

特集

# 弁護士会は生成AIの 技術革新・リスクとどう向き合うべきか

# 第1章 はじめに

生成AI分野の技術進歩は目覚ましく、官公庁・民間企業を問わず社会のあらゆる領域に亘り生成AIの利活用が促進される一方、幻覚（ハルシネーション; Hallucination）、知的財産権の侵害、個人情報・機密情報の漏洩、学習データに起因するバイアス、差別偏見の助長のおそれなど、生成AIを取り巻く様々な問題点への対応も急務とされる。

我が国では、AI技術に関するイノベーションの促進とリスク対応の両立を目指し、2025（令和7）年5月28日、人工知能（Artificial Intelligence; AI）分野に特化した初の法律「人工知能関連技術の研究開発及び活用の推進に関する法律」（以下、「AI活用推進法」という。）が参議院本会議で可決・成立した。

本稿では、弁護士会としても、技術革新の促進を阻害しないようリスク対応の在り方を早急に議論することが望ましい旨を提言するものである。

生成AIが急速に社会に遍く普及しつつある現状において、弁護士が生成AIの仕組みやリスクを理解していない、あるいは生成AIを使ったことすらないとなれば、生成AIの問題点にかかわる法律相談や事件

処理において適切な対応ができず、弁護士法2条、職務基本規程7条等により求められる弁護士の職責に悖るとの批判を免れない。また、生成AIに関する理解不足は弁護過誤の原因にすらなり得る。

他方で、法律事務所所属の弁護士とインハウスロイヤーとを問わず、生成AIを適切に利活用することができれば、業務のクオリティ向上・業務の効率の向上が期待できるし、何より使ってみることでその特性や問題点の理解も促進されると思われる。そのような観点からは、生成AIの有用性やリスクを弁えつつも、日常生活や業務の中で積極的に使ってみようとする姿勢が望ましい。

本稿では、2019（平成31）年3月29日統合イノベーション戦略推進会議「人間中心のAI社会原則」を参考に、弁護士が生成AIを適正かつ有効に利活用するための基本的な考え方を示した上、生成AIについて最低限の知識を習得し、その有用性とリスクを踏まえつつ、利活用する方法を積極的に模索すべき旨を提言するものである。

## 第2章 生成AIの基本的な仕組みの理解は重要である

### 第1 生成AIとは何か？

生成AI（Generative AI）とは、新しいデータ、コンテンツ、プログラム等を生成する汎用的な能力を備えた人工知能（Artificial Intelligence；AI）モデルを基盤とするシステムや当該システムをアプリケーション等に組み込んだサービス全般を指す。生成AIは、プロンプトといわれる指示文を入力すると、テキスト、画像、音楽、動画など、様々な形式の（マルチモーダル；multi-modal）新しいコンテンツを作り出すことができるので、自然言語生成だけでなく、画像生成、

音楽作曲、映像生成などの広範なコンテンツ生成に応用される。

AIには、人間があらかじめ定めたルール（if-then）に従って動作するルールベースのAIと、大量のデータからデータ間の関係についてのパターンやルールを自動的に学習して新たなデータに対して予測を行う機械学習・ニューラルネットワークベースのAIがあるが、本稿で述べる生成AIは、後者のタイプに属する。

### 第2 生成AIの仕組み・特性の理解の重要性

生成AIが急速に社会に遍く普及しつつある現状において、弁護士業務を適切に遂行する上では、弁護士がごく基本的な生成AIの知識を習得しておくことが望ましいという観点に基づき、第3及び第4においては、生成AIに関する基本的な概念整理や説明を行うものである。

汎用的なツールである生成AIの技術進歩は著しく、その有用性は日進月歩の進化を遂げつつあるが、他方で後述の第3章のとおり、ハルシネーションのような原理的な問題や、情報漏洩の危険性といった問題も抱えており、これらの問題を完全に解決することは困難である。

上記のような解決困難な課題を抱えていることを知らずに生成AIを利用することは、弁護士の生成AIに対する過度な信頼・依存を招きかねない。他方で、課題があるから一切使用しないという姿勢では、弁護士が社会一般から取り残されることになりかねない。

生成AIの仕組み・特性をその概要だけでも知っておけば、リスクを回避しつつ生成AIの効用を十分に享受できる。また、一般社会での生成AI利活用がますます普及し社会にあまねく浸透しつつある今日状況では、生成AIの仕組み・特性の概要は現代人の必

須の教養であるといってもよい。後述する「人間中心のAI社会原則」も、「教育・リテラシーの原則」の中で、「AIの利用者側は、AIが従来のツールよりはるかに複雑な動きをするため、その概要を理解し、正しく利用できる素養を身につけていることが望まれる。」と表明している。

弁護士は、基本的人権を擁護し、社会正義を実現するという使命（弁護士法1条1項）に基づき、法律事件という形で社会における森羅万象を業務上取り扱う存在であるといっても過言ではないが、そのために常に、深い教養の保持と高い品性の陶冶に努めなければならないとされている（弁護士法2条、弁護士職務基本規程7条）。生成AIが急速に社会で普及しつつある現在に至っては、生成AIの仕組みや特性を習得することは、もはや弁護士に課せられた責務といえよう。また、生成AIの技術進歩の速度は著しく、数か月単位という短い期間で状況が刻々と変化し得ることに鑑み、適宜の知識のアップデートも必要となろう。

ただし、自動車を運転するために自動車の動作に関する技術的事項すべてに精通する必要がないのと同様、ごく基本的な仕組みと特性を理解していれば、生成AIの技術的原理のすべてに精通する必要はない。

## 第3 ChatGPTなどの対話型文書生成AI

### 1 対話型文書生成AIの有用性

ChatGPTなどの対話型文書生成AIは、人間の目から見ても自然で流暢な言葉や表現で回答を生成してくれるだけでなく、あたかも会話の文脈を理解しているかのように、過去の応答や会話の履歴を踏まえた適切な回答を生成し、長い対話や複雑な問い合わせにも柔軟に対応してくれる。しかも、質問応答、文書要約・翻訳、文書生成などの様々な自然言語処理のタスクに対応できる汎用的な能力を持ち、自然科学、人文科学、社会科学から日常会話まで、多岐に亘るトピックや質問に対応できる。書籍のタイトルなどのアイデアの提案、ディベート、ブレインストーミングといったタスクにも対応できる。物語の筋書きやショートストーリーの生成も可能である。

最新の主要な生成AIサービスでは、ウェブ検索機能（生成AIの知識を拡張し、回答の正確性を強化する）、ファイルアップロード機能（長いプロンプトを入力する手間を省略できる）、Deep Research機能（インターネット上の情報に基づき詳細なリサーチをしてくれる）などの便利な機能が備わっている。

### 2 一般的に便利とされている活用法

#### (1) 文書の生成・添削・校正

メールの文案作成、マーケティング文の作成、日報・議事録の作成、講演やスピーチのアウトライン作成など。

文法・スペルミスだけでなく、内容の改善点も指摘してくれる。

#### (2) 文書や概念の要約・翻訳

長い文章を簡潔で要領を得た短い文章にまとめてくれる。

行数や文字数を指定して要約することもできる。

#### (3) 壁打ち・ブレインストーミング

人間が生成AIと対話しながら思考を深めることが可能である。

生成AIが人や物になりきって会話するなどのロールプレイも可能である。

#### (4) リサーチ・論点の洗い出し

ディベートができる。複数の利害関係人の要望を調整することで妥結点・折衝点を提案するなど、交渉の

シミュレーションにも利用できる。

#### (5) アイデアの提案

見出し・キャッチコピー、イベントのプランニング、プレゼン資料（タイトル、目次案、ページ配分、トークスクリプト）などを作成してくれる。例題、ダミーデータ、乱数の生成といったこともしてくれる。

### 3 対話型文書生成AIと大規模言語モデル<sup>\*1</sup>

ChatGPTのような対話型文書生成AIは、「入力された単語（及びこれまでに出力された単語）の次に来る確率の高い単語を次々と出力する」（next word predictionないしnext token prediction）という、比較的単純な原理に基づく言語モデル（「Transformer型言語モデル」といわれる。）を基盤モデルとするものである。<sup>\*2</sup>

大規模言語モデルは、大量のテキストデータを用いて訓練されたニューラルネットワークによる言語モデルである。入力されたコンテキスト（単語列）から、次の単語の確率分布（次に続く単語として適切であると予測された候補と、その生成確率の組）を予測する。候補となる単語のうち、より自然で適切と予測された単語に、高い生成確率が割り当てられるように訓練される。巨大な言語データを用いた大規模言語モデルは、様々な文章中の文脈に応じた単語の意味、単語間相互の関係性（意味の関連性・類似性）、言語的な規則・構造、文章の文脈における単語の使われ方など、人間が過去に言語を用いたありとあらゆるパターンを認識するよう訓練されており、人間の言葉の構文や意味・文脈を高い精度で捉えていることが分かっている。大規模言語モデルは、テキスト生成、翻訳、質問応答、チャットボットのような対話システムなど、自然言語処理タスクに応用される。

<sup>\*1</sup> 岡野原大輔『岩波科学ライブラリー319 大規模言語モデルは新たな知能か——ChatGPTが変えた世界』岩波書店・2023年。岡崎直観＝荒瀬由紀＝鈴木潤＝鶴岡慶雅＝宮尾祐介共著『IT Text 自然言語処理の基礎』オーム社・2022年。

<sup>\*2</sup> ChatGPTは、オープンソースのモデルではなく、いわゆるプロプライエタリな（技術的詳細が公開されていない）モデルであることには留意すべきである。とはいえ、ChatGPTも、基本的には次の単語を予測するモデルの範疇にあると考えてよいと思われる。

他方で、対話型文書生成AIには、後記第3章のとおり、架空の情報を捏造してしまう「幻覚」（ハルシネーション; Hallucination）という現象や、個人情報・機密情報の漏洩リスク、著作権などの知的財産権侵害のおそれ、有害コンテンツの生成など悪用のおそれ、偏見・バイアスを含む応答のおそれなどの問題点も指摘されている。

#### 4 大規模推論モデル<sup>\*3</sup>

なお、最近では、単なるパターン認識と次単語予測の枠組みを超え、初見の問題を解くための論理を構築できる大規模推論モデル（Large Reasoning Model;

LRM）をベースとした対話型文書生成AI（「深く思考するAI」）の研究開発が盛んにおこなわれている。その代表例としては、OpenAI のGPT-5（従来のChatGPT o-series（o1、o3-proなど）がさらに進化を遂げたモデル）、Anthropic のClaude Opus/Sonnet4、GoogleのGemini 2.5 Pro、Gemini Flash Thinkingなどが挙げられる。LRMは、複雑な思考ステップを要する問題に直面したときに、思考経路を探索・評価しつつ回答を生成するよう、強化学習の技術を用いて訓練されており、法律問題への推論能力を備えたモデルの研究開発が大いに期待される場所である。

## 第4 DALL-Eなどの画像生成AI

### 1 拡散モデル<sup>\*4</sup>

画像生成AIとして著名なOpenAIのDALL-Eシリーズは、拡散モデルをベースにしていると言われている。「拡散モデル」とは、画像に徐々に一定の確率法則に従ったノイズを加えてノイズ画像（砂嵐のような画像）とする過程（順過程）を逆再生し、ノイズ画像から元の画像を確率的に予測して復元する（逆過程）ように訓練された画像生成モデルである。順過程でノイズを加えていく各ステップを、逆過程が正しく巻き戻せるように学習する。拡散モデルは、画像だけでなく、音声・動画の生成モデルとしても利用される。

### 2 画像生成AIの有用性と問題点

DALL-Eなどの画像生成AIは、自然言語のプロンプトを入力すれば、あたかもプロのクリエイターが作成

したかのようなクオリティの高い画像コンテンツや動画コンテンツが比較的短時間で生成できる。

他方で、画像生成AIを用いれば、素人でも、プロンプト指示のみによって高品質の画像・動画コンテンツが容易かつ短時間で生成できることに伴う弊害も指摘されている。すなわち、後記第3章のとおり、著作権を侵害するコンテンツが大量に生成されるおそれ、人間のクリエイターから仕事を奪い、人間の尊厳や創作のインセンティブを奪ってしまうおそれのほか、有害コンテンツ生成への悪用、ディープフェイクの蔓延のおそれ等の問題点も指摘されている。

\*3 Fengli Xu, et al. Toward Large Reasoning Models: A Survey of Reinforced Reasoning with Large Language Models. arXiv:2501.09686v3 [cs.AI] (2025).

\*4 数理科学2024年10月号No.736「特集 生成AIのしくみと数理」（鈴木大慈「拡散モデルの理論」参照）。岡野原大輔『岩波科学ライブラリー328 生成AIのしくみ——〈流れ〉が画像・音声・動画をつくる』岩波書店・2024年。

# 第3章 生成AIの限界・リスクとそれらへの対策が急務である

## 第1 生成AIの限界・リスクについて

現在の生成AI技術の急速な進展と社会実装の拡大に伴い、その活用が多様な領域に及んでいく一方で、以下のような限界・リスクへの対応が喫緊の課題となっている。今後、生成AIの利活用を促進しつつも、これらの諸問題への制度的・技術的な対応を<sup>\*5</sup>早急に整備することが要請される。

この点につき、AI戦略会議・AI制度研究会の2025（令和7）年2月4日付「中間とりまとめ」は、以下に挙げた様々な限界・リスクを含む生成AIのリスクへの対応について、「イノベーション促進とリスクへの対応の両立を確保するため、法令とガイドライン等のソフトローを適切に組み合わせ、基本的には、事業者の自主性を尊重し、法令による規制は事業者の自主的な努力による対応が期待できないものに限定して対応していくべきである。」（10頁）と述べている。

### 1 AI倫理（AI Ethics）

生成AIに限らず、AI技術の普及に伴い検討・解決しなければならない倫理的な問題が提起されている。

具体的には、バイアス・偏見（Bias；学習データ等に含まれる差別的表現等により、差別・偏見を助長する出力を行う可能性がある）、公平性の欠如（Fairness；性別・人種など特定の属性に不利な出力

を行うことで、公平性が損なわれるおそれがある）、プライバシー侵害（Privacy；個人情報の不適切な取扱いがされる可能性がある）、説明可能性の欠如（Explainability；生成過程がブラックボックスであり、なぜその出力がされたのかが説明困難な場合がある）、偽情報（Misinformation；信憑性のない情報を尤もらしく生成する危険性がある）、労働の置換（Labor Replacement；一部の職種がAIに代替され、雇用に影響が生じる可能性がある）、有害コンテンツの生成（Harmful Content Generation；差別的・暴力的・違法な情報の生成が可能である）、ディープフェイク（Deep Fakes；顔・声などをAIが模倣し、なりすまし等に悪用される危険性がある<sup>\*7</sup>）などの問題が挙げられる。

上記の諸問題を検討することは、生成AIの適正な利活用にとって必須の事項である。

### 2 幻覚（ハルシネーション；Hallucination）

対話型文書生成AIが、学習データの中に存在しない架空の情報を捏造し、あたかも真に存在するかの如くごく自然に、捏造した事実を正確な事実と織り交ぜながら回答を生成してしまう「幻覚」（Hallucination）という現象が広く知られている。生成AIが生成した虚偽情報がインターネットやSNS上で拡散されてしまうと、真偽不明の情報が大量に流布し、何か真実かの区別が極めて困難となり、混乱が生じるおそれがある。生成AIが生成した虚偽情報は自然で尤もらしく、専門家でも区別が難しい場合すらあるため、極めて致命的な問題であると指摘されている。確率モデルに必然的に伴う原理的な問題であるため、現時点の最新技術

<sup>\*5</sup> 技術的な対応について、現在、AI生成物であることを識別できる技術（電子透かし、メタデータの埋込み、コンテンツ認証情報との紐づけなど）、知的財産権を侵害するおそれのあるデータ・コンテンツのAI入出力を抑制する技術（類似画像検索機能など）、AIの学習に備えて行われる自動収集プログラムによるコンテンツ収集を拒絶する技術（「robots.txt」の記載による制限、ID・パスワードによるアクセス制限）、画像に特殊な画像処理を施すことでAIの学習を妨げる技術（学習を妨害するノイズを加えるなど）、画像生成に寄与した学習データを推定する技術などの様々な取り組みがされており、既に商用化されたサービスも一部ある。2024年5月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」（[https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2024/0528\\_ai.pdf](https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2024/0528_ai.pdf)）35～43頁。

<sup>\*6</sup> [https://www8.cao.go.jp/cstp/ai/interim\\_report.pdf](https://www8.cao.go.jp/cstp/ai/interim_report.pdf)

<sup>\*7</sup> ディープフェイクは、EU・AI法案3条（60）において、「AIで生成又は操作された画像、音声又は映像コンテンツであって、実在の人物、物体、場所その他の存在物や事象に類似させたり、ある人物が本物又は真実であるかのように偽って表示し得るもの」と定義されている（2024年5月・AI時代の知的財産権検討会編「AI時代の知的財産権検討会 中間とりまとめ」参照）。

でも、この問題は解決されていない。

例えば、弁護士が生成AIに法律問題への回答を指示した場合、生成AIの回答に含まれる適用法令や引用裁判例には誤りが含まれることが多い。<sup>\*8</sup> OpenAI社のChatGPT利用規約（日本語版、最終更新：2024（令和6）年12月11日）が、「当社の本サービスを利用する場合、お客様は次の事項を理解し、同意するものとします。」とした上で、「信用、教育、雇用、住宅、保険、法律、医療、その他の重要な決定など、個人に法的又は重大な影響を与える可能性のある目的において、その個人に関連するアウトプットを使用してはなりません。」と記載しているのも、主に生成AIの「幻覚」現象を念頭に置いていると思われる。<sup>\*9</sup>

### 3 個人情報や機密情報の漏洩のおそれ

ユーザーが無意識に個人情報や機密情報をプロンプトとして入力してしまい、生成AI経由でこれらの情報が漏洩してしまうおそれがある。生成AIにプロンプトとして入力した情報は、単に生成AIの回答に用いられるだけでなく、一連の入出力情報がシステムのクラウドサーバに記録される場合や、当該生成AIの基盤モデルの追加学習に利用される場合がある。また、入出力情報が生成AIの基盤モデルの追加学習に利用された結果、当該生成AIを第三者が利用した際に当該入出力情報が出力される可能性も全くないとはいえない。

個人情報保護委員会は、2023（令和5）年6月2日、個人情報取扱事業者や行政機関等に対し、生成AIサービスの利用に際しての個人情報の取扱いに関する注意点を示しつつ、個人情報の保護に関する法律（平成15年法律第57号）に従って個人情報を適正に取り扱っていただきたい旨の注意喚起を行った。<sup>\*10</sup>

デジタル庁・デジタル社会推進会議幹事会事務局は、2025（令和7）年2月6日、「DeepSeek 等の生成AIの業務利用に関する注意喚起（事務連絡）」<sup>\*11</sup>を公表し、政

府等による生成AIの業務利用について、①約款型サービスに関し、原則として要機密情報を取り扱うことはできない旨、機密情報を取り扱わない場合であっても、リスクを考慮した上で利用可能な業の範囲をあらかじめ特定し、個々の利用に当たっては、利用手続に従って、利用目的（業務内容）や利用者の範囲などの利用者からの申請内容を許可権限者が審査した上で利用の可否を決定し、その利用状況について管理することが必要である旨、②要機密情報を取り扱わない場合であっても、例えば、国外にサーバ装置を設置している場合は、現地の法令が適用され、現地の政府等による検閲や接収を受ける可能性があることが利用の可否を判断する際に考慮すべきリスクである旨などを示した。

弁護士は、依頼者や一定の事件関係者に対する秘密保持義務（弁護士法23条、弁護士職務基本規程23条）を負うところ、生成AIにプロンプトとして秘密保持義務の対象となる情報をそのまま入力した場合、秘密保持義務違反となる可能性がある。

### 4 著作権など知的財産権の侵害のおそれ

#### (1) 生成AIと著作権

生成AIを巡っては、知的財産権の中でも、特に著作権の侵害に関する懸念が各方面から強く表明され、生成AIと著作権との関係が活発に議論されているが、生成AIと著作権の関係に関する判例及び裁判例の蓄積は未だない現状にある。このような状況を踏まえて、研究者・弁護士・裁判官などの有識者委員で構成される文化庁・文化審議会著作権分科会法制度小委員会は、2024（令和6）年3月15日、「AIと著作権に関する考え方について」<sup>\*12</sup>を公表し、生成AIと著作権の関係につき関係者の懸念を解消すべく、著作権法の解釈に当たっての一定の考え方を示したところである。この「考え方」は、後述する「人間中心のAI社会原則」における「人間中心の原則」が示すように、「人がAIを高度な道具として補助的に用いることが原則」であるとの考えに基づき、「人が道具としてAIを使用するもの

\*8 実際に米国では、ChatGPTを利用した弁護士が、ChatGPTの回答に含まれていた架空の裁判例を準備書面に引用し、ペナルティを受けた事例が起きている。

\*9 <https://openai.com/ja-JP/policies/row-terms-of-use/>

\*10 個人情報保護委員会「生成AIサービスの利用に関する注意喚起等」([https://www.ppc.go.jp/files/pdf/230602\\_alert\\_generative\\_AI\\_service.pdf](https://www.ppc.go.jp/files/pdf/230602_alert_generative_AI_service.pdf))。

\*11 [https://www.digital.go.jp/assets/contents/node/basic\\_\\_page/field\\_\\_ref\\_\\_resources/d2a5bbd2-ae8f-450c-adaa-33979181d26a/e7bfeba7/20250206\\_\\_councils\\_\\_social-promotion-executive\\_\\_outline\\_\\_01.pdf](https://www.digital.go.jp/assets/contents/node/basic__page/field__ref__resources/d2a5bbd2-ae8f-450c-adaa-33979181d26a/e7bfeba7/20250206__councils__social-promotion-executive__outline__01.pdf)

\*12 [https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94037901\\_01.pdf](https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94037901_01.pdf)

であり、これに伴う行為の責任はAIを道具として用いる人に帰属するということを前提としている。」と表明している。

この「考え方」のみならず、一般に、生成AIと著作権との関係については、AI（学習済みモデル）作成のための学習や生成AIを用いたソフトウェア・サービスの開発を行う「開発・学習段階」と、生成AIにより生成物を出力し、その生成物を利用する「生成・利用段階」に分類した上で検討が行われている。<sup>\*13</sup>

生成AIの開発・学習に伴う複製については、著作権法30条の4、同47条の5による権利制限の可能性や各要件の解釈論などが議論されている。生成・利用に伴う複製については、主に著作権侵害の要件である類似性・依拠性の判断方法が議論されているほか、生成AIに対する生成指示として著作物を含むプロンプトを入力する行為に伴う著作物の複製等について、著作権法30条の4の適用の可否及び適用範囲も議論されている。<sup>\*14</sup>

そのほか、ジブリ風の画像のような、クリエイターの作風・画風を模倣した画像が生成できる画像生成AIの開発・学習や当該画像生成AIを用いた画像の生成・利用について、作風・画風というアイデアの模倣に過ぎず著作権侵害が否定されるか、あるいは表現の模倣として著作権侵害が肯定され得るかといった困難な問題も提起されており、未だ議論の一致をみない。<sup>\*15</sup> コンテンツを無断で学習されることに対する新聞社、作家、クリエイターの反発や不満を解消するための合理的な制度の検討については未だ不十分な点も多い。この点については、非享受目的か享受目的か、AI生成物とクリエイターの作品が市場で競合し得るか否かといった著作権法30条の4の解釈論のみならず、生成AIの開発インセンティブ確保とクリエイターの権利保護を両立し得る制度設計が必要であろう。

\*13 上野達弘＝奥邨弘司編『AIと著作権』（勁草書房・2024年）など。

\*14 前掲「考え方」では、生成物の生成のためにプロンプトを情報解析することに伴う著作物の複製等については著作権法30条の4の適用が考えられるとしつつ、「生成AIに対する入力に用いた既存の著作物と類似する生成物を生成させる目的で当該著作物を入力する行為」は、非享受目的と享受目的が併存するため、同条の適用はないとしている。

\*15 前掲・上野達弘＝奥邨弘司編『AIと著作権』221～236頁。

## (2) 生成AIとその他の知的財産権

生成AIと著作権以外の知的財産権との関係については、研究者・弁護士・企業担当者などの有識者委員で構成される特許庁・AI時代の知的財産権検討会が、2024（令和6）年5月、「AI時代の知的財産権検討会中間とりまとめ」<sup>\*16</sup>を公表したところである。

この「中間とりまとめ」は、我が国経済社会の発展につなげるために生成AIを用いた様々なAIツールの開発・提供・利用を促進する一方で、生成AIが文章、画像、動画等を取り込むことでマルチモーダル化するに至っている現状において、クリエイターの権利保護等の観点に留意しつつ、知的財産権全体を俯瞰する観点から、著作権だけにとどまらず、それ以外の知的財産権との関係についても整理する必要があるとして策定されたものである。

また、「中間とりまとめ」は、生成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つつ、必要な方策を検討するに当たっては、単に法律上の観点から整理するだけにとどまらず、技術による対応策や、創作活動の持続可能性の観点から不可欠となる契約等による対価還元についても視野を広げて検討し、それぞれを有機的に連携させながら権利保護とAI技術の進歩を後押しするという方向性を見いだしていくことも必要と考えられるとしている。

「中間とりまとめ」では、生成AIと意匠法（意匠権）との関係、生成AIと商標法（商標権）との関係、生成AIと不正競争防止法との関係のほか、技術による対応、契約による対応（対価還元）、AI技術の進展を踏まえた発明の保護の在り方等が検討の対象とされている。また、声優をはじめとする人の声を学習させ、本人類似の音声を生成できるAIが無断で開発されるといった事例を踏まえ、人の声を保護する法律の適用関係について、肖像権・パブリシティ権による保護の有無のほか、著作隣接権等の知的財産権による保護の有無が検討の対象とされている。

\*16 [https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2024/0528\\_ai.pdf](https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2024/0528_ai.pdf)

## 5 偽情報の蔓延・有害コンテンツ生成、サイバー犯罪への悪用のおそれ

### (1) 偽情報の蔓延

生成AIを悪用して、虚偽の内容を含むニュース記事（フェイクニュース）を意図的に生成し、SNSや電子メールを通じて拡散するなど、偽情報を巧妙に作出する者が跋扈し、虚偽情報・偏向情報が蔓延しかねないという危険性が指摘されている。

特に、ChatGPT GPT-4oやChatGPT O3など最新の生成AIの能力の飛躍的増大により、ますます巧妙かつ説得力を持った虚偽の情報が蔓延するおそれがある。

### (2) 有害コンテンツの生成

生成AIを用いて、特定の個人、集団または社会にとって有害な様々な種類のコンテンツを生成することが可能である。例えば、自殺など自傷行為を奨励するアドバイス、性的な内容・暴力的なコンテンツ、ハラスメントやヘイトスピーチなどを含むコンテンツ、武器・兵器の製造方法など暴力攻撃の手段を提供するコンテンツ、その他違法行為を助長・奨励し、あるいは違法行為の手段に関する情報を提供するコンテンツである。このような生成AIの能力を悪用して、有害コンテンツが生成される危険性が指摘されている。

有害コンテンツの生成に対しては、生成AIの基盤モデルの学習方法、具体的には人が好む応答に高い報酬・評価を与える「人間のフィードバックに基づいた強化学習」（Reinforcement Learning from Human Feedback; RLHF）の手法を用いて、有害なコンテンツの生成を拒否する技術を開発するなどの安全対策が鋭意進められており、違法行為をなすハードルは高まりつつあるが、これらを完全に防止することは技術上困難であると考えられている。

### (3) サイバー犯罪への悪用

悪意のあるユーザーが生成AIを悪用して、システムの脆弱性に関する説明の指南を求めてこれを生成させる危険性、あるいは、極めて巧妙かつリアルな内容のフィッシング詐欺のメールやスパムメッセージを生成し、大量のユーザーに送信する危険性など、生成AIがサイバー犯罪に悪用される危険性が指摘されている。

また、不正にファイルをダウンロードさせるプログラムコードを生成AIに生成させ、ユーザーから個人情報・機密情報を引き出すといった悪用がされる危険

性も指摘されている。

さらに、生成AIに対し、開発者が想定していない特殊かつ巧妙なプロンプト指示を行うことで、チャットボットが保有する機密情報や本来公開の対象とされていないデータを引き出すという悪用（Prompt Injection）がされる危険性も指摘されている。

生成AIを利用したサイバー攻撃に対しては、強化学習の手法等によりタスク実行を拒否する技術の開発が鋭意進められているものの、完全な防止策はなく、不正使用を継続的に監視する、迅速かつ反復的なモデル改善を継続するといった対処療法に依存せざるを得ないといわれる。

## 6 バイアス（Bias）を含む応答

対話型文書生成AIの基盤モデル（LLM）は、大量の学習データに含まれる言語の構造や使い方のパターン認識を学習しているため、生成AIの応答も必然的に認識したパターンを反映するものとなる。それゆえ、生成AIの学習データの中に、人種、性別、宗教、社会的・経済的身分その他に関する差別的なステレオタイプや偏見（Bias）を含む文章が存在する場合、生成AIは、そのような偏見を反映した応答を生成する可能性がある。例えば、生成AIは、ある特定の職業に男女構成比の偏りがある場合、当該職業に就いている人の性別を、当該職業の多数派を構成する性別であると判断してしまう傾向にあり、例えば、生成AIがソフトウェア・エンジニアを男性、小学校の先生を女性であると判断する傾向が報告されたことがある<sup>\*17</sup>。

「偏見」を含む応答を生成する生成AIが広く普及し、社会のあらゆる領域であまねく使用されるようになると、特定の差別・偏見やそれらを含む世界観が強化・再生産されるおそれがある。

弁護士が生成AIを利用するに際しても、生成AIの回答が上記のようなステレオタイプや偏見が反映されたものでないか否かを必ず確認・検討することが不可欠である。

\*17 S. Bubeck, et al. Sparks of Artificial General Intelligence: Early experiments with GPT-4, Microsoft Research. 2023, P.87.

## 第2 問題点に対する各国の対策（EU ACT、G7広島AI原則）

### 1 米国

米国では、連邦全体の包括的なAI規制は存在しない状況である。これまでの動きとしては、2023（令和5）年10月30日に、バイデン大統領が、「人口知能の安全、安心、信頼できる開発と利用に関する大統領令」（大統領令14110）<sup>\*18</sup>を発令した。同大統領令は、一定のAIシステムの開発者に対して報告義務を課するようなハードローとしての側面も有していた。なお、同大統領令が公布されるまでは連邦全体におけるAIに関するハードローは存在しなかった。

しかし、トランプ大統領は就任初日に、「有害な大統領令及び措置の初期的撤回」（大統領令14148）<sup>\*19</sup>を発令した。さらに、続く2025（令和7）年1月23日、「人工知能におけるアメリカの主導権を妨げる規制の撤廃」（大統領令14179）<sup>\*20</sup>を発令し、イノベーションの障壁となっているAI政策を撤回し、イノベーションを促進することを示した。これら大統領令により、バイデン時代の前大統領令は撤回されることとなり、規制緩和路線に再び進むこととなったと評価できる。

他方、カリフォルニア州のカリフォルニア州AI透明性法<sup>\*21</sup>やニューヨーク州のニューヨークRAISE法<sup>\*22</sup>など、州レベルでは、消費者保護や透明性確保等の観点でハードローによる規制が進められている。

このように米国では、AI規制について、連邦と州レベルで規制状況が異なっている状況である。

### 2 EU

EUにはAI法（The Artificial Intelligence Act（AI Act））が存在する。AI法は、世界に先駆けたAIに関

するハードローである。AI法は、EU内の人権の自由の確保が最優先で、技術は人間中心であるとする思想を出発点とする。AI法は、2021（令和3）年4月21日に公表され、欧州議会が2024（令和6）年3月13日に可決し、EU理事会が同年5月21日に最終承認した。特定の規定を除き、発効（同年8月1日）から2年後に適用されることになる。

AI法の特徴としては、①リスクベースアプローチの採用②遵守しない場合のリスクが比較的大きい③幅広く適用される点が挙げられる。①は4つのリスクに応じて、規制の種類を定めるものである。②については、最大3500万ユーロまたは全世界売上高の7%のいずれか高い方の金額が罰則として規定されていることが挙げられる。③については、EU域内にAIシステムを提供する域外企業も適用対象とするというものである。

### 3 日本

日本においては、AI開発を後押しし日本の競争力を高める思想の下、2023（令和5）年以前はAIに関するハードローを定める動きはく、2024（令和6）年4月19日には、総務省と経済産業省が共同して「AI事業者ガイドライン（第1.0版）」（ソフトロー<sup>\*23</sup>）を公表した（最新版は、2025（令和7）年3月28日発表の「AI事業者ガイドライン（第1.1版）」<sup>\*24</sup>）。もっとも、前述のAIの様々なリスクを鑑み、2024（令和6）年2月に、自由民主党AIの進化と実装に関するプロジェクトチームワーキンググループが「責任あるAI推進基本法（仮称）」<sup>\*25</sup>を発表した。これを契機に、同年7月には、政府内では、AI戦略会議のもとで法規制の在り方を議論するAI制度研究会が立ち上がり、2025（令和7）年5月28日には、AI活用推進法が成立するに至った。

AI活用推進法は、イノベーション促進とリスク対応の両立を目指し、活用事業者の国の情報収集等の協

\* 18 <https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>

\* 19 <https://www.whitehouse.gov/presidential-actions/2025/01/initial-rescissions-of-harmful-executive-orders-and-actions/>

\* 20 <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>

\* 21 [https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill\\_id=202320240SB942](https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942)

\* 22 <https://www.nysenate.gov/legislation/bills/2025/S6953/amendment/B>

\* 23 [https://www.meti.go.jp/shingikai/mono\\_info\\_service/ai\\_shakai\\_jisso/pdf/20240419\\_1.pdf](https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_1.pdf)

\* 24 [https://www.meti.go.jp/shingikai/mono\\_info\\_service/ai\\_shakai\\_jisso/pdf/20250328\\_1.pdf](https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20250328_1.pdf)

\* 25 「責任あるAIの推進のための法的ガバナンスに関する素案」（<https://www.hirataku.com/uploads/65d41b304bf0992d09793b08.pdf>）参考。

力義務を除き、民間事業者に対して直接の義務を創設するのではなく、協力義務違反に対する罰則の規定も存在しない。

もっとも、附帯決議において、「AIのリスクへの対応について、常に最新の知見の情報収集に努め、必要な対応について不断の検討を行うこと。また、既存の法令やガイドライン等によっては対応が困難な新たなリスクが顕在化した場合においては、そのリスクの程

度に応じて規制の度合いを変えるリスクベースアプローチに基づいた規制的措置の導入も含め検討し、その結果に応じて必要な措置を講ずること。<sup>\*26</sup>」とされている。法令による規制は事業者の自主的な努力による対応が期待できないものに限定して対応していくのを原則として、新たなリスクが顕在化した場合、ハードローの制定が検討されるものと考えられる。

### 第3 弁護士会としてなすべき対策

これまで述べてきたように、生成AIの利活用には一般的に留意すべき問題やリスクがあり、その解決や対処には技術の発展に委ねるべき部分と、法令やガイドライン等の規範に委ねるべき部分とがある。この点、我々弁護士がその業務上で生成AIを安心して利活用するにあたっては、上述したような一般的なリスクを、弁護士業務の遂行過程における個別のリスクと置き換えて理解する必要がある。それらの理解は技術の発展によって補えるものではなく、また、弁護士業務に特化した法令の制定も現実的でない。となると、弁護士が自らの手でリスク回避のための指針（ガイドライン）を設ける必要があり、具体的には弁護士会がそれらの指針を示すことで、個々の弁護士の懸念を払しょ

くすることについて責任を持つことが望ましい。

この点、生成AIをめぐる技術の発展は日進月歩であり、生成AIの利活用を通じ従前では想定されてこなかったような新たな論点が今後も生まれる可能性は否定できない。それらの観点から、弁護士会としては、上述したガイドライン等の制定にとどまらず、生成AIをめぐる技術の発展の動向や各国の規制の動向その他利活用をめぐる様々な事例を注視しつつ、適時に会員たる弁護士に情報を提供し、必要に応じて上記ガイドライン等の柔軟な改訂と周知に努めることが望まれる。

\*26 「人工知能関連技術の研究開発及び活用の推進に関する法律案に対する附帯決議」([https://www.shugiin.go.jp/internet/itdb\\_rchome.nsf/html/rchome/Futai/naikaku2C59590D86DDF9A049258C73000C0C74.htm](https://www.shugiin.go.jp/internet/itdb_rchome.nsf/html/rchome/Futai/naikaku2C59590D86DDF9A049258C73000C0C74.htm))。

## 第4章 生成AIの適正な利活用により、業務効率化のみならず、業務のクオリティ向上を目指すべきである

### 第1 統合イノベーション戦略推進会議決定「人間中心のAI社会原則」<sup>\*27</sup>

「人間中心のAI社会原則」は、その基本理念の一つに「人間の尊厳が尊重される社会（Dignity）」を置いている。この理念は、「我々は、AIを利活用して効率性や利便性を追求するあまり、人間がAIに過度に依存したり、人間の行動をコントロールすることにAIが利用される社会を構築するのではなく、人間がAIを道具として使いこなすことによって、人間の様々な能力をさらに発揮することを可能とし、より大きな創造性を発揮したり、やりがいのある仕事に従事したりすることで、物質的にも精神的に豊かな生活を送ることができるような、人間の尊厳が尊重される社会を構築する必要がある。」というものである。

その上で、「人間中心のAI社会原則」は、「人間中心の原則」について、「AIの利用は、憲法及び国際的な規範の保障する基本的人権を侵すものであってはならない。AIは、人々の能力を拡張し、多様な人々の多様な幸せの追求を可能とするために開発され、社会

に展開され、活用されるべきである。AIが活用される社会において、人々がAIに過度に依存したり、AIを悪用して人の意思決定を操作したりすることのないよう、我々は、リテラシー教育や適正な利用の促進などのための適切な仕組みを導入することが望ましい。」と述べている。

弁護士が生成AIを利活用する際の心構えや原則も、上記基本理念及び「人間中心の原則」と同様、「弁護士が生成AIを道具として使いこなすことによって、弁護士の様々な能力をさらに発揮し、より大きな創造性を発揮し、あるいはやりがいのある仕事に従事すべき」こと、「弁護士による生成AIの利用は、憲法等が保障する基本的人権を侵すものであってはならない」こと、「弁護士が生成AIに過度に依存したり、弁護士がその意思決定を生成AIによってコントロールされたりしてはならない」ことに主眼を置くべきものとなるだろう。

### 第2 弁護士が生成AIを利活用する際の心構えと原則

前記の「人間中心のAI社会原則」を踏まえると、弁護士が生成AIを利活用する際の心構えと原則は、以下のとおりであると考えられる。

#### 1 生の事件情報をそのままプロンプト入力することは控えるべきである

現時点では、生成AIに生の事件情報をそのまま入力することは、秘密保持義務、個人情報保護法などの規律に違反するおそれがあることを踏まえると、いくら便利であるからといっても、これは慎むべきである。

生成AIに事件情報を入力する場合は、適宜の匿名化処理を施す、事件情報を一定程度抽象化するなどの

対応が必要となろう。API連携等がなされた生成AIベンダーのシステムを利用するという対応も考えられるが、相応の費用を要するため、一定規模の法律事務所や企業でない限り導入は難しいかもしれない。

#### 2 生成AIが出力した回答をそのまま利用することは控えるべきである

尤もらしい架空情報を生成することがある生成AIの幻覚（Hallucination）現象を踏まえると、生成AIが出力した結果をそのまま案件処理に利用することも慎むべきである。ウェブ検索機能を利用すること、あるいは生成AIの幻覚現象を緩和する技術として知られる検索拡張生成（RAG; Retrieval Augmented Generation）を併用したサービスを利用することは有

<sup>\*27</sup> <https://www8.cao.go.jp/cstp/ai/aigensoku.pdf>

用であるが、それでも生成AIの回答が正確である保証は全くなく、現に法律問題への回答をさせた場合には適用法令を誤り、あるいは引用する裁判例が架空のものであることも多い。そのため、生成AIが出力した結果を妄信することは禁物であり、その回答内容については必ずソースや場合によっては原典・一次資料を確認することが重要である。

法令のリサーチや訴訟提出書面の作成などの弁護士のコアタスクをAIに丸投げするなど、AIへの過度な依存は絶対に避けなければならない。あくまでも弁護士自身が自分の頭で事件の内容を咀嚼し、納得しながら弁護士業務を進めるべきであって、生成AIはそのための道具に過ぎないことは強く自覚していなければならない。弁護士業務を行う際に生成AIの回答を利用するという判断の最終責任は弁護士にある。

### 3 弁護士各人の業務内容に応じた有益な利活用方法を模索・創意工夫すべきである

現時点では、生の事件情報をそのまま入力すること、生成AIの回答をそのまま案件処理に利用することは、原則として慎むべきである。上記のような限界はあるものの、上手に使える、弁護士業務やその周辺業務において有用なツールとなり得る。大量かつ多様やデータを学習した生成AIは、その汎用性ゆえに、弁護士

業務やその周辺業務にも大いに活用できる可能性を秘めている。近年は生成AIの推論能力も大幅に向上しつつあり、一定の限度では法律推論を生成することが可能であることから、本来的な弁護士業務を強力にサポートしてくれる可能性も大きい。

生成AIというと業務効率化の側面にスポットが当たることが多いが、単なる業務効率化に留まらず、書面のクオリティその他の業務の成果を向上させるような積極的な利活用の方法を模索すべきである。弁護士業務の内容ややり方は弁護士ごとに千差万別であると考えられるが、生成AIの積極的な利活用に資する共通の視点として重要であるのは「良い質問をする」、「良いプロンプトを入力する」ことであろう。「良い質問をする」、「良いプロンプトを入力する」とは、生半可なプロンプトエンジニアリングのような小手先のテクニックを弄することではない。今直面している法律問題において何が解決すべき問題なのか、それを解決するにはどうすべきかを常に念頭に置き、文章（プロンプト）の形で表現することは、法律家が法的思考を深めるために重要である。悩んでいる問題に対しプロンプトを考えながら頭を整理する過程で、事件に対する自分の理解が深まっていく。法律家としての自らの思考を深化させるための着想・アイデアを得る（ブレインストーミング）の使い方が有効である。

## 第3 東京弁護士会・日本弁護士連合会ガイドラインの遵守

### 1 はじめに

生成AIの利活用を推進すべきであるといっても、生成AIの限界・危険性を踏まえた適正な利活用をするべきことはいうまでもない。この点、東京弁護士会や日本弁護士連合会が生成AIの利活用につき、どのような視座を持っているかを確認、理解し、その動向について注目することは、弁護士会の重要な政策団体である法友会としても極めて有用である。以下では東京弁護士会、日本弁護士連合会の順に、生成AIないし生成AIサービスとの向き合い方につき俯瞰する。

\*28 なお、いずれの取組についても本稿執筆時点で明らかにされているものであり、逐次進行中のものである点は留意されたい。

### 2 東京弁護士会の取組み

東京弁護士会では、弁護士業務改革委員会新業務研究部会が中心となり、「弁護士業務における生成AIサービスの適正利用ガイドライン」（以下「東弁ガイドライン」という。）を制定し、2025（令和7）年3月27日付で施行している<sup>\*29</sup>。東弁ガイドラインは、生成AIサービスを円滑かつ適正に活用し、弁護士業務の改善に資することを目的として制定されたものであり、前記第3章の生成AIの危険性を踏まえ、入出力情報が外部に漏洩するリスクのない生成AIサービスの選定基準、誤情報（ハルシネーション）のおそれとファクト

\*29 現時点では公開されていないものの、東京弁護士会会員であれば、会員サイト上の弁護士業務改革委員会のページにて閲覧可能である。

チェックの不可欠性、生成AIサービスの利用に際し依頼者の意思を尊重しその意向に配慮した対応をすべきこと、日本弁護士連合会の弁護士情報セキュリティ規程（会規第117号）に基づく適切な安全管理措置を講じるべきこと等を述べており、これまで本稿で述べてきた点及び次項で述べる留意点を概ね踏襲した内容になっている。なお、東弁ガイドラインの特徴は、生成AIの特殊性に鑑みた利活用の注意点への言及にとどまらず、生成AIサービスを利用した弁護士業務における報酬算定の考慮要素についても触れている点にある。すなわち、「生成AIサービスの利用には一定のコストが伴ううえ、これを弁護士業務に適切に利用すること自体が高度な専門スキルとして評価されるべき場合もある。そのため、生成AIサービスの利用によって依頼者に対してより高い付加価値を提供できる場合には、その点を考慮し、従来と同等またはそれ以上の報酬体系を採用することも選択肢として否定されるべきではない。」（東弁ガイドライン「2（4）」）とあり、生成AIサービスの利活用が、弁護士業務の価値向上に資する可能性について述べており、この点は我々の弁護士業務の在り方につき、大いに示唆に富んだ内容といえる。

なお、東弁ガイドラインは、その遵守が義務付けられるものではなく、綱紀又は懲戒の直接的な基準とされることも想定していないとのことであるが、弁護士が生成AIサービスを実務上利用するに際し、遵守することが望ましい準則が記載されており、生成AIの利活用の際には参照しておくことが望ましい。

### 3 日弁連の取組み

日本弁護士連合会でも、AI戦略ワーキンググループを中心に、生成AIを利用する際の行動規範が必要であるとの問題意識から、これまでガイドラインの策定が進められてきたところ、2025（令和7）年9月に「弁護士業務における生成AIの利活用等に関する注意事項」（以下「注意事項」という。）が策定された<sup>\*30</sup>。その詳細な内容について本稿での言及は避けるものの、概ね本稿で述べるところの留意点について問題意識を同じくしていると理解できる。注意事項もまた、綱紀・

懲戒の直接の基準とされることは想定されていないが、PCやスマートフォンを利用するほぼ全ての弁護士を対象として、生成AIを利活用する場面が広がることを念頭に、一定の注意点についてまとめたものであることからすれば、やはり弁護士がその業務遂行にあたり十分にその内容を了知し理解することが強く望まれる。

### 4 小括

以上のとおり、東弁、日弁連のいずれにおいても、生成AIの利用に関するガイドライン等が策定、施行されており、われわれ弁護士が業務を行う上で特に指針となるべき情報に接することができるようになったと言える。なお、いずれのガイドライン等についても、技術の進歩や利活用の実態などを踏まえて逐次改定することが想定されているようである。弁護士としてはそれらの改定の動向につき留意し、場合によっては自己の経験を踏まえ、当該改定に情報提供その他の態様で関与するようなことも有意であろう。

いずれにせよ、今後は、弁護士が業務で生成AIの利活用を図る際に、東弁ガイドラインや注意事項の内容を踏まえつつ、個別具体的な検討と判断を行うことが求められる。そこでは例えば、生成AIサービスを利用して得られた生成物や生成AIを用いた解析結果等を訴訟上証拠として提出することの可否や適否、提出する際の留意点などについて、弁護士倫理等の観点から真摯に検討することが必要になってくる。具体的には、本稿ですでに指摘したハルシネーションなどへの留意が想定されるが、弁護士自身が生成AIを利活用する場合にとどまらず、依頼者その他の利害関係者により提供される証拠や資料が生成AIを用いて生成されたものか否かといった視点も場合によっては必要になるであろう。いっぽう、弁護士による生成AIの利活用を経た弁護活動は、従来の弁護士業務とは異なる付加価値を提供できる可能性もある。単に生成AIの利活用に伴う懸念や弊害にのみとらわれることなく、それらの懸念を払しょくすることで産み出せる新たな意義の可能性についての考察を深めることこそ、生成AIの利活用に際して求められる姿勢と言えよう。

\*30 日弁連会員サイト上において閲覧可能であるが、会員外への交付、配布及び転送等はいできない。

## 第4 法律の非専門家が生成AIに法律問題への回答を求めて安易に使用することの危険性についての周知の必要性

### 1 はじめに

本項は、直接には弁護士による生成AIの適正な利活用に言及するものではないが、法律問題への回答を求める主体が弁護士である場合と非専門家である場合の比較対照を通じて、非専門家が法律問題について安易に生成AIを使用することの危険性に言及するものである。弁護士による生成AIの利活用と関連する問題であることから、本項において記述することとした。

### 2 生成AIの「社会的追従性」について

対話型文書生成AIは、ユーザーに対して過度に同調したり、ユーザーの感情を無批判に受け入れて共感を示したりする「社会的追従性」という特性があると指摘されている<sup>\*31</sup>。すなわち、生成AIは、ユーザーの感情を「忖度」して応答を生成する傾向がある。この特性は、前述した生成AIの基盤モデルの学習方法（人が好む応答に高い報酬・評価を与えるRLHFの手法によって獲得されたと考えられている。

### 3 法律の非専門家による生成AIの誤った使用の危険性

法律の非専門家が、自己が当事者となった法律問題についての回答を求めて生成AIを使用する場合は、法律の専門的知識の欠如に加え、自己に都合のよい事実・証拠に着目しやすく、不都合な事実・証拠には気づきにくいといったバイアスが生じることも十分に考えられる。非専門家たるユーザーが法律知識の欠如とバイアスに起因して生成AIに偏頗な質問（プロンプト）を入力し、当該質問から窺われる感情に「忖度」した偏頗な回答を得てこれを受容し受け入れてしまうとすれば、そのようなユーザーの思考傾向と「社会的追従性」という生成AIの特性とが相俟って、ユー

ザーが生成AIとの繰り返しの対話を通じて、自己の主張の正当性について誤った思い込みを強めてしまう危険性が生じ得る。

この点につき、OpenAI社のChatGPT利用規約は、「アウトプットは常に正確であるとは限りません。お客様は、本サービスからのアウトプットを、真実又は事実に基づく情報の唯一の情報源として、又は専門家のアドバイスの代わりとして依拠すべきではありません。」「…法律…など、個人に法的又は重大な影響を与える可能性のある目的において、その個人に関連するアウトプットを使用してはなりません。」と規定しており、法律問題に対する生成AIの回答を無批判に受け入れてそのまま使用し、専門家への確認を怠るといった安易な使用行為を禁止しているが、そのような禁止事項もユーザーの思い込みの前には無力であろう。

そのため、弁護士会としては、法律の非専門家が生成AIに法律問題への回答を求めて安易に使用することの危険性を周知すべきである。

### 4 弁護士法72条との関係について

なお、法律の非専門家が生成AIの利活用を通じ法律問題への回答を行うなどといったことをすれば、それは非弁護士が法律事務に関するアドバイスをおこなうことと同等の行為であり、弁護士法72条が禁止する行為となり得ることは留意が必要である。同条該当性については主に上述した法律事務該当性のほか、報酬を得る目的やいわゆる「事件性」といった要件の充足有無が問題となり得るところ、非法律専門家による生成AIを用いたサービスの提供あるいは利活用がこれらの要件をいずれも充足する場合には、いわゆる非行為として同法の規制対象となり得るのである。

この点法務省は、2023（令和5）年8月1日、「AI等を用いた契約書等関連業務支援サービスの提供と弁護士法第72条との関係について」（以下「法務省ガイドライン」という。）を公表し、契約書等を扱う業務をAIを用いて支援するサービスに関して、ユーザーに

\*31 Myra Cheng, et al. Social Sycophancy: A Broader Understanding of LLM Sycophancy. arXiv:2505.13995v1 [cs.CL].

\*32 このようなバイアスは、法律の専門家であっても、自身がトラブルに陥った場合には生じ得るものであり、必ずしも非専門家に固有な問題ではないことには留意すべきである。

\*33 <https://www.moj.go.jp/content/001400675.pdf>

よる非定型的な入力内容に応じ、個別具体的な解決案を提示するようなリーガルテックサービスを「鑑定…その他の法律事務」に該当し、弁護士法72条に抵触する疑義が生じ得るサービスと位置付けた。

かかるガイドラインは、弁護士法72条に関する従来の最高裁判例の立場を超えるものではないものの、AIの利活用に際しての同条の考え方につき一定の視座を与えたものと理解することができる。そこでは、非法律専門家による生成AIの利活用がそのまま一律に弁護士法72条に抵触するわけではないものの、例えば、生成AIを用いて一定の情報を得たうえで、当該情報をもとに、個別の事案に応じた法的リスクの有無やその程度を提示するような法的サービスを有償で提

供するような場合には、同条に抵触しうることを示唆している。

そのため、弁護士会としては、法律の非専門家が生成AIに法律問題への回答を求めて安易に使用することは弁護士法72条の趣旨に抵触し、自己の法的権利を毀損しかねない行為であることの周知に努めるべきである。

以上に対し、弁護士が法律問題への回答を求めて生成AIを使用することは、弁護士法72条の問題を一切生じさせるものではなく、その活用方法によっては弁護士業務の成果物等のクオリティを向上させることに資するものである。

---

\*34 少なくとも現時点の技術水準で用いられるAIであれば、たとえ非定型的な入力があったとしてもその内容を踏まえた自律的判断に基づいた回答をするのではなく、当該入力に応じた統計処理を通じた類型的な回答をするに過ぎないと整理すれば、「個別の事案における契約に至る経緯やその背景事情、契約しようとする内容等を法的に処理」することにはならず、「鑑定…その他の法律事務」に該当しない、と整理することができる。

## 第5章 まとめ

### 第1 生成AIの限界・リスクへの対策について法的側面からの検討を継続すべきであり、その際には、技術革新の促進を阻害しないような法令の解釈と立法提言を行うことが重要である

本稿の第3章において、生成AIの限界・リスクとそれらへの対策が急務であることと、我が国を含む各国の検討・対策及び弁護士会の検討・対策の内容を論じた。

生成AIが有する汎用的な能力は、今後においても短期間のうちに指数関数的に増強されることが容易に予測されるが、そのような生成AIの能力の増強に応じ、生成AIの限界はますます浮き彫りになると共に、生成AIがもたらし得るリスクもますます増大する可能性がある。前記AI戦略会議・AI制度研究会の2025（令和7）年2月4日付「中間とりまとめ」も、国内において、「性的な動画コンテンツの出演者の顔を芸能人の顔にすり替えインターネット上に公開した事件」、「生成AIを用いてコンピュータウイルスを作成した事件」、「生成AIを用いて女性のキャラクターを作って現金をだましとった事件」が実際に発生しているとして、警鐘を鳴らしているところである。<sup>\*35</sup>特に、人の生命・身体・財産といった基本的な権利利益に対し重大

な問題を生じさせる可能性が高いAI技術については、当該リスクの内容やその社会的な影響の重大性に応じた規律の必要性の有無を検討することも必要である。<sup>\*36</sup>

以上のような観点から、生成AIの各種の限界やリスクに対しては、その技術的進歩に応じた制度的・技術的な対応が急務となるが、弁護士会は、そのうち制度的な対策、特に法令解釈・立法提言など法的側面からの対策を早急に検討すべきである。

もっとも、生成AIに関連する諸法令の解釈及び立法提言を行うに際しては、AI技術に関するイノベーションの促進とリスク対応の両立を目指すことを旨とするAI活用推進法を尊重することが必要である。

本稿は、弁護士会が生成AIの限界やリスクに対する法的側面からの検討を行うに際しては、法令やソフトローによる対応を進めている各分野の所轄府省庁と歩調を合わせる形で、技術革新の促進を阻害しないような諸法令の解釈と立法提言を行うことが必要である旨を提言するものである。

### 第2 弁護士は、生成AIの基本的な仕組み、生成AIの利便性、限界・リスクを踏まえつつ、個々の弁護士業務における積極的な利活用を模索すべきである

本稿の第2章において、生成AIの基本的な仕組みの理解は重要であることを論じ、第4章において、生成AIの適正な利活用により、業務効率化のみならず、業務のクオリティ向上を目指すべきであることを論じた。

弁護士がその業務において何ら予備知識を持たずに生成AIを使用するとなれば、秘密保持義務をはじめとする法令違反のおそれや、弁護士職務基本規程違反

を含む弁護士倫理上のリスクが生じ得る。もっとも、生成AIの基本的な仕組みや生成AIの限界・リスクを踏まえていれば、安全・安心な活用が可能となることから、生成AIを積極的に利活用することに何ら躊躇する理由はないはずである。

今後において、各々の弁護士が、生成AIの基本的な仕組みや生成AIの限界・リスクを踏まえつつ、各人により千差万別であろう弁護士業務の内容・性質に応じ、日進月歩の進化を遂げつつある生成AIが持つ汎用的な能力の最大限の活用を模索していけば、依頼者に対するより良質なリーガルサービスの提供と弁護

\*35 前掲・「中間とりまとめ」8頁。

\*36 前掲・「中間とりまとめ」11頁参照。

士業務の大幅な効率化が可能となり、もって弁護士業務を大きく変革させる十分な可能性を秘めているといっても過言ではない。

本稿は、以上のような観点から、個々の弁護士が、生成AIのごく基本的な技術的概要と生成AIの利便性、限界及びリスクを理解しつつ、各種のガイドラインを遵守しながら、「人間中心のAI社会原則」における「人間中心の原則」の趣旨に則り、「弁護士が生成AIを道具として使いこなすことによって、弁護士の様々な能力をさらに発揮し、より大きな創造性を発揮し、ある

いはやりがいのある仕事に従事すべき」との発想の下に、弁護士業務への利活用の方法を積極的に模索すべきである旨を提言するものである。

生成AIの技術進歩は著しく、数か月単位で新しい技術が公表されるような状況にあることを踏まえ、各種のガイドラインについては継続的に検証を重ね、必要に応じて改訂を行うとともに、会員に対する適宜のタイミングで研修を行うことも併せて検討されるべきである。