第4章 デジタルテクノロジーの課題と現状の対応策

第1節 AIの進化に伴う課題と現状の取組

進化してきたAIは我々の生活に便利さをもたらす一方で、活用に当たっては留意すべきリスクや課題も存在している。これまで、AI全般についても、不適切なデータや偏ったデータを学習に使用することでモデルのバイアスや誤差が増加し、予測の信頼性が低下する点や、多くの従来の機械学習モデルについてブラックボックス（透明性の欠如）となっていてその内部動作が理解しにくく、重要な意思決定の場面で問題を引き起こす可能性が指摘されていた。これに加え、生成AIが爆発的に発展・普及する中で、特有の課題・リスクも明らかになってきた。以下に生成AIが抱えるリスク・課題を技術的/社会・経済的な観点から概観する。

① 生成AIが抱える課題

2024年4月に総務省・経済産業省が策定した「AI事業者ガイドライン（第1.0版）」では、（従来から存在する）AIによるリスクに加えて、生成AIによって顕在化したリスクについて例示している（図表Ⅰ-4-1-1）。例えば、従来から存在するAIによるリスクとして、バイアスのある結果及び差別的な結果が出力されてしまう、フィルターバブル及びエコーチェンバー現象[*1]が生じてしまう、データ汚染攻撃のリスク（AIの学習実施時の性能劣化及び誤分類につながるような学習データの混入等）、AIの利用拡大に伴う計算リソースの拡大によるエネルギー使用量及び環境負荷[*2]等が挙げられている。また、生成AIによって顕在化したリスクとしては、ハルシネーション等が挙げられる。生成AIは事実に基づかない誤った情報をもっともらしく生成することがあり、これをハルシネーション（幻覚）と呼ぶ。技術的な対策が検討されているものの完全に抑制できるものではないため、生成AIを活用する際には、ハルシネーションが起こる可能性を念頭に置き、検索を併用するなど、ユーザーは生成AIの出力した答えが正しいかどうかを確認することが望ましい。また、生成AIの利用において、個人情報や機密情報がプロンプトとして入力され、そのAIからの出力等を通じて流出してしまうリスクや、ディープフェイクによる偽画像及び偽動画といった偽・誤情報を鵜呑みにしてしまい、情報操作や世論工作に使われるといったリスク、既存の情報に基づいてAIにより生成された回答を鵜呑みにする状況が続くと、既存の情報に含まれる偏見を増幅し、不公平あるいは差別的な出力が継続/拡大する（バイアスを再生成する）リスクがあること等も指摘されている。

同ガイドラインでは、このような「リスクの存在を理由として直ちにAIの開発・提供・利用を妨げるものではない」としたうえで、「リスクを認識し、リスクの許容性及び便益とのバランスを検討したうえで、積極的にAIの開発・提供・利用を行うことを通じて、競争力の強化、価値の創出、ひいてはイノベーションに繋げることが期待される」としている。

＊1 「フィルターバブル」とは、アルゴリズムがネット利用者個人の検索履歴やクリック履歴を分析し学習することで、個々のユーザーにとっては望むと望まざるとにかかわらず見たい情報が優先的に表示され、利用者の観点に合わない情報からは隔離され、自身の考え方や価値観の「バブル（泡）」の中に孤立するという情報環境を指す。「エコーチェンバー」とは、同じ意見を持つ人々が集まり、自分たちの意見を強化し合うことで、自分の意見を間違いないものと信じ込み、多様な視点に触れることができなくなってしまう現象を指す。これらへの対応については、第6章第1節2. 参照

＊2 同ガイドラインにおいては、エネルギー管理にAIを導入することで、効率的な電力利用も可能となる等、AIによる環境への貢献可能性もある点も指摘されている。

図表Ⅰ-4-1-1　生成AIの課題

	リスク	事例
従来型AIから存在するリスク	バイアスのある結果及び差別的な結果の出力	●IT企業が自社で開発したAI人材採用システムが女性を差別するという機械学習面の欠陥を持ち合わせていた
	フィルターバブル及びエコーチェンバー現象	●SNS等によるレコメンドを通じた社会の分断が生じている
	多様性の喪失	●社会全体が同じモデルを、同じ温度感で使った場合、導かれる意見及び回答がLLMによって収束してしまい、多様性が失われる可能性がある
	不適切な個人情報の取扱い	●透明性を欠く個人情報の利用及び個人情報の政治利用も問題視されている
	生命、身体、財産の侵害	●AIが不適切な判断を下すことで、自動運転車が事故を引き起こし、生命や財産に深刻な損害を与える可能性がある ●トリアージにおいては、AIが順位を決定する際に倫理的なバイアスを持つことで、公平性の喪失等が生じる可能性がある
	データ汚染攻撃	●AIの学習実施時及びサービス運用時には学習データへの不正データ混入、サービス運用時ではアプリケーション自体を狙ったサイバー攻撃等のリスクが存在する
	ブラックボックス化、判断に関する説明の要求	●AIの判断のブラックボックス化に起因する問題も生じている ●AIの判断に関する透明性を求める動きも上がっている
	エネルギー使用量及び環境の負荷	●AIの利用拡大により、計算リソースの需要も拡大しており、結果として、データセンターが増大しエネルギー使用量の増加が懸念されている
生成AIで特に顕在化したリスク	悪用	●AIの詐欺目的での利用も問題視されている
	機密情報の流出	●AIの利用においては、個人情報や機密情報がプロンプトとして入力され、そのAIからの出力等を通じて流出してしまうリスクがある
	ハルシネーション	●生成AIが事実と異なることをもっともらしく回答する「ハルシネーション」に関してはAI開発者・提供者への訴訟も起きている
	偽情報、誤情報を鵜呑みにすること	●生成AIが生み出す誤情報を鵜呑みにすることがリスクとなりうる ●ディープフェイクは、各国で悪用例が相次いでいる
	著作権との関係	●知的財産権の取扱いへの議論が提起されている
	資格等との関係	●生成AIの活用を通じた業法免許や資格等の侵害リスクも考えうる
	バイアスの再生成	●生成AIは既存の情報に基づいて回答を作るため既存の情報に含まれる偏見を増幅し、不公平や差別的な出力が継続/拡大する可能性がある

（出典）「AI事業者ガイドライン（第1.0版）」別添（概要）

1 主要なLLMの概要

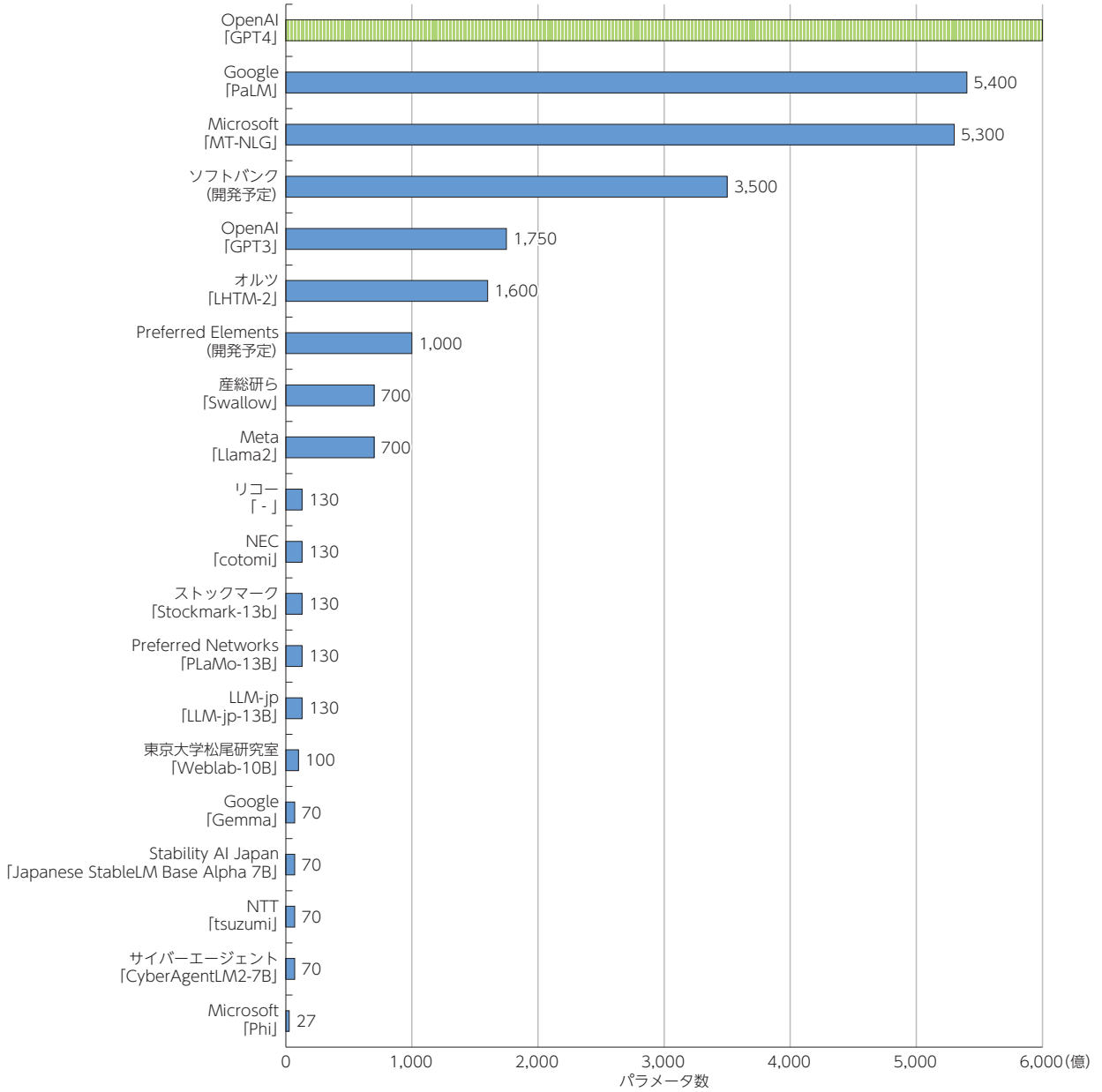
生成AIの基盤となる大規模言語モデル（LLM）の開発では、マイクロソフトやグーグルなど米国ビックテック企業などが先行している状況にある。

しかし、日本以外の企業・研究機関がクローズに研究開発を進めたLLMを活用するだけでは、LLM構築の過程がブラックボックス化してしまい、LLMを活用する際の権利侵害や情報漏えいなどの懸念を払拭できない。日本語に強いLLMの利活用のためには、構築の過程や用いるデータが明らかな、透明性の高い安心して利活用できる国産のLLM構築が必要となる[*3]。すでに日本の企業においても、独自にLLM開発に取り組んでおり、ここではその動向を紹介する。

ビッグテック企業が開発したLLMと比べると、日本では、中規模モデルのLLMが開発されている傾向が見られる（図表Ⅰ-4-1-2）。

*3　産業技術総合研究所プレスリリース「産総研の計算資源ABCIを用いて世界トップレベルの生成AIの開発を開始ー産総研・東京工業大学・LLM-jp（国立情報学研究所主宰）が協力ー」（2023年10月17日），<https://www.aist.go.jp/aist_j/news/pr20231017.html>（2024/3/22参照）

図表Ⅰ-4-1-2　各モデルのパラメータ数



（出典）企業HP、ニュース記事などの情報を基に作成*4

2 国産LLMの開発

ア　NICTによる国産LLMの開発*5

2023年7月に、国立研究開発法人情報通信研究機構（NICT）は、ノイズに相当するテキストが少ない350GBの高品質な独自の日本語Webテキストを用いて、400億パラメータの生成系の大規模言語モデルを開発した旨を発表した。発表によれば、NICTの開発したLLMについてはファインチューニングや強化学習は未実施であり、性能面ではChatGPT等と比較できるレベルではないものの、日本語でのやり取りが可能な水準に到達しているとしており、今後は、学習テキストについて、日本語を中心として更に大規模化していくこととしている。また、GPT-3と同規模の

＊4　OpenAI「GPT4」のパラメータ数は非公表。
＊5　国立研究開発法人情報通信研究機構，「日本語に特化した大規模言語モデル（生成AI）を試作～日本語のWebデータのみで学習した400億パラメータの生成系大規模言語モデルを開発～」2023年7月4日<https://www.nict.go.jp/press/2023/07/04-1.html>（2024/3/22 参照）

1,790億パラメータのモデルの事前学習に取り組み、適切な学習の設定等を探索していく予定である。さらに、より大規模な事前学習用データ、大規模な言語モデルの構築に際し、ポジティブ・ネガティブ両方の要素に関して改善を図るとともに、WISDOM X、MICSUS等既存のアプリケーションやシステムの高度化等にも取り組む予定としている（2024年5月現在、NICTではさらに開発を進め、最大3,110億パラメータのLLMを開発するなど、複数種類のLLMを開発しパラメータや学習データの違いによる性能への影響等を研究している）。

イ　サイバーエージェントが開発した日本語LLM「CyberAgentLM」*6*7

2023年5月、サイバーエージェントが最大68億パラメータの日本語LLMを開発したことを発表した。2023年11月には、より高性能な70億パラメータ、32,000トークン対応の日本語LLM「CyberAgentLM2-7B」と、チャット形式でチューニングを行った「CyberAgentLM2-7B-Chat」の種類を公開した。日本語の文章として約50,000文字相当の大容量テキストを処理可能である。商用利用が可能なApacheLicense2.0で提供されている。

ウ　日本電信電話（NTT）が開発した日本語LLM「tsuzumi」

2023年11月にNTTが開発した、軽量かつ世界トップレベルの日本語処理能力を持つLLMモデル「tsuzumi」が発表された。「tsuzumi」のパラメータサイズは6～70億と軽量であり、クラウド提供型LLMの課題である学習やチューニングに必要なコストを低減できる。「tsuzumi」は英語と日本語に対応しているほか、視覚や聴覚などのモーダルに対応し、特定の業界や企業組織に特化したチューニングが可能である。2024年3月から商用サービスが開始されており、今後はチューニング機能の充実やマルチモーダルの実装も順次展開される見込みである*8。

2 生成AIが及ぼす課題

前述のような生成AI自身が抱える制約事項のほか、生成AIの進展・普及には、それに伴う社会的・経済的な課題も多く、国内外のテック事業者、プラットフォーム事業者、業界団体や政府等による対策検討が進められている。

1 偽・誤情報の流通・拡散等の課題及び対策

「ディープフェイク」とは、「ディープラーニング（深層学習）」と「フェイク（偽物）」を組み合わせた造語で、本物又は真実であるかのように誤って表示し、人々が発言又は行動していない言動を行っているかのような描写をすることを特徴とする、AI技術を用いて合成された音声、画像あるいは動画コンテンツのことをいう。近年、世界各国でこれらディープフェイクによる情報操作や犯罪利用が増加しており、その対策には各方面からの取組が行われているものの、いたちごっこの様相を呈している。

*6　サイバーエージェント，「サイバーエージェント、最大68億パラメータの日本語LLM（大規模言語モデル）を一般公開―オープンなデータで学習した商用利用可能なモデルを提供―」2023年5月17日，<https://www.cyberagent.co.jp/news/detail/id=28817>（2024/3/22参照）
*7　サイバーエージェント，「独自の日本語LLM（大規模言語モデル）のバージョン2を一般公開―32,000トークン対応の商用利用可能なチャットモデルを提供―」2023年11月2日，<https://www.cyberagent.co.jp/news/detail/id=29479>（2024/3/22参照）
*8　NTT，「NTT独自の大規模言語モデル「tsuzumi」を用いた商用サービスを2024年3月提供開始」2023年11月1日，<https://group.ntt/jp/newsrelease/2023/11/01/231101a.html>（2024/3/22参照）

ア　ディープフェイクによる課題

（ア）AIにより生成された偽・誤情報の流通・拡散

生成AIの進歩により、非常に高品質なテキスト、画像、音声、動画を生成することが可能になり、リアルで信憑性の高い偽・誤情報を作成することが可能になった。ディープフェイク技術を用いれば、実在する人物が実際には言っていないことを本当に話しているかのような動画を簡単に作成することができる。我が国でも、生成AIを利用して作られた岸田総理大臣の偽動画がSNS上で拡散した事例が発生した[*9]。2024年1月1日に発生した能登半島地震の際にも、東日本大震災の時の津波映像や静岡県熱海市で2021年に起きた大規模土石流の映像などをあたかも能登半島地震と結びつけた投稿がSNS上で多数投稿され、大量に閲覧・拡散された[*10]。2020年には、新型コロナウイルス感染症と5G電波との関係を謳う偽情報が携帯電話基地局の破壊活動を招く[*11]など社会的影響も生じさせている。

SNSなど様々なデジタルサービスが普及し、あらゆる主体が情報の発信者となり、インターネット上では膨大な情報やデータが流通するようになったが、このような情報過多の社会においては、供給される情報量に比して、我々が支払えるアテンションないし消費時間が希少となるため、それらが経済的価値を持って市場で流通するようになる。このことはアテンション・エコノミーと呼ばれ、プラットフォーム事業者が、受信者のアテンションを得やすい刺激的な情報を優先表示するようになるなど、経済的インセンティブ（広告収入）により偽・誤情報が発信・拡散されたり、インターネット上での炎上を助長させたりする構造となっている。

偽・誤情報の拡散は世界的に問題となっており、2024年1月、世界経済フォーラムは、社会や政治の分断を拡大させるおそれがあるとして、今後2年間で予想される最も深刻なリスクとして「偽情報」を挙げた[*12]。特に2024年は、米国をはじめ、バングラデシュ、インドネシア、パキスタン、インド等、50か国余りで国政選挙が予定されている。既にインドネシア大統領選の際のディープフェイク動画の流布や、米大統領選の予備選の前に偽の音声でバイデン米大統領になりすます悪質な電話等、生成AIを利用したディープフェイクによる情報操作の事例が確認されている（**図表Ⅰ-4-1-3**）。

＊9　首相にそっくりの声で卑わいな発言をさせた動画で、民放のニュース専門チャンネルのロゴが表示され、岸田首相の話が緊急速報として生中継されているかのような印象を与えるものだった。読売新聞オンライン「生成AIで岸田首相の偽動画、SNSで拡散…ロゴを悪用された日テレ「到底許すことはできない」」2023年11月4日，<https://www.yomiuri.co.jp/national/20231103-OYT1T50260/>

＊10　日本経済新聞オンライン版「能登半島地震の偽映像、SNSで拡散　送金募集も」2024年1月2日，<https://www.nikkei.com/article/DGXZQOCA020JZ0S4A100C2000000/>（2024/3/22参照）

＊11　日本経済新聞オンライン版「欧州5G基地局破壊、影の犯人は「コロナ拡散」のデマ」2020年4月25日，<https://www.nikkei.com/article/DGXMZO58443970U0A420C2XR1000/>

＊12　世界経済フォーラム「混乱、偽情報、分裂の時代を乗り切るために」2024年1月15日，<https://jp.weforum.org/agenda/2024/01/no-wo-ri-rutameni-fo-ramu-sa-dhia-zahidhi/>
NHK NEWS WEB「“「偽情報」が最も深刻なリスクに”「ダボス会議」前に報告書」2024年1月11日，<https://www3.nhk.or.jp/news/html/20240111/k10014317071000.html>（2024/3/22参照）

図表Ⅰ-4-1-3　生成AIを利用したディープフェイクによる情報操作の事例

年月	国	内容
2021年2月	日本	●宮城県と福島県で震度6強の地震が発生した際に、記者会見を行った当時の加藤勝信官房長官の顔画像が、笑みを浮かべているように改竄された偽画像が出回った。
2022年3月	ウクライナ	●ロシアのウクライナ侵攻後、ゼレンスキー大統領がウクライナ軍に降伏を呼びかける偽動画がソーシャルメディア（SNS）上で拡散した。
2022年9月	日本	●大型の台風15号が上陸した際に、静岡県で多くの住宅が水没したとする偽画像がTwitter（現X）で拡散した。
2023年3月	米国	●画像生成AIを利用して、トランプ前大統領が逮捕されたという偽画像が生成され、Twitter（現X）で拡散された。
2023年5月	米国	●国防総省の近くで爆発が起きたとする偽画像がソーシャルメディア（SNS）上で拡散し、ダウ平均株価が一時100ドル以上も下落した。
2023年11月	日本	●岸田文雄首相が性的な発言をしたように見せかける偽動画がソーシャルメディア（SNS）上で拡散した。
2023年11月	アルゼンチン	●アルゼンチン大統領選で、AIを使ったとされる偽動画がソーシャルメディア（SNS）上で出回った。
2024年1月	台湾	●台湾総統選の際に、蔡英文総統の私生活について虚偽の主張をしている偽動画が作成・投稿された。
2024年1月	米国	●ニューハンプシャー州で、大統領選挙の予備選が控えている週末に、バイデン大統領の声を模したなりすまし電話が、予備選への投票を控えるように呼びかけた。

（出典）BBC News Japan（2024）[*13]等を基に作成

（イ）その他犯罪利用

生成AIが、情報操作のみならず、犯罪に利用されるケースも増えている。米国OpenAIのチャットボット（自動会話プログラム）であるChatGPTに用いられているものと同じAIが悪用され、「悪いGPT（BadGPT）」や「詐欺GPT（FraudGPT）」と呼ばれる不正チャットボットによってフィッシング詐欺メールが量産されている。このようなハッキングツールは、OpenAIがChatGPTを公開した2022年11月の数か月後には闇サイト上で確認されるようになり、ChatGPT公開後の12か月間で、フィッシング詐欺メールは1,265%増加し、一日平均約3万1,000件のフィッシング攻撃が発生しているという試算もある[*14]。

ディープフェイクを利用した犯罪には、AIの画像生成能力を悪用した恐喝行為もある。SNS等で共有された一般的な写真画像をAIで不適切な内容に変換し、被害者を脅迫するというもので、米国連邦捜査局（FBI）は、被害者には未成年の子供も含まれると警告している[*15]。

イ　ディープフェイクによる情報操作や犯罪利用への対策

（ア）欧州連合（EU）

偽・誤情報に関する法規制で先行するのは欧州連合（以下「EU」という。）である。2022年11月に発効[*16]した「デジタルサービス法（The Digital Services Act）」[*17]（以下「DSA」という。）は、超大規模オンラインプラットフォーム（VLOP[*18]）などに対して、自身の提供するサービスのリスク評価（偽情報に関するものを含む）やリスク軽減措置の実施を義務付けており、違反企業には最大で世界年間売上高の6%の制裁金が科されることとなっている。実際に、EUの執行機関である欧州委員会（以下「EC」という。）は、イスラエルに対するハマス等によるテロ攻撃に関わる違法コンテンツの拡散等を踏まえ、X（旧Twitter）がDSAを遵守していない可能性があるとして、違法コンテンツの拡散への対応のほか、プラットフォーム上の情報操作への対抗措置の有効性

＊13 BBC NEWS Japan,「【米大統領選2024】バイデン氏に似せた自動音声通話が予備選を妨害、米ニューハンプシャー州」, 2024年1月23日<https://www.bbc.com/japanese/68065455>（2024/2/28参照）
＊14「【焦点】生成AI「悪いGPT」の時代へようこそ」,『ダウ・ジョーンズ米国企業ニュース』2024年3月1日号
＊15 Federal Bureau of Investigation, “Malicious Actors Manipulating Photos and Videos to Create Explicit Content and Sextortion Schemes”, <https://www.ic3.gov/Media/Y2023/PSA230605>（2024/2/28参照）
＊16 同法は2023年8月からVLOP等に対して適用が開始され、2024年2月から全ての規制対象事業者に対して適用が開始されている。
＊17 European Commission, “The Digital Services Act package”, <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>（2024/2/28参照）
＊18 Very large online platformの略。オンラインプラットフォームサービスのうち、EU域内での利用者が4,500万人（EU域内人口の10%）以上のサービスを指す。

等の領域について、2023年12月に正式な調査を開始した[*19]。プラットフォーム上の情報操作への対抗措置に関し、ECは、特に、投稿に第三者が匿名で注釈を加える「コミュニティ・ノート」という機能等の有効性に焦点を当てる方針であるとしている。2024年3月、欧州議会は、AIに関する世界初の包括的な法的枠組みと位置づける「AI法（AI Act）」[*20]の最終案を可決し、同年5月にEU理事会にて正式承認され、同法が成立した。同法は一部ディープフェイクに関する規制も含み、2026年頃には本格的に適用される見込みである。

（イ）英国

英国では、2023年10月に発効された「オンライン安全法（Online Safety Act 2023）」[*21]に、虚偽であると知っている情報を受信者に心理的または身体的危害を与えることを意図してインターネット上で送信した者に、6か月の禁錮刑を科す内容が含まれている。特に、相手に苦痛、不安や屈辱等を与える加害意図や、自分が性的満足を得ようとする意図があったと立証されれば、最高刑が懲役2年となる。

（ウ）米国

米国においては、2023年7月、バイデン政権が、AI開発を主導するGoogle、Meta PlatformsやOpenAI等の7社[*22]から、AIの安全性や透明性向上に取り組む自主的なコミットメントを得たと発表した[*23]。同年9月には、新たにIBM、Adobe、NVIDIA等8社[*24]が合意し[*25]、同15社はディープフェイク対策として、真贋を示す目印をデータに忍ばせて識別を可能にする「電子透かし」等、AIによる生成を識別するための技術開発を推進している[*26]。また、米国の一部の州において、ポルノや選挙活動等の特定の目的下でのディープフェイクに関する規制が見られる。例えば、カリフォルニア、テキサス、イリノイ、ニューヨーク等9州では、相手の同意の無いディープフェイクを用いたポルノ画像や動画の配布を刑事犯罪として規定しているほか、テキサス州やカリフォルニア州では、公職の候補者に対するディープフェイク等の発信に係る規制法を設けている。なお、米国連邦法においては、国防総省や全米科学財団等の連邦機関に対し、ディープフェイクを含む偽情報に関する調査研究の強化等を求める法律が制定されている[*27]。他方、民間事業者に対しては、1996年成立の「通信品位法（Communications Decency Act）」第230条（通称Section 230）において、プロバイダは第三者が発信する情報に原則として責任を負わず、有害な内容の削除に責任を問われないと規定されているが、バイデン政権では、偽・誤情報に関してプラットフォーム事業者に一定の責任を求めるよう、法改正しようとする方向で議論が行われている。

＊19 European Commission, “PRESS RELEASE18 December, Commission opens formal proceedings against X under the Digital Services Act”, <https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709>（2024/2/28参照）
＊20 European Commission, “AI Act”, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>（2024/3/2参照）
＊21 Legislation.gov.uk, “Online Safety Act 2023”, <https://www.legislation.gov.uk/ukpga/2023/50/enacted>（2024/3/2参照）
＊22 Amazon、Anthropic、Google、Inflection、Meta Platforms、Microsoft、OpenAI
＊23 The White House, “FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI”, <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/>（2024/3/8参照）
＊24 Adobe、Cohere、IBM、NVIDIA、Palantir、Salesforce、Scale AI、Stability
＊25 The White House, “FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI”, <https://www.whitehouse.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>（2024/3/8参照）
＊26「米・AI動画識別の仕組み開発で各社合意　バイデン大統領が発表　「対策を進める」」,『NHKニュース』2023年7月22日号
＊27 2020年12月成立、2021年度の国防予算に関する「2021会計年度国防授権法」、「敵対的生成ネットワークの出力の識別に関する法律（IOGAN法：Identifying Outputs of Generative Adversarial Networks Act）」。

（エ）日本

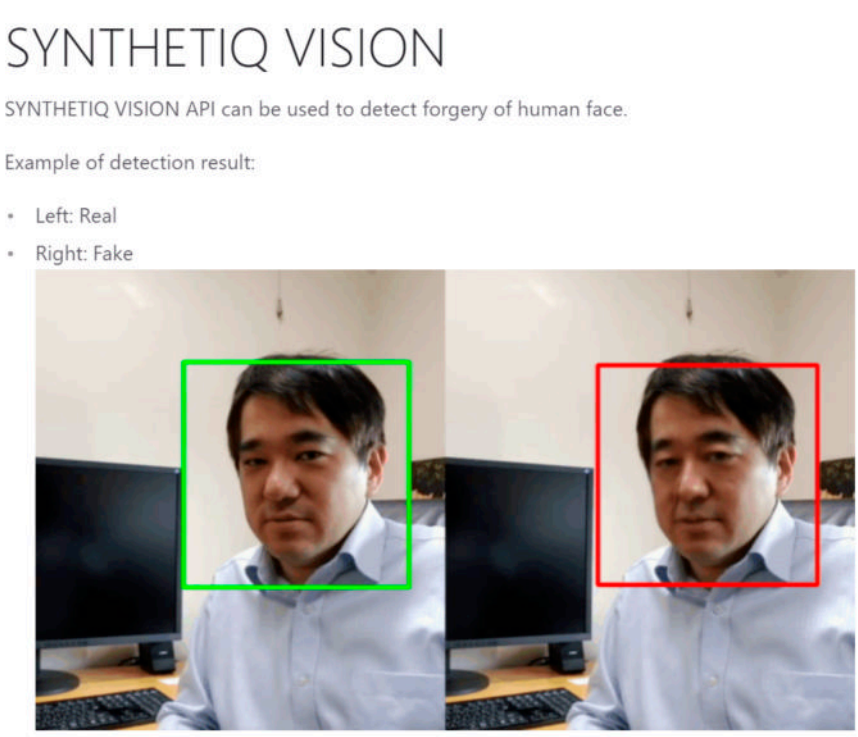
我が国におけるデジタル空間の情報流通の健全性確保に向けては、総務省が2023年11月から「デジタル空間における情報流通の健全性確保の在り方に関する検討会」を開催しており、2024年（令和6年）夏頃までに一定のとりまとめを公表予定である*28。

技術的な対策としては、インターネット上のニュース記事や広告などの情報コンテンツに、発信者情報を紐付けるオリジネータープロファイル（OP、Originator Profile）技術の研究開発が進んでいる。この技術により、なりすましや改変が見える化されることで、Web利用者が透明性の高いコンテンツを閲覧できるようになる、フェイクニュースや安易な関心獲得による広告収益が得られにくくなり、適正なWebメディアやコンテンツの配信者の権利利益侵害を低減できるようになる、広告枠が設置されるWebコンテンツの発信者が明確になることで、広告主が安心して広告出稿ができるようになるといった効果が期待される*29。

また、国立情報学研究所（以下「NII」という。）がフェイク技術対策に関する研究に早期から取り組んでおり、2021年9月には、AIにより生成されたフェイク顔画像を自動判定するツール「SYNTHETIQ VISION：Synthetic video detector」を開発した（**図表Ⅰ-4-1-4**）。これは真贋判定をしたい画像をサーバーにアップロードすると、同ツールがフェイクかどうかを判定するものである。現在NIIでは、更に進んだディープフェイク対策技術「Cyber Vaccine（サイバーワクチン）」を開発中であり、これが実現すると、真贋判定だけでなく、どこが改竄されたのか等の情報も得ることができるようになると期待されている*30*31。

＊28 総務省，「デジタル空間における情報流通の健全性確保の在り方に関する検討会」，<https://www.soumu.go.jp/main_sosiki/kenkyu/digital_space/index.html>
＊29 https://originator-profile.org/ja-JP/
＊30「Breakthrough　特集1－無人防衛2－〔第4部：ディープフェイク対策〕－ディープフェイクを見抜くツール　改ざんを自動修復するワクチンも」，『日経エレクトロニクス』2024年1月20日号
＊31 ただし、これらの対策には、真贋判定ツールの精度という課題もある。OpenAIによると、同社が自主開発した判定ツールが生成AI（主にChatGPT）製の文書を正しくAIによるものと判定する確率は26%で、逆に人間が書いた文書を誤って生成AIによると判定してしまう「偽陽性」の確率も9%あったという。そのため、この程度の精度では実際には有効な判定ツールとはならず、同社は当該ツールの提供を中止している。今後テキストや画像、音声等の生成AIと、それらの判定ツールが互いに競い合う形で双方の技術改良が進んでいく可能性が高いため、そのような技術を使っても、フェイク情報を正確に判別するのは難しいと見られている。

図表Ⅰ-4-1-4　SYNTHETIQ VISION



（出典）国立情報学研究所 シンセティックメディア国際研究センター[*32]

2 著作権を含む知的財産権等に関する議論

生成AIの生成物は、主に、文章、画像、音楽・音声の3種類である。これらは、大量のデータからその特徴を学習し、プロンプト（入力）に応じて適切な結果を出力する「機械学習」の手法を用いて開発されている。この際、データを収集・複製し、学習用データセットを作成したり、データセットを学習に利用して、AI（学習済みモデル）を開発することがオリジナルデータの制作者等の権利を侵害しないかという開発・学習段階の論点がある。また、生成AIを利用して画像等を生成したり、生成した画像等をアップロードして公表、生成した画像等の複製物（イラスト集など）を販売する際に、既存の画像等の作品と類似したものを使ってしまう等の場合に、既存作品の制作者の権利の侵害等になることがある（生成・利用段階の論点）。

ア　生成AIの進展・普及に伴う著作権を含む知的財産権等に関わる問題提起

生成AIに関連する著作権や肖像権の侵害問題は国際的に注目されており、多くの訴訟が発生している。米国では、2022年11月、GitHub Copilotの開発に関連して、学習に使用しているオープンソースコードがプログラマーの著作権を侵害している可能性があるとして、Microsoft、GitHub、OpenAIに対する集団訴訟が提訴された[*33]ほか、2023年7月には、米国の作家3名がOpenAIとMeta Platformsの2社を提訴した訴訟も発生した。同集団訴訟は、ChatGPTの機械学習に作家の著作物が無断で使用されたことによる損害賠償を請求するもので、同訴訟の結果、

＊32 https://www.synthetiq.org/
＊33 3社はGitHub Copilotがオープンソースのコードから得られた知識を使用しており、著作権侵害は行っていないと主張し、裁判所に対して訴訟の棄却を求めている。Reuters, “OpenAI, Microsoft want court to toss lawsuit accusing them of abusing open-source code”, <https://www.reuters.com/legal/litigation/openai-microsoft-want-court-toss-lawsuit-accusing-them-abusing-open-source-code-2023-01-27/>（2024/2/27参照）

OpenAIは学習データから著作物を削除するのではなく、著作権侵害で訴えられた場合の訴訟費用を負担することを表明することとなった*34。

新聞社、通信社等のメディアでのAIの活用は慎重なものとなっている。米国のAssociated Press（AP通信）は2023年7月にOpenAIとの提携を発表し、生成AIをニュース報道に生かす方法等について共同で研究する契約を結んだが、8月にはAIを配信可能なコンテンツ作成のために使用しないとした。一方、New York TimesはAIによる記事の無断使用でOpenAIとMicrosoftを訴え、これが報道機関による初の訴訟提起となった*35。日本国内においても、新聞・通信各社は、生成AIによる報道記事の無断使用について、生成AIによる記事の無断使用は許容できず、根本的な法改正に向けた検討を求める意見を表明している。

日本では、生成AI技術の発展と急速な普及に伴って権利者やAI開発者から著作権などの知的財産権の侵害に関する懸念の声が上がったことを踏まえ、2024年3月、文化審議会著作権分科会法制度小委員会において、「AIと著作権に関する考え方について」がとりまとめられるとともに*36、（著作権を含む）知的財産権との関係について、2024年5月、AI時代の知的財産権検討会より、「AI時代の知的財産権検討会　中間とりまとめ」が公表された*37。

イ　著作権を含む知的財産権等の侵害リスクに対する取組

生成AIの利用に際しての著作権等の権利侵害対策に向けては、データ・コンテンツの権利保持者とAI事業者双方が、互いの契約の中で対応を行うこと等が考えられる。技術的には、生成AI生成物であることの表示を可能とする電子透かしの実用化や、OpenAIによる知的財産権を侵害する恐れのあるデータ・コンテンツのAI入出力を抑制する仕様の提供等がある一方で、New York Times、CNN、Bloomberg、Reuters、日本経済新聞等の国内外のメディア側も、OpenAI等AI事業者のGPTボットのブロックを行う等の対策で自衛している*38。

技術を活用しながら著作権侵害の法的リスクに対してコミットする取組もある。Microsoftは、大規模言語モデル（Large Language Model：LLM）を組み込んだ自社の生産性向上ツール「Microsoft Copilot」に対する法的リスクに対して責任を負う、「Copilot Copyright Commitment」を2023年9月に発表している。Microsoft Copilotで生成した出力結果を使用して、著作権上の異議を申し立てられた場合、Microsoftが責任をとる仕組みとなっている*39。著作物を使用しない、あるいは許諾済みの著作物を活用する方法で著作権等侵害のリスクを回避する方法もある。例えば、Adobeが提供する「Adobe Firefly」は、オープンライセンス等、著作権の問題の無い画像を学習段階で利用しており、著作権侵害の心配なく生成した画像の商用利用が可能としている。

＊34 生成AI活用普及協会，「AIの著作権はどうなる？生成AIで生成した画像やイラストの著作権や適法性、注意したいポイントを徹底解説」2023年12月28日，<https://guga.or.jp/columns/ai-copyright/>（2024/3/2参照）
＊35 Reuters, “OpenAI, Microsoft want court to toss lawsuit accusing them of abusing open-source code”, <https://www.reuters.com/legal/litigation/openai-microsoft-want-court-toss-lawsuit-accusing-them-abusing-open-source-code-2023-01-27/>（2024/2/27参照）
＊36 文化審議会著作権分科会法制度小委員会「AIと著作権に関する考え方について」（令和6年3月15日），<https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94037901_01.pdf
＊37 AI時代の知的財産権検討会「AI時代の知的財産権検討会 中間とりまとめ」（2024年5月），<https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2024/0528_ai.pdf>
＊38 AI時代の知的財産権検討会「AI時代の知的財産権検討会 中間とりまとめ」（2024年5月），<https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2024/0528_ai.pdf>
＊39 AIキャラクターに著作権はある？ 違反したらどうなる？ 弁護士に聞いた，<https://webtan.impress.co.jp/e/2023/12/19/46093>（2024/3/2参照）