R5-583

生成系 AI の倫理的・法的・社会的課題を踏まえた今後の利用可能性に関する調査研究

（一社）科学技術と経済の会　調査研究部　担当部長　兼　主任研究員　　田中　祐耕

「生成(系)AI(Generative AI)[1]」は、膨大なデータベースをよりどころに、求める内容に対して優れたインタラクティブ性でたちどころに応えるツールとして注目され、その利用が急速に広がるところとなった。一方で、その得意不得意、利用条件があまり知られていなかったり、その特性に対応したルールが未整備であったりして、一部誤解や混乱を招いたりする事案も生じたりしている。

これらを未然に防止し、その性能を十分に発揮できるよう、よりよく効果的に使用できるようにしていくためには、どのような配慮が必要となるか、このような問題意識を背景に、本調査に取り組んだ。

本調査では、まず生成（系）AI の現状動向について、とらえるべき対象の全体像を押さえるべく、その種類や実現手法、および企業の活用例等の観点から概要を整理した。そして、そのうえで今般、主題とした生成(系)AIの倫理的・法的・社会的課題と対応について、顕在化した事案や業界一般での現時点の認識状況、現行の国内外の法規制状況等を調査し、整合関係や要対応事項等の分析・考察した。

本調査にあたっては当会主催研究会への講師招聘、有識者へのインタビュー実施、当会会員との意見交換およびアンケート実施、関連展示会出展企業への照会・シンポジウムの聴講、文献調査等により実態調査、分析を行った。

1．生成（系）AI の発展経緯、現状動向と活用状況

生成（系）AI は、人手を介さず自ら知識を取得するディープラーニング（Deep Learning：DL）と称する優れた AI 学習技術や、画像処理で画期的な高速処理を可能にした GPU（Graphics Processing Unit）などのハードウェア技術を背景に、人とコンピュータがあたかも会話するかのようにやりとりできる技術として大いに関心を持たれ、現在のブームを迎えるに至っている。

人の問いかけに対し、文書のみならず画像、動画や音声、楽曲あるいはそれらを複合的に出力できるマルチモーダル型の生成(系)AI が登場し、対話やレポート作成、演奏、映画制作等、様々な領域で利用が試みられ成果を上げるようになった。

現在では企業においても内部業務サポートアシスタントあるいは商用サービスとして、実証あるいは実導入が始まっている。例えば、前者では業務改善のための過去取り組み事例の検索や、製品設計・開発時のリスク洗い出しによる関係部門間での連携への応用、後者においては、ネット出品時の利用者登録支援や旅行プランニングのサポートなどである。

企業や地方公共団体における生成(系)AI 活用の全体状況については、2023 年にシンクタンクが約 2,400 社を対象に実施した調査によれば、ビジネスパーソンの約半数が、生成(系)AI の業務応用を認知し高い期待と関心をもっており、実際全体の2割弱から既利用、試行中、検討中との回答を得ている。

その一方で、生成（系）AI の確率・統計的な手法から得られる出力結果と AI に対するこれまでの通念とのギャップ等から、その生成内容と期待や目論見との間に齟齬が生まれ、不安や懸念を抱かれるケースが顕在化するようになった。

例えば、故意・過失等による不適切な使用で生成された不正確で誤った出力結果から、誤認・誤解を生じ、信頼性の低下さらには社会的な混乱を招くようなリスクを内包している点である。これらについて欧米では関係当局が調査を開始し、必要な留意点を発表して注意を喚起するなど早い動きを見せている。日本においてもこれに応える形で、昨年のG7デジタル・技術大臣会合でのアジェンダの一つとし、生成（系）AIに関する議論の場をもつことについての合意を得るに至っている。

---

[1] 生成系 AI、生成 AI 等いくつかの用法があり、これらをまとめ以下、生成(系)AIと表記した。

2．生成(系)AIの倫理的・法的・社会的課題と対応

生活上経済的苦境に追い込まれた中で ChatBOT との対話が自殺の引き金になったり、ネット上に流布している誤った歴史的解釈等から人種・民族的差別を想起する画像が生成されたりするなど倫理的・法的・社会的な側面からの懸案事項が知られるようになった。

これらについては、IT 新技術の導入時にも同様のリスクがあり、その対処が必要であることが先行研究で発表されている。例えば、顔認識技術での判定基準設計や、画像検索技術でのラベリングにおける不適切さが偏見・差別を引き起こす要因となっており、個々のケースに正しく対応するとともに、それらのリスクを念頭に、設計段階に参照すべき基本原則を定めるべきといった指摘がなされている。

生成（系）AIに関しても、法制、政策、社会問題等の専門家が加わって、そのリスクに関するカテゴライズやその影響度、発生メカニズム等の研究が行われ、整理が進められている状況にある。

リスクカテゴリ	発生メカニズム	発生リスクの有害性種別
I. 差別、排除、及び有害性 (Discrimination, Exclusion and Toxicity)	反感、不平、悪意を反映した話し言葉の誇張、極論、ボ暴言などをそのままに学習データとして言語モデルが取り込む	社会的偏見を持つステレオタイプの形成 排他的な規準 悪意を含む表現 マイノリティグループへの不適切な扱
II. 情報ハザード (Information Hazards)	学習データに含まれていた個人情報や機密データなどから、さらに秘匿されるべき機微な内容を含む情報を推測する	プライベートデータ 機密情報の混入による示唆、漏洩
III. 誤情報による害 (Misinformation Harms)	学習データの欠陥、欠如等から、誤認、曲解する	誤った情報の拡散 誤報を根拠とした実際の誤作動 違法、非倫理的な行動の勧奨
IV. 悪用(Malicious Uses)	ユーザーや製品開発者の故意による言語モデルの悪用	扇動、ネガティブキャンペーン強化 大規模な詐欺の模倣 サイバー攻撃の支援 違法探索、検閲
V.人間とコンピューターの相互作用による害(Human Computer Interaction Harms)	会話内容を直接取り込むことによる不適切な学習	擬人化システムへの過度の依存 プライバシー情報抜取のためのチャネル形成 人種・民族、性別等に関する有害なステレオタイプの形成
VI. 自動化、アクセス、及び環境への害 (Automation, Access, and Environmental Harms)	適切でない特定グループのみに利益を供与するアプリケーションを設定し、流布する	不公正の助長 創造性の奪取 コンピュータ資源アクセスの制約、能力発揮の制限

このようなリスクに対し、如何に対応するかについても同時に重要課題として研究も進められている。例えば、予防接種の効果を心理学に応用した「接種理論」に基づき、ゲームとしてフェイクニュース生成を体験し、その仕組みの理解を「ワクチン」として、実際のフェイクニュースに接したときにも鵜呑みにしない「心の強さ」を持てるようにするものがある。

これら各種研究成果を踏まえ、各国でも適切なガバナンスを可能とするべく、政府・当局において法的整備のための枠組み作りがすすめられている。例えば英国では生成（系）AI を含む AI 規制に関する

#	原則	内容
1	安全性、セキュリティ、堅牢性 (Safety, security and robustness)	AIシステムは、そのライフサイクルを通じて、堅牢、セキュア、かつ安全でなければならず、リスクは常に特定、評価、管理されなければならない。
2	適切な透明性・説明可能性 (Appropriate transparency and explainability)	AIシステムの開発者・実装者は、関係者に対してAIシステムがいつ、どのように、どのような目的で使用されているかの情報を十分に提供し（透明性）、関係者に対して、AIシステムの判断形成過程について十分な説明を提供しなければならない（説明可能性）。
3	公正さ (Fairness)	AIシステムは、そのライフサイクルを通じて、個人/法人の法的権利を侵害してはならず、個人を不公正に差別してはならず、不公正な市場結果をもたらしてはならない。
4	説明責任とガバナンス (Accountability and governance)	AIシステムの供給と使用について効果的な監視を確保するガバナンス体制が構築されなければならず、AIシステムのライフサイクルを通じて明確な結果に対する説明責任が伴わなければならない。
5	不服申立て・是正ができること (Contestability and redress)	AIによる判断や結果が有害であり、又は重大なリスクを伴う場合、それによって影響を受ける者に対して不服を申し立て、是正する機会を提供しなければならない。

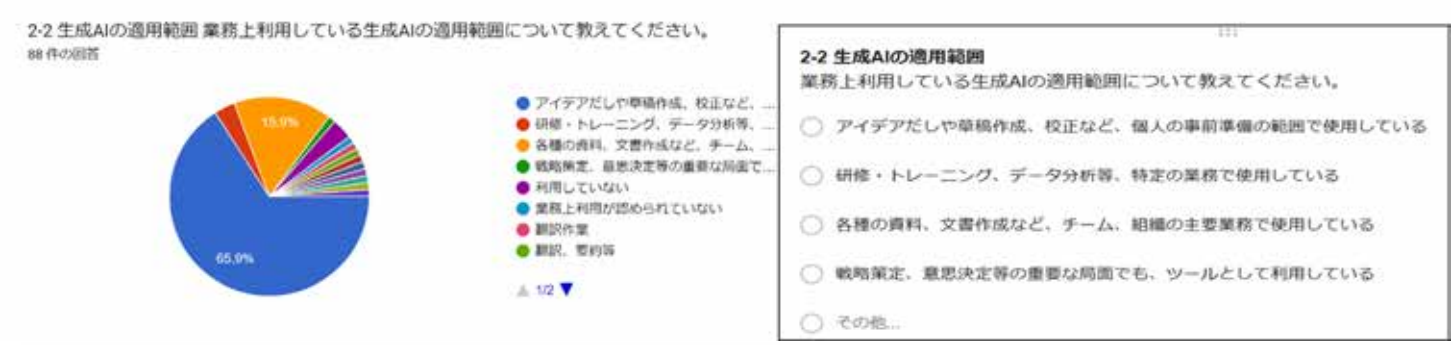
5原則を定め、人権擁護等、基本的な価値を保護しつつ、規制の不透明さを縮減し、産業のイノベーシ

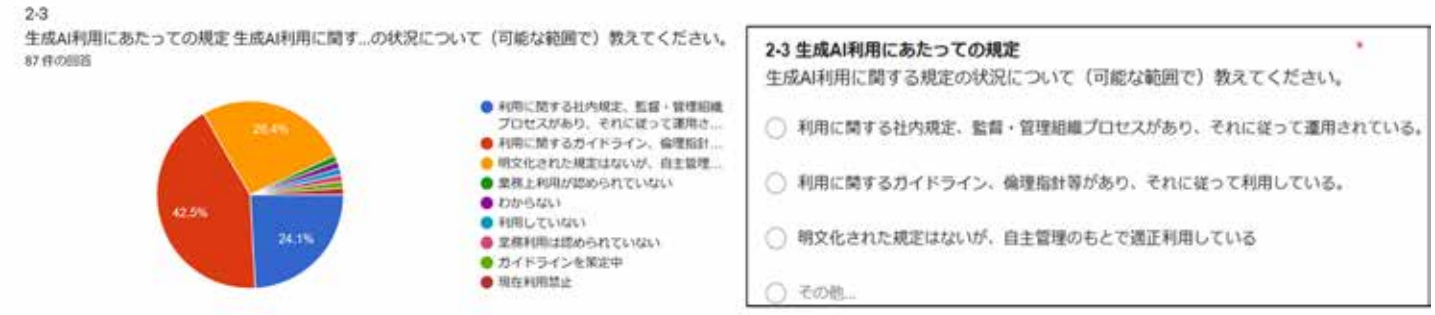
ョンに資するアプローチを図ろうとしている。

日本においても、2024 年2月に与党プロジェクトチームにおいて、議員、アカデミア、実務者参加のもと「責任ある AI 推進基本法（案）」をまとめ、社会的影響力の大きい「特定 AI 基盤モデル」の定義、開発企業に対する第三者の脆弱性検証等の体制整備要件の設定、技術進歩に遅れない官民共同規制方式の採用などが述べられている。

3．産業界現場での関心・反応の状況

これらに対し、現場一般では実際に生成(系)AI とどの程度接触し、どのような印象を得ているかについて、当会会員企業116社を対象にアンケート調査を実施した。当会趣旨を理解し、研究会出席頻度も比較的高い層を中心に 89 件の有効回答を得た。回答者は50～60歳代、男性、管理職および役員が中心となった。





生成（系）AI の業務利用については、現状7割程度がアイデアだし等、個人作業用途が7割を占めていること、業務利用規定に関しては、ガイドライン、倫理指針等が整えられているところは4割程度であることがわかった。

4．今後に向けての対応、まとめ

これまでの発生事案や課題研究、法整備等に向けての議論等を踏まえ、今後に向け考慮すべき観点としては以下の様なものが挙げられる。すなわち、生成（系）AIへの技術的対応だけでなく、改めて前提となる人間・社会の課題としてのアプローチをとること、生成（系）AI の扱い方、接し方についても、人と人とのコミュニケーションの在り方に立ち返り考慮することが必要である点である。例えば、情報発信側と受信側で受け渡される内容の意味合いが異なっていたり、普遍的なルールとしての道徳と、現実の具体的な状況での振舞い方には個々の場合によって差があったりする。情報の受け取り方に差異が生じてしまうことは避けられない側面もある。しかし、それゆえにこそ丁寧な対応が必要であり、その積み重ねが生成（系）AI の利用に際して生じる課題を解決し、未然防止策を検討する端緒となるのではとも考えられる。

生成（系）AI は発展途上の技術であり、そのポテンシャルが十分に発揮され役立つものとなるよう、その動向を引き続きフォローしていきたい。

以上