IFI Policy Recommendation

IFI Policy Recommendation No.37　January 2026

透明性を起点とした AI の適正性確保に向けた提言

工藤 郁子
大阪大学社会技術共創研究センター 特任准教授

実積 寿也
中央大学総合政策学部 教授

江間 有沙
東京大学東京カレッジ/未来ビジョン研究センター 准教授

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

透明性を起点としたAIの適正性確保に向けた提言

工藤 郁子（大阪大学）
実積 寿也（中央大学）
江間 有沙（東京大学）

エグゼクティブサマリー

透明性は、AIガバナンスにおける基盤的構想であり、その適正性を確保するために必要不可欠である。AI透明性を促進する仕組みとして、法規制型アプローチとインセンティブ型アプローチの2つがあるが、主要国でAIガバナンスに関する姿勢が異なり、法規制の国際的整合性を図ることが短期的には難しい現状を背景に、より柔軟なインセンティブ型アプローチが注目されている。そして、法規制の観点からは見落とされがちだが、機関投資家もAI透明性について重要な役割を担っている。2023年以降にAI透明性に関する株主提案が増加したことはその例である。

インセンティブ型アプローチの観点からAI透明性に関する施策について整理すると、ビジネスユーザー向けの限定開示と規制当局向けの限定開示だけでなく、機関投資家などが取得可能な一般公開情報も重要であることがわかる。そのため、企業が自主的に一般公開している「透明性レポート」の取組みをさらに促進すべきである。

この点において、2023年の先進国首脳会議（G7）において日本が議長国として立ち上げた国際枠組み「広島AIプロセス（HAIP: Hiroshima AI Process）」における報告枠組み（Reporting Framework）は注目に値する。

こうした検討結果を踏まえて、日本政府に対して、以下の7点を提言する。

1. **AI透明性に関する既存法令の解釈明確化と「法の欠缺」の調査推進**
   AIの透明性に関する多くの問題には現行法で一定の対応が可能だが、AI固有の状況への解釈が不明確な場合がある。そこで、解釈適用の明確化を提案する。また、現行法では対応しきれない新たなリスクにより法制度的空白も生じ得る。そのため、AI推進法に基づく調査を推進し、制度整備を行うことを提案する。

2. **政府調査の適正手続きに関する指針策定**
   日本政府がAI事業者などに調査協力を求める際、法の支配や適正手続きといった基本原則を守り、恣意的な権限行使を避けるべきである。そこで、ガバメントアクセスに関する国際的なガイドラインなどを参照しながら、政府自身の透明性を高める指針を策定することを提案する。

3. **AI透明性を進展させる経済的インセンティブの強化に向けた制度構築**
   AI透明性に関する事業者の取組みを促進すべく、経済的インセンティブを強化することを提案する。具体的には、ESG投資などの非財務評価にAI透明性について反映する仕組みの構築、「AIトラスト銘柄」の創設、公共調達における加点、消費者一般に対するAIリテラシー教育の支援などが考えられる。

4. **企業の自己宣言に対する履行確保メカニズムの導入に関する検討**
   企業が自主的に公開する透明性レポートなどを「社会への自己宣言」と位置づけ、その履行状況を第三者機関が確認するメカニズムの導入を提案する。虚偽や欺瞞があれば是正対象とすることで、透明性を装った不誠実な行為を防ぎつつ、企業の自由を尊重した実効的な説明責任を確保できる。

5. **広島AIプロセス報告枠組みの活用による相互運用性確保**
   国際的な整合性を確保するため、海外の主要なAI事業者が既に参加する広島AIプロセス報告枠組みを国内制度として活用することを提案する。これにより、国外事業者にも透明性確保を求めやすくなり、国内企業との不公平を防ぎながら、実効性を高められる。既存の質問票を日本語化することを基軸としつつ、日本の実情に合わせた補足質問を加えることも検討すべきである。

6. **国内利用者のアクセス性を高める透明性レポートの統合プラットフォーム整備**
   AI事業者が公開する透明性レポートを集約し、検索・閲覧できるプラットフォームを整備することを提案する。機関投資家や市民社会による分析が容易になり、HAIP不参加企業も含めた包摂的なエコシステムを構築できる。また、AI透明性に関する知見共有のためのコミュニティ形成を支援すべきである。さらに、国内外のベストプラクティスを周知・顕彰するイベントを開催することも提案する。

7. **AI透明性に関する国際的「クロスウォーク」と相互承認の推進**
   国際的な相互運用性を高めるべく、各国の透明性基準を比較し対応関係を整理する「クロスウォーク」を行うことを提案する。これにより、AI事業者が複数の基準に沿って対応しやすくなることが期待できる。さらに、HAIP報告枠組みでの透明性レポートを各国規制当局への報告として相互承認する仕組みを検討することも推奨する。

目次

1. はじめに：社会の信頼性のためのAI透明性　4
2. AI透明性に関する2つのアプローチ 4
　2-1. 法規制型とインセンティブ型 4
　2-2. 機関投資家というステークホルダー 5
3. AI透明性のための施策 5
　3-1. 開示と公開 5
　3-2. AI透明性の施策に関する3類型 6
4. 諸外国におけるAI透明性に関する制度 9
　4-1. 欧州 9
　4-2. 米国 10
　4-3. 中国 11
　4-4. 韓国 12
5. 日本におけるAI透明性に関する制度構築の方向性 12
　提言1　AI透明性に関する既存法令の解釈明確化と「法の欠缺」の調査推進 13
　提言2　政府調査の適正手続きに関する指針策定 14
　提言3　AI透明性を進展させる経済的インセンティブの強化に向けた制度構築 14
　提言4　企業の自己宣言に対する履行確保メカニズムの導入に関する検討 14
6. 「広島AIプロセス」報告枠組みの活用 15
　提言5　広島AIプロセス報告枠組みの活用による相互運用性確保 16
　提言6　国内利用者のアクセス性を高める透明性レポートの集約プラットフォーム整備 16
　提言7　AI透明性に関する国際的クロスウォークと相互承認の推進 17
7. おわりに：AI透明性が評価される社会に向けて 17
Appendix 1. 概念整理 18
Appendix 2. 汎用目的AIモデルに関する情報開示 19

1. はじめに：社会の信頼性のためのAI透明性

- AIは社会経済の基盤技術として急速に普及し、産業競争力の向上に寄与する一方、さまざまな課題を顕在化させている。社会に受容可能な形でAIを開発・利用するためには、安全性を確保して社会からの信頼を獲得する仕組みが不可欠である。
- その起点となるのが、透明性である。AIシステムに関する情報を適切な範囲で可視化し、外部から検証可能にすることで、説明責任やリスク管理を実現できる。そして、透明性は、法規制の適用に必要な前提情報の確保に資するだけでなく、企業の自主的取組を促す市場メカニズムとも結びつくという面でも重要である。
- 本提言では、透明性をAIガバナンスの基盤的な「プロトコル」として再定義する。そして、法的規制と市場原理を接続する多層的な制度設計について、ステークホルダーが協働するエコシステムの構築に向けたプロトタイプのひとつとして提示する。

2. AI透明性に関する2つのアプローチ

2-1. 法規制型とインセンティブ型

- AIに関する透明性を確保する仕組みとして、大きく分けて2つのアプローチがある。法規制型アプローチとインセンティブ型アプローチである。
- 法規制型アプローチは、情報開示を法令によって義務付け、AI開発者やAI提供者などにコンプライアンスを求め、違反した場合には高額の課徴金などの制裁を加える。
- インセンティブ型アプローチは、レピュテーションや社会的期待などによって実践を促す任意的な枠組みであり、市場インフラや情報流通の仕組み（プロトコル）として理解される[1]。
- もっとも、これら2つのアプローチは対立するものではなく、補完し合う関係にある[2]。また、AIのリスク水準に応じて使い分けることも可能である。つまり、比較的リスクが高いAIには法規制型アプローチを採用し、比較的リスクが低いAIにはインセンティブ型アプローチを採用することもできる[3]。
- AIシステムは、国境を越えて利用されることが多い。各国のAI関連法制が異なる中で、法規制型アプローチには国際的ハーモナイズや規制協力が課題となる。そのため、短期的な実効性や国際的な拡張性を重視する観点から、インセンティブ型アプローチへの関心が高まっている。

---

[1] O’Reilly, T. (2025). Disclosures. I do not think that word means what you think it means. *Asimov’s Addendum, Substack*. March 27, 2025.

[2] 金融商品取引法や会社法に基づく「法定開示」と、企業が自主的に公開する「任意開示」の関係などを想起すると理解しやすい。なお、両者の中間的な形態として、証券取引所が定めた有価証券上場規程に基づく開示（「適時開示」）という選択肢があることも参考になるだろう。

[3] EU AI法では「限定的なリスクのAIシステム」には透明性に関する義務を課しているが、「最小限のリスクのAIシステム」は行動規範の策定を奨励するに留めて、自主規制に委ねている。

2-2. 機関投資家というステークホルダー

- インセンティブ型アプローチに関する注目すべき動向として、2023年以降、AI透明性に関する株主提案が増加していることが挙げられる。
- 例えば、2024年にMicrosoft社に対して生成AIの偽・誤情報リスクの年次報告を求める株主提案があった[4]。同様に、Googleの親会社にあたるAlphabet社やFacebookを運営するMeta社に対しても偽・誤情報リスクの報告を求める提案が上程された。
- また、Netflix社に対しても、AIガバナンスに関する透明性レポートを公開するよう求める株主提案が行われた[5]。
- これらの株主提案はいずれも過半数には届かなかったため否決されたが、Netflixでは43%、Microsoftでは36%の賛成を集めており、比較的高い支持率となっている。
- こうした株主提案には、機関投資家や議決権行使助言会社など投資関連ステークホルダーが重要な役割を担っている[6]。そして今後、日本企業にも波及する可能性も指摘されている[7]。
- 法規制型アプローチでは見落とされがちだが、機関投資家なども、AI透明性について重要な役割を担っていることが、インセンティブ型アプローチの視点からは明確になる。
- 透明性の確保が、コンプライアンスだけでなく、資本市場を通じて評価される経営資源となれば、AI開発者などの説明責任に関するインセンティブが強化される。

3. AI透明性のための施策

3-1. 開示と公開

- AI透明性の向上のための施策には様々なものがある。まず基礎的なこととして、「情報開示（limited disclosure）」と「情報公開（publication）」の区別が挙げられる【表1参照】[8]。
- 透明性（transparency）というと、誰にでもアクセス可能になるという情報公開の印象が強いかもしれない。しかし近年では、特定の関係者に限定してアクセスを可能にする情報開示の方が主要な施策になりつつある。これは、営業秘密、知的財産権、セキュリティ、個人情報保護などに配慮する必要があるためである。

---

[4] Open MIC (2024). 2024 Campaign: Report on Generative Artificial Intelligence Misinformation and Disinformation Risks.

[5] Maiden, B. (2024). AI ethics proposal attracts support among Netflix shareholders. *Governance Intelligence.* June 25, 2024.

[6] Cavé, A. et al. (2024). Next-Gen Governance: AI's Role in Shareholder Proposals. *Harvard Law School Forum on Corporate Governance.* May 6, 2024.

[7] 今村桃子（2025）．「株主総会の争点にAIリスク　バークシャーで監督委提案、日本に波及も」日本経済新聞（2025年6月11日）．

[8] ここでは、金融商品取引法における「開示」というよりも、行政法における「開示」を参照した。情報公開制度における情報「公開」請求では不特定多数に情報提供される一方、個人情報保護制度における情報「開示」請求では本人のみが閲覧可能になる。そのため、「公開」と「開示」は使い分けられていると考えられる。

- 情報公開の方であると強調するために「public transparency」と記載するなど[9]、レトロニムのような表現がされることも散見される。

【表1】開示と公開

カテゴリー	内容
開示（limited disclosure）	規制当局やユーザーなど特定の関係者に限定して、ある情報をアクセス可能にすること
公開／公表（publication）	不特定多数の一般市民や社会全体を対象として、ある情報を広く一般にアクセス可能な状態にすること

3-2. AI透明性の施策に関する3類型

- AI透明性に関する施策は、情報提供先に注目すると、「当局開示（regulatory disclosure）」、「利用者開示（user disclosure）」、「一般公開（publication）」の3つに大別できる【表2参照】。あくまで例示であって、相互に排他的なものでも網羅的なものでもない。
- 法規制型アプローチで特に注目されるのは、当局開示と利用者開示である。他方、インセンティブ型アプローチの観点からは、市場からの評価の促進という意味で、一般公開も重要となる。

【表2】AI透明性に関する施策例

カテゴリー	提供先（例）	情報内容（例）	提供方法（例）
当局開示	規制当局	技術仕様、リスク管理体制、内部統制、インシデント情報など	安全性計画書、影響評価書、監査報告書、インシデント報告書など （※営業秘密やセキュリティとの関係で限定的なアクセス）
利用者開示	AI提供者・AI利用者	AIシステムの内部構造、動作原理、想定用途・禁止用途など	技術仕様書、リスク評価書、API仕様書、利用規約など （※営業秘密やセキュリティとの関係で限定的なアクセス）

[9] 例えば、EU AI法における汎用目的AIモデルの行動規範（Code of Practice）における措置10.2には「public transparency」という小見出しが付いている。

	被影響者 （※プロファイリングの対象者など）	AIシステムの出力に関する理由・根拠など	簡潔なFAQ、インタラクティブな説明用UI、人間による追加説明など （※個人情報保護との関係で限定的なアクセス）
一般公開	被影響者 （※他者が生成したコンテンツの閲覧者など）	AI生成コンテンツであることの明示など	ラベル、ウォーターマークなど
	データ権利者	自身のデータがAI学習に利用されているか否か、など	データカード、モデルカード、透明性レポート、サステナビリティ報告書など
	市民社会	AIシステムの環境負荷、安全対策状況など	
	機関投資家	経営資源とリスクなど	

(1) 当局開示

- 技術動向が制度整備の速度を上回る状況にあるため、ルールの形成と執行にあたって、ステークホルダーからの関連情報が必要となる。
- 同時に、営業秘密、知的財産権、セキュリティ、個人情報保護などに配慮する必要もあるため、開示範囲を適切に限定することが求められる。
- **規制当局**：法令や公的権限に基づき、AIシステムのルールの策定、監督、執行、評価を担う行政機関または独立規制機関。技術仕様、リスク管理の実態、苦情対応状況などの情報を求めており、AI開発者・AI提供者・AI利用者・市民社会などとの連携が重要となる。

(2) 利用者開示

- AIの製品やサービスが利用者のもとに届くまでには、複数の事業者が関わり合うバリューチェーン（価値連鎖）が形成されることが多い。そのため、上流から下流までの関係性を踏まえて、リスク管理をデザインする必要がある。
- この点において、エンドユーザーである個人利用者だけでなく、ビジネスユーザーである事業者間の情報開示もまた重要となる。同時に、営業秘密、知的財産権、セキュリティ、個人情報保護などに配慮する必要もあるため、開示範囲を適切に限定することが求められる[10]。
- **AI開発者**：AIシステムを開発する事業者（AIを研究開発する事業者を含む）を指す。バリューチェーン上流にいる者として、AI透明性について、第一次的な主体と

[10] バリューチェーン上のステークホルダーは、ライセンスや利用規約などを通じて、何らかの契約関係を結んでいることが想定される。

なる[11]。EU AI法においては「下流プロバイダー（downstream provider）」や「デプロイヤー（deployer）」などに対する透明性義務が課されている。

- **AI提供者**：AIシステムを、アプリケーション、製品、既存のシステム、ビジネスプロセス等に組み込んだサービスとして、AI利用者、場合によっては業務外利用者に提供する事業者を指す。AI開発者とAI利用者の中間に位置し、上流と下流をつなぐ主体となる[12]。
- **AI利用者**：事業活動において、AIシステムまたはAIサービスを利用する事業者を指す。AIシステムのビジネスユーザーも、一定の範囲で情報提供者となり得る点に留意する必要がある[13]。例えば、人事評価に関するAIサービスを利用した人事部門の担当者（AI利用者）は、評価対象になった一般従業員（被影響者）に対して、AIの出力結果に関する説明責任を果たすことが期待される。
- **被影響者**：AIシステム等を直接的には利用していないものの、その出力結果や運用によって意思決定・評価・サービス提供等の形で実質的な影響を受ける者を指す。AIによるプロファイリングや自動意思決定の対象となるケースなどで、AIシステムの出力に関する理由・根拠などの情報を必要とする【Appendix 1 参照】。

(3) 一般公開

- 利用者開示や当局開示だけでなく、一般公開情報を増やしてAI透明性を促進することも必要である。こうすることで、法規制によるモニタリングと市場からの評価という二重のメカニズムが働き、AIのガバナンスがより持続的に機能することが期待できる。
- AI開発者等にとっては、機関投資家に対する適切な情報公開によって資金調達や事業拡大の機会を得られる点がインセンティブとなる。また、被影響者、データ権利者、市民社会、学識経験者などとの対話によって、技術開発の方向性を社会的要請と整合させることができる。
- **被影響者**：AIシステム等を直接的には利用していないものの、他者が出力したAI生成コンテンツ（例：ディープフェイク）を閲覧してその内容により行動や判断が影響されるケースが想定される。そのため、AI生成コンテンツであることを明示したラベルやウォーターマーク（電子透かし）などを必要としている。
- **データ権利者**：AIの学習や運用に利用される個人情報のデータ主体やコンテンツの権利者を指す。AI学習等に用いられるデータは、明示的な許諾を得て提供される場合もあれば、無許諾で利用されている場合もある。例えば、生成AIの学習対象となる文章・画像・音声などは、著作権法の権利制限規定の対象となったり、フェアユース（公正利用）の範囲内となったりして、無許諾であっても適法に利用可能と解されている。よって、権利者は自身のデータがAIの学習に使用されているか否かを

[11] 総務省・経産省（2025）「AI事業者ガイドライン（第1.1版）第3部のAI開発者に関する事項では、「検証可能性の確保」、「関連するステークホルダーへの情報提供」、「AI提供者への共通の指針の対応状況の説明」などが記載されている。

[12] 前掲「AI事業者ガイドライン（第1.1版）」第4部のAI提供者に関する事項では、「関連するステークホルダーへの情報提供」や「AI利用者への共通の指針の対応状況の説明」が記載されている。

[13] 前掲「AI事業者ガイドライン（第1.1版）」第5部のAI利用者に関する事項において、「関連するステークホルダーへの情報提供」や「関連するステークホルダーへの説明」が記載されている。

確認することすら困難な場合も想定される。そのため、データ取得経路・利用目的・異議申立て手続きなどの情報を求めている。

- **市民社会・学識経験者**：AIの開発・提供・利用が公共の利益、人権の尊重、環境の持続可能性に適合しているかを監視し、社会的議論を形成するNGOなどを指す。必要な情報は多岐に渡り、モデル設計思想・学習データの出所・リスク管理などの技術的情報、法制度や監督体制に関する情報、差別やデータの不適切利用などに関する事例・証言などを必要としている。また、AIの開発・運用について環境負荷の高さが指摘されているため、モデル訓練・運用に伴う電力消費量やデータセンターの再エネ電力比率などの情報を求めている場合もある。
- **機関投資家**：AI関連の開発・提供・利用に関わる企業やプロジェクトに資金を提供する者を指す。近年は、AIの持続可能性や倫理的側面も投資判断の要素となっており、財務情報だけでなく、ガバナンスやリスク管理に関する非財務情報を求めている。具体的には、モデルの技術成熟度や性能評価、リスク管理体制（権利処理の状況なども含む）、外部監査や第三者評価の有無などである。

4. 諸外国におけるAI透明性に関する制度

4-1. 欧州

- 2024年に成立したEUのAI法は、前掲3カテゴリーのうち、利用者開示と当局開示の義務を重視している[14]。一般公開について、多くの場合に任意となっているが、高リスクAI、限定的リスク、汎用目的AIモデルについて、それぞれ一定の義務を課している。
- AI法以外によって規律されている部分もあり、EU法全体としてAI透明性を担保している。

(1) 高リスクAIシステム

- **当局開示**：規制当局への情報提供義務として、技術文書の作成・更新（11条）や重大インシデントの報告（73条）などが定められている。ただし、一般公開することを想定したものではない。
- **利用者開示**：高リスクAIシステムのプロバイダーは、デプロイヤーがそのシステムを適切に理解し、安全かつ責任ある方法で使用できるように、システムの目的、性

[14] 2025年12月に内閣府「AI時代の知的財産権検討会」事務局が「生成AIの適切な利活用等に向けた知的財産の保護及び透明性に関するプリンシプル・コード（仮称）（案）」を公表した。当該資料には「（AI推進法の趣旨を踏まえつつ）ＥＵ ＡＩ Ａｃｔにおける取組（透明性の確保のための措置や著作権保護のための措置）及びコーポレートガバナンスの分野におけるスチュワードシップ・コード等の取組（コンプライ・オア・エクスプレイン）を参考に」措置の原則を定めたとしている。もっとも、EU AI法では「限定的リスク」と「汎用目的AIモデル」というカテゴリーを設けて、それぞれのリスクに応じた規律を設けているが（いわゆる「リスクベース・アプローチ」）、内閣府のプリンシプル・コード（案）ではそのようなカテゴリーが設けられておらず、生成AI一般を対象にしているように見える。また、EU AI法では、「開示」と「公開」の違いを意識しながら、マルチステークホルダーが時間をかけて検討・調整しているが、内閣府のプリンシプル・コード（案）は、その概要開示対象事項を「利用者及び権利者を含めたすべての者が閲覧可能な状態にする」としていることから、本提言でいう「一般公開」を念頭においているようである。そのため、EU AI法とは実質的に異なるアプローチを採用しているように見える。

能、限界、出力の正確性や精度指標などを明示する義務を負っている（13条）。これらは、使用説明書や画面上の表示などの形で個別に通知されることが想定されており、一般公開する義務は課されていない。

- **一般公開**：高リスクAIシステムをEU市場で上市する前に、AIプロバイダーは欧州委員会が管理するEUデータベースに登録しなければならない（49条）。システムの名称、提供者名、用途、適合性評価情報などの概要は一般公開されるが、営業秘密や機密情報に当たる詳細情報は非公開とされている。

(2) 限定的リスク

- **一般公開**：AI生成物であることの識別や、対話相手が人間でなくAIであるとの明示が義務付けられている（50条）。特に、ディープフェイクなどのAI生成コンテンツに関しては、ウォーターマークやメタデータ付与などによって識別可能にすることが求められている。
- なお、これらは出力物自体へのマーキングを想定しており、一覧できる形式の透明性レポートとして一般公開する仕組みではない。

(3) 最小リスク

- EU AI法上の法的義務は設けられていない。
- 透明性の確保や説明責任の向上は、事業者が自主的に策定する行動規範によって推進されることが想定されており、努力義務のような位置づけとなっている。

(4) 汎用目的AI（GPAI: General-Purpose AI）モデル

- **当局開示：**技術文書の作成・更新・提出に加えて（53条）、システミックリスクを伴う場合はインシデント報告義務などが課されている（55条）。
- **利用者開示：**下流プロバイダーに対する情報開示義務が課されている（53条）。なお、テンプレートとして「Model Documentation Form」が策定されており、情報開示の簡易化・標準化が図られている【Appendix 2 参照】。
- **一般公開**：訓練に用いたコンテンツに関する「十分に詳細な要約」を公開する義務（53条1項(d)）が定められている。なお、情報公開は、テンプレートに基づいて実施されることになっており、事業者と規制当局の双方の対応コスト低減が目指されている。

(5) AI法以外の関連規定

- 個人データの保護に関するGDPR（一般データ保護規則）は、説明責任の原則を定めた上で（5条2項）、人間による介入なしに機械の処理のみに基づいて個人に関する決定を行う「自動化された意思決定」の対象とならない権利を有するとしているため（22条）、データ主体である利用者は一定の情報開示を受けられる。
- プラットフォーム事業者等の責任に関するDSA（デジタルサービス法）は、レコメンデーションシステム（コンテンツ推奨機能）の透明性確保を求めており（27条）、利用者はパラメーターなどについて情報開示を受けられる。また、透明性レポートの公表義務も規定しており（15条）、違法コンテンツ対応やアルゴリズム管理の実績が年次で一般公開される。

4-2. 米国

- 米国では、分野別規制、業界主導ガイドライン、州法などによってAI透明性に関する施策が行われている。
- モザイク型で制度が形成されているため、事業者にとっては対応の複雑さが課題となっている。

(1) イリノイ州「AI動画面接法」

- **利用者開示**：2020年に施行されたイリノイ州法「AI動画面接法（Artificial Intelligence Video Interview Act）」では、応募者の動画をAIで分析する場合に事前の通知と同意取得を義務付けている。特に、特に表情分析については詳細な説明が求められている。
- **当局開示：**選考通過をAIのみで判断する場合、AI評価の下で異なる人口統計学的カテゴリー（人種・性別など）で候補者がどのように評価されたかを示すデータを州政府に提出することが義務付けられている。
- **一般公開**：提出されたデータの集計結果は、州政府の報告書として公表される。

(2) バイデン政権における大統領令と自主的コミットメント

- **当局開示**：2023年10月に公表された「AIの安心、安全で信頼できる開発と利用に関する大統領令」では、安全保障、公衆衛生、安全性に重大なリスクをもたらす基盤モデルを開発する事業者に対して、モデルのトレーニングを行う際の政府への通知、テスト結果の政府への共有などを義務付けていた。しかし、トランプ新政権は当該大統領令を撤回している。
- **一般公開**：主要なAI事業者に自主的な取組みを求めた「Voluntary AI Commitments」において、AI生成コンテンツの表示を展開することや、社会的リスクを含むAIモデル等の能力・限界などについて公開レポートを作成することが求められていた。

(3) 2024年に廃案になったカリフォルニア州法案（SB1047）

- **当局開示**：2024年に廃案になったカリフォルニア州法「フロンティアAIモデルのための安全で安心な技術革新法（SB1047）」では、計算総量が10の26乗フロップ以上のAIモデルを開発する企業に対し、安全性計画を策定して規制当局に提出する義務やインシデント報告義務を設けていた。
- **一般公開**：モデルの用途や使用制限を記載した透明性レポートを公開することも求められていた。
- しかし、ギャビン・ニューサム知事が拒否権を発動したため、廃案となった。シャットダウン機能（「キルスイッチ」）の実装を義務付ける条項などが事業者にとって過度な負担になることを懸念したためと指摘されている。

(4) 2025年に成立したカリフォルニア州法（SB53）

- **当局開示**：2025年に成立したカリフォルニア州法「フロンティアAIモデル透明性法（SB53）」も、計算総量が10の26乗フロップ以上のAIモデルを開発する企業に対し、安全性計画を策定して規制当局に提出する義務やインシデント報告義務を設けている。

- **一般公開**：透明性レポートを自社ウェブサイトで一般公開することを義務付けている。ただし国家安全保障や営業秘密に関する情報は非公開対応となる。
- 廃案となったSB1047は責任追及を重視していたが、本法では透明性確保へと重点を移し、実行可能な義務に修正されたとされている。

(5) AIコンパニオンに関するカリフォルニア州法（SB243）

- **利用者開示**：2025年に成立したカリフォルニア州法「AIコンパニオンチャットボット規制法（SB243）」は、人間ではなくAIと対話していることを未成年の利用者に通知する義務を課している。
- **当局開示**：事業者は、カリフォルニア州の自殺予防オフィスに年次報告書を提出することが義務付けられている。
- **一般公開**：提出されたデータは、個人情報やユーザーに関する識別子を除外した上で州政府から公表される。
- なお、本法成立の背景には、近年、AIチャットボット利用後の未成年の自傷・自殺やAI依存に関する報道がされていることがある。

4-3. 中国

- 中国では、国家主導で透明性に関する義務を法制化しており、一貫性・統一性が目指されている。
- 前掲3カテゴリーのうち、当局開示と一般公開を重視している点に特徴がある。また、エンドユーザーを念頭においた利用者開示は手厚いが、ビジネスユーザー同士に関する利用者開示の措置が相対的に少ない。

(1) アルゴリズム推薦管理規定

- **利用者開示**：2021年に成立した「インターネット情報サービスにおけるアルゴリズム推薦管理規定」では、アルゴリズムの「中核的仕組み」について、適切な形式で利用者に通知することが義務付けられている。
- **当局開示**：アルゴリズムの詳細を政府に提出することが義務付けられている。
- **一般公開**：政府に提出した情報のうち、概要については、一般公開されている。2025年5月までに575件が官報で公開されている[15]。

(2) ディープシンセシス管理規定

- **利用者開示**：2022年に成立した「インターネット情報サービスにおけるディープ・シンセシス管理規定」では、AI生成コンテンツであることを明示するラベリング義務が課されている。
- **当局開示**：アルゴリズムの詳細を政府に提出することが義務付けられている。
- **一般公開**：政府に提出した情報のうち、概要については、一般公開されている。2025年7月までに389件が官報で公開されている[16]。

---

[15] 中国网信网（2025）「国家互联网信息办公室关于发布互联网信息服务算法备案信息的公告」

[16] 中国网信网（2025）「国家互联网信息办公室关于发布第十二批深度合成服务算法备案信息的公告」

(3) 生成AIサービス管理暫定弁法

- **当局開示**：2023年に成立した「**生成AIサービス管理暫定弁法**」では、違法コンテンツの生成の報告義務などが課されている。
- なお、ディープシンセシス管理規定に則って、生成コンテンツにラベルを付与する義務も規定されている。さらに、関連する国家標準として「生成AIサービスのコンテンツのラベリング方法（生成式人工智能服务内容标识方法）」、「生成AIサービス安全基本要件（生成式人工智能服务安全基本要求）」なども公表されている。

4-4. 韓国

- 韓国では、AIに関する包括的な規制と利活用の枠組み「AI基本法」が2024年に成立したが、同法では透明性義務も規定している（31条）。
- **当局開示**：リスクの評価や履行結果を当局に報告することが義務付けられている。
- **利用者開示**：AI事業者は、高影響AIや生成型AIを利用した製品・サービスを提供する際、その運用がAIによるものであることを利用者に事前に通知する必要がある。
- **一般公開**：AI生成コンテンツについて、「AIによって生成された」ものであることを利用者に明示することが義務付けられている。ただし、芸術的・創作的な表現の場合は、表現の自由を損なわない方法での表示が認められている。
- なお、学習データに関する情報の透明性義務については見送られている。韓国新聞協会などからは、学習データ公開や説明責任の強化を法改正で追加すべきとの意見も提出された。文化体育観光部も、AIに機械学習で学習させたデータの目録の公開に係る規定を入れるべきである等の主張を行っていたという経緯がある[17]。

5. 日本におけるAI透明性に関する制度構築の方向性

- 日本では、AIに関する透明性確保の取組みは、既存法令による義務に加えて、ガイドラインなどの任意的枠組みを組み合わせることによって進められてきた。今後の制度設計においても、法規制型とインセンティブ型を組み合わせた多層的構造を整備することが求められる。
- AI透明性を法規制に基づく義務としてだけでなく、インセンティブについても考慮し、ステークホルダー間のコミュニケーションの基盤として位置づけることが重要である。
- 開示・公開された情報をもとに相互に学び合い、ガバナンスを改善し、価値を循環させるエコシステムの構築が、日本の制度設計の方向性となる。
- 日本政府に対して、以下の4点を提言する。

[17] 藤原夏人(2025)「韓国における AI 基本法の制定」国立国会図書館調査及び立法考査局 編『外国の立法』304. pp. 41-69.

提言1　AI透明性に関する既存法令の解釈明確化と「法の欠缺」の調査推進

- 既存の法規制の明確化を提案する。AI透明性についても、既存の法体系（消費者保護法、労働関連法、個人情報保護法、行政手続法など）によって、一定程度対応可能である。不当表示や不意打ちなどを招く行為の多くは既に禁止されており、行政による是正措置・業務改善命令等による一般的な救済メカニズムも整備されている。
- もっとも、法令は抽象的であり、AIシステムに関する解釈適用について疑義が生じる場合もある。そのため、既存法令の解釈について明確化することが、予見可能性を高める意味でも望ましい。
- 同時に、既存法令の適用範囲を超える新たなリスクが発生する場合も想定される。そのため、AI推進法に基づいて政府が調査を行い、「法の欠缺」について積極的に探索すべきと提案する。
- 諸外国におけるAI透明性に関する制度を踏まえると、例えば、以下の類型について、さらなる調査研究が必要であると考えられる【表3参照】。

【表3】AI透明性についての調査研究例

リスク（例）	調査研究内容（例）
大規模モデルなど社会的影響が大きいAIシステムの制御不能	・AIシステムの解釈可能性に関する利用者開示の現状（営業秘密など配慮すべき論点などを含む） ・AIシステムの監査可能性に関する当局開示のあり方（規模に依存しない高性能なAIが開発されていること等を踏まえ、どのような要素を考慮すべきかなどを含む） ・一般公開される透明性レポート等のベストプラクティス
AIシステムの開発・学習・利用過程での権利侵害、データを公開することに対する経済的インセンティブ低下	・大規模モデルなどにおいて訓練に用いられたデータに関する公開のあり方 ・データ権利者などへの収益還元プログラムなどの取組みの現状
AI生成コンテンツに関する誤認や不信	・ディープフェイクに関する現状 ・AI生成コンテンツに関する表示のあり方
人事や与信などにおけるバイアス（差別・偏見）　の助長	・人事や与信などにおけるAIシステム利用の現状 ・人事や与信などにおける説明可能性に関するベストプラクティス ・AI利用に関する利用者開示のあり方
AI依存や自傷行為の誘発など人間とAIの負の相互作用	・対話型AI、AIコンパニオン、AIパーソナライズ機能などにおいて不適切な出力を防ぐ安全機能「ガードレール」などに関する現状 ・未成年者等に対する保護策の現状

	・利用者開示のあり方（利用者本人だけでなく保護者への通知について別途考慮すべきことなどを含む）
環境負荷の増大	・環境負荷の可視化に関する現状やベストプラクティス ・情報公開における環境指標のあり方

提言2　政府調査の適正手続きに関する指針策定

- 日本政府は、調査協力を企業などに要請する際、広島AIプロセス等でも確認された、法の支配、適正手続き、民主的責任行政などの基本原理を遵守すべきである。
- 政府による恣意的な権限行使を抑止し、事業者等の予見可能性低下や委縮効果を生じさせないように、政府側も透明性を高める指針を策定することを提案する。
- 策定の際には、2021年の「データ、プライバシー及び法の支配へのガバメントアクセス」に関する世界プライバシー会議（GPA）決議、2022年のOECD閣僚会合で採択された「民間部門が保有する個人データに対するガバメントアクセスに関する宣言」なども参照すべきである。

提言3　AI透明性を進展させる経済的インセンティブの強化に向けた制度構築

- 情報開示・情報公開の促進として、経済的インセンティブの強化をすべきである。
- AI透明性に関する情報公開について、ESG投資やグリーンファイナンスといった非財務的評価指標に反映される仕組みを機関投資家のコミュニティと連携して構築することを、日本政府に提案する。
- また、証券取引所などと連携して、AI透明性に優れた企業を「AIトラスト銘柄」として選定することも検討に値する。その場合は、企業価値とAI透明性に関する取組みとの関係を定量的に分析するフレームワークを開発して、効果検証をすることも重要となる。
- このほか、公共調達の評価などにおける加点や、市場圧力を十分に機能させるための一般消費者向けの啓発活動の推進、AIリテラシー教育プログラムの開発なども選択肢となる。

提言4　企業の自己宣言に対する履行確保メカニズムの導入に関する検討

- AI透明性を適正性・信頼性に架橋すべく「自己宣言の履行確保メカニズム」の導入を提案する。
- 後述する「広島AIプロセス」の報告枠組みのように、企業が自主的に公開する透明性レポートにおいて、AIの安全対策やリスク管理体制などを明示することが広がりつつある。こうした取組みを社会に対する約束や自己宣言と捉え、自己宣言の内容の真実性・履行状況を監督当局や第三者機関が確認し、虚偽または不履行が判明し

た場合には、欺瞞的行為として是正の対象とする。こうすることで、不誠実な主体が「透明性」を欺瞞的な訴求の手段として利用することを防止できる。

- これは、遵守内容そのものを政府が決定するのではなく、企業等が自ら宣言した自主規制を守る義務を課す点に特徴がある。こうしたメカニズムの導入により、企業による研究の自由や営業の自由を確保しつつ、説明責任の実効性を担保できる。
- なお、この提案は、技術的中立性の原則を前提としており、AIという特定の技術に対して加重的に規制を課すものではない。金融分野などで既に一般化している自主的開示と説明責任の確保の仕組みをAIシステムにも適用するものである。これにより、健全な競争条件を確保し、信頼性の高い事業を行う企業が正当に評価される市場環境を整備することが期待できる。

6. 「広島AIプロセス」報告枠組みの活用

- 前記の方向性を実践する上で重要になるのが、「広島AIプロセス（HAIP: Hiroshima AI Process）」である。2023年に日本がG7議長国として立ち上げたHAIPは、生成AIを含む高度なAIシステムに関する国際的なルール作りを行うために立ち上げられた。HAIPは、以下の3つの要素で構成されている【表4参照】。

【表4】広島AIプロセス（HAIP）の3要素

全てのAI関係者向けの広島プロセス国際指針（Principles）	人間中心性、説明責任、透明性といった基本的価値を示す共通原則
高度なAIシステムを開発する組織向けの国際行動規範（Code of Conduct）	指針を実装するための実務的な手引き
報告枠組み（Reporting Framework）	各組織がどのようにAIを設計・運用し、リスクに対応しているかを開示するフォーマット

- HAIPの報告枠組みは、透明性及び説明責任を促進するため、国際行動規範の遵守状況をAI開発者等が自主的に報告・公開するための手法として、2025年2月から正式に運用が開始された。
- 「リスク管理」「インシデント管理」「安全性向上のための研究開発投資」など7セクション・全39項目で構成される質問票に基づき、各組織が回答を提出する仕組みである。事務局を務める経済協力開発機構（OECD）のウェブサイトで誰でも閲覧できる。
- 日本においては、この国際的な報告枠組みを国内制度と接続することが重要である。広島AIプロセスを単なる外交的成果にとどめず、国内にも根付かせ、法規制とインセンティブの橋渡しを行うことが、今後のAIガバナンス政策の鍵となる。
- 日本政府に対して、以下の3点を提言する。

提言5　広島AIプロセス報告枠組みの活用による相互運用性確保

- 日本政府に対して、国内制度としてHAIPの報告枠組みを活用することを提案する。
- AI推進法において、AIの研究開発及び活用の適正な実施を図るため、国際的な規範の趣旨に即した指針を整備することとされている。ここで「国際的な規範」として念頭に置かれているのは、主として「全てのAI関係者向けの広島プロセス国際指針（Principles）」や「高度なAIシステムを開発する組織向けの国際行動規範（Code of Conduct）」であると見られる[18]。つまり、国際行動規範の遵守状況のモニタリングすることに役立つ「報告枠組み（Reporting Framework）」との接合性は必ずしも明らかになっていない。
- 国際的な相互運用性の観点からは、HAIPの報告枠組みを国内制度としても活用することが効果的である。
- 国内で利用されるAIシステムの多くは国外事業者が提供しているが、地理的要因等からコンプライアンスの協力を得にくい国外事業者に対しても制度の実効性を確保する必要がある。仮に、いわゆる「域外適用」が難しいとすると、事業者に対して透明性義務を課すような法規制について国内事業者のみを対象とすることになり、国内事業者が一方的に不利益を被るばかりか、国内利用者保護もが十分に図れない。そのため、国外事業者についても、国内事業者と同じく制度の対象とする必要がある。
- この点、HAIPの報告枠組みには、既にGoogle、Microsoft、Salesforce、OpenAI、Anthropicなどの主要な国外事業者が参加しており、連携の基盤が形成されている。
- また、HAIP報告枠組みの活用は、質問票が既に公表されていることから、提言2において言及した、政府からのアクセスに関する適正手続きを確保するという観点からも望ましい。
- そこで、HAIPの報告枠組みの質問票を日本語に翻訳することを基本としつつ、日本における個別事情に応じた質問を補完的に追加することを検討すべきである。例えば、プライバシーポリシーの公開有無については、HAIPの報告枠組みの質問票に既に記載があるが、それに追加して、プライバシーポリシーが日本語で公開されているか否かなどを尋ねることなどが想定される。
- なお、筆者らは、AIガバナンスに資する透明性レポートのハンドブックおよびHAIPの報告枠組みにおける実例を参照したワークシートを作成しており、透明性レポートを新たに作成・公開したい事業者の支援ツールを用意している。HAIPの報告枠組みに参加していない国内事業者であっても、参画しやすい環境が整備されつつある。

提言6　国内利用者のアクセス性を高める透明性レポートの集約プラットフォーム整備

- 透明性レポートの集約プラットフォームを設置することを、日本政府に提案する。
- HAIPの報告枠組みでの提出物をはじめとする、企業が公開した透明性レポートを統合的に閲覧・検索可能にするデータベースを整備すれば、機関投資家や市民社会に

[18] 内閣府（2025）「ＡＩ法に基づく適正性確保に関する指針の整備ついて」

よる分析を可能にする。また、HAIPの報告枠組みに参画しない組織についても、包摂することができる。

- さらに、AI透明性に関する知見を共有して意見交換を行うコミュニティの形成を支援することも検討に値する。
- 加えて、国内外のベストプラクティスを周知・顕彰するイベントを日本政府が主催することなども選択肢となる。その場合は、HAIPの「フレンズグループ」と呼ばれる56の有志国・地域のステークホルダーにもアウトリーチを行い、東南アジア諸国連合（ASEAN）やアフリカ開発会議（TICAD）などグローバルサウスとの連携と交流を深める舞台とすべきである。

提言7　AI透明性に関する国際的クロスウォークと相互承認の推進

- 国際的なクロスウォークの実施を提案する。
- EUのAI法、韓国のAI基本法など、AI透明性に関する異なるフレームワークを比較して、対応関係を整理すれば、相互運用性が高まる。事業者が国内外の基準に沿って透明性に関する取組みを推進しやすくなることが期待できる。
- さらに、HAIPの報告枠組みにおいて透明性レポートを公開することが、各国における規制当局への定例報告などとして認められるよう、国内制度の相互承認に関する仕組みを検討することも選択肢となる。

7. おわりに：AI透明性が評価される社会に向けて

- 本提言は、AIガバナンスにおける基盤的概念である透明性の取組みについて、コンプライアンスのチェックボックスに押し込めるのではなく、市場と社会の信頼によって価値が循環するエコシステムとして拡張しようと試みた。
- 7つの提言において、法的規制という「縦のガバナンス」と、自主的取り組みを促す「横のインセンティブ」をレイヤーとして重ね、社会実装に耐えうる多層構造のガバナンスを示している。
- 特に、広島AIプロセスの報告枠組みを国内制度として活用する仕組みは、日本のみならず、グローバルサウスなどの有志国・地域でも展開可能であり、国際戦略上も重要な基盤となり得る。
- AIの社会実装が急速に進む中、イノベーションを止めずにどう安全性や信頼性を組み込むかが問われているが、透明性は、その起点となり得る。

Appendix 1. 概念整理

- **透明性（transparency）**：AIシステムに関する情報を可視化し外部にアクセス可能にすること。
- **説明可能性（explainability）**：AIシステムが出力した特定の判断について、人間が理解できる理由や根拠を提示すること[19]。例えば、AIによるプロファイリング結果を利用者や被影響者に対して通知する場面などで特に重要になる。
- **解釈可能性（interpretability）**：AIシステムのモデルの構造や動作原理を人間（特に専門家）が理解できる程度を示すこと。例えば、生成AIモデルの制御不能リスクを検証するといったモデルの内部表現や学習過程が問題になる場面などで特に重要になる。
- **監査可能性（auditability）、検証可能性（verifiability）**：AIシステムの設計・運用・結果を人間（特に専門家）が確認できる状態にあること。
- **答責性／説明責任（accountability）**：AIシステムの利用により生じた結果について、開発者や運用者が応答し、必要に応じて責任をとること。説明することだけでなく、責任を引き受けるために、誰が責任を負うのかを明確にしたり、異議申立ての手続きを整えたりすることなどを含む。

説明可能性（explainability）と解釈可能性（interpretability）を区別して用いたが、国際的に合意された明確な定義はない。そのため、両者を区別しない場合や、区別していても意味が異なる場合もある点に注意が必要である【表5参照】。特に、EUと米国の国立標準技術研究所（NIST: National Institute of Standards and Technology）では概ね逆の意味として用いられているケースが見られる。

【表5】説明可能性と解釈可能性の比較

機関・文書	Explainability （説明可能性）	Interpretability （解釈可能性）
OECD（2019） *OECD Principles on AI*	**AIの出力や決定**について、人々がその理由やロジックを理解できるようにすること。	用語としては明示されていない。透明性や説明可能性に統合していると見られる。
OECD.AI Policy Observatory / ONE AI	**AIの挙動や出力**を人間に理解できる形で説明する能力。	用語としては明示されていない。透明性や説明可能性に統合していると見られる。
EU（2019） *Ethics Guidelines for Trustworthy AI*	**AIシステムが出力した判断の理由・根拠**を人間が理解できる形で説明できる能力。	**モデルの内部構造や動作原理**を人間（専門家）が理解できる程度。
ISO/IEC 22989:2022	**AIシステムの内部メカニズムや出力**が人間に理解できる形	**AIモデルの内部論理や構造**を人間が理解できる程度。

[19] OECDの主要な政策文書において explainability は用いられているが、interpretability は用いられておらず、本提言で示したような区別をしていない点に注意が必要である。

	で説明される程度。	
ISO/IEC TS 6254:2025	**AI出力の理由やプロセス**を人間が理解できる形で説明する能力。	**モデルの構造・挙動**を専門家が理解可能である程度。 （解釈可能性は説明可能性を支える技術的基盤と位置づけ）
米国 NIST（2021） *NIST IR 8367*	モデルの内部メカニズムや実装を記述する能力 （どのようにその出力が生じたかを説明する）	人間がAI出力を理解し、モデルの判断を予測・理由づけできる程度。 （なぜその出力が意味を持つか・設計上の意図と関連して理解できるかを扱う）
米国 NIST（2023） *AI Risk Management Framework*	システムの挙動や出力がどのように導かれたかを説明できること	出力の意味を人間が理解し、それがどのような文脈・目的に沿うかを理解できる程度
機械学習やXAI分野の研究論文	モデル構造やパラメータ、特徴変数の重みや内部構造を、人が直接理解できるようにすることを指すことが多い。 ただし、あまり区別せず、ほぼ同義的な概念として扱う場合もある。	ブラックボックスモデルに後付けで説明を与えること（特徴寄与、局所説明、対応例提示など）を指すことが多い。 ただし、あまり区別せず、ほぼ同義的な概念として扱う場合もある。

Appendix 2. 汎用目的AIモデルに関する情報開示

- 一般公開について、HAIPの報告枠組みを活用することを提案した。
- 当局開示と利用者開示について、大規模な基盤モデルに関しては、EU AI法における汎用目的AIモデルの行動規範が参考になる。
- 透明性に関するテンプレートとして「Model Documentation Form」が策定されており、以下のような項目が設けられ、情報開示の標準化が図られている【図1参照】。

【図1】Model Documentation Form

Model Documentation Form

This Form includes all the information to be documented as part of Measure 1.1 of the Transparency Chapter of the Code of Practice. 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 and NCAs shall be treated in accordance with the confidentiality obligations and trade secret protections set out in Article 78 AI Act.

Date this document was last updated: Click or tap to enter a date. **Document version number:** Click or tap here to enter text.

General information		AIO	NCAs	DPs
Legal name for the model provider:	Click here to add text.	☒	☒	☒
Model name:	The unique identifier for the model (e.g. Llama 3.1-405B), including the identifier for the collection of models where applicable, and a list of the names of the publicly available versions of the concerned model covered by the Model Documentation.	☒	☒	☒
Model authenticity:	Evidence that establishes the provenance and authenticity of the model (e.g. a secure hash if binaries are distributed, or the URL endpoint in the case of a service), where available.	☒	☒	☐
Release date:	Click or tap to enter a date. Date when the model was first released through any distribution channel.	☒	☒	☒
Union market release date:	Click or tap to enter a date. Date when the model was placed on the Union market.	☒	☒	☒
Model dependencies:	If the model is the result of a modification or fine-tuning of one or more general-purpose AI models previously placed on the market, list the model name(s) (and relevant version(s) if more than one version has been placed on the market) of those model(s). Otherwise write 'N/A'.	☒	☒	☒

Model properties						AIO	NCAs	DPs
Model architecture:	A general description of the model architecture, e.g. a transformer architecture. *[Recommended 20 words]*					☒	☒	☒
Design specifications of the model:	A general description of the key design specifications of the model, including rationale and assumptions made, to provide basic insight into how the model was designed. *[Recommended 100 words]* If any other please specify:					☒	☒	☐
Input modalities:	☐Text	☐Images	☐Audio	☐Video	If any other please specify:	☒	☒	☒
For each selected modality please include maximum input size or write 'N/A' if not defined	Maximum size: …	Maximum size: …	Maximum size: …	Maximum size: …	Maximum size: …	☒	☐	☒
Output modalities:	☐Text	☐Images	☐Audio	☐Video	If any other please specify:	☒	☒	☒
For each selected modality please include maximum output size or write 'N/A' if not	Maximum size:	Maximum size:	Maximum size:	Maximum size:	Maximum size:	☐	☐	☒

著者による仮訳

Model Documentation Form
このフォームには「Code of Practice（行動規範）」の透明性チャプターにおける「措置1.1」の一環として、文書化すべきすべての情報が含まれています。 右欄にあるチェックマークは、文書化された情報の提供先が以下いずれかであることを示しています。**AIオフィス（AIO）、各国の規制当局（NCAs）、下流プロバイダー（DPs）**。なお、DPsは、汎用目的AIモデルを自らのAIシステムに組み込もうとするAIシステムのプロバイダーのことを指します。

DPs向けの情報は、積極的に提供されるべきものですが、AIOまたはNCAs向けの情報は、AIOからの要請があった場合に限り提供されます。この要請は、AIOが自らの職権判断で行う場合や、NCAsからの要請に基づいてAIOが行う場合があります。 そのような要請には、法的根拠と目的が明記され、AIOがAI法（AI Act）に基づく職務を遂行するため、または、NCAがAI法に基づく監督権限を行使するために当該時点で厳密に必要とされるフォームの項目のみに限定されます。特に、モデルのプロバイダーとAIシステムのプロバイダーが異なる場合に、汎用目的AIモデルを基盤とする高リスクAIシステムの適合性を評価するための情報が対象となります。
Model Documentation Formに含まれる情報のうち、AIOおよびNCAsと共有される要素については、AI法78条に定められた秘密保持義務および営業秘密保護に則って取り扱われます。

最終更新日：[クリックまたはタップして日付を入力]
バージョン番号：[クリックまたはタップしてテキストを入力]

一般情報		AIO	NCAs	DPs
モデルプロバイダーの正式名称	[ここをクリックしてテキストを入力]	☑	☑	☑
モデル名	モデルの一意な識別子（例：Llama 3.1-405B）のことです。該当する場合は、モデル群（コレクション）の識別子も含めてください。さらに、本Model Documentationの対象となる当該モデルの一般公開版の名称一覧を記載してください。	☑	☑	☑
モデルの真正性（authenticity）	可能な場合には、モデルの出所（provenance）および真正性（authenticity）を証明する証拠を示してください（例：バイナリが配布される場合は安全なハッシュ値、サービスとして提供される場合はURLエンドポイントなど）。	☑	☑	
リリース日	[クリックまたはタップして日付を入力] モデルがいずれかのディストリビューション・チャネルを通じて初めてリリースされた日付です。	☑	☑	☑
EU域内市場のリリース日	[クリックまたはタップして日付を入力] モデルがEU域内市場に投入された日付です。	☑	☑	☑
モデルの依存関係	モデルが、以前に市場に投入された1つまたは複数の汎用目的AIモデルを改変またはファインチューニングした結果である場合は、そのモデル名（および複数のバージョンが市場に投入されている場合は該当バージョン）を記載してください。 該当しない場合は、「N/A」（該当なし）と記入してください。	☑	☑	☑

モデルプロパティ		AIO	NCAs	DPs
モデルのアーキテクチャ	モデルのアーキテクチャについての一般的な説明を記入してください（例：トランスフォーマー型アーキテクチャ）。 ※ 推奨文字数：20words程度	☑	☑	☑
モデルの設計仕様	モデルの主要な設計仕様について、一般的な説明を記入してください。その際、設計の背景となる考え方（rationale）や前提条件（assumptions）など、モデルがどのような設計思想に基づいて構築されたのかを基本的に理解できるようにしてください。 ※ 推奨文字数：100words程度 [その他の場合、具体的に記載]	☑	☑	
入力モダリティ	☐ テキスト ☐ 画像 ☐ 音声 ☐ 動画 [その他の場合、具体的に記載]	☑	☑	☑
選択した各入力モダリティの最大サイズ。定義されていない場合は「N/A」と記入	[最大サイズ] [最大サイズ] [最大サイズ] [最大サイズ] [最大サイズ]	☑		☑
出力モダリティ	☐ テキスト ☐ 画像 ☐ 音声 ☐ 動画 [その他の場合、具体的に記載]	☑	☑	☑
選択した各出力モダリティの最大サイズ。定義されていない場合は「N/A」と記入	[最大サイズ] [最大サイズ] [最大サイズ] [最大サイズ] [最大サイズ]			☑
モデルの総サイズ	モデルに含まれるパラメータの総数を、少なくとも2桁の有効数字で記入してください。（例：7.3×10^10 パラメータ）	☑		
モデルの総パラメータ数がおおよそどの範囲に属するか	☐ 1-500M ☐ 500M-5B ☐ 5B-15B ☐ 15B-50B ☐ 50B-100B ☐ 100B-500B ☐ 500B-1T ☐ >1T		☑	☑

モデルのディストリビューションとライセンス	AIO	NCAs	DPs

ディストリビューション・チャンネル	モデルがEU域内市場でディストリビューションまたは利用可能となる方法の一覧を記入してください。 （例：既存のソフトウェアスイートや専用ソリューションを通じたエンタープライズ向けまたはサブスクリプション型のアクセス；API経由の公開アクセスまたはサブスクリプション型アクセス；統合開発環境、デバイス専用アプリケーション、ファームウェア、オープンソースリポジトリを通じた公開または専有アクセスなど） 可能な場合は、モデルへのアクセス方法に関する情報へのリンク、アクセスレベル（例：ウェイトレベルでのアクセス、ブラックボックスアクセス）も併せて記入してください。	☑	☑	
ライセンス	モデルが下流プロバイダーに対して利用可能になるディストリビューション方法の一覧を記入してください。 （例：既存のソフトウェアスイートや専用ソリューションを通じたエンタープライズ向けまたはサブスクリプション型のアクセス；API経由の公開アクセスまたはサブスクリプション型アクセス；統合開発環境、デバイス専用アプリケーション、ファームウェア、オープンソースリポジトリを通じた公開または専有アクセスなど）			☑
	モデルのライセンスへのリンクを記入してください（リンクがない場合は、AI法91条に基づくAIOからの要請に応じてライセンスの写しを提供してください）。 ライセンスが存在しない場合は、その旨（「ライセンスなし」など）を明記してください。	☑	☑	
	モデルが下流プロバイダーに対して提供される際に適用されるライセンスの種類またはカテゴリーを記入してください。 例えば、フリーかつオープンソースのライセンス（モデルを公開共有でき、提供者が自由にアクセス・利用・改変・再配布（または改変版の再配布）を行えるもの）、制限付きライセンス（利用目的などに一定の制約を設けるもの。倫理的利用を確保するための制限など）、プロプライエタリ（専有）ライセンス（モデルのソースコードへのアクセスを制限し、利用・配布・改変に制約を課すもの）など。			☑

	ライセンスが存在しない場合は、下流プロバイダーへのアクセス方法（例：利用規約を通じて提供されるなど）について説明してください。			
	追加的に提供される関連アセットがある場合は、その一覧を記入してください。 （例：学習データ、データ処理コード、モデル学習コード、推論コード、モデル評価コードなど） それぞれのアセットについて、アクセス方法（どのように入手・利用できるか）と適用されるライセンス（存在する場合）も明記してください。	☑		☑

利用		AIO	NCAs	DPs
利用ポリシー	適用される許容利用ポリシー（Acceptable Use Policy）へのリンクを記入してください。 リンクがない場合は、ポリシーの写しを添付するか、ポリシーが存在しない旨を明記してください。	☑	☑	☑
想定される利用目的	以下のいずれか、または両方について記入してください。 (i) プロバイダーが意図する利用目的 （例：生産性向上、翻訳、創作コンテンツ生成、データ分析、データ可視化、プログラミング支援、スケジューリング、カスタマーサポート、各種自然言語処理タスクなど）。 (ii) プロバイダーが制限または禁止している利用目的（EU法または国際法（AI法5条を含む）で禁止されているもの以外で、提供者独自に禁止している用途）。 これらの内容は、使用説明書、利用規約、販売資料、広報・宣伝資料、技術文書などに記載された情報に基づいて記入してください。 モデルが提供されるライセンスの性質上、(i)または(ii)の特定が不適切または不可能な場合は、「N/A」（該当なし）と記入してください。 ※推奨：200words程度	☑	☑	☑
汎用目的AIモデルを組み込むことができるAIシステムの種類および性質	以下のいずれか、または両方について、一覧または説明を記入してください。 (i) 汎用目的AIモデルを組み込むことができるAIシステムの種類・性質。 (ii) 汎用目的AIモデルを組み込むべきではないAIシステムの種類・性質。	☑	☑	☑

	例えば、以下のようなシステムが含まれます：自律型システム、会話支援システム、意思決定支援システム、クリエイティブAIシステム、予測システム、サイバーセキュリティ関連システム、監視システム、人間とAIの協働システム。 ※推奨：300words以内			
モデル統合のための技術的手段	汎用目的AIモデルをAIシステムに統合する際に必要となる技術的手段について、一般的な説明を記入してください。例えば、統合のための手順書（instructions for use）、必要なインフラ、ツールやAPI、開発環境、ソフトウェア開発キット（SDK）など。 ※推奨：100words程度			☑
必要なハードウェア	モデルを利用するために必要となるハードウェア（該当する場合はそのバージョンを含む）について記入してください。 該当しない場合（例：モデルがAPI経由で提供される場合など）は、「N/A」（該当なし）と記入してください。 ※推奨：100words程度			☑
必要なソフトウェア	モデルを利用するために必要となるソフトウェア（該当する場合はそのバージョンを含む）について記入してください。 該当しない場合は、「N/A」（該当なし）と記入してください。 ※推奨：100words程度			☑

トレーニング・プロセス		AIO	NCAs	DPs
学習プロセスの設計仕様	学習プロセスに関する主な手順や段階について、一般的な説明を記入してください。その際、以下の内容を含めてください：使用した学習手法および技術、主要な設計上の選択肢、学習時に置いた前提条件、モデルが最適化の対象としている目的、また、必要に応じて、各種パラメータの重要性や関連性など。 例えば、以下のように記述してください。「モデルはランダムに初期化されたウェイトを持ち、Adamオプティマイザーを用いた勾配ベースの最適化によって2段階で学習された。第1段階では、大規模な事前学習コーパスにおいてクロスエントロピー損失を最小化し、次単語予測を1エポック行った。第2段階では、人間の嗜好データセットを用いて10エポックの学習を行い、人間の価値観	☑	☑	

	に整合するようモデルを調整し、ユーザーからのプロンプトへの応答性を高めた」。 ※推奨：400words程度			
意思決定の根拠	モデルの学習において行われた主要な設計上の選択について、その内容（how）および理由（why）を記入してください。 ※推奨：200words程度	☑	☑	

学習・テスト・検証に利用されたデータの情報		AIO	NCAs	DPs
データの種類／モダリティ 該当するすべての項目を選択	☐ テキスト ☐ 画像 ☐ 音声 ☐ 動画 [その他の場合、具体的に記載]	☑	☑	☑
データの出所 該当するすべての項目を選択 各カテゴリーの定義については、AIOが提供する「汎用目的AIモデルの学習内容に関する公開要約テンプレートを参照	☐ ウェブクローリングによるデータ ☐ 第三者から入手した非公開データセット ☐ ユーザーデータ ☐ 公開データセット ☐ その他の手段で収集されたデータ ☐ プロバイダー自身またはその委託により作成された非公開の合成データ [その他の場合、具体的に記載]	☑	☑	☑
データの取得および選定方法	学習・テスト・検証データを取得および選定するために用いた方法について説明してください。 データの注釈付けに使用された方法およびリソース、ならびに該当する場合には合成データを生成するために使用されたモデルおよび方法を含めてください。 第三者から以前に取得されたデータが含まれる場合、AI法53条1項(d)に基づく公開訓練データ要約の開示をすでに行っていない場合に限り、提供者がどのようにデータ利用権を取得したかを説明してください。 ※推奨：300words程度	☑	☑	
データポイント数	学習、テスト、および検証データのそれぞれについて、データポイント数を記入してください。 また、データポイントの単位（例：トークンまたは文書、画像、動画の時間（時間数）またはフレーム数）を定義してください。 数値は少なくとも1桁の有効数字にしてください（例：3×10^{13}トークン）。		☑	
	学習、テスト、および検証データのそれぞれについて、データポイント数を記入して	☑		

	ください。 また、データポイントの単位（例：トークンまたは文書、画像、動画の時間（時間数）またはフレーム数）を定義してください。 数値は少なくとも2桁の有効数字にしてください（例：1.5×10^{13}トークン）。			
データの範囲および主な特徴	学習、テスト、および検証データの範囲と主な特徴についての一般的な説明を記入してください。 該当する場合には、以下の要素を含めてください：分野（医療、科学、法律など）、地理的範囲（グローバル、特定地域に限定など）、言語、モダリティの範囲（例：テキスト、画像、音声、動画など）。 ※推奨：200words程度	☑	☑	
データキュレーション手法	取得したデータをモデルの学習、テスト、および検証データに変換する際に行われたデータ処理の一般的な説明を記入してください。 例えば、以下のような処理を含みます：クリーニング（広告などの無関係なコンテンツなど不要な内容の除去）、トークン化などの正規化、バックトランスレーションなどの拡張など。 ※推奨：300words程度	☑	☑	☑
不適切なデータソースを検出するための措置	モデルの想定される利用目的を考慮し、不適切なデータソースの存在を検出するためにデータの取得または処理の過程で実施された方法について、もし実施していれば、その説明を記入してください。 以下のような内容を含めてください：違法コンテンツ、児童性的虐待に関する資料（CSAM）、同意のない親密画像（NCII）、不法な処理につながる個人データなど。 ※推奨：400words程度	☑	☑	
識別可能なバイアスを検出するための措置	学習データに存在する識別可能なバイアスの偏りに対処するために、データの取得または処理の過程で実施された方法について、もし実施していれば、その説明を記入してください。 ※推奨：200words程度	☑	☑	

（学習中に使用された）計算資源		AIO	NCAs	DPs
学習時間	測定対象となる期間を説明してください。 また、その期間の長さが以下のいずれの範囲に該当するか（1か月未満、1〜3か月、3		☑	

	～6か月、6か月超）について記入してください。			
	測定対象となる期間の説明に加えて、その期間の長さを、実時間（日数換算、例：9×10^{1}日）およびハードウェア換算（日数換算、例：4×10^{5} Nvidia A100日、2×10^{5} Nvidia H100日）の双方で、少なくとも1桁の有効数字を用いて記入してください。	☑		
学習に使用された計算量	学習に使用された計算量の実測値または推定値を、浮動小数点演算回数（floating point operations）で示し、その桁数（例：10^{24}FLOPs）まで示して記入してください。		☑	
	学習に使用された計算量の実測値または推定値を、計算処理回数（例：2.4×10^{25} FLOPs）で示し、少なくとも2桁の有効数字を用いて記入してください。	☑		
測定方法	AI法53条5項に基づく、測定および計算方法を詳細化する委任法令が採択されていない場合、学習に使用された計算量を測定または推定するために用いた方法論を記入してください。	☑	☑	

（学習および推論における）エネルギー消費量		AIO	NCAs	DPs
学習に使用されたエネルギー量	学習に使用されたエネルギー量の実測値または推定値を、メガワット時（MWh）で記入してください。少なくとも2桁の有効数字で示してください（例：1.0×10^{2} MWh）。 計算資源やハードウェア提供者からの重要情報が不足しており、学習に使用されたエネルギー量を推定できない場合は、「N/A」（該当なし）と記入してください。	☑	☑	
測定方法	AI法53条5項に基づく、測定および計算方法を詳細化する委任法令が採択されていない場合、学習に使用されたエネルギー量を測定または推定するために用いた方法論を記入してください。 モデルのエネルギー消費量が不明な場合は、使用された計算資源に関する情報に基づいて推定してください。 学習に使用されたエネルギー量を、計算資源やハードウェアの提供者からの重要情報の欠如によって推定できない場合、プロバイダーは不足している情報の種類を開示してください。 ※推奨：100words程度	☑	☑	
推論に使用された計算量（ベンチマーク値）	推論に使用された計算量のベンチマーク値を、浮動小数点演算回数（floating point operations）で記入してください。少なく	☑	☑	

	とも2桁の有効数字で示してください（例：5.1×10^{17}FLOPs）。			
測定方法	AI法53条5項に基づく、測定および計算方法を詳細化する委任法令が採択されていない場合、推論に使用された計算量を測定または推定する際に用いた計算タスク（例：10万トークンの生成）およびハードウェア構成（例：Nvidia A100を64基使用）について記入してください。	☑	☑	

本項目は、推論時のエネルギー消費に関するものであり、学習時のエネルギー消費と合わせて「モデルのエネルギー消費」（AI法附属書XI第2(e)項）を構成する。推論時のエネルギー消費はモデルそのもの以外の要因にも依存するため、本項目で求められる情報は、モデル自体にのみ依存する関連情報、すなわち、推論に使用された計算資源に関する情報に限定される。