



| | |
|--------------|---|
| Title | 生成AIの倫理的・法的・社会的課題（ELSI）論点の概観：2023年4～8月版：グローバルな政策動向を中心に |
| Author(s) | 岸本, 充生; カテライ, アメリア; 井出, 和希 |
| Citation | ELSI NOTE. 2023, 30, p. 1-67 |
| Version Type | VoR |
| URL | https://doi.org/10.18910/92475 |
| rights | |
| Note | |

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka



生成AIの倫理的・法的・社会的課題（ELSI）論点の概観:

2023年4～8月版

ーグローバルな政策動向を中心にー

Authors

| | | |
|-----------|-----------------------------|-------------------|
| 岸本 充生 | 大阪大学 社会技術共創研究センター（ELSIセンター） | センター長 (2023年8月現在) |
| カテライ アメリア | 大阪大学 社会技術共創研究センター（ELSIセンター） | 特任助教 (2023年8月現在) |
| 井出 和希 | 大阪大学 感染症総合教育研究拠点/ELSIセンター | 特任准教授 (2023年8月現在) |

本 ELSI NOTE は 2023 年 4 月に公開された、ELSI NOTE No.26『生成 AI（Generative AI）の倫理的・法的・社会的課題（ELSI）論点の概観：2023 年 3 月版』（カテライ アメリア、井出和希、岸本充生）

（<https://doi.org/10.18910/90926>）の続編になります。両方セットで読むことを推奨します。

脚注に記載されている URL の多くは 2023 年 8 月半ば時点でアクセスできたものであり、その後にアクセスができなくなる可能性があることに注意してください。また一部に購読を必要とする URL も含まれていることをご了承ください。

目次

| | |
|-----------------------------------|-----------|
| はじめに：グローバルなリスクガバナンスの構図 | 3 |
| 1. 生成 AI 事業者の動向..... | 5 |
| 1.1. OpenAI 社の動向 | 5 |
| 1.2. OpenAI 社以外の企業の動向 | 7 |
| 1.3. 生成 AI 事業者による政策提案や自主的取組 | 10 |
| 1.4. 生成 AI モデルの比較 | 13 |
| 2. 権利者団体の反応と訴訟 | 13 |
| 2.1. 国内 | 13 |
| 2.2. 米国 | 17 |
| 2.3. その他 | 20 |
| 2.4. 訴訟 | 21 |
| 3. 各国・機関の政策動向 | 23 |
| 3.1. 日本 | 24 |
| 3.2. 欧州（EU） | 30 |
| 3.3. 米国 | 34 |
| 3.4. 英国 | 40 |
| 3.5. カナダ | 44 |
| 3.6. オーストラリア | 44 |
| 3.7. 中国 | 45 |
| 3.8. 国際機関 | 45 |
| 3.9. 非政府機関 | 48 |
| 4. その他の ELSI 論点 | 51 |
| 4.1. 悪用 | 52 |
| 4.2. 偽情報（幻覚） | 53 |
| 4.3. バイアス・ステレオタイプ | 55 |
| 4.4. 著作権 | 57 |
| 4.5. デュアルユース | 59 |
| 4.6. 労働者の搾取 | 60 |
| 4.7. 人類破滅リスク | 60 |
| 5. 生成 AI のためのデータセット | 61 |
| 5.1. 画像生成 AI モデル | 61 |
| 5.2. テキスト生成 AI モデル | 62 |
| 5.3. ウェブスクレイピングの是非 | 65 |
| 6. おわりに | 66 |

本 ELSI NOTE は、2023 年 4 月に公表した『生成 AI（Generative AI）の倫理的・法的・社会的課題（ELSI）論点の概観：2023 年 3 月版』¹の続編という位置づけであり、**2023 年 4 月から 8 月半ばまでの情報をまとめたものである**。これらの情報の多くは時系列に並べてある。また、それぞれの情報については**必ず出所（URL）を明記してあるので引用する際には念のためソースを確かめることを推奨する**。

はじめに：グローバルなリスクガバナンスの構図

生成 AI の登場によって、AI の安全性を巡るルール作り、すなわちリスクガバナンスの議論が盛んになり、主導権を競う構図ができあがりつつある。主要なアクターとしては、G7 議長国である日本、英国、欧州（EU）、米国、中国に加えて、国際機関、生成 AI のプロバイダーとしての大手テック企業、グローバルに活動する NPO（消費者団体、プライバシー保護団体、学術団体）などが挙げられる。代表的なものを以下にまとめた。詳細は 3 章に記載されている。

<日本>2023 年 5 月に日本の広島で開催された G7 サミットの首脳声明では、「責任ある AI」の推進に向けて、「広島 AI プロセス」の創設が盛り込まれた。実際に実務者レベルの作業部会が 5 月末に開始され、閣僚会議を経て年内に G7 としての見解をまとめる予定である。

<英国>英国は、規制的アプローチを推進する欧州（EU）とは対照的に、AI に関しては規制に頼らず、自主的な取り組みを中心に進める、すなわちイノベーション促進的なアプローチ（pro-innovation approach）をとることを宣言してきたが、2023 年 6 月、Sunak 首相は英国において AI の安全性に関する「初のグローバルサミット」を開催すると発言し、軌道修正を試みているように見える。

<欧州（EU）>これまで AI の規制を牽引してきた欧州（EU）では、2022 年 6 月に AI Act の議会修正版が公表され、理事会と委員会を合わせた三者協議（trilogue）による合意に年内に到達することが見込まれている。しかし、施行までには時間がかかるため、自主的な行動規範も策定しようとしている。

<米国>米国ではこれまで拘束力のない「AI 権利章典のための青写真」と自主的な「AI リスク・マネジメント・フレームワーク」を中心に議論が進められてきたが、大手テック企業によ

¹ <https://doi.org/10.18910/90926>

る自主的取り組みを中心とする共同規制的なアプローチをとりつつ、AI の安全性に関する大統領令を策定中であり、さらには、米国が責任あるイノベーションを先導できるよう、超党派の法案が提出される見込みである。

米国と欧州（EU）は、米国・EU 貿易技術評議会（Trade and Technology Council, TTC）の場で自主基準としての AI 行動指針（Code of Conduct）を早期に作成すべく議論を開始しており、最初に主要概念の定義集を作成している。

< **中国** > 中国は世界で初めて生成 AI 規制を導入した。生成 AI サービスの管理措置の原案が 2023 年 4 月に公表され、7 月に最終案が出て、8 月から施行されている。国家主権を脅かすようなコンテンツを禁止する文言が入っていることが特徴である。中国をグローバルな AI 規制枠組みにどのような形で組み入れるかがグローバル・ガバナンスの鍵となる。

< **国際機関** > これら以外に、複数の国際機関やネットワークもリスクガバナンスを主導しようとしている。経済協力開発機構（Organisation for Economic Co-operation and Development, OECD）は、生成 AI のリスクをカバーできるように、「AI に関する OECD 原則」を更新する予定である。世界保健機関（World Health Organization, WHO）、国連教育科学文化機関（United Nations Educational, Scientific and Cultural Organization, UNESCO）、国連安全保障理事会なども生成 AI のリスクガバナンスの提言を行っている。

< **GPAI** > 12 月にはインドにおいて「AI に関するグローバル・パートナーシップ（Global Partnership on Artificial Intelligence, GPAI）」の会合も予定されている。

第 1 章では生成 AI 事業者の動向、第 2 章では、これらの動向に対する権利者団体の反応と主な訴訟について、第 3 章ではグローバルな観点から各国・地域の政府、国際機関、非政府機関による政策や提案を詳細にまとめた。第 4 章では、その他の ELSI 論点について、第 5 章では生成 AI モデルのトレーニング・データセットに関してこれまで分かっていることがらをまとめた。

1. 生成 AI 事業者の動向

第1章では生成 AI プロバイダーの動向をまとめた。1.1 節はこの分野をリードしている OpenAI 社の動向をまとめた。1.2 節は Google 社や Meta 社を含む他社の動向をまとめた。1.3 節には生成 AI プロバイダーが公表した政策文書や自主的取組についてまとめた。1.4 節では生成 AI モデル間の比較が行われた研究を紹介する。この研究には含まれていないが Adobe 社のように著作権問題に取り組むことで差別化を図る例も出てきた。

1.1. OpenAI 社の動向

- OpenAI 社は「バグ報奨金プログラム」を開始することを 2023 年 4 月 11 日に発表した。Bugcrowd 社と連携し、Open AI 社のシステムで発見された脆弱性、バグ、またはセキュリティ上の課題を報告したプログラム参加者に報酬が与えられる取り組みである²。
- OpenAI 社は 4 月 24 日、ブランドに関するガイドラインを公表した³。用語については、「**GPT」は使用できず、「** powered by / built on / developed with GPT」等の表記に必要があるとしている。
- OpenAI 社は 4 月 25 日、ChatGPT で履歴をオフにする機能を導入したことを発表した⁴。オフにすると、会話履歴はモデルの訓練と改善に利用されなくなる。不正使用を監視するためのみの理由で 30 日間保存される。
- OpenAI 社の CEO である Altman 氏は 5 月 24 日、EU による AI 規制を遵守できない場合は欧州で「事業を停止する」可能性があるとして述べた⁵。しかし、Altman 氏はその後、勧告を撤回し、OpenAI 社は欧州に協力するつもりであり、離脱する予定はないと主張した⁶。
- OpenAI 社の Altman 氏は複数の国を訪問し、生成 AI の規制の必要性を訴えた。しかし、OpenAI 社が欧州の AI Act を通じた規制を阻止するロビー活動に積極的に関与していたことが

² <https://openai.com/blog/bug-bounty-program>

³ <https://openai.com/brand>

⁴ <https://openai.com/blog/new-ways-to-manage-your-data-in-chatgpt>

⁵ <https://time.com/6282325/sam-altman-openai-eu/>

⁶ <https://twitter.com/sama/status/1672460302712193025>

6月20日にTIME誌で報道された⁷。特に、本物のように見えながら偽物のテキストや画像を生成する生成AIシステムを高リスクに分類するAI Actの改正案に反対したことが明らかになった。また、生成AIを「基盤モデル（foundation models）」といった別のカテゴリーとして定義づけられることを求めた。

●OpenAI社はヒト型ロボットの研究開発を行っている1X Technologies社（ノルウェー）に2,350万ドル出資した⁸。同社は、既に米国、欧州で利用されているEVEやAI搭載ロボットのNEOの製造を進めている。

●OpenAI社が2023年1月に発表した、AI生成の文章かどうかを判別するツール「AI Text Classifier」を「精度が低い」という理由で7月20日にひそかに終了していたことが報道された⁹。

●OpenAI社はAP通信と7月13日、ニュース製品やサービスにおける生成AIの潜在的な使用事例を検討するため、一部のニュースコンテンツと技術へのアクセスを共有することで合意に達した¹⁰。

●OpenAI社は、GPT-4はテキストだけでなく画像の入力も受け付ける「マルチモーダルAI」として開発されていたが、画像認識機能が特定の個人を識別するとプライバシー侵害になる懸念から、顔認識機能に制限をかける措置を取っていることが7月18日、報じられた¹¹。

OpenAI社は、Be My Eyes社というスタートアップと共同で、画像認識技術を使って目の不自由なユーザーに画像を説明するアプリを開発していたが、最近アプリが顔情報を提供しなくなり、プライバシー保護のために顔が隠されているというメッセージが表示されるようになったという。

⁷ <https://time.com/6288245/openai-eu-lobbying-ai-act/>

⁸ <https://techfundingnews.com/openai-backs-1x-technologies-in-23-5m-funding-to-commercialise-humanoid-robotics/>

⁹ <https://gigazine.net/news/20230726-openai-ai-text-classifier-shut-down/>

¹⁰ <https://www.ap.org/press-releases/2023/ap-open-ai-agree-to-share-select-news-content-and-technology-in-new-collaboration>

¹¹ <https://arstechnica.com/information-technology/2023/07/report-openai-holding-back-gpt-4-image-features-on-fears-of-privacy-issues/>

●OpenAI 社は 8 月あたりに、学習データの収集を行う Web クローラー（GPTBot）について、Web サイトの robots.txt を用いて GPTBot のアクセスを許可しないように設定する方法をウェブサイトで公開した¹²。

1.2. OpenAI 社以外の企業の動向

●Hugging Face 社は 2023 年 4 月 26 日、HuggingChat¹³のリリースを発表した。オープンソースであることを特徴とする。LAION（詳しくは 5 章を参照）によって生み出された Open Assistant と名付けられたオープンソースのチャットアシスタントと対話できるユーザーインターフェースを用いている¹⁴。Open Assistant のモデルは、650 億のパラメータを持つ Meta の大規模言語モデル「Llama」をベースとしている（そのため商業的な利用は許可されていない）。

●Google 社は、テキスト生成 AI である Bard が日本語に対応したことを発表した¹⁵。Bard は、Google の大規模言語モデル（LLM）である PaLM 2¹⁶ を利用している。一つの質問に対して、複数の回答案を提示するため、ユーザーは自分のスタイルに合った回答案を複数の候補から選べるのが特徴である。注意事項として下記の点が記載されている¹⁷。

- Bard は試験運用中の技術であるため、生成される情報は不正確または不適切な場合がありますが、Google の見解を述べるものではありません。
- 医学上、法律上、金融上、またはその他の専門的な助言として、Bard の回答に依拠しないでください。
- Bard との会話に機密情報またはセンシティブな情報を含めないでください。

¹² <https://platform.openai.com/docs/gptbot>

¹³ <https://huggingface.co/chat/>

¹⁴ <https://venturebeat.com/ai/hugging-face-launches-open-source-version-of-chatgpt-in-bid-to-battle-openai/>

¹⁵ <https://japan.googleblog.com/2023/05/bard.html>

¹⁶ <https://japan.googleblog.com/2023/05/palm-2.html>

¹⁷ https://support.google.com/bard/answer/13594961?hl=ja#things_to_know

●Google 社は同日、Bard が、カナダと EU 加盟国（27 か国）などを除く 180 の国と地域でアクセス可能となったと発表したと ARS Technica が報じた¹⁸。Google 社は、カナダや EU 加盟国、その他の除外された地域に Bard をまだ導入していない理由を説明しなかった。

●Stability AI 社は 4 月 19 日、新しいオープンソースの言語モデル StableLM をリリースした¹⁹。アルファ版は 30 億パラメータと 70 億パラメータのモデルが用意されているという。開発者は、CC BY-SA-4.0 ライセンスの条件に従って、商用または研究目的で、StableLM ベースモデルを自由に使用することができるとされている。4 月 28 日には「世界初のオープンソース RLHF LLM チャットボット StableVicuna をリリース」したと発表した²⁰。RLHF とは人間のフィードバックからの強化学習を指す。StableVicuna は、Meta 社が開発した LLaMA 13B をベースとするチャットボット AI「Vicuna-13B」をさらに調整して訓練したものである。

●Adobe 社は、3 月 21 日に、Adobe Stock の画像、オープンライセンスコンテンツ、著作権が失効したパブリックドメインコンテンツを利用した「商用利用にも安全なコンテンツを生成するように設計」された生成 AI モデルである Adobe Firefly を発表した²¹。その際に、ベータ版を終了した時点で Adobe Stock のクリエイターへの報酬モデルを開発中であるとしていた。5 月 10 日には、Google 社の Bard に Firefly が導入されたことがアナウンスされた²²。同社はウェブメディア VentureBeat に対し、商用利用しても安全であることを確信しているものの、現在 Adobe Stock で実施しているのと同様に、企業顧客に対して、万が一著作権の侵害で訴えられた場合、Adobe 社が法的訴えを引き受け、それらの請求に対していくらか金銭的な補償を提供する予定であると語った²³。6 月 8 日には、企業など法人向けの Adobe Firefly エンタープライズ版を発表した²⁴。7 月 12 日には、Firefly のウェブサービスのグローバル展開として、

¹⁸ <https://arstechnica.com/gadgets/2023/05/google-bard-hits-over-180-countries-and-territories-none-are-in-the-eu/>

¹⁹ <https://ja.stability.ai/blog/stability-ai-launches-the-first-of-its-stablelm-suite-of-language-models>

²⁰ <https://ja.stability.ai/blog/stablevicuna-open-source-rlhf-chatbot>

²¹ <https://news.adobe.com/news/news-details/2023/Adobe-Unveils-Firefly-a-Family-of-new-Creative-Generative-AI/default.aspx> 日本語版の FAQ において「Firefly の学習には、Adobe Stock の画像、オープンライセンスのコンテンツ、著作権の切れた一般コンテンツが使用されました」と書かれている。

²² <https://blog.adobe.com/en/publish/2023/05/10/adobe-firefly-adobe-express-google-bard>

²³ <https://venturebeat.com/ai/adobe-brings-firefly-commercially-safe-image-generating-ai-to-the-enterprise/>

²⁴ <https://news.adobe.com/news/news-details/2023/Adobe-Brings-Firefly-and-Express-to-Enterprises/default.aspx>

100 言語以上のテキストプロンプトをサポートするとともに、さらに 20 か国語にローカライズされ、フランス語、ドイツ語、日本語、スペイン語、ポルトガル語版が利用可能になったことが発表された²⁵。

●Google 社は 7 月 13 日、ようやく欧州で Bard のサービスを開始した²⁶。EU 加盟 27 か国とブラジルを含む 50 か国で新たに利用可能になった。これは Bard のこれまでで最大のアップデートに合わせたものであった。6 月には Google 社は、EU 一般データ保護規則（General Data Protection Regulation, GDPR）をどのように遵守する予定なのか、アイルランド・データ保護委員会（Data Protection Commission, DPC）に詳細に説明することができなかったため、欧州での Bard の発売延期を余儀なくされていた。当局との対話の結果、また、Google 社は、Bard が EU で運用開始されてから 3 ヶ月後にレビューを実施し、DPC に報告書を提出することに同意しているという。

●ストックフォト企業の大手である Shutterstock 社は、2023 年 1 月に OpenAI 社の DALL-E 2 エンジンを搭載した画像ジェネレーターを発表した²⁷。Shutterstock の画像ライブラリでカスタムトレーニングされ、そのプロセスで使用された作品のアーティストにロイヤリティが支払われる。7 月、同社の生成 AI ツールを使用する企業顧客に対して、ライセンス契約の定型文²⁸に補償条項を追加し、著作権やその他の法的な不正使用に関するユーザーからの法的請求を引き受けることを約束した²⁹。

●Meta 社は 6 月 18 日、大規模言語モデル（LLM）の第 2 バージョン、Llama 2（ラマ 2）を発表した³⁰。これまでのバージョンと異なり、商用利用が自由にできることが特徴である。パラメータの数が異なる 3 種類（7B、13B、70B）が公表された。

²⁵ <https://news.adobe.com/news/news-details/2023/Adobe-Firefly-Expands-Globally-Supports-Prompts-in-Over-100-Languages/default.aspx>

²⁶ <https://www.euractiv.com/section/artificial-intelligence/news/google-launches-generative-ai-model-bard-in-europe/>

²⁷ <https://voicebot.ai/2023/01/25/shutterstock-launches-ai-text-to-image-generator-powered-by-openais-dall-e/>

²⁸ <https://www.shutterstock.com/license>

²⁹ すでに Adobe 社が Firefly について同様の発表を行っている。ライバルである Getty Images 社とは対照的なアプローチである。

³⁰ <https://ai.meta.com/llama/>

●Google 社は 2023 年 1 月に音楽をテキストから自動的に生成する AI モデル、MusicLM を発表し³¹、5 月 10 日から AI Test Kitchen に登録した人が順次試すことが可能となった³²。2 つのバージョンの楽曲が作成され、気に入った方を選ぶことでモデルが改良されるという。モデル自体は公開されていない。MusicCaps データセットは公開されており（ライセンスは CC BY-SA 4.0）、ミュージシャンから提供された 5521 の音楽が含まれており、それぞれに英語の aspect list（例：ハイピッチの女性ボーカルメロディ）とミュージシャンが書いたフリーテキストのキャプションが付けられている³³。1 月に公開された論文の第 8 節 “Broader Impact” にはモデルとユースケースのリスクとして、生成されたサンプルがトレーニングデータに存在するバイアスを反映するために、トレーニングデータに十分に代表されていない文化に関する音楽生成の適切さに関する疑問を生じさせると同時に、文化の盗用についての懸念も引き起こしかねないことが指摘された。また、ユースケースについても、創造的なコンテンツの潜在的な不正流用のリスクを認識しているとした。そして、これらの音楽生成のリスクに対する今後の研究の必要性を強く主張するとともに、現時点ではモデルを公開する予定はないとした。

1.3. 生成 AI 事業者による政策提案や自主的取組

●OpenAI 社は 2023 年 4 月 5 日、“Our approach to AI safety”と題する文書を公表し、生成 AI システムの安全性をどのようにして確保しており、今後さらに確保する予定であるか説明した³⁴。新しいシステムのリリース前には、内部での厳格なテストに加えて、外部の専門家にフィードバックを求め、人間のフィードバックによる手法でモデルの動作を改善するという体制を構築しているとする。例えば、GPT-4 の場合、トレーニング終了後、一般公開前に、組織全体で 6 カ月以上かけて安全性のチェックが実施されたという。また、個人情報については、「可能な限りトレーニング・データセットから個人情報を削除し、個人情報の要求を拒否するよう

³¹ <https://google-research.github.io/seanet/musiclm/examples/>

³² <https://blog.google/technology/ai/musiclm-google-ai-test-kitchen/>

³³ <https://www.kaggle.com/datasets/googleai/musiccaps>

³⁴ <https://openai.com/blog/our-approach-to-ai-safety>

にモデルを微調整し、システムからの個人情報の削除要求に対応するように努めています。」という。

●OpenAI 社は 5 月 22 日付の自社ブログにおいて、Altman 氏ら 3 名の署名記事「スーパーインテリジェンスのガバナンス」を公表した³⁵。今後 10 年以内に、ほとんどの領域で AI システムが専門家の技量を超え、現在の大企業に匹敵する生産活動を行うようになることが考えられるとしたうえで、必要となることとして、すべての企業が協調して AI 能力の成長を一定に制御することに合意すること、スーパーインテリジェンスに対する、IAEA（International Atomic Energy Agency, 国際原子力機関）のような監査や試験を実施できる国際機関を設けること、スーパーインテリジェンスを安全なものにするための技術的能力を研究開発することを挙げた。

●Microsoft 社は 5 月 25 日、「AI を統治する：未来へのブループリント」と題する 40 ページの報告書を公表した³⁶。第一部では、政府が AI に関する政策、法律、規制を検討すべき 5 つの方法が提示された。

1. 政府主導の新しい AI 安全枠組み（=NIST の AI リスク・マネジメント・フレームワーク（AI RMF）³⁷）を施行し、またその上に構築する
2. 重要インフラを制御する AI システムに効果的な安全ブレーキを要求する
3. AI のための技術アーキテクチャに基づく広範な法的及び規制のフレームワークを開発する
4. AI の透明性を促進し、学術および一般の人々からのアクセスを確保する
5. 新しい技術に付随する不可避な社会的課題に対処するための効果的ツールとして AI を使用するための新しい官民パートナーシップを追求する

第 2 部では、Microsoft 社の AI に対する社内の取り組みに焦点を当て、同社が責任ある AI の運用と文化の構築の両方に取り組んでいることが紹介された。

³⁵ <https://openai.com/blog/governance-of-superintelligence>

³⁶ <https://blogs.microsoft.com/on-the-issues/2023/05/25/how-do-we-best-govern-ai/>

³⁷ <https://www.nist.gov/itl/ai-risk-management-framework>

●OpenAI 社は 7 月 6 日、「フロンティア AI 規制：公共の安全に対する新たなリスクの管理」と題する未査読の論文（プレプリント）を公表した³⁸。公共の安全に深刻なリスクをもたらすのに十分な危険な能力を保有する可能性のある、非常に能力の高い基盤モデルを「フロンティア AI（frontier AI）」と呼び、自主規制だけでは人々を十分に保護できる可能性は低いため、政府による介入が必要になると主張している。介入の仕方としては、安全性基準を策定し更新し続ける仕組み、規制当局に情報が可視化される仕組み、安全性基準への準拠が確保される仕組みが挙げられた。次に、フロンティア AI の開発のための初期の安全性基準として次のような事項を挙げた。

- 危険な能力と制御可能性についての評価の結果を用いた包括的なリスクアセスメントを実施する。
- モデルに対して独立した精査を実施するために外部専門家に参加してもらう。
- フロンティア AI モデルが、リスクアセスメント結果に基づき、どのようにして実装できるかについて、標準化されたプロトコルに従う。
- モデルの能力に関する新しい情報を常にモニターし、すぐに対応する。

●7 月 26 日、Anthropic 社、Google 社、Microsoft 社、OpenAI 社の 4 社は、フロンティア AI（frontier AI）モデルの安全で責任ある開発を確保することに焦点を当てた新しい業界団体、フロンティア・モデル・フォーラム（Frontier Model Forum）の設立を発表した³⁹。フォーラムの基本的な目的は以下の通りである：

1. フロンティアモデルの責任ある開発を促進し、リスクを最小化し、能力と安全性についての独立した標準化された評価を可能にするために、AI の安全性研究を推進する。
2. フロンティアモデルの責任ある開発と実装のためのベストプラクティスを特定し、一般市民が技術の性質、能力、限界、影響を理解できるようにする。
3. 政策立案者、学識経験者、市民社会、企業と協力し、信頼と安全性に関するリスクに関する知識を共有する。

³⁸ <https://openai.com/research/frontier-ai-regulation>

³⁹ <https://blog.google/outreach-initiatives/public-policy/google-microsoft-openai-anthropic-frontier-model-forum/>

4. 気候変動の緩和と適応、がんの早期発見と予防、サイバー脅威との闘いなど、社会が直面する最大の課題への対応に役立つアプリケーションの開発努力を支援する。

1.4. 生成 AI モデルの比較

●Stanford 大学の研究者らによる研究「基盤モデルのプロバイダーは EU の AI Act 案に準拠しているか？」が 2023 年 6 月 15 日、公開された⁴⁰。10 種類の基盤モデルについて、データソース、計算リソース、モデル自体、展開方法という 4 つのカテゴリーから計 12 項目（もともと抽出されたのは 22 項目）についてそれぞれ欧州議会による修正案に基づき 5 段階（0～4 点）で評価された。満点は 48 点で、最高点は Hugging Face 社の Bloom で 36 点（75%）であった。OpenAI 社の GPT-4 は 25 点、Meta 社の Llama は 21 点、Google 社の PaLM2 は 27 点であった。多くの基盤モデルで 0 または 1 点であった分野が 4 つあり、(i)著作権で保護されたデータ、(ii)計算／エネルギー、(iii)リスク／軽減、(iv)評価／試験であった。

2. 権利者団体の反応と訴訟

第 2 章では、第 1 章で挙げた生成 AI 事業者の迅速な事業展開に対して、権利者団体側の反応を取り上げる。各種の権利者団体から多くの提言が発表されている。画像生成 AI システムに対しては訴訟が米国と英国ですでに 2 件起こされていたが、新たにテキスト生成 AI システムに対しても複数の訴訟が起こされている。

2.1. 国内

●イラストレーターや漫画家などおよそ 30 人で作る「クリエイターと AI の未来を考える会」⁴¹が 2023 年 4 月 27 日に記者会見を開き、提言を発表した⁴²。提言文書は「画像生成 AI の適正

⁴⁰ <https://crfm.stanford.edu/2023/06/15/eu-ai-act.html>

⁴¹ <https://support-creators.com/>

⁴² <https://www3.nhk.or.jp/news/html/20230427/k10014051061000.html>

使用及びそれに伴う著作権制度の整備等に関する提言」というもので、2月17日付で第1版を、4月15日付で第2版を作成・公表した。第2版における提言は以下の5項目からなる。

1. Stable Diffusion に代表される画像生成 AI の機械学習における著作物の使用は、著作権法 30 条の 4 に規定する著作物の権利制限の対象外とすること。

・画像生成 AI の学習における著作物の使用は著作権の原則に従いオプトイン方式とし、著作物から著作物の使用許可を事前に得ること。

2. 画像生成 AI の使用において、AI の機械学習に使用した著作物の著作者に対し、学習への使用及びその AI が消費者に使用された回数等に応じた使用料を支払うこと。

3. 著作権は、これまで通り、思想または感情の創作的表現に与えること。

・画像生成 AI 生成物においては、すべて又は大部分が AI 生成物である制作物を著作権の保護の対象とせず、創作的寄与が明確に認められるものののみ保護の対象とすること。

4. 画像生成 AI 生成物は、AI 生成物であること及びその起源の明示を義務付けること。

5. 画像生成 AI を、著作権法にとどまらず、人権侵害や安全保障を含む幅広い観点でリスク評価し、国際的な枠組みに沿って適切に規制すること。

●5月8日、日本芸能従事者協会は会見を開くとともに、生成 AI 技術の開発にあたって肖像や声、演技に関する権利なども法律で明文化すること、学習データの内容を開示する義務、クリエイターへの対価の支払い義務などを求める要望書を国に提出した⁴³。また、5月9日には UNI Global Union（ユニグローバルユニオン）「メディア・芸術・芸能部門」から日本政府宛てに「AI の急速な出現に関する日本芸能従事者協会との連帯声明」を提出したことを発表するとともに、「AI リテラシーに関する全クリエイターのアンケート」を開始した⁴⁴。

●5月17日、一般社団法人日本新聞協会は「生成 AI による報道コンテンツ利用をめぐる見解」を公表した⁴⁵。論点として、1) 言論空間の混乱と社会の動揺、2) 個人情報保護上の懸

⁴³ <https://www3.nhk.or.jp/news/html/20230509/k10014061001000.html>

⁴⁴ <https://artsworkers.jp/news/ai-statement-202305/>

⁴⁵ <https://www.pressnet.or.jp/statement/20230517.pdf>

念、3) 現行著作権法や法改正に至る過程の問題点、4) 報道機関の著作物等をめぐる課題、5) 不透明な運用実態、権利者への不十分な情報開示、が挙げられた。

●日本芸能従事者協会は6月8日、クリエイターらに実施したアンケートの最終結果を公表した⁴⁶。アンケートはインターネット上で5月8～28日に行い、2万6891人が回答した。

●日本俳優連合は5月10日付で「AI対策プロジェクト」を発足させ⁴⁷、6月13日には「生成系AI技術の活用に関する提言」を公表した⁴⁸。提言は下記の5点である。

- 国内外での意見交換を活発に行うとともに、EUによるAI ACTの考え方に大いに賛同し、これを参考にしたガイドラインの策定を行う
- 学習素材は著作者が許可を与えたもののみを使用可能として、著作権法新30条4の運用の見直しを諮る
- 機械学習の結果によるアウトプットのイラスト、写真、映像、音声など全てのものにAIによる生成物であることを必ず明示する
- AIの「表現」分野への進出については、一定のルールを設ける。具体的には、人間の代替としてのAIによる表現※をしてはならないと規定する(※ただし情報の伝達は可、福祉的な利用はこの限りではない)
- 「声の肖像権」の設立を目指す。これまでは声と表現が切り離されて使われることが想定されていなかったため、外見だけではない「声の肖像権」の新設を急ぐ

続いて、8月8日、俳優や声優といった実演家に対して「生成AIに関する緊急アンケート（無記名方式・全15問）」を開始した⁴⁹。アンケートは、一般社団法人日本芸能マネジメント事業者協会の協力のもとで実施している。締め切りは8月27日である。

●一般社団法人 日本音楽作家団体協議会（Japan Federation of Authors and Composers Association, FCA）は6月15日、「AIによる著作物利用に関するFCAの見解」を公表した

⁴⁶ <https://artsworkers.jp/questionnaire/20230608/>

⁴⁷ <https://www.nippairen.com/ja/news/post-12520.html>

⁴⁸ <https://www.nippairen.com/about/post-14576.html>

⁴⁹ <https://www.nippairen.com/ja/news/post-17502.html>

⁵⁰。FCA は、「現行著作権法における権利制限規定を見直して、創作者の権利を阻害することなく AI 技術の発展と調和を図ること」、「G7 デジタル・技術閣僚宣言に基づき早急に生成 AI に関する議論の場を設け、創作者をステークホルダー（利害関係者）の一員として協議に参加させること」を政府に求めた。

●一般社団法人 日本音楽著作権協会（Japanese Society for Rights of Authors, Composers and Publishers, JASRAC）は 7 月 24 日、7 月 5 日の理事会での決議に基づき、次の 4 項目からなる「生成 AI と著作権の問題に関する基本的な考え方」を公表した⁵¹。

- 1 人間の創造性を尊重し、創造のサイクルとの調和を図ることが必要です。
- 2 フリーライドが容認されるとすればフェアではありません。
- 3 AI には国境がないので、国際的な調和を確保すべきです。
- 4 クリエーターの声を聴き、懸念の解消を図るべきです。

●8 月 17 日、一般社団法人日本雑誌協会、一般社団法人日本写真著作権協会、一般社団法人日本書籍出版協会、一般社団法人日本新聞協会は連名で、「生成 AI に関する共同声明」を公表した⁵²。生成 AI は、著作権法が目的とする文化の発展を阻害する恐れがあるとして、第 30 条の 4 ただし書きの解釈を明確にし、著作権法改正の必要性を見極める必要があるとした。

●同日、一般社団法人 日本美術著作権連合⁵³は「生成 AI に関する声明」を公表した⁵⁴。生成 AI の利用に関する何らかのルール、例えば、「学習対象とされることに対するオプトアウト、データセットに関する透明性の確保、AI の生成物であることの表示、学習や利用などの各段階におけるガイドラインの制定など」が作成されるべきであるとした。

⁵⁰ <https://fca-rights.jp/>

⁵¹ https://www.jasrac.or.jp/release/23/07_3.html

⁵² https://www.pressnet.or.jp/statement/copyright/230817_15114.html

⁵³ 所属団体は、一般社団法人日本美術家連盟、公益社団法人日本グラフィックデザイン協会、一般社団法人日本児童出版美術家連盟、一般社団法人日本理科美術協会、一般社団法人日本出版美術家連盟、一般社団法人東京イラストレーターズ・ソサエティ、一般社団法人日本図書館設計家協会である。

⁵⁴ http://www.jart.tokyo/jart_wp/wp-content/uploads/2023/08/5bbd651f0394f60cbdbf80279ca1b5ac.pdf

2.2. 米国

●アーティスト、パフォーマー、作家、アスリートなどの 40 以上の団体が参加し、人間の創造性と達成感を支援する AI を求めるキャンペーン「Human Artistry Campaign」が開始され、オースティンで開催された SXSW（South by Southwest）⁵⁵の場で 2023 年 3 月 16 日に発表された。キャンペーンは同時に、人間の創造性と達成感を支援するための AI 応用のコア原則をまとめた⁵⁶。

1. 技術は古くから人間の表現力を高めてきており、AI も同様であるだろう。
2. 人間が創り出した作品は、これからも私たちの生活に欠かせない存在であり続けるだろう。
3. 著作権で保護された作品の使用、およびプロの演奏家の声や肖像の使用には、認可、ライセンス、および関連するすべての州・連邦法を遵守する必要がある。
4. 政府は、AI 開発者が許可や補償なしにクリエイターを搾取できるような、新たな著作権やその他の知的財産の免除を創設すべきではない。
5. 著作権は、人間の知的創造性のユニークな価値のみを保護すべきである。
6. AI の成功やクリエイターの保護には信頼性（Trustworthiness）と透明性が不可欠である。
7. 政策形成において、クリエイターの利益が代表されなければならない。

Human Artistry Campaign の開始は大きな反響を呼び、ジャーナリズム、写真、声優、そしてソングライター、作曲家、出版社など多様な団体が新規に加わっている。

●米国最大の新聞業界団体である News/Media Alliance は 4 月 20 日、生成 AI システムがジャーナリズムやクリエイティブなコンテンツを利用する際の包括的なガイダンスを提供する新しい AI 原則を発表した⁵⁷。生成 AI 開発者が会員のコンテンツを利用する際の、（パブリッシャー

⁵⁵ サウス・バイ・サウス・ウエスト。毎年 3 月にテキサス州オースティンで開催されるテクノロジーとアートの祭典。

⁵⁶ <https://www.humanartistrycampaign.com/>

⁵⁷ <https://www.newsmediaalliance.org/release-news-media-alliance-releases-new-generative-artificial-intelligence-ai-principles/>

の) 知的財産 (の保護と補償)、 (パブリッシャーとユーザー双方への) 透明性、 (アウトプットに対する) アカウンタビリティ、 (競争上の) 公平性、 (プライバシーについての) 安全性、 設計 (への組み込み) に関する問題を取り上げている。関係者は「出版社は、そのコンテンツが生成 AI 技術の開発に貢献する多大な価値に対して、公正に補償されなければなりません。これは単純な価値の交換です。」と発言している。

●全米脚本家組合 (Writers Guild of America : WGA) は、5 月 2 日からストライキを開始した⁵⁸。この決定は、映画テレビ制作者協会 (Alliance of Motion Picture and Television Producers, AMPTP) 傘下の Netflix 社、Amazon 社、Apple 社、The Walt Disney Company 社、Warner Bros. Discovery 社、NBC Universal 社、Paramount 社、Sony 社の各社との 6 週間にわたる交渉の末に下された。AI の利用を制限する拘束力のある合意を求めていたが拒否されたことも交渉決裂の原因となったという。

- 映画テレビ制作者協会 (AMPTP) は 8 月 23 日、脚本家組合に対する 8 月 11 日提案の詳細を発表した⁵⁹。挙げられた項目の中には、「生成 AI の使用をめぐる脚本家のための画期的な保護措置」として、生成 AI によって製作された脚本は文学的素材とみなされないこと、脚本の一部が生成 AI によって制作された場合でも脚本家の報酬などが影響を受けないことが挙げられた。

●アメリカ SF ファンタジー作家協会 (Science Fiction and Fantasy Writers Association, SFWA) 理事会は 6 月 13 日、AI/ML (機械学習) 利用に関するステートメントを公表した⁶⁰。

1. クリエイターは、その作品の使用に対して補償されなければならない。
2. 作品に対するクリエイターの貢献は、クレジットされなければならない。
3. クリエイターのプライバシーは、特に未発表の作品については保護されなければならない。
4. 大衆小説の執筆と出版は、重要な規範を持つビジネスである。

⁵⁸ <https://www.wga.org/news-events/news/press/writers-guild-of-america-calls-strike-effective-tuesday-may-2>

⁵⁹ <https://press.amptp.org/node/19237>

⁶⁰ <https://www.sfwawriters.org/2023/06/13/current-statement-on-ai-ml-use/>

●映画俳優組合-米国テレビ・ラジオ芸術家連盟（SAG-AFTRA, Screen Actors Guild - American Federation of Television and Radio Artists）は 7 月 13 日、理事会を開催し、14 日午前 0 時からストライキを実施することを決め、すべての SAG-AFTRA 組合員に対し、裏方仕事も含む、TV／演劇契約に基づくすべての業務の提供および遂行を中止するよう指示した⁶¹。ストライキは 1980 年以來。SAG-AFTRA はハリウッドの有名俳優ら約 16 万人が加入している。映画テレビ制作者協会（AMPTP）との交渉は 6 月 7 日に始まったが、交渉は決裂した。AI については、「クリエイティブな職業に存亡の危機を投げかけており、すべての俳優とパフォーマーは、同意と報酬なしに彼らのアイデンティティと才能を搾取されることから彼らを保護するという契約文言を得るに値する」とした⁶²。

●米国の作家のための非営利団体である Authors Guild（全米作家協会）は、OpenAI 社、Alphabet 社、Meta 社、Stability AI 社、Microsoft 社の各 CEO に対し、AI のトレーニングのために、同意の取得、クレジットの付与、公正な補償なしに著作物を使用しないよう求めるオープンレターを作成し、これに 1 万人以上の作家が署名し、7 月 17 日に送付した⁶³。具体的に要求している措置は次の 3 つである。

- 課題 1：AI が人間の著作物を模倣し、同意、補償、クレジットを得ずに利用する。
⇒提案：AI 開発者は、生成 AI において著作物の使用許可を得るべきである。
- 課題 2：生成 AI は人間の著作物を模倣するために、同意、補償、クレジットなしに訓練されてきた。
⇒提案：AI 開発者は、"in-the-style of "（誰々のスタイルでの）作品については許可を求め、対価を支払うべきである。
- 問題 3：AI が生成した作品が人間の作品と競合することによる市場の希釈化
⇒提案：AI が生成した作品には明確にラベルを付けるべき。AI が生成した作品は著作権による保護を受けるべきではない。AI 企業はトレーニングデータの開示を義務付けられるべき。

⁶¹ <https://www.sagaftra.org/sag-aftra-strike-order-tvtheatricalstreaming-contracts-0>

⁶² <https://www.sagaftra.org/message-sag-aftra-president-and-chief-negotiator>

⁶³ <https://authorsguild.org/news/thousands-sign-authors-guild-letter-calling-on-ai-industry-leaders-to-protect-writers/>

2.3. その他

●欧州の Authors' Rights Initiative に参加する 43 の協会や労働組合が 2023 年 4 月 19 日付で、欧州 AI Act への要望を共同声明として公表した⁶⁴。「生成 AI」と「一般目的 AI (General-purpose artificial intelligence, GPAI)」が区別されるべきこと、生成 AI は製品サイクル全体にわたり、特に基盤モデルのプロバイダーに焦点を当てて規制されるべきことを挙げたうえで、生成 AI のための基盤モデルを欧州市場に投入する際にプロバイダーが満たすべき要件を次のように挙げた。

- 使用されるトレーニング素材の完全な透明性
- トレーニング素材が真実性、正確性、客観性、多様性の観点から十分なレジリエンスがあり、特にトレーニング素材全体に占める十分な割合が欧州起源かつ専門的な情報源であることを示す文書
- 個人データと非個人データの両方について、トレーニング素材の収集・使用の法的根拠があるというエビデンス
- AI が生成・普及したすべてのコンテンツ、特に個人の権利や著作権の侵害、誤情報、差別に対する責任
- AI 生成コンテンツを人間生成コンテンツよりもアルゴリズムやその他の方法で推奨したり、後者を中傷したりしないこと。また、ユーザーが AI コンテンツに過度に依存しないための合理的措置
- AI アウトプットの生成と普及を構造的に分離（基盤モデルのプロバイダーは同時にプラットフォームサービスを運営してはならない）
- AI システムの推論の最低限は、欧州にあるコンピューティングインフラで実行されなければならない、国内のデータ処理の割合を徐々に増加させる。

●Center for Artistic Inquiry and Reporting から 5 月 2 日、「AI イラストレーションの出版を制限せよ」と題するオープンレターが発表され、署名が呼びかけられた⁶⁵。クリエイターに対

⁶⁴ <https://urheber.info/diskurs/call-for-safeguards-around-generative-ai>

⁶⁵ <https://artisticinquiry.org/AI-Open-Letter>

する補償や同意なしに著作権で保護されているものを含む画像が大量に活用されることを「史上最大の美術品強盗」と呼び、出版物における AI イラストの使用の制限を求めている。

●8月9日、多くの報道機関・団体（フランス通信（AFP）、EPA 通信、欧州出版社評議会、ガネット社（USA TODAY ネットワーク）、Getty Images、全米報道写真家協会、全米著述業組合、News/Media Alliance、AP 通信社、Authors Guild（全米作家協会））が署名したオープンレター「統一された AI 規制と実践を通じてメディアに対する国民の信頼を守る」を公表した⁶⁶。次のような規制あるいは産業界としての措置を求めた。

- AI モデルの作成に使用されるすべてのトレーニングセットの構成に関する透明性。
- トレーニングデータと出力におけるコンテンツの使用とコピーに対する知的財産権者の同意。
- メディア企業が、AI モデルの運営者や開発者と、運営者の知的財産へのアクセスや使用の条件について集团的に交渉できるようにすること。
- 生成 AI モデルおよびユーザーに対し、そのアウトプットやインタラクションに AI が生成したコンテンツが含まれることを明確、具体的、かつ一貫性をもって明示することを義務付けること。
- 生成 AI モデルのプロバイダーに対し、そのサービスにおけるバイアスや誤情報を排除するための措置を講じるよう求めること。

2.4. 訴訟

●2023年6月28日付で、カリフォルニア州北部地区連邦地方裁判所に、イニシャルで表された16人が同様の他の大勢を代表するという形で、OpenAI 社や Microsoft 社を相手に集団訴訟を起こした⁶⁷。OpenAI 社のテキスト生成 AI と画像生成 AI の両方を対象としており、パーソナルデータに財産権を認めたうえで、インターネットのスクレイピングによって同意も通知もなく勝手に収集することで、パーソナルデータの窃盗（theft）により不当に利益を得ているとし

⁶⁶ <https://drive.google.com/file/d/1jONWdRbwbS50hd1-x4fDvSyARJMCgRTY/view>

⁶⁷ <https://www.vice.com/en/article/wxjgx/openai-and-microsoft-sued-for-dollar3-billion-over-alleged-chatgpt-privacy-violations> 本文は、
https://storage.courtlistener.com/recap/gov.uscourts.cand.414754/gov.uscourts.cand.414754.1.0_1.pdf

て、不当利益の返還や補償を求めた。第Ⅱ章では一般的リスクが列挙され、第Ⅲ章では被告の行為が財産権とプライバシー権を侵害している点が述べられ、第Ⅳ章では子どものプライバシーと安全上のリスクが取り上げられた。これらは 15 の法律に違反しているとしたうえで、これらに基づき、各種損害賠償額に加えて、製品（＝ChatGPT 等）への商業的アクセスおよび商業的開発を一時的に凍結するという形で、被告が以下の一部または全部の完了を裁判所が満足する形で証明できるまでの差止を求めた。

1. 独立した組織（AI 協議会）の設立
2. 出力に対する責任を負うこと、またデータに対する公正な補償
3. AI 協議会が効果的なサイバーセキュリティ対策を実施
4. 収集するデータを明確かつ正確に開示
5. オプトアウトの機会を設けることと、不正に取得されたデータとそれらに基づくアルゴリズムの削除
6. 技術が人間の知能を凌駕し、人間に危害を加えることを防止する技術的安全対策
7. 情報ネットワークへの脅威管理プログラム
8. 不正行為を補償するための金銭基金（AI 基金）の設立
9. データの窃盗から得られた利益の公平で公正な「データ配当」のための第三者管理者の任命
10. 保持と使用が正当化されない個人情報削除/破棄/消去されたことを確認
11. さらにあらゆる正当な是正措置を要求すること

●SF 及びホラー作家の Paul Tremblay 氏と小説家の Mona Awad 氏によって 6 月 28 日、カリフォルニア州北部地区連邦地方裁判所に訴訟が起こされた⁶⁸。提出された訴状によると、ChatGPT の学習データセットは、「同意もクレジットも対価もなしに OpenAI 社によってコピーされた」書籍やその他のテキストから来ていると指摘している。OpenAI 社は ChatGPT の学習データセットを公開していないが、2020 年の論文では GPT-3 の学習データセットを紹介しており、トレーニング・データセットの 15%を占めている 2 つの書籍データセットのうちの 1 つ(Books2)は、著作権で保護された何千もの作品を違法に出版する「影の図書館（shadow

⁶⁸ <https://news.bloomberglaw.com/ip-law/openai-facing-another-copyright-suit-over-ai-training-on-novels>

library)」から来ていると考えられると主張している。そして ChatGPT に各原告が書いた本を要約するように指示すると、非常に正確な要約を生成したことから著者らの本が学習データセットに含まれていると確信してたとしている。

●コメディアンで作家の Sarah Silverman 氏、作家の Christopher Golden 氏と Richard Kadrey 氏の 3 人は 7 月 7 日、OpenAI 社の ChatGPT と Meta 社の Llama に著作権を侵害されたとしてそれぞれに対して訴訟を起こした⁶⁹。原告らは、彼らの著作の海賊版がウェブサイト上の「影の図書館 (shadow library)」から入手され、両社の AI モデルのトレーニング素材として使用されたと主張している。ChatGPT については自分たちの著書を要約したことをその証拠として挙げた。また、Llama についてはモデルのトレーニングに使用したデータセットから、著者の書籍にアクセスできたと主張している。彼らの訴訟にはそれぞれ、著作権侵害、過失、不当利得、不正競争など様々な種類の 6 つの訴因が含まれており、法定損害賠償、利益の返還などを求めている。

●7 月 19 日、William Orrick 連邦地裁判事は、サンフランシスコで行われた公聴会で、アーティスト集団が Stability AI 社、Midjourney 社、DeviantArt 社を相手取って起こした訴訟の大半を棄却する意向を示した⁷⁰。Orrick 判事によると、アーティストたちはオープンソースである Stability AI 社の生成 AI のソースコードにアクセスできることから、著作権侵害の疑いについて「より多くの事実を提供」できるはずだと述べたという。

3. 各国・機関の政策動向

2023 年 4 月以降は、各国政府や国際機関の政策的な動きが多くなってきた。日本で開催された G7 会合も 1 つのきっかけとなった。第 3 章の要約は第 1 章を参照。

⁶⁹ <https://www.theverge.com/2023/7/9/23788741/sarah-silverman-openai-meta-chatgpt-llama-copyright-infringement-chatbots-artificial-intelligence-ai>

⁷⁰ <https://www.reuters.com/legal/litigation/us-judge-finds-flaws-artists-lawsuit-against-ai-companies-2023-07-19/>

3.1. 日本

●自由民主党のデジタル社会推進本部「AI の進化と実装に関するプロジェクトチーム」は3月30日、「AI ホワイトペーパー～AI 新時代における日本の国家戦略～」を取りまとめ、案として公開した⁷¹。4月13日の自民党政調審議会では了承が得られた。項目ごとに、多数の提言が行われた。第5章は「AI 規制に関する新たなアプローチ」と題され、諸外国の検討状況を参照しつつ法規制を含む対応の検討、国際的なルール作りの議論への参加、アナログ規制の見直しの促進、規制緩和手続きの使い勝手の向上、生成 AI に関する知財ルールのガイドラインの策定、公教育における AI の利用指針の策定などが提言された。

●OpenAI 社の CEO である Sam Altman 氏が来日し、4月10日、自民党の「AI の進化と実装に関するプロジェクトチーム」の会合に参加した。塩崎彰久衆議院議員によると、Altman 氏は日本に対して7点提案したとされる⁷²。

1. 日本関連の学習データのウェイト引き上げ
2. 政府の公開データなどの分析提供等
3. LLM（大規模言語モデル）を用いた学習方法や留意点等についてのノウハウ共有
4. GPT-4 の画像解析などの先行機能の提供
5. 機微データの国内保全のため仕組みの検討
6. 日本における OpenAI 社のプレゼンス強化
7. 日本の若い研究者や学生などへの研修・教育提供

●4月14日に開催された衆議院内閣委員会で、立憲民主党の中谷一馬議員が ChatGPT に関する質問を行い、その中で規制を行う可能性の有無を問いただした⁷³。高市早苗大臣はこれに対して「科学技術政策を担当するという立場からは現時点においてただちに使用を禁止にするなどの規制を行うつもりはございません」と回答し、そのうえで「しかしながら、この生成 AI につきましては一定の懸念は指摘されております」とし、経済安全保障の観点から技術情報の流

⁷¹ <https://www.taira-m.jp/2023/03/ai-ai.html>

⁷² <https://japan.cnet.com/article/35202408/>

⁷³ 衆議院インターネット審議中継からアーカイブを視聴。 <https://www.shugiintv.go.jp/jp/>

出などのリスクを低減する対策を検討する必要があることを指摘した。続いて、松野博一官房長官も「現状、規制する考えはございません」と回答した。

●日本政府は、Chat GPT をはじめとする生成 AI の「有効活用」に向けた検討チームを立ち上げることが 4 月 14 日に報道された⁷⁴。そして政府が立ち上げた「AI 戦略チーム（関係省庁連携）」⁷⁵の第 1 回会合が 4 月 24 日に開催された。ChatGPT など生成 AI を利活用する場合の留意点などが議題であった。村井英樹総理補佐官の下で様々な課題に対して迅速に対応していくための関係省庁の実務者級（審議官・課長級）のチームという位置づけである。

●総務省、経済産業省、デジタル庁は共同で、群馬県高崎市において「G7 群馬高崎デジタル・技術大臣会合」を 4 月 29 日と 30 日に開催した⁷⁶。6 つのテーマのうちの 1 つが「責任ある AI と AI ガバナンスの推進」であり、アクションプランが承認された。

●デジタル庁の「第 8 回デジタル社会推進会議幹事会」が 5 月 8 日付で書面開催され、関係省庁に対して、「ChatGPT 等の生成 AI の業務利用に関する申合せ」が決定された⁷⁷。ChatGPT などは「約款型外部サービス」に該当するとして、政府機関等のサイバーセキュリティ対策のための統一基準に従うこととされた。同日、各都道府県及び各指定都市の情報セキュリティ担当課に事務連絡として通知された⁷⁸。

●5 月 9 日、岸田首相は首相官邸で AI に関わる企業関係者や研究者らとの「AI に関する次世代リーダーとの車座対話」を開催した⁷⁹。

●5 月 11 日に総理大臣官邸で 1 回目の「AI 戦略会議」⁸⁰が開催された。構成員は 8 名の有識者と岸田首相を始めとした政府側参加者からなる。資料 2 は「AI を巡る主な論点」であり、3 つの論点とそれぞれの各論が示されている。

⁷⁴ <https://www3.nhk.or.jp/news/html/20230414/k10014039071000.html>

⁷⁵ https://www8.cao.go.jp/cstp/ai/ai_team/ai_team.html

⁷⁶ <https://www.digital.go.jp/news/efdaf817-4962-442d-8b5d-9fa1215cb56a/>

⁷⁷ <https://www.digital.go.jp/councils/social-promotion-executive/councils/191f444c-37fe-4c38-9909-09d9ccdb23af/>

⁷⁸ https://www.soumu.go.jp/menu_hourei/tsutatsu/t_tsutatsu.html

⁷⁹ https://www.kantei.go.jp/jp/101_kishida/actions/202305/09kurumaza.html

⁸⁰ https://www8.cao.go.jp/cstp/ai/ai_senryaku/ai_senryaku.html

➤ 論点 1 AI の利用

- ・ 日本の AI 利用は遅れていないか？
- ・ 民間、教育、公的分野等では、どのような点に留意し、どのように利用を進めるべきか？

➤ 論点 2 懸念・リスク

- ・ プライバシーの侵害、犯罪への使用など人権や安心を脅かす行為にどう対処するか？
- ・ 機密情報の流出、サイバー攻撃の巧妙化などセキュリティ上のリスクにどう対処するか？
- ・ 誤情報、虚偽情報、偏向情報等が蔓延する問題にどう対応するか？
- ・ AI が知的財産権を脅かしていないか？
- ・ 透明性をどのように確保すべきか？
- ・ AI の利用に当たっての責任をどのように考えるか？
- ・ 諸外国におけるルール形成、国際的な規律・標準の検討などにどのように対応するか？

➤ 論点 3 AI の開発

- ・ 日本の AI 開発力は遅れていないか？ どこをどのように強化すればよいのか？

●5 月 11 日、自由民主党デジタル社会推進本部（本部長・平井卓也衆院議員）の「AI の進化と実装に関するプロジェクトチーム（PT）」は、「AI ホワイトペーパー～AI 時代における日本の国家戦略」を発表した⁸¹。研究開発や社会実装の推進に加えて、諸外国の AI 規制の検討状況を分析しつつ、「①重大な人権侵害、②安全保障、③民主主義プロセスへの不当介入など、AI 新時代において法規制を含む対策が必要と考えられる分野につき具体的な検討を行うこと」を提言し、「各国と連携を図りながら AI 利用を巡る国際的なルール作りの議論に積極的かつ戦略的に参画すること」とした。

⁸¹ <https://www.jimin.jp/news/policy/205802.html>

●5月16日に文科省で「デジタル学習基盤特別委員会(第1回)」が開始され、その資料6「生成AIの学校現場での取扱いに関する今後の対応について」では、学校現場に対するガイドライン ver1.0 を、政府全体の議論も踏まえ、夏前を目途に公表する予定とされた⁸²。

●5月19～21日に開催されたG7広島サミットにおいて、5月20日に「広島首脳コミュニケ」が採択された⁸³。前文において、「我々が共有する民主的価値に沿った、信頼できる人工知能(AI)という共通のビジョンと目標を達成するために、包摂的なAIガバナンス及び相互運用性に関する国際的な議論を進める。」と書かれ、＜デジタル＞の箇所には下記のように記載された。

「我々は、法的拘束力を有する枠組みを尊重しつつ、AIの標準の開発におけるマルチステークホルダーアプローチの更なる推進にコミットし、責任あるAIの推進のため、透明性、開放性、公正なプロセス、公平性、プライバシー及び包括性を推進する手続の重要性を認識する。我々は、信頼できるAIという共通のビジョンと目標を達成するためのアプローチと政策手段が、G7諸国間で異なり得ることを認識しつつも、AIガバナンスに関する国際的な議論とAIガバナンスの枠組み間の相互運用性の重要性を強調する。我々は、マルチステークホルダー型の国際機関を通じて、信頼できるAIのためのツール開発を支援し、マルチステークホルダープロセスを通じて、標準化機関における国際技術標準の開発及び採用を促す。我々は、国や分野を超えてますます顕著になっているAIの機会及び課題について直ちに評価する必要性を認識し、OECDなどの国際機関が政策展開の影響に関する分析を検討し、人工知能グローバルパートナーシップ(GPAI)が実践的なプロジェクトを実施することを奨励する。この観点から、我々は、関係閣僚に対し、生成AIに関する議論のために、包摂的な方法で、OECD及びGPAIと協力しつつ、G7の作業部会を通じた、広島AIプロセスを年内に創設するよう指示する。これらの議論は、ガバナンス、著作権を含む知的財産権の保護、透明性の促進、偽情報を含む外国からの情報操作への対応、これらの技術の責任ある活用といったテーマを含み得る。我々は、デジタル・技術大臣会合における「AIガバナンスの相互運用性を促進等するためのアクションプラン」を歓迎する。」

⁸² https://www.mext.go.jp/b_menu/shingi/mext_01431.html

⁸³ <https://www.g7hiroshima.go.jp/documents/>

●政府は5月26日、「AI戦略会議」（座長・松尾豊東大大学院教授）の第2回会合を開き、現時点での論点を整理し、「懸念されるリスク」として、①機密情報の漏洩や個人情報の不適正な利用、②犯罪の巧妙化・容易化、③偽情報などが社会を不安定化・混乱、④サイバー攻撃が巧妙化、⑤学校現場における生成AIの扱い、⑥著作権侵害、⑦AIによって失業者が増える、の7点を挙げた⁸⁴。

●個人情報保護委員会は6月2日、「生成AIサービスの利用に関する注意喚起等について」と題する文書を発表し、利用者（個人情報取扱事業者、行政機関等、一般の利用者）向けとOpenAI社向けに注意喚起を行った⁸⁵。後者については、OpenAI社に対して「あらかじめ本人の同意を得ないで、ChatGPTの利用者及び利用者以外の者を本人とする要配慮個人情報を取得しないこと」を要請した。

●政府は6月9日に開催された知的財産戦略本部会合において、「知的財産推進計画2023」（案）を公表した⁸⁶。「III. 知財戦略の重点10施策」の3点目が「急速に発展する生成AI時代における知財の在り方」となっており、（1）生成AIと著作権、と、（2）AI技術の進展を踏まえた発明の保護の在り方、が記述されている。（1）に関する施策の方向性として「生成AIと著作権との関係について、AI技術の進歩の促進とクリエイターの権利保護等の観点に留意しながら、具体的な事例の把握・分析、法的考え方の整理を進め、必要な方策等を検討する。」とされた。

●G7の個人データ保護機関のトップによる「第3回G7データ保護・プライバシー機関ラウンドテーブル」の会合が6月20日と21日に東京都内で開かれ、コミュニケ（信頼性のある自由なデータ流通（DFFT）の具体化と規制協力の強化に向けて）、生成AIに関する声明、行動計画が採択された⁸⁷。

⁸⁴ https://www8.cao.go.jp/cstp/ai/ronen_honbun.pdf

⁸⁵ https://www.ppc.go.jp/news/careful_information/230602_AI_utilize_alert/

⁸⁶ <https://www.kantei.go.jp/jp/singi/titeki2/230609/gijisidai.html>

⁸⁷ https://www.ppc.go.jp/enforcement/cooperation/international_conference/g7_roundtable_202306/

- 文化庁の文化審議会著作権分科会（第 68 回）（第 23 期第 1 回）が 6 月 30 日に開催された⁸⁸。資料 4 の審議事項②「著作権法制度に関すること」の 2 点目に「生成 AI と著作権に関する論点整理について」が挙げられ、会議でも生成 AI に関する議論が行われた。
- 文部科学省が 7 月 4 日、初等中等教育局長名で、都道府県教育委員会や指定都市教育委員会の教育長等宛に、「初等中等教育段階における生成 AI の利用に関する暫定的なガイドライン」を発出した⁸⁹。位置づけとしては、「主として対話型の文章生成 AI について、学校関係者が現時点で生成 AI の活用の適否を判断する際の参考資料として、令和 5 年 6 月末日時点の知見をもとに暫定的に取りまとめるもの」とされている。
- 文部科学省が 7 月 13 日、文部科学省 高等教育局 専門教育課及び大学教育・入試課から、各国公立大学法人担当課などに対して、「大学・高専における生成 AI の教学面の取扱いについて（周知）」を発出した⁹⁰。
- 経済産業省は 7 月 21 日、「生成 AI 開発支援スキーム検討委員会」を立ち上げ、第 1 回会議を開催したことを発表した⁹¹。
- 8 月 4 日に第 4 回の AI 戦略会議（イノベーション政策強化推進のための有識者会議）が開催された⁹²。5 月に立ち上げられた「広島 AI プロセス」について、9 月頃にオンラインでの閣僚級会合を開催し中間報告のとりまとめを行い、10 月 9 日には IGF（インターネット・ガバナンス・フォーラム）京都 2023 においてマルチステークホルダー会合を実施し、年末までに再度オンラインでの閣僚級会合を開催し、成果のとりまとめを行うというロードマップが示された。

⁸⁸ <https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/bunkakai/68/index.html>

⁸⁹ https://www.mext.go.jp/content/20230704-mxt_shuukyo02-000003278_003.pdf

⁹⁰ https://www.mext.go.jp/b_menu/houdou/2023/mext_01260.html

⁹¹ <https://www.meti.go.jp/press/2023/07/20230721003/20230721003.html>

⁹² https://www8.cao.go.jp/cstp/ai_senryaku/4kai/4kai.html

3.2. 欧州（EU）

●2023年4月5日付の Le Monde 紙は、前日の4日までに ChatGPT に対する個人データの使用に関する申立てが少なくとも2件、フランスのデータ保護当局である CNIL に寄せられたことを報じた⁹³。

●欧州データ保護会議（EDPB）は4月13日、ChatGPT サービスに関して、イタリアのデータ保護当局が OpenAI 社に対して行った強制措置や、加盟各国のデータ保護当局が実施する可能性のある強制措置に関して情報交換したうえで、ChatGPT に対する措置に関して協力を促進するための専門タスクフォースを立ち上げることを決定した⁹⁴。

●スペインのデータ保護当局（AEPD）は4月13日、OpenAI 社に対する調査手続きを開始したことを発表した⁹⁵。この調査は、OpenAI 社のデータ保護に関する実践やプロセスに焦点を当てており、OpenAI 社が個人データの取扱いに関して適切な対策を講じているかどうかを評価するとしている。また、前の週に AEPD は欧州データ保護会議（EDPB）に対して、一般データ保護規則（GDPR）との関連について欧州レベルでの調和した対応のために会議で取り上げることを要請したことも明らかにした。

●ドイツの規制当局が、OpenAI 社のプライバシー対策と EU の一般データ保護規則（GDPR）の遵守について調査を開始したと4月24日、報じられた⁹⁶。OpenAI 社に対する質問状をまとめ、6月11日までに回答を得ることを期待しているという。

●OpenAI 社はイタリア当局にレターを送付し、4月11日付で指示された指令を遵守するための方策について説明したことが、4月28日、イタリアのデータ保護当局のウェブサイトで明らかにされた⁹⁷。当局はこれらの対策を歓迎し、同社に対して、情報キャンペーンを実施しイタリア国民にオプトアウトの権利や年齢認証などを周知することを求めた。ただし、EDPB が設

⁹³ https://www.lemonde.fr/pixels/article/2023/04/05/chatgpt-premieres-plaintes-aupres-de-la-cnil-contre-le-logiciel-d-intelligence-artificielle_6168425_4408996.html

⁹⁴ https://edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en

⁹⁵ <https://www.aepd.es/es/prensa-y-comunicacion/notas-de-prensa/aepd-inicia-de-oficio-actuaciones-de-investigacion-a-openai>（スペイン語）

⁹⁶ <https://www.barrons.com/news/germany-launches-data-protection-inquiry-over-chatgpt-cdd15588>

⁹⁷ <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9881490#english>

置したタスクフォースのもとでの事実確認活動は継続する予定である。これを受けて OpenAI 社はイタリア国内向けサービスを再開した。主な改善点は以下のとおり。

- アルゴリズムのトレーニングのためにどのような個人データがどのような取り決めのものと処理されているかについて記載。
- アルゴリズムのトレーニングのためのデータ処理からオプトアウトする権利を付与。
- データ主体が不正確と思われる情報の消去を求めることができる仕組みを導入する一方、不正確な情報を修正することは現時点では技術的に不可能であることを明記。
- 「正当な利益 (legitimate interests)」という法的根拠に基づき、個人データの処理を継続。
- イタリア人登録ユーザー向けのページに、18 歳以上であること、または 13 歳以上で保護者の同意を得ていることを確認するためのボタンを追加。
- 13 歳未満のユーザーのアクセスを遮断するために、サービスのサインアップページで生年月日を明記してもらう要求や、13 歳以上 18 歳未満のユーザーに対して保護者の同意の確認を求める要求を追加。

● 欧州議会では数カ月の議論の末、4 月 27 日に、欧州委員会が 2 年前に提案した「AI Act」の修正案の内容について暫定的な政治合意に達したと報道された⁹⁸。当初は一部の委員は、生成 AI モデルの学習に使用される著作権で保護された素材を全面的に禁止することを提案していたが、透明性の要件に賛成して断念したという。

● 2023 年 5 月 11 日、欧州議会の域内市場委員会 (Internal Market Committee) と市民の自由委員会 (Civil Liberties Committee) は、AI を規制するルール「AI Act」の修正案の草案を、賛成 84 票、反対 7 票、棄権 12 票で採択した⁹⁹。新たに生成 AI に関する条項が付け加えられた (詳細は後述)。

⁹⁸ <https://www.euractiv.com/section/artificial-intelligence/news/meps-seal-the-deal-on-artificial-intelligence-act/>

⁹⁹ <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>

●フランス CNIL は 5 月 16 日、AI、特に ChatGPT のような生成 AI アプリケーションに関するプライバシーに関する懸念に対応するアクションプラン（行動計画）を発表した¹⁰⁰。行動計画は 4 つの柱からなる。

- AI システムの機能と人々への影響について理解する
- プライバシーに配慮した AI の開発を可能にし、その方向に導く
- フランスおよび欧州の AI エコシステムにおけるイノベティブなプレイヤーを束ね、支援する
- AI システムを監査・管理し、人々を保護する

●欧州委員会の副委員長である Margrethe Vestager 氏は 5 月 31 日、スウェーデンで開催された米国・EU 貿易技術協議会（TTC）の第 4 回会合後の記者会見で、立法措置を待たずに、AI に関する行動規範（AI Code of Conduct）の草案を数週間以内に作成する考えを示した¹⁰¹。そのための専門家グループを設置したという。EU と米国の作業グループは、65 の用語を収録した用語集の第一版を公表した¹⁰²。まず用語を選択し、AI ライフサイクル、計測、技術システム属性、ガバナンス、信頼に値する、の 5 分野に分類された。

●欧州委員会は 5 月 30 日、スタッフ向けの内部ガイドライン「オンラインで利用可能な生成 AI ツールの使用に関するスタッフ向けガイドライン」を発行したことを Euractiv が報じた¹⁰³。欧州委員会スタッフのみがアクセスできる内部情報システム上で利用可能である。欧州委員会の Information Management Steering Board が 4 月 27 日に採択した。挙げられたリスクと対策は以下のとおり。

1. 機微情報や個人データが AI プロバイダーに送信されることから、AI モデルに入力してはならない
2. 不完全なデータセットやアルゴリズム設計のために、間違った回答やバイアスのある回答が生み出されることを前提に、常に回答を批判的に評価すべき

¹⁰⁰ <https://www.cnil.fr/en/artificial-intelligence-action-plan-cnil>

¹⁰¹ <https://www.reuters.com/technology/eu-tech-chief-calls-voluntary-ai-code-conduct-within-months-2023-05-31/>

¹⁰² <https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence>

¹⁰³ <https://www.euractiv.com/section/artificial-intelligence/news/eu-commission-issues-internal-guidelines-on-chatgpt-generative-ai/>

3. AI モデルの訓練に知的財産権で保護されたコンテンツが含まれる可能性があるために、スタッフは常に、AI が生成した出力が知的財産権を侵害していないかどうかを批判的に評価すべき & 公的文書に直接使用してはならない
4. AI モデルは常に利用可能とは限らないため、スタッフは、重要で時間的制約のある業務において、これらのツールに頼ってはならない

●オランダのデータ保護当局（AP）は 6 月 7 日、OpenAI 社に対して、個人データの取り扱いに関して説明を求めているとしたうえで、今後いくつかの措置を講じる予定であるとした¹⁰⁴。具体的には、どのようにインターネットから個人データを収集し、使用しているか、人々の入力アルゴリズムのトレーニングに使用されているか、個人データがアウトプットに含まれるか、不適切なアウトプットを修正したり、削除したりできるかどうか、などについてである。

●欧州議会で 6 月 14 日、AI Act の修正案が採決された¹⁰⁵。生成 AI モデルを含む「基盤モデル」が明示的に法案に追加され、基盤モデルは「大規模なデータで訓練され、出力の汎用性を考慮して設計され、広範な特徴的タスクに適応可能な AI システムモデル」と定義されたうえで、新規に設けられた第 28b 条第 2 項において、基盤モデルのプロバイダーには次のような義務が課された。

- A) 適切な設計、試験及び分析を通じて、健康、安全、基本的権利、環境、民主主義及び法の支配¹⁰⁶に対する合理的に予見可能なリスクを、独立した専門家の関与など適切な方法で、開発前及び開発期間中に特定・低減・緩和すること、並びに開発後に残存する軽減不可能なリスクを文書化することを通して示す。
- B) 基盤モデルのための適切なデータガバナンス措置、特にデータソースの適切性、ありうるバイアス、適切な軽減策を検討するための措置の対象となるデータセットのみを処理し、組み込む。
- C) 独立した専門家の関与によるモデル評価、文書化された分析、構想・設計・開発中の広範な試験などの適切な方法を通じて評価された、適切なレベルのパフォーマンス、

¹⁰⁴ <https://autoriteitpersoonsgegevens.nl/actueel/ap-vraagt-om-opheldering-over-chatgpt>（オランダ語）

¹⁰⁵ https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

¹⁰⁶ 2021 年 4 月に公表された欧州委員会による AI Act 案では保護すべきものが「健康」「安全」「基本的権利」であったが、2023 年 6 月に採択された議会修正案では、「民主主義と法の支配」と「環境」が加わり、5 項目になった。これらの追加は、生成 AI の普及に対応したものであると考えられる。

予測可能性、解釈可能性、信頼性、安全性及びサイバーセキュリティを、そのライフサイクル全体を通じて達成するために、基盤モデルを設計・開発する。

- D) エネルギー使用、資源使用及び廃棄物を削減し、エネルギー効率及びシステム全体の効率を向上させるために、関連する既存の EU 法及び国内法を損なうことなく、適用可能な基準を活用して、基盤モデルを設計及び開発する。（以下、略）

また、第 4 項では、生成 AI を「複雑なテキスト、画像、音声又は映像のようなコンテンツを、様々なレベルの自律性をもって生成することを特に意図した AI システム」と定義したうえで、「使用される基盤モデルのプロバイダー、及び基盤モデルを生成 AI システムに特化させるプロバイダー」は追加で次のような義務が課された。

- A) 第 52 条 (1) に概説される透明性の義務を遵守すること。
- B) 一般に認識されている最新の技術水準に沿い、かつ、表現の自由を含む基本的権利を損なうことなく、EU 法に違反するコンテンツの生成に対する適切なセーフガードを確保するような方法で、基盤モデルを訓練し、場合によっては、設計し開発すること。
- C) 著作権に関する EU 法又は国内法に違反することなく、著作権法に基づいて保護されるトレーニングデータの使用についての十分に詳細な概要を文書化し、一般に公開すること。

A) の透明性の義務とは「AI システムと相互作用することを適時に、明確かつ分かりやすい方法で通知するように設計及び開発されることを確保しなければならない。」とするものである。このあと、欧州議会、EU 理事会、欧州委員会の三者が最終的な法案を作成する三者協議（trilogue）へと向かっている。

3.3. 米国

- バイデン大統領は 2023 年 4 月 4 日、ホワイトハウスで大統領科学技術諮問委員会（President's Council of Advisors on Science and Technology, PCAST）と会見し、AI が危険

かどうかはまだ分からないとしながらも、テクノロジー企業には製品を公開する前に安全性を確認する責任があると強調した¹⁰⁷。

●米国商務省の電気通信情報庁（National Telecommunications and Information Administration, NTIA）は4月11日、AIのアカウントビリティ・ポリシーに関する意見募集を開始した¹⁰⁸。書面でのコメントの締め切りは6月12日。アカウントビリティを支える実践としてアセスメント（Assessment）と監査（Audit）に着目し、34の大きな問いと、それらにさらに枝分かれした小問を挙げた。

●連邦取引委員会（Federal Trade Commission, FTC）の議長である Lina Khan 氏は4月25日、司法省（Department of Justice, DOJ）公民権局、消費者金融保護局（Consumer Financial Protection Bureau, CFPB）、雇用機会均等委員会（Equal Employment Opportunity Commission, EEOC）の担当者とともに、「自動化システムにおける差別や偏見に対する執行努力に関する共同声明」を発表した¹⁰⁹。AIを利用した自動化システムを対象とし、差別が生じる原因として、1）データ及びデータセット、2）モデルの不透明性とアクセス、3）設計と使用、が挙げられた。

●バイデン政権は5月4日、AIにおける米国の責任あるイノベーションをさらに促進し、人々の権利と安全を守るための新しい措置を発表した¹¹⁰。同日、ハリス副大統領らは米国企業4社（Alphabet 社、Anthropic 社、Microsoft 社、OpenAI 社）のCEOと会談した。発表の内容は以下の3点である。

- 全米科学財団（National Science Foundation, NSF）が新たに7つのAI研究機関を立ち上げるために1億4000万ドルの資金を提供すること
- OpenAI 社や Stability AI 社を含む既存の生成AIシステムの開発者らが公開の第三者評価に参加することを約束したこと

¹⁰⁷ <https://www.reuters.com/technology/biden-discuss-risks-ai-tuesday-meeting-with-science-advisers-2023-04-04/>

¹⁰⁸ <https://ntia.gov/issues/artificial-intelligence/request-for-comments>

¹⁰⁹ <https://www.ftc.gov/news-events/news/press-releases/2023/04/ftc-chair-khan-officials-doj-cfpb-eeoc-release-joint-statement-ai>

¹¹⁰ <https://www.whitehouse.gov/ostp/news-updates/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/>

- 行政管理予算局（Office of Management and Budget, OMB）が、AI システムの開発・調達・利用において連邦省庁が従うべき具体的な方針を定めた政策ガイダンス案を作成・公開し、今年の夏にパブリックコメントを求めること

●5 月 13 日、大統領科学技術諮問委員会（PCAST）は、生成 AI 技術に関して、機会とリスクを評価し、可能な限り公正で責任ある形で安全に開発・展開されるための最善の方法について大統領に提言することを目的とした生成 AI に関する作業グループを立ち上げた¹¹¹。公開セッションを開催するとともに、パブリックコメントの募集を開始した。具体的なテーマとして次の 5 つが挙げられた。

1. 説得力のある画像、音声、テキストが簡単に大量に生成できる時代に、検証可能で信頼できる情報に確実にアクセスするにはどうすればいいか？ あるメディアが、本当にそのソースから発信されたものであると確信できるのだろうか。
2. 悪意ある行為者が AI を使って市民の信念や理解を操作することに、私たちはどのように対処すればよいか？
3. AI が生成した偽情報を検知し、対抗するために、どのような技術、政策、インフラを開発することができるのか？
4. 民主主義の根幹である、選挙で選ばれた代表者と国民との関わりを、AI が生み出すノイズにかき消されないようにするにはどうすればいいのか？
5. 科学、政治、産業、教育のリーダーを含むすべての人が、AI が生成した誤報、なりすまし、操作を識別するために必要なスキルを身につけるにはどうすればよいか？

●5 月 16 日、上院司法委員会のプライバシー・技術・法律に関する小委員会における公聴会「AI の監督：AI のためのルール（Oversight of A.I.: Rules for Artificial Intelligence）」において、OpenAI 社の Altman 氏を含め 3 名が証人として参加した¹¹²。

- Sam Altman 氏（OpenAI 社 CEO）
- Gary Marcus 氏（ニューヨーク大学名誉教授）
- Christina Montgomery 氏（IBM 社）

¹¹¹ <https://www.whitehouse.gov/pcast/briefing-room/2023/05/13/pcast-working-group-on-generative-ai-invites-public-input/>

¹¹² <https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-rules-for-artificial-intelligence>

●ホワイトハウス科学技術政策局（Office of Science and Technology Policy, OSTP）は5月23日、個人の権利と安全を守り、米国民に成果をもたらす責任あるAI研究・開発・実装を進める新たな取り組みを発表した¹¹³。

- AI研究開発（R&D）に対する連邦政府の投資を集中させるためのロードマップ（2019年発表の「国家AI研究開発戦略計画」）を更新
- 重要なAI問題に対する一般からのインプットを求める新たな要請（Request for Information, RFI）を開始（7/7締め切り）
- 教育現場におけるAIに関するリスクと機会に関する新しい報告書「AIと将来の教育と学習：洞察と勧告」を発表

RFIにおける問い掛けは、「権利・安全・国家安全保障を保護する」「衡平を進め、市民の権利を強化する」「民主主義と市民参加を支持する」「経済成長と良い仕事を促進する」「公共サービスを革新する」等の大項目のもとで計29問からなる。

●米国議会調査局（Congressional Research Service, CRS）は5月23日付で、報告書「生成AIとデータプライバシー：入門編（Generative Artificial Intelligence and Data Privacy: A Primer）」を発表した¹¹⁴。政策の検討材料として、1）既存のデータプライバシーと関連法、2）提案されているプライバシー法案、3）既存の政府機関の権限、4）データ・スクレイピングの規制、5）代替技術アプローチの研究開発、が挙げられた。また、「生成AIと著作権法」と題する報告書も5月11日付で第四版に更新されている¹¹⁵。

●上院司法委員会（Senate Committee on the Judiciary）の人権及び法小委員会は6月13日、「AIと人権」と題する公聴会を開催した¹¹⁶。証人は次の4名であった。ディープフェイクによる被害者も含まれている。

- Geoffrey Cain氏（Foundation for American Innovation）

¹¹³ <https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/23/fact-sheet-biden-harris-administration-takes-new-steps-to-advance-responsible-artificial-intelligence-research-development-and-deployment/>

¹¹⁴ <https://crsreports.congress.gov/product/pdf/R/R47569>

¹¹⁵ <https://crsreports.congress.gov/product/pdf/LSB/LSB10922>

¹¹⁶ <https://www.judiciary.senate.gov/committee-activity/hearings/artificial-intelligence-and-human-rights>

- Alexandra Reeve Givens 氏 (Center for Democracy & Technology)
- Aleksander Madry 氏 (マサチューセッツ工科大学 (MIT))
- Jennifer DeStefano 氏 (AI ディープフェイク誘拐／恐喝詐欺の被害者)

●6月22日、Gina Raimondo 商務長官は、米国標準技術研究所 (National Institute of Standards and Technology, NIST) が AI に関する新たな公開作業グループ (WG) を立ち上げ、NIST の「AI リスク・マネジメント・フレームワーク (AI RMF)」に基づいて、急速に進歩するこの技術に取り組むことを発表した¹¹⁷。WG の短期、中期、長期の目標が示されており、最初は NIST の AI RMF を生成 AI 技術の開発支援にどのように利用できるかを説明するガイダンスに関する意見を集める。中期的には、生成 AI に関連する試験、評価、測定に関する NIST の作業を支援する。長期的には現代の重要な課題に生成 AI が対処できる機会を探るという。

●連邦取引委員会 (FTC) は 6 月 29 日、「生成 AI に競争上の懸念」と題する、競争局と技術局スタッフによるブログ記事を掲載した¹¹⁸。生成 AI は、データ、人材、計算資源といった主要な構成要素の 1 つ以上を支配することで、様々な競争上の懸念を引き起こす可能性があるとし、競争局は技術局と緊密に連携しながら、あらゆる手段を駆使して不公正な競争方法を特定し、対処していくとしている。技術局 (office of technology) は 2023 年 2 月に FTC 内に新設された部署である¹¹⁹。

●連邦取引委員会 (FTC) が OpenAI 社に対して、同社の AI モデルに関連するリスクにどのように対応しているかについての記録を求める 20 ページにわたる文書 (demand for records)¹²⁰を送付したことを 7 月 13 日付の Washington Post 紙が報じた¹²¹。FTC は、OpenAI 社が不公正な慣行に関与し、消費者に「風評被害 (reputational harm)」をもたらしたかどうかを調

¹¹⁷ <https://www.nist.gov/news-events/news/2023/06/biden-harris-administration-announces-new-nist-public-working-group-ai>

¹¹⁸ <https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns>

¹¹⁹ <https://www.ftc.gov/news-events/news/press-releases/2023/02/ftc-launches-new-office-technology-bolster-agencys-work>

¹²⁰ https://www.washingtonpost.com/documents/67a7081c-c770-4f05-a39e-9d02117e50e8.pdf?itid=ik_inline_manual_4

¹²¹ <https://jp.reuters.com/article/ftc-openai-investigation-idCAKBN2YT0R1>

<https://www.washingtonpost.com/technology/2023/07/13/ftc-openai-chatgpt-sam-altman-lina-khan/>

査している。FTC が OpenAI 社に行った質問の 1 つが、同社製品が「実在の個人について、虚偽、誤解を招く、または中傷的な発言をする (false, misleading, or disparaging)」可能性があるととして、同社がとった対策に関するものであるという。FTC は Washington Post 紙に対して本件についてのコメントを拒否した。

●バイデン-ハリス政権は 7 月 21 日、Amazon 社、Anthropic 社、Google 社、Inflection 社、Meta 社、Microsoft 社、OpenAI 社の大手 AI 企業 7 社をホワイトハウスに招集し、AI 技術がもたらすリスクの管理について、これらの企業から自発的なコミットメントを確保したことを発表した¹²²。コミットメントは、AI の未来にとって基本的でなければならない 3 つの原則、すなわち安全 (safety)、セキュリティ (security)、信頼 (trust) を強調するものであるとした。

- バイオ、サイバー、その他の安全分野など、誤用、社会的リスク、国家安全保障上の懸念を含む分野において、モデルやシステムの社内外でのレッドチーム活動¹²³にコミットする。
- 信頼と安全のリスク、危険な、あるいは出現しつつある能力、セーフガードを回避しようとする試みに関して、企業と政府の間で情報共有に努める。
- 独自かつ未発表のモデルウェイトを保護するため、サイバーセキュリティと内部脅威に対する安全策に投資する。
- 第三者による課題や脆弱性の発見と報告を奨励する。
- AI が生成した音声または映像コンテンツについて、ロバストな証明、電子透かし、またはその両方を含め、利用者が、AI が生成した音声または映像コンテンツかどうかを理解できる仕組みを開発し、実装する。
- モデルまたはシステムの能力、限界、および適切な使用領域と不適切な使用領域について、公正性やバイアスへの影響など社会的リスクに関する議論を含めて公表する。
- 有害なバイアスや差別の回避、プライバシーの保護など、AI システムがもたらす社会的リスクに関する研究を優先する。

¹²² <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>

¹²³ 情報セキュリティの文脈で用いられることの多い用語で、攻撃側の視点からシステムを評価し、脆弱性やリスクを特定する活動を指す。

- 最先端の AI システムを開発・実装し、社会の最も大きな課題の解決に貢献する。

同時に、「現在進行中の仕事はまだたくさんある。」としたうえで、「バイデン＝ハリス政権は現在、大統領令を策定中であり、米国が責任あるイノベーションを先導できるよう、超党派の法案を追求する予定である。」と念を押している。これを受けて 7 月 26 日、フロンティア・モデル・フォーラム（Frontier Model Forum）の設立が発表された（1.3 節を参照）。

●7 月 25 日、上院司法委員会のプライバシー・技術・法に関する小委員会が公聴会「AI の監視：規制の原則」を開催した。AI に関する有識者 3 人が証言を行い、規制のあり方や今後の展望について提言した。書面¹²⁴と映像¹²⁵が公開されている。要約は Techcrunch の記事¹²⁶を参照。

- Dario Amodei 氏（Anthropic 社 CEO）
- Yoshua Bengio 氏（ケベック AI 研究所（Mila）創設者兼科学ディレクター、モントリオール大学教授）
- Stuart Russell 氏（UC Berkeley 教授）

3.4. 英国

●情報コミッショナー局（Information Commissioner's Office, ICO）は 2023 年 4 月 3 日、データ保護法は、処理する個人データが一般にアクセス可能なソースに由来する場合にも適用されることを確認したうえで、個人データを処理する生成 AI を開発または使用している場合は、以下の質問を自問する必要があるとした¹²⁷。

- 個人データを処理するための法的な根拠は何ですか？

¹²⁴ <https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-principles-for-regulation>

¹²⁵ <https://www.youtube.com/watch?v=hm1zexCjELo>

¹²⁶ <https://techcrunch.com/2023/07/25/ai-leaders-warn-senate-of-twin-risks-moving-too-slow-and-moving-too-fast/>

¹²⁷ <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2023/04/generative-ai-eight-questions-that-developers-and-users-need-to-ask/>

- あなたは、コントローラー（controller）、共同コントローラー（joint controller）、あるいはプロセッサ（processor）ですか？
- データ保護影響評価（DPIA）を準備しましたか？
- 透明性はどのように確保するのですか？
- セキュリティリスクはどのように軽減されますか？
- 必要のないデータ処理をどのように制限するのですか？
- 個人の権利の要求にはどのように対応するのですか？
- 生成 AI を使用してもっぱら自動化された意思決定を行うのですか？

●Sunak 首相は 4 月 24 日、英国を 2030 年までに科学技術大国としての地位を確固とするため、大規模な言語モデルを含む基盤モデルを開発するための「基盤モデルタスクフォース」に 1 億ポンドの初期立ち上げ資金を提供することを発表した¹²⁸。

●競争・市場庁（Competition and Market Authority, CMA）は 5 月 3 日、LLM を含む AI 基盤モデルの開発と利用が市場競争と消費者保護に与える影響について初期レビューを開始することを発表した¹²⁹。ステークホルダーからの意見とエビデンスを募集しており、2023 年 6 月 2 日までの提出を求めた。これらに基づき、9 月に報告書が発表される予定である。

●Sunak 首相は 6 月 7 日、今秋、主要国、大手ハイテク企業、研究者が一堂に会し、AI がもたらす最も重大なリスクを評価・監視するための安全対策に合意することを目的とした「初のグローバル AI サミット」を英国政府が開催すると発表した¹³⁰。

●政府は 6 月 18 日、「基盤モデルタスクフォース（Foundation Model Taskforce）」の議長に Ian Hogarth 氏を任命したことを発表した¹³¹。タスクフォースの今後数ヶ月間の重要なミッションは「今年後半に英国で開催される AI の安全性に関する初の世界サミットに向けて、最先端の安全性研究を進めること」とされている。

¹²⁸ <https://www.gov.uk/government/news/initial-100-million-for-expert-taskforce-to-help-uk-build-and-adopt-next-generation-of-safe-ai>

¹²⁹ <https://www.gov.uk/government/news/cma-launches-initial-review-of-artificial-intelligence-models>

¹³⁰ <https://www.gov.uk/government/news/uk-to-host-first-global-summit-on-artificial-intelligence>

¹³¹ <https://www.gov.uk/government/news/tech-entrepreneur-ian-hogarth-to-lead-uks-ai-foundation-model-taskforce>

●英国の 24 大学からなる Russell Group は 7 月 4 日、「教育における生成 AI ツールの利用に関する原則」を公表した¹³²。5 つの原則は次のとおり。本文ではそれぞれに解説がくわえられている。

1. 大学は、学生やスタッフが AI リテラシーを身につけられるよう支援する。
2. スタッフは、学生が学習体験において効果的かつ適切に生成 AI ツールを使用できるように支援する能力を備える。
3. 大学は、生成 AI の倫理的な利用を取り入れ、平等なアクセスを支援するために、教育と評価を適応させる。
4. 大学は、学問の厳正さと公正さが維持されるようにする。
5. 大学は、技術とその教育への応用が発展するのに合わせて、協働してベストプラクティスを共有する。

●英国議会の通信・デジタル委員会（Communications and Digital Committee）」は 7 月 7 日、大規模な言語モデルと、英国がその機会とリスクに確実に対応できるようにするために今後 1～3 年間に何が必要かを検討する調査を開始することを発表した¹³³。委員会は、2023 年 9 月 5 日までに書面による情報を募集している。質問リストは次のとおり。

<能力と傾向>

1. 今後 3 年間で、大規模言語モデルはどのように発展するか。
 - a) この分野の予測には不確実性がつきものであることを踏まえ、将来の予測の理解と信頼性を向上させるために何ができるか。
2. 今後 3 年間の最大のチャンスとリスクは何か。
 - a) この文脈において、リスクについてどのように考えるべきか。

<国内規制>

3. AI 白書は（他の政府政策とともに）大規模言語モデルにどの程度適切に対処しているか？ニーズに合わせて調整した規制アプローチが必要か。

¹³² <https://russellgroup.ac.uk/news/new-principles-on-use-of-ai-in-education/>

¹³³ <https://committees.parliament.uk/committee/170/communications-and-digital-committee/news/196281>

a) オープンソースモデルが普及することの含意は？

4. 英国の規制当局は、大規模言語モデルに対応するのに十分な専門知識とリソースを有しているか。

5. リスクに対処し、機会を生かすために、非規制&規制オプションは何か？

a) そのようなオプションは実際にどのように機能するのか、またそれを実施する上での障壁は何か。

b) AI ライフサイクルのどの段階での介入が最も効果的か？

c) 意図しない帰結のリスクにはどのように対処できるか？

< 国際的背景 >

6. 英国のアプローチは、他の国、特に EU、米国、中国のアプローチと比較してどうか。

a) より広範な戦略的国際競争は、大規模言語モデルの規制のあり方にどの程度影響するか。

b) 規制がばらばらになる可能性はどの程度か。その結果はどのようなものか。

● 情報コミッショナー局（ICO）は「イノベーション・アドバイス・サービス」において、これまでに尋ねられた質問と回答（Q&A）を用意しているが、このリストに、「顧客や見込み顧客へのメール回答案を作成するために生成 AI システムを使用する場合、“正当な利益（legitimate interests）”は適切な法的根拠となりますか？」という Q を追加した¹³⁴。この組織は、生成 AI ツールが作成したメール回答案をスタッフが確認するが、特別カテゴリーデータ（要配慮個人情報に相当する）が含まれることがあるため、データ処理の法的根拠が必要になる。これに対して回答は、「正当な利益（legitimate interests）に依拠するためには、組織は、正当な利益が、個人の利益、権利、または自由によって上書きされない限り、生成 AI ツールの使用が、組織が特定した正当な利益の目的のために必要であることを証明しなければならない。このため、組織は、生成 AI ツールの使用が比例的（proportionate）であり、顧客や見込み顧客の権利を侵害しないことを証明する必要がある。」とした。ICO は「正当な利益」を証明するための方法として、「正当な利益アセスメント（Legitimate Interests Assessment, LIA）」を用意している¹³⁵。

● 英国政府は 8 月 24 日、世界で最初に「AI 安全サミット」を 11 月 1～2 日に Buckinghamshire

¹³⁴ <https://ico.org.uk/for-organisations/advice-and-services/innovation-advice/previously-asked-questions/>

¹³⁵ <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/legitimate-interests/>

の Bletchley Park で開催されると発表した¹³⁶。

3.5. カナダ

●カナダのプライバシーコミッショナー局（Office of the Privacy Commissioner of Canada, OPC）は 2023 年 4 月 4 日、同意のない個人情報の収集・使用・開示を主張する苦情を受けて、OpenAI 社に対する調査を開始した¹³⁷。

●OPC は 5 月 25 日、ケベック州、ブリティッシュ・コロンビア州、アルバータ州のプライバシー保護当局と合同で、OpenAI 社を調査すると発表した¹³⁸。調査内容として挙げられたのは以下の点である。

- カナダに拠点を置く個人の個人情報の収集、使用、開示について、ChatGPT を通じて、有効かつ意味のある同意を得たかどうか。
- オープン性と透明性、アクセス、正確性、アカウントビリティに関する義務を遵守しているかどうか。
- 合理的な人間がその状況において適切、合理的、または合法的と考える目的のために個人情報を収集、使用、および/または開示しているか、及び、この収集がこれらの目的に必要な情報に限定されているかどうか。

3.6. オーストラリア

●Ed Husic 産業・科学大臣の依頼により、国家科学技術会議（National Science and Technology Council）が 2023 年 3 月 24 日、「生成 AI に関する速報研究レポート（Rapid Research Report on Generative AI）」を作成した¹³⁹。この報告書は以下の 2 つの問いに答えたものである。

¹³⁶ <https://www.gov.uk/government/news/iconic-bletchley-park-to-host-uk-ai-safety-summit-in-early-november>

¹³⁷ https://www.priv.gc.ca/en/opc-news/news-and-announcements/2023/an_230404/

¹³⁸ https://www.priv.gc.ca/en/opc-news/news-and-announcements/2023/an_230525-2/

¹³⁹ <https://www.chiefscientist.gov.au/GenerativeAI>

- 今後 2 年、5 年、10 年の間に、大規模言語モデル（LLM）とマルチモーダル基盤モデル（MFM）の学習技術を適用する機会とリスクは何か？
- AI の潜在的な機会と影響に対処するため、ChatGPT のようなモデルが登場して以来、他の先進国が国際的に実施している戦略の例にはどのようなものがあるか？

●政府は上記レポートに基づき、6 月 1 日から 8 月 4 日まで、「オーストラリアにおける安全で責任ある AI」と題するディスカッションペーパーを公表し、国民から意見を募集した¹⁴⁰。

3.7. 中国

●政府から 2023 年 4 月 11 日付で「国家インターネット情報局による「生成 AI サービス管理方法（意見募集案）」のパブリックコメント募集のお知らせ」が公表された¹⁴¹。

●政府は 7 月 13 日、生成 AI サービスを規制する新しい規則「生成 AI サービスの管理に関する暫定措置」を公表した¹⁴²。この規則は 8 月 15 日から施行された。草案では「研究開発」も規制の対象であったが、一般に利用可能な生成 AI サービスのみが対象に変更された。また、モデルの訓練に使用したデータの正確性を検証することも企業に求めていたがこれも削除されたようである。第 4 条の 1 では、生成 AI を用いて生成されたコンテンツについて、「社会主義の核心的価値観を堅持し、国家権力の転覆、社会主義体制の転覆を扇動し、国家の安全と利益を危うくし、国のイメージを損ない、分離独立を扇動し、国家の団結と社会の安定を損ない、テロリズムと過激主義を助長し、民族憎悪と民族差別を助長し、暴力、わいせつ、ポルノ、虚偽の有害情報など、法律と行政法規で禁止されているコンテンツを生成してはならない」としている。

3.8. 国際機関

●OECD から 2023 年 4 月、「AI 言語モデル：技術的、社会経済的、および政策的検討」と題する報告書が公表された¹⁴³。第 1 章では各国での AI 言語モデルに関する動向、第 2 章では AI

¹⁴⁰ <https://consult.industry.gov.au/supporting-responsible-ai>

¹⁴¹ http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm（中国語）

¹⁴² http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm（中国語）

¹⁴³ https://www.oecd-ilibrary.org/science-and-technology/ai-language-models_13d38f92-en

言語モデルを取り巻く環境、そして第3章では OECD で 2019 年に採択された AI 原則¹⁴⁴に即して AI 言語モデルの課題（様々なリスクも含めて）が取り上げられた。

●OECD は 5 月末、「AI に関する OECD 原則」を更新し、生成 AI をカバーする予定であると、Mathias Cormann 事務総長が日本経済新聞に語ったことが報じられた¹⁴⁵。「AI に関する OECD 原則」は 2019 年 5 月に起草され、翌月の G20 首脳会議で首脳の間で共同コミュニケの付属文書として採択された。

●UNESCO は 3 月 30 日、AI 研究のモラトリアムを求める Future of Life Institute からの声明を受け、各国に対して UNESCO が 2021 年に制定した「AI の倫理に関する勧告」¹⁴⁶を直ちに実施することを求めた。「AI の倫理に関する勧告」は 193 の加盟国により全会一致で採択された規範の枠組みであり、生成 AI の発展で必要とされている保護手段を含んでいるとした¹⁴⁷。

●UNESCO は 4 月 14 日、高等教育機関が ChatGPT 等の生成 AI にどう向き合えばよいかについてのクイックスタートガイド（Quick Start Guide）を公開するとともに、インタラクティブセミナーを実施することを発表した¹⁴⁸。課題と倫理的含意として、学術的公正、規制の欠如、プライバシー懸念、認知バイアス、ジェンダーと多様性、アクセシビリティ、商業化が挙げられている。UNESCO は 2023 年中に「A Manual for Higher Education Stakeholders（高等教育関係者のためのマニュアル）」を公表予定である。

●UNESCO は 5 月 25 日に初の世界教育大臣会合をオンラインで開催し、生成 AI ツールが教育にもたらす影響やリスクが議論された¹⁴⁹。また、450 以上の学校と大学を対象とした新しい UNESCO の調査によると、生成 AI アプリケーションの使用に関する組織の方針および/または正式なガイダンスを策定しているのは 10%未満であることがわかったという。

¹⁴⁴ <https://oecd.ai/en/ai-principles>

¹⁴⁵ <https://asia.nikkei.com/Editor-s-Picks/Interview/OECD-pursues-guidelines-on-regulating-generative-AI-leader>

¹⁴⁶ https://www.mext.go.jp/unesco/009/1411026_00004.htm; https://unesdoc.unesco.org/ark:/48223/pf0000381137_eng

¹⁴⁷ <https://www.unesco.org/en/articles/artificial-intelligence-unesco-calls-all-governments-implement-global-ethical-framework-without>

¹⁴⁸ <https://www.iesalc.unesco.org/en/2023/04/14/chatgpt-and-artificial-intelligence-in-higher-education-quick-start-guide-and-interactive-seminar/>「クイックスタートガイド」については日本語訳を作成した人がいる。<https://note.com/pogohopper8/n/n7ab895a74734>

¹⁴⁹ <https://www.unesco.org/en/articles/ai-unesco-mobilizes-education-ministers-around-world-co-ordinated-response-chatgpt>

●WHO は 5 月 16 日、AI が生成した大規模言語モデルツール（LLM）を、人間の幸福、人間の安全、自律性を保護・促進し、公衆衛生を保全するために使用することに注意を払うよう呼びかけた¹⁵⁰。懸念点として挙げられたのは以下のとおりである。

- AI を訓練するために使用されるデータが偏っている可能性があり、健康、公平性、包摂性にリスクをもたらす誤解を招く/不正確な情報を生成する
- LLMs は、エンドユーザーにとって権威があり、もっともらしく見える回答を生成するが、これらの回答は、特に健康に関する回答では、完全に間違っていたり、深刻な誤りを含んでいたりすることがある
- LLMs は、そのような利用について事前に同意が得られていない可能性のあるデータによって訓練されている可能性があり、また、ユーザーが応答を生成するために提供する機微データ（健康データを含む）を保護しない場合がある
- LLMs は、一般の人々が信頼できる健康コンテンツと区別することが困難なテキスト、音声または映像コンテンツの形で、非常に説得力のある偽情報を生成し拡散するために悪用される可能性がある
- WHO は、AI やデジタルヘルスを含む新技術を活用して人々の健康を改善することに尽力しつつ、テクノロジー企業が LLMs の商業利用に取り組む中で、政策立案者が患者の安全と保護を確保することを勧告している

●国連安全保障理事会（UN Security Council）は 7 月 18 日、国際平和と安全保障に対する AI の潜在的脅威に関する初の会合を開催した。映像¹⁵¹が公開されている。15 名からなる理事会は、グテーレス国連事務総長、AI スタートアップ企業 Anthropic の共同設立者である Jack Clark 氏、AI 倫理とガバナンスのための中国・英国研究センターの共同ディレクターである Zeng Yi 教授から説明を受けた。

グテーレス国連事務局長は、加盟国に対して、1) 責任ある AI の設計・開発・使用に関する国家戦略の策定、2) AI の軍事利用をめぐる規範・規則・原則を策定する多国間プロセスに関与、3) テロ対策を目的とした AI を含むデータ駆動技術の利用を規制し、監視メカニズムを強

¹⁵⁰ <https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health>

¹⁵¹ <https://media.un.org/en/asset/k1j/k1ji81po8p>

化するためのグローバルな枠組みへの合意、を求めた¹⁵²。具体的には、今年末までに「AI に関するマルチステークホルダー諮問機関（Multistakeholder Advisory Body on Artificial Intelligence）」を設立するために、個人の資格でメンバーを務める専門家の指名のための公募を開始した¹⁵³。推薦受付の締め切りは 8 月 31 日。また第 1 回会合に向けて以下のテーマ（これら以外のテーマも排除されているわけでない）に関する論文を 9 月 30 日締め切りで募集している。

- グローバル AI ガバナンスの重要課題：AI ハイレベル諮問委員会にとって重要なテーマや優先課題、特にグローバル・ガバナンスを通じて特に取り組むべきテーマについての研究および提言。
- グローバル AI ガバナンスにおける現在の取り組み：二国間、多国間、地域間のイニシアチブを網羅し、異なる哲学的アプローチ、批判的視点、提言を含む分析。
- グローバル AI ガバナンスのモデル：既存のモデルの分析、または新しいモデルの提案、その他の提案の調査や分析。

3.9. 非政府機関

●2023 年 3 月 31 日付で、ベルギーの研究者らによるオープンレター「操作可能な AI への準備はまだできていない一対策を急げ」が公表された¹⁵⁴。こちらは、ベルギーの男性がチャットボットとの会話の中でそそのかされて自殺をしたとされる事件¹⁵⁵などを受け、操作（manipulative）リスクに焦点を当てたものである。本レターでは、AI システムに関連するリスクをただちに周知し、AI 開発者が事前に確実にリスクを特定し、テストし、対処する責任を負うことを促す啓発キャンペーンを行うことを呼び掛けている。また、AI が人々の基本的権利に対して及ぼしうる影響に関する研究にもっと投資することも訴えている。

¹⁵² <https://press.un.org/en/2023/sgsm21880.doc.htm>

¹⁵³ <https://www.un.org/techenvoy/content/artificial-intelligence>

¹⁵⁴ <https://www.law.kuleuven.be/ai-summer-school/open-brief/open-letter-manipulative-ai>

¹⁵⁵ <https://www.brusselstimes.com/belgium/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt>

●3月22日にオープンレターを公表した Future of Life Institute は4月12日、先進 AI システムによるリスクに対抗するための政策立案者向けの勧告を発表した¹⁵⁶。

- 強力な第三者監査と認証を義務化すること。
- 計算能力へのアクセスを規制すること。
- 国家レベルで能力のある AI 機関を設立すること。
- AI による被害に対する責任を確立すること。
- AI モデルの漏洩を防止し、追跡するための措置を導入すること。
- 技術的な AI 安全性研究のための資金提供を拡大すること。
- AI 生成コンテンツや推奨事項の識別および管理のための基準を策定すること。

●AI Now Institute の研究者らによる、「汎用目的 AI (GPAI)」には深刻なリスクがあり、来るべき EU の AI Act で除外されてはいけないという政策提言が4月13日、公表された¹⁵⁷。EU の規制当局に対して下記の5つの指針を示した。本提言には50を超える組織や個人が署名している。

- 汎用目的 AI は広範なカテゴリーである。EU の AI Act が将来も通用するものであるためには、チャットボット／大規模言語モデル (LLM) に限定されるのではなく、様々な技術に適用されなければならない。
- 汎用目的 AI モデルには固有のリスクがあり、実証された広範な害が発生している。これらのリスクが下流の様々な用途に持ち越されうるという事実は、開発段階での規制の必要性を軽減するどころか、むしろ高めている。
- 汎用目的 AI は、関係するステークホルダーが多様であることを考慮できるように、アプリケーション段階だけでなく、製品サイクル全体を通して規制されなければならない。最初の開発段階が重要であり、これらのモデルを開発する企業は、使用するデータや設計上の選択についてアカウンタブルでなければならない。
- 汎用目的 AI の開発者が、標準的な法的免責事項を用いて責任を放棄することを認めるような規制アプローチは見当違いである。汎用目的 AI の開発者（多くの場合、十分な

¹⁵⁶ https://futureoflife.org/wp-content/uploads/2023/04/FLI_Policymaking_In_The_Pause.pdf

¹⁵⁷ <https://ainowinstitute.org/news/gpai-is-high-risk-should-not-be-excluded-from-eu-ai-act>

リソースを持つ大企業）を免責し、代わりに、すべてのリスクを軽減するリソース、アクセス、能力を持たない下流のアクターだけに責任を負わせる危険な抜け穴を生じさせることになる。

- 表面的なチェックボックス作業となることが予想される、汎用目的 AI に対する限定的な評価・精査方法を支持するような規制は避けるべきである。これは活発に議論されている研究分野であり、市民社会、研究者、その他の業界外からの参加者が参加する幅広い検討の対象となるべきものである。

●Center for AI Safety は 5 月 30 日、Statement on AI Risk（AI リスクに関する声明）を公表した¹⁵⁸。「AI による人類絶滅のリスクを軽減することは、パンデミックや核戦争といった他の社会的規模のリスクと並んで、世界的な優先事項であるべきです。」というシンプルなものである。Sam Altman 氏ら多数の AI 専門家が署名している。Future of Life Institute（FLI）はこの Center for AI Safety の声明に対して賛意を表明した¹⁵⁹。そのうえで下記のような提言を行った。

- 核拡散防止条約（Treaty on the Non-Proliferation of Nuclear Weapons, NPT）や生物兵器禁止条約（Biological Weapons Convention, BWC）のような国際協定を締結する。
- 国際原子力機関（IAEA）のような政府間組織を設立する。
- 国レベルでは、厳格な監査や免許制度を確立し、強力な AI システムに適用する。新薬の市場導入に関する食品医薬品局（Food and Drug Administration, FDA）の承認プロセスと同様な事前承認と上市後の監督を義務付ける。
- 生命科学や原子力分野でのデュアルユース研究の公表への制約と同様に、強力な AI システムについては、厳格な安全性と倫理要件を満たさなければ、オープンソースでの公表は禁止する。
- 大規模な汎用 AI システムの現在の最先端を大幅に超えるような、極めて強力な AI システムの開発を一時停止する。

¹⁵⁸ <https://www.safe.ai/statement-on-ai-risk>

¹⁵⁹ <https://futureoflife.org/ai-policy/fli-on-a-statement-on-ai-risk-and-next-steps/>

●Accountable Tech、AI Now Institute、EPIC（Electronic Privacy Information Center, 電子プライバシー情報センター）は 8 月 10 日、企業による「自主規制」への「解毒剤」として、「ゼロ・トラスト AI ガバナンス枠組み（Zero Trust AI Governance Framework）」を共同発表した¹⁶⁰。枠組み提案は、次の 3 つの原則により大まかに整理されている。

1. 時間が最も重要である - 既存の法律を強力に執行することから始める。
3. 大胆で、運用しやすく、明確に線引きできるルールが必要である。
3. AI システムのライフサイクルの各段階において、そのシステムが有害でないことを証明する責任は企業にあるべきである。

●100 を超える産業、学術、非営利、メディア組織から選ばれたパートナーを擁するグローバルネットワークである Partnership on AI (PAI) は 2022 年から大規模モデルの責任ある実装のための共有プロトコル（Shared Protocols）を確立するためのマルチステークホルダープロセスを進めてきた¹⁶¹。これは 2022 年 12 月に作られた「安全クリティカル AI 運営委員会（Safety Critical AI Steering Committee）」がリードしている。2023 年 4 月に生成 AI に関する新しい取り組みを開始し、6 月には「包摂的 AI のためのグローバルタスクフォース」を立ち上げた。2023 年 10 月に、共有プロトコル（Shared Protocols）の第一版を公開し、パブリックコメントを募集する予定である。

4. その他の ELSI 論点

前 ELSI NOTE では、画像生成 AI とテキスト生成 AI に分けて、ELSI 論点を、データ、著作権、バイアス、プライバシー&セキュリティ、情報環境へのインパクト、自然環境へのインパクト、高度な AI システムで生じるリスクの 7 項目に分けて整理した。本稿では、悪用（4.1）、偽情報（幻覚）（4.2）、バイアス・ステレオタイプ（4.3）、著作権（4.4）、デュアルユース（4.5）、労働者の搾取（4.6）、人類破滅リスク（4.7）を取り上げた。

¹⁶⁰ <https://accountabletech.org/media/accountable-tech-ai-now-epic-release-zero-trust-ai-governance-framework/>

¹⁶¹ <https://partnershiponai.org/a-call-for-ai-providers-to-deploy-models-responsibly/>

4.1. 悪用

●米国アリゾナ州で 2023 年 4 月、地元テレビニュース局が、Jennifer DeStefano という母親が見知らぬ電話番号からの電話をとると、15 歳の娘が「ママ！」と言って泣き叫ぶ声が聞こえ、男から娘の返還のために 100 万ドルの身代金を支払うよう要求された（実際には 5 万ドルまで下げた）というニュースを報じた¹⁶²。しかし、娘は実際には誘拐されておらず無事であった。ボイスクローン技術を使ったものだと考えられ、3 秒の声データがあれば作成可能だという。DeStefano 氏はその後、6 月 13 日に上院司法委員会の公聴会に証人として呼ばれた（3.3 節参照）。

●ドイツの女性週刊誌 Die Aktuelle は 4 月 15 日号の表紙に笑顔のミハエル・シューマッハ氏の写真を掲載し、「ミハエル・シューマッハ、初のインタビュー」という見出しをつけた¹⁶³。しかし、7 度の F1 チャンピオンに輝いたシューマッハ氏は、2013 年 12 月にスキー事故で頭部に重傷を負い、自宅に戻って以来、病状は家族によって非公開にされている。この記事は charatcter.ai という AI プログラムを使って作成され、健康状態や家族に関するシューマッハの「引用」を人工的に生成したという。家族は雑誌に対して法的措置を取ることを計画しているという。

●中国で女性歌手、孫燕姿（ステファニー・スン）の音声データを利用した「AI 孫燕姿（ステファニー・スン）」の楽曲が動画配信プラットフォームに多数公開されていることが報じられた¹⁶⁴。彼女は 2017 年以来楽曲をリリースしていない。

●Drake と The Weeknd によるものとされる新曲「Heart on My Sleeve」が TikTok と Spotify で公開され、ネット上で広まった。しかしこの曲は TikTok のユーザーである Ghostwriter977 が、Drake と The Weeknd の作品を AI に学習させ、彼らの声、歌詞、音楽スタイルを完璧に模倣したものであった¹⁶⁵。彼らのレコード会社であるユニバーサルミュージックグループからのクレームによって TikTok や Spotify などのプラットフォームからは削除された。

¹⁶² <https://www.azfamily.com/2023/04/10/ive-got-your-daughter-scottsdale-mom-warns-close-encounter-with-ai-voice-cloning-scam/>

¹⁶³ <https://www.bbc.com/sport/formula1/65333115>

¹⁶⁴ <https://www.afpbb.com/articles/-/3464304>

¹⁶⁵ <https://hls.harvard.edu/today/ai-created-a-song-mimicking-the-work-of-drake-and-the-weeknd-what-does-that-mean-for-copyright-law/>

●5月22日、Twitter（現 X）で Verified Blue Check を受けたブルームバーグの偽アカウントが、明らかに AI が生成した、米国ワシントン DC のペンタゴンで爆発が起きたフェイク画像を投稿した¹⁶⁶。実際にはペンタゴンで爆発はなかったがこのフェイク画像によって株価が下落する事態が生じた。当該偽アカウントはすでに削除されている。

●共和党の大統領候補である DeSantis 氏の支援キャンペーン団体が6月5日、ライバルのトランプ氏に対する攻撃広告動画において、AI で生成したディープフェイクと専門家が認識するものを使用したことが報じられた¹⁶⁷。感染症対策の中心人物であるファウチ氏と一緒に写っている写真の画像に不自然な点があることや、画像の逆引き検索で1つも見つからないことも証拠として挙げられた。生成 AI の登場で、選挙キャンペーンにおけるパロディと偽情報の境界があいまいになることも指摘されている。また、DeSantis 氏を支持する団体は、トランプ氏を攻撃する新しいテレビ CM で、生成 AI で再現したトランプ氏の声を使用している¹⁶⁸。

●Amazon 社の Kindle Unlimited のヤングアダルト・ロマンス・ベストセラーリストが6月26～27日に AI が生成したナンセンスな本で埋め尽くされたことが報じられた¹⁶⁹。28日の朝の時点で、Amazon 社はこれらの本に対して措置をとったようだが、AI が生成したナンセンスな本をプラットフォームにスパム送信し、それを収益化する方法が見つけられたことを示唆している。

4.2. 偽情報（幻覚）

●ChatGPT が「視観」という新語を発明したらしいという話題が twitter（現 X）でとりあげられ、その単語が出てくるウェブサイトを集めたページが作られた¹⁷⁰。他にも同様の ChatGPT による造語があるようである。

¹⁶⁶ <https://www.insider.com/ai-generated-hoax-explosion-pentagon-viral-markets-dipped-2023-5>

¹⁶⁷ <https://www.theverge.com/2023/6/8/23753626/deepfake-political-attack-ad-ron-desantis-donald-trump-anthony-fauci>

¹⁶⁸ <https://www.politico.com/news/2023/07/17/desantis-pac-ai-generated-trump-in-ad-00106695>

¹⁶⁹ <https://www.vice.com/en/article/v7b774/ai-generated-books-of-nonsense-are-all-over-amazons-bestseller-lists>

¹⁷⁰ <https://okumuralab.org/~okumura/misc/230611.html>

●Google 社の検索エンジンが、2023 年 5 月初めに、Edward Hopper の検索結果のトップ結果を AI 生成の偽物にしていたことが報じられた¹⁷¹が、今度は、Johannes Vermeer（フェルメール）で検索すると、代表作「真珠の耳飾りの少女」を AI が再現したものがトップで表示される事態が生じた¹⁷²。ちなみに、この AI 版「真珠の耳飾りの少女」は、本物がフェルメールの回顧展のために貸し出されている間に、Mauritshuis 美術館がその代わりに飾ることにしたファンによって制作された 4 点の 1 つが、AI アーティストの Julian van Dieken が制作したこの作品であった。3,500 点以上の応募の中から選ばれたこともあり、この決定はアートコミュニティの間で議論を呼んだ。

●米国で 5 月末、Schwartz 弁護士が裁判所に提出した準備書面が「存在しない判例の引用で埋め尽くされている」こと、これは ChatGPT の回答をそのまま引用したことによることが報道された¹⁷³。この事件は、ある男性が、2019 年 8 月のエルサルバドルからニューヨークへのフライト中に、金属製の配膳カートが膝に当たって負傷したとして、航空会社 Avianca を提訴したものである。Avianca 社が、時効が成立していることを理由に訴えを却下するよう求めたのに対して、被害男性の弁護士である Schwartz 氏は、マルチネス対デルタ航空、ジシャーマン対大韓航空、ヴァルゲーゼ対中国南方航空といった判例を引用して、訴訟の続行を認めるべきとの主張を支持する 10 ページの準備書面を提出した。Avianca 社の弁護士がこれらの裁判例を見つけられなかったため、裁判所から資料を提出するよう命じられ、判決文の概要を提出したがこれも存在しないものであった¹⁷⁴。6 月 8 日、本件についての公聴会が開かれ、Schwartz 弁護士は裁判官から尋問を受けた。「ChatGPT が事件を捏造できることを理解していなかった」「ChatGPT に騙され続けている。恥ずかしい限りです」などと彼は裁判官に語ったという¹⁷⁵。ChatGPT のことを、標準的なデータベースよりも広い検索ができる「スーパーサーチエンジン

¹⁷¹ <https://futurism.com/top-google-result-edward-hopper-ai-generated-fake>

¹⁷² <https://futurism.com/top-google-result-johannes-vermeer-ai-generated-knockoff>

¹⁷³ <https://reason.com/volokh/2023/05/27/a-lawyers-filing-is-replete-with-citations-to-non-existent-cases-thanks-chatgpt/>

¹⁷⁴ Gigazine による日本語での紹介記事 <https://gigazine.net/news/20230529-lawyer-filing-with-non-existent-cases-chatgpt/>

¹⁷⁵ <https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html>

のようなもの」だと勘違いしていたという。6月23日、裁判官は、Schwartz氏ら2人の弁護士とその事務所が連帯責任で支払う5,000ドルの罰金を言い渡した¹⁷⁶。

●初の(?) ChatGPTによる名誉棄損訴訟が米国で提起されたことが報じられた¹⁷⁷。これはWalters氏がOpenAI社を相手に、ジョージア州Gwinnettカウンティの裁判所に提出した申立てで、ジャーナリストのRiehl氏が別の裁判の原告の訴状をChatGPTに要約させたが、そこに書かれているWalters氏に関する記述はすべて虚偽であり、OpenAI社はこのことを知っていたか、知るべきであったことが名誉棄損に当たるとした¹⁷⁸。

●ブリティッシュ・コロンビア大学のKlein教授は、ChatGPTなどの生成AIの不正確なアウトプットがhallucinations（幻覚）と呼ばれる傾向を批判している¹⁷⁹。「幻覚」は心理学からの専門用語であるため、生成AIの擬人化を促し、生成AIの限られた能力の過大評価につながるものが理由として挙げられている。

4.3. バイアス・ステレオタイプ

●Bloombergの記事は、画像生成ソフトのStable Diffusionを使って、14の職種（米国で一般的に「高給」とされる7つの職種と「低給」とされる7つの職種）と、犯罪に関連する3つのカテゴリーについて、5100枚の労働者の画像を生成・分析した結果、人種と性別のバイアスのパターンを見出したことを報告した¹⁸⁰。高給とされる職種では明るい肌色の被写体が多く、低給とされる職種では暗い肌色の被写体が多くなった。米国の裁判官の34%が女性であるにもかかわらず、「裁判官」というキーワードで生成された画像のうち、女性だと認識できた画像はわずか3%だった。米国のファストフードの従業員の70%が白人であるにもかかわらず、モデルは70%の確率で肌の色が濃い人を生成した。米国の刑務所の受刑者の半分以下が有色人種であるにもかかわらず、「inmate（囚人）」というキーワードで生成された画像は、肌の黒い人

¹⁷⁶ <https://arstechnica.com/tech-policy/2023/06/lawyers-have-real-bad-day-in-court-after-citing-fake-cases-made-up-by-chatgpt/>

¹⁷⁷ <https://reason.com/volokh/2023/06/06/first-ai-libel-lawsuit-filed/printer/>

¹⁷⁸ Gigazine による日本語での紹介記事 <https://gigazine.net/news/20230609-chatgpt-hallucination-first-defamation-suit/>

¹⁷⁹ <https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein>

¹⁸⁰ <https://www.bloomberg.com/graphics/2023-generative-ai-bias/>

が肌の白い人の5倍多くなった。「terrorist（テロリスト）」というキーワードに対して、宗教的な頭巾をかぶった黒い髭を持つ画像ばかりを生成した。つまり、現実世界に存在するステレオタイプや格差を再現するだけでなく、それらを拡大させていることが分かった。そして、AI がもたらす最大の脅威は、人類の絶滅ではなく、不平等を拡大する可能性であると警告した。

●スタンフォード大学の Santurkar らの研究（Santurkar et al. 2023）は、9つの代表的なテキスト生成 AI モデルを対象に、それらのアウトプットが、Pew Research 社の世論調査に見られる米国の人口統計グループの意見とどの程度一致しているかを定量化した¹⁸¹。モデル間に大きなばらつきがあり、一般的に、インターネットデータだけで訓練されたモデルは低学歴・低所得・保守的な視点に偏る傾向があるのに対して、人間のフィードバックによって改良されたモデルはよりリベラル、高学歴・高収入の視点に偏る傾向があるという。ただし、テキスト生成 AI が現在の一般大衆の意見と完全に一致させることを目標とすべきか否かが問題ではなく、モデルにはそれぞれ政治的バイアスがあることを意識することが重要だと著者らは指摘している。

●同じスタンフォード大学の Bianchi らの研究（Bianchi et al. 2023）は、誰でも利用できる画像生成 AI ツールである Stable Diffusion モデルを使って、特性（例：魅力的な、貧しい）、職業（例：ソフトウェアエンジニア、ハウスキーパー）、物体（例：衣服、家）といった単純なプロンプトによって生成される人物の描写に、ジェンダー、人種、文化や性的指向を含む様々な種類のバイアス（ステレオタイプ）が反映されるとし、建物や車などの背景にある要素の描写にも現れていることを明らかにした¹⁸²。画像生成 AI モデルが社会で広く利用されると、社会経済的地位、人種、性別などのアイデンティティの複数の要素の交差に関する固定概念（ステレオタイプ）が再生産される懸念を指摘している。

¹⁸¹ <https://hai.stanford.edu/news/assessing-political-bias-language-models> Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P. and Hashimoto, T. (2023). Whose Opinions Do Language Models Reflect? arXiv:2303.17548 arXiv:2303.17548

¹⁸² Bianchi F, Kalluri P, Durmus E, Ladhak F, Cheng M, Nozza D, et al. Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale. FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. pp. 1493–1504. <https://doi.org/10.1145/3593013.3594095>

4.4. 著作権

●Financial Times 紙が 2023 年 4 月、ユニバーサルミュージックグループが、Spotify や Apple を含むストリーミング・プラットフォームの運営主体に対して、AI プラットフォームサービスが著作権のある楽曲からメロディや歌詞をスクレイピングするのをブロックするよう電子メールで指示したことを報じた¹⁸³。

●集英社・週刊プレイボーイが 5 月 29 日に“AI グラビア”写真集「生まれたて。」を発売した¹⁸⁴。「生まれたて。」は AI で生成した「さつきあい」という実在しないモデルによる実写風画像をまとめたグラビア写真集であり、電子書籍として 499 円で販売された。画像は週刊誌「週刊プレイボーイ」の編集部が生成したとされる。

ところが、集英社・週刊プレイボーイ編集部が 6 月 7 日、画像生成 AI を使った写真集『生まれたて。』を販売終了することを公式サイトで発表した¹⁸⁵。「本企画について発売後よりたくさんのご意見を頂戴し、編集部内で改めて検証をいたしました。その結果、制作過程において、編集部で生成 AI をとりまく様々な論点・問題点についての検討が十分ではなく、AI 生成物の商品化については、世の中の議論の深まりを見据えつつ、より慎重に考えるべきであったと判断するにいたしました。」とのことであった。

●武蔵野美術大学が、教育ベンチャー「スタディプラス」とネットサービス会社「カヤック」とともに「武蔵野 AI 美術大学 AI 絵画アワード」¹⁸⁶を主催し、6～7 月に作品を募集し、400 点以上の応募があったが、急きょ中止になったことが報道された¹⁸⁷。ウェブサイトには「協議の結果、「受賞作品なし」と判断いたしました」と書かれ、その理由として「運営および審査過程において、現在の画像生成 AI 技術によって生み出された作品から、特定の作品を選出し発表することは困難であると判断いたしました」と記載されている。Stable Diffusion を元にしたアプリ「AI ピカソ」を使うことを条件としていたが、アプリの説明として、「著作権が切れ

¹⁸³ <https://www.ft.com/content/aec1679b-5a34-4dad-9fc9-f4d8cdd124b9>

¹⁸⁴ <https://www.itmedia.co.jp/news/articles/2305/29/news148.html>

¹⁸⁵ <https://www.grajapa.shueisha.co.jp/post/202306/>

¹⁸⁶ https://www.musabi.ac.jp/ai_award/

¹⁸⁷ <https://www.yomiuri.co.jp/national/20230805-OYT1T50207/>

た画像のみを学習させたサービス」としていたにもかかわらず、著作権のある画像も学習させているのではないかと外部から指摘が寄せられたことで、7月13日に説明を修正し、謝罪したという。

●New York Times 紙が8月3日、利用規約 (terms of service) ¹⁸⁸を更新し、機械学習や AI システムを訓練するためにコンテンツをスクレイピングすることを禁止した¹⁸⁹。

2. サービスのコンテンツ

2.1 本サイトを含む本サービスのコンテンツは、お客様の個人的かつ非商業的な使用を目的としています。(中略) 非商業的利用には、New York Times 社からの書面による事前の同意なしに、以下に関連してコンテンツを利用することは含まれません：(1) 機械学習または人工知能 (AI) システムのトレーニングを含むがこれに限定されないソフトウェアプログラムの開発、または (2) コンテンツを含むアーカイブまたはキャッシュされたデータセットを他の個人または団体に提供すること。

4. サービスの禁止された利用

4.1 ユーザーは、当社または第三者に損害を与える可能性のある行為を行うために、本サービスにアクセスまたは使用したり、アクセスまたは使用を試みたりしてはならないものとします。(中略) 利用者は、NYT の書面による事前の承諾がない限り、以下の行為を行わないものとします： (中略)

(2)本サービスからコンテンツ、データまたは情報をデータマイニングまたはスクレイピングするように設計された、ロボット、スパイダー、スクリプト、サービス、ソフトウェア、または手動/自動のデバイス、ツール、またはプロセスを使用すること、またはその他の自動化された手段を使用して本サービスからコンテンツ、データまたは情報を使用、アクセス、収集すること；

(3) 機械学習または人工知能 (AI) システムのトレーニングを含むがこれに限定されない、ソフトウェアプログラムの開発のためにコンテンツを使用すること。(中略)

¹⁸⁸ <https://help.nytimes.com/hc/en-us/articles/115014893428-Terms-of-Service>

¹⁸⁹ <https://arstechnica.com/information-technology/2023/08/the-new-york-times-prohibits-ai-vendors-from-devouring-its-content/>

4.2 本サービスの禁止された利用に関与した場合、ユーザーおよびユーザーを支援する者に対して、民事、刑事、および/または行政上の罰則、罰金、または制裁が課される場合があります。

NPR は 8 月 16 日、New York Times 紙の弁護士が、報道に関連する知的財産権を保護するために、連邦著作権法に基づき、OpenAI 社を提訴するかどうかを検討していると報道した¹⁹⁰。

4.5. デュアルユース

●The Intercept は 2023 年 5 月 8 日付の記事で、OpenAI 社が生成 AI 技術の軍事利用を許すかどうか、そして企業が自主的に定めた原則の実効性はどの程度なのかという問題提起を行った¹⁹¹。3 月に開催された新興技術と安全保障に関する会議において、国家地理空間情報局 (NGA) の副長官が ChatGPT の利用に前向きな発言を行ったことをとりあげ、これが OpenAI 社の利用方針 (Usage Policies) ページの文言 (禁止された利用方法のリストに、兵器開発、軍事や戦争を含む「身体的危害を及ぼす高いリスクのある活動」が含まれている) に反するのではないかと The Intercept が、OpenAI 社 CEO の Sam Altman 氏に尋ねたところ広報担当者 Alex Beck 氏を紹介されたが、会議での発言に関するコメントや質問には一切答えなかったという。利用方針ページには「お客様の製品または使用方法がこれらのポリシーに従っていないことが判明した場合、当社はお客様に必要な変更をお願いすることがあります」や「違反が繰り返されたり、深刻だった場合は、アカウントの停止や終了を含むさらなる措置がとられる場合があります」と書かれている¹⁹²。

●米国の国防総省 (Department of Defense, DoD) は 8 月 10 日、生成 AI タスクフォース “Lima” の設立を発表した¹⁹³。DoD 内の Chief Digital and Artificial Intelligence Office (CDAO) が率いる。CDAO は 2022 年 6 月に業務を開始し、DoD 全体の AI 能力の統合と最適化に取り組んでいるという。

¹⁹⁰ <https://www.npr.org/2023/08/16/1194202562/new-york-times-considers-legal-action-against-openai-as-copyright-tensions-swirl>

¹⁹¹ <https://theintercept.com/2023/05/08/chatgpt-ai-pentagon-military/>

¹⁹² <https://openai.com/policies/usage-policies>

¹⁹³ <https://www.defense.gov/News/Releases/Release/Article/3489803/dod-announces-establishment-of-generative-ai-task-force/>

4.6. 労働者の搾取

●OpenAI 社のコンテンツ・モデレーションがアウトソーシングされ、搾取的な労働条件が批判された Sama 社の従業員を含む、African Content Moderators Union（アフリカコンテンツモデレーター連合）が 2023 年 5 月に組織された¹⁹⁴。

●Wall Street Journal 紙の Podcast サービスである The Journal が 7 月 1 日、「ChatGPT から暴力と虐待を排除した隠れた労働力」と題するストーリーを公開し、OpenAI 社が Sama 社を通して有害情報の取扱い作業を委託したケニアのデータ労働者当事者へのインタビューを掲載した¹⁹⁵。コンテンツモデレーションフィルターを構築する作業にケニアに在住する約 50 人のデータ労働者が従事した。ケニアが選ばれたのは、英語を話す高学歴の労働力が多く、無線 LAN のインフラが整備されているにもかかわらず、失業率が非常に高く、賃金が低いことが挙げられた。フィルターを有効に機能させるためには、AI に有害情報が何かを知らせなければならず、それは人間が行わなければならず、Sama 社の従業員の仕事であった。インタビューで明らかになったことは、OpenAI 社は Sama 社に 1 時間あたり 12.50 ドルのサービス料を支払ったが、チームの人々の時給は 1.5 ドルから 3.75 ドル程度だったことである。OpenAI 社の広報担当者は、テキストを確認する作業員がそのごく一部しか得ていないことは知らなかったと述べた。Sama 社のスポークスウーマンは、この 12.50 ドルの手数料はインフラ費用など他のものもカバーしており、コンテンツモデレーターには生活賃金を決定するための公認の公式に従って賃金が支払われていると述べた。

4.7. 人類破滅リスク

●Nature 誌の 2023 年 6 月 29 日付の Editorial「AI の既知のリスクについて話すときだ」は、AI 専門家を中心にたびたび聞かれる「AI が人類を破滅させる」という話は、ハイテク企業の思惑にまんまとはまり、AI が今まさに引き起こしている社会的危害に対する効果的な規制を妨げ

¹⁹⁴ <https://time.com/6275995/chatgpt-facebook-african-workers-union/>

¹⁹⁵ <https://www.wsj.com/podcasts/the-journal/the-hidden-workforce-that-helped-filter-violence-and-abuse-out-of-chatgpt/ffc2427f-bdd8-47b7-9a4b-27e7267cf413>

ると主張した¹⁹⁶。そうした終末論の問題点として次の2点を挙げた。1点目は、終末論がもたらす、AIが万能の機械であるというイメージは、国家間のAI開発競争を煽り、結果として規制を弱め、業界にとって有利な状況を生み出す。2点目は、AIのリスクや規制について、企業幹部や技術者からなる同質的なグループが会話を支配し、他のコミュニティが取り残されることになる。

●BCS（British Computer Society, 英国コンピューター協会）が英国政府と産業界に向けたAIに関するオープンレターを作成・公開し、署名を集めている¹⁹⁷。

「AIは人類にとって存亡の危機ではなく、その開発と使用に関する重要な決定を正しく行うことができれば、善のための変革的な力となる。英国は、堅牢な行動規範、国際的な協力関係、十分なリソースのある規制によって支えられ、AIの役割における専門的・技術的な基準を設定するための先導的な役割を果たすことができる。そうすることで、「Coded in Britain」は高品質で倫理的、包摂的なAIの世界的な代名詞となることができる。」

7月18日付の記事では、1300人以上の署名が集まったことが報告された¹⁹⁸。情報倫理やAI倫理で有名なLuciano Floridi教授も署名した1人である。

5. 生成AIのためのデータセット

最後の章は、生成AIモデルのためのトレーニング・データセットを取り上げる。OpenAI社は画像生成AIについてもテキスト生成AIについても最新のものについてはそれらのトレーニング・データセットの中身を明らかにしていない。

5.1. 画像生成AIモデル

●2021年1月、OpenAI社は、画像と画像を説明するテキストのペア4億組を学習させた、汎用画像分類モデル（ニューラル・ネットワーク）であるCLIP（Contrastive Language-Image

¹⁹⁶ <https://www.nature.com/articles/d41586-023-02094-7>

¹⁹⁷ <https://www.bcs.org/sign-our-open-letter-on-the-future-of-ai/>

¹⁹⁸ <https://www.bcs.org/articles-opinion-and-research/bcs-open-letter-calls-for-ai-to-be-recognised-as-force-for-good-not-threat-to-humanity/>

Pre-Training) を公開した¹⁹⁹。しかし、OpenAI 社は、CLIP のコードとモデルの重み(model weight)はオープンソース化したものの、データセットについては公開しなかった。

●LAION²⁰⁰は、LAION-400M を使って CLIP モデルを再現することに成功した。LAION は、Large-scale Artificial Intelligence Open Network の略称で、ドイツの非営利団体である。

●LAION は、2022 年 3 月に約 58 億 5000 万個の超巨大な、CLIP を使ってフィルタリングされた「カラー画像」と「テキスト」のペアの研究用データセットである「LAION-5B」を公表した²⁰¹。58.5 億のうち、23 億は英語、22 億は 100 以上の他の言語、10 億は特定の言語を割り当てられないテキストを含むサンプルからなる。

●「この大規模なデータセットは未キュレーションであることに留意してください。このデータセットが未修正であることは、収集されたリンクが人間の閲覧者にとって強い不快感を与えるコンテンツや不穏なコンテンツにつながる可能性があることを意味していることを念頭に置いてください。従って、デモリンクの利用は慎重に、かつ自己責任でお願いします。」と注記がある。

5.2. テキスト生成 AI モデル

●C4 データセット：2023 年 4 月 19 日付のワシントンポスト紙は「ChatGPT のような AI を賢く見せているウェブサイトの秘密リストの内部」を題する記事で、Allen Institute for AI の研究者と協力し、ウェブ解析会社 Similarweb のデータを用いて Google の C4 データセットを分析した結果を報告した²⁰²。

●このうち約 1/3 のウェブサイトは分類できず（多くはアクセスできず）、残りの 1000 万のウェブサイトについて、11 のカテゴリー²⁰³とサブカテゴリーに分類し、「トークン（単語やフレーズ）」の出現数に基づいてランク付けした。上位 3 位は、世界中で発行された特許のテキ

¹⁹⁹ <https://openai.com/research/clip>

²⁰⁰ <https://laion.ai/>

²⁰¹ <https://laion.ai/blog/laion-5b/>

²⁰² <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/>

²⁰³ Business & Industrial (16%) , Technology (15%) , News & Media (13%) , Arts & Entertainment (11%) , Science & Health (9%) , Hobbies & Leisure (8%) , Jobs & Education (7%) , Home & Garden (6%) , Travel (6%) , Community (5%) , Law & Government (4%)

ストを含む patents.google.com で 0.46%、無料のオンライン百科事典である wikipedia.org で 0.19%、購読制のデジタルライブラリーである scribd.com で 0.07% であった。

●C4 は、非営利団体コモンクロール（Common Crawl）が 2019 年 4 月に行ったスクレイプとして始まった。C4 とは、Colossal Clean Crawled Corpus の略で、1500 万のウェブサイトのコンテンツからなる。Common Crawl は Washington Post 紙に対し、最も重要で評判の良いサイトを優先するようにしているが、ライセンスや著作権のあるコンテンツを避けようとはしていない、と述べている。C4 は、Google の T5 や Facebook の LLaMA など、有名な英語 LLM の学習に使用されたことで知られる。

●C4 データセットは、オープンソースの「List of Dirty, Naughty, Obscene, and Otherwise Bad Words」を使用してフィルターをかけている。しかし、性的でない LGBTQ コンテンツを排除してしまったり、逆にポルノサイトなどを残してしまったりしていることも確認されている。また、白人至上主義者のサイト stormfront.org、反トランスのサイト kiwifarms.net、個人に対する標的型嫌がらせキャンペーンを組織することで知られる匿名掲示板 4chan.org などの問題あるコンテンツをフィルターが除去できていないことも発見された。

●コモン・クロール（Common Crawl）²⁰⁴：2008 年以降にウェブサイトのクロールを開始し、そのアーカイブとデータセットをフリーで提供している非営利団体である。コモン・クロールのフィルタリングバージョンは、2020 年に発表された OpenAI 社の GPT-3 の学習に使用されたとされる。Brown et al. (2020)²⁰⁵ の Table 2.2 によると、次のとおりである（Table 1）。コモンクロールに加えて、書籍データやウィキペディアなどからデータセットが作成されていることが分かる。

²⁰⁴ <https://commoncrawl.org/>

²⁰⁵ Tom B. Brown et al. Language Models are Few-Shot Learners.

<https://doi.org/10.48550/arXiv.2005.14165>

Table 1 OpenAI 社の GPT-3 の学習に使用されたデータセット

| データセット | 量（トークンの数） | トレーニングミックスにおける重み | 300B トークンのトレーニング時の経過時間 |
|-------------------------|-------------|------------------|------------------------|
| Common Crawl (filtered) | 410 billion | 60% | 0.44 |
| WebText2 | 19 billion | 22% | 2.9 |
| Books1 | 12 billion | 8% | 1.9 |
| Books2 | 55 billion | 8% | 0.43 |
| Wikipedia | 3 billion | 3% | 3.4 |

●Llama（Large Language Model Meta AI）²⁰⁶：2023 年 2 月に Meta 社が発表した大規模言語モデル。いくつかのサイズ（7B、13B、33B、65B パラメータ）で提供されている。学習のためのデータセットは誰でも利用できるもののみを利用したとされる。

Table 2 Meta 社の LLaMA の学習モデルに使用されたデータセット

| データセット | サンプリング割合 | Epochs | Disk size |
|---------------|----------|--------|-----------|
| CommonCrawl | 67.0% | 1.10 | 3.3 TB |
| C4 | 15.0% | 1.06 | 787GB |
| GitHub | 4.5% | 0.64 | 328GB |
| Wikipedia | 4.5% | 2.45 | 83GB |
| Books | 4.5% | 2.23 | 85GB |
| ArXiv | 2.5% | 1.06 | 92GB |
| StackExchange | 2.0% | 1.03 | 78GB |

²⁰⁶ <https://ai.meta.com/blog/large-language-model-llama-meta-ai/>

5.3. ウェブスクレイピングの是非²⁰⁷

- データサイエンス企業 hiQ Labs 社（現在は活動していない）が、LinkedIn の公開プロフィールに掲載されたデータをスクレイピングにより収集し、分析して販売するビジネスを展開していた。
- これに対して、LinkedIn が、「コンピューター詐欺と濫用に関する法律（Computer Fraud and Abuse Act, CFAA）」違反であるなどとして、hiQ に無許可のデータスクレイピングの停止を要求する停止命令書（cease-and-desist letter）を送達し、訴訟を予告するとともに、hiQ が LinkedIn のサイトにアクセスできないよう技術的措置を講じたことを hiQ に通知した。その後、hiQ は 2017 年に LinkedIn を相手取り、LinkedIn のアンチスクレイピング行為が不正競争、経済的不法行為にあたると主張し、差止命令と宣言的救済²⁰⁸を求める訴訟を起こした。
- 2017 年に仮処分命令が連邦地裁によって出されたが、LinkedIn は控訴裁判所に控訴した。2019 年、第 9 巡回控訴裁判所は地裁判決を支持した。
- 2020 年、LinkedIn は最高裁に、再審理を求めたが、2021 年 7 月、最高裁判所は第 9 巡回控訴裁判所に差し戻した。
- 2022 年 4 月 18 日、第 9 巡回控訴裁判所は、2019 年の仮処分命令を支持し、LinkedIn の主張を退けた。しかし、これは最高裁判所によって取り消され、さらなる検討のために差し戻された。
- 2022 年 11 月、略式判決（summary judgment）において、カリフォルニア州北部地区連邦地方裁判所は、LinkedIn のデータスクレイピングと偽のプロフィールを禁止するウェブサイトのユーザー契約の条項は、契約違反の請求において強制執行可能であるとの判決を下した。
- これを受けて 12 月、LinkedIn と hiQ Labs は、6 年にわたる訴訟の末、データスクレイピングに関連するすべての請求を解決するための同意判決（consent judgment）と終局的差止命令（permanent injunction）で合意した。すなわち、hiQ が、ユーザー規約に違反して会員のプ

²⁰⁷ 以下の話は脚注の記事を参照した。 <https://www.privacyworld.blog/2022/12/linkedin-data-scraping-battle-with-hiq-labs-ends-with-proposed-judgment/>

²⁰⁸ 「いまだ具体的事件としては熟していないものの、原告がその権利について不安・懸念をもつときに、権利関係、法的地位を宣言することにより紛争の終結を目指してなされる制定法上の救済であり、拘束力は当事者間のみですが、同種訴訟がなされても同様の判断が下されることを宣言するのの」 [https://www.shugiin.go.jp/internet/itdb_kenpou.nsf/html/kenpou/chosa/1571002sengoku.pdf/\\$File/1571002sengoku.pdf](https://www.shugiin.go.jp/internet/itdb_kenpou.nsf/html/kenpou/chosa/1571002sengoku.pdf/$File/1571002sengoku.pdf)

ロフィールドデータをスクレイピングした際に作成されたすべてのソースコード、データ、アルゴリズムを破棄し、スクレイピングを停止することを求める永久差止命令に同意した。

6. おわりに

生成 AI を巡っては、2023 年 4 月以降、各国、地域間、国際機関、非政府機関などから、規制措置を含めた、政策に関して具体的な提案が次々と出されている。年末に向けて、何らかの国際的な合意が得られる可能性もあるが、いったん 2023 年 8 月後半までの動きをまとめておくことにした。また、ELSI の各論については、議論が整理されてきた面もあれば、新たな展開が生じている面もある。ただし、学術文献については、本稿では、報道されたものを除くと十分に取り上げることができなかった。学術文献のレビューは別途実施することとする。

また、生成 AI は、これまでの AI 技術に比べて、非専門家が容易に主体的に利用することができるため、誤用や悪用、また想定外の使われ方などが今後も次々と現れることが予想される。そうした点についても、あらかじめ予期し、対処できるような仕組みも検討したい。

ELSI NOTE No.30

令和 5 年 9 月 3 日

**生成 AI の倫理的・法的・社会的課題（ELSI）論点の概観: 2023 年 4～8 月版
ーグローバルな政策動向を中心に**

| | | |
|-----------|-------------------------------|----------------------|
| 岸本 充生 | 大阪大学 社会技術共創研究センター（ELSI センター） | センター長 (2023 年 8 月現在) |
| カテライ アメリア | 大阪大学 社会技術共創研究センター（ELSI センター） | 特任助教 (2023 年 8 月現在) |
| 井出 和希 | 大阪大学 感染症総合教育研究拠点/社会技術共創研究センター | 特任准教授 (2023 年 8 月現在) |

Overview of Ethical, Legal, and Social Issues (ELSI) in Generative AI: August 2023 Edition**-Focusing on Global Policy Trends**

| | |
|-----------------|--|
| Atsuo KISHIMOTO | Research Center on Ethical, Legal and Social Issues, Osaka University |
| Amelia KATIRAI | Research Center on Ethical, Legal and Social Issues, Osaka University |
| Kazuki IDE | Center for Infectious Disease Education and Research (CiDER)/Research Center on Ethical, Legal and Social Issues, Osaka University |



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

〒565-0871 大阪府吹田市山田丘 2-8
大阪大学吹田キャンパステクノアライアンス C 棟 6 階
TEL 06-6105-6084
<https://elsi.osaka-u.ac.jp>



大阪大学