

# 研究・教育における生成AIの 利活用の方法とその考え方

高等教育研究叢書

177 2025年3月

野内 玲・井野瀬 久美恵 編



広島大学

高等教育研究開発センター

研究・教育における生成 AI の利活用の方法と  
その考え方

野内 玲・井野瀬 久美恵 編

広島大学高等教育研究開発センター



## 目次

序文	野内 玲（広島大学）	1
第1章	研究活動における生成 AI の利用と研究者の責任	
	野内 玲（広島大学）	3
第2章	技術の発展と研究公正 – 法哲学の観点から –	
	大屋 雄裕（慶應義塾大学）	13
第3章	産業界における AI 倫理の現状と課題	
	樋笠 堯士（多摩大学）	25
第4章	AI 活用に伴う法的課題	
	樋笠 知恵（信州大学）	37
第5章	論文執筆と発表における生成 AI の適正利用と課題	
	岡林 浩嗣（筑波大学）	47
第6章	中高の現場における生成 AI の現状と課題	
	南里 翔平（市川中学校・高等学校）	61
第7章	人文社会科学研究における生成 AI	
	藤井 基貴（静岡大学）	73
第8章	生成 AI で研究不正は変わるのか？	
	井野瀬 久美恵（人間文化研究機構）	85



## 序文

野内玲  
(広島大学)

本書は一般財団法人公正研究推進協会（APRIN）2023 年度全国公正研究推進会議（2024 年 2 月 16 日開催）（以下、推進会議）における登壇者による当日の発表内容及びその後の検討成果を言説化・文章化したものである。

推進会議の全体テーマは「教育・研究機関における生成 AI の取扱いについて」であり、当日は基調講演等の会合全体としての講演に続き、研究領域に分かれた分科会（生命医科学系、理工学系、人文・社会科学系、中等教育系）と、推進会議のサテライトイベントとして研究者向け、研究倫理教育に携わる研究支援者・事務職員向けの会合が開催された。本書の中心にあるのは、これらの中でも人文学・社会科学分科会のテーマ設定「技術革新の中の研究公正」である。同分科会では生成 AI によって研究活動の何が変わり、何が変わらないのかを議論し、研究公正という概念が内包する「普遍的な要素」を浮き彫りにすることであった。しかし限られた時間枠しかなかったため、必ずしもその目的を達成できた議論が行えたとは言えない。そこで、人文学・社会科学系分科会以外で登壇した発表者の見解をも集積することで、多面的に、研究活動自体と生成 AI（ないしは AI 技術一般）の関わり、さらには研究公正への視座を導こうと企画されたのが本書である。哲学・法哲学・法学・教育学・生命科学・生命倫理・医療倫理・中等教育教員という多様な背景をもつ著者らが集い、生成 AI が教育・研究現場に与えた影響や、各自の専門的観点から AI への基本的向き合い方を検討する学際的内容になった。章立てとしては、はじめに野内から生成 AI が学術界隈に与えた影響を総論的に概観し、各論として各先生方の論考が続いていく。本書全体の見取り図を確認してから読みたいという読者は、終章である第 8 章 4 節において井野瀬が各章の概要を振り返っているので、該当箇所を参照してから第 1 章に戻ると良いだろう。

本書は『高等教育研究叢書』シリーズという刊行物であるため、高等教育への関連についても言及したい。生成 AI に関してはその技術的発展が早く、大学における学術研究や教育、事務業務、企業等の産業活動などあらゆる場面で活用が進んでいる。生成 AI は研究の公正性、ELSI/RRRI といった幅広いテーマを内包しており、論文等研究成果の発表にも関わることから、研究者の業績やそれらの集積でもある大学ランキングといった指標にも影響を及ぼす。これらの観点からも高等教育分野としても正面から取り組むべき課題であることは間違いない。

一つの科学技術が多方面にもたらした影響を後に通時的な形で検証するためには、共時的な形で記述することが必要である。本書はそうした影響のうち、研究公正という一側面を切り取って描き、アーカイブ化する。もちろん網羅的な議論を展開するまでには至っていないであろうが、本書が今後の研究公正の推進の一つの礎となることを期待している。

なお、本書を作成するにあたり、一般財団法人公正研究推進協会には執筆者一同がそれぞれの発表内容を振り返ることができるよう、推進会議の動画記録を提供していただいた。ここに謝意を表す。

# 第 1 章 研究活動における生成 AI の利用と研究者の責任

野内玲  
(広島大学)

## 1. はじめに

生成 AI を研究活動で使用する際には検討・注意すべき様々な問題があることが知られている。そういった問題には生成 AI それ自体に特有のものも、研究者の行動規範（研究公正）に関する既存の考え方を再検討する機会となるものもある。科学研究における生成 AI への対応としては、生成 AI に関する問題の根本を現時点で認識可能な範囲で考慮した上で限定的な利用を許容しており、全面的に使用を禁止する組織は少ない。社会にこのような新しい技術が登場するたびに、その使用について議論が起きるのは避けられない。重要なのは、その技術によって何が違って何が変わらないのかという点の見極めが、新しい技術の受容において重要なことであろう。

このような観点を踏まえ、本章では生成 AI の中でも ChatGPT という大規模言語モデル（Large Language Models, LLM）について、2023 年に展開された学術的動向を振り返りたい（2 節）。それを踏まえ、生成 AI を一つの鏡として設定し、その鏡に照らされた研究公正の姿から、根本にある責任という概念について視点を素描する（3 節）。

## 2. 生成 AI をめぐる学術界の動向

本節では生成 AI が登場して以降、アカデミアでなされた同技術に関する動向を概観する。イラスト、音楽、動画を生成する AI もあるが、最も多くの研究者に関わるとされる ChatGPT（テキストやコードの生成など）を引き合いにする。

### 2. 1 研究面の対応 ～著者か謝辞か～

まず、時系列に沿って、論文の出版という研究活動の一側面に関する生成 AI の使用について展開された議論を国際的かつ幅広い文脈で取り上げたい。なお、生成 AI による出力結果の真偽の問題やハルシネーションの存在は、生成 AI それ自体の特性に由来するものとして懸念が共有されている。そもそも生成 AI で研究の実施成果の根幹に関わるデータを生成することも可能であるが（Taloni et al., 2023; Nature, 2023）、データの捏造・改ざんに関わる使い方は少ないものとしておきたい（論文に挿入する説明用画像の生成という補助的な使い方が関わる事例を後に示す）。また、ChatGPT とやり取りをした結果、架空の論文を引用したり、他人のテキストを意図せず盗用してしまったりする問題は、自分



で出典を確認するだけで対応可能である。そのため、本章では研究公正に関する問題全般ではなく、オーサーシップの観点を主に検討したい。

ChatGPT が 2022 年 11 月に登場した直後の 12 月 19 日, Kung et al. (2022) は ChatGPT 自体のパフォーマンスに関するプレプリント論文を執筆し, ChatGPT を共著者に含めて発表した (なお, このプレプリントに基づく論文が 2023 年 2 月 9 に *PLOS Digital Health* 誌に掲載された際に, ChatGPT は著者から除外された。本章の参考文献にはプレプリント版の書誌情報を示したが, その末尾にはジャーナル掲載版の DOI も併記した)。その後, ChatGPT は著者になれるか, 謝辞に入れるべきか, という議論がさまざまなジャーナル上で展開されていくことになる。

2023 年 1 月, Thorp (2023) は *Science* 誌において生成 AI の使用を禁止するという強いメッセージを提出した。生成 AI, 機械学習等から作成された文章の使用は本質的に禁止すべきであり, その理由は生成 AI によって出力された文章は厳密な意味でオリジナルなものではないため, 生成 AI がオーサーとなるには不適だからだという。

2023 年 5 月, 国際医学雑誌編集者委員会 (International Committee of Medical Journal Editors, ICMJE) による投稿論文の執筆要件に関する勧告 (ICMJE Recommendation) の Section II, A, 4 に, 投稿時にカバーレター・原稿内に AI 技術の使用を明記すべきだという加筆がされた (この勧告は研究公正に関する教材でオーサーシップに関する議論をするときには広く参照されている)。この勧告で示されているように, 生成 AI がオーサーであるか否かという論点は AI の道具性という観点から検討がなされていく。

ChatGPT を使うと, 英語の翻訳や校正も簡易に実行できる。英語を母国語としない研究者にとって, 英語で論文を執筆することには非常にコストがかかる (Amano et al., 2023)。もし校正機能を無料で使用できるなら英語での論文投稿において非常に心強い。materials & methods での研究手順の説明など, 標準的な言い回しがあるような研究領域においてはなおさらである。Berdejo-Espinola and Amano (2023) は 2023 年 9 月 *Science* 誌にて, AI ツールを利用した校正は非英語母国語研究者にとって英語論文の執筆の際のハードルを低くするために有用であり, 「不平等」を埋めるための使用を認めるべきだという提案を行なった (ただし, 英文校正前後のファイルを提出するなど透明性は確保する必要がある)。この記事のすぐ次には Thorp and Vinson (2023) から「いずれ許容されるにしても, 科学コミュニティ全体での明確な議論を待つべきだ」というエディターらしい見解を示す note が置かれている。なお, 補足すれば, こうした非英語母国語者への配慮は ChatGPT の登場前から存在する。例えば, Taylor & Francis 社は投稿規程の中で Editage 社の校正サービスへ誘導し, 支援の一環として割引も行っている。

ChatGPT が発表されて一年が経過した 2023 年 12 月, *New England Journal of Medicine* の AI に関するシリーズにおいて, Koller et al. (2023) は「著者が内容に完全に責任を持ち, 謝辞に LLM の使用を示すなら問題ない (ただし共著者としては NG)」と

いう見解を示した。ここでの重要なポイントは、著者としての責任というワードである。この点については後述する議論の中でもフォローしたい。

## 2. 2 研究の質保証

ここまでは著者が論文を投稿する際に注意すべき事項に関連した議論を確認した。次に、投稿論文を受け付けるジャーナル側の対応を確認したい。多くの研究者は査読者として投稿論文に向き合う立場にもなるためである。まず、査読プロセスの中で AI に査読を任せるといった可能性は容易に思いつくであろう。また、研究の再現性問題に関するツールとして AI を活用する提案もある (Ball, 2023)。具体的には、AI によって生成された可能性のあるテキストや画像を診断するためのツールの導入である (Springer Nature, 2024)。

一方で、こうした活用方法は生成 AI に入力する情報の機密性という観点からしたら回避すべきという見方もある。Elsevier 誌は、出版社の責任として、投稿された論文の原稿は機密文書として扱い、生成 AI にアップロードしないという方針を示している (Elsevier n.d.)。また、学術ジャーナルではないが米国国立科学財団も、研究費申請課題の情報を生成 AI にアップロードしないよう評価者及び申請者に注意している (National Science Foundation, 2023)。欧州委員会が 2024 年 3 月に発表した「研究における責任ある生成 AI の使用に関するガイドライン」も同様の趣旨のもと、生成 AI の使用についてはある程度の制限と監視の元での積極的な活用が示されている (European Commission, 2024)。

## 2. 3 総括的見解

以上、生成 AI を学術論文の投稿及び質保証に用いる場合の議論を概観した。およそ一年間程度の議論を、学術コミュニティの中で生成 AI に対する考え方が次第に深められてきたと表現することもできなくはない。しかし個々の論点は明確なものであり、生成 AI によって改めて新たな問題が生じたという評価には至らないであろう。それはなぜか。ChatGPT はその名のとおりチャット形式の入力インターフェイスにより使用者のハードルを大きく引き下げたが、そのベースである LLM 技術は以前より存在する。すなわち、本技術それ自体が抱える問題をあらかじめ認識することは容易だったからである。

Van Dis et al. (2023) は 2023 年 2 月という ChatGPT 登場の比較的初期の段階で 2.2 節から 2.3 節の議論よりも総括的な視点からまとめた議論を提示していた。すなわち、LLM は科学研究を加速させる可能性がある一方、研究の質と透明性を低下させ、生身の研究者としての我々の自律性を根本的に変化させてしまう可能性があるという指摘である (彼らは ChatGPT に限定した論点を提示したわけではないため、ここでは LLM とする)。彼らが示した生成 AI の使用に関する 5 つの優先事項の概要を示す (前節まで確認された議論への視座も含まれている)。これらは 2.3 節で示した議論を先取りしたもの (補完的なもの) としても読むことができる。すなわち、(1) 人間による検証を維持すること、(2)

説明責任 (accountability) のためのルールを構築すること (LLM の不正な使用の検出を主眼とすると検出と検出回避の連鎖になってしまうため、研究コミュニティと出版社は LLM の公正で誠実な使用法を提示すべき)、(3) 真にオープンな LLM を開発すること (LLM のソースやトレーニングセットは公開されていない、学術界のオープンサイエンス、透明性の追求という方向性と両立しない。オープンソースの AI 技術を開発すべき)、(4) AI による恩恵を容認すべきであること (生成 AI は学習した内容から統計的に関連する語句を並べるだけで文章の意味を理解していないと考える人は多い。しかし将来、AI が自律的に創造的活動を実現する可能性はあるため、科学コミュニティは研究者にとって本質的なスキルとは何かを検討する必要がある)、(5) 議論の場を拡大すること (様々な研究分野で LLM を実際に使用した上で議論を行い、その議論を継続する。例えば、LLM が科学における非英語母国語者の困難を支援するといった一面はあるが、その一方で、機能等に制限のない有償の LLM を使うことのできる「裕福な」研究者がその恩恵により預かることができ、格差を拡大するという懸念もありうる)。以上の 5 つの事項のうち、最後に示した LLM による格差といった点は 3 節でも扱う。

さて、ここまでは主にテキスト生成に関する LLM を念頭に置いたが、その他のタイプの生成 AI に関する問題事例も参考のため提示したい。Guo et al. (2024) は Midjourney という画像生成 AI で作成されたイラストを論文で使用した。しかしそれには科学的に不正確な表現が含まれており、編集委員の検討の結果 (ラットの身体描写について過剰なデフォルメを行っており、読者に誤った印象を与える懸念)、撤回された。この論文の掲載誌 *Frontiers in Cell and Developmental Biology* (もしくは *Frontiers* 誌シリーズ全体) のポリシーでは、生成 AI で作成された生成物の使用を禁止しないが、その内容について著者自身によるチェックが必要だとされている。その一方で Guo らによる論文の撤回事例は、むしろ著者自身のチェックおよび判断だけでは不十分であることを示している。

### 3. 研究公正の観点から見る生成 AI の導入

#### 3. 1 そもそも研究活動とは

ChatGPT が登場したわずか一年程度の中で、大学や学術誌などは機敏に対応を発表した。このような経緯と議論を踏まえ、研究公正に沿った研究の実践において生成 AI をどのように位置づけていくべきかを考えることは、今後の責任ある学術活動の成果発信に関する展望を描くために重要であろう。

まず、研究活動とはどのような活動なのかを改めて概念分析しておきたい。研究者はさまざまな活動を行う。いわゆる人文系の研究者は論文や文献を読みとき、独自の解釈や学説、見解を導き出す。社会科学系の研究者は、フィールドワークやアンケート、インタビューを基にしたデータを定量的・定性的に分析して、何らかの仮説を構築したり、仮説を検証したりし、独自の主張に結びつけていく。自然科学系の研究者であれば、実験や観察、

シミュレーションなどを行い、仮説の構築・検証・反証などを行うであろう。もちろん、こうした分野による分類にはステレオタイプな見方も反映されており、昨今は現実の研究者の研究活動を研究分野という枠で括ることは難しくなっていることは言うまでもない。例えば統計分析など、一つの研究手法を武器に、学際的な活動を行う研究者もいる。

いずれにせよ、研究活動を通じて得られた成果について、指導教員や自分の同僚、共同研究メンバーらと意見交換、学会発表時の質疑や議論、さらには論文投稿時の査読コメントから修正のための着想を得て、研究者は最終的に口頭発表や投稿論文や著作を発表していく。ここで着目したいのは、投稿までに受ける著者自身以外からの多様なフィードバックの扱いである。自然科学や工学系、医学系のようにラボ単位・複数研究者での共同研究を実施する研究領域では、相互の意見交換により論文を構築し、適宜、共著者となる。一方で、人文系の中には、論文の加筆修正にどれだけ関わっても共著者として名前が載ることがない分野もある（例えば、指導教員が大学院生と一緒に投稿論文の構成を立案し、ほぼ書き直しレベルで原稿に赤を入れても、教育の一環であるため名前は出ない、など）。出版までに意見交換を行った同僚や研究仲間の名前を謝辞や「あとがき」で示すこともあるだろうが必須ではなく、そうしたことに同意の上での関与であれば、著者と協力者間でオーサiershipの問題が生じることはない。しかしながら、研究公正の教科書的な観点からは、本来なら著者として名を連ねるべき人間を共著者としない場合、ゴーストオーサiershipと呼ばれる問題事例になる。人文系の「隠れた貢献者」文化はこれに相当しうるものである。学際的な共同研究の機会も多い昨今、異なる文化間で生じるトラブルの回避という予防的な観点から、当該分野で認められているとしても懸念がないとはいいがたい。

### 3. 2 責任主体としての研究者

こうした隠れた貢献者という観点は、研究環境の格差の問題にも波及する。研究活動の中で何らかの支援サービスを使うことは通常認められており、英文校正をジャーナルの規程レベルで推奨する例は先述した通りである（その際には使用した旨を謝辞に記載する）。生成 AI はこれらサービスの延長にあるものと位置付けることができるだろう。しかしながら、これらサービスを使用するために必要な費用を支払うための研究費があるか（生成 AI の場合はより高度な機能を使用できる有料版にアップデートするか）、周囲に研究の相談をできる同僚はいるか、指導教員は学生指導に十分な時間を割く余裕があるか、さらには使用可能な解析ソフトウェアや機器を機関（もしくはラボ）で所有・契約しているかといった点で、研究者の研究環境はそもそも平等ではない。過去には計算をすることそれ自体にもコストがかかった。今ほど高速なコンピュータがなかった数十年前なら、スーパーコンピュータをどれくらい使用できる環境にあるか、もっと昔には、手回し計算機を回してくれるスタッフを雇用できるかという話である。前述した Van Dis et al. (2023) は生成 AI が用途によっては研究者間に存在する格差の是正にも拡大にもつながると指摘した。

しかし研究環境の格差は織り込み済みとし、純粹に成果物のみを評価するのを当然としていたこれまでの慣習からすれば、生成 AI の道具的使用それ自体は回避すべき事項ではない（資金に余裕がなければ無いりの工夫やテーマ設定を行うのみである）。

現状、生成 AI を使用した場合には謝辞にそのことを記載するというルールが標準的であろう。しかしながら問題はまだ残っている。例えば、謝辞記載を怠ること（もしくは意図的に隠すこと）で、研究者の研究力が過剰に示されてしまうという懸念があるため、生成 AI の使用を禁止すべきだという意見はありうる。しかし、この意見は生成 AI ではなく、研究の実施過程それ自体の透明性の確保という観点から吟味すべきである。上述した隠れた支援者の存在（ゴーストオーサiershipや研究領域ごとの慣習の問題）を考えたら、単に生成 AI だからという理由だけでアカデミアから除外されるのは筋が通らない。むしろ隠れた支援者という慣習を見直すべきという主張を顕在化させる意味で、生成 AI の登場は研究公正の「伝統的な」問題を再燃させると位置付けることも可能である。結局のところ、生成 AI はその多様な利活用法や有能さによって懸念面が過剰に演出されているだけであり、生成 AI は数多ある研究支援ツールの一環でしかない。もちろん、サービスの提供前に適切な使用法を確立するための議論をすべきであるが、新興技術の ELSI 検討という意味では後追いとなるケースもやむを得ない。



図 1 研究活動の多様な形態と成果物の発表に際した判断と責任

また、道具としての側面を強調することは、生成 AI と研究者が根本的に違うことを認めることでもある（図 1）。研究者は道具の使用者であり、生成 AI は使用される側である。ここではさまざまな形で得られた出力結果に対して、個々の何らかの判断をすることのできる主体であるか、さらにはその判断に責任を負うことのできる主体であるかどうかが一



つのポイントとなろう。研究者はそういった判断が可能だが、(現状では)生成 AI には不可能である。さらには論文や出版物の内容に関する責任を負うのは研究者であって、ツールは責任を負うことはできない。論文等の成果物に研究不正が含まれていた場合、その行為に関連していた生成 AI が罰則や措置を受けることになるだろうか。この場合に可能になるのは、せいぜい、当該不正行為を行った研究者による生成 AI の「使用」を一定期間禁止するといった措置であろう。AI 自体への罰則は想定できない。

しかし上記のような想定ができないのは、研究活動に関する現在の環境(文脈)というバイアスが関係している。生成 AI との向き合い方についての原理的・根源的な検討を行うための呼び水として、科学哲学における人工知能の議論を参照したい。鈴木(2023)は道具としての人工知能と自律的な主体としての人工知能とを区別し、道具としての AI の社会実装やその際の人間観という、本章よりも広い文脈で検討している。より将来的な発展を見据えた検討として、出口(2024)は生成 AI によるテキスト探索は表層的なものでしかないという経験を踏まえた上で、人が生成 AI を道具として利用するという「主人-奴隷モデル」ではなく、人と AI が対等な「共冒険者」として共同していくこと、そして両者の相互参照的営為によって表層的ではない「人類知の豊饒化」が成し遂げられる可能性を提案している。この相互参照的営為という観点には、生成 AI の将来的な発展だけでなく、それを活用する人間側の発展も視座に含めることができる。見落とされがちなことではあるが、人間自体も周囲の環境に応じてソフト的にもハード的にも変化するのである。

こうした議論を研究公正の文脈に置き換えれば、生成 AI は研究公正の既存の問題を再燃させるだけといった程度では留まらない未来が見えてくる。すなわち、査読プロセスへの活用、投稿論文の再現性のチェック、翻訳校正への利用など、AI があくまで道具としてのみ使用されることを念頭に置くことができるのは、アカデミアにおける成果発表の媒体のあり方や、研究者の業績評価が現在のままであり続ける限りにおいてである。しかしながら、AI の浸透により研究活動における現行の制度や慣習の枠組みが変革される可能性はある(呉羽・久木田, 2020)。そうなれば、研究者(人)と AI の関係やそれぞれの役割、責任も見直す必要があるだろう。こうした検討は ELSI/RRI の方向性に準じて、学界にとどまらない社会的文脈で検討すべき事項である。

#### 4. おわりに

本章では主に 2023 年から 2024 年初頭にかけて学界が検討した、生成 AI との向き合い方に関する議論を概観した。冒頭で提示した「生成 AI によって何が変わって、何が変わらないのか」といった問いかけに対して、生成 AI は研究者の活動を支援するツールという位置付けで理解することができ、根本的な変革が生じているとは言えないと応えることができるであろう。もちろん、それは永続的な風を意味するものではない。AI それ自体の発展と共に、その自律性や帰属可能な責任が見直される余地が多分にあることは念頭に

置かれるべきであろう。なお、そうした近未来を具体的に検討したものとして、本書にも岡林論文（本書第5章）がある。本章の内容を生命系実験科学分野からボトムアップな形で補完し、さらには本章が検討しきれなかった（思いつきもしなかった）側面に光を照らす充実した論考としてお読みいただけるだろう。

## 【参考文献】

- 呉羽真・久木田水生（2020）「AI と科学研究」，稲葉振一郎他（編）『人工知能と人間・社会』勁草書房所収
- 鈴木貴之編（2023）『人工知能とどうつきあうか』勁草書房
- 出口康夫（2024）「共冒険者としての AI」，2023 年度全国公正研究推進会議人文学・社会科学系分科会（開催概要），（一財）公正研究推進協会（最終閲覧日 2024 年 10 月 25 日）  
[https://www.aprin.or.jp/wp/wp-content/uploads/2024/03/bunkakai\\_jinbun.pdf](https://www.aprin.or.jp/wp/wp-content/uploads/2024/03/bunkakai_jinbun.pdf)
- Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., González-Trujillo, J. D., Montaña-Centellas, F., Paudel, K., White, R. L., & Veríssimo, D. (2023). The manifold costs of being a non-native English speaker in science. *PLoS Biology*, 21(7), e3002184. <https://doi.org/10.1371/journal.pbio.3002184>
- Ball, P., (2023). Is AI leading to a reproducibility crisis in science?. *Nature* 624, 22-25.  
<https://www.nature.com/articles/d41586-023-03817-6>
- Berdejo-Espinola, V., & Amano, T. (2023). AI tools can improve equity in science. *Science*, 379 (6636), 991. <https://doi.org/10.1126/science.adg9714>
- Elsevier (n. d.) Elsevier Policies/Policies - Publishing ethics. (accessed Oct 25, 2024)  
<https://www.elsevier.com/about/policies-and-standards/publishing-ethics>
- European Commission. (2024). Living guidelines on the responsible use of generative AI in research. *The European Research Area Forum*, 20 March, 2024.  
[https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc\\_en?filename=ec\\_rtd\\_ai-guidelines.pdf](https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf)
- Guo, X., Dong, L. and Hao, D. (2024) Cellular functions of spermatogonial stem cells in relation to JAK/STAT signaling pathway. (2024). *Frontiers in Cell and Developmental Biology*, 12. <https://doi.org/10.3389/fcell.2024.1386861> Retracted.
- ICMJE Recommendations. (2024). International Committee of Medical Journal Editors. (accessed Oct 25, 2024) <https://www.icmje.org/recommendations/>
- Koller, D., Beam, A., Manrai, A., Ashley, E., Liu, X., Gichoya, J., Holmes, C., Zou, J., Dagan, N., Wong, T. Y., Blumenthal, D., & Kohane, I. (2023). Why we support and encourage the use of large language models in NEJM AI submissions. *NEJM AI*, 1(1). <https://doi.org/10.1056/aie2300128>

- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2022). Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models. *medRxiv* (Cold Spring Harbor Laboratory). <https://doi.org/10.1101/2022.12.19.22283643>  
Now published in *PLOS Digital Health* doi: 10.1371/journal.pdig.0000198
- National Science Foundation. (2023). Notice to research community: Use of generative artificial intelligence technology in the NSF merit review process. Dec, 2023. (accessed Oct 25, 2024) <https://new.nsf.gov/news/notice-to-the-research-community-on-ai>
- Nature. (2023). ChatGPT generates fake data set to support scientific hypothesis. *Nature*. 2023 Nov;623(7989):895-896. doi: 10.1038/d41586-023-03635-w.
- Springer Nature. (2024). Springer Nature unveils two new AI tools to protect research integrity. (accessed Oct 25, 2024) <https://group.springernature.com/gp/group/media/press-releases/new-research-integrity-tools-using-ai/27200740>, 2024.6.12
- Taloni A, Scordia V, Giannaccare G. (2023). Large Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research. *JAMA Ophthalmol*. 2023;141(12):1174–1175. doi:10.1001/jamaophthalmol.2023.5162
- Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science*, 379(6630), 313. <https://doi.org/10.1126/science.adg7879>
- Thorp, H. H., & Vinson, V. (2023). Editor's note. *Science*, 379(6636), 991. <https://doi.org/10.1126/science.adh3689>
- Van Dis, E. a. M., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, 614(7947), 224–226. <https://doi.org/10.1038/d41586-023-00288-7>





## 第2章 技術の発展と研究公正

### —法哲学の観点から—

大屋雄裕

(慶應義塾大学法学部)

#### 1. じこはおこるさ

ウィリアム・ハスキソンから話を始めよう。19世紀初頭のイギリスでトーリー党内自由主義派の指導者として知られ、いくつかの内閣で閣僚も務めた政治家としてその肖像画がロイヤル・ポートレート・ギャラリーにも収蔵されている人物だが、残念なことに現在彼をその側面で記憶している人はほとんどいないだろう。もう一つの側面、すなわち世界初の鉄道事故死者としての存在の方が際立ってしまったからだ。

1830年、最初の実用的鉄道路線と評されるリヴァプール・アンド・マンチェスター鉄道（L&MR; Liverpool and Manchester Railway）が開業した。9月15日に開かれた開業式典に招かれたハスキソンは、途中の休憩地点で列車が一時停止した際にやはり招かれていた当時の首相・ウェリントン公爵に挨拶するために客車を降りて線路上に立っていたところ、おそらくはその速度を見誤ったために近付いてきた対向列車にはねられてしまい、急遽搬送される途上で命を落としたのである。鉄道の歴史はまさに血塗られた誕生を遂げたということになるだろう。イギリスで創られた幼児向け人形アニメ「きかんしゃトーマス」は「じこはおこるさ」（Accidents Happen）という極めてアナーキーなタイトルの主題歌で幕を開けるのだが、それはまさに鉄道技術の実態を示していたと言えるかもしれない。

このように、新たな技術の誕生は同時に新たなリスクの発生を意味している。リスクがあればそれに対する社会的対応が求められ、それはしばしば法的な規制——禁止や制限、あるいは免許制度の導入など——へとつながることになるだろう。だがこの法的規制という手法が特に近年の先端技術とは相性の良くない側面を持っていることについては注意を促す必要がある。

#### 2. 法という規制手法

##### 2. 1 事後規制としての法

法とはなんだろうか？ この問いを主に扱う法概念論は法哲学者が法の誕生以来約2千5百年にわたって取り組んできたものだが、有力な立場の一つは「法命令説」、すなわち法の基本的な性質は制裁の威嚇を伴う命令だというものである。たとえば刑法（明治40年

法律 45 号) 199 条は「人を殺した者は、死刑又は無期若しくは五年以上の懲役に処する」と定めており、このような制裁を回避するためには殺人行為を避けよと我々に命令していると考えられるだろう。現在ではこのようなモデルは単純すぎて法の全体を捉え切れていないという評価が通説であるが、しかし法の根幹的な意義がこのような命令にあるというモデルはゆらいでいない<sup>4</sup>。

その際のポイントは、制裁というものは何らかの行為の後でしか加えようがないという点にある。法律は本質的に手遅れでしかないという性質を帯びており、何らかの望ましくない行為がまず起きてしまうことそれ自体を止めるような能力は持っていないのである。法の本質はこのような事後規制にある。

## 2. 2 事前公開の要求

ただし現代の我々の法システムは、古代中国において唱えられていた法家思想とは異なり、国家が法に定められたルールに従って行動することだけではなく、事後規制を指示する法を人民に対して事前に公開することを要求している。ドイツではこれを法治国原理 (Rechtsstaatliches Prinzip) と呼び、我が国もこれに準拠していると言ってよいが、その理由としては第一に人民に対する「不意打ち」を避けることが挙げられる。事前公開を義務付けることにより、ある行為が行なわれたあとで実はその行為が禁じられていたということを明かして処罰するようなことは不可能になるだろう。逆に人民の観点からすれば、公開されている法の内容のみを遵守すれば処罰を避けることができるという安心感を抱くことができる。法の公開は、人民の自由を保障するためにあると考えることができる。

第二に、あらかじめ公開して知り得る状態にあった以上、にもかかわらず法を知らなかったという事実は罪を免れる理由にならないと定めることで、人民に法を知るコストを負担させることもその機能である。「法の不知はこれを赦さず」という法諺があり、現行刑法でも 38 条 2 項本文 (法律を知らなかったとしても、そのことによって、罪を犯す意思がなかったとすることはできない。) がそれに対応したものだと考えることができるだろう。

だが実際には、これらよりも重要な理由があると考えられる。それは法が、その制裁が実現することを希望していないという問題である。前述のように、刑法 199 条が一定の制裁によって我々を威嚇していると考えよう。将来に対する予期能力を持ち、みずからにとって不利な結果が実現することを避けようとする傾向がある自律的な主体である我々は、その威嚇を受けて、よほどのことがない限りは殺人を犯すことを回避しようとする行動を変容させるだろう。結果的にこの社会における殺人の件数は減少し、それに対応する処罰は (理想的な条件のもとでは) ゼロになると期待されるし、それこそが法の狙うところなのである。法による制裁は、それが予告されることによって実現しなくなることを期待するという逆説的な機能を果たしている。このように事後規制の予告は、制裁対象となる行為から我々をあらかじめ遠ざけるという事前規制へと我々の予期能力を通じて転

化することになるだろう<sup>2</sup>。

ただしそのためには、第一に我々がそこで想定されていたような自律的主体であり、第二に「ああすればこうなる」という推論を行なうための理性や制裁が科される条件についての情報を伝達するための言語を共有しており、さらに言えば苦痛や死を避けることが望ましいというような一定の価値観が一致していることが求められるだろう。「進者往生極楽／退者無間地獄」（進めば極楽・退けば地獄）の旗を掲げていた一向門徒を死の威嚇によって阻止することは可能だっただろうか？　ここではひとまず、この転化が自動的・必然的に起きるわけではなく一定の条件が必要になるものだとことを確認しておこう。

## 2. 3 リスク配分における転化

リスク配分についても、事後規制の事前規制への転化と同様のメカニズムを想定することができる。一定の事故が発生した場合、その帰結としての損害はひとまず被害者に対して発生することになる。これがリスク配分に関するデフォルト値だと考えることができるだろう。しかし発生した被害が放置されるこのような状態（泣き寝入り）を我々は適切だと思えないので、通常は加害者から被害者への財の移転を命じることで社会的な損害の再配分を行なうという、事後の救済システムが用意されている。その典型は過失・故意の存在を前提とする不法行為に対する損害賠償（民法（明治 29 年法律 89 号）709 条）だが、無過失の場合も含めて損害を社会的に救済する補償制度<sup>3</sup>、強制保険<sup>4</sup>も同様の機能を備えていると考えることができる。

ところで我々が予期的な行為者であるとすれば、一定の行為からそのような事後救済への責任が発生すると予想すればその行為を回避しようとするだろう。自動車事故については運転者が責任を負うことが予想されるため、事故を起こさないよう運転にあたっては十分に注意するだろう。他方、鉄道については鉄道営業法（明治 33 年法律 65 号）37 条により鉄道地内への立入りが禁止されており、その結果として事故が起きた場合には運転士ではなく立ち入ったものの責任が第一に問われることになっている。鉄製のレールと車輪によって走行している鉄道は走行中の抵抗が極めて小さく、ブレーキをかけても急停止することができないため、運転側に注意義務を課しても事故を回避できる可能性はほとんどない。このような技術的な性質を踏まえ、むしろ歩行者や自動車の方に注意義務を課すため、このような責任の配分が行なわれていると考えることができるだろう。このように法は、事後の責任配分を予告しておくことによって人々の予期を形成し、事故の可能性に対する注意・回避義務を事前に配分していると理解することができる。

科学技術のガバナンスについて言えば、このように一定の守るべき条件を客観的リストとして提示し、その遵守を確認することでリスクの発生を回避しようとしてきたのが研究倫理における「チェックリスト主義」、すなわち客観的な義務のリストを満たしていることを各自が確認することを求めるようなガバナンスの手法だと考えることができる。

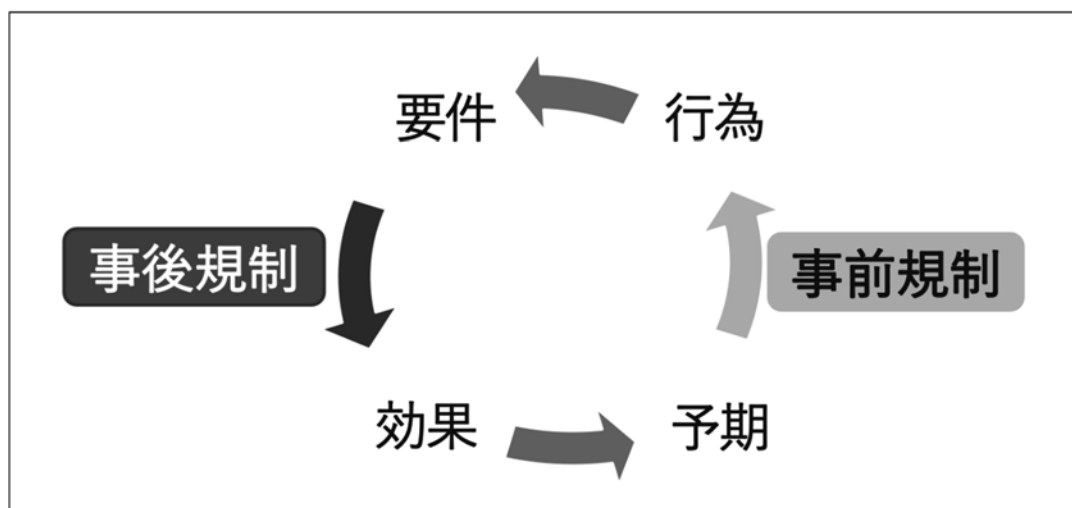


図1 法による規制の特徴

法による規制の特徴は、図1の左側に示されている要件と効果の結合、すなわち一定の要件が満たされた場合にはそれに相当する効果が生じることを予告する事後規制のみが文章として提示されているにもかかわらず、実際には右側に示している予期と行為の結合、すなわち効果の発生を予期しそれを（制裁の場合には）回避するように行為を変容させるという事前規制が機能することを期待しているという点に求められる。法文に書かれていない事前規制効果こそが、法律の期待するものなのである。

### 3. AI と法の危機

このようなサイクルが法の機能を示していると考えた場合、これが十分に働くためには二つの要素が求められると考えることができる。第一は行為の結果を予測することができるという予見可能性であり、第二は不幸な結果が予測された場合にそれを回避することができるという意味における回避可能性である。そしてこの二要素が現代の技術、特に生成系 AI（Generative AI）を含むモダン AI<sup>5</sup>においては失われている点に注意する必要があるだろう。

#### 3. 1 予見可能性

予見可能性から考えよう。典型的には深層学習を用いるモダン AI においては、基本的にはニューラル・ネットワークという生物の神経系を模倣した仕組みを採用し、判断において注目する特徴量や判定の閾値について適切と考えられる学習データに合致するように AI システム自体が変容させていくような学習過程が採用されている。2023 年以降急速に発展し社会的な議論を呼んだ生成系 AI についても、このように学習に焦点をあてるなら

ばモダン AI の一種であり、そこに本質的な違いはないということになるだろう。そして重要なのは、判断のプロセスを管理するオールドファッションド AI の考え方ではなく結果を管理するモダン AI においては、その故に推論過程の形式化や閾値の明確な設定ができない領域を取り扱えるようになった一方で、判断過程の相対的なブラックボックス化が生じたという点にある<sup>6</sup>。

象徴的なエピソードを紹介しよう。Google の開発した囲碁 AI である AlphaGo が 2016 年に韓国の李世<sup>イ・セドル</sup>乜、2017 年には中国の柯潔というトップクラスの棋士に勝利したことは大きな注目を浴びた。この際、特に李世乜との対局の中盤に AlphaGo が定石を外れほとんどの棋士には理解できない手を打ったこと、しかし結果は AlphaGo の勝利に終わり、問題手が有効に機能していたことが事後に判明した点は、人間には理解できないがおそらくは「正解」であろうものを AI が生み出した事態として指摘されてきた。

しかしおそらく更に問題なのは、その後に Google が作り出した AlphaGo Zero というシステムである。本来の AlphaGo は従来知られていた定石や棋譜といった人類の知的遺産を学習させることで強化を試みたのだが、AlphaGo Zero はそれらを一切用いずに囲碁のルールのみをプログラムし、その後は自分相手の対局を繰り返すことによって学習させたところ、従前の AlphaGo より早く強くなってしまったというのである。おそらくこれまでの人類は囲碁というゲームを十分には正しくプレイできておらず、我々に知られていなかった知的成果がそこで発見されたということになるのではないだろうか。

このようなシステムを念頭に置いたならば、そこでは AI システムがなにを生み出すかという問題に対する開発者・利用者の予測可能性が低下しており、法が機能するために必要な予見可能性の実効性も大きく低下していると言えるだろう。

### 3. 2 回避可能性

回避可能性についてはどうだろうか。問題は、仮にモダン AI のように開発者・利用者の指示が遵守されるタイプの AI であってもそれらが相互に接続されネットワーク化されることにより、問題動作を阻止したり操作できる可能性が低下しているという点にある。

2010 年 5 月 6 日、ニューヨーク株式市場において、S&P500・ダウ=ジョーンズ指標などの株価指数がわずか数分のあいだに約 9% 暴落し、その後約 30 分かけてほぼ回復するという急激な変動が発生した（フラッシュ・クラッシュ flash crash と呼ばれる）。9 月 30 日に発表された SEC/CFTC の報告書はその理由の一部を、自動取引システムを利用することによりミリ秒単位での売買を繰り返す高頻度取引（High-Frequency Trading）に求めている（SEC/CFTC, 2010）。

経済学者ケインズが『雇用・利子および貨幣の一般理論』のなかで、株式市場における投資家の行動を美人投票——「100 枚の写真の中から最も美人だと思う人に投票してもらい、最も投票が多かった人に投票した人達に賞品を与える新聞投票」にたとえたことを思

い出そう (Keynes, 1936)。自分がもっとも美しいと思う人ではなく多くの人が美しいと思う人がそこでは勝利するのであり、したがって投資家も自分の評価ではなく他の多くの投資家の評価を模倣するようになるだろう。株価についても同様に、多くの投資家が肯定的に評価する＝買うことで上がるのだから、その傾向をいち早く探知し模倣することで自分も値上がりの利益を得ることができるだろう。

だが問題はこのように他者の行動を模倣するアルゴリズムが組み込まれた自動売買システム同士がネットワーク化された状況で、そのうち1台が何らかの事情で誤動作を起こしたらどうなるかという点にある。誤って発生した動作を他のシステムが模倣し、それを探知した別のシステムが模倣するといった連鎖が発生することにより、誤動作が自己増幅したパニックへと成長することになるだろう。しかもかつてのように売買注文を人間が媒介していた時代と異なり、これらのシステムの相互動作はミリ秒の世界で発生している。そこで生じた異常に人間が気付き一定の阻止行動を取ろうとする前に、事態はすでに破局的な段階に至ってしまうのではないだろうか。自動化されたシステムたちの世界の前で、回避可能性の実効性もまた大きく低下していると考えることができる。

### 3. 3 現代科学技術の課題

問題はつまり、科学技術の複雑化・高速化が生み出す課題の状況がこれまでとは大きく異なっている点にある。かつてそのガバナンスが問題となった大規模科学技術としては、たとえば原子力発電が挙げられるだろう。しかし原子力技術は大規模・複雑・専門的だが全体として決定論的なプロセスであり、どのような放射性物質をどの程度の量置けばその結果がどうなるかについては十分に予測可能である。そのもとで規制の課題となったのは技術を理解するための専門知識・能力が規制される事業者側に遍在しており、規制する側である議会や行政庁とのあいだに非対称性があることであった。このために被規制側に協力を求めることによってその知識を吸い上げるような行政手法が広く採用され、それは同時に両者のあいだに馴れ合いや癒着を生じさせたと批判されることになったのである<sup>7</sup>。

これに対し機械学習以降のモダン AI 技術は確率論的なプロセスであり、そこでなにが起きたかを分析することは現実的にほぼ不可能だと言ってよい<sup>8</sup>。そのような状況において、規制側・被規制側を問わずどのような問題が起きるかを事前に予測することそれ自体が困難になっているのである。

このような状態で、たとえば AI システムによる人種差別を防ぐためにはどのようなことが可能なのだろうか。従来の処方箋は、その基礎となる人種情報をデータベースから排除することであった。就職のための履歴書から顔写真や人種欄を排除するような取り組みが、その一例と考えられる。しかし、血液型や星座に基づいた採用選考のように不合理な差別の場合にはその根拠情報を削除することで十分な対策になるだろうが、異なる取扱いをすること自体に理論的な正しさが存在する合理的な差別に対して対抗することはできな



いと考えられる。

アメリカ社会を例にとろう。黒人に属する個人の教育水準が統計的に低く、平均所得が低く、ローンを供与した場合の滞納率が高いことは基本的に事実であり、このような対象に対して与信を拒否することは理論的に、少なくとも経営の効率化というそれ自体適切と考えられる観点からすれば合理的である。しかし我々は他方でこのような実態が努力不足など個々人に帰責することのできる要因だけではなく、むしろ歴史的に形成され累積されてきた不当な取扱いの帰結だと考えており、そのため仮に合理的だろうがそのような事実に基づいて異なる取扱いをすることを規範的に禁じているのである。

このような場合、仮に人種に関する情報を基礎となるデータベースから排除しても、それに替わる代理変数、すなわち郵便番号やファーストネームなど人種と関連する情報が見つかってしまうことが予測される<sup>9</sup>。このように、我々人間にとってまだ知られていない代理変数が AI によって発見され活用されてしまう危険性に対抗するためには、センシティブな情報をあらかじめ除去しておくという事前規制ではなく、判断結果とセンシティブ情報のあいだに相関がないことを確認するという事後規制の手法によるしかないと考えられる（大屋・東藤, 2023）。そしてそこでの問題は、事後に相関がないことを検証するためには当該センシティブ情報がデータベース内に存在する必要があるのも、それを除去するという事前規制の手法と根本的に対立するという点にある（大屋, 2018）。我々はここで、従来型の事前規制と新たな事後規制という両立しないガバナンス手法のあいだでの選択を迫られているのである。

#### 4. 事前規制の副作用

ここまで述べてきたように法という手段による規制が難しくなりつつあるにもかかわらずそれを強行するとすればどのような問題が起きると考えられるだろうか。以下ではそれを、①技術進歩の抑制、②萎縮効果（Chilling Effect）、③実効性、④逃避可能性の4点に分けて議論することにしよう。

##### 4. 1 技術進歩の抑制

第一に、見通せないリスクの存在はしばしば、何があっても大丈夫な規制の導入というゼロリスク実現への要求を引き起こす。たとえば 1895 年に日本初の路面電車が京都で開業した際に導入された「告知人」制度、すなわち先導役が電車の前を走りながら通行人に警告するシステムをその一例と考えることができるかもしれない<sup>10</sup>。しかし当然のことではあるが、そのような規制を我々が現在に至るまで堅持していたとすれば新幹線をはじめとする鉄道技術の進歩もそれがもたらす社会的恵沢も実現しなかっただろう。

また、規制を遵守するために必要となる直接的コストだけが問題になるわけではない。法を事後規制として機能させるためには、違反行為を発見しそれに制裁を加えるための制



度——典型的には警察等の捜査機関，裁判所を中心とする刑事司法制度，そして囚人を閉じ込めて管理するための監獄制度が挙げられる——が必要になり，それを維持するためには一定の行政コストが必要になる。社会を監視し，必要であれば制裁を加えるために必要となる取締りコストについても，十分に視野に入れておく必要があるだろう。

#### 4. 2 萎縮効果

第二に，どのような場合に発動されるか不明確なのだが強い制裁が科されるような規制が導入された場合，万が一にもその制裁が発動されることを避けて人々は過度に防衛的な対応を取ることになるだろう（萎縮効果）。憲法学においては特に，「品位ある言動」のように曖昧で解釈の余地が大きい文言で言論・表現を規制することから生じる萎縮効果の問題が注目されてきた。たとえば男性から女性に向けられた場合には問題にならないような愛情表現について，逆に女性が男性に向けて発した場合に「品がない」「ふしだら」と評されてきたような事例を想起すればそれは，多数派が「自然」だと考える基準によって少数派の言動が裁断され，処罰される根拠として機能する可能性があるだろうし，それを予期すれば少数派の言動はその危険を意識して回避するようなもの，萎縮したものにならざるを得ない。さらにそれはしばしば，制度を導入した側では問題のない行為と想定していた領域にも及び，意図を超えた過度の抑制が実現することになってしまうというのである。

これについても，それが技術開発それ自体を抑制する効果のみならず，その領域に対する投資自体を抑制することによって将来的な可能性をも消滅させる危険性があることに留意しておくべきだろう。

#### 4. 3 実効性

第三に，自律的な主体は規制を予期したら制裁をもたらす行為を自発的に回避するだろうとここまで我々は前提してきたが，それが常に正しいわけではないという点には注意する必要がある。たとえばかつての日本では，未成年者飲酒禁止法（当時，大正 11 年法律 20 号）<sup>11</sup>によって 20 歳未満のものの飲酒が禁じられているという規範的な状態があったが，他方で大学生は飲酒しても構わないという広汎な社会的合意があり，入学後の合宿などで当たり前のようにビールが提供されるという実態があった。その状況で，にもかかわらず規範を遵守して飲酒を拒否するという対応が現実的に選択可能だっただろうか。過去の自分の行為を正当化しようとしているわけではないのだが，新たな人間関係に入っていくというタイミング，そこで妙に目立つような行動を取りたくないというプレッシャー，まずビールが注がれてしまうといった事実から生じる同調圧力などを考慮すればそれは疑わしいだろう。違法行為を追求される危険性よりそうしないことから予想される社会的ダメージが大きければ，自律的な個人が法に反する行為を選ぶ可能性は十分あると，我々は言わなければならない。

同様に、法に反する行為から得られるメリットが十分大きい場合も考えられる。詐欺で捕まって 5 年の実刑を科されるとしても 10 億円の利益を隠匿することができれば年収 2 億円の仕事に成功したのと同じだという思考を、プロの詐欺師たちはすると言われている。たとえ結果的に死刑に処されることを覚悟しても憎い相手に復讐するということもあるだろう。すでに指摘したとおり法律が予告する制裁は常に事後に加えられるものであり、その可能性を引き受けてもなお法に違反することを選ぶ自由は物理的には必ず存在しているのである（「犯罪を犯す自由」）。このような事態が発生する場合、法規制の実効性は大きく損なわれることになるだろう。

だが逆に、このように法律という規制手段がそれに反する余地を残しているからこそ、たとえばベトナム戦争に反対して徴兵を拒否することで逮捕されボクサーとしてのライセンスを剥奪されたモハメド・アリのように、みずからの良心に反する公的な命令に従わず、そのことによって社会へ異議を申立て民主政に再考を促す行為（市民的不服従 *Civil Disobedience*）が可能になるとも考えることができる<sup>12</sup>。法に反する余地それ自体には、民主政の健全性を保障する最後の手段としても重要な価値がある。しかしその故に、軽すぎる制裁、発見されない行為、尊厳が踏みにじられたケースのように人々の法に反する行動が何らかの意味で合理的になる場合には、それがむしろ誘発されることになるだろう。萎縮効果においては不明確で曖昧な規制が問題となったが、実効性の観点からは明確だが不合理な規制が法に対する潜脱を誘発する可能性が問われることになる。

#### 4. 4 逃避可能性

第四に、現在の我々が住んでいる主権国家システム、すなわち世界がそれぞれの主権国家の領域へと分割され、その内部においては当該国家の法律が原則として適用されるような社会においては、それぞれの国家が持つ利害の対立・相違が大きな問題になると考えられる。個々の国家は各自の立場から利益になるように法律を定め強制しており、利害が異なる国家の法執行に積極的に協力したり、他国の利益実現に積極的に貢献したりすることを目的として法を制定することは（例外的な事態を除けば）ないだろう<sup>13</sup>。

そして問題は、特にインターネットを介したサービスについてはどこの国から・どの国に対してもそれを提供することができるという点にある。このような状況である国家が非常に不条理な（と被規制側に感じられる）規制を導入したとすれば、たとえば AI の開発や提供を行なっている企業はその本拠地を当該国家の外へと移転することを選択するだろう。現代のグローバル社会において、自分たちにとって有利で快適な規制を導入している国家へと「逃避」することは当事者にとって常に有利な選択として存在している。すでに被規制側にとって不条理な規制は（市民的不服従などによって）抵抗しなければならない対象ですらなく、単にそこから逃避し無視すればよいものと評価されるようになっていると言ってよい。かつてのように重厚長大産業の設備が特定の土地に固定され、その領域を

支配する国家の影響力を逃れがたかった時代とは異なり、現代においては規制の適切さが厳しく問われる状況にあると考えることができる。

## 5. 新たなガバナンスへ

このような状況で、事前規制的なチェックリスト主義によって現代の先進的な科学技術をガバナンスすることは極めて難しくなっている。そのなかで、にもかかわらず有効な規制のあり方を模索するとすればそれは同時併走的なガバナンスでしかあり得ないのではないかというのが——ある意味においては法律家としての無力感の吐露にすぎないかもしれないが——筆者の暫定的な結論である。部外者が外部から遵守されるべきチェックリストを提供するのではなく、研究を推進している当事者自身がその適切性や一定の価値基準に照らして擁護可能かといったことについてみずから考え取り組んでいくしかないのではないだろうか。

その例として「人間中心の AI 社会原則」（統合イノベーション戦略推進会議, 2019）を考えることもできるだろう。そこでは、「人間の尊厳が尊重される社会（Dignity）」「多様な背景を持つ人々が多様な幸せを追求できる社会（Diversity & Inclusion）」「持続性ある社会（Sustainability）」が基本理念として、それが実現することを我々自身が望む社会の姿として提示しており、それに基づいて AI の開発者・利用者自身が考え、適切に振る舞うことが求められているのである。2023 年 5 月に開かれた G7 広島サミットを受けて展開されている「広島 AI プロセス」も同様であり、そこでは日米欧が AI 規制に関するハーモナイゼーションを進めていくにあたって「民主的価値」を共有していることが確認され、その具体的内容として法の支配、人権、適正手続き、多様性、公平性、無差別、民主主義、人間中心主義などが含まれるということを示すことで、規制の適切性を保つことが試みられている。

また、そのような環境のなかでたとえば筆者のように法律の立場から関与しているものも一定の技術の存在を前提してそれを外在的に裁くのではなく、むしろ協力者・協働者として研究者・開発者と相互対話を通じて知識・経験を共有し、その適切なあり方についてともに考えていくことが求められているだろう。前述の「人間中心の AI 社会原則」などはマルチ・ステークホルダー・プロセス、すなわち研究・開発側だけではなく利用側、さらに法律・経済などに関する有識者を包含した会議体において相互に対話しながら共通の基盤を見いだしていくという営みを通じて起草されたものだが、このようなプロセスがより広く展開される必要があるだろうと考えている。

最後に、特に現在のモダン AI については事前にその動作結果を予測することが難しくなっていることを前提として、事後的に対応できる可能性を残しておくこと——たとえば回復不能な被害を生じさせないような防護措置を事前に講じておくこと、実効性のある被害回復を実現するために撤回可能性（Retrievability）やその対象特定を可能にする追跡可

能性 (Traceability) を確保することなどがガバナンスの重要な手段になるのではないかと  
いうことを指摘しておきたい。

先端科学技術において研究公正を担保するためにはこのように、事前から事後へという  
展開を組み込んだガバナンスの将来像について検討する必要があると考えられる。

- 
- 1 もっとも重要な指摘は、制裁の予告に裏付けられた命令（義務賦課ルール）というモデルでは、婚姻や遺言のように特定の行為を禁止するのではなく一定の方式に従って行為した際にメリットを与えるような法制度を説明できないというものである (Hart, 1961)。
  - 2 法が持つ事後規制としての性質との対比によって、事前規制を本質とするアーキテクチャの権力の性格を明らかにしたのがアメリカの憲法学者ローレンス・レッシングである (Lessig, 1999)。それが我々の予期を通じて事前規制へと転化することについては、筆者がすでに指摘している (大屋, 2007)。
  - 3 予防注射の副反応に対する補償制度がその典型である。予防注射については関係者の過失・故意が存在しなくとも一定の副反応が発生することが避けられない一方、全員が受けることから社会全体にもたらされるメリットが非常に大きいため、被害を社会的に埋め合わせるシステムが導入されたと考えられることができる。
  - 4 自動車損害賠償責任保険が典型であり、加害者が無資力であるために自動車事故の被害者が救済されないという事態を避けるために設けられている。
  - 5 ここでは、第二次 AI ブームまでで扱われていたルールベースの AI (推論・判断の規則を明示的にプログラムする AI) をオールドファッシュョンド AI, 第三次 AI ブームにおいて急速に発展した機械学習技術を用い、顔認証のように判断の過程や根拠が明示できない領域においても活用されるようになった代償としてブラックボックス化が生じている AI をモダン AI と位置付けている。
  - 6 相対的と表現しているのは、一般的に同種の判断を人間が行なっている場合の推論プロセスも形式化できないのであり、ブラックボックス化が AI に固有の問題だというような誤解を避けるためである。
  - 7 2011 年の福島第一原子力発電所事故以降に発足した原子力規制委員会・規制庁においては、この点への反省から事業者との打ち合わせ等を原則としてすべて公開するという方針が採られている。
  - 8 念のために言えば、あくまでプログラムに基づく電子計算機の動作である以上すべての計算経過を記録し事後的に分析することが理論的には可能である。ただしそのためには膨大な記録容量とそれを分析するための計算能力が必要となり、ごく特殊な事例に限った分析についてはともかく、リアルタイムで動作を分析したり点検したりすることは現実的に不可能になっている。
  - 9 実際にアメリカでは黒人が強い集住傾向を有するために郵便番号と相関を持っているこ

とが知られている。また、ファーストネームを黒人らしいものやインド系らしいものから白人らしいものに変えることで、たとえば就職面接に呼ばれるかどうかの結果が異なることが主張されている。

<sup>10</sup> イギリスでも、赤旗を持った先導役が自動車の前を走って警告することが 1865 年から 96 年まで求められていた (The Locomotive Act 1865)。

<sup>11</sup> 2022 年に民法上の成年年齢が 18 歳に引き下げられたことを受け、「二十歳未満の者の飲酒の禁止に関する法律」に改称された。

<sup>12</sup> 前述のローレンス・レッシングは、事前規制を本質とし「従わない可能性」それ自体を存在させないような規制手法としてのアーキテクチャの権力が、法や規範のように伝統的なコントロール手法と大きく異なることに注意を促している (Lessig, 1999)。

<sup>13</sup> もちろん「人類共通の敵」*hostis humani generis* と位置付けられた海賊への対応のように、おおむねすべての主権国家にとって共通の利害が成り立ちそれを反映した法が成立するようなケースも存在しないわけではない。

## 【参考文献】

大屋雄裕 (2007) 『自由とは何か：監視社会と「個人」の消滅』筑摩書房

大屋雄裕 (2018) 「人民の、人民による、人民のための情報：個人情報と自由と範囲」『社会保障研究』3 巻 3 号, 352-364 頁

大屋雄裕, 東藤大樹 (2023) 「AI マッチングにおける二つの公平性：労働法的検討への基礎として」『季刊労働法』282, 106-115 頁

統合イノベーション戦略推進会議 (2019) 「人間中心の AI 社会原則」(最終閲覧日：2024 年 9 月 30 日) <https://www8.cao.go.jp/cstp/ai/aigensoku.pdf>

Hart, H.L.A. (1961). *The Concept of Law*. Oxford University Press. = 長谷部恭男 (訳) (2014) 『法の概念』筑摩書房《原著第 3 版の邦訳》

Keynes, John Maynard. (1936). *The General Theory of Employment, Interest and Money*. Palgrave Macmillan. = 山形浩生 (訳) (2012) 『雇用, 利子, お金の一般理論』講談社

Lessig, Lawrence. (1999). *CODE and Other Laws of Cyberspace*, Basic Books. = 山形浩生・柏木亮二 (訳) (2001) 『CODE：インターネットの合法・違法・プライバシー』翔泳社

SEC/CFTC (U.S. Commodity Futures Trading Commission & U.S. Securities & Exchange Commission). (2010). Findings Regarding the Market Events of May 6, 2010 Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues. Sep. 30, 2010. (最終閲覧日：2024 年 9 月 30 日) <https://www.sec.gov/news/studies/2010/marketevents-report>



## 第3章 産業界における AI 倫理の現状と課題

### —自動運転分野の ELSI の動向を手がかりに—

樋笠 堯士

(多摩大学経営情報学部)

#### 1. AI 倫理の国際的な潮流

2024 年 5 月 21 日（現地時間）、EU において「AI 法（Artificial Intelligence Act）」が成立した（2 年かけて段階的に適用される）。これは、世界初の包括的な AI 規制法で、かつ、域外適用もある。つまり、EU 域内に拠点を持たない日本企業であっても、EU 域内に製品やサービスを提供する事業者は同法の適用対象になる。すなわち、EU 域内に取引先・顧客を持つすべての外国企業も対象である。日本企業ももちろん含まれる。同法は、AI を「禁止 AI」「高リスク AI」「低リスク AI」などに分けた規定となっている。たとえば、自動運転のシステム等も、高リスク AI に当てはまる。高リスク AI に該当した場合には提供者には事前に適合性評価手続を経る義務が生じ、リスク管理システムの構築やデータガバナンス、技術文書などの作成や記録保持、透明性・利用者への情報提供やサーバーセキュリティなどが必要になる（樋笠, 2024）。これらを満たさないと EU に関する場所での取引ができない恐れがある。もちろん、前述の通り、EU 域内に拠点のない企業であっても、AI 規制法については確認して対応を講じる必要がある。

さらに、遡ること 2023 年 12 月 1 日には、G7 によって、高度な AI システムの設計・開発・導入・提供及び利用をカバーする「全ての AI 関係者向けの広島プロセス国際指針」が誕生している（総務省, 2023）。同指針においても、十分な透明性の確保、個人データの保護の必要性、開発だけでなく、利用者を含むすべての AI 関係者が、自身そして必要に応じて他者のデジタル・リテラシーを向上させる機会を求めるべき旨が規定されている。

なお、2024 年 6 月 14 日の G7 会合においては、ローマ教皇までもが参画し、AI の軍事的利用や偽情報の拡散を防ぐため、AI を倫理観に沿って使用するための国際的枠組みやガイドラインの創設へ向けた協議が開始され、国際標準の策定する手続きに移行している。そこでは、「透明性があり、労働者の権利と人権を尊重し、より良いサービスの提供に資する、人間中心の倫理的な AI の重要性（the importance of a human-centric, ethic AI that is transparent and respects workers' and human rights, thus helping deliver better services）」が強調されている（G7, 2024）。

このように国際的には、AI への規制および事業者に対する透明性の担保やリスクの把

握、対応などの義務付けがなされる潮流があるといえる。

かかる潮流に付随するのは、AI をめぐる「倫理」問題である<sup>1</sup>。生成 AI、とりわけ大規模言語モデル（LLM）に関して、利用者の個人情報漏洩、利用者による著作権侵害等の法令違反、権利侵害や、悪用の恐れなども各国の先行研究では指摘がなされている（表 1）。

表 1 LLM のリスク 出典：筆者作成

Laura Weidinger et al. <sup>①</sup>	Kumar, Sachin et al. <sup>②</sup>	Open AI GPT-4 Technical Report <sup>③</sup>	
Discrimination, Hate speech and Exclusion	Discrimination, Toxicity, and Exclusion	Hallucinations	Harmful content
Information Hazards (情報漏洩含む)	Factual Errors, Misinformation, and Disinformation	Harms of representation, allocation, and quality of service	Disinformation and influence operations
Misinformation Harms (ユーザの非倫理的または有害な行動を動機付ける捏造情報を提供)	Privacy Violations	Proliferation of conventional and unconventional weapons	Privacy
Malicious Uses	Other Underexplored Issues (Malicious Usesを含む)	Cybersecurity (refuse malicious cybersecurity requestsを含む)	Potential for risky emergent behaviors
Human-Computer Interaction Harms		Interactions with other systems	Economic impacts
Environmental and Socioeconomic harms		Acceleration	Overreliance

利用者の個人情報漏洩、誤った情報の発信、プライバシー侵害、利用者による著作権侵害等の法令違反、権利侵害や、悪用の恐れなどが想定されている。ユーザー側の「使い方」について踏み込んだアプローチが重要であることが看取される。

このように考えると、産業界一般に言えることは、コンプライアンス、サイバーセキュリティ、ESG/CSR と全く被らない領域として、「AI 倫理対応」の領域が生まれていることである。

各社の AI 活用方針・AI 指針策定とその実行プロセスを明確に、透明性を持たせること、そして、利用者による誤用・悪用・法令違反をいかに防ぐかが重要である。

先述した広島プロセス国際指針においても、「産業界、政府、市民社会、学界を含む、高度な A I システムを開発する組織間での責任ある情報共有とインシデントの報告に向けて取り組む」ことが要求されており、AI を開発・事業で利用する産業界では、「市民社会」に対する情報共有やリスクを含むインシデントを公開していく必要がある<sup>2</sup>。

2. 日本の対応

広島プロセス国際指針の「産業界、政府、市民社会、学界を含む、高度な A I システムを開発する組織間での責任ある情報共有とインシデントの報告に向けて取り組む」では、まず、取り上げられる主体は、産業界である。たとえば、AI のみならず、法学、生命医学、生態学、SDGs、消費者行政など多様性に配慮した様々な分野の専門家から構成される「AI

倫理外部委員会」を設置し、AI 倫理に関するさまざまな議題について、社長、副社長含めての議論を行い、その結果を取締役に共有する企業も現れている<sup>3</sup>。

そして、産業界の次にくる主体は「政府」である。さらに市民社会が続く。内閣府(2019)では、「・・・中略・・・そのため、政府、企業、大学、研究機関、一般の人々等、多様なステークホルダーが協働してルール、制度、標準化、行動規範等のガバナンスについて問題を設定し、影響を評価し、意思決定を行うと共に実装できる体制が整っていることが必要である。また、社会的に声の上げにくい人たちを含む、多様なステークホルダーの声を拾い上げて、常に最先端の社会的、技術的な課題に取り組む体制を構築できていることが求められる。こうしたガバナンスの実現に際しては、法律によるのみならず、技術的手段を含む企業の自主的な取組によるなど、柔軟かつ実効的な方法がとられていることが求められる。また、ガバナンスのための国際的な整合が重要であり、各国におけるガバナンスに加えて、国境を越える問題に対処するための国際協力体制が整っていることが求められる。」と規定されているところ、「一般の人々」との協働が想定されている。ここでいう一般の人々とは、おそらく AI のユーザーも含むであろう。市民社会(ユーザーも含む)と、産業界とをつなぐのは、「政府」である。それゆえ、政府・行政による AI 倫理指針の制定による効果・方向づけは大きいと考えられる。たとえば、日進月歩の勢いで技術開発と社会実装が進む自動運転の世界で、産業界と市民社会をつなぐ、政府の役割はなんであろうか。AI が運転する自動運転車にどのようなリスクがあり、どのような挙動をし、どのような利益・不利益を生むのか、ジレンマ問題(＝トロッコ問題)などの倫理的な問題が多く含まれる分野において、日本では、自動運転の倫理に関する指針は出されていない。それゆえ、外国に眼を向ける。

### 3. 諸外国の自動運転の倫理対応

#### 3. 1 ドイツ

ドイツでは、2017 年に、旧デジタルインフラストラクチャー省が「自動運転及びコネクテッド・カーに関する倫理規則 (BMVI, Ethik-Kommission Automatisiertes und Vernetztes Fahren, Bericht 2017)」を制定している。20 ある項目の中で、たとえば、【第 7】では、「すべての技術的な予防措置でも、避けられないことが明らかである危険な状況において、人命の保護は、法益衡量において最も優先される。それゆえ、それによって人的損害を避けられ得る限りで、技術的に実現可能な範囲内で動物の損害あるいは物の損害を甘受するようにプログラミングされるべきである。」とされ、開発者にとっても、市民にとっても分かりやすい優先順位が示されている。さらに、【第 8】では、「生命対生命のような真のジレンマにおける決定は、関係者の「予測できない」行動様式を含んだ具体的な実際の状況に左右される。それゆえ、かかる決定は、一義的に規範化できず、また、倫理的に疑う余地のないようプログラムすることもできない。技術システムは、事故を避ける



ために設計されなければならない。しかし、道徳的に判断する能力を有する答責的な運転者の決定を置き換えたり、あるいは、それを先取りし得るような、複雑あるいは直感的な事故の評価に向けた規範化はできないのである。人間の運転者が、一人あるいはそれ以上の人間を救うために緊急状況下で一人の人間を殺してしまった場合、たしかに、その運転者は違法に行為したものであろう。しかしながら、必ずしも責任ある行為とはいえない。回顧的に、特別な事情も含めてなされるこのような法的な判断は、容易には、抽象的、一般的な事前判断に置き換えられ得ず、それゆえ、ふさわしいプログラミングにも置き換えることができない。したがって、望ましいのは、独立の公的機関（例えば、自動輸送システムに関わる事故調査連邦事務局、あるいは自動運転およびコネクテッド・トランスポート保全連邦庁）により、体系的に諸経験を整理することである。」と規定されている。つまり、ジレンマ問題については、プログラミングするな、ということが産業界に対して示されており、開発側は、たとえば、5 人を救うために 1 人を犠牲にする、といった状況設定と対処法を実装しないで済むことになる。また、市民側が持つであろう、「誰を犠牲にするのかを開発者が勝手に決めてしまう」という懸念についても、本指針により、不安の払拭がなされている。

さらに、同規則の 9 において、「回避することができない事故状況において、個人的な特徴（年齢、性別、身体あるいは精神上的の素質）によるあらゆる格付け（重みづけ）は厳格に禁止される。被害者同士の数を差し引き計算することも禁止である。人的被害数を減少させる一般的なプログラミングは支持されうる。乗り物のリスクの発生に関与する者は、関与しない者たちを犠牲にしてはならない。」と規定される点は特徴的である。人数比較という差し引き計算（Aufrechnung）を厳格に禁ずる姿勢は、過去のドイツの連邦憲法裁判例において、基本法 1 条 1 項の「人間の尊厳は不可侵である」を根拠に、「一括した生命対生命の衡量（pauschale Abwägung von Leben gegen Leben）」が許されずとした 1975 年判決や<sup>4</sup>、ハイジャックにより死を免れない乗客と地上の潜在的犠牲者の生命を差し引きしてはならないこと等を理由に、航空機を撃墜できるとした航空安全法に対する 2006 年の違憲判決<sup>5</sup>、等と整合的である。なお、この箇所（個人的な特徴による格付け禁止）は、驚くべきことに、2021 年 7 月に発効したドイツの改正道路交通法の 1e 条 2 項 2 号にも採用されている（樋笠, 2022）。

このようなドイツの基本権の立場によると、規則 8 が「生命対生命のような真のジレンマにおける決定は、関係者の予測できない行動様式を含んだ具体的な実際の状況に左右される。それゆえ、かかる決定は、一義的に規範化できず、また、倫理的に疑う余地のないようプログラムすることもできない。」と規定されていることも当然である。ジレンマ問題への対処については、生命体生命で優劣をつけられないため、プログラミング自体が禁止されているのである<sup>6</sup>。

さらにドイツは、2022 年に AFGBV（定められた運行区域における自律走行機能付き自

動車の許可と使用に関する省令)の ANLAGE 1 Teil 1 1.1 において、ジレンマの際に誰が優先されるべきかを示している。

#### ANLAGE 1 Teil 1 1.1

##### 一般的な衝突回避

他の道路利用者や無関係の第三者との衝突は、以下の方法で回避できる場合は回避しなければならない。

##### 1. 急ブレーキ

2. 周辺交通の他の参加者、他の無関係な第三者、または自律走行機能付き自動車の乗員を危険にさらさない回避操作がありうる。

自律走行機能付き自動車の乗員の生命を危険にさらすことを回避するための衝突が、周囲の交通の他の参加者または無関係の第三者の生命を危険にさらすことでしか回避できない場合（人命に対するやむを得ない代替リスク）には、周囲の交通の他の参加者および無関係の第三者よりも、自律走行機能付き自動車の乗員の保護を優先してはならない。

乗客を優先して周囲の交通の他の参加者または無関係の第三者の生命を危険にさらすことは禁じられている。これにより、開発および運用を行う自動運転の事業者において、乗客優先の挙動は組み込まず、周囲の交通の他の参加者または無関係の第三者に配慮したプログラムを組むことになる。各社バラバラの基準で動かず、行政が産業界に対し、AI 倫理をしっかりと示した例といえる。また、本条項の存在により、市民社会における危惧感・不安感は減少すると考えられる。

日本では、このような規定、指針を持たないまま自動運転の実装が進み続けている。ここで、ひとつ事例を想定したい。AI システムが運転する自動運転 (Lv.4 自動運転) バスが 60km/h で左車線を走行中に、前方 20m に人が倒れ込んでくるのを検知した。この時点で、仮に急制動 (フルブレーキ) を行うと約 19m で車両が制動され、ぎりぎりぶつからなくて済む。しかし、急制動に伴う横の重力が 1G を超えると仮定する (図 1)。

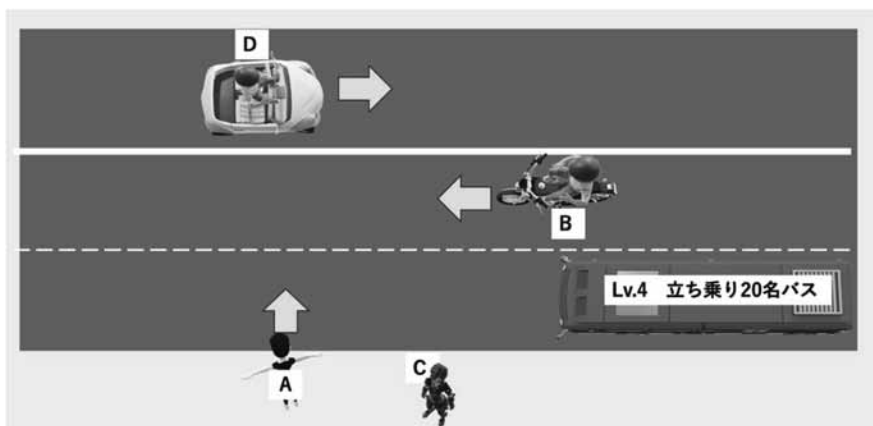


図1 リアルジレンマ状況 出所：筆者作成

20名の立ち乗りの乗客にとって、急制動で1Gの重力がかかることは、車内転倒による骨折等が想定される。ここで、重力を抑えるために制動（ブレーキ）をゆっくり通常制動としてかけるルートと、右に車線変更して他の道路利用者であるBに衝突するルート、左に進路を変え、歩道に乗り上げ無関係のCに当たる選択肢も想定される<sup>7</sup>。ここで、日本の企業は、どの選択肢をとるか迷うことになる。もっとも、ドイツでは、上述のAFGBVにより、急制動の選択肢一択である。よって、ドイツでは乗客が負傷する。しかし、法律を遵守したうえで起きた結果に対し、起訴がなされるとは考えにくい。そもそも法律を守るようプログラムして、守った以上、「過失」もないといえよう。また、市民社会的にも、そのような選択が条文をもって明示されているため、予想可能である。しかし、日本では、乗客を守らなかったバス会社は、批判を免れないであろうし、何らかの注意義務違反（過失）が認められうるかもしれない（表2）。また、企業ごとに、急制動を選ぶもの、車線変更するもの、など別れた場合には、交通社会は予測不可能で、不安定なものになりうる。

表2 リアルジレンマ状況に対応する挙動ごとの帰結 出所：筆者作成

選択肢	結果	批判
急制動	乗員が重傷等 場合によっては死亡	・バス運営会社に対する訴訟 ・乗客を守らない挙動をするバスという評価
通常制動	Aが死亡ないし重傷	・乗客を優先したとの評価
車線変更	Bが死亡ないし重軽傷	・交通参加者を犠牲にしたとの評価
歩道に乗り上げ	Cが死亡	・交通弱者を犠牲にしたとの評価

### 3. 2 EU

「EU 協調型自動運転車の倫理：交通安全・プライバシー・公平性・説明性・責任に関する勧告」は、2020 年 9 月に、EU から出された CAV（自動運転車）に関する倫理の提言であり、ドイツの倫理規則の延長線上に位置づけられるものである<sup>8</sup>。

勧告 6 の説明においては、「CAV の開発と展開を規制する際、政策立案者はジレンマ状況における CAV の行動が、不平等に関する勧告 5 に記載されたリスク配分の原則を遵守することで有機的に生まれることを受け入れることができる。このようなリスク配分の原則を守ることで、CAV の行動が基本的な倫理的・法的原则に抵触しないことが保証される。ジレンマに基づいた状況によって引き起こされた広範な社会的議論と、CAV が開発者によってプログラムされ、透明性のない、あるいは倫理的・社会的に受け入れがたい基準に基づいて衝突を選択するのではないかという社会的懸念を考慮するに、研究者、政策立案者、製造者および配備者は、勧告 6 について一般の人々を安心させ、その潜在的な影響について一般の人々を巻き込んだ検討プロセスを行うべきである。」とし、一般人をプログラミングの基準策定に関与させるべきとの方向性が見て取れる。

なお、勧告 6 の注釈部分では「起こりうるすべてのジレンマについて望ましい結果を定義するのではなく、CAV が衝突に至った全過程において、本報告書に記載されているすべての主要な倫理的・法的原则を遵守し、提言 5 から生じるリスク管理の原則を遵守し、ジレンマの発生を防止する合理的かつ実行可能な先行行動が存在しなかった場合には、ジレンマ状況における CAV の挙動はデフォルトで許容され则认为。このことは、CAV のメーカーや配備者が自信を持ってシステムを配備できるようにするために必要なことであり、速度を落として予防的な操縦（manoeuvres）を行うことが、安全上のリスクを減らす最善の解決策である。」と規定されており、NON-MALEFICENCE, BENEFICENCE, DIGNITY, PERSONAL AUTONOMY, RESPONSIBILITY, JUSTICE, SOLIDARITY, INCLUSIVE DELIBERATION の原則を守り、交通弱者へのリスク是正をしていれば、どんな結果でも「許容」されることになる。

以上、ドイツ・EU の状況に鑑みれば、市民社会と産業界との間をつなぐため、また、産業界を正しく導くために、政府・行政が AI の倫理指針を出すべきであるといえる。とはいえ、日本の自動運転分野のように、政府・行政が AI 倫理指針を出さない場合は、産業界で策定するしかなくなる。

## 4. 中立的幫助

### 4. 1 Winny 事件

現代の産業界における AI 倫理指針の策定を考えるにあたり、参考になるのは最決平成 23 年 12 月 19 日である<sup>9</sup>。これは、Winny 事件と呼ばれ、ファイル共有ソフト「Winny」に絡む著作権法違反（公衆送信権の侵害）を問われたが、被告人が無罪となった刑事事件

である。Winny の利用者（著作権法違反の当事者）だけではなく、Winny のアプリケーションソフトウェア開発者も、逮捕・起訴された事案である。ユーザーの使い方が悪ければ法律違反も引き起こしうる、中立的なソフトであるが、悪用したものがいた事案である。まさに、現代の生成 AI と同様の状況である。

最高裁は、「もっとも、Winny は、1, 2 審判が価値中立ソフトと称するように、適法な用途にも、著作権侵害という違法な用途にも利用できるソフトであり、これを著作権侵害に利用するか、その他の用途に利用するかは、あくまで個々の利用者の判断に委ねられている。また、被告人がしたように、開発途上のソフトをインターネット上で不特定多数の者に対して無償で公開、提供し、利用者の意見を聴取しながら当該ソフトの開発を進めるという方法は、ソフトの開発方法として特異なものではなく、合理的なものを受け止められている。新たに開発されるソフトには社会的に幅広い評価があり得る一方で、その開発には迅速性が要求されることも考慮すれば、かかるソフトの開発行為に対する過度の萎縮効果を生じさせないためにも、単に他人の著作権侵害に利用される一般的可能性があり、それを提供者において認識、認容しつつ当該ソフトの公開、提供をし、それを用いて著作権侵害が行われたというだけで、直ちに著作権侵害の幫助行為に当たると解すべきではない。」と判示した。「著作権侵害に利用するか、その他の用途に利用するかは、あくまで個々の利用者の判断に委ねられている」とし、企業側ではなく、ユーザー側に責任があるかのような文言も用いられている。この状況下で、開発者側が有罪になる場合について、最高裁は、「かかるソフトの提供行為について、幫助犯が成立するためには、一般的可能性を超える具体的な侵害利用状況が必要であり、また、そのことを提供者においても認識、認容していることを要するというべきである。すなわち、ソフトの提供者において、当該ソフトを利用して現に行われようとしている具体的な著作権侵害を認識、認容しながら、その公開、提供を行い、実際に当該著作権侵害が行われた場合、当該ソフトの性質、その客観的利用状況、提供方法などに照らし、同ソフトを入手する者のうち例外的とはいえない範囲の者が同ソフトを著作権侵害に利用する蓋然性が高いと認められる場合で、提供者もそのことを認識、認容しながら同ソフトの公開、提供を行い、実際にそれを用いて著作権侵害(正犯行為)が行われたときに限り、当該ソフトの公開、提供行為がそれらの著作権侵害の幫助行為に当たると解するのが相当である。」と判示している。提供者（開発・企業側）が、当該ソフトを（ユーザーが）利用して現に行われようとしている具体的な著作権侵害を認識（＝理解して）、認容（＝それでも構わないと思って）しながら、そのソフトの公開や提供することが、著作権侵害の「幫助」にあたるとされている。このような、それ自体が違法ではない幫助を、中立的幫助という。たとえば、包丁を売る店で売られた包丁が殺人に使われたとしても、直ちに包丁屋の店員が幫助となるわけではない。中立的幫助行為にあたるが、主観的に幫助の故意（＝ここでは、殺人行為を認容する内心のこと）を有していないため、店員は無罪となる。

実際に被告人が幫助にあたる行為をしたか否かについて、最高裁は、「そして、被告人の本件 Winny の提供方法をみると、違法な ファイルのやり取りをしないようにとの注意書きを付記するなどの措置を採りつつ、ダウンロードをすることができる者について何ら限定をかけることなく、無償で、継続的に、本件 Winny をウェブサイト上で公開するという方法によっている。これらの事情からすると、被告人による本件 Winny の公開、提供行為は、客観的に見て、例外的とはいえない範囲の者がそれを著作権侵害に利用する蓋然性が高い状況の下での公開、提供行為であったことは否定できない。」とし、Winny の提供行為それ自体は、客観的に（他の者による著作権法違反の）幫助行為にあたるとした。もっとも、主観的に幫助の故意（侵害的利用の高度の蓋然性の認識・認容）がないので、結果としては「無罪」となった。

なお、本最高裁決定には、大谷判事の反対意見がある。同意見は、「多数意見は、被告人の幫助の故意を消極的、否定的に評価する事情として、開発スレッドへの書き込みに自らソフトの開発・提供の意図を書き込んでいたとか、著作権者側の利益が適正に保護されることを前提とした新たなビジネスモデルの出現を期待していたとか、侵害的利用についてこれをしないよう警告のメッセージを発していたという点を挙げるが、これらは被告人に法益侵害の積極的意図が無かったという事情としてはもっともであるにしても、これらの事情が必ずしも法益侵害の危険性の認識・認容と抵触し、これを否定することにはならないと考えられる。提供行為の法益侵害の危険性を認識しているからこそ、このような利用が自らの開発の目的や意図ではなく、本意ではないとして警告のメッセージとして発したものと考えられる。被告人は、このようなメッセージを発しながらも、侵害的利用の抑制への手立てを講ずることなく提供行為を継続していたのであって、侵害的利用の高度の蓋然性を認識、認容していたと認めざるを得ない。」とし、侵害的利用についてこれをしないよう警告のメッセージを発していたことが、犯罪に使われてしまう認識を充足し、かつ、侵害的利用の抑制への手立てを講ずることなく提供行為を継続していたことが、開発者の主観的に、「著作権侵害を認識（＝理解して）、認容（＝それでも構わないと思って）」していたと評価されている。

したがって、多数決で無罪とはなっているものの、侵害的利用の抑制への手立てを講じないことは、「中立的幫助の故意が存在する」との評価側になり得る事情であることは明白である。よって、たとえば、生成 AI を一般公開し、利用できるようにする企業においては、「このような悪用がなされそうだ」という法的リスクを認知していることが多いため、侵害的利用の抑制への手立てを講じることが必要である（図 2）。



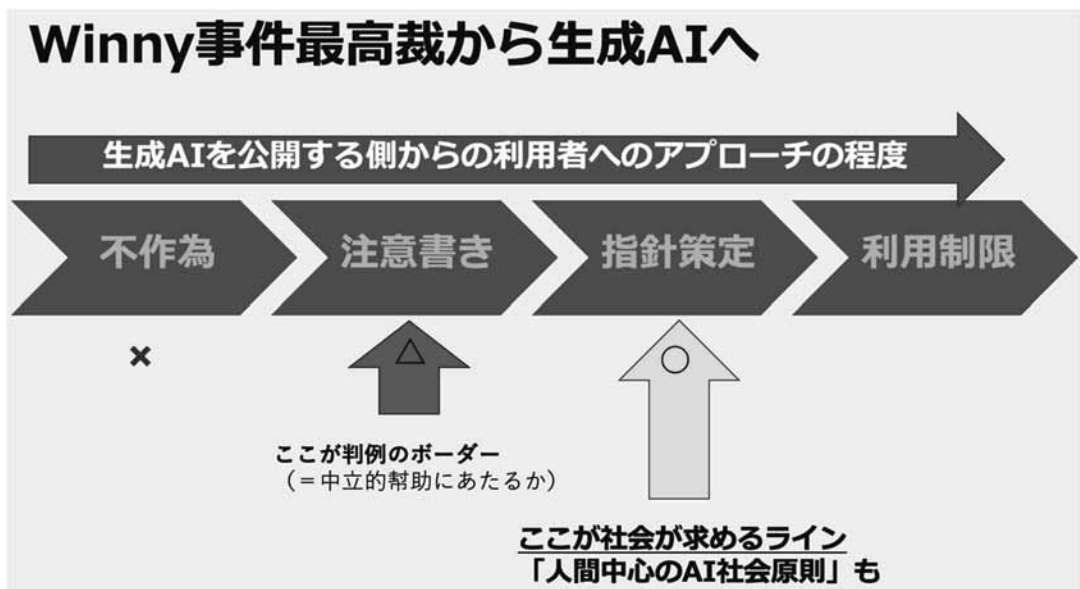


図 2 Winny 事件最高裁から生成 AI へ 出所：筆者作成

では、侵害的利用の抑制への手立てとは、どのようなものが想定されるか。たとえば、公開時に誰でもダウンロードできるのではなく、一定の資格保有者、あるいは本人確認が取れた者などに制限をするのも一つの手である。悪用の抑止に資するといえる。とはいえ、多くの一般人に当該ソフトを使用してもらいたい企業側としては、制限という選択肢は取りづらい。そこで考えられるのが、企業自ら、さらにユーザーをも縛る指針（ガイドライン）の策定と、その行動の実行である。

開発し、公開し、「あとは利用者の良心に委ねる」のは誤りである。悪用・乱用・誤用の可能性がありながらも、企業側がルール・指針を作らないことには、一定の「不作為」の責任があるといえる。それが中立的幫助である。したがって、指針策定時には、一般的な、「平等性」「公平性」や「透明性」などに加えて、犯罪利用等を防ぐためのルール・改善のプロセス規定が必要となろう。

## 5. 帰結

したがって、行政・政府が AI 倫理指針を策定しない場合には、産業界側が自ら AI 倫理指針を策定する必要がある。それは、市民社会と産業界をつなぐためだけではなく、翻って、企業を守ることに寄与する。AI 倫理指針を策定して、侵害的利用の抑制への手立てを講じることは、中立的幫助を避けることになるからである。「平等性」「公平性」や「透明性」などに加えて、犯罪利用等を防ぐためのルール・改善のプロセスを規定した AI 倫理指針を策定することが、目下、AI を利用する産業界にとって必要なことである。

- 
- <sup>1</sup> 倫理については、樋笠堯士（2023d）を参照。
- <sup>2</sup> たとえば、医薬品においては、開発の段階から市販後に至るまで常にリスクを適正に管理するため、「医薬品リスク管理計画」（RMP：Risk Management Plan）のシステム化がなされている。個々の医薬品について安全性上の検討課題を特定し、使用成績調査、市販直後調査等による調査・情報収集や、医療関係者への追加の情報提供などの医薬品のリスクを低減するための取組を、医薬品ごとに文書化して公開している。
- 独立行政法人医薬品医療機器総合機構（最終閲覧日 2024 年 10 月 15 日）  
<https://www.pmda.go.jp/safety/info-services/drugs/items-information/rmp/0002.html>
- <sup>3</sup> 富士通グループ AI 倫理外部委員会（最終閲覧日 2024 年 11 月 19 日）  
<https://global.fujitsu/ja-jp/technology/key-technologies/ai/aiethics/governance>
- <sup>4</sup> BVerfGE 39,1.
- <sup>5</sup> BVerfGE 115,118.
- <sup>6</sup> 樋笠（2023c）を参照。
- <sup>7</sup> 現状の車両ルール、国土交通省の保安基準では、急制動が原則であり、操舵回避は許容されていない。
- <sup>8</sup> もっとも、Recommendation レベルなので、EU において法的拘束力は持たない（＝規則・指令・決定よりも下）
- <sup>9</sup> 刑集第 65 巻 9 号 1380 頁

## 【参考文献】

- 総務省（2023）「全ての AI 関係者向けの広島プロセス国際指針」（最終閲覧日 2024 年 10 月 15 日）  
<https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document03.pdf>
- 内閣府（2019）「人間中心の AI 社会原則（統合イノベーション戦略推進会議決定）（平成 31 年 3 月 29 日）」（最終閲覧日 2024 年 10 月 15 日）
- 樋笠堯士（2022）「自動運転レベル 4 における関与者の義務と責任およびデータ記録－ドイツの改正道路交通法を手がかりに－」『経営情報研究』No.26, 49-68 頁
- 樋笠堯士（2023a）「自動運転の倫理に関する ISO39003 の分析と自動運転レベル 4 への応用可能性」『国際交通安全学会誌』Vol.48, No.2, 132-140 頁
- 樋笠堯士（2023b）『自動運転レベル 4 どうしたら社会に受け入れられるか』学芸出版社。
- 樋笠堯士（2023c）「自動運転の倫理」『自動車技術』1 月号（No.77），48-53 頁
- 樋笠堯士（2023d）「自動運転レベル 4 の実装に伴う販売店の倫理」『自動車販売』2023 年 5 月号，2-7 頁
- 樋笠堯士（2024）「AI の国際的潮流と自動運転の社会実装～事故減少のための運転支援や社会的受容性の醸成～」『自動車販売』7 月号，2-8 頁



<https://www8.cao.go.jp/cstp/ai/aigensoku.pdf>

Gill, Tripat. (2021). Ethical Dilemmas Are Really Important to Potential Adopters of Autonomous Vehicles. *Ethics and Information Technology* 23 (4): 657–673.

G7. (2024). G7 meeting with Outreach Countries and International Organizations (Borgo Egnazia, 14 June 2024) (最終閲覧日 2024 年 10 月 15 日)

<https://www.mofa.go.jp/mofaj/files/100685413.pdf>

Kumar, Sachin, Vidhisha Balachandran, Lucille Njoo, Antonios Anastasopoulos, and Yulia Tsvetkov. (2023). “Language Generation Models Can Cause Harm: So What Can We Do About It? An Actionable Survey.” In *Proceedings of EACL-2023*, 3299–3321

Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. (2022). Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 214–229

<https://doi.org/10.1145/3531146.3533088>

Open AI. (2023). GPT-4 Technical Report, Appendix “GPT-4 System Card,” 41-78  
<https://arxiv.org/abs/2303.08774>

## 第4章 AI 活用に伴う法的課題

### —プライバシー権、平等権、著作権、特許権—

樋笠知恵

(信州大学医学部医学科公正研究推進講座)

#### 1. はじめに

2022年に登場した生成AIをはじめとするAIは、文章作成、画像作成、自動翻訳、医療における診断支援、自動車・ドローンの自律運転、自律型ロボットなど様々な分野で幅広く活用されており、私たちの日常と切り離せない存在となっている。他方で、AIの活用に伴い、さまざまな法的課題が生じていることも事実である。なお、ここでの「法的課題」とは必ずしもネガティブなものではなく、AIと人間が共存していく現在と未来を創造していくための前向きな議論であると筆者は捉えている。

本章では、今後AIの技術がより発展しその活用の機会が増加していくことに備えて、改めて議論しておくべきと思われる法的問題として、①プライバシー権侵害、②平等権侵害、③著作権侵害、④特許法における「発明」と「発明者」適格について、考察を行う。

#### 2. プライバシー権侵害

生成AIを含むAIの活用に伴いプライバシー権に関する議論が必要とされていることは、すでに、さまざまなところで指摘されている<sup>1,2</sup>。プライバシー権は、日本国憲法第13条後段の幸福追求権（「生命、自由及び幸福追求に対する国民の権利については、公共の福祉に反しない限り、立法その他の国政の上で、最大の尊重を必要とする。」）に含まれる権利である。

近年、AI活用と密接に関連する形で進められているDX化におけるデータ収集の際などに、人の行動履歴や健康状態、思想・信条、趣味嗜好等が詳細に把握されることによって、個人のプライバシー権に対するリスクは高まっている<sup>3</sup>。もっとも、新たな価値を創造するためのデータ活用の推進と個人の権利利益の保護は、微妙な均衡状態を保ちながら実現していくほかない。また、生成AIの活用についても、AIの学習の際に各種データを読み込ませ、市場に投入されたAIツールを人々が利用し、市場投入後もAIが継続的に学習する際に、プライバシー権侵害の問題が生じ得る<sup>4</sup>。もっとも、プライバシー権自体の捉え方が、学問上、あるいは社会認識として、変化していく可能性もある。引き続き議論を注視していく必要があるだろう。

個人情報に関する法制度との関連では、1980年に示された「プライバシー保護と個人デ

ータの国際流通についてのガイドライン」(OECD)における8原則<sup>5</sup>を受け、1988年にわが国で最初の個人情報保護法として行政機関個人情報保護法が制定された後、1998年にプライバシーマーク制度が創設され民間部門での取り組みが始まり、2003年に現在の個人情報保護法が制定されている。その後、情報通信技術の発展やパーソナルデータの利活用によって国境を超えたデータ流通が活発化したことを背景に2015年に改正が行われ、個人情報の保護に関する国際的動向や情報通信技術の進展を踏まえた3年ごとの見直しとして、2020年に再び改正法が成立し、2022年4月1日に同改正法が施行されている。わが国の個人情報保護法および同法に基づく規則等は、今後のAI活用とDX化に関連する形で引き続き改正を必要とすることになるだろう。

### 3. 平等権侵害

AI活用に伴い、憲法第14条第1項で保障された平等権<sup>6</sup>の侵害の問題が生じ得る。このことは、プライバシー権侵害同様、これまでさまざまなところで指摘されてきた<sup>7</sup>。平等権は最も重要な人権の一つであるが、歴史上、人種、信条、性別などの特定の事由のみを理由とした差別が行われてきたことは事実であり、AIもこのような差別に関する問題とは無縁ではない。

AIによる差別の主な原因としては、①アルゴリズム設計のバイアス、②データの代表性の欠如、③データに内在する既存の社会のバイアス、の3つが指摘されている。①は、アルゴリズムの設計段階で開発者が差別的な意図をもってアルゴリズムを設計することなどにより生じるとされている。これには、意図せずに差別的なアルゴリズムが設計される場合も含まれる。②は、AIの利用段階で代表性に欠ける偏ったデータを学習してしまうことによるとされている。③は、データの元となっている社会の構造自体に不公平なバイアスがある場合に、データに従来社会の差別的な構造が反映され、それを学習したAIが差別を再生産することによって生じる。これら①～③に共通する大きな特徴は、「AIによる差別」の多くはAI自体が生み出したものではなく、人間が生み出した差別構造をAIが学習し再生産した差別だということである(成原, 2021; 樋笠, 2024)。実際、既に、採用時のフィルタリングにおいて女性が差別される事例や、対話型AIを利用した際に人種差別が起こる事例が確認されている。そこで、公平性の確保に関する対策として、公平性配慮データマイニングという研究トピックも立ち上がっており、グループ公平性・個人公平性などの公平性基準を定義し、それを用いて不公平さを検出する手法も提案されている。

### 4. 著作権侵害

生成AIの活用が進むなか、AIの開発事業者や利用者などからどのような場合に著作権侵害となるのかが不明確でリスクが大きいといった懸念の声が上がったことから、文化庁(2024)は著作権侵害に関する見解を示した。生成AIによる生成物の生成段階と生成物

の利用段階では、生成物の生成行為（著作権法における複製等）と、生成物のインターネットを介した送信などの利用行為（著作権法における複製、公衆送信等）に際して既存の著作物の著作権侵害が発生する可能性がある。

従来の AI が関与しない事案において、裁判所は、著作権者から許諾を得ておらず、権利制限規定（家庭内複製や引用）にも該当しないにも関わらず、①自分の著作物が他人の既存の著作物と同一、又は類似しており（類似性）、②その自分の著作物を既存の著作物に依拠して複製等した（依拠性）場合に、著作権侵害を認めてきた。このような①②の判断枠組みが生成 AI を利用した場合にも妥当するか否かが問題となるが、この点、文化審議会著作権分科会法制度小委員会は、「従前の人間が AI を使わずに行う創作活動の際の著作権侵害の要件と同様に考える必要がある。」、「生成 AI による生成物についても、その生成・利用段階において、既存の著作物との類似性及び依拠性が認められれば、当該既存の著作物の著作権者は、生成物の生成行為や利用行為が、既存の著作物の著作権侵害に当たるとして、当該行為の差止請求や損害賠償請求を請求し得る。また、故意による著作権侵害に対しては、刑事罰の適用があり得る。」としている。その上で、「類似性」については、「AI 生成物と既存の著作物との類似性の判断についても、人間が AI を使わずに創作したものについて類似性が争われた既存の判例と同様、既存の著作物の表現上の本質的な特徴が感得できるかどうかということ等により判断されるものと考えられる。なお、ここでいう「表現上の本質的な特徴」に具体的に当たるものについては、個別具体的な事例に即し、判断されることに留意する必要がある」と述べ、他方、「依拠性」については、「このような既存の判例・裁判例を踏まえると、生成 AI が利用された場合であっても、権利者としては、被疑侵害者において既存著作物へのアクセス可能性があったことや、生成物に既存著作物との高度な類似性があること等を立証すれば、依拠性があるとの推認を得ることができる。」と示している。以上の見解は、法令解釈の権限を持つ裁判所によるものではないものの、著作権法を所管する文化庁が携わる委員会によるものであることから、現時点での実務上の指針の一つとなり得ると考えられる。

## 5. 特許法における「発明」と「発明者」適格

### 5. 1 現行特許法での解釈運用の可能性

まず、AI を利用した発明について現行特許法の規定の枠内での解釈運用が可能かどうかについては、中山（2021）から「現状では、その多くは、AI 技術自体に関する発明又は人間が AI を「道具」として創作する発明のいずれかであろうと思われる。この場合、人間が発明者であることに変わりはなく、AI 関連発明に関与する人間のうち発明者は誰か、換言すれば、どのような人間の関与が発明者に求められるのが問題となる。その意味において、この問題は、従来からの発明者の認定をめぐる問題の一場面であり、現行法の枠内での解釈問題である。」などといった指摘がされていたところ、特許庁（2023）も、「AI を利

用した発明についても、現段階では現行法制度上の発明者の要件の考え方で対応可能ではないか。」との認識を示した。

## 5. 2 「発明」の定義と発明者

特許法第2条第1項は、「この法律で「発明」とは、自然法則を利用した技術的思想の創作のうち高度のものをいう。」と規定している。ここで「技術的思想」とは、技術的課題を解決するための具体的な手段をいうと考えられている。他方、特許を受けることができる者については、特許法第29条第1項柱書で、「産業上利用することができる発明をした者は、…その発明について特許を受けることができる。」とされている。裁判例では、「発明者」たるためには、発明完成の経緯からみて発明の特徴的部分である技術的思想の創作に実質的な貢献をしたことが求められており<sup>8</sup>、通説的見解もおおよそ裁判例の傾向と一致する<sup>9</sup>。これらの議論と関連して、自然人が発明の過程でAIを活用した場合に、「発明」にあたるのか、そして発明者は誰なのか、という問題が生じる。

## 5. 3 AIを発明者とすることの可否

それでは、発明の創作過程でAIを活用した場合、AIを発明者として記載し特許を取得することはできるのだろうか。特許法第29条第1項柱書は、「産業上利用することができる発明をした者は、……その発明について特許を受けることができる。」と規定するのみで「発明者」の定義自体はないが、特許法第36条第1項第2号が、願書に発明者の氏名及び住所又は居所を記載することとしているのに対して、特許法第36条第1項第1号が「出願人」については氏名又は名称及び住所又は居所を記載することとしていることからすれば、特許法は、「発明者」として自然人のみを想定していると考えられる（特許庁、2021）。

他方で、自然人による指示がある程度抽象的な場合には、そもそも、その自然人を発明者として認めてよいかが問題となるが、この点、「人間の指示が多少抽象的でもAIによる具体的解決手段特定の蓋然性をあわせて考慮すれば、当該指示を出した人間を発明者と認定し得る余地はあり得る」との考え方も示されている（中山、2021、69）。

## 5. 4 発明の創作過程におけるAI活用の現状

デロイトトーマツファイナンシャルアドバイザーリー合同会社（2024）では、発明の創作過程におけるAI活用の現状が示されている。同社が実施した調査（以下、DTFA調査）は、AIを利活用した創作について特許法上の保護の在り方について検討がなされてきた一方で、創作過程におけるAIの利活用が拡大することが見込まれ、それによって生まれた発明を含む特許出願が増えることが予想されることなどを背景として、諸外国の状況も踏まえ、(1)最新のAIの技術水準や、発明の創作過程におけるAIの利活用の状況を調査し、(2)創作過程におけるAIの利活用の拡大により生じる特許審査実務上の課題を整理・検

討し、さらに、(3) AI による自律的な発明の取扱いに関する課題を整理・検討することで、AI を利活用した創作の特許法上の保護の在り方を検討する上での基礎資料を作成することを目的として、公開情報調査、国内アンケート調査、国内ヒアリング調査、海外質問票調査、海外ヒアリング調査、および委員会での議論を行ったものである。

DTFA 調査報告書によれば、「発明者」適格と関連する国内アンケート調査（125 者にアンケートを送付<sup>10</sup>、うち 41 者から回答<sup>11</sup>）の結果は、次の通りである。発明の創作過程における AI の活用状況については、現在活用している（14 者）、現在活用していないが今後活用する予定がある（15 者）、現在活用しておらず今後活用する予定もない（12 者）との回答が得られている<sup>12</sup>。回答者のおよそ 70%が、発明の創作過程で AI を現在利用しているか、もしくは、今後活用する予定があり、少なくともこの調査結果からは、AI を活用した発明に伴い生じる課題に取り組む必要性の高さが読み取れる。また、現在活用していると回答した 14 者に対して AI 活用の具体的な態様を訪ねたところ、課題候補の抽出（7 者）、課題解決手段候補の抽出（13 者）、用途候補の抽出（4 者）、効果の検証（7 者）との回答が得られており<sup>13,14</sup>、発明の創作過程の複数の段階で AI が活用されていることが分かっている。

## 5. 5 東京地裁令和 6 年 5 月 16 日判決

さて、このような状況の下、特許法にいう「発明」が自然人によるものに限られるかどうかについて、2024 年 5 月 16 日に東京地裁判決<sup>15</sup>が出されている。事案の概要は次の通りである。原告は、国際出願をした上、特許庁長官に対し特許法第 184 条の 5 第 1 項所定の書面に係る提出手続をし、国内書面における発明者の氏名として、「ダバス、本発明を自律的に発明した人工知能」と記載した。これに対し、特許庁長官は、原告に、発明者の氏名として自然人の氏名を記載するよう補正を命じたものの、原告が補正をしなかったため、同条の 5 第 3 項に基づき、本件出願を却下する処分（以下、「本件処分」という）を行った。そこで、原告は、特許法にいう「発明」は AI 発明を含むものであり、AI 発明に係る出願では発明者の氏名は必要的記載事項ではないから、本件処分は違法である旨主張して、本件処分の取消しを求めた。

本事案において、東京地裁は次のように判示した。「知的財産基本法 2 条 1 項は、「知的財産」とは、発明、考案、植物の新品種、意匠、著作物その他の人間の創造的活動により生み出されるもの（発見又は解明がされた自然の法則又は現象であつて、産業上の利用可能性のあるものを含む。）、商標、商号その他事業活動に用いられる商品又は役務を表示するもの及び営業秘密その他の事業活動に有用な技術上又は営業上の情報をいうと規定している。上記の規定によれば、同法に規定する「発明」とは、人間の創造的活動により生み出されるものの例示として定義されていることからすると、知的財産基本法は、特許その他の知的財産の創造等に関する基本となる事項として、発明とは、自然人により生み出さ



れるものと規定していると解するのが相当である。

そして、特許法についてみると、発明者の表示については、同法 36 条 1 項 2 号が、発明者の氏名を記載しなければならない旨規定するのに対し、特許出願人の表示については、同項 1 号が、特許出願人の氏名又は名称を記載しなければならない旨規定していることからすれば、上記にいう氏名とは、文字どおり、自然人の氏名をいうものであり、上記の規定は、発明者が自然人であることを当然の前提とするものといえる。また、特許法 66 条は、特許権は設定の登録により発生する旨規定しているところ、同法 29 条 1 項は、発明をした者は、その発明について特許を受けることができる旨規定している。そうすると、AI は、法人格を有するものではないから、上記にいう「発明をした者」は、特許を受ける権利の帰属主体にはなり得ない AI ではなく、自然人をいうものと解するのが相当である」。したがって、「特許法に規定する「発明者」は、自然人に限られるものと解するのが相当である。」

以上のように、東京地裁は、知的財産基本法第 2 条第 1 項の「発明」が自然人により生み出されることを想定した規定であると解されること、特許法においても、同法第 36 条第 1 項第 2 号における「氏名」が自然人を想定していること、同法第 29 条第 1 項の「発明をした者」は特許を受ける権利の帰属主体になり得る自然人をいうと解するのが相当であることを理由として、特許法における発明者は自然人に限られることを明らかにしている。

本判決は、AI を発明者として認めることの可否について裁判所が初めて見解を示したものとして意義がある。AI を活用した発明とその発明にかかる特許出願が増加していく流れのなかで、発明者となり得るのは誰かというこの問題は今後新たな論点（AI が自律的に発明をするようになった場合はどうかなど）を交えながら、成熟していくと思われる。ひとまず、本判決は現時点での一定の方向性を示したものとし、今後も、その動向を注視していくことが必要だろう。

## 6. おわりに

本章では、AI が私たちの日常と切り離せない存在となった今、改めて議論しておくべきと思われる法的課題として、①プライバシー権侵害、②平等権侵害、③著作権侵害、④特許法における「発明」と「発明者」適格についての考察を行った。ここでの「法的課題」とは、AI と人間が共存していく現在と未来を創造していくための前向きな議論である。さまざま文脈でこれらの権利侵害のリスクが指摘されるなか、著作権侵害については文化庁から一つの考え方が、特許権の「発明者」適格については裁判所から一定の方向性が示されるなど、議論は進みつつある。新しい AI 技術の登場や技術革新などとともに個人の権利利益の捉え方が変化していくなか、どのような関わり方（開発、利用）であれ AI に関与する者がそれぞれの視点で引き続き議論を行っていくことが求められるだろう。



<sup>1</sup> 内閣府（2019）では、基本理念として「プライバシー確保の原則」が示されており、「AIを前提とした社会においては、個人の行動などに関するデータから、政治的立場、経済状況、趣味・嗜好等が高精度で推定できることがある。これは、重要性・要配慮性に応じて、単なる個人情報を扱う以上の慎重さが求められる場合があることを意味する。パーソナルデータが本人の望まない形で流通したり、利用されたりすることによって、個人が不利益を受けることのないよう、各ステークホルダーは、…パーソナルデータを扱わなければならない。」と述べられている。また、内閣府（2023）においても、プライバシーに関するリスクが指摘されている。

<sup>2</sup> 詳しくは、樋笠（2024）を参照のこと。

<sup>3</sup> 同様の指摘をし、企業のプライバシーガバナンスに関する指針となるものとして、総務省・経済産業省（2023）がある。

<sup>4</sup> 生成 AI とプライバシー権侵害について、詳しくは樋笠（2024）を参照のこと。

<sup>5</sup> ①収集制限の原則、②データ内容の原則、③目的明確化の原則、④利用制限の原則、⑤安全保護の原則、⑥公開の原則、⑦個人参加の原則、⑧責任の原則の 8 つ。

<sup>6</sup> 日本国憲法第 14 条第 1 項は、「すべて国民は、法の下に平等であつて、人種、信条、性別、社会的身分又は門地により、政治的、経済的又は社会的関係において、差別されない。」と規定する。

<sup>7</sup> 内閣府（2019）においては、「AI の利用によって、人々が、その人の持つ背景によって不当な差別を受けたり、人間の尊厳に照らして不当な扱いを受けたりすることがないように、公平性及び透明性のある意思決定とその結果に対する説明責任（アカウンタビリティ）が適切に確保されると共に、技術に対する信頼性（Trust）が担保される必要がある。」「AI の設計思想の下において、人々がその人種、性別、国籍、年齢、政治的信念、宗教等の多様なバックグラウンドを理由に不当な差別をされることなく、全ての人々が公平に扱われなければならない。」との指摘がある。

<sup>8</sup> 例えば、知財高判令和 3 年 3 月 17 日、東京地判平成 17 年 9 月 13 日など。

<sup>9</sup> 通説的見解は、発明者たるためには、一般に発明の特徴的部分（技術的思想）の創作に現実的に関与した者である必要があるとする。高林龍（2020）『標準特許法第 7 版』（有斐閣）72 頁、三村量一（2006）「発明者の意義」『金融商事判例 1236 号』（経済法令研究会）123 頁など。

<sup>10</sup> 国内アンケート調査の対象は、1.特許情報から抽出（「AI に関する特許検索式」を用いて、過去 10 年間に AI 関連出願が 1 件以上ある企業を抽出した後、経済情報プラットフォーム SPEEDA の業界分類に基づき、各業界の出願件数上位 5 者をアンケート対象企業として選定。また、「生成系 AI に関する特許検索式」を用いて、過去 10 年間に生成系 AI 関連出願が 1 件以上ある企業をアンケート対象企業として選定。）した合計 103 者と、2.公開情報から抽出（公開情報に基づき AI を開発、又は導入している企業をアンケート対象

企業として選定。)した 22 者の合計 125 者。

<sup>11</sup> アンケートを送付した計 125 者のうち、41 者（素材・素材加工品 8 者、医薬・バイオ 6 者、広告・情報通信 4 者、不動産 4 者、機械・電気製品 3 者、公共サービス 3 者、消費財 3 者、大学・研究機関 2 者、中間流通 2 者、輸送機械 2 者、建設 1 者、消費者サービス 1 者、食品 1 者、法人サービス 1 者）が回答。

<sup>12</sup> DTFA 調査報告書 74 頁を参照。

<sup>13</sup> DTFA 調査報告書 75 頁を参照。

<sup>14</sup> それぞれの態様ごとの人間の関与については、DTFA 調査報告書 77 頁～78 頁参照のこと。

<sup>15</sup> 東京地判令和 6 年 5 月 16 日（最終閲覧日 2024 年 10 月 14 日）

[https://www.courts.go.jp/app/files/hanrei\\_jp/981/092981\\_hanrei.pdf](https://www.courts.go.jp/app/files/hanrei_jp/981/092981_hanrei.pdf)

### 【参考文献】

総務省（n.d.）「プライバシー保護と個人データの国際流通についてのガイドラインに関する OECD 理事会勧告（1980 年 9 月）（仮訳）」（最終閲覧日 2024 年 10 月 14 日）

[https://www.soumu.go.jp/main\\_content/000196320.pdf](https://www.soumu.go.jp/main_content/000196320.pdf)

総務省・経済産業省（2023）「DX 時代における企業のプライバシーガバナンスガイドブック ver1.3（2023 年 4 月）」（最終閲覧日 2024 年 10 月 14 日）

[https://www.meti.go.jp/policy/it\\_policy/privacy/guidebook\\_ver1.3.pdf](https://www.meti.go.jp/policy/it_policy/privacy/guidebook_ver1.3.pdf)

デロイトトーマツファイナンシャルアドバイザーリー合同会社（2024）「AI を利活用した創作の特許法上の保護の在り方に関する調査研究報告書（令和 6 年 3 月）」（最終閲覧日 2024 年 10 月 14 日）

[https://www.jpo.go.jp/system/patent/gaiyo/sesaku/ai/document/ai\\_protection\\_chousa/zentai.pdf](https://www.jpo.go.jp/system/patent/gaiyo/sesaku/ai/document/ai_protection_chousa/zentai.pdf)

特許庁（2021）「発明者等の表示について（令和 3 年 7 月 30 日）」（最終閲覧日 2024 年 10 月 14 日）

<https://www.jpo.go.jp/system/process/shutugan/hatsumei.html#:~:text=1.%20%E3%80%82>

特許庁（2023）「AI 技術の進展を踏まえた発明の保護の在り方について（2023 年 11 月 7 日特許庁審査第一部調整課審査基準室）」（最終閲覧日 2024 年 10 月 14 日）

[https://www.kantei.go.jp/jp/singi/titeki2/ai\\_kentoukai/gijisidai/dai3/siryou3.pdf](https://www.kantei.go.jp/jp/singi/titeki2/ai_kentoukai/gijisidai/dai3/siryou3.pdf)

内閣府（2019）「人間中心の AI 社会原則（2019 年 3 月 29 日統合イノベーション戦略推進会議決定）」（最終閲覧日 2024 年 10 月 14 日）

<https://www8.cao.go.jp/cstp/ai/aigensoku.pdf>

内閣府（2023）「AIに関する暫定的な論点整理（2023年5月26日AI戦略会議）」（最終閲覧日2024年10月14日）

[https://www8.cao.go.jp/cstp/ai/ronten\\_honbun.pdf](https://www8.cao.go.jp/cstp/ai/ronten_honbun.pdf)

中山一郎（2021）「AI関連発明の発明者」『別冊パテント26号』（日本弁理士会），49-50頁

成原慧（2021）「AIと人権—「AIによる差別」と公平性」，ひろげよう人権，クローズアップ（最終閲覧日2024年10月14日）

[https://www.jinken-net.com/close-up/20210802\\_2587.html](https://www.jinken-net.com/close-up/20210802_2587.html)

樋笠知恵（2024）「AI活用に伴う“ELSI”と研究におけるAI活用」『横幹』18(1), 18-25頁

文化庁（2024）「AIと著作権に関する考え方について（2024年3月15日文化庁審議会著作権分科会法制度小委員会）」（最終閲覧日2024年10月14日）

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



## 第5章 論文執筆と発表における生成 AI の適正利用と課題

### — 生命系実験科学分野の研究活動を例に —

岡林浩嗣  
(筑波大学)

#### 1. はじめに

近年の生成 AI の急速な発展は、大学の研究教育の内容や、そのあり方自体に大きな影響を与えている。特に大学研究機関の研究活動においては、研究倫理や研究公正の観点から研究者や学生に求められる行動上の規範を想定する必要があるが、生成 AI は研究活動のあらゆる点に活用され得ることから、今後の生成 AI の加速的な実用化や高度化を見据え、その利活用が研究活動に与える影響を功罪共に想定しておくことは重要である。

本章は、一般財団法人公正研究推進協会の 2023 年度全国研究公正推進会議サテライトイベントにおける講演「学術論文の執筆と発表における生成 AI の適正利用と課題 —生命系実験科学分野の研究活動を例に—」の内容に基づき、特に実験生命科学分野の研究活動における生成 AI の利用の可能性を想定し、その活用に伴って発生し得る問題点と課題について、筆者なりの考えをまとめたものである。本章では最初に実験科学分野での生成 AI 利用がもたらす研究公正上の影響や様々な問題点について個別に概説し、次に生成 AI の利用により生ずる研究者の行動変容について述べる。さらに、生成 AI を研究活動へ積極的に利用する場合の条件に加え、生成 AI の利用を前提とし、研究成果の品質管理の観点から研究者に求められる寄与の考え方についても議論する。最後には、AI による支援を利用する場合に求められる研究活動上のルールや生成 AI の有効活用の方策に加え、近年のオープンサイエンスの推進を前提とし、研究成果の評価に関して今後求められる留意点について簡単に述べる。

なお、生成 AI とその学問的な利用に関する既存の論考や先行文献などについては、野内論文（本書第1章）を参照されたい。

#### 2. 生成 AI の利用推進により懸念される実験科学分野の問題

生成 AI を自動化された支援機能と捉える場合、作業者にとって本質的ではなく煩雑で時間のかかる手順を AI が代行し、無駄な労力を省くことによって作業効率をあげる方向で活用されるのが一般的であり、研究活動においてもそれは同様である。特に創造的な要素の強い研究活動の中での生成 AI の利用には懸念もあるものの、研究の本質を歪めない範囲で生産性の向上や作業の効率化を目的とした生成 AI の活用には望ましい面もある。

研究活動への生成 AI ツールの導入と健全な範囲での利用促進にあたっては、例えば研究活動を「実験や調査によるデータ取得」「解析」「解釈」「論文執筆」等に分けて考え、それらの活動において「煩雑で時間がかかり」さらに「研究者自身の知的寄与無しで自動化することが正当化できる」範囲が何なのか、が重要になる。研究活動にはその分野ならではのデータの扱いや解析、解釈を含めた独特な研究の進め方があること、さらに研究活動自体に内在する本来的な自由さや多様性を考えるならば、生成 AI の利用に伴って「何を AI に任せてもその研究の本質が失われないか」という根源的な問題については、その研究を推進する研究者自身が自ら考え、他人に説明できるだけの論拠が必要になろう。

一方で研究活動には、それを通じて得られる発見による栄誉や、成果としての論文の量と質がその研究者自身のキャリアアップや収入増に繋がること、さらに社会実装や産業応用により金銭的利益を生む可能性があることから、不可避免地に「競争」としての側面がある。この競争を勝ち抜くことを目的として、研究活動の生産性を上げる為に生成 AI を利用するという発想は自然なものではあるが、ここで研究活動の「品質管理」が問題となる。ここでいう品質管理とは、いわゆる「サービスそのものの品質管理」ということであり、研究成果を生み出すプロセスとしての研究活動の品質管理は、生み出される研究成果の妥当性を左右する決定的な要素でもある。

ある研究成果が妥当であるかどうかの判定は、論文発表における査読がそうであるように、外形的にはピア・レビュー（専門家同士の評価）によって定まるものの、論文の発表時点では、その論文を構成する「データの取扱い」「解析」「解釈」「論文（文章）」の妥当性は当該論文を発表する研究者側の判断に委ねられているといえる。論文発表における競争的側面を過度に重視するあまり、その研究の妥当性を大きく歪める不適切な行為や、明らかな捏造・改ざん・盗用等の「研究不正（Research misconduct）」があることは古くから認識されている（Steneck, 2006）。この様な不正を避け、研究活動やその成果の妥当性を担保する為にも研究活動の品質管理が必要とされているが、故意の有無を問わず研究不正が生じ得ることや、明確な研究不正には至らない「不適切な研究行為（Questionable research practices: QRPs）」が広く想定され得ることを含め、どの様な品質管理を行えばこれらの不正行為等を防止できるかを統一的に明示することは困難である。

研究活動に関わる生成 AI の利用においては、これらの不正を抑止する方向に活用されることが期待される一方で、残念ながら、生成 AI を利用して意図的な不正を気付かれないうように実施するような「不正の巧妙化」も想定され得る。その他にも「研究の生産性を上げるための生成 AI の利用である」と理解して実施した操作が、結果的に研究成果の妥当性や競争的環境を歪める様に作用する可能性もある。生成 AI が研究活動における不正を増加させ、研究活動そのものの本質を歪める方向で利用される可能性については幅広く、具体的に検討される必要があろう。

## 2. 1 AI の利用による特定不正行為の巧妙化

特定不正行為という用語は、文部科学省による研究不正対応ガイドラインで定義づけられており、特に罰則の対象となる研究不正として「研究成果の中に示されたデータや調査結果等の捏造、改ざん及び盗用」を指す（文部科学省，2014）。特定不正行為として認定されるかどうかは必ずしも故意の有無だけで決まる訳では無く、不正の有無に関する調査は実験等の生データ（数値的計測データや画像データなどの電子データを含む）や実験ノートに基づいて検証される為、発覚しないように生データそのものを操作できれば、外形的に研究不正を行ったという判定を行うことは極めて困難になる。

「生成 AI を利用して発覚しないように不正を行う」という意図の下で行われやすい捏造・改ざんとしては、計測により取得するはずの数値データや画像データを生成 AI によってそれらしく創り出す、という行為が想定される。特に数値データを創り出す場合には、必ずしも生成 AI を用いなくとも、自らが望んだ結果に合致するように矛盾の無い数値データを捏造することは比較的容易であり、これが生成 AI の支援によりさらに巧妙化する可能性がある。画像データの捏造については、生成 AI が創り出した写真か否かを判別するツールが存在すれば排除できるものの、チェックツールを回避するような画像を創り出せる生成 AI の誕生もあり得る。盗用については、現時点（2024 年 2 月時点）で既に盗用チェック用の Web ツール等が一般化している状況下では、今後新たに発表される論文では明らかな盗用は減少する方向に進むと推測されるものの、既存論文等の自動的な切り貼りによる文章の生成や、生成 AI に論文の作成を完全に任せることにより、著者本人による内容の確認や十分な寄与が無い論文が盗用チェックツールを回避する形で発表される可能性はある。この様に、生成 AI の不適切な利用により研究者の不正が巧妙化し、外形的には不正であると判定することが極めて困難な論文が大量に生成される可能性があることは容易に想像できる。

そもそも、研究者自身が目視で判断できないほど巧妙な捏造画像を多用し結論が歪められた論文や、既存論文を参照して文章の切り貼りにより自動的に生成された論文等が大量に公開された場合、「科学的に何の根拠も無い多くの論文」が「科学的に誠実かつ正確に表現された少数の論文」を覆い隠すことで、例えるなら俗に言うところの「悪貨が良貨を駆逐する」問題が発生する。また、生成 AI が一般化した時代では、多数派を占める低品質な論文の内容をさらに別の AI が学習することで、web 上から発信される情報の質が劣化し、専門家以外の者（非専門家）による科学的知見に関する認識の歪みが際限なく拡大する様な悪循環に陥りかねない点にも注意が必要である。

## 2. 2 生成 AI の利用によるバイアスとデータ再現性に関する問題

前項で論文執筆における生成 AI の不適切な利用に触れ、web 上の情報の劣化による非専門家の認識の歪みについて例を挙げたが、研究者自身が有効活用の範囲であると認識し



て生成 AI を利用する場合でも、この様な認識の歪みやある種のバイアスに陥らないよう、注意が必要になる。例えば、研究者が時間に追われ、論文そのもののドラフト作成を生成 AI に任せ、その検証が疎かになることで、誤った認識や偏った情報に基づく論文が発表されやすくなる可能性がある。特に、論文のイントロダクションやディスカッションの項目を執筆する場合に、生成 AI に当該分野の研究の状況や代表的な先行研究論文の選択を任せてしまうと、その生成 AI 自体の性能の問題や、学習対象や参照対象の偏りに研究者自身が気付かないまま、その論文の結論としては必ずしも間違っていないが、その研究の位置付けや既存の研究との関係性を誤認したまま論文が発表される可能性がある。

もちろん、学術論文の質はピア・レビューにより担保されるべきであり、学術誌への投稿時の査読において、前提となる情報収集の不十分さなどの指摘がなされるのが一般的ではある。さらに、仮に生成 AI を用いなかったとしても、研究者の能力や論文執筆にかかる時間の多寡により、論文の前提の認識が歪む可能性はある。しかし、論文執筆の一部を生成 AI に任せることが常態化することで、研究者側が自らの頭で論理的に思考し、検証するというプロセスを放棄しがちになり、結果的に、研究活動においてこれまで求められてきた先行研究への敬意を含む、研究者としての誠実さが失われやすくなることは確かであろう。その結果として、特定の結論が過大に評価されたり、考慮すべきマイナーな結論が無視されたりしやすくなる傾向が、研究者自身の意向とは無関係に増幅され得ることを十分に理解しておく必要がある。

その他にも、既に広く継続的に議論されているペーパーミルの様なサービスによるいわゆる「無意味な論文」の問題がある (Else & Van Noorden, 2021)。この様な商業目的の偽科学論文や、再現性の乏しい論文の結果が過剰に評価され、前項でも述べたように生成 AI がこれを助長する可能性がある。ただし、テキストマイニングによる研究結果の正否に関わる評価については、今後の AI 等の支援により正確性を増す可能性もある。例えばインフォマティクス系の（膨大な生命科学系の計測データとその解釈に基づく）実験データから導かれる結論や、あるいは実験生命科学において報告される原著論文の内容は必ずしも真実という意味での「正解」とは限らず報告例に過ぎないことを正しく判別する前提で生成 AI を高度化することが可能であれば、むしろ精度が上がることもあり得るであろう。

結局のところ、生成 AI を論文執筆に用いる場合には、生成 AI が得意とする「文章の要約」的な機能の限界を正しく理解し、生成 AI に「丸投げ」することにより生じる歪みが研究活動に与える影響を想定しつつ、研究者自身が責任を持ってテキスト等を検証することを通じて、これを適切に利用する必要がある。また、生成 AI を利用した場合にはその旨を論文上に正しく公開することも、適正な生成 AI 利用の大前提であるといえる。

## 2. 3 論文発表における伝統や「専門知」のあり方の変容

さて、研究者自身が十分な注意を払うことで、生成 AI を論文執筆に部分的に活用する

ことができ、さらに学術的に信頼のおける生成 AI（例えば学習対象が信頼のおける論文データベースに基づいており、各学協会等も公的に指定する AI）が広く利用できる環境になったと仮定した場合、論文発表に関わるどのような習慣や行動が変容するだろうか。実験生命科学分野を例とすると、以下の想定が可能である。

論文のイントロダクションの文章は生成 AI によりドラフトを作成させた上で、研究者が十分な検証を行い、AI が選択したリファレンス以外の論文を広く確認した上で、ほぼ完全に書き直したものを最終稿とする。これにより、たたき台としての文章がある方が思考しやすくなる場合には、研究者が正しく活用すれば、生成 AI の利用により従来よりも公平に幅広い背景から研究の前提を概観できる可能性がある。これは生成 AI を対話の相手として利用する方法であり、仮想的なディスカッション相手を増やすという意味で有用な面がある。

さらに、自らの論文作成過程における生成 AI の支援以外にも、発表後の論文を広く認知させるために、生成 AI が web 上の情報を学習データとして利用する際に自らの論文がいち早く学習対象となるよう、オープンアクセスやプレプリントとしての発表が今まで以上に加速化することもあり得る。後に述べる研究データのオープン化の流れにより、このような公開化の方向性は既に定まっているともいえるが、自らの論文をいち早く、広く認識してもらうことを研究者が従来以上に重視するようになる可能性があることは確かである。

また、特に実験科学分野においては、オープンサイエンスの推進に伴って大規模な AI による既存研究データの再解析・再解釈が行われる例が増加するであろう。その結果、イントロダクションやディスカッションの文章としての独創性よりも、発表される計測データやその解析処理と解釈の内容だけが重要になり、従来の原著論文の形式自体も変化し、計測データとその処理、解釈のみを記述するデータペーパーがより重要になる可能性もある。

論文発表に関わるこのような変化が起きた場合、「生成 AI によりどの様に認識されるか」という視点によって「論文の評価」のあり方も変化する可能性がある。例えば、従来の論文集合が提供する知見を AI が大規模な総説的論文データとして保持しており、新たな原著論文が報告された場合には、従来のデータセットに対してどれだけ新たな知見を加えるか、その広がりや精度を高めるかという視点で AI が判定することで、その論文の価値を評価する基準が変わり、その成果の評価も短期間に行われるようになる可能性がある。このような評価が可能になると、通常の意味での総説論文という形式は（十分に信頼のおける生成 AI を利用できる場合に限り）AI による要約が代替する時代になるかもしれない。さらに、前述の通り、論文発表の形式そのものが、生成 AI により認識され再解釈・マイニングされ易い構成へと変化する可能性もある。

信頼のおける生成 AI が利用できる環境下ではこのような想像もできるものの、研究者としては改めて、従来の研究者が世代を超えて伝えてきた専門知のあり方自体が変化することの意味とその功罪を考えておくべきであろう。仮に生成 AI を全く利用しなかったとし

ても、一定の時間内に一個人が調べられる論文の数には限界があり、さらに英語以外の言語で発表されている各国の論文の内容まで含むとすれば、分析対象が「誠実に記載された論文」に限られる限りは、AI等の自動ツールを用いた方が幅広い知識を得やすい、ということはあるかもしれない。そもそも、web検索で自分の知らなかった論文を見出しその内容を理解する、ということを生成AIにより効率良く代替できるなら、これは研究者自身の理解不足を支援する目的での生成AIの利用として有用である。この様なAIの利用は通常のwebブラウザを用いた検索操作の拡大版に過ぎないと考えれば、むしろこの様な生成AIの利用は何ら問題無いという判断もあり得るかもしれない。

ただし、この様な生成AIの効果的な利用が可能であるためには、本当に信頼でき、学協会が推薦するようなAIが存在することが前提である。また、どの様なAIを利用したとしても、2.2の項で述べた問題等により生成AIの提示する情報に過度なバイアスがかかっている可能性を認識しそれに対処出来ることが、今後の専門家として必須の能力となるかもしれない。特に実験科学分野においては、その論文の成果に対する信頼度を左右するのは実験そのものであり、信頼性の根拠として論文の報告内容を鵜呑みにせず批判的に検証・検討できるだけの専門知が、今後益々重要になることは確かであろう。

なお、生成AIの利用促進に伴って想定されるこれらの変化においても一つ重要なのは、この様な研究者の姿がそのまま、教育を通じて大学院生向けに刷り込まれてゆく可能性があることである。研究者が単に生成AIに質問し、その結果を転用しているだけのように見える行動を取り、専門家としての検証を怠っている場合には、そこから学ぶ次世代の研究者の認識はさらに甘く、弛緩したものとなる。研究者が本来有すべき検証能力や批判能力が世代を経る毎に劣化してゆく可能性をも客観的に理解した上で、今後の生成AIの正しい有効活用法を探ることは、現在の研究者に課せられた使命であるともいえる。

### 3. 生成AIの利用により想定される研究活動の変化

前項までに述べたように、研究活動における生成AIの利用には、研究不正や研究者自身の認識を気付かぬうちに歪めてしまう問題点もある一方で、論文執筆やその内容の評価という点では有効に活用可能な側面もある。特に今後、AI関連技術の急速な進歩が予想される状況下では、研究活動のあらゆる側面に生成AI等の支援が介入するのが当たり前の時代になるであろう。その様な想定の下、研究者が自発的に生成AIを利用する場合、実験科学分野の研究活動のどのような面で利用可能であり、それにより研究活動全体として何が変化するか、その功罪は抜きにして簡単にまとめてみたい。

#### 3. 1 自発的な生成AIの利用による研究者の行動変容

研究活動を「計画立案」「実験」「論文作成」「投稿・査読」に分けて考えた場合、それぞれ下記の様な行動が想定される。

#### ① 研究計画立案

- ・当該分野の先行研究の確認と新規性の予備的チェックを AI に任せる。

#### ② 実験の実施とデータ管理

- ・AI により実験プロトコルを既存研究から自動抽出し、改変して利用する。
- ・実験データの保存と分類において AI による支援を利用する。

#### ③ データ解析・図の作成・論文執筆

- ・定型化済みの実験についてはデータ解析と図の作成を AI で自動化する。
- ・イントロダクションのドラフト作成を AI に任せる（前項のとおり）。
- ・実験手法の記述（Materials & Methods 等）を AI に任せる。
- ・引用する先行研究論文のリスト化（書式修正や引用漏れの有無）を AI に任せる。
- ・日本語で記述した論文のドラフト版英訳を AI に任せる。

#### ④ 論文の投稿と査読

- ・論文投稿前の図表の確認や根拠データとの対照、盗用の有無を AI で確認する。
- ・論文査読時において不明な点を、生成 AI との対話を利用して確認する。

ここでは、AI に文章のドラフト生成を任せた後の検証と編集、改稿（リライト）などを研究者自身が責任を持って行うことを前提に、実験科学分野における生産性向上の支援として想定し得る生成 AI の利用の例を挙げた。また、研究に関わるデータ量増大への対処や、それにより研究者自身が認識不足やミスにより研究不正に陥らない為のチェック機能としての AI 利用なども含まれている。どの行動も、研究者が悪意をもって利用すれば QRP となり得る面があることから、生成 AI の利用にあたってはそれを正しく御し、節度をもって利用する姿勢が研究者に求められることは言うまでも無い。

### 3. 2 生成 AI のもたらす変化とその利点をどう生かすか

それでは、このような生成 AI 等による支援や解釈が加わることで、研究活動にどのような変化が想定されるだろうか。まず、研究活動にかかる時間の短縮、研究データを取得してから発表するまでのプロセスの効率化が期待される。特に、盗用に関する責任の問題を十分に理解した上で生成 AI を適切に利用することにより、例えば自らの執筆する文章のパターンを踏まえた論文のドラフトを短時間で創出し、これを十分な検証の下でリライトすることで、文章作成にかかる時間が減少する可能性がある（本人の文章力や英語力などにも依存するが）。その結果、これをポジティブに捉えるなら、実験デザインに関する充分な

計画立案や、データ解析・解釈に関するより深い思索など、研究の本質的価値を高める重要なプロセスに注力し易くなる可能性がある。

また、既に 2. 3 で述べたことの部分的な繰り返しにもなるが、論文発表後の AI によるデータの再解釈・マイニング等を想定するなら、特に実験科学分野における原著論文の形態と評価のあり方自体が変化することが想定される。特に、元々実験科学分野においては、（とりわけ非英語圏の研究者にとっては）記述する英文の巧拙そのものよりも実験データとその解釈の部分が重要であるが、今後はこのような傾向が加速化し、データペーパー化のさらなる進行など、論文発表の形式自体が変化する可能性がある。このことは特に、異なる言語間での論文発表のあり方にも影響を与えると考えられ、例えば英語以外の言語で書かれた論文でもデータとその解析に関わる部分のみを英語化し、AI による再解析やデータマイニングにも容易に対応出来る様な形で公開しておけば、その成果は広く認識され、評価される状況になる可能性もある。

以上の想定を踏まえ、生成 AI の利点を正しく生かす基準としては下記のようにまとめられる。即ち生成 AI を利用する上では、あくまで [A] 研究の効率化・支援のツールとしての利用に留めること、[B] 仮想的な対話の相手として利用しアウトプットに対する検証を怠らないこと、[C] その生成 AI の背景（学習対象や特徴）を理解して適切な AI を選択すること、の 3 点が重要であるといえる。

#### 4. 生成 AI の利用を前提とした成果創出において研究者の寄与をどう考えるか

さて、ここまでは主に論文作成に関わる生成 AI の問題点に加え、仮に積極的に利用した場合の利点等を含め、論文自体の内容についてフォーカスしてきたが、ここからは生成 AI の利用と研究者の実質的寄与に関して問題となり得る要素について述べてみたい。

##### 4. 1 研究者自身が表現すべき事実／新規性／寄与とは何か

特に生命系の実験科学分野の論文に求められるものとして、論文の構成要素は下記のようにまとめられる。まず、イントロダクションで適切な範囲の先行研究に言及しているか、結果では実験を分かり易く端的に説明し合理的な解釈と時系列情報を提供し、結果を誤解無く正しく伝え得る図表を用いているか、実験手法等の項では実験内容とその手法や試料の情報が過不足なく示されているか、が重要である。ディスカッションでは実験結果の総括とこれに基づく合理的な推論に加え、既存研究との比較で不足している視点や想定される今後の展開などが記載されていることが望ましい。つまり原著論文においては当然ながら、実験デザインと結果・その解釈が正しく伝わるのが最も重視されることになる。

それでは、原著論文の著者が表現すべき事項と、その創出に関わる寄与をどう考えるべきだろうか。まず実験科学分野の原著論文で最も重要な「表現すべき事実」とは「実験事実 (Experimental fact)」であり、これを正確に伝達することが求められる。そしてその



データの示し方は、その分野で広く認められる常識に従う必要がある。さらに、原著論文では新規性が問われることから、ここで「表現すべき新規性」とは、先行研究との関連性を適切に把握し表現した「実験デザイン・結果の解釈手法・解釈結果・推論」等の新規性であるといえる。また、原著論文においては特に近年、その研究への寄与を正しく記載することが求められることから、これらの実験事実と新規性を発見し発表に至るプロセスにおいて、「著者らの具体的な寄与と責任の範囲」が論文内に明示されていることが望ましい。このことを前提に、生成 AI の利用とその原著論文発表に関わる著者の寄与との関係について考えてみよう。

#### 4. 2 英文校正等の既存サービスとの比較と生成 AI の適正利用の範囲について

研究計画の立案から原著論文の執筆、発表に至るプロセスのどこに、どの程度の寄与が研究者に求められるかという判断は、時代により異なるといえる。生成 AI を容易に利用可能な状況下で研究者に求められる研究上の実質的な寄与とは何だろうか。

例えば我が国では、実験生命科学の分野では英文校正が広く利用されていることから、そもそも研究者に対して **Native speaker** に近い正確な英文構成能力を求めている訳では無いことは確かであろう。また、英文校正ほどでは無いにせよ、英文への翻訳サービスなども利用されており、研究者自身が研究の効率化を目的としてこれらのサービスを用い、他人の力を借りる事は許されている。その他、実験データの解析という面でも、特に各種オミクスデータ（生体分子の網羅的な情報）の取得や解析の外部委託は既に一般化したサービスとして用いられており、時にはデータ正確性の確保のために、あえて外注による解析を選ぶ場合もある。さらに、これらの外部委託などのサービスを利用した場合には、少なくとも現在ではできる限り、論文中にその旨を明示することが求められている。

これらの英文校正や翻訳の委託や web ツールによる英文翻訳、外部への委託解析等と生成 AI の利用では、何が違うのだろうか。少なくとも、研究成果の公表に向けた活動を効率化することを目的として、研究者以外の力を借りる行為であることは確かである。そういう意味では、生成 AI の利用についても、論文中に明示的にそれと分かる様に述べている限り、その利用には何ら問題無いという考え方があろう。

一方で、最も大きな違いとしては、「研究者側の寄与の度合い」と「検証可能性の有無（もしくは容易さ）」を挙げることが出来る。英文校正や翻訳においては、作業前の文字列は通常、研究者自らが一定の寄与の下で作成したものであり、校正・翻訳作業前の内容と作業後の内容の比較検証は容易に可能である。委託解析についても、サンプルは研究者側が用意したものであり、十分なサンプル量があれば再検証が可能である。しかし、生成 AI を用いた論文本文の執筆の場合は、研究者側は質問を投げかけるだけであり、自らの十分な寄与の下でその文字列を作成したというプロセスが欠けている他、生成 AI がその様な文章を作成するに至った根拠も不明確であり、バイアスの有無等を容易には検証できない。

このような生成 AI の利用上の問題は将来的に改善され、検証が容易になり信頼性が向上する可能性もあるとはいえる。しかし一例として、AI が構築した文字列がたまたま他人の発表済みの論文と同じ文章になっていた場合には、AI を利用して後から発表した研究者個人の責任が問われる（生成 AI に責任を取らせることが出来ない）ことから分かる様に、どれだけ生成 AI が進化しても、執筆者自身がその文章表現の本質的部分を提供し（研究者自身の充分な寄与があり）、さらに正確さや妥当性を、専門知をもつ者としての立場から検証できない限りは、英文校正と同様に扱うことは難しいと判断せざるを得ない。

このことは、4. 1 に挙げた原著論文に求められる「実験事実 (Experimental fact) を正確に伝達する」という最重要な品質管理を担保するために必要な、「著者らの具体的な寄与と責任の範囲」をいかに明示するか、ということでもある。結局のところ生成 AI の利用にあたっては、研究者自身が主で生成 AI が従である、という関係性を超え、便利さに流されて生成 AI に頼ってしまうことが問題であり、その様な異常な状態を回避する為に、「研究者側の充分な寄与」と「検証可能性」という基準に立った生成 AI の利用と、その旨の明示的な説明が必須とされていると考えるべきである。そしてこの基準は、3. 2 で挙げた生成 AI の利点を正しく生かす為の基準 3 点とも完全に合致している。

なお、このような生成 AI の利用に関するルールは、論文を掲載する側である学術雑誌の出版社から明確に指定され始めている。例えば日本分子生物学会の学会誌である **Genes to Cells** 誌の投稿規定には“Use of Generative AI”の項に、英文校正や文章のブラッシュアップのための利用は Acknowledgements 欄に利用の旨を明記することを条件として認めるものの、生成 AI を用いて文章自体を創り出すことは認めておらず、論文中の図 (Figure) についても、概念図としてのイラスト作成以外での生成 AI の利用を認めていない。

## 5. AI 支援の一般化を前提とした論文発表と研究の評価について

最後に、生成 AI による研究活動支援ツールが今後もさらに洗練され一般化するであろう状況を踏まえ、今後の論文発表に関して想定される新たな枠組みの例や、研究者の評価において想定されるべき要素の例を、あくまで筆者個人の考えとして示したい。

### 5. 1 学術誌や論文発表のあり方に関する問題について

先に述べたように、学術雑誌の投稿規定においては、既に生成 AI の利用に関するルールは明示されつつある。その他、近年は多くの学協会などでも、生成 AI の利用に関するポリシーを制定し公開している例が増えている。例えば日本教育工学会では「生成 AI 等の利用開示と著者の説明責任」の項において、生成 AI が責任を負う主体となり得ないこと、生成 AI の利用者が説明責任を負うこと、発表の際には何の生成 AI を何の目的で利用したのかを明示すること、等の基準を公開している（日本教育工学会、2024）。

一方で、生成 AI による判断結果の検証可能性等を含めた今後の進歩や、学術界からみ



ても十分に信頼できる統合的な AI サービスが広がる可能性を踏まえるならば、生成 AI の効果的な利用を図りつつ研究成果の信頼性を向上する為に想定される論文発表上の対策や新たな論文発表の形態として、以下の例が挙げられる。

まず、将来的に高度な生成 AI が自動的に、専門家から見ても十分な信頼性をもつ総説論文や部分的な解説的文章(キャプションなど)を創出する時代が来ることを想定すると、学術誌側にも研究成果の信頼度向上を目的とする新たな論文発表形式が期待される。例えば、既存研究の解説またはシステマティックレビューとしての総説論文以外に、既存の研究により成り立っている科学的知見を総説的に概観し、その本質的要素を実験的に再検証する様な「検証論文」の発表を、単に新規性の観点とは別の枠組みで評価し、掲載するなどの枠組みが考えられよう。

さらに、生成 AI の利活用を前提とした学術誌自体のあり方としては、原著論文の発表に付随して雑誌者側で新たな項目を追加したり、従来とは異なる枠組みで原著論文を公開したりするなど、従来の常識にとらわれない多様な方策も検討されるべきである。

例えば、発表される原著論文に対し、あえて生成 AI を用いた解説項目を雑誌社側が独自に追加する、という方策が考えられる。あくまで雑誌社側や関連学会が推奨する生成 AI などが形成されてからの話にはなるものの、これにより著者が限られた文字数の中で説明しきれない背景等の情報が自動的に補完されることになる。あえて生成 AI に一任する形で解説項目を論文本体と同時に提示する枠組みは、読者に対して当該論文に関連する幅広い知見を提供する効果がある他、論文発表時点での当該研究の先進性や独創性を含め、論文発表における競争的側面を公平かつ客観的に管理する上で役立つかもしれない。

その他、学術雑誌においてオリジナルデータを報告する為の様式として、既存の原著論文やデータペーパーの枠組みの延長に、多様な論文発表の様式を検討することもあって良いだろう。例えば、定型化できる実験手法(いわゆる **Materials & Methods** 等の内容)の記述はあらかじめ所定のフォーマットに数値や固有名詞を入力することで提供し、原著論文としての盗用チェックの対象としない、等の方策もあり得る。

また、近年の翻訳 AI のめざましい発展状況を活用し、非英語圏の研究者が英語表現の類似や借用により盗用に陥ることを根本的に回避する方策として、例えば日本人研究者が原著論文を発表する際に日本語での論文発表を可能とし、英文への翻訳は自動翻訳ツールで提供する(ただし図表やデータは最初から英語版として提供する)、などの発表形態をもつオープンアクセス誌があっても良いかもしれない。少なくとも実験データとその解析結果について広く英語圏を含めて流通する枠組みがあり、論文発表が英文誌でなければならぬとする前提を無くすことができれば、海外のマイナーなオープンアクセス誌に掲載料を払って論文を掲載するよりも低コストで、本当に重要な実験データを早く安全に公開できるかもしれない。これはかなり空想的な提案ではあるが、そもそも我が国の研究者にとって有利に、低コストで論文成果を正確に発表できる枠組みを提供するという視点に基づ

くなら、今後はあえて日本語での発表という選択肢も検討の余地があるのではないかと思考する次第である。

## 5. 2 今後の実験科学分野の研究活動で評価されるべき点について

現在、特にいわゆる西側先進諸国を中心としたオープンサイエンスの推進が進められており、我が国でも学術論文の即時オープン化と同時に、そのデータを AI 等による再解析・再解釈やデータマイニングなどに活用出来る様に、根拠データなどもできる限り広く公開する旨の方針が策定されている（内閣府，2024）。具体的な制度整備については機関単位で進行中だが、少なくとも今後の原著論文の発表においては、その論文の根拠データを可能な限り公開し共有する流れになっている。そもそも、仮に生データを（別の研究で解析中などの理由により）秘匿したとしても、その論文で述べられている結論は将来的に AI によるテキストマイニングの対象となることは避けられない。この様な状況下では、原著論文に最も期待されるのは、高い再現性と誠実な論考の提供である。

大学等に所属する研究者は特に近年、その研究活動を定量的に評価されており、例えば時間に追われた結果として不十分な解析に留まる論文の発表や、論文数を稼ぐためにいわゆる「サラム論文」的な論文発表を行う等の行為が生じやすい環境にある。今後、AI を活用した大規模なテキストマイニングやデータマイニングを通じ、オープンサイエンスの推進を真に有効なものとするには、従来よりも高い再現性の担保や、再解析に適したデータセットの提供にも配慮する必要がある。この様な背景の下、研究者には原著論文の質の向上のみならず、「信頼度の低い論文の大量生産は害悪である」ことの十分な理解が求められていることを、我々は改めて認識すべきである。また、研究者を評価する大学や研究機関の側にも、評価対象期間や評価方法の再検討などを通じ、否定的な結果を報告する論文や検証的な研究論文等を含めて広く評価の対象とするなど、研究活動の評価指標の改善と多様化を進めることが求められているといえよう。

## 6. おわりに

本章はあくまで、生成 AI に関する詳細な専門知識を持たない一般的な実験生命科学者からのボトムアップ的な視点で、研究活動への生成 AI の利用をイメージし、特に研究公正の観点から生じ得る問題に着目し、現時点で想定される利活用方法と留意点等について広く概説したものである。生成 AI は強力なツールではあるが、研究活動に利用される場合には、限界や危険性を十分に理解した上でその特性を活用し、研究者自身を主とする明確な意図や目的の下で利用していることを研究者が自覚することが重要である。また、自覚的に生成 AI を用いたとしても研究者側が気付かぬうちに従来の研究活動のあり方や競争的環境そのものを歪める可能性があることも想定し、研究を取り巻く状況を継続的に概観し、その健全性を保つ努力も必要となる。

生成 AI の技術開発とその実用化や社会実装は加速度的に進んでおり、研究者にとって十分な信頼性をもつ研究支援 AI ツールの創成なども不可避免に進むものと思われる。本章では主題から外れるために明確に取り上げなかったが、学術論文のように公開される文書ではない、例えば研究費申請書などについては、外から見えないが故に、生成 AI 等に過度に依存した文書作成が今後行われる可能性もある。実際のところ、研究活動において研究者自身の能力や具体的な寄与、成果を正当に評価するという前提で、生成 AI をどの様に利用することが許され、何が不適切なのかについては、ある意味で QRP に含まれる論点であり、一概に明確な答えを出す事は困難である。生成 AI の利用による「それらしい嘘」の蔓延など、明らかな情報の歪みも問題であるが、今後はむしろ、研究活動そのものに与える影響を含め、生成 AI の妥当な利用の範囲に関する議論こそ、より深められるべきであろう。

## 【参考文献】

内閣府（2024）「学術論文等の即時オープンアクセスの実現に向けた基本方針」（統合イノベーション戦略推進会議 令和6年2月16日決定）の実施にあたっての具体的方策（最終アクセス日：2024年10月30日）

[https://www8.cao.go.jp/cstp/openscience/r6\\_0221/hosaku.pdf](https://www8.cao.go.jp/cstp/openscience/r6_0221/hosaku.pdf)

日本教育工学会（2024）「一般社団法人日本教育工学会における生成 AI 利用に関する基本的な考え方（2024年5月18日）」（最終アクセス日：2024年10月30日）

<https://www.jset.gr.jp/news/news-4587/>

文部科学省（2014）「研究活動における不正行為への対応等に関するガイドライン（平成26年8月26日、文部科学大臣決定）」（最終アクセス日：2024年10月30日）

[https://www.mext.go.jp/a\\_menu/jinzai/fusei/index.htm](https://www.mext.go.jp/a_menu/jinzai/fusei/index.htm)

Else H., & Van Noorden R. (2021). The fight against fake-paper factories that churn out sham science. *Nature*, 591 (7851), 516-519.

<https://doi.org/10.1038/d41586-021-00733-5>

Genes to Cells, Author Guidelines (n.d.). “Use of Generative AI”（最終アクセス日：2024年10月30日）

<https://onlinelibrary.wiley.com/page/journal/13652443/homepage/forauthors.html>

Steneck, N. H. (2006). Fostering integrity in research: definitions, current knowledge, and future directions. *Science and Engineering Ethics*, 12 (1), 53-74.

<https://doi.org/10.1007/pl00022268>



## 第6章 中高の現場における生成AIの現状と課題

南里 翔平

(市川中学校・高等学校)

### 1. はじめに

本章は2024年2月16日に東京大学を会場として開かれたAPRIN（全国公正研究推進会議）で報告した内容を基に、会議後に進展があったものを一部加えてまとめたものである。APRIN2023年大会は生成AIをテーマに各分野で議論が行われた。しかし前年度までは生成AIがメインテーマに扱われたことはなく、APRINの議題を見ても、2023年は私たちの社会にとってのターニングポイントと言っても過言ではない。

中学校や高校の現場では、2012年の中央教育審議会の答申において「アクティブ・ラーニング」の言葉が使われ始めてから、授業内外の様々な活動においてタブレット端末等のICT機器の利用が推進されている。中央教育審議会の答申から約10年経って、デバイスやアプリケーションの選択、授業の構成といった対応が、学校毎に、生徒の発達段階や習熟度に併せてできるようになりつつある。

さらに文部科学省はICT機器所有の有無によって生徒や学校、地域毎の教育格差を生まないために、生徒1人1台端末を目標とした「GIGAスクール構想」を2018年度から進めている。この影響もあり、現在は多くの学校において生徒1人ずつが自分の端末を保有できている。ちょうど2020年から2021年頃には新型コロナウイルスが世界的に流行し、一定期間の休校を余儀なくされた学校が多くあった。生徒1人1台端末を確保できていた多くの学校ではオンライン授業などを行い、コロナ禍でも教育を停めない対応ができた。しかし、1人ずつが端末を保有しているからこそ、新しく台頭した生成AIは中高生にあっという間に浸透してしまっている。いわば黒船的に現れた生成AIに対して、中学校や高校の現場でどのように付き合ったらよいかを考えるために、2023年の現状を報告する。

### 2. 生成AIをめぐる教育現場の現状

#### 2. 1 ChatGPTの公開と国際的な動き

OpenAIにより2022年11月30日に「GPT-3.5」をベースとしたChatGPTが公開され、また2023年3月14日に「GPT-4」が公開された。同時期にGoogleのBard（現Gemini）やMicrosoftのBingも公開され、この頃から我々の日常で生成AIという言葉聞く機会が一気に増えた。

このような流れの中で2023年5月に開かれたG7広島サミットでは、「我々は、急速な技術革新が社会と経済を強化してきた一方で、新しいデジタル技術の国際的なガバナンス

が必ずしも追いついていないことを認識する。技術進化が加速する中、我々は、共通のガバナンスの課題に対処し、世界的な技術ガバナンスにおける潜在的なギャップや分断を特定することの重要性を確認する。(中略)我々は、関係閣僚に対し、生成 AI に関する議論を年内に行うために、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 プロセスを創設するよう指示する。」(G7 広島サミット, 2023) との G7 首脳声明が出され、生成 AI に関する国際的なルールの検討を行うために、広島 AI プロセスが立ち上げられた。

広島 AI プロセスでは生成 AI に関する国際ガバナンスを形成するために、「高度な AI システムを開発する組織向けの広島プロセス国際指針」と「高度な AI システムを開発する組織向けの広島プロセス国際行動規範」が策定された。この国際行動規範は安全、安心、信頼できる生成 AI を世界に普及させるための、開発者向けの行動規範である。この規範の中では生成 AI の開発組織に対して、「高度な AI システムの設計、開発、導入、利用をカバーするために、必要に応じて、ライフサイクルのすべての段階にこれらの行動を適用すべきである。」(広島 AI プロセス, 2023) としている。国際行動規範は 11 項目からなり、9 番目の項目では、世界の課題のうち、気候変動や教育等に対処するため、高度な AI システムの開発を優先することが求められている。そして国際行動規範は開発者に対して、学生や労働者を含む一般市民への教育と訓練を促進すべき、としている。このような国際的な動きを背景として、教育現場における生成 AI に関する議論が急速に求められるようになった。

## 2. 2 国内の動き

ChatGPT の公開や G7 サミットなどの国際的な動きを受けて、日本国内の教育関係者の間でも生成 AI に関する議論が始まった。本節では文部科学省と各大学の動きについて述べる。

### 文部科学省の対応

文部科学省初等中等教育局は G7 広島サミットの首脳声明を受けて、2023 年 7 月に「初等中等段階における生成 AI の利用に関する暫定的なガイドライン」を策定した。このガイドラインは生成 AI の開発が飛躍的なスピードで進展していく中で、教育現場においても、児童や生徒に対して一定の配慮をした上で、様々な活用の可能性を考えるための参考資料としてまとめられたものである。ガイドラインの位置づけとしても「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である(一律に禁止や義務づけを行う性質のものではない)。」(文部科学省, 2023) としている。

ガイドラインは、1. 本ガイドラインの位置づけ、2. 生成 AI の概要、3. 生成 AI の教育利用の方向性、4. その他の重要な留意点、の全 4 節からなる。このうち第 3 節では初等中



等教育の中で生成 AI を利用する際の基本的な考え方が述べられている。学習指導要領では「情報活用能力」を学習の基盤となる資質・能力と位置づけ、その技術を学習や日常生活に活用することの重要性を強調している。生成 AI の仕組みを理解することは、近い将来それを活用していく世代にとって重要だからである。一方で生成 AI は発展途上の技術であり、著作権侵害や偽情報の拡散など、このほかにも多数の潜在的なリスクが指摘されている。そのため、教育現場において生成 AI を活用していく際には、児童生徒の発達段階を十分考慮することを求めている。これらを踏まえ、生成 AI の利用に関する適切な判断を行うためには教師の側にも一定の AI リテラシーが求められる、としている。

また、ガイドラインの中では AI 時代の教育において忘れてはならないこととして、「真偽の程は別として手軽に回答を得られるデジタル時代であるからこそ、根本に立ち返り、学ぶことの意義についての理解を深める指導が重要となる。また、人間中心の発想で生成 AI を使いこなしていくためにも、各教科等で学ぶ知識や文章を読み解く力、物事を批判的に考察する力、問題意識を常に持ち、問を立て続けることや、その前提としての「学びに向かう力、人間性等」の涵養がこれまで以上に重要になる。」（文部科学省、2023）と述べられている。生成 AI は人類が新たに獲得した道具の一つであるから、それが台頭しても、教育を通して生徒に何を伝えたいのかという視点を指導する側が忘れてはならない。

一方で、ガイドライン中でも述べられている通り、中学生や高校生のスマートフォン所有率は 2022 年時点で 9 割を超えており、児童・生徒が生成 AI にふれる機会は多く、利用規約で制限されているに関わらず、ゲーム感覚で生成 AI を利用してしまう中高生は多い。実際に筆者の周辺でも、授業内の考察課題の解答を生成 AI に求めようとする高校生が現れており、今後このような流れには歯止めが効かなくなるものと思われる。新しい技術に対して大人が慎重に対応しようとしても、中高生の動きには追いつけないのが実際ではないかと思う。

文部科学省はこのガイドラインを踏まえ、生成 AI を活用した教育活動や校務のパイロット的な取り組みを行う「生成 AI パイロット校」を指定した。パイロット校は全国の公立学校から 2023 年度に 52 校、2024 年度に 66 校指定されている（文部科学省、2024）。これらの学校では各種生成 AI ツールの利用規約を遵守した上で、生成 AI の性質やメリット・デメリット、AI には自我や人格がないこと、生成 AI にすべてを委ねるのではなく自己の判断や考えが重要であること、などを事前に十分理解させた上で教育活動への利用を進めている。また、個人情報や機密情報の保護に細心の注意を払いながら、働き方改革の一環として校務での活用も進められている。

2023 年度のパイロット校には実証内容分析のために、2024 年 3 月 8 日～3 月 14 日にアンケートが実施された。教育利用に関するアンケートのうち、「教育活動に生成 AI を利用する中で、児童・生徒に対して最も当てはまると思う内容」はどれか、という質問では、学習意欲が低下した、創造性が低下したという回答はほとんど存在しなかった。しかし、



生成 AI の回答をうのみにしてしまうケースや、仕組みを理解せずに使用してしまったケースは一定数存在したことから、ファクトチェックの習慣づけを行うことが重要であると指摘されている。また、生徒の発達段階に応じて生成 AI の仕組みについて学習する機会を用意することが必要であることも指摘された。一方で、生成 AI を利用すると自分の思った回答が必ずしも得られるわけではないため、その点で試行錯誤することが、学習意欲や創造性の育成につながっているのではないかと、との考察もされている。

校務・活用に関するアンケートでは、パイロット校として生成 AI を活用することで、その懸念点や利便性を教員が理解することができた反面、使用に慣れるにはさらに一定の時間がかかるとの結果が得られている。校務での活用例としては、保護者向け文書などの定型文の作成や、教材や練習問題のたたき台作成に活用できるとのことである。実践は自ら行うが、アイデアやヒントは教員コミュニティ内で周知し、それを共有するための研修を行っていくことが重要であると指摘されている。

生成 AI パイロット事業は 2023 年度までの成果として以下のようにまとめている。成果としては、生成 AI のうち、約款型で提供されている汎用基盤モデル(例:ChatGPT, Gemini, Claude など)を使用する中で、教育現場で活用可能なプロンプトや生成 AI を活用した各教科での実践例が生まれつつある、としている。また現状の課題として、生成 AI に入力した個人情報生成 AI に学習されるリスクがあるなどの課題があるが、実証環境が未整備である点や、パイロット校以外からも様々な実証事例が存在していることから、成果や課題を検証していくにはパイロット校以外から創出される事例も収集する必要があることなどが挙げられている。

このガイドラインに関しては、2024 年度にも検討会議が進められており、8 月～9 月には有識者からのヒアリングを行っている。有識者の 1 人である国立情報学研究所の新井紀子氏は、生成 AI が教育に及ぼす影響について報告を行っており、生成 AI の出力を教育現場で「正しく活用」するには極めて高度なファクトチェック力が求められるが、現在の児童生徒の読解力ではそれは極めて困難であることを指摘している(新井紀子, 2024)。また、教員が出力のファクトチェックを代行するとなると教員の多忙さに一層の拍車がかかることから現実的ではないとし、教育目的で生成 AI を「正しく活用」することは現実的には不可能であると論じている。

ガイドラインに関して、今後は 2024 年の 9 月～11 月に「暫定的なガイドライン」の改定案の検討が行われ、秋～冬頃にガイドライン改訂版(ver.2.0)が取りまとめられる予定である。

## **大学の対応**

OpenAI により 2022 年 11 月 30 日に「GPT-3.5」をベースとした ChatGPT が公開されたことを受け、東京大学の太田邦史理事・副学長は、東京大学の HP 上で 2023 年 4 月 3

日に「生成系 AI (ChatGPT, BingAI, Bard, Midjourney, Stable, Diffusion 等) について」と題した声明を発表した (太田, 2023)。

声明の中では生成系 AI の仕組みを踏まえた懸念点を示しつつ、社会に対する影響について、2023 年時点の法律や制度では対応できないことなどを指摘している。このことから失業者の増大や産業構造の変化、社会の階層化の進行など、さまざまな悪影響が生じる可能性を示唆している。しかし、イタリアやアメリカのように、生成系 AI を有害な存在として利用禁止するだけでは問題は解決せず、また秘密裏な開発により、地下で悪用される可能性があることを示した。このことから、どのようにしたら問題を生じないようにできるのか、その方向性を見出すべく行動することが重要であるとしている。そして、東京大学の構成員に対しては、なぜこのタイミングで大規模言語モデルに創発が起きたのか、その原因を考察したり、良い利用方法や、新しい法制度、社会システム・経済システムを開発したりしていくことを求めている。

生成 AI については東京大学以外にも、京都大学、東北大学、慶應義塾大学、早稲田大学など、国内の多くの大学から声明が出されているが、いずれも利用しながらその利用法を考えることを求めているものが多い。

### 2. 3 生成 AI の利用規約

生成 AI の利用規約は未成年者に対して利用制限を設けているものが多い。例えば ChatGPT を運営している OpenAI は、利用者の最低年齢を 13 歳以上とした上で、18 歳未満の場合には親権者又は法定後見人の許可を得る必要があるとしている。Google 社の Gemini (旧 Bard) も、使用する際には 18 歳以上である必要があるとしている。Microsoft の Bing も成人または法的責任に足る年齢に達している者の使用に制限している。このことから、日本の中学生や高校生が自分でアカウントを作成して生成 AI を利用することは、規約上は難しい。しかし先述の通り、中学生や高校生の生成 AI 利用には歯止めが効かず、教育関係者には利用させながら利用法を考えていくことが求められる。そのため、生成 AI パイロット校を始めとした、教育活動で生徒に生成 AI を利用させることを想定している学校では、予め親権者の同意を得ている場合がある。しかし、教員によっては未成年者が生成 AI を利用する際に年齢制限があることを認識していないこともあり、文部科学省を始めとした関係機関から、その周知が図られることが望まれる。

### 3. 中学校・高校の現場における利用例

2023 年頃から学校現場での利用が始まった生成 AI であるが、約 2 年弱の間にも様々な実践例が現れつつある。私たちにとって生成 AI は突然現れた優秀な助手のようなものである一方で、新しい価値は創造してくれない。そのことを理解していれば、過去の事例検索や校務負担の軽減などに強力な手助けをしてくれる。

### 3. 1 生成 AI パイロット校での利用例

文部科学省（2024）では、生成 AI パイロット校での取り組みとして次のようなものが紹介されている。大阪市立高殿小学校では情報の授業の中で AI に関する基本的な知識を確認した上で、生成 AI により作成した記事と、実際の記事を比較して、AI の活用法を学習させた。児童からの反応として「インターネット上の情報をすぐに信じるのではなく、資料や自らの経験と照らすことが重要だ」との感想が得られたことから、生成 AI に対するファクトチェック力を養うためにも、生成 AI が活用できることがわかる。

宮城県岩沼市立岩沼北中学校 3 年生の英語の授業では、生徒が作成した英文を生成 AI により添削させる取り組みを行っている。また、音声入力により生徒自身の発音を確認させることもでき、生成 AI を活用してよりよい英文の作成や発音ができるようになっている。茨城県立竜ヶ崎第一高等学校では情報の授業で Python を用いたアプリケーションの作成時に生成 AI を活用してコードの作成を行っている。プログラミングコードも含め、言語の翻訳は生成 AI が得意とするところであるから、パイロット校以外でも利用が進むと良いのではないか。

### 3. 2 市川中学校・高等学校での使用例

筆者が勤めている市川中学校・高等学校は千葉県市川市にある私立の中高一貫校である。生徒は中高合わせて約 2300 人おり、教職員は約 250 人いる。2023 年 12 月に教職員を対象に生成 AI を使用しているか調査を行った。結果は生成 AI を使用していると回答があったのは 15 名前後で、使用しているのは ChatGPT がほとんどであった。主な用途は翻訳、要約、語学壁打ち、Excel などの関数作成、数式の解答、数学の作図（Python や TeX のコード作成）、論文検索、小論文指導などであった。以下に使用例のいくつかを紹介する。

#### TF 問題の作成

英語科の教員より、TF 問題（○×問題）の作成に生成 AI を使用しているとの回答があった。筆者は担当教科が理科（地学）であり、ChatGPT に「火山に関する TF 問題を作成して」とお願いして作成してもらった例を表 1 に示す。このような問題を簡単に作成してくれるので校務の負担は軽減される。しかし内容を確認すると「4. 火山は決して眠っていない」は、現在は定義として使われていない休火山のことを指していて、解答は T（正答）を求めているのか、活動を止めた火山もあるので F（誤答）なのかよくわからないなど、そのまま生徒に出題するのは問題があるものもあるので、内容の確認が必要である。

表 1 ChatGPT (GPT-3.5) で作成した TF 問題の例

<p>テーマ：火山</p> <ol style="list-style-type: none"> <li>火山は、地球上の断層帯の上に位置している。 ・正しい / 誤り</li> <li>火山は、溶岩や火山灰の噴出によって形成される地形である。 ・正しい / 誤り</li> <li>火山ガスには、主に窒素や酸素が含まれている。 ・正しい / 誤り</li> <li>火山は常に火山活動を示しており、決して眠っていない。 ・正しい / 誤り</li> <li>火山噴火は、周囲の地域に破壊的な影響を及ぼす可能性がある。 ・正しい / 誤り</li> </ol>
--

### 論文タイトルの思案

2023 年は関東大震災の発生から 100 年の節目であった。これをきっかけにして、校内の生徒・教職員の有志を募って「関東大震災ゼミ」を開いた。表 2 はそのゼミでの活動を生徒が外部で発表するための要旨を作成する際に、文章の要約を ChatGPT にやってもらった例である。生徒が作成した文章を ChatGPT にそのまま入力し、「1,000 字程度で要約して」とお願いすれば、表 2 のような要約文がたちまちに返ってくる。しかし作成された文章を見ると、ゼミの参加者として感じたこと、あるいは発表者として伝えたいことは文章に織り込まれていないので、ChatGPT が作成した文章を参考に、改めて生徒が要約文を作成した。このように ChatGPT を利用して生徒の思考を深めることができた。

また、発表のタイトルも ChatGPT に考えてもらった (表 2)。回答されたものに納得がいかないときに、何度でも回答を要求できるのは機械の良いところである。ChatGPT が用意した回答はどれも物足りなさを感じたため、最終的にはこれらを参考に、生徒が「発生から 100 年の関東大震災を多角的に捉える」というタイトルを考案した。このように、生成 AI にアイデアを出してもらってそれを人が吟味しより良いものを生み出すという使い方は有効であると思う。

### 大学入試の志望理由書作成

大学の総合型入試を受験したい生徒の志望理由書の作成指導に ChatGPT を利用した。この生徒は志望理由書を書くための考えがまとまらず、なかなか書き出せずにいた。そこで、生徒の前で ChatGPT に表 3 のように質問を入力し、志望理由書を作成してもらった。すると ChatGPT がずらずと文章を作成する様子を見た生徒は、「私、こんなに書けません」と圧倒されていた。ChatGPT が作成した文章をよく見ると、志望校の HP に書いて

ありそうな文言が見られるので、よく調べて書かれた文章のように見える。しかし表 3 の文章には作成者自身の経験に基づく具体的な内容がないため、生成 AI によって書かれた平均的な文章であると指導している教員にはわかる。そこで ChatGPT が作成した文章の至らない点を示しながら、機械的に書かれた文章だからこそ文章の構造はしっかりしているため、表 3 の文章の骨組みを参考に自身で書くことを生徒に勧めた。

表 2 ChatGPT (GPT-3.5) による要約とタイトルのアイデア出し

You

つぎの文章を 1000 字程度で要約してください

2023 年は関東大震災の発生から 100 年の節目の年であった。関東大震災は 1923 年 9 月 1 日午前 11 時 58 分に神奈川県西部を震源として発生した大正関東地震（気象庁，2023）に起因した災害の総称であり，死者約 10 万 5000 人（内閣府，2006）と甚大な被害を及ぼす未曾有の大災害であった。この災害では火災や津波といった地震による被害だけでなく朝鮮人，中国人の虐殺などの事件など，社会全体に影響を及ぼした。（全文約 2300 字，以下略）

ChatGPT

2023 年は関東大震災発生から 100 年の節目であった。市川学園市川中学校・高等学校では「関東大震災ゼミ」を開催し，生徒と教員が地震動，気象，火災，医療，交通，朝鮮人虐殺などのテーマで議論を行った。研究目的は，高校生が多角的な視点から関東大震災を学ぶことで，今後の震災発生時の行動や防災意識に活かすことである。テーマごとにスライド発表と被災地巡検を行った。新聞記事や建築基準法の変遷にも着目した。災害教育の重要性を認識するとともに，学校での教育や災害伝承碑の拡充が必要だと考えられる。

You

いま要約した文章にタイトルを付けてください

ChatGPT

「災害 100 年後の挑戦：関東大震災ゼミからの洞察」

You

他の候補はありますか

ChatGPT

1. 「関東大震災 100 年後：学びと課題」
2. 「未来への指針：関東大震災の遺産と防災意識」
3. 「記憶と教訓：関東大震災ゼミからのメッセージ」

4. 「震災の影響と変革：学びと考察」
5. 「防災への架け橋：関東大震災ゼミの知見」

表 3 ChatGPT (GPT-3.5) による大学志望理由書の作成

You

日本の 18 歳の女子高校生として、筑波大学生命環境学群地球学類の志望理由書を 600 字以内で書いてください。

ChatGPT

もちろん、お手伝いします。以下が、架空の 18 歳の女子高校生の方が筑波大学生命環境学群地球学類を志望する理由を 600 字以内でまとめたものです。

私は自然と環境に興味を抱き、その中で人間と自然の調和を追求したいと考え、筑波大学生命環境学群地球学類を志望いたします。

高校生活を通じて、地球環境の変化が私たちの生活に与える影響に関心を抱きました。気候変動、生態系の変化、そして持続可能な開発の必要性が明確になる中、私はその中でどのようにして人間が自然と調和していくべきかを深く考えるようになりました。筑波大学の生命環境学群地球学類は、その課題に真剣に向き合い、未来に向けた解決策を模索する場所だと感じました。

筑波大学はその独自のカリキュラムや研究環境で知られています。特に、地球学類は地球環境に関する包括的な知識を提供しており、その中で私が追求したい環境問題に焦点を当てることができます。また、優れた教育体制や充実した研究施設が学びの深化を支えています。

私は生命環境学群地球学類で学び、地球上のさまざまな環境課題に取り組むことで、将来、地球環境問題に対する解決策を提案し、社会に貢献したいと考えています。筑波大学で培われる専門知識と実践力を身につけ、リーダーシップと協働のスキルを発展させ、地球環境の持続可能な未来を築く一翼を担うことが私の目標です。

私はこの志望理由書を通じて、筑波大学の地球学類に入学することで自らの目標を追求し、社会に貢献できる人材に成長することを約束いたします。どうぞよろしくお願いいたします。

大学の総合型入試を受験する場合には、その分野を研究していくことの意欲と、それを裏付ける経験を、志望理由書や口頭試問によって鮮やかに語れるかを問われていると考えている。この生徒にはまず字数制限を考えずに文章を膨らませてもらい、例えば、どのような授業を受けたときに気象学を学びたいと思ったのか、自分が行った研究によってどの



ような事がわかったのか、何が大変だったのか、研究活動のどこに面白みを感じているのかなどを、具体的に回答できるようになるまで添削と面談を繰り返した。最終的に 600 字の制限に収めるために言葉を吟味してまとめた文章が表 4 である。

表 4 生徒が作成した志望理由書

私は、大気の局地的かつ突発的な現象を研究するために筑波大学地球学類を志望する。

幼い頃は、頭上に広がる空と宇宙の境界や、天気が自然に変わる理由などを不思議に思っていた。中学生のときに受けた授業で、大気が地球を覆っていて、その状態によって発生する気象現象が異なると知り、空という概念に具体的なイメージを持ち、腑に落ちた。これをきっかけに大気や気象を含む地球の活動に関心を持った。

高校 2 年では、SSH 課題研究の機会を得て、2022 年 6 月に関東を中心に発生した降雹の解析を行った。雹は気象庁で常時観測されていないので Twitter でデータを集めた結果、tweet の分布から降雹の有無を推定できた。約 9 万件の tweet を 1 件ずつ分析するのは非常に時間がかかったが、天気の変化や被害などの詳細な情報を得られた。また、夏休みには三宅島で風の現地観測も行った。これらの活動を通して気象を専門的に学びたいと思った。

大気科学に興味を持つうちに貴学の気象研究所を知り、学部生から専門機関の研究を意識して学べる点に惹かれた。また、地球学類は大気科学の他にも地球学の幅広い分野が網羅され、かつそれぞれが体系化されており、分野の枠を超えた研究ができる。大気や地形、地理など広い視点から地球環境とその進化過程や人間活動との関わりを探究し、特に大気科学を専門的に研究するために筑波大学生命環境学群地球学類を志望する。

生徒が実際に作成したこの文章は、いずれ広島大学高等教育研究開発センターの HP に『高等教育研究叢書』（PDF 版）の一部として掲載されることになる。数年後に大規模言語モデル（LLM）が更新されると、この文書をインプットとし、理論上は生成 AI がこの文章を書くことができるようになる。それでもなお普遍的な価値として残るのは、この文章がこれを作成した生徒本人の体験に基づいて書かれている、ということである。この文章が生成 AI ではなく本人により書かれたのかどうかを確認するならば、例えば「中学生の時の授業担当の先生の名前は」、「tweet の分析は 1 人で行ったのですか」など、当事者でなければわからないことを口頭試問により確認すれば良い。生成 AI の台頭によってこれまで以上に物事の価値を判断する眼を養う必要が増してきたのではないだろうか。



#### 4. まとめ

中川（2024）は、以下のような複数の指摘・提言を行なっている。生成 AI が誘発する現状固定化のリスクとして、生成 AI が作成した回答の正誤や善悪、効用などを評価する能力を磨く必要があるが、そのためには生成 AI 以外の場所から予備知識を入手する必要がある。そして、生成 AI にまだ短い期間しか接していないはずの 2023 年時点で成人である人は、基本的に生成 AI 以外の予備知識を大量に持っているはずなので、これが可能である。しかし今後は物心ついた時から AI 漬けで育つ人が増えるため、AI のみによる知識獲得に頼ることになり、視野を狭めることになるという懸念がある。したがって、初等中等教育の関係者は、今後はこのような若い世代への生成 AI の回答の評価方法の指導法を開発する必要がある。また、生成 AI はその仕組み上、過去のある時点の社会で正しいとされる概念に基づいて回答を作成しているため、生成 AI のみを使い続けることで、社会の現状固定化を促進する可能性が高い。

平（2023）は学校現場における生成 AI の対応策として、まずは教員として ChatGPT のような生成 AI の仕組み、およびその生成テキストの特徴を理解しておく必要があることを指摘している。そして、生成 AI の活用もしくは剽窃防止に取り組むなら、学生に対してあらかじめポリシーを明示しておく必要があることも提言している。筆者の勤務校でも生徒にレポートなどの提出を求めるときは、事前に「執筆に生成 AI を使用しないこと」という注意喚起を行っている。生徒と良好な関係を保つためにも必要な配慮だと思う。

終わりに、人類と生成 AI との関係において 2023 年はターニングポイントとなったはずである。これからの社会を担っていく中学生や高校生に、2024 年の最初の授業で次のような提言をした。すなわち「AI 時代」を生きる今の中学生や高校生は、これまで以上に手軽に知識を得られるようになった。しかし、生きていく上で自分に蓄えておかなければならない知識はほとんどなく、知識を持っていなくてもヒトとしての生存は可能である。この原点に立つと、知識は教わるものではなく自分で意識的に獲得するものである、ということがわかるのではないか。では学校の授業では何を学ぶのかというと「ものの見方、考え方、はたらき」である。授業では世界の視点を獲得してほしい。そして、無限である世界から新たな視点を切り出すことは AI には不可能で、人間にしかできない。そもそも AIこそ人類が発明した道具、人工知能（Artificial Intelligence）である。生徒には世界を観察し、その変化に気付く力を身につけてほしい。

授業や学校生活を通して生徒に何を伝えたいのかという教育の根源が、生成 AI の台頭によって改めて問われている。

#### 【参考文献】

新井紀子（2024）「生成 AI の教育に及ぼす影響について」（文部科学省初等中等教育段階における生成 AI の利活用に関する検討会議（第 4 回）配布資料）（最終閲覧日 2024 年

9 月 30 日)

[https://www.mext.go.jp/content/20240920-mxt\\_jogai01-000037908\\_001.pdf](https://www.mext.go.jp/content/20240920-mxt_jogai01-000037908_001.pdf)

太田邦史 (2023) 「生成系 AI (ChatGPT, BingAI, Bard, Midjourney, Stable, Diffusion 等) について」(東京大学情報システムの総合案内サイト utelcon) (最終閲覧日 2024 年 10 月 8 日)

<https://utelecon.adm.u-tokyo.ac.jp/docs/20230403-generative-ai>

G7 広島サミット (2023) 「G7 首脳声明」(最終閲覧日 2024 年 10 月 2 日)

[https://www.mofa.go.jp/mofaj/gaiko/summit/hiroshima23/documents/pdf/Leaders\\_Communique\\_01\\_jp.pdf?v20231006](https://www.mofa.go.jp/mofaj/gaiko/summit/hiroshima23/documents/pdf/Leaders_Communique_01_jp.pdf?v20231006)

平和博 (2023) 『チャット GPT vs. 人類』文藝春秋

中川裕志 (2024) 「生成 AI が誘発する現状固定化のリスク」『科学』岩波書店 94 (2), 118-123 頁

広島 AI プロセス (2023) 「高度な AI システムを開発する組織向けの広島プロセス国際行動規範」(最終閲覧日 2024 年 10 月 2 日)

<https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document05.pdf>

文部科学省 (2023) 「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」(最終閲覧日 2024 年 10 月 2 日) [https://www.mext.go.jp/content/20230710-mxt\\_shuukyo02-000030823\\_003.pdf](https://www.mext.go.jp/content/20230710-mxt_shuukyo02-000030823_003.pdf)

文部科学省 (2024) 「初等中等教育段階における生成 AI に関するこれまでの取組み」(最終閲覧日 2024 年 10 月 8 日)

[https://www.mext.go.jp/content/20240725-mxt\\_jogai01-000037149\\_21.pdf](https://www.mext.go.jp/content/20240725-mxt_jogai01-000037149_21.pdf)

## 第7章 人文社会科学研究における生成 AI

藤井 基貴  
(静岡大学)

### 1. はじめに

2024 年 10 月 8 日、スウェーデン王立科学アカデミーは、機械学習においてニューラルネットワークを利用可能にする基礎的な発見と発明を行った、ジョン・ホップフィールド（プリンストン大学名誉教授）とジェフリー・ヒントン（トロント大学名誉教授）らにノーベル物理学賞を授与すると発表した。さらに翌日、グーグル・ディープマインド社のデミス・ハサビス最高経営責任者（CEO）ら 3 名による、タンパク質の立体構造を予測する AI「アルファフォールド (AlphaFold)」の開発に対して、ノーベル化学賞の受賞が報じられた。

これにより 2024 年の「物理学賞」、「化学賞」のノーベル賞はいずれも AI と深い関係ある研究成果が選ばれることとなった。1901 年の創設以来、一世紀を超える長いノーベル賞の歴史のなかでも、AI に関する研究が選ばれたのは今回が初めてのことであり、さらに多様な自然科学の研究領域にあって同様の研究領域が二つの部門で同時受賞したことも異例であった。このことはまた、AI の技術が急速に進展し、私たちの実生活に着実に浸透しつつあること、さらに現実社会における AI の影響が日増しに増大していることを改めて印象づけるところともなった。

「物理学×AI」、「化学×AI」にとどまらず、AI は自然科学全般において新たな研究ツールとして活用され、学問の加速と拡張に寄与し続けている。この傾向は人文社会科学における主要な研究分野にも及んでいる。例えば、文献の読解・解釈やその考察を行う研究、人間および社会の動態を多角的に分析し、数学や統計学を用いて社会構造の変容を論じる研究、異なる地域・時代の文化や人間活動を比較し、関連する言語を習得・活用して進められる研究など、多様な人文科学的・社会科学的な研究においても、AI はその威力を発揮し続けている。AI の適用はこれらの分野においても、従来の研究方法を補完し、新たな知見の創出に貢献しているといえるだろう。

2021 年に発表された第 6 期科学技術・イノベーション基本計画では、自然科学と人文社会科学を統合する「総合知」が強調され、学際的なアプローチを通じて新たな知の創造が目指されている。この「総合知」は、単なる技術革新だけでなく、社会的・倫理的な側面も考慮した持続可能な社会の実現に向けた重要な指針となっており、これを受けて「人間の本質」や「幸せ」といった根源的な問いに対して、AI を活用して人文社会科学と脳神

経科学の研究者が協力し、これらの問いに取り組む学際研究も進行している。さらに、こうした AI 研究を支えるデータサイエンスの重要性も高まっており、さまざまな大学でデータサイエンス学部も新設されてきた。こうした動向からも、社会全体で AI 技術の社会実装を支援し、AI を活用した新しい研究分野の開拓とその活況を見て取ることができるだろう。

また、1990 年代の国際的な研究プロジェクトである「ヒトゲノム計画 (Human Genome Project)」において注目された「倫理的・法的・社会的課題 (ELSI)」に関する研究や、「責任ある研究イノベーション (RRI)」といった概念への再評価・再検討の動きも注目されなければならない。現代社会における科学技術の適切な実装には、単に研究開発を進めるだけでは不十分である。研究開発と共に、研究倫理や研究公正の制度を確立することが、科学技術の健全な発展のための二本柱であることは明白といえよう。

加えて、AI の活用は大学教育を根本から見直すことにもつながっている。本章の構想を練っている段階で、アメリカの大学で宗教学を教えている友人から次の個人的なメッセージを受け取った。

I'm working with a student who is using AI in a very positive way on a great project related to Biblical texts. Unfortunately, many of my students are using it to write papers and I really don't want to spend my time grading a computer!

(訳) 聖書のテキストに関連した素晴らしいプロジェクトで、AI を非常に有益な方法に活用している学生と一緒に仕事をしています。ただし、残念ながら、学生の多くはレポートを書くために AI を使っていて、私はコンピュータを採点することに時間を費やしたくはないのです！

AI の研究開発は倫理や安全性とどのように両立しうるのか、また人文社会科学において研究や教育活動における「有益な方法 (positive way)」をどのように生み出し、共有しうるのか。さらに AI をめぐる新たな課題に対して人文社会科学の知はどのような貢献を果たしうるのか。こうした AI をめぐる新たな人文社会科学的研究の「問い」は尽きないものの、以下では生成 AI の活用とその倫理的課題に焦点をしばって議論を進めてみたい。なお、本章内で掲載している幾つかの図表は 2024 年 2 月 16 日に開催された「2023 年度全国公正研究推進会議」での筆者の報告資料からの抜粋となる。

## 2. 生成 AI と学問領域

### 2. 1 AI の歴史と現在

AI に関する研究は 1946 年に世界初のコンピュータ ENIAC（エニアック）の開発に端を発する。1956 年、アメリカ・ダートマス大学で開催された会議（通称「ダートマス会議」）において「AI : Artificial Intelligence」という言葉が初めて用いられ、それから 1970 年代の初めまでが第一次 AI ブームとされる。AI にはもともと 2 種類の方式が構想されており、一つは「記号処理型の AI (Symbolic AI)」, もう一つは「AI の父」と呼ばれるヒントンらが注目した「ニューラルネット」であった。AI 研究の創世記にあつては、前者の記号処理型が主流となったものの、コンピュータの性能が十分ではなく、人間が記号処理のルールをすべてコンピュータに移植しなければならないため、時間と労力が見合わった成果があらわれず、次第にブームは陰りを見せることとなる。

その後、1980 年代に「エキスパートシステム」が登場し、第二次 AI ブームが始まる。「エキスパートシステム」とは分野を絞ることで、効率的に専門知識のコンピュータへの移植を目指すものであり、法学や医学への応用が期待されていた。日本でも 1982 年に「第五世代コンピュータプロジェクト」が発足し、570 億円もの予算が通産省より投入されている。しかしながら、専門性を絞ったとしても知識やルールは膨大にあり、第一次ブームと同様に、それらを入力する作業の繁雑さに直面して、ブームは下火となっていった。

これに対して、ヒントン氏らはコンピュータや AI 自体が、現実世界のデータから学び、賢くなっていくという方式の研究開発に傾注していた。この発想は人間や動物の神経細胞（ニューロン）の接続機能に着目して生み出されたもので、神経細胞の接続機能やネットワークをコンピュータ上で再現しようとするところから「ニューラルネット」と呼ばれるようになる。その構築にあたっては、コンピュータ (AI) に膨大な情報（ビッグデータ）を読み込ませ、そのパターンを解析して、学習させていかなければならない（機械学習）。ただし、この技術が確立するにいたるにはコンピュータ・プロセッサの処理能力の高速化及び高度化を待たなければならなかった。

2009 年、トロント大学ヒントン研究室の大学院生たちは、アメリカの半導体メーカー「Nvidia (エヌビディア)」が家庭用ゲーム機に提供していた「GPU (Graphics Processing Unit)」と呼ばれるグラフィックス・カードに注目する。ウクライナ出身のアレックス・クリジェフスキーとロシア出身のイリア・サツキバーの二人の大学院生は GPU を用いることでニューラルネットを多層化させ、処理能力を格段に向上させることに成功し、その成果をまとめた論文を発表する (Alex Krizhevsky et. al, 2012)。同論文はこれまでに 10 万回以上引用されており、AI 研究における最重要論文の一つとなっている。その後、GPU を活用した多層ニューラルネットによる機械学習は「ディープラーニング (深層学習)」と呼ばれるようになり、これにより第 3 期 AI ブームが到来した。

2011年にはGoogle社内に「グーグル・ブレイン (Google Brain)」と呼ばれる研究部門が設置され、ヒント教授も臨時研究員として招聘される。2013年、Googleはヒント氏を含むトロント大学の研究チームを引き抜き、2014年にはイギリスに拠点を置く新興ベンチャー「ディープマインド」を買収して、アルファフォールドの前進となる「AlphaGo」の開発を加速させて、新時代のAI研究開発に先鞭をつけた。その成果は冒頭のノーベル賞受賞へと結実している。ただし、Googleの一人勝ちと思われたAIの研究開発については、その後、生成AI「ChatGPT」を開発したOPEN AIを始め、新たな新興企業が続々と誕生しており、群雄割拠の時代に入りつつある。

ここで本章が中心的に取り上げる生成AIについて整理しておく、生成AIとは端的には「学習された大量の情報・データをもとにして 再構成された情報・データを生み出せるAI」ということになるだろう。その対象は文字情報だけにとどまらず、画像・音声・音楽・動画などにも及ぶ。現在の生成AIの研究開発はGoogleを傘下とするAlphabet, OPEN AI, Amazon, Metaなどの巨大ITグローバル企業に主導されるものの、用途に応じたツール開発を加速させるスタートアップベンチャー等によって、多角化・細分化・個別化の様相を呈している。

また、ビッグデータからの統計的推論を進めてきた「深層学習」についても今後は注意が必要となると言われている。というのも、これまでのAI研究開発のように人間が生み出したデータを読み込ませるのではなく、生成AIの活用などを通してAIがつくったデータをAI自身が読み込んで学習してしまうという時代がすでに始まっているからである。この問題は「garbage in, garbage out. (ゴミを入れたら、ゴミが出てくる)」とも形容されており、AIのデータの質そのものと直結する問題となるため、良質なデータをいかに維持できるかがAI開発の新たな課題となっている。また、研究活動においても生成AIから生み出されたデータの真偽や妥当性をどのように評価、採用できるかが一層問われることになるだろう (図1, 図2)。

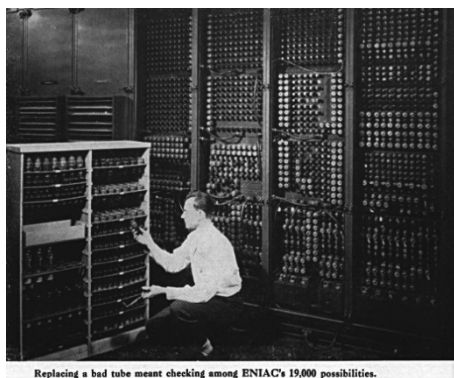


図1) 世界最初のコンピュータの画像



図2) 生成AIに「世界最初のコンピュータ」と入力して生成された画像



## 2. 2 生成 AI の活用に関するアンケート

大学における生成 AI の活用について、2024 年 5 月に教育系および福祉系の学部を擁する A 大学の教員に簡易なアンケート調査を実施させていただいた。アンケートに協力くださった教職員は 72 名であり、回答数は 68 である。なお、倫理的配慮として、対象となる大学及び回答者の了解をえた上で、個人情報保護の観点から匿名で行い、回答の強制はないことを口頭で説明するとともに、趣旨に賛同いただいた回答のみを紹介する。

問 1：研究の中で生成 AI（ChatGPT 等）を活用したことはありますか？

	項目	回答数	割合
1	はい	3	5%
2	いいえ	56	82%
3	無回答	9	13%

問 2：生成 AI の使用について学生に活用方法の説明，注意，指導をしたことはありますか？

	項目	回答数	割合
1	はい	7	10%
2	いいえ	51	75%
3	無回答	10	15%

問 3：上記で「はい」の方におたずねします。上記についてどのような内容であったか可能な範囲でお聞かせください。

- ・生成 AI の説明，倫理の説明
- ・盗用や剽窃，コピーにならないよう指導しました。
- ・「いいえ」と回答しましたが，授業時のレポートを含めて説明をする機会を設けたいと感じました。
- ・レポート作成の際には，教員側は使ったことがすぐに分かるので，使用しないように話した。また，使っても学生さんの論理的思考や論述の力がつかないことをあわせて説明した。
- ・具体的な活用法と使用されている事例を紹介し，さまざまな場面において私たちの手助けとなることを説明すると同時に，内容の信ぴょう性や著作権の問題，セキュリティ

ィ上の問題等を説明し、他大学でどのような対応をとっているのかなどを説明した。

- ・ AI についての基本的な知識，生成 AI の使用と著作権などへの配慮
- ・ まずは，使ってみること。今後注意すべきことについて簡単に話したことはあります。
- ・ レポート作成の際に，コピーしてはいけない。

同大学では，教員のほとんどが人文社会科学領域の教育・研究活動に携わっている。2024 年 5 月の時点では研究における生成 AI の利用率は 5%，学生に生成 AI の活用方法について直接的な指導を行っている教員は 10%という結果であった。また，指導内容に関する自由記述からは生成 AI の効果的活用についてよりも，剽窃，注意を要する事案，使用リスクに関するものが多かったことが読み取れる。国内大学の対応状況については、九州大学 IR 室で学術推進専門委員をつとめる森本氏がまとめており，2024 年 9 月の時点で日本国内の大学において ChatGPT/生成 AI への対応を表明した大学は 400 校と報告されている（森本，2024）。

一方で，全国大学生協が 2023 年秋に全国の国公立 30 大学の学部生約 1 万人に実施した第 59 回学生生活実態調査によれば，ChatGPT などの生成 AI を継続的に利用している学生はおよそ全体の 3 割にのぼっているという。また，利用経験がある学生についても 46.7%という結果が示されており，利用目的については「論文，レポート作成の参考」，「翻訳，外国語作文」，「相談，雑談相手」が上位となっている（全国大学生生活協同組合連合会，2024）。

### 3. 人文・社会科学研究における生成 AI 活用の可能性と倫理的課題

#### 3. 1 生成 AI 活用の可能性

生成 AI を活用した新たな成果は人文社会科学領域の研究においても続々と発表されている。以下では主たる活用方法について整理してみたい。

第一はデータ分析の効率化である。生成 AI を活用することで，大量のテキストデータを迅速に処理し，テーマやトピックの抽出，内容の要約，さらには分析を瞬時に行うことが可能となる。たとえば，社会科学における大規模なアンケートデータや歴史的文書の分析において，生成 AI は短時間でデータの傾向や主要テーマを把握するために有用であり，研究者が迅速に仮説を立て，今後の研究計画を明確化する手助けとなっている。

第二に，生成 AI は単にデータの分析にとどまらず，テキストそのものの生成も可能である。この機能により，研究者は論文の下書き，仮説の提示，分析結果のまとめ，さらには文章の推敲といった支援を受けることができる。自然科学の分野で実験結果の報告書の作成等で活用されているように，人文社会科学の分野でも，歴史学や文学研究において文献レビューの草稿作成や分析のまとめに生成 AI を用いることが増えてくるだろう。

第三には、翻訳および要約機能の活用を挙げることができる。特に、外国語を使用した研究において、生成 AI は海外の資料や文献の読解および翻訳において不可欠なツールとなっている。また、この技術の活用により、国際共同研究も促進される。実際に、筆者の周囲でも、海外の学生や研究者が生成 AI を用いて日本語で書かれた論文を理解し、その結果として留学の受け入れや共同研究の申し込みに至った事例が複数存在している。

第四に、教育業務において、特に卒業論文等の論文指導の場面で生成 AI を有効に活用することが可能である。もちろん、生成 AI の利用に関する基本的な注意事項や従来の研究作法の根本について学生に指導した上で、論文の構成案やアイデアの検討段階で生成 AI を活用することで、効率的な指導が期待できる。また、誤字脱字の修正や文章表現の改善を生成 AI に事前に任せることで、指導者は内容的な指導に注力でき、教育業務の効率化が図られる。しかしながら、生成 AI の活用には研究倫理上の重要な課題が存在することも忘れてはならない。教育活動においても生成 AI を効果的に用いることが期待される一方で、以下に述べるように、その限界や倫理的な課題についても理解し、適切に対応することが求められる。

### 3. 2 生成 AI 活用の倫理的課題

根本的な課題として、生成 AI を活用する際には、民間および軍事の両分野で利用可能な技術開発を伴うデュアルユースの問題について十分に考慮されなければならない。以下ではこの大きな問題を踏まえつつも、現在の AI 開発における本質的な課題に焦点を当てて、人文社会科学における生成 AI の活用に関する倫理的課題を整理・検討しておく（図 3）。

第一に、生成 AI が学習したデータのソースは必ずしもすべて公開されておらず、それらのデータから再構成された情報やデータがどのように生成されているのかについては、依然としてブラックボックスとなっていることがある。このため、生成 AI が学習したデータに偏りが存在する場合、その生成された情報やデータにも偏見や誤りが含まれる可能性が高く、結果として情報やデータに一定のバイアスが生じる恐れがある。この問題に関連して、生成 AI が学習に使用する多くの情報・データは、主に先進諸国で生成されたものであると推定されており、いわゆるグローバル・ノース由来の情報源が主流となっていることも懸念される。これにより、新興国や途上国の情報やデータが不足し、先進諸国における議論を前提とした結論が導き出される傾向があるとも指摘されている（チリツィ・マルワラ他、2024）。

第二に、著作権および知的財産権に関する問題が挙げられる。生成 AI が学習に使用するデータが著作権で保護されたものである場合、研究において生成 AI が生成したデータや情報を利用することは、著作権侵害のリスクを伴う可能性がある。また、未発表の研究データやアイデアを入力してしまうと、それらのデータやアイデアが他者によって無断で

活用されるリスクも生じる。

第三に、プライバシーおよび個人データの取り扱いに関する問題がある。例えば、生成 AI のユーザーが入力した個人情報が、本人の同意なしに収集され、それがマーケティングや広告目的で利用されることが懸念される。さらに、一度入力された情報については、削除や再利用を求めても、生成 AI がその要求に応じない設計となっている場合があり、その結果、個人データが不正確な形で再構成され、公開される可能性がある。

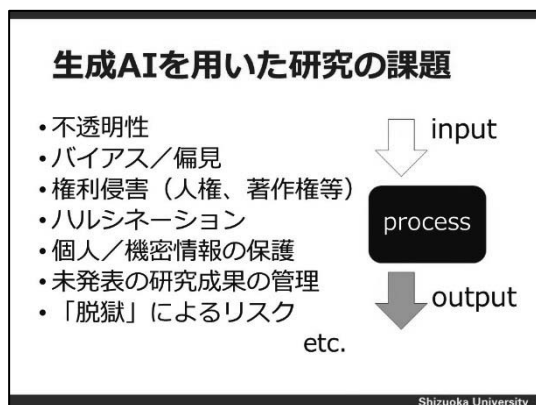


図 3) 生成 AI を用いた研究の課題

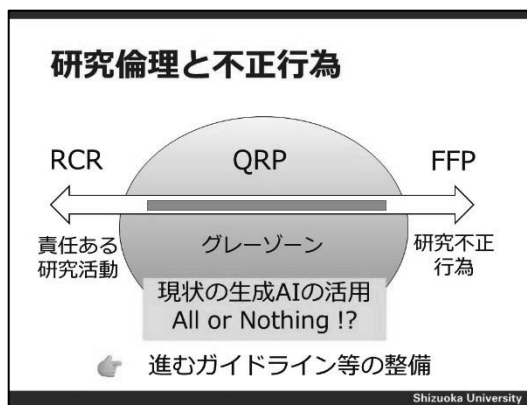


図 4) 研究倫理と不正行為

さらに、技術の進展により発生頻度は減少しているものの、生成 AI が誤った情報を生成する「ハルシネーション」現象にも注意を払う必要がある。加えて、生成 AI は外部からのハッキング行為や、AI システムの倫理的制限を回避しようとする「脱獄」といったリスクにも常にさらされている。このような状況下で、安全性を確保するためには、技術革新を継続的に進めるとともに、データの取り扱いに関する明確な倫理的基準を設定し、AI 開発側も透明性を担保する努力が必要不可欠であろう。

あわせて、これらの倫理的リスクの多くは、従来の研究倫理の枠組みにおいて、「責任ある研究活動 (RCR)」と「研究不正行為 (FFP)」の間に置かれる「好ましくない研究行為 (QRP)」と重なる部分が多い (図 4 参照)。我が国における研究倫理および研究公正に関するこれまでの指針や運用は、倫理的に優れた「よい実践 (Good Practice)」を推進する志向倫理的な方向性よりも、生成 AI の利用を制限し、規制を強化する予防倫理的な傾向に陥ることが少なくなかった。したがって、これからの生成 AI 活用に関するガイドラインの策定および改訂作業においては、「べからず集」に偏ることなく、研究者がいかにか「よい実践」を実現できるかに焦点を当てることも重要である。すなわち、単にリスクを回避するための規制ではなく、生成 AI を適切かつ責任を持って活用し、研究の質や倫理を高めるための積極的なアプローチが創発されるよう配慮すべきであろう。

#### 4. 生成 AI の活用に向けた取組

本章の執筆に際し、2023 年秋から約半年間にわたり、勤務校を中心として同僚や研究仲間に対するインタビュー調査を行う機会をえた。その過程で一つの気付きが得られた。それは、自然科学領域の研究者の多くが生成 AI を単なる便利なツールとして認識しており、その利用に際しては「取扱説明書」があれば十分であるという、ある種の楽観的な姿勢が広く共有されている点である。一方で、人文社会科学領域の研究者の間では、「AI は全能であり得るのか」や「AI に人格を付与することが可能なのか」といった根源的な問いについて語られることがあり、生成 AI については単なる外部の道具というよりも、自律性を備える可能性や、人間を凌駕する存在となるリスクを伴うものとして、倫理的・道徳的な観点から総合的な検討が必要であるという警戒感が示されることが少なくなかった。

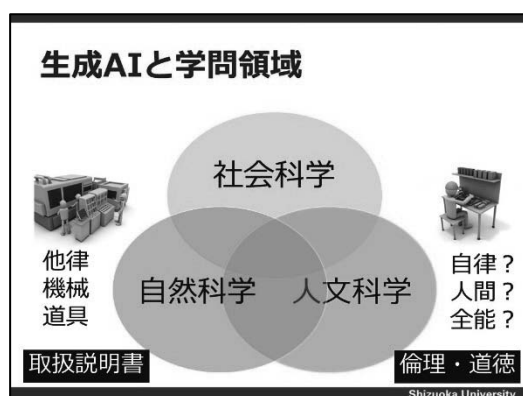


図 5) 生成 AI と学問領域

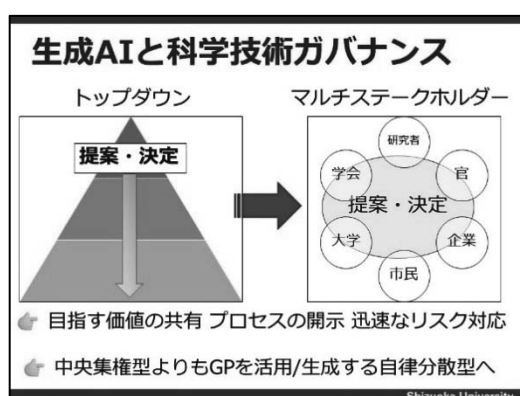


図 6) 生成 AI と科学技術ガバナンス

この違いは、生成 AI を巡る自然科学と人文社会科学の単純な「分断」を示すものではなく、むしろ、両者が学術研究を支える上で互いに補完的な役割を果たしていることを示唆していると受けとめたい。自然科学の実務的な視点と、人文社会科学における倫理的・哲学的な考察の相互補完は、生成 AI の利用に関する多角的な理解を深める上で、重要な循環構造をなしていると考えている（図 5）

また、加速度を増して進展する AI 研究においては、迅速かつ適切なリスク対応がますます重要な課題となる。従来、大学や研究組織においては中央集権型やトップダウン型のガバナンスが主流であったが、このモデルでは動きの速い AI 技術の進展やそれに伴うリスクに対して迅速に対応することが困難になることが懸念される。特に、複雑化するリスク要因に対しては、従来の一元的な意思決定プロセスでは対応が後手に回る可能性が高い。

これに対し、近年の企業経営において広く採用されているマルチステークホルダー型ガバナンスが、AI 研究や技術開発におけるリスク管理のための有効な手法として注目されている（図 6）。このガバナンスモデルでは、問題が発生した際に複数の関係者が一堂に会し、



協働して即座に問題解決にあたる体制を構築することが可能である（経済産業省，2024）。また，この体制は各部門や個別の研究組織における自律的な運営を支援し，中央集権型モデルに比べて柔軟かつ迅速な対応を促進することにもつながることが見込まれる。加えて，このマルチステークホルダー型ガバナンスの導入は，大学の組織運営に対しても大きな変革をもたらすと考えられる。大学が AI 技術の進展に対応するためには，従来の縦割りのなガバナンスから脱却し，研究者，行政，産業界，市民社会といった多様なステークホルダーが協働する体制への転換が求められる。これにより，リスク対応をより迅速かつ効果的に行うとともに，大学の運営における柔軟性と自律性を高めることも期待できるだろう。

## 5. おわりに

2024 年 10 月 8 日，ノーベル物理学賞を受賞したヒントン氏は，オンラインによる記者会見を開き，その中で AI 技術に内在するリスクについて強調した（日本経済新聞，2024）。

ヒントン氏によれば，現在 AI 開発に携わるほぼすべての企業が「より優れたモデルの開発」に全力を注いでおり，安全性に関する取り組みが十分に考慮されていないという。会見では特に「安全性に対する取り組みはモデル開発と同様に重要である」との強い懸念が表明された。さらに，教え子であるサツキバー氏の動向に関する言及もあった。サツキバー氏は 2024 年 5 月に OPEN AI を退社し，安全性を最優先とする AI 企業を新たに設立している。その背景には，OPEN AI 創業者であるアルトマン氏が，技術の急速な発展を重視する「効果的加速主義」(e/acc) を支持し，安全性よりも利益を優先しているとして社内での大きな対立を招いたことがある。ヒントン氏はこのことについて，アルトマン氏の立場に明確な批判を展開している。

アメリカにおける AI 開発の安全性をめぐる議論は，長期的な視野に立ち，技術のリスクを慎重に検証しながら進めるべきだとする「効果的利他主義」(EA) と，技術革新の加速を促進する「効果的加速主義」との対立として顕在化し，ときに先鋭化している。この対立は，AI 技術の急速な進展に伴い，企業内外における重要な課題として浮上しており，我が国にあっては今後の AI 研究と開発におけるガバナンスや倫理的な方向性に大きな影響を及ぼすと考えられる。

このことに関連して，ノーベル物理学賞を受賞したヒントン氏と化学賞を受賞したハサビス氏の両名が，いずれも Google の関係者である点についても付言しておきたい。高額な報酬によって優秀な研究者を集め，技術開発を加速させる企業文化は，技術革新を促進する一方で，科学技術の寡占化や学術文化との乖離を招きやすい。前者の寡占化については，米 Google 社はかねてよりデジタル広告をめぐる反トラスト法（独占禁止法）違反の可能性が指摘されており，2024 年 8 月 5 日に米連邦地裁は米 Google 社に対して反トラスト法（独占禁止法）に違反するという判断を下している。その一方で，国内市場における寡占の解消は，同時に国際競争力の低下を引き起こすリスクも伴うため，経済界からは慎



重な対応が求められていることも事実である。このように国内における競争環境を維持しつつ、適切な規制を施すことは、今後の科学技術をめぐる企業ガバナンスの重要な課題となる。

日本においても、企業と大学との関係において、特に AI などの先端技術を扱う分野では「超学際的研究 (transdisciplinary research)」という形態での共同研究が活発に進展している。たとえば、ある企業と大学が共同で AI アルゴリズムの開発を行い、これを自動運転技術や医療診断に応用する研究プロジェクトがあるとする。このような共同研究においては、異なる分野の知識を統合し、産業界と学術界の知的資源を結集することで、新たな技術革新が期待されるだろう。しかしながら、こうした共同研究の推進に伴い、企業文化と学術文化の相違に起因する新たな倫理的課題が生じる可能性がある。たとえば、企業は商業的利益の最大化を優先する一方、大学は学術的自由や研究の透明性を重視する傾向が強く、この価値観の違いが研究の方向性やデータの公開方針に影響を及ぼすことがある。そこではまさにアメリカにおける「効果的加速主義」と「効果的利他主義」との対立の相似形が現れることになるかもしれない。さらに、共同研究では、知的財産権や研究成果の帰属に関する問題が発生しやすく、これが研究の公正性や倫理性に関する懸念を引き起こす要因となりうる。こうした倫理的課題に対応するためには、企業と大学の間での明確な合意形成が不可欠であり、さらに研究倫理に関する具体的な指針の整備が求められる。「超学際的研究」においては、このようなリスクの適切な判断と対応が極めて重要であり、それは社会全体にとっても喫緊の課題となる。加えて、これらの新たな課題に対し、人文社会科学がどのような学術的貢献を果たすかも、重要な問いとして浮上するだろう。

## 【参考文献】

喜連川優 (2024)『生成 AI の論点 - 学問・ビジネスからカルチャーまで』青弓社

経済産業省 (2024)「AI 事業者ガイドライン (第 1.0 版)」(最終閲覧日 2024 年 10 月 30 日)

<https://www.meti.go.jp/press/2024/04/20240419004/20240419004-1.pdf>

鈴木貴之編 (2023)『人工知能とどうつきあうか』勁草書房

全国大学生生活協同組合連合会 (2024)『第 59 回学生生活実態調査 概要報告』(最終閲覧日 2024 年 10 月 30 日)

[https://www.univcoop.or.jp/press/life/pdf/pdf\\_report59.pdf](https://www.univcoop.or.jp/press/life/pdf/pdf_report59.pdf)

チリツィ・マルワラ・中村征樹 (2024)「国際連合大学学長 チリツィ・マルワラ氏 インタビュー」『学術の動向』328 号, 1~20 頁

西垣通・河島茂生 (2019)『AI 倫理 - 人工知能は「責任」をとれるのか -』中公新書ラクレ

日本経済新聞 (2024)「ノーベル賞ヒントン氏『AI が賢くなると支配権握る』」最終閲覧日 2024 年 10 月 30 日)

<https://www.nikkei.com/article/DGXZQOGN090500Z01C24A0000000/>

古川直裕・吉永恭子（2024）『責任ある AI とルール』一般社団法人金融財政事情研究会  
森本銀河（2024）「ChatGPT/生成 AI への対応を表明した国内の大学一覧」（最終閲覧日  
2024 年 10 月 30 日）

<https://docs.google.com/spreadsheets/d/1cDOqaIdu9JKOYuF0ThG33oixE09m210z/edit?gid=1530250804#gid=1530250804>

Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton. (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, Vol.25, 1097-1105.

#### 【謝辞】

本章は JSPS 科研費 19K02500 及び 23H00979 の助成を受けたものです。

## 第8章 生成 AI で研究不正は変わるのか？

井野瀬 久美恵  
(人間文化研究機構)

### 1. はじめに

研究不正をメディアが大々的に取り上げるのは理工系や医学系の分野かもしれないが、不正行為自体は人文学・社会科学分野も無関係ではない。それどころか、2015 年度以降、毎年公表される文部科学省の不正研究事案（正式名称「文部科学省の予算の配分又は措置により行われる研究活動において不正行為が認定された事案」）では、人文学・社会科学分野は常に約半数、時に過半数を占めている（文部科学省、2024）。

特定不正行為（捏造 Fabrication, 改ざん Falsification, 盗用 Plagiarism のことで、これらの英単語の頭文字から FFP と称される）のなかでは「盗用」が圧倒的に多いが、その中身は、単純な引用方法のミスや典拠不記載、データの無断借用に留まらない。「盗用」には研究の発想、アイデアがどうやって得られたかも問われ、学会報告だけでなく、その後の懇親会、あるいは内輪の研究会なども含まれる。アイデアの盗用疑惑が生じないように、論文にする場合には注や謝辞を有効に活用することも、2015 年度以降、大学や研究機関に義務づけられた研究倫理教育や研究の公正性向上のための研修（e-learning を含む）を通して、人文学・社会科学分野の研究者に広く知られるようになってきた。文科省研究公正推進室（2015 年 4 月設置）によるチェックもあって、2015 年からの 10 年間のうちに、各機関では、規程改定や告発窓口の設置といった研究公正推進の体制づくりはほぼ完了したとっていいだろう。今求められているのは、この 10 年間の検証であり、次のステージの模索である。

「次のステージ」を考える際に、生成 AI の問題は不可避、不可欠である。奇しくも 2023 年 5 月の G7 広島サミットを受けて、生成 AI をめぐる国際ルールを検討する「広島 AI プロセス」が立ち上げられた。そこで策定された「包括的政策枠組み」と「作業計画」に基づいて、現在、さらなるプロセスが進行中である。広島大学高等教育研究開発センターの『高等教育研究叢書』の一冊として、『研究・教育における生成 AI の利活用の方法とその考え方』を上梓する意義はここにもある。

上記の問題意識に基づき、終章となる本章では、まずは人文学・社会科学分野における研究倫理、研究公正をめぐるここ 10 年間の動きを簡潔にふり返り、そのうえで、本書の議論を多少なりとも整理しながら、いささか身にあまる「生成 AI で研究公正は変わるのか」という問いへのヒントを得たいと思う。

## 2. 人文学・社会科学分野の現状とアップデートの必要性

「研究活動における不正行為への対応等に関するガイドライン」(2014年8月26日文部科学大臣決定)が本格的に適用された2015年4月以降、大学等研究機関では、研究不正防止、研究者の研究倫理向上のための体制整備が進められてきた。常勤・非常勤を問わず所属する教員や研究者、および職員や学生(院生)への多様な研修を通じて、この10年間のうちに、「FFPとは何か」をはじめ、FFP以外に研究不正を疑われるグレイゾーン、すなわち「問題ある研究行動(Questionable Research Practices, QRPs)」に関する知識や情報、共同研究における利益相反の考え方、ピアレビューのあり方などの認知も、(程度の差はあるにせよ)前進したと思われる。

その一方で、人文学・社会科学分野の研究者には、研究不正(の可能性)が感知しづらい内容もある。以前に発表した自分の論文や著作の一部を適切な典拠を示さずに使う「自己盗用」はその好例であろう。同じ文章(あるいはデータ)が複数の論文で再利用されることで、研究成果が量的に水増しされ、質的に希薄化することが危惧され、論文で重視されるオリジナリティへの疑義を生むからなのだが、過去に自分が発表した研究の一部、自分が書いた文章なのになぜ「盗用」が問われるのか、理解に苦しむシニア研究者は少なくない。さらには、授業負担とともに(時にはそれ以上に)「役職負担」によって研究時間が奪われることで、自己盗用に対する思慮や慎重さを失うシニア研究者は実際に存在する<sup>1)</sup>。

シニア研究者が若手育成を担当している現状を勘案すると、より憂慮されるのは、研究不正に対する認識の甘さや誤認識が再生産される可能性である。文部科学省の委託事業調査報告書(未来工学研究所, 2021)によると、大学院博士後期課程の学生のうち、研究不正関連の相談をしたことのある学生は57%、相談相手は指導教員が72%で圧倒的に多く、大学等研究機関の窓口はわずか9%であった。「研究不正の再生産」という観点からすると、研究歴の長い研究者こそ研究不正に敏感であらねばならない。

にもかかわらず、指導する学生の論文盗用で訴えられる教員も少なくなく、パワハラか研究不正かの境界は大学関係者を悩ませるグレイゾーン(QRPs)ともなっている。研究不正に関わる認識は常にアップデートする必要がある。

もうひとつ、2019年に東洋英和女学院大学の臨時理事会が懲戒解雇を決議した研究不正事例に触れておきたい(東洋英和女学院大学, 2019)。学院トップで学会の重鎮、すでにその分野で名を成したベテラン研究者が、なぜ歴史上実在しない「神学者カール・レーフラー」という人物の、ありもしない「今日の神学にとってのニーチェ」なる論文を「捏造」せねばならなかったのか、「捏造」できてしまったのはなぜか。人文学関係者に与えた衝撃は大きく、今なお理解に苦しむ。その立証に多くの時間を割かねばならなかった調査委員会メンバーの物理的・精神的負担は想像するに余りある。調査対象人物には「立証妨害」も認定されたというから言葉もない。通常の犯罪とは異なり、研究不正については、大学等研究機関を通じた学術コミュニティの自浄努力に任されている。研究者が研究公正と研

究倫理を問われるのは、まさにこの原点ゆえである。

この事案が、人文学・社会科学分野の「評価」と関わる問題点を鋭く突いていたことにも留意しておきたい。人文学・社会科学分野においては、個人で研究、データや統計の処理、執筆を行うことが多く、過去の単著論文を集めて一冊の単著として刊行することが慣例となっている。分野横断の共同研究が進められている現在なお、この基本はさほど変わってはいないように思われる。ゆえに、論文段階で研究不正に気がつかなければ、単著公刊の後で、告発を受けて初めて不正に気づくことになる。著作の回収、絶版の公表など、出版社の負担も小さくない。大学等研究機関にポストを得るために単著刊行をめざしてきた若手研究者の場合、その打撃はいかばかりだろうか。

上記は人文学・社会科学分野のごく一部で起こっている事象ではあるのだが、研究公正と研究倫理に関わる知識や情報、わけてもそのアップデートが、研究の質向上のみならず、実は研究者自身の未来を守るものであることは明らかだろう。研究公正の重要性はまさにここにある。では LLM（大規模言語モデル）や生成 AI は、上記の基本認識とどう関わっているのだろうか。生成 AI で研究公正の何が変わり、何は変わらないのだろうか。

### 3. AI による自動翻訳とスキル売買——アウトソーシングされる論文

ChatGPT に代表される生成 AI は、突然未知の世界からやってきたエイリアンではない。20 世紀末以降のデジタル化の進展のなかで、人文学・社会科学分野の研究公正に関しても何が起こるかの予兆は感じることはできた。ここでは翻訳ソフトの使用、並びに研究スキル売買をめぐる問題に触れておきたい。

デジタル化以前の 1980 年代までは、きわめて狭い領域の専門家以外、海外ではその領域においてどのような研究がなされているのかをつかむことは難しかった。それに乗じて、海外の研究者の業績、著作や論文（の一部）を日本語に翻訳し、出典を明記せず、自分の著作、論文として紀要等の雑誌論文（ひいてはそれを集めた論文集）に発表することが行われていた。明らかに「盗用」だが、上記の事情もあって、なかなか発覚しないこともあった。この種の研究不正を陳腐化させたのが、20 世紀末以降のデジタル化の進展であり、とりわけ 2010 年代以降の手軽な翻訳ソフトの開発であった。

デジタル化が進む 2014 年に設立された日本翻訳大賞を引き合いに出すまでもなく、外国語文献の翻訳は、それ自体が人文学・社会科学分野の重要な研究対象、業績の形であり続けている。その一方で、AI による自動翻訳ソフトの使用が、翻訳という作業の意味を変えつつあることもまた事実である。研究者は翻訳ソフトの使用を明記する必要があるのか。どこまでならばソフトの使用は許されるのか。スペルミスや文法といったテクニカルな部分であれば問題ないのだろうか。

ソフト開発は翻訳に留まらない。先行する文献の調査、統計処理やデータ分析、グラフやイラストの作成、ひいては論文の構成や文章の校正、投稿支援に至るまで、ありとあら

ゆるソフトが開発されている。誤解を恐れずに言えば、すべて外注、アウトソーシングで論文が書けてしまう時代に突入している。それを、「どこまでが自分のオリジナルなのか、自分でもわからない時代」と言い換えてしまうと、言い過ぎだろうか。

言い過ぎではないことを強く印象づけたのが、「研究スキル売買」を報じた『毎日新聞』の記事であった（2021年9月13日）。「研究スキル売買 8万円で購入した私立大教員の不安と背景」<sup>2</sup>というショッキングな見出しでこれを報じた鳥井真平記者には、2021年度のAPRIN全国公正研究推進会議の人文・社会科学分科会において、「研究スキル売買の問題点」という題目で報告いただいた<sup>3</sup>。分科会の司会進行を務めた私には、少なくとも以下の4つの問題点が明らかになったと考えている。

- 1) 匿名同士の売買は、いわゆるゴーストオーサーシップに相当し、のちの検証を不可能にする。
- 2) スキル購入の契機として、指導（ないしは研究室の上司）にあたる教員の指導不足がある。「育てる気がない」ととられかねない状況はハラスメントも疑われる。
- 3) 研究者ポストに就けず、経済的な問題を含めて、スキル提供者側には「自分のスキルを活かして何が悪い？」という意識がある。
- 4) 上記2) 3) にもかかわらず、売買する双方の研究者に「うしろめたさ」がつきまとう。

全体として、研究スキル売買は、どこまでが研究不正となるのかのグレイゾーンが増えたことを意味する。

研究ではアイデア、発想こそが中核であり、著作のオリジナリティは発想にこそ問われることは広く認識されている。だからといって、「中核」以外ならば、資料やデータの提供、分析やその図像化、論文の構成までも外注していいものだろうか。「標準的な実験手順の記述にオリジナリティは必要ない」という理工系とは異なり、言葉の選び方や表現、いや「書くこと」自体にこだわりのある人文・社会科学系の研究者は、この問いにどう答えるだろうか。生成AIが日常化しつつある今は、スキル売買問題につきまとった「うしろめたさ」を感じなくてもよくなった——これが生成AIで変わった認識だとすれば、あまりに寂しい。

2023年度APRINの人文・社会科学分科会では、「その先」を展望するヒントを得たいと考えた。本書はその成果でもある。最後に本書全体を見直して、結びとしたい。

#### 4. 本書全体をふり返る

第1章の野内論文で目をひくのは、AIの道具性、研究の責任性、オーサーシップをめぐる問題を総合的に考えながら、論文作成過程で当然視されてきた「隠れた支援者」という慣習をあぶり出したことにある。ひとつの論文が完成するまでに、研究者は構想、資料分



析、執筆、それらの修正や微調整などで、ひとり思考をめぐらせるのみならず、仲間たちと議論し、助言を受けて、試行錯誤を繰り返す。その結果である論文は、まさしく複数の協力者、支援者の産物でもある。そこに生成 AI が加わったとて、論文作成過程そのものが変化するわけではないだろう。

生成 AI が問題ではないとすれば、何が問題なのだろうか。野内論文が示唆するのは、「生成 AI＝ツール」、「研究の責任・判断主体＝研究者＝人」という前提が変わりうる可能性である。生成 AI のみならず、「人もまた技術と時代の産物」という視点から、この双方向性の上に絶えず研究公正を置き直すことこそが重要なのであろう。

双方向性が問われるのは人と AI の関係性だけではない。第 2 章の大屋論文は、AI には「結果の管理」はできても「プロセスの判断」は不明であり、よって、AI 使用を規制する側・される側の双方に予測困難な状況をもたらす可能性が大きく、研究公正に関しても、撤回と追跡の可能性を担保したガバナンス、事前規制から事後規制への法の発想と立ち位置の転換が必要だと語っている。言い換えれば、生成 AI をめぐる問題とは、どこまでならば許され、どこからがアウトなのかという境界線を探ることではない、ということだろう。人の歴史は新しい技術との向き合い方の歴史でもあったのだから…。

第 3 章の樋笠堯士論文は、自動運転研究における倫理指針への対応の遅れを問題視しながら、研究開発とその社会実装において生まれる社会的責任、ELSI（倫理的・法的・社会的課題）に焦点を当てた。政府による研究開発支援においても、大学と産業界との共同研究に際しても、そして研究公正と市民社会との接点としても、ELSI は共有すべきキーワードである。

第 4 章の樋笠知恵論文は、タイトルに謳われた「法的課題」、すなわちプライバシー、知財（著作権）、特許と関わる法的な議論から構成される。特許法での「発明」の定義は人に限られており、ゆえに「生成 AI は発明者になれない」という明快な切り口は、「生成 AI はオーサー＝執筆者になれない」と読み替えることで、第 1 章で扱われたオーサーシップの問題とも重なる。

第 5 章の岡林論文は、生命系実験科学分野を中心に、研究活動において生成 AI を積極的に活用する段階（作業）を具体的に提示した上で、生成 AI と共に進める研究の未来を模索するものである。信頼という研究の最も大切な部分にも AI が絡むという想定は示唆に富む。

第 6 章の南里論文は、「広島 AI プロセス」が包含する教育を真正面からとりあげ、中等教育（中学、高校）の教育現場から生成 AI の可能性と諸問題に切り込む。AI パイロット校での実践例、特に大学入試の志望理由書の分析は興味深い。生成 AI は、志望大学の HP をはじめとするネット上の情報から志望理由をうまくまとめる一方で、「自己認識」のない AI が作成した理由書には個人の経験が欠落している——この指摘に、思わずはっとさせられた。現在の就活では、多くの企業がエントリーシートの分析・判断を AI に委ねている

が、学生もまた、AI を活用してエントリーシートを書いている（のだろう）。簡単な質問に答えればAI は即座に学生の強みを見つけ出し、かつそれを文章化までしてくれるのだ。思わず、「目には目を」ならぬ「AI にはAI を！」という言葉が浮かんだ。

しかしながら、長らく大学教員を務めてきた者としては、そうした行動のなかで、物事を他人事ではなく、自分事として考える姿勢はますます希薄化していくように感じる。大学の卒業論文指導において、「教育業務の効率化」の観点から生成 AI の有用性を捉えることには限界があるのではないだろうか。そうではなくて、教員の経験自体が、生成 AI の日常化という時代の変化に晒され、変わるべきなのだろうか。

第 7 章の藤井論文で興味深いのは、AI 開発において、技術リスクを考慮しながら進める「効果的利他主義」と技術革新を優先させる「効果的加速主義」という 2 つのせめぎ合いであり、それが企業と大学の共同研究においては、双方の文化的落差と重なって、倫理的課題を顕在化させる可能性への言及である。ここでも、生成 AI 以前から学術と社会の双方で人文学・社会科学分野に問われ続けてきた問題——この分野の研究の学術的・社会的貢献とは何か——が問われている。

## 5. 結びにかえて

本書全体をふり返るなかで、「生成 AI で研究公正は変わるのか」の先が少し見えてきた気がする。

ひとつには、生成 AI の活用において問われるのは、研究の基本だということである。研究とは何か、研究者とは誰か、研究のオリジナリティを担保するものは何か…。ゆえに、生成 AI によって研究公正自体が変わるわけではない。

もうひとつは、生成 AI を展開させつつある 2020 年代という現在が、人を、研究者をどのようにつなげようとしているのか、そこを熟慮する必要があるということである。そもそも、20 世紀末、Windows95 が IT 時代の幕開けを告げて以降のデジタル化の進展と呼応するかのように、研究の公正性が前景化されてきたのはなぜなのだろうか。歴史研究者である私にとって最も興味深いのは、この時代性——なぜ今研究の公正性なのか、である。

それは、ベルリンの壁崩壊（1989）、ソ連の解体（1991）に象徴される冷戦体制終結後を模索した時代であり、それまで米ソ対立という「大きな物語」のもとで抑えられ、封印されてきた「小さな物語」、すなわち、地域や民族、そして個人と関わる過去が、噴出しはじめる時代であった。それらは、「大きな物語」に収まらないという意味から言えば、「新しい過去」ともいえる。

その発掘に大きく貢献したのが、IT 革命であり、オープンサイエンスの推進であった。わけても、奴隷貿易や奴隷制度の過去についてはアナログ・データのデジタル化が加速度的に進み、それまで現地に行かないと見られなかった歴史資料が、ネット上で誰もが簡単にアクセスできるようになった。アメリカの大学が連携するウェブサイトの「奴隷の航海

(Slave Voyages)」, ロンドン大学ユニヴァーシティ・カレッジを中心とする「英領奴隷所有遺産プロジェクト (通称 Legacy Project)」はその好例である (井野瀬, 2023)。これらをベースに, 現在, 奴隷貿易や奴隷制度の過去への謝罪や和解, 賠償や補償が進められている。

こうした動きに敏感に反応したのが, IT 環境の整備とともに育った若者たち, 1990 年代後半から 2000 年代に生まれた Z 世代と呼ばれる若者である。2015 年 3 月, 南アフリカ, ケープタウン大学で始まった学生たちの草の根運動, 「ローズ・マスト・フォール (Rhodes Must Fall)」運動では, 旧来のカリキュラムに色濃く残る植民地主義に反発する学生たちが, イギリス植民地主義の象徴であるセシル・ローズ (1853-1902) の彫像の撤収を大学当局に求めた。「ローズ・マスト・フォール!」の叫びは SNS を通じてイギリスにも伝播し, ローズの遺産を元に設立された奨学金制度を持つオクスフォード大学でも学生たちの声が響き渡った。

2020 年 5 月には, ミネソタ州ミネアポリスで, アフリカ系アメリカ人, ジョージ・フロイドの死は, 警官に頸部を圧迫される様子が SNS でリアルタイムに拡散したことで, 「ブラック・ライヴズ・マター (Black Lives Matter)」の声を世界中に広げた。若者たちは, 奴隷貿易, 奴隷制度, 植民地主義の過去を現在の人種主義と直結させて, コロンブスやリー將軍 (南北戦争の南軍リーダー), エドワード・コルストン (17 世紀後半から 18 世紀にかけて奴隷貿易と関わったブリストル出身のイギリス商人) らの彫像を倒し, スプレーで落書きして回った (井野瀬, 2021)。

かくごとく, 20 世紀末以降の IT 革命, そして現在進行中のデジタル革命は, 「過去と現在の対話」を大きく変えつつある。そのなかで, 歴史教科書や郷土史を彩る「偉人」たちに突きつけられたのは, 歴史的正義・公正性という言葉である。その詳細に触れる紙面的余裕はもはやないが, 2024 年 10 月末, 2 年に一度開催されている英連邦首脳会議の声明に, 奴隷貿易によって被害を受けた旧植民地への賠償協議の開始が盛り込まれたことは, 「過去と現在の対話」の変質を如実に物語るだろう。賠償方法のひとつとして, かつてイギリスが植民地から略奪した文化財の返還も含まれるため, イギリスを含み, かつて植民地支配を行った欧米の博物館では, 新たな「対話」の準備が進められている (飯田, 2024)。

こうした動きは「知の脱植民地化」と呼ばれている。「脱植民地化」とは, 「植民地化されていた国々が宗主国から独立する過程」を意味するが, それは政治や経済のみならず, 文化や社会, それを支え担う人びとの精神を含めて考える必要がある。言い換えるならば, 私たちが対峙する人類知は, 所与としてそこに存在し, 存在し続けるわけではない。私たちはなぜその知識, 情報を知っているのか, なぜその知が今ここ (たとえば大英博物館の展示室) にあって, そこ (たとえばアフリカ諸国) にはないのか, それを対話すべき時代になったといえる (井野瀬, 2025)。

知に対して誠実である (あろうとする) 姿勢は, 研究公正という考え方の根幹に接続さ

れるべきものである。私たちが知的に誠実であるがゆえに、生成 AI が取り込む情報もまた更新されていく。この新旧の知識の交流から、生成 AI は何を出力していくのだろうか。私たちは、「知の脱植民地化」に代表される歴史の大きなうねりを通じて、研究公正が生成 AI とともに向かおうとしている行き先を改めて考える必要がある。

---

<sup>1</sup> 文部科学省（n.d.）が公開している各年度の研究不正事案（2016-02）によれば、ある大学の外国学部政治学教授の調査過程で既発表論文との比較検証をした結果、「当該論文全体の 44.9%が自己盗用（重複）」と特筆されている。当該事案の詳細は次の URL を参照のこと（最終閲覧日 2024 年 11 月 4 日）。

[https://www.mext.go.jp/a\\_menu/jinzai/fusei/1374208.htm](https://www.mext.go.jp/a_menu/jinzai/fusei/1374208.htm)

<sup>2</sup> 当該新聞記事は以下の URL を参照のこと（最終閲覧日 2024 年 11 月 4 日）。

<https://mainichi.jp/articles/20210911/k00/00m/040/097000c>

<sup>3</sup> 一般財団法人公正研究推進協会「2021 年度全国公正研究推進会議」を参照（最終閲覧日 2024 年 11 月 4 日）。[https://www.aprin.or.jp/seminar/seminar\\_detail/aprinkaigi2021](https://www.aprin.or.jp/seminar/seminar_detail/aprinkaigi2021)

## 【参考文献】

飯田卓（2024）「民族誌博物館のデコロナイゼーション：ヨーロッパと日本の博物館コレクションの形成をふまえて」『国立民族学博物館研究報告』vol.48-4, 475-525 頁

井野瀬久美恵（2021）「コルストン像はなぜ引き倒されたのか——都市の記憶と銅像の未来」『歴史学研究』1012 号, 41-50 頁

井野瀬久美恵（2023）「歴史研究と生成 AI」『史學雑誌』第 132 編第 7 号, 38-40 頁

井野瀬久美恵（2025）『奴隷・骨・ブロンズ—脱植民地化の歴史学』世界思想社

東洋英和女学院（2019）「関係者の皆様へ」（最終閲覧日 2024 年 11 月 4 日）

<https://www.toyoeiwa.ac.jp/2019/05/2019052402.html>

未来工学研究所（2021）「我が国の研究倫理教育等に関する実態調査・分析業務」（令和 2 年度 文部科学省 科学技術調査資料作成委託事業）

[https://www.mext.go.jp/content/20210524\\_mxt\\_kiban02\\_000015039\\_0.pdf](https://www.mext.go.jp/content/20210524_mxt_kiban02_000015039_0.pdf)

文部科学省（n.d.）「文部科学省の予算の配分又は措置により行われる研究活動において不正行為が認定された事案（一覧）」（最終閲覧日 2024 年 11 月 4 日）

[https://www.mext.go.jp/a\\_menu/jinzai/fusei/1360484.htm](https://www.mext.go.jp/a_menu/jinzai/fusei/1360484.htm)

文部科学省（2024）「研究不正事案の認定・公表状況（2015 年度～2023 年度）」科学技術・学術政策局研究環境課研究公正推進室公正な研究活動の推進に関する有識者会議（第 27 回 2024 年 5 月 31 日）配付資料 6（最終閲覧日 2024 年 11 月 4 日）

[https://www.mext.go.jp/content/20240531-mxt\\_kibanken01-000036776\\_05.pdf](https://www.mext.go.jp/content/20240531-mxt_kibanken01-000036776_05.pdf)

# Methods and Approaches for the Use of Generative AI in Research and Educational Activities

Rei Nouchi<sup>\*1</sup> and Kumie Inose<sup>\*2</sup>

This publication reflects presentations from the National Conference on the Promotion of Research Integrity FY2023, organized by the Association for the Promotion of Research Integrity (APRIN) on February 16, 2024. The Conference featured keynote speeches and other talks, followed by subcommittees divided into research fields, including life and medical sciences, science and engineering, humanities and social sciences, and secondary education. Satellite events were also held for researchers and research administrative staff involved in promoting research integrity.

A central theme, led by the Humanities and Social Sciences Subcommittee, was “Research Integrity in Technological Innovation.” This subcommittee aimed to discuss generative AI's impact on research practices and the core values of research integrity. However, due to time constraints, the discussion did not necessarily achieve this objective. Therefore, this publication compiles perspectives from presenters across subcommittees to provide a multidimensional analysis of the relationship between research activities and generative AI (or AI technology general), as well as its implications for research integrity.

The authors, representing diverse fields such as philosophy, legal philosophy, law, pedagogy, bioscience, bioethics, medical ethics, and secondary education, have collaborated to examine the impact of generative AI on education and research. They explore fundamental approaches to handling generative AI in their respective disciplines. The interdisciplinary nature of this publication enhances its value as a resource for understanding the evolving role of AI in higher education and research.

As part of the Reviews in Higher Education series, this publication highlights AI's extensive role in research, education, and university operations, emphasizing its significance in higher education. While not exhaustive, this work aims to be a valuable resource for promoting research integrity in the face of technological advancements.

We appreciate APRIN's support in providing conference recordings that contributed to this publication.

---

<sup>\*1</sup> Associate Professor, Hiroshima University

<sup>\*2</sup> Auditor, National Institutes for the Humanities

## 執筆者紹介

◎井野瀬 久美恵

人間文化研究機構・監事

大屋 雄裕

慶應義塾大学・教授

岡 林 浩嗣

筑波大学・講師

◎野内 玲

広島大学・准教授

樋笠 堯士

多摩大学・准教授

樋笠 知恵

信州大学・助教

藤井 基貴

静岡大学・准教授

南里 翔平

市川中学校・高等学校・教諭



研究・教育における生成 AI の利活用の方法とその考え方  
(高等教育研究叢書 177)

2025 (令和 7) 年 3 月 31 日

---

編 者 野内 玲・井野瀬 久美恵

発行所 広島大学高等教育研究開発センター

〒739-8512 広島県東広島市鏡山 1-2-2

電話 (082)424-6240

<https://rihe.hiroshima-u.ac.jp>

印刷所 株式会社 タカトープ rint メディア

〒730-0052 広島市中区千田町 3 丁目 2-30

電話 (082)244-1110

---

ISBN978-4-86637-054-5



---

Methods and Approaches for the Use of Generative AI  
in Research and Educational Activities

---

RESEARCH INSTITUTE FOR  
HIGHER EDUCATION  
HIROSHIMA UNIVERSITY

ISBN978-4-86637-054-5