究センター 特任准教授
工藤郁子
2025年3月19日

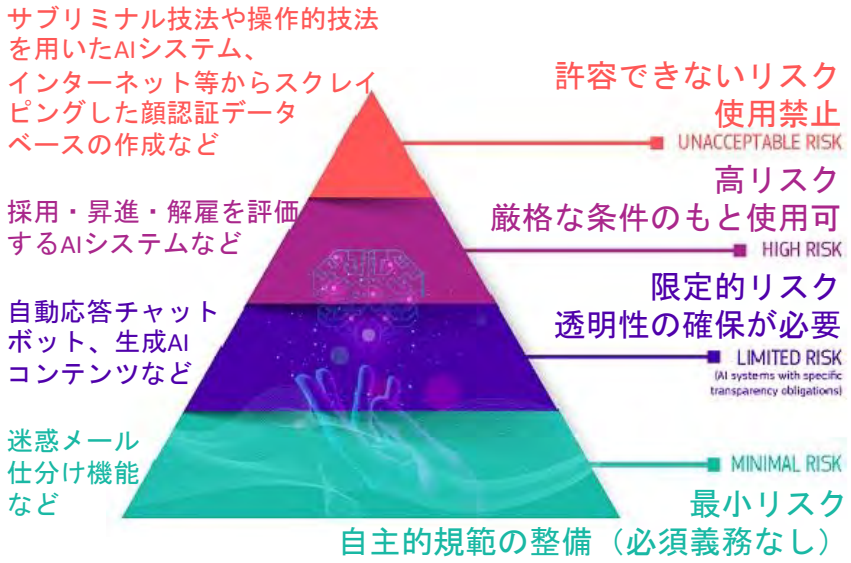
大阪大学 社会技術共創研究センター
Research Center on Ethical, Legal and Social Issues

Osaka University
Research Center on
Ethical, Legal and
Social Issues

EU AI法

本セミナー第2回
報告資料より再掲

- ハードロー：世界で初めて包括的にAIを規制する法律
- リスクベースアプローチ：リスクの大きさに合わせた対応（右図参照）
  - 「リスク」とは、害が発生する確率とその害の重大性の組み合わせ(3条2号)
- 域外適用：日本企業も対象になりうる
  - EU域内で、AIシステムを市場投入、サービス提供している、または汎用目的AIモデルを市場投入しているプロバイダー（EUに所在しているかは問わない）（2条1項(a)）
  - アウトプットがEU域内で利用される場合、第三国に所在するAIシステムのプロバイダー及びデプロイヤー（2条1項(c)）
- 違反には多額の罰金（99～101条）

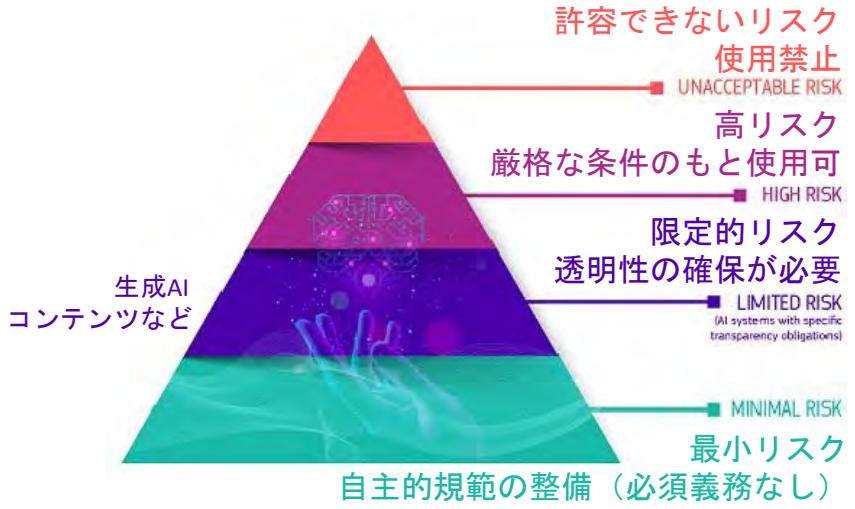


EUのウェブサイトの資料をもとに発表者作成
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

EU AI法：汎用目的AIモデル

本セミナー第2回
報告資料より再掲

- 4つのリスク類型に加えて、汎用目的AI (GPAI: general-purpose AI)モデルも別枠で規制
  - 生成AIのモデル等（not 生成AIが出力したコンテンツ）
  - 経緯：2021年4月の法案公開後、 2022年11月以降のChat GPT等の普及を受けて軌道修正
- 汎用目的AIモデル：顕著な汎用性を示し、多様で独立したタスクを適切に実行する能力を持ち、様々な下流のシステムやアプリケーションに統合可能なAIモデル（3条63号）
  - 「大量のデータを用いて学習されたAIモデルを含む」（が、それに限られないと解釈可能な文言）
  - 市場投入前の研究開発やプロトタイピング活動のために使用されるAIモデルは除外



汎用目的AIモデル
GPAI: general-purpose AI

EUのウェブサイトの資料をもとに発表者作成
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

EU AI法：汎用目的AIモデル

本セミナー第2回
報告資料より再掲

- 特定の条件を満たす汎用目的型AIモデルは、「システミックリスク」があると見なされ義務が加重
  - システミックリスクは、例えば、化学兵器や生物兵器の開発障壁の低下、自律型汎用AIモデルに対する制御不能、有害な差別や偽情報の大規模化などを想定（前文110）
  - システミックリスクの有無は適切な技術ツールや手法で判断されるとされ、
    （1）学習の合計計算量が10の25乗FLOPSの場合は有と推定されることに加え、
    （2）EU当局による指定もある
  - EU当局は、システミックリスクがある汎用目的AIモデルは、現時点では少数の企業によって開発されているものの、時間の経過とともに変化する可能性があるとしている

システミックリスクがない場合の汎用目的AIモデルに関する義務（53条）	システミックリスクがある場合の汎用目的AIモデルに関する義務（55条）
• 当局への情報提供（技術文書） • 下流プロバイダーへの情報提供 • EU著作権法遵守ポリシーの策定 など	• 53条の義務に加えて • 標準化されたプロトコルによるモデル評価 • システミックリスクのアセスメントとリスク低減措置実施 • インシデント発生時における当局への報告 • サイバーセキュリティに関する保護措置の実施 など

EU AI法：汎用目的AIモデル

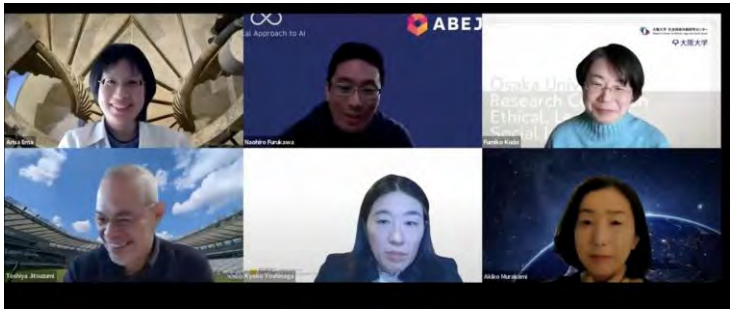
- 日本企業にも影響
- 直接：既存の汎用目的AIモデルを調整・ファインチューニングした主体が、新しいモデルのプロバイダーとして位置付けられる可能性あり（EU AI OfficeによるFAQ*参照）
  - CoPを起草する有識者チームは、ファインチューニングによって、新たに許容できないリスクが生じる可能性がないのであれば、義務の対象外とすべきと提言
  - EU AI Office が本問題に関するガイドラインを近日中に公開することを期待
- 間接：（基盤モデルでなくても）汎用目的AIの開発受託をする場合、取引先から、開示義務の対象となる情報の提供や文書作成について協力を要請されることも想定
- 未来：「大規模で回復不能な損害の差し迫った脅威を防ぐため、またはその悪影響を軽減するために、より短い期間で発行・採用できるよう、CoPの緊急更新を可能にする（CoPの定期的な見直しやCoP実施サポートとしての解釈表明以外の）第3のメカニズムが必要」との提言（CoP 第3草案 Appendix 2より）

* https://digital-strategy.ec.europa.eu/en/faqs/general-purpose-ai-models-ai-act-questions-answers

ブリュッセル効果

本セミナー第1回
より

- 古川先生：日本企業に影響がないというわけではないことに注意が必要だが、GDPRに比べて、ブリュッセル効果は大きくないだろう
- 吉永先生：AIスタートアップなどは、結局は自分たちが開発しやすいところに移動していく。GDPRに比べると、ブリュッセル効果は大きくないのではないか
- 実積先生：AI法が世界各国で、最低限の規制や政府調達基準の策定において参照されることで、影響を広げうる
- 工藤： 内閣府「AI制度研究会」構成員の一人として、日本における制度のあり方を検討する中で、国際的な相互運用可能性を考える上でEUの動向に関心を持っている



https://www.tc.u-tokyo.ac.jp/ai1ec_event/13586/

EU AI法とCoP

本セミナー第2回
報告資料より再掲

- 汎用目的AIモデルについて「行動規範（CoP: Codes of Practice）」と呼ばれるガイドラインを策定（56条）
  - 法解釈を具体化するもので法的拘束力はないが、CoPへの準拠は~~「法令遵守の推定」として機能する~~AI法への準拠を実証する手段となるが、AI法への適合の推定は提供されない
  - なお、ハイリスクAIなどを対象とする「CoC: Code of Conduct」（95条）とは別
  - さらにちなみに、G7日本議長国の下で始動した広島AIプロセスで策定されたのは「Code of Conduct」
- CoPは、3ラウンドのマルチステークホルダー協議を経て、2025年5月までに策定予定
  - EU AI Office は、行動規範の策定に参画するマルチステークホルダーを、 2024年8月25日までEU内外から公募（現在は受付終了）　→日本からの参加者はごく少数
  - 2024年9月30日、マルチステークホルダーが集まるキックオフ総会に、汎用目的AIモデルプロバイダー、下流プロバイダー、産業界、市民社会、学界、独立した専門家など約1,000人が参加
  - 4つのワーキンググループ（WG）で議論しながら、草案をブラッシュアップ中

Code Of Practice Plenary

WG 1
Transparency and copyright-related rules

Nuria Oliver
(Spain)
WG 1 Co-Chair
Transparency

Alexander Peukert
(Germany)
WG 1 Co-Chair
Copyright

- Vice Chair (Transparency): Rishi Bommasani (US)
- Vice Chair (Copyright): Céline Castets-Renard (France)

WG 2
Risk identification and assessment for systemic risk

Matthias Samwald
(Austria)
WG 2 Chair
Risk assessment

- Vice-Chair: Marta Ziosi (Italy)
- Vice-Chair: Alexander Zacherl (Germany)

WG 3
Technical risk mitigation for systemic risk

Yoshua Bengio
(Canada)
WG 3 Chair
Technical risk mitigation

- Vice-Chair: Daniel Privitera (Italy and Germany)
- Vice Chair: Nitarshan Rajkumar (Canada)

WG 4
Governance risk mitigation for systemic risk

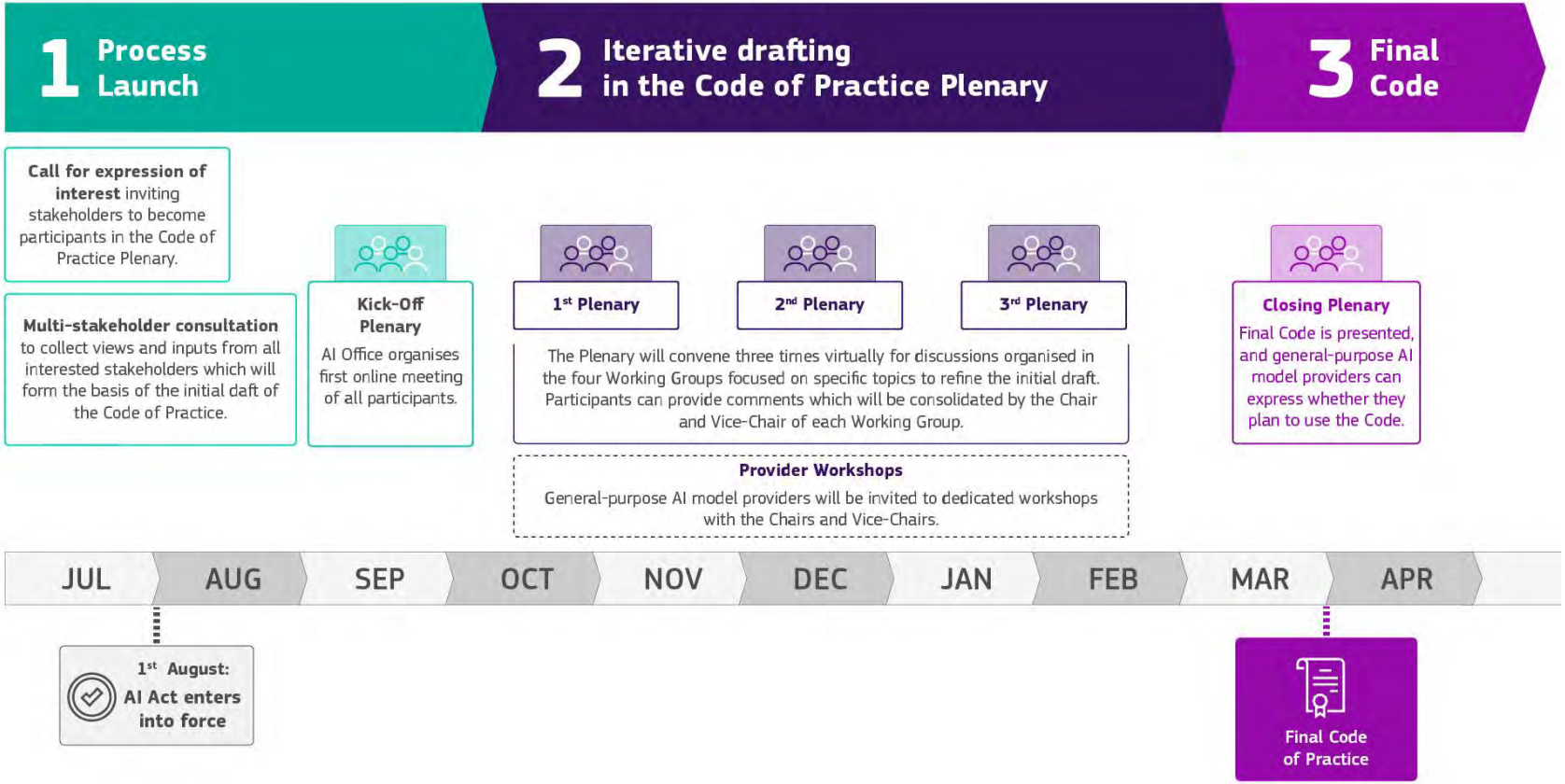
Marietje Schaake
(Netherlands)
WG 4 Chair
Internal Governance

- Vice Chair: Markus Anderljung (Sweden)
- Vice Chair: Anka Reuel (Germany)

（via https://digital-strategy.ec.europa.eu/en/news/meet-chairs-leading-development-first-general-purpose-ai-code-practice ）

EUROPEAN ARTIFICIAL INTELLIGENCE OFFICE

TIMELINE OF THE CODE OF PRACTICE DRAFTING PROCESS



(via https://digital-strategy.ec.europa.eu/en/news/ai-act-participate-drawing-first-general-purpose-ai-code-practice)

CoP第1草案フィードバック提出時の雑感

本セミナー第1回
報告資料より再掲

- まだざっくりした案で全体的に詰め切れていない
  - Measure（対策）の記述密度が低い項目も多く、コメントしづらい
  - 例えば、 重大インシデントは主にアプリケーションレイヤーで生じると思うが、下流プロバイダー等の報告義務がどうなっているか、現案からは不明瞭など
  - 逆に言うと、域外適用を確保したいニーズなどがある場合、その提案ができる余地が大きい
- EU AI法の枠内であるはずだが、上乗せ／横出し規制になりかねない表記ぶり？
  - 例えば、条文上、情報開示の範囲は知的財産や営業秘密を考慮することになっているが（ 53条(1)(b) ）、「Measure 2. Documentation for downstream providers」のみを読むと、そのような考慮がされているか若干怪しい
- （解釈上の疑義が明瞭になれば法学研究者としては満足だが）日本企業等のニーズがわからないと、日本の研究開発やビジネスにとって有益なフィードバックにならない
  - 削るべきところを削れるチャンスがあるが、自分だけの知見では活かしきれない

CoP第2草案フィードバック提出時の雑感

本セミナー第2回
報告資料より再掲

- KPIが多数追加されるなど、記述密度が向上した
  - 例えば「Commitment 18. 内部告発の保護」の記述が前回比でかなり厚くなった
- フィードバックが（意外と？）ちゃんと反映された
  - 例えば「コーポレート・ガバナンスの実践を考慮すべき。特に経営責任と執行責任を分離した記述にするべき」と指摘したところが反映された（Commitment 14. システミックリスクの責任分担）。 デジタルサービス法との用語統一も図られた（KPI 14.1.）
  - 知的財産や営業秘密とのバランスに配慮した記述も、各所でだいぶ増えた
- やはり上乗せ／横出し規制になりかねないところだった
  - 「（EU当局に開示するモデル報告書を）一般に公表するのは過度な負担であり、AI法の要件をこえているとのフィードバックを受けて、緩和する形で改訂した」との説明（Commitment 21. Public transparency）
- あまり説明なく前回と比べて義務が重くなっているように見える箇所も散見された

CoP進捗

2024年12月11日　本セミナー第1回
（CoP第1草案の解説）

2025年1月15日　本セミナー第2回
（CoP第2草案の解説）

2025年3月19日　本セミナー第3回
（CoP第3草案の解説）

Timeline

This is a tentative timeline that may be subject to changes depending on the progression of the drafting process.

10 December 2024	AI Board debriefed on progress with the Code of Practice (CoP)
11 December 2024	European Parliament invites AI Office, Chairs and Vice-Chairs
19 December 2024	AI Office publishes the second draft of the CoP (first draft) EU Survey launched simultaneously for participant feedback
Week of 13 January 2025	Working Group Meetings (dates under "next steps")
Week of 20 January 2025	Provider workshops AI Board: GPAI subgroup meeting with Chairs
27 January 2025	Summary Plenary
11 March 2025	Third draft of the Code of Practice
Week of 17 March 2025	Working group meetings
Week of 24 March 2025	Fourth provider workshops
28 March 2025	Summary plenary
March 2025 (date to be confirmed)	AI Board: GPAI subgroup meeting with Chairs AI Board: full meeting
From May 2025	Final version of the first CoP to be presented in a Closing Plenary and published AI Office and AI Board assess the CoP and publish the assessment The Commission may approve CoP via an Implementing Act

via https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice#timeline

CoP進捗

2024年12月11日　本セミナー第1回
（CoP第1草案の解説）

2025年1月15日　本セミナー第2回
（CoP第2草案の解説）

WG1 model template 意見募集

2025年3月19日　本セミナー第3回
（CoP第3草案の解説）

Timeline

This is a tentative timeline that may be subject to changes depending on the progression of the drafting process.

10 December 2024	AI Board debriefed on progress with the Code of Practice (CoP)
11 December 2024	European Parliament invites AI Office, Chairs and Vice-Chairs
19 December 2024	AI Office publishes the second draft of the CoP (first draft) EU Survey launched simultaneously for participant feedback
Week of 13 January 2025	Working Group Meetings (dates under "next steps")
Week of 20 January 2025	Provider workshops AI Board: GPAI subgroup meeting with Chairs
27 January 2025	Summary Plenary
11 March 2025	Third draft of the Code of Practice
Week of 17 March 2025	Working group meetings
Week of 24 March 2025	Fourth provider workshops
28 March 2025	Summary plenary
March 2025 (date to be confirmed)	AI Board: GPAI subgroup meeting with Chairs AI Board: full meeting
From May 2025	Final version of the first CoP to be presented in a Closing Plenary and published AI Office and AI Board assess the CoP and publish the assessment The Commission may approve CoP via an Implementing Act

via https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice#timeline

CoP進捗

2024年12月11日　本セミナー第1回
（CoP第1草案の解説）

2025年1月15日　本セミナー第2回
（CoP第2草案の解説）

WG1 model template 意見募集

CoP第3草案の公開遅延の告知（当初予定は2/17～週）

2025年3月19日　本セミナー第3回
（CoP第3草案の解説）

Timeline

This is a tentative timeline that may be subject to changes depending on the progression of the drafting process.

10 December 2024	AI Board debriefed on progress with the Code of Practice (CoP)
11 December 2024	European Parliament invites AI Office, Chairs and Vice-Chairs
19 December 2024	AI Office publishes the second draft of the CoP (first draft) EU Survey launched simultaneously for participant feedback
Week of 13 January 2025	Working Group Meetings (dates under "next steps")
Week of 20 January 2025	Provider workshops AI Board: GPAI subgroup meeting with Chairs
27 January 2025	Summary Plenary
11 March 2025	Third draft of the Code of Practice
Week of 17 March 2025	Working group meetings
Week of 24 March 2025	Fourth provider workshops
28 March 2025	Summary plenary
March 2025 (date to be confirmed)	AI Board: GPAI subgroup meeting with Chairs AI Board: full meeting
From May 2025	Final version of the first CoP to be presented in a Closing Plenary and published AI Office and AI Board assess the CoP and publish the assessment The Commission may approve CoP via an Implementing Act

via https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice#timeline

CoP進捗

2024年12月11日　本セミナー第1回
（CoP第1草案の解説）

2025年1月15日　本セミナー第2回
（CoP第2草案の解説）

WG1 model template 意見募集

CoP第3草案の公開遅延の告知（当初予定は2/17～週）

3週間遅れでCoP第3草案の公開（意見提出は3/30まで）

2025年3月19日　本セミナー第3回
（CoP第3草案の解説）

Timeline

This is a tentative timeline that may be subject to changes depending on the progression of the drafting process.

10 December 2024	AI Board debriefed on progress with the Code of Practice (CoP)
11 December 2024	European Parliament invites AI Office, Chairs and Vice-Chairs
19 December 2024	AI Office publishes the second draft of the CoP (first draft) EU Survey launched simultaneously for participant feedback
Week of 13 January 2025	Working Group Meetings (dates under "next steps")
Week of 20 January 2025	Provider workshops AI Board: GPAI subgroup meeting with Chairs
27 January 2025	Summary Plenary
11 March 2025	Third draft of the Code of Practice
Week of 17 March 2025	Working group meetings
Week of 24 March 2025	Fourth provider workshops
28 March 2025	Summary plenary
March 2025 (date to be confirmed)	AI Board: GPAI subgroup meeting with Chairs AI Board: full meeting
From May 2025	Final version of the first CoP to be presented in a Closing Plenary and published AI Office and AI Board assess the CoP and publish the assessment The Commission may approve CoP via an Implementing Act

via https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice#timeline

EUで何が起きている？

- 国際競争力の強化のためのAI規制の見直し？
- 「ドラギレポート」の規制の簡素化（simplify EU rules）提言
  - フォンデアライエン欧州委員長も受入れの姿勢
  - 2025年2月、AI民事責任指令（AI liability directive）廃案
- 2025年2月「AI Action Summit」での徴候
  - マクロン仏大統領がインタビューで規制簡素化を主張
  - ヴァンス米副大統領のAI規制批判
  - 英米の共同声明署名拒否
  - 英AISIの名称変更（Safety → Security）
- とはいえ、AI法は発効済みで大きな方針変更は（でき）ない

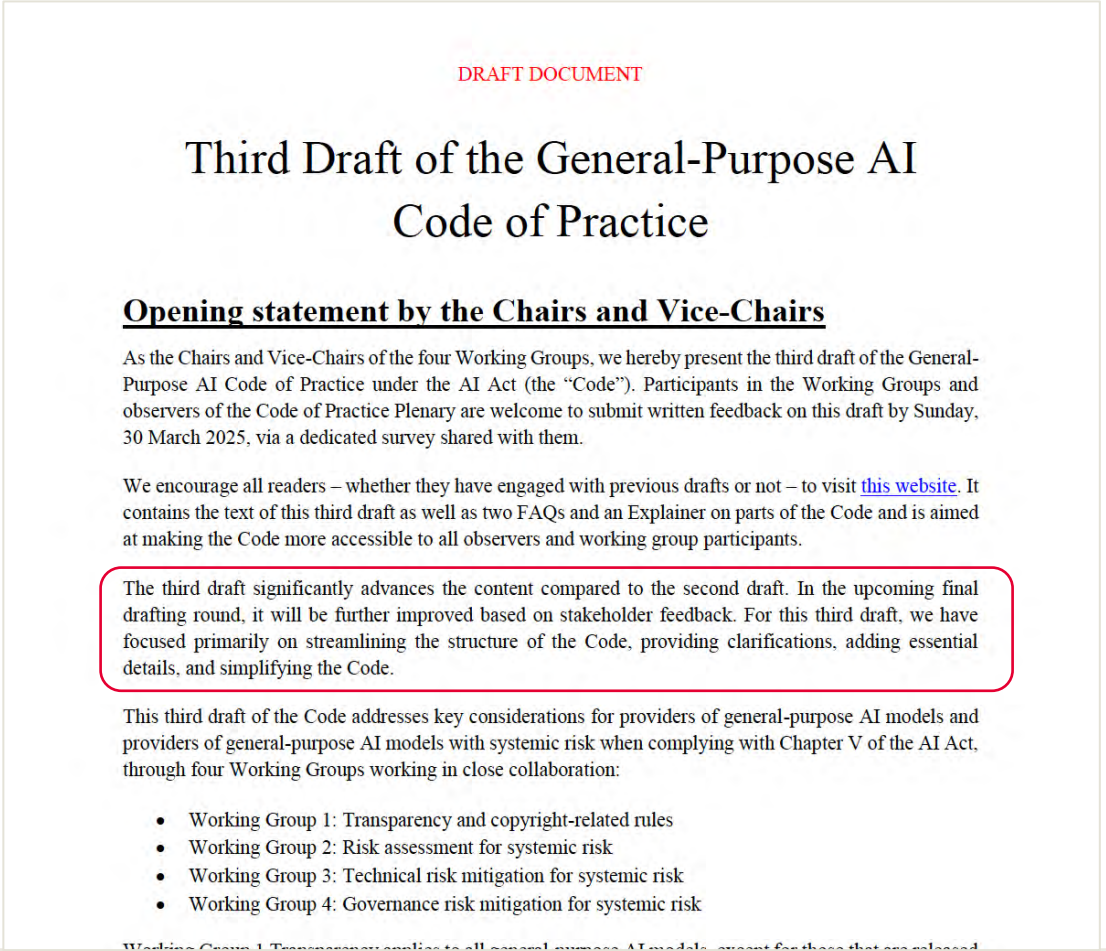
via
https://news.ntv.co.jp/category/international/54c53a4ba3004af2aa1b1a3edfd0e565
https://www.cnn.co.jp/world/35229256.html

CoP第3草案

〈仮訳〉

第3草案は、第2草案と比べて内容が大幅に前進しています。今後の最終草案作成ラウンドでは、関係者のフィードバックに基づいてさらに改善される予定です。

第3草案では、主としてCoPの構造を合理化し、明確化を図り、必要不可欠な詳細を追加し、**CoPを簡素化すること（simplifying the Code）**に重点を置きました。



DRAFT DOCUMENT

Third Draft of the General-Purpose AI Code of Practice

Opening statement by the Chairs and Vice-Chairs

As the Chairs and Vice-Chairs of the four Working Groups, we hereby present the third draft of the General-Purpose AI Code of Practice under the AI Act (the "Code"). Participants in the Working Groups and observers of the Code of Practice Plenary are welcome to submit written feedback on this draft by Sunday, 30 March 2025, via a dedicated survey shared with them.

We encourage all readers – whether they have engaged with previous drafts or not – to visit this website. It contains the text of this third draft as well as two FAQs and an Explainer on parts of the Code and is aimed at making the Code more accessible to all observers and working group participants.

The third draft significantly advances the content compared to the second draft. In the upcoming final drafting round, it will be further improved based on stakeholder feedback. For this third draft, we have focused primarily on streamlining the structure of the Code, providing clarifications, adding essential details, and simplifying the Code.

This third draft of the Code addresses key considerations for providers of general-purpose AI models and providers of general-purpose AI models with systemic risk when complying with Chapter V of the AI Act, through four Working Groups working in close collaboration:

- Working Group 1: Transparency and copyright-related rules
- Working Group 2: Risk assessment for systemic risk
- Working Group 3: Technical risk mitigation for systemic risk
- Working Group 4: Governance risk mitigation for systemic risk

CoP第3草案

- 21のコミットメント（第2草案）→18のコミットメント（第3草案）
  - taxonomyは用語集（glossary）として巻末付録へ。WG2～WG4の検討内容を「Safety and Security」として統合・再編
- KPIsは削除
  - 「CoP最終採択版にKPIが含まれることを期待すべきではない」とのこと
- （PDFに加えて）interactive website も公開
  - Website 改善のために Github でプルリクもできる
- 透明性に関する文書作成のためのテンプレート（Model Documentation Form）を提供
- システミックリスクをある程度明確化、新規参入企業への支援としての側面を強調
  - 4種類のリスクを提示：サイバー攻撃への利用、生物兵器等の開発障壁低下、心理的操作の高度化、制御喪失
  - 現時点で「システミックリスクがある汎用目的AIモデルのプロバイダー（GPAISR）は 5～15 社程度であると想定」しており、それらの大企業は既に対策済みであるところが多いとの認識。他方、「新規参入（newcomer）企業もAI法におけるGPAISR義務の対象となる可能性があるので、そのコンプライアンスを容易にするように設計した」との記載
- 巻末付録で、EU AI Office に対して2年ごとの定期的なCoP見直しを推奨

interactive website

via https://code-of-practice.ai/?section=summary

CoP第3草案：18のコミットメント

1. 汎用目的AIモデルのプロバイダー全体（システミックリスクがない場合も含む）

Transparency Section

- Commitment 1.1. 透明性に関する文書作成

Copyright Section

- Commitment 1.2. 著作権法遵守ポリシー

2. システミックリスクがある汎用目的AIモデルのプロバイダー（GPAISR）

Safety and Security Section

- Commitment 2.1. 安全性・セキュリティのフレームワーク
- Commitment 2.2. モデル開発中を含むモデルライフサイクル全体にわたるシステミックリスクのアセスメントと低減
- Commitment 2.3. システミックリスクの特定
- Commitment 2.4. システミックリスクの分析
- Commitment 2.5. システミックリスクの許容可能性の判断
- Commitment 2.6. 安全性対策
- Commitment 2.7. セキュリティ対策
- Commitment 2.8. 安全性とセキュリティのモデルレポート
- Commitment 2.9. 適切性（adequacy）アセスメント
- Commitment 2.10. システミックリスクの責任分担
- Commitment 2.11. 独立性のある外部評価
- Commitment 2.12. 重大インシデントの報告
- Commitment 2.13 内部告発者に対する報復防止
- Commitment 2.14. 当局への通知
- Commitment 2.15. 記録文書作成
- Commitment 2.16. 公衆への透明性

〈※ 全て仮訳です〉

CoP第3草案： 4部構成のSafety and Security

- Commitment 2.1. 安全性・セキュリティのフレームワーク

→　フレームワークを策定する

- Commitment 2.2. モデル開発中を含むモデルライフサイクル全体にわたるシステミックリスクのアセスメントと低減
- Commitment 2.3. システミックリスクの特定
- Commitment 2.4. システミックリスクの分析
- Commitment 2.5. システミックリスクの許容可能性の判断

→　策定したフレームワークに従ってリスクアセスメントをする

- Commitment 2.6. 安全性対策
- Commitment 2.7. セキュリティ対策

→　策定したフレームワークに従って技術的リスクを低減する対策を行う

- Commitment 2.8. 安全性とセキュリティのモデルレポート
- Commitment 2.9. 適切性（adequacy）アセスメント
- Commitment 2.10. システミックリスクの責任分担
- Commitment 2.11. 独立性のある外部評価
- Commitment 2.12. 重大インシデントの報告
- Commitment 2.13 内部告発者に対する報復防止
- Commitment 2.14. 当局への通知
- Commitment 2.15. 記録文書作成
- Commitment 2.16. 公衆への透明性

→　ガバナンスリスクを低減する対策を行い、フレームワークの遵守をサポートする仕組みを整える

〈※ 全て仮訳です〉

Thanks!

CoPが求める
情報開示・情報公開

中央大学　実積寿也

Prepared by T.Jitsuzumi for 3/19/2025.

EUAI法前文107

- GPAIモデルの事前学習および学習に使用されるデータについて透明性を高めるため、著作権法で保護されたテキストやデータを含む、GPAIモデルの提供者によるGPAIモデルの学習に使用される内容の十分詳細な要約の作成および公開が適切である。
- 営業秘密および企業秘密の保護の必要性に十分配慮しつつ、この要約は、例えば、モデルの訓練に使用された主なデータ集合またはセット（大規模な民間または公共のデータベースやデータアーカイブなど）を列挙し、使用されたその他のデータソースに関する説明を記載することなどにより、EU法に基づく権利を行使し、行使できるようにする正当な利害関係者（著作権者など）を支援するために、技術的に詳細なものではなく、一般的に包括的な範囲とすべきである。
- AIオフィスが要約用のテンプレートを提供することは適切であり、そのテンプレートは、シンプルかつ効果的で、かつ、プロバイダーが要求される要約を説明形式で提供できるものでなければならない。

Translated by DeepL

EUAI法53条 GPAIモデル提供者の義務

1.汎用AI（GPAI）モデルの提供者は、

a) そのモデルの訓練及び試験プロセス並びに評価結果を含む当該モデルの技術文書を作成し、最新の状態に維持するものとし、その技術文書には、AI事務局及び国内の権限のある当局の要請に応じて提供することを目的として、少なくとも附属書XIに定める情報を記載するものとする。

d) AIオフィスが提供するテンプレートに従って、…一般に公開する。

2. 第1項（a）および（b）に定める義務は、AIモデルへのアクセス、利用、修正、および配布を許可する自由かつオープンソースのライセンスのもとでリリースされ、重み付け、モデルアーキテクチャに関する情報、およびモデル利用に関する情報を含むパラメータが一般に公開されているAIモデルのプロバイダーには適用されない。この例外は、システミックリスクを有するGPAIモデルには適用されない。

7. 本条に従って取得された情報または書類（企業秘密を含む）は、第78条に定める守秘義務に従って取り扱われるものとする。

システミックリスクを持たない
オープンソースモデルは対象外

企業秘密には配慮

Systemic riskという概念

AI法第3条

65)「システミックリスク」とは、GPAIモデルの広範な影響力に特有のリスクであり、その影響力により欧州連合市場に重大な影響を及ぼす、または、公衆衛生、安全、公共の安全、基本的人権、または社会全体に実際に悪影響を及ぼす、または合理的に予測可能な悪影響を及ぼすリスクを意味し、バリューチェーン全体に規模を拡大して伝播する可能性がある。

AI法第55条　システミックリスクを持つGPAIモデル提供者の義務

1. 第53条および第54条に列挙された義務に加え、システミックリスクを有するGPAIモデルの提供者は、
   b) システミックリスクを有するGPAIモデルの開発、市場への投入、または使用に起因する可能性のある、その発生源を含む、欧州連合レベルでのシステミックリスクの可能性を評価し、緩和する。

CoPが定めるコミットメント

I. Commitments by providers of GPAI models

- Transparency section
  - Commitment 1.1. Documentation
- Copyright section
  - Commitment 1.2. Copyright policy

II. Commitments by providers of GPAI models with systemic risk

- Safety and security section
  - Commitment 2.1. Safety and Security Framework
  - ～
  - Commitment 2.14. Notifications
  - Commitment 2.15. Documentation
  - Commitment 2.16. Public transparency

CoP 3rd Draft

コミットメント 1.1. 文書化

- 署名者は、AI法第53条第1項（a）および（b）の義務を履行するため、「措置1.1.1に従ってモデル文書を作成し、最新の状態に維持すること」、「措置1.1.2に従って、GPAIモデルをAIシステムに統合しようとするAIシステム提供者に、また求めに応じてAI事務局に、関連情報を提供すること」、「措置1.1.3に従って、文書化された情報の品質、セキュリティ、および完全性を確保すること」にコミットする。これらの措置は、AI法第53条第2項に規定する条件を満たすオープンソースAIモデルの提供者には適用されないが、そのモデルがシステムリスクを有するGPAIモデルである場合はこの限りではない。

透明性は、AI法第53条第2項に定める条件を満たす無償かつオープンソースのライセンスの下で公開され、かつ、システミック・リスクを伴わないGPAIモデルに分類されるものを除き、すべてのGPAIに適用される。

Translated by DeepL

CoP 3rd Draft

措置1.1.1. モデル文書の作成と最新状態の維持

- 署名者は、GPAIモデルを市場に投入する際に、以下の「モデル文書」を作成することを約束する。
- 署名者は、比較可能で検証可能な文書化を可能にすることを目的として、...「計算リソースおよびエネルギー消費」のセクションで要求される情報を報告することを約束する。
- モデル文書に含まれる情報の関連する変更があった場合、署名者は、新しい情報を反映させるためにモデル文書を更新することに同意し、モデルが市場から撤退してから10年が経過するまでの期間、モデル文書の旧バージョンを保管する。

Translated by DeepL

CoP 3rd Draft

措置1.1.2. 関連情報の提供

- 署名者は、GPAIモデルを市場に投入する際、...AIオフィスおよび下流プロバイダーに連絡先を公開することを約束する。
- ...厳密に必要なモデル文書の1つ以上の要素についてAI事務局から要請があった場合、署名者は、...、最新のモデル文書に含まれる関連要素、またはその他の必要な追加情報を提供することに同意する。
- 署名者は、...最新のモデル文書に記載され、下流のプロバイダーを対象とした情報を提供することを約束する。...。
- 署名者は、上記のすべての行動を適時に実行することを約束する。
- 署名者は、文書化された情報の全部または一部を一般に開示し、透明性を高めることを検討することが推奨される。...

Translated by DeepL

CoP 3rd Draft

措置1.1.3. 情報の品質、完全性、セキュリティの確保

- 署名者は、文書化された情報が品質と完全性を確保され、AI法の義務遵守の証拠として保持され、意図しない改変から保護されることを確実にすることを約束する。情報の作成、更新、および品質とセキュリティの管理に関して、署名者は確立されたプロトコルと技術標準に従うことが推奨される。

Translated by DeepL

CoP 3rd Draft上の類似規定

システムリスクを有するGPAIモデルの提供者に対する義務

- Commitment 2.15. Documentation　[文書作成・保存義務]
  - Measure 2.15.1. Documentation regarding GPAISRs
- Commitment 2.16. Public transparency [情報公開義務]
  - Measure 2.16.1 Publication of Frameworks and Model Reports (or similar documents)

スケジュール

- 1月17日：AIオフィスによるテンプレート草案をWG1に提示
- 1月31日：フィードバック締切
- 3月12日：CoP第三次草案提示
- 3月30日：CoP第三次草案へのフィードバック締切
- 5月2日：テンプレートおよびGPAIガイドラインの採択期限
- 8月1日：GPAI規則の適用開始

イマココ

- AI法は2024年8月1日に施行され、規程の最終版は2025年5月2日までに準備する必要がある

AIオフィスのテンプレート作成方針

- 目的：正当な利害関係を有する当事者（権利保有者、データ対象者など）がEU法の下で権利を行使することを促進する
- 対象範囲：トレーニング前段階からfine tuningまでを対象
- 有効性：目的を達成するのに十分な詳細さ
- 簡潔性：技術的詳細さを求めず、理解容易で、法的専門知識は不要
- 企業秘密とのバランス：開示対象はデータのみ。企業秘密は不開示
- 負担の適正化
  - 説明文は要約形式
  - Fine-tuningの場合は追加学習データのみが公開対象
  - 中小企業に適正な負担
  - 最新情報への更新要求。小規模更新の場合は6か月以内毎に更新

1月17日提示のテンプレート草案

3. AI Office's approach to the template
Section 1 General information

1.1 Model and provider identification

- Provider's name and contact
- Authorized representative
- Model identifier
- Base model(s)

1.2. Date of placement on the market and knowledge cut off date

1.3. Overall training data size, modalities and characteristics

	Modalities	Overall size
	☐ Text	Number of tokens or bytes
	☐ Image	Number images (or pairs with other media)
	☐ Video	Number of minutes (or pairs with other media)
	☐ Audio	Number of minutes (or pairs with other media)
	☐ Other	____[please specify]

- Description of the **linguistic, regional, demographic and other relevant characteristics** of the overall training data:

_____

Text	Image	Video	Audio
☐ Fictional texts, literature	☐ Photography	☐ Movies, shows, performances	☐ Music
☐ Scientific and educative texts	☐ Paintings & fine-arts	☐ Animated video content	☐ Narrative and fiction (e.g. audiobooks)
☐ News, journalism and opinions	☐ Infographics	☐ Video game & immersive footage (e.g. 3D)	☐ Non-fiction educative audio content
☐ Legal and official documents	☐ Illustration & graphic design	☐ Documentaries	☐ Radio shows and podcasts
☐ Social communication (e.g.messages)	☐ Social / personal images	☐ Video news and journalism	☐ Social communication (phone calls, voice messages)
☐ Promotion, advertising, product and service reviews	**Special**	☐ User content, short videos	☐ Other (e.g. sounds and ambient)
☐ Other text	☐ Source code	☐ Other video content (e.g. experimental art, video effects)	
	☐ Structured data (e.g. calendar, maps)		
	☐ Other, describe:		

1月17日提示のテンプレート草案

3. AI Office's approach to the template
Section 2 List of data sources

2. List of Data Sources

2.1. Publicly accessible datasets:

- Overall size per modality and number of all datasets (number of synthetic datasets)
- List of 'main/large' datasets (above 5% of the overall data in this category) with unique identification, links + period of collection

2.2. Private non-publicly accessible datasets of third parties:

- Data licensed by rightholders or their representatives: Overall size per modality
- Datasets acquired from other third parties: Overall size per modality and number of datasets (number of synthetic datasets)
- List of 'main/large' private data sets acquired from other third parties (above 5% of the overall data in this category), unique identifiers and links (if available) and narrative description + period of collection

2.3. Data crawled and scraped from online sources:

- Overall size per modality, period of scraping
- Identification of crawlers, their purpose and behaviour;
- Explanation what content has been targeted;
- List of top 10 % of all internet domain names per type of data modality (e.g., text, image).
- For SMEs top 5 % or 1 000 internet domain names regardless of data modalities, whichever is lower, unless the model is with systemic risks.

2.4. User-sourced data (collected by provider incl. prompts):

- Overall size per modality
- List of providers' services/products

2.5. Self-sourced synthetic data(sets):

- Overall size per modality
- Name of AI model

2.6. Data acquired by the provider through other means:

- Overall size per modality
- Means of acquisition

1月17日提示のテンプレート草案

3. AI Office's approach to the template
Section 3 Relevant data processing

3. Other Relevant Data Processing Aspects

2.1. Respect of copyright and related rights

- Measures implemented to respect reservations of rights from the text and data-mining exception under Art.4(3) DSM Directive **during** data collection incl. specification of the opt-out protocols and solutions honoured by the provider
- Measures implemented **after** data collection is completed to identify and remove content for which rights have been reserved by the rightsholders

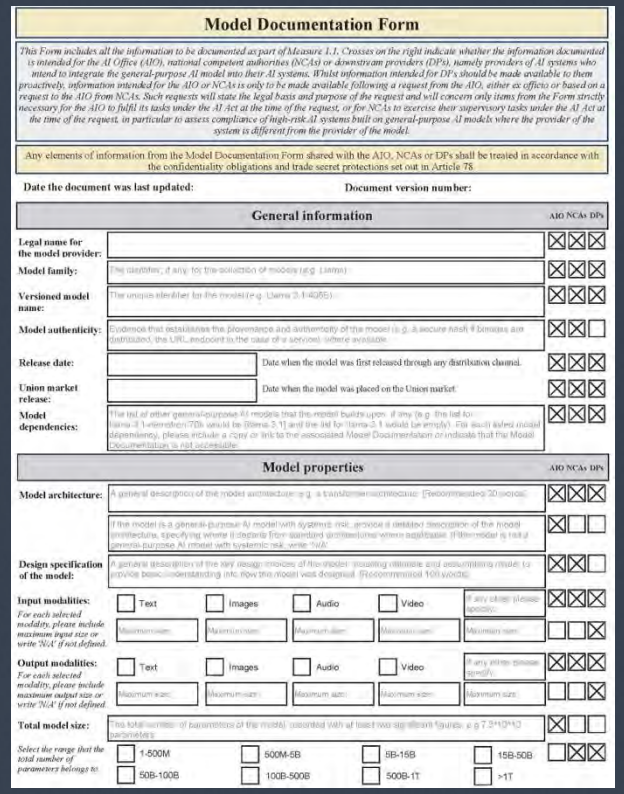
2.2. Removal of unwanted content

- Describe content deemed unwanted by the provider as part of the training data
- List the measures taken to avoid and/or remove such content (such as blacklists, keywords, and model-based classifiers)
- Measures applied by the curators of listed datasets may be mentioned, but do not need to be listed exhaustively

EU AI事務局提案への意見提出（2025/1/29）

- モデル情報と学習データ
  - 識別情報：下流事業者が上流の公開情報を複製・公表することを奨励
  - カットオフ日：基盤・ファインチューニングモデルとも開示必須。ただし中小企業向けに年次・半期ごとの開示を推奨
  - 学習データ：全件確認は困難なため、サンプル抽出による開示を容認
- データソース情報の開示
  - 開示対象：NDA保護データは免除
  - 公開データ：合成データの開示は元提供者が公開している場合に限定
  - スクレイピングデータ：全件確認は非現実的。5%基準の引き上げを検討
- データ処理（TDM(Text and Data Mining)オプトアウト・不要データ除去）
  - TDMオプトアウト：クロール期間と除外データ割合を開示
  - 不要データの除去：具体的な手法の開示は不要。
- バランスが重要：透明性確保と企業負担の軽減を両立する制度設計が必要

2025年3月12日付け再提案



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.

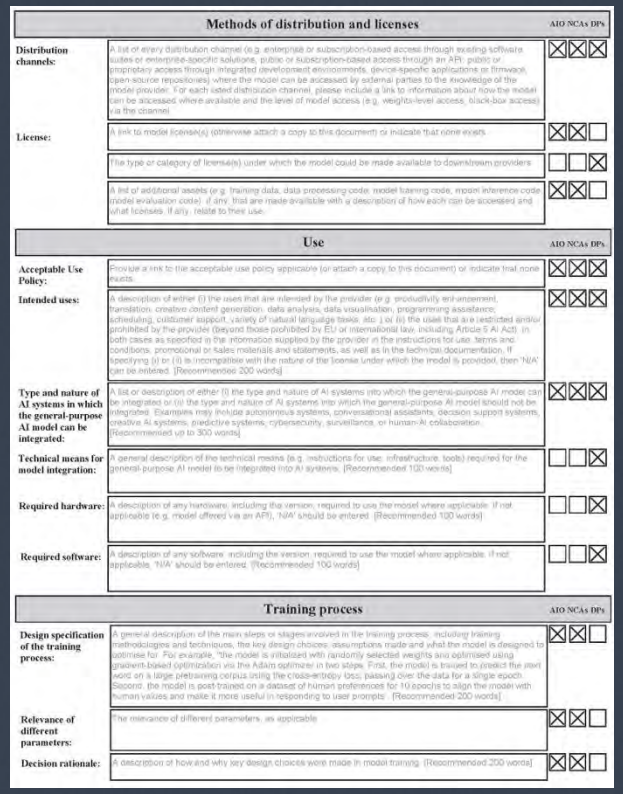
Date the document was last updated: Document version number:

General information AIO NCAs DPs

Legal name for the model provider:
Model family: The name(s), if any, for the series/collection of models (e.g. Llama)
Versioned model name: The unique identifier for the model (e.g. Llama 3.1405B)
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-variethen-70b would be Llama 3.1T and the list for Llama-3.1 would be empty). For such 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)
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)
Total model size: The total number of parameters of the model, rounded with at least two significant figures, e.g. 7.3*10^9 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

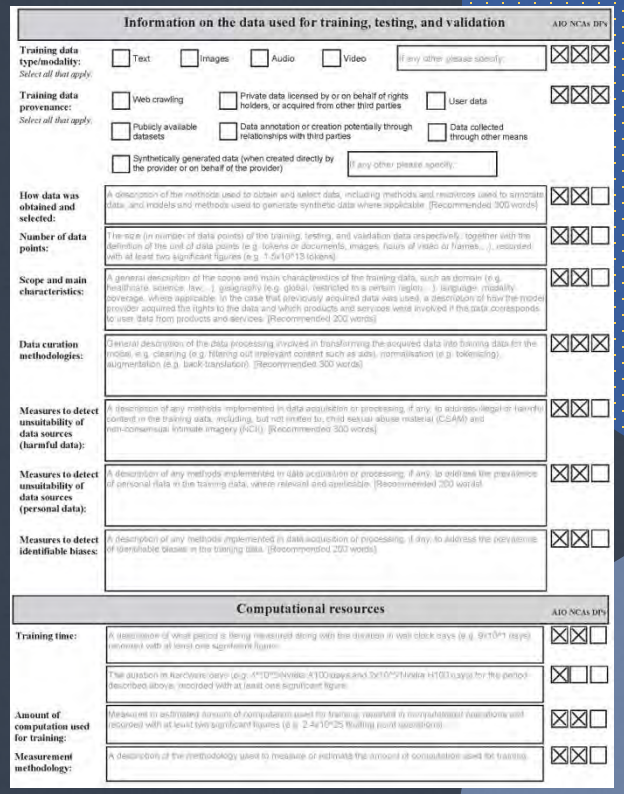
Distribution channels: A list of every distribution channel (e.g. enterprise or subscription-based access through existing software suites or enterprise-specific solutions, public or subscription-based access through an API, public or proprietary access through integrated development environments, device-specific 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 (ii) or (i) 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]

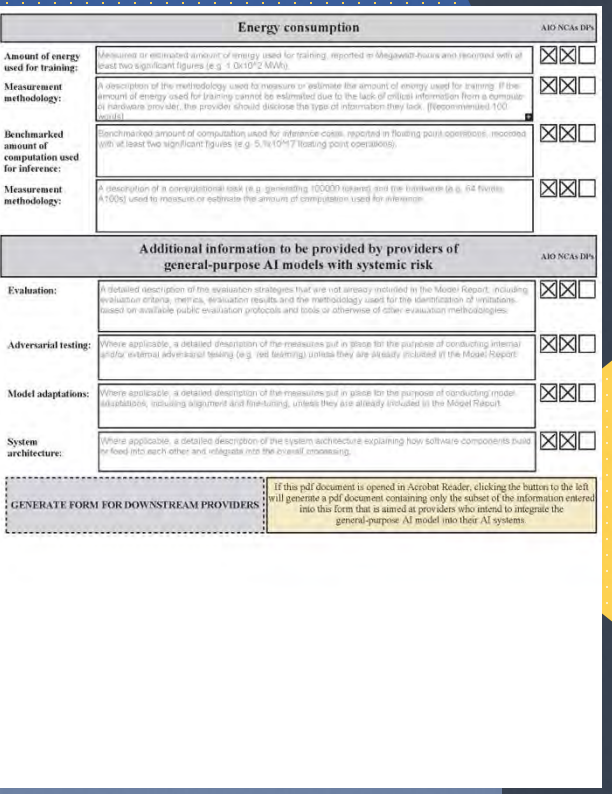


Information on the data used for training, testing, and validation AIO NCAs DPs

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...), rounded 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...), geographic (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), normalization (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) rounded 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, rounded 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 rounded 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

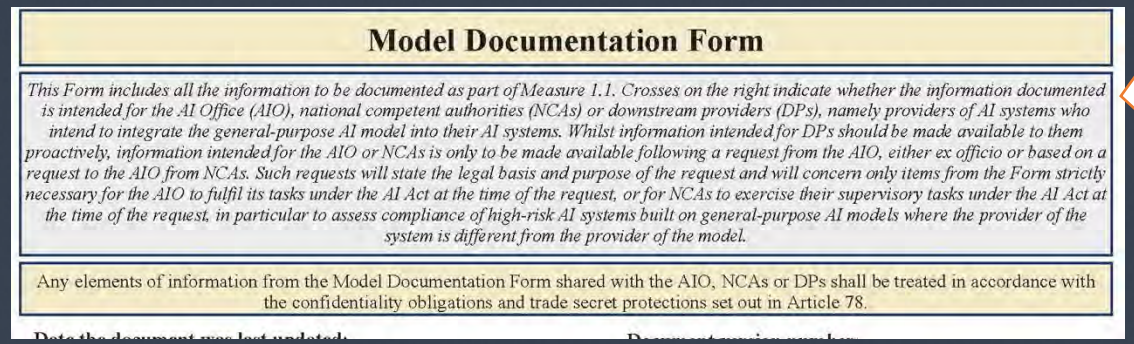
Amount of energy used for training: Measured or estimated amount of energy used for training, reported in Megawatt-hours and rounded with at least two significant figures (e.g. 1.0x10^2 MWh).
Measurement methodology: A description of the methodology used to measure or estimate the amount of energy used for training. If the amount of energy used for training cannot be estimated due to the lack of critical information from a computer 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 cases, reported in floating point operations, rounded 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

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.

- モデルのmodificationやfine tuningの場合、比例性を守るため、プロバイダーの義務は当該修正部分に限定される



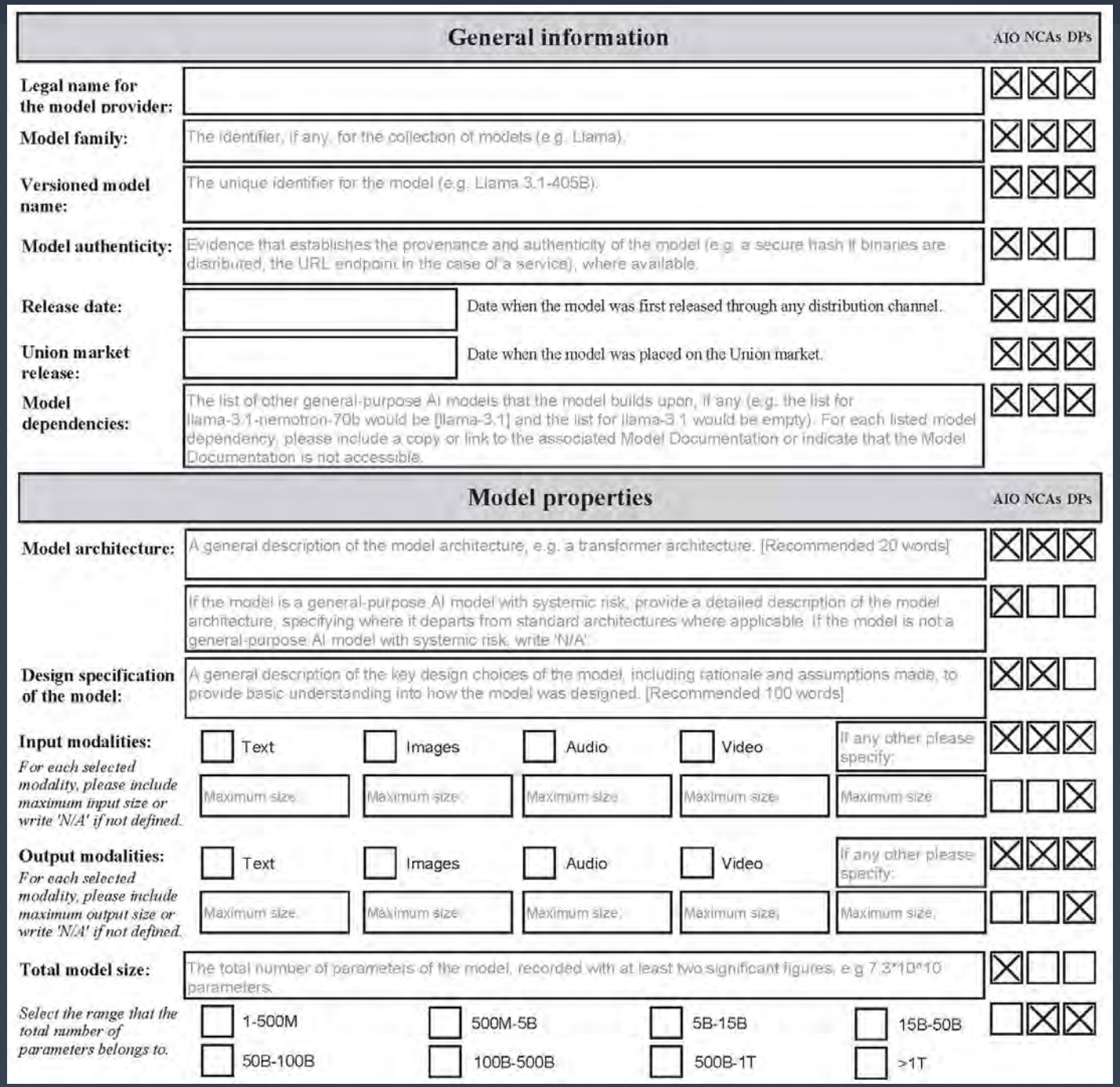
Model Documentation Form

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.

Translated by DeepL

本フォームには、措置1.1の一部として文書化されるべきすべての情報が含まれています。
右側のチェックマークは、文書化された情報がAIオフィス（AIO）、国家当局（NCAs）、または下流プロバイダー（DPs）、すなわち汎用AIモデルをAIシステムに統合しようとするAIシステムプロバイダーのいずれを対象としているかを示します。
DP向けの情報をDPに積極的に提供すべきであるのに対し、AIOまたはNCAs向けの情報を提供するのは、AIOが職権により、またはNCAsからAIOへの要請に基づいて要請があった場合のみである。
このような要請には、法的根拠および要請の目的が明記され、フォームの項目については、要請時点でAI法に基づく職務をAIOが遂行するために、または要請時点でAI法に基づく監督職務をNCAsが遂行するために、特に、汎用AIモデルを基に構築された高リスクAIシステムのコンプライアンス評価（システム提供者とモデル提供者が異なる場合）に厳密に必要なもののみが対象となる。



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 — Maximum size ☐ Images — Maximum size ☐ Audio — Maximum size ☐ Video — Maximum size If any other please specify: — Maximum size	☒	☒	☒
Output modalities: *For each selected modality, please include maximum output size or write 'N/A' if not defined.*	☐ Text — Maximum size ☐ Images — Maximum size ☐ Audio — Maximum size ☐ Video — Maximum size If any other please specify: — Maximum size	☒	☒	☒
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	☐	☒	☒

TOKYO COLLEGE

ブリュッセル効果への対応：日本企業はEU-AI法にどう備えるべきか3

NTTグループのEU-AI法対応

2025年3月19日

日本電信電話株式会社

技術企画部門 AIガバナンス室

根本 宗記

NTTグループのAIガバナンス

組織、体制

NTTグループは、日本中心のグローバル企業

国内 約19万人、グローバル 約15万人

CAIO(Chief AI Officer)を配置

適切なAI利用の推進とAIリスクへの対応を統制するため

AIガバナンス室を設置

AIリスクマネジメントの特徴

タイプ： アクセルとガードレール型（not ブレーキ）

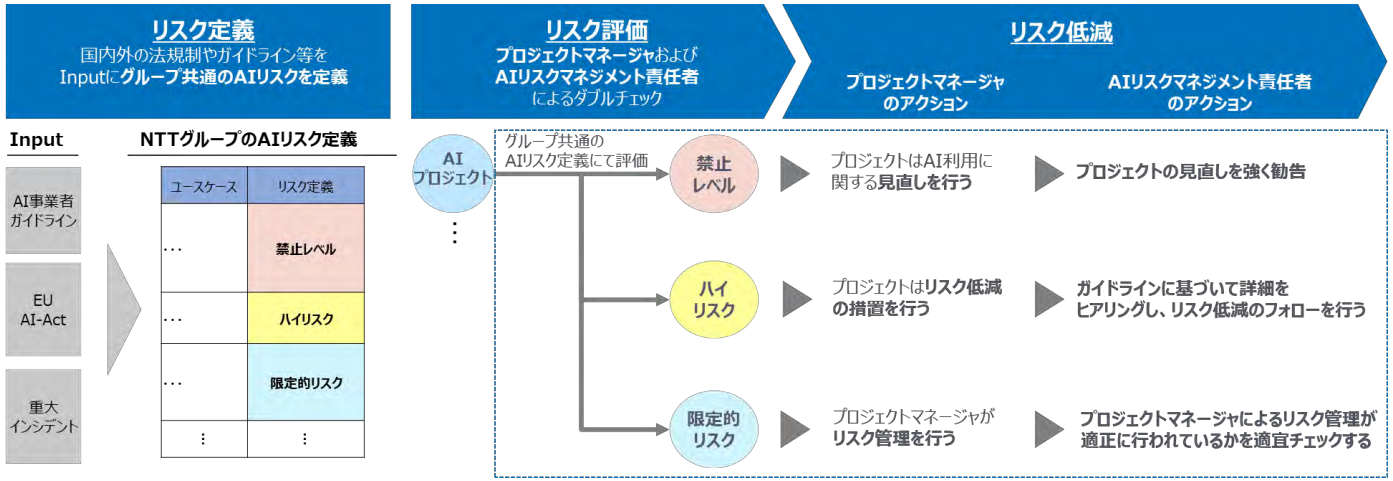
AI利用を進めるために如何にリスクを低減するか

定義： NTTグループでAIリスク定義を統一

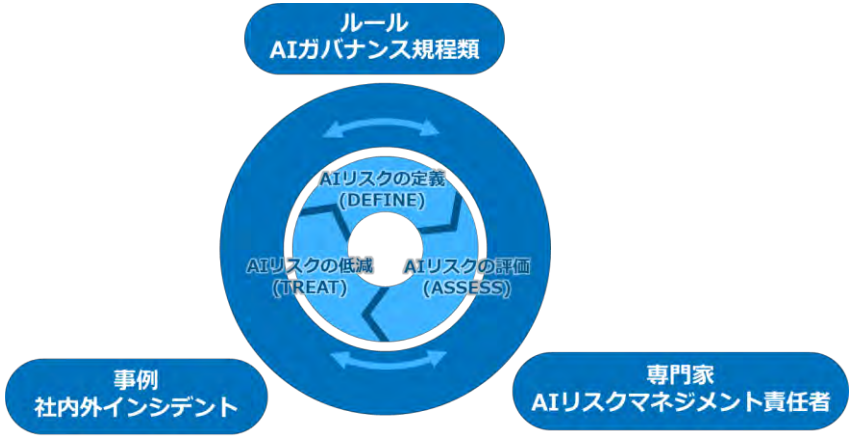
プロセス： リスクベースアプローチによる段階的評価

運用： ルールと事例と専門家による運用

リスクベースアプローチによる段階的評価



ルールと事例と専門家による運用



国際機関・政府との連携状況

関係先	主なトピック
G7/OECD	✓ 広島AIプロセス モニタリングメカニズム検討、パイロット報告対応 ✓ OECD HAIP ReportingFramework ローンチイベント参加 @パリ ✓ 広島AIプロセス フレンズグループ パートナーズコミュニティ イベント参加 @東京
日本政府	✓ AI制度研究会 に構成員参画、ヒアリング対象として日本のAI制度に関する意見表明
米国政府	✓ 2024/12 Center for Strategic and International Studies（CSIS） AIガバナンス イベントにてNTTグループメンバが多数登壇
EU政府	✓ EU-AI法 GeneralPurpose AI Code of Practice WGに個人の資格で参画

(1)EU-AI法への対応

NTTグループにおける対応状況

2025年8月2日施行に向けた対応

役割に応じた責務を確認し、適切な対応がとれるようAIガバナンス規程類を改訂予定

AIモデル開発者(ベースモデル): 条件を抽出し、個別にフォロー
AIモデル開発者(追加学習): 条件・責務をコメント
AIサービス提供者: 責務を確認し、必要な条件を抽出
AI利用者: 個別対応は難しい領域のため、慎重な確認が必要

⇒ 方向性は変わっていない。Third Draftを元に法務担当とも対応の具体化を検討中。階層の深いグループ会社における法務担当も含めたマネジメント方針が論点に

(2)Second Draft GeneralPurpose AI Code of Practiceへの主なコメント内容

コメント観点	コメント内容	Third Draftの記載内容
AIモデル開発者(追加学習)の責務	追加学習を行ったものがどのような責務を負うのか明確にすべき	First Draftからコメント実施 Third Draftでも記載は確認できず
AIサービス提供者、AI利用者の責務	対象者が広範囲に渡るため、責務を負うのであれば明確にすべき	記載は確認できず
国際整合性	国際合意に至った広島AIプロセス モニタリングメカニズムより踏み込んだ責務を負わせる箇所(AIモデル開発者に対する開示内容等)については、どのようなリスク低減につながるのか明確にしてもらいたい	記載は確認できず
中小企業、OSSに対する要件緩和の見直し	特定キャラクターの出力を目的としたOSS AIモデルなど、小規模・OSSであっても、権利侵害を拡散する懸念が強いものも多い。モデルの規模や、OSSであるかはリスクの度合いとの関連性は薄い。中小企業やOSSに対する要件緩和は見直すべき	著作権侵害について中小企業への緩和措置は削除になった模様。
デジタルコンテンツの権利保護の強化	海賊版サイトの学習禁止などのコンテンツ権利保護は、中小企業・OSSにおいてもしっかりと適用すべき	Copyright Measure Ⅰ.2.2.(b)で具体化。 海賊版サイト情報は慎重に取り扱う必要あり。

(3)AIリスクマネジメントにおけるポイント

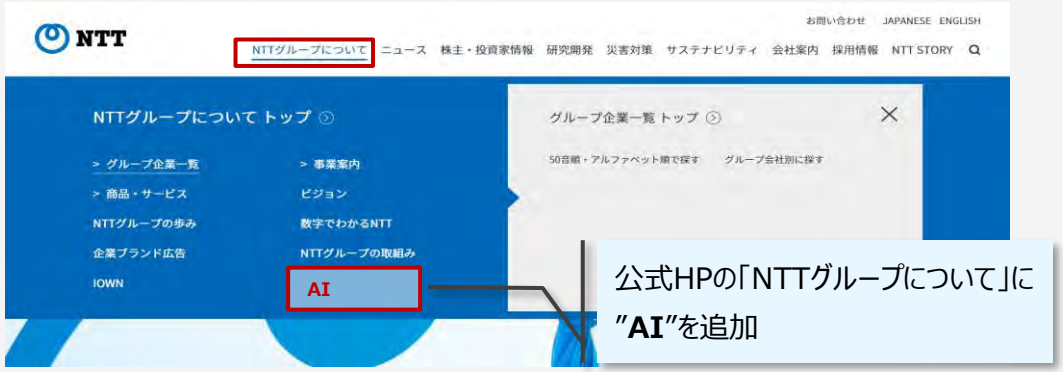
活動ポイント	内容
① 自社・グループにおけるリスク定義の整備	自社・グループにおいて、まず共通したリスク定義を行い、その内容に従い、運用の対応を整備・検討していくことが重要
② リスクベースアプローチでのリスク低減	リスクベースアプローチで、リスクが高いPJにしっかりとリスク低減対応を行う。2段階でリスク評価を行うプロセスは有効
③ 各ロールを意識した運用	プロセスやガイドラインの整備を行う際は、各ロール(AIのモデル開発者、サービス提供者、利用者)の役割を意識するのがよい

社外向けHPの公開について

■NTTの公式HP（https://group.ntt/jp/）にて、AIに関するページを2月に公開

AIページへの導線とページ概観

NTT公式HP TOP
（NTT / NTTグループ | 日本電信電話株式会社）



NTT公式HP AI
（https://group.ntt/jp/group/AI/）

Innovating a Sustainable Future for People and Planet