AI出力物をめぐる三つの情報開示策とその変容可能性

渡辺 智暁　国際大学GLOCOM主幹研究員・教授・研究部長

1 はじめに

本稿は『智場』のGLOCOM創立35周年記念号のための論考である。筆者はGLOCOMを人文的な知にも、社会科学や理工系の知にも関わる越境性がある点が興味深い研究所だと思ってきた。また、情報社会論や、近年でいえばAIガバナンスの議論のように、テクニカルで緻密な議論だけでなく、批判的思考や大胆な想像力など、さまざまな知的なアプローチが貢献できるテーマを扱う点も刺激的だと思ってきた。筆者がそれらをうまく体現できているわけではないが、本稿では少しそれに近い議論を試みたい。

題材として取り上げるのは生成AIの出力物に関連した三つの異なる情報開示策やその政策課題で、これをやや巨視的な観点から論じてみたい。各国の政策動向の分析や既存の法制度との整合性などテクニカルな検討ではなく、長期的にどのような社会変動が起きつつあるのかを考えるなかから政策課題の性質について考察することを試みる。そのような考察によって長期間にわたる息の長い示唆を与えることができることを願っているが、当たらなかった予測が情報社会論には大量に存在しているため、楽観はせず、本稿で提示する可能性を考えただけで得られる示唆にも意義が見いだせるように努めた。

取り扱う課題はいずれも、生成AIの出力物と人間が創作・制作したものとの判別に関するものである。それを解決するために何らかのメタデータ・追加情報を付与し、それによってある情報資源がAI出力物か人間の制作物かの情報開示を行う、というようなことが政策としての共通点になっている。反面、個々の政策で論じられる領域や文脈はかなり異なっている。

2 生成AIと情報環境の変容

本稿で注目したいのは、生成AIの出力が、人間の創作物・制作物などと判別しにくくなっているというよく知られた事態である。文章、画像、音声、動画などの領域でこれが起きている。以下ではこれを「判別困難性」と呼ぶ。この判別困難性は、生成AIを相談相手のように用いる利用行為、大学の課題レポートや投稿小説などの評価・審査上の困難、ディープフェイクやフェイクニュースの深刻化など、さまざまな問題と関わっている。AIの出力がいかにも機械的で人間の表現とかけ離れていれば、こうした問題の多くはここまで大きくはならなかっただろう。

この判別困難性とあわせて重要なのが、生成AIの出力が人間に比べて高速であり、大量に生成でき、しかも手軽で安価に利用できることである。判別困難性だけでなく、この効率性の高さが、情報環境の変

化を加速させている。たとえば、文芸作品の執筆・投稿やSNS上の大量発信において、以前なら相応の時間や動機が必要だった行為が、安価な副収入目的、単なる実験や悪ふざけとしても容易にできるようになった。人間が時間をかけて作ること（したがって、それ相応の動機を持った者だけがそうした行為に関与すること）を暗に前提していた制度や慣行が、生成AIによって揺らぎ始めているとも言えるところがある。

たとえば示唆的な例として、小説の投稿を受け付けていた複数の雑誌がチャットGPT（ChatGPT）の台頭からほどなくして受付を停止し、その理由としてAIを利用した質の低い投稿が増えたことを挙げていたことと、最近では（AI利用も公式に認めている）星新一賞においてAIを利用した作品が実際に上位を占め、審査員にも区別がつかなかったというコメントがあったことを挙げることができる[*1]。

もっとも、こうした変化がどの程度の速度と規模で進むか、あるいは停滞や縮小が起こるかは不明瞭である。生成AIの料金設定、競争環境、AIの性能向上のペース、開発や利用に必要になる計算資源や電力の供給、政策論議や訴訟の行方、開発・運用を支える投資など、多くの要因がこれに影響するだろう。ただ、少なくとも一定の拡大は続く可能性が高いと仮に前提してみる。そのうえで以下では、まず判別困難性に関わる三つの具体的な政策課題を取り上げ、その後で、それらが長期的にはどのように変容しうるかを考える。

3 AIによって生成された写実的表現の識別可能性担保

（1）課題と対策の概要

判別困難性をめぐる政策課題のなかで、近年最も広く論じられているのは、ディープフェイクやAI生成コンテンツを含む偽・誤情報への対策である。もちろん偽・誤情報は生成AIに限らず存在するが、写実的な画像・音声・映像が容易に生成できるようになったことで、問題の性質は変わりつつある。人は一般に、動画や音声、写真のような表現を、単なる文章よりも高い信憑性（しんぴょう）を持つものとして受け止めがちである。そのため、AIによる写実的表現が虚偽の内容を伴って流通すると、誤認を招く危険が大きい。

この問題に対しては、主として二つの方向がある。一つは、AI出力物であることの表示やラベル付けを求める法的・制度的対応である[*2]。もう一つは、コンテンツそれ自体に電子透かしや来歴情報を付与し、どのような機器・ソフトウェア・主体によって生成・編集されたかをたどれるようにする技術的対応や、発信者が信頼に足る組織であるかどうかを特定しやすいようにすることである[*3]。いずれも、AI出力物であること、あるいは逆に信頼できる発信源によるものであることを識別可能にする点で共通している。

（2）困難と限界

こうした取り組みの有効性やその度合いについては、いろいろと限界も考えられる。第一に、何をAI出力物として扱うかの定義が容易でない。AIが生成した画像や文章を人間が部分的に修正した場合、それはなおAI出力物なのか。AI出力を下敷きにして別の人間が加工した場合はどうか。政策目的に照ら

して定義されることが理想だが、線引きは容易ではない。

第二に、こうした開示策は悪意ある主体に対しては順守が期待できず、実効性が限られがちである。偽・誤情報流布や情報操作を意図する主体は、そもそも表示義務を課されても守らず、情報が自動的に埋め込まれてもそれを削除する動機を持ちうる。この種のメタデータの削除や偽装が容易であれば、ＡＩ出力物を使った偽・誤情報の流布がもたらす被害の抑制策としての効果も限られたものになるだろう。その場合には、これらの対策は問題を完全に防ぐというより、少なくともコストを引き上げ、大規模な悪用をやや困難にする措置として位置づけるほうが現実的であろう（設計がまずいと、悪意を持った最も厄介な主体は規制できないが、幅広く一般の人々の言論活動に負担を与える、という可能性もある点にも注意が必要だろう）。

第三に、来歴情報の付与は匿名性やプライバシーとの緊張関係を生む場合がある。報道関係者への匿名の情報提供や公益通報のように、情報の出所を秘匿することに価値がある場面もある。より一般的に匿名での言論の自由を保護することにも意義がある。そのため、来歴の可視化を進めるとしても、どの程度の情報を、誰に対して、どの条件で開示するのかには注意深い検討・設計が求められることになろう。もっとも、こうした限界を考慮しても、信頼されるべきコンテンツには一定の来歴情報が付与されるという慣行が定着すれば、情報環境の信頼性を部分的に改善できる可能性はあるだろう。

ただし、ディープフェイクやフェイクニュースの影響は、真偽が判明すればそれで自動的に解消するわけではない点には注意が必要だろう。風刺、誹謗中傷、アダルトコンテンツ、党派的言論のように、偽・誤情報であることが知られつつ流通・消費される場合もある。問題はメタデータやラベルによる対策だけで解決できる範囲よりも広いと考えるべきであろう。

4　創作物へのＡＩ利用の有無を開示すること

（1）課題と対策の概要

第二の課題は、教育、資格試験、コンテスト、投稿作品審査などにおいて、応募作品や答案へのＡＩ利用の有無がわからなければ、受験者や応募者の資質などが評価できない、という問題である。これは偽・誤情報対策とは異なり、写実性や真偽の問題（現実と虚構の判別が困難になっている問題）ではなく、誰の能力や努力の成果なのか否かの判別が困難になっていることの問題とも言える。たとえば、文章作成能力を測る課題であれば、ＡＩの全面的利用を禁じたり、利用した場合には申告を求めたりすることに一定の合理性がある。

この種の方針は、無断のＡＩ利用を、ある種の剽窃に近いものとして扱う場合もある。ただし、通常の剽窃と異なり、ＡＩ出力物には明確な「被害者」が存在しないことも多い。その意味で問題なのは、他人の成果を奪うことより、自らの能力や努力について不当な評価を得ようとする点や、それによって競争上優位に立とうとする点にあるとも言えるだろう。

（2）困難と限界

この領域での開示や禁止にも困難が多く考えられる。第一に、どの範囲のＡＩ利用を問題視するか、線引きはここでも難しいように思われる。ＡＩによるレポート丸ごとの代筆禁止はよいとして、着想の補助、文献検索、文献の要約・翻訳、自分の原稿への反論の提示、推敲（すいこう）や添削の支援はどうか。さらに、検索エンジンや文書作成ソフト、入力補助、字幕生成など、日常的なデジタル環境にもＡＩが埋め込まれている。何を禁止し、何を許容するかは自明ではない。

第二に、違反の発見や証明が難しい。ＡＩ利用について正直な申告を求める方式は、隠したい動機を持つ主体には弱い。作成履歴の提出を求めるなどの対策も考えられるが、それで全面的に防げるわけではない。ＡＩの出力を見ながら本人が打ち直すことや、完成原稿に対してＡＩに改善提案を求め、それを自分で打ち込むことで反映させることは十分可能であろう。また、その履歴を評価者が細かく確認するコストが高すぎる場合もある。資格試験の中にはデジタル機器が使えないように規則と監視によって制限し、能力を測るアプローチがとれるものもあるが、これは実施コストが高くなりがちである。

第三に、そもそも何の能力を測るためにＡＩ利用を制限するのかも問われうる点であろう。ＡＩを使わずに文章を書ける能力を測りたいのか、それともＡＩを用いつつ情報を評価・編集・改善する能力を重視するのか。生成ＡＩが今後さらに普及し、高性能かつ低コストになるなら、実社会ではＡＩを用いて文章や情報処理を行う場面が増えるだろう。そのとき、ＡＩ抜きの執筆などの能力だけを重視することがどこまで妥当かは再検討を要する課題のように思われる。

もっとも、ＡＩ利用を前提とする方向にも反論は考えられる。社会の制度や慣行はすぐには変わらず、当面は依然としてＡＩを使わない能力が要求される場面が残るかもしれない（手書きの履歴書や押印といった慣行が想起される）。また、教育の目的は労働市場への適応だけではなく、伝統の継承、市民的能力の育成など多面的である。したがって、ＡＩを使わずに文章を構想し、書き上げる訓練にもなお一定の意義があると考えられる。芸術の領域でも、人間が挑戦し達成することそれ自体に価値が見いだされる場合がある。ＡＩや電子機器に依存しない社会的機能・能力の維持も、電磁パルス（ＥＭＰ）等の広域的リスクを考えるならばここに付け加えてもよいかもしれない[*4]。

5　著作権のないＡＩ出力物を識別可能にすること

（1）課題と対策の概要

第三の課題は、これまでの二つとは逆向きの性格を持つ。前二者ではＡＩ出力物はどちらかといえば警戒や制限の対象とされているが、ここで課題になるのは、むしろＡＩ出力物が著作権法の制約のない資源として利用可能である場合に、そこから得られるメリットを広く享受できるように、それを識別可能にすることである。

日本の著作権法の下では、多くのＡＩ出力物には著作権が認められないと考えられる。ＡＩ自体は著作者になれず、ＡＩ利用者の関与が創作的寄与といえるほど強くない場合、その出力物は著作物ではないか

らである。そうだとすると、ＡＩ出力物は、著作権の保護を受けない情報資源として自由に利用できる可能性がある。しかし、ＡＩ出力物と人間の創作物の判別困難性があるため、「自由に使えるＡＩ出力物であろう」と思って使ったものが実は人間の著作物であり、結果として著作権侵害に至るおそれがある。逆にいえば、著作権のないＡＩ出力物を識別可能にすることは、自由に利用可能な資源を社会のなかで増やすことにつながる。

この課題は、デジタル・アーカイブやオープンデータの領域にある種の先例を見いだせる[*5]。これらの領域では、実際にはパブリック・ドメインに属するはずの、あるいは著作権の保護対象でないはずの情報資源に対しても、あたかも権利が存在し、その権利をアーカイブ提供機関やデータ提供機関が保有しているかのような形で利用許諾が与えられ、利用条件が付されることが国際的に見られる。対照的に、たとえばＥＵのデジタル・アーカイブ機関群を横断的に閲覧できるプラットフォームEuropeanaではライツステートメンツ（rights statements）と呼ばれるメタ情報を付加し、著作権を考慮せずに自由に使える資料についてはそれと判別しやすいようにしている[*6]。

（2）困難と限界

この政策にも困難がある。第一に、ここでもＡＩ出力物の範囲をどう定義するかが問題になる。とりわけ、既存の著作物に強く依拠し、結果として類似した出力や、人間が部分的に創作的な加工を加えた場合には、著作権の有無は単純に判断できない。既存の著作物と類似した出力物があるとして、そこに著作権侵害に当たるほどの類似性があるかどうかも、一律の基準にしにくい。

第二に、ＡＩ出力物であることを隠したい動機を持つ主体への有効性も限界がある。企業や個人が、広告やウェブサイト、イラストなどに用いた出力物を、あたかも自らの著作物であるかのように扱いたいと考えることは十分ありうる。それらをＡＩ出力物だと明示すれば、著作権法上の排他的権利を主張できず、他人による再利用が容易になるためである。

第三に、代替策として人間の著作物の側を登録・証明する制度も考えられるが、これにも限界がある。登録を安価・簡便にするとＡＩ出力物などを人間製と偽って大量に登録する行為など濫用を招くおそれがあるだろう。コストを高くすると登録が進まない。現行の日本の著作物登録制度がそれほど広く使われていないことを踏まえると、大きな期待はしにくいように思われる。

第四に、虚偽の来歴記録を作成できないようにし、自動的に生成されるようにすると、プライバシー侵害や国家による監視などの被害をもたらすリスクが高くなる点も課題だろう。言論活動自体への萎縮効果も発生しかねない。

それでも、ＡＩ出力物をある程度識別可能にすることには意義があるだろう。大多数の利用者が正直に開示し、著作権のない資源が流通しやすくなれば、それだけでも情報資源の利用可能性は広がるからである。被害を食い止めるタイプの政策は困難や限界に関してシビアな検討を強いられがちだが、便益を拡大するこの種類の政策は、ある程度有効であればそれでよいという妥協もしやすいところがあるだろう。

6 生成AIと情報環境の変容

以上、生成AIの出力物が人間の創作物などと判別困難であることに関連する課題と対応策を手短に概観してきた。最後に、冒頭に挙げた長期的な文脈に照らして、こうした課題が大きく変容する可能性を指摘してみたい。

人間の創作・制作には時間がかかるが、AIの出力はかなり高速であり、バリエーションのある出力物を大量に生成することができる。そうすると、人間の創作物や制作物よりも、多くの情報資源は安価になり、市場競争が働くところではAI出力物が選ばれることも多いと思われる。また、安価な情報が大量に作成され、流通することで、受け手側もAIを用いて選別・受容することが考えられるだろう。特に取引にあたっての情報のやりとりや情報探索、学習など情報の認知的な処理のみが必要な文脈では、大量の情報を人間が確認・評価・整理するのは時間がかかり、人間の関与がボトルネックとなってしまうため、AIが担う部分が増える可能性があろう。いわば情報の大量発信・大量受信双方にAIが用いられるような状況で、人間はそれぞれのAIの後ろにいるような構図である（芸術・娯楽・儀礼的コミュニケーションなど人間が体験・関与することに意味がある領域では少し事情が異なる可能性があるだろう）。AIがどのように統治され、社会で利用されるべきであるかについての議論では人間中心であることがしばしば重視されるが、すでに人間の関与を減らすほうが物事を円滑・効率的に処理できる場面が増えつつあるようにも見える現状との関係で、「人間中心」とは何なのか、それを意味ある形で維持できるのかは、AIが大量の情報の送受信両側に用いられる可能性を考えても、引き続き問われ続ける課題になるように思われる。ポストモダニズムが見た「作者の死」がわかりやすい形で実現しつつあるように見えたり、人間が道具として言葉を使って発言をするのではなく、言語の秩序自体が個々の文章を生み出していてM・フーコーの言う「人間の終焉」が到来しているように見えたりする時、従来の人間観もまた変更を迫られる可能性もあろう[*7]。

そこまで大きな変容を考えなくとも、さまざまな制度の形成によって政策課題が変動していく可能性はあるだろう。偽・誤情報流通の課題については、AI生成偽・誤情報が大量に流通し、大きな影響力を持つ可能性もあるものの、反対に偽・誤情報の影響が弱くなる可能性もあるだろう。AIを用いた情報収集などに際して、信頼性の高い発信者であることが広く認知されやすい大手報道機関や政府機関などが重んじられ、それ以外の発信者のリーチや影響力が相対的に低下する形で緩和される可能性が考えられるためだ。信頼できない情報が増えた結果、それを排除する技術も発達した領域として電子メールを挙げることができる。スパムや詐欺メールは、そのような技術によって影響力は弱くなっただろう。これがネット上の言論やニュース流通の領域でも起こる場合、偽・誤情報対策としては問題が緩和され望ましいわけだが、信頼性が確かではない一般人や著名ではない言論人などの言論の衰退が起こること自体は望ましいこととは限らない。つまり別の課題が浮上する可能性がある。また、受け手がAIである場合には、画像や動画、音声などが「写実的」であるからといって説得力があるとは評価されなくなる可能性も考えられるだろう。AI間の情報伝達に関しては、人間にとって最適な表現形式とは異なる形式が発達していくことも考えられるだろう。もっとも、人は情報収集や分析のためだけにニュースを消費しているわけではないから、AIが普及・浸透した社会にあっても、良くも悪くも娯楽性のあるニュースコンテンツが流通し、人が直

接の受け手となってそこから影響を受け続けている可能性もあるだろう。

小説の投稿受付やその審査については、大量応募を前提に仕組みをする、人間の審査にかかるコストに充当するべく応募料を徴収するなどの対策が有力になると思われる。また、学校教育の文脈では文章などの作成のためにＡＩを利用するようなアウトプットのためのスキルだけでなく、情報の収集・評価・整理などインプットにもＡＩを使うスキルを養うことが重要になる領域が増え、ＡＩ利用を禁止する、ないし利用について開示を求めるべきケースは減っていくと思われる。

自由に利用可能な著作物については、ＡＩ出力物が量的に人間の創作物をはるかに上回るようになる世界でも、人間の創作的な関与がされていないものだけを識別できることは重要なままにとどまるかもしれない。もっとも、人間の創作物であれ、ＡＩの創作物であれ、それを人間が直接受け取る機会が減り、ＡＩが受け取ることが増えるとしたら、創作のインセンティブ自体が減ってしまう可能性も考えられなくはない。本稿では十分検討できなかったが、このようにＡＩが普及・浸透した社会では、人間は他の人間とのコミュニケーションを多くとり続けるだろうか。コミュニケーションの相手も、ＡＩが多く担うようになるだろうか。後者であれば、人間をターゲットとした創作活動自体が減るかもしれない。

こうした状況が訪れることになるかどうか、それがどの程度近い将来でありうるかについての検討は、本稿では概ね省略しているわけだが、ＡＩの普及・浸透というさほど小さくないと思われる可能性にいくらかの仮定を追加して検討してみただけでも、本稿で取り上げたような政策課題の変容や、政策目的の変更も伴うような別の政策への要請が浮上する可能性がうかがえることは留意に値するように思われる。

現在までのところ、ＡＩをめぐるガバナンスは日本ではアジャイルであることが重んじられる傾向にあり、変化に即応することが比較的容易である。また、本稿で取り上げた課題の中でも実際に議論や制度形成が最も進んでいるように見える偽・誤情報対策が典型だが、単一の主体による効果的な対策を打つことが難しく、さまざまな困難や限界があるため、さまざまなステークホルダーが協力・連携しながら問題の緩和に取り組むといった形をとっていることもある。だが、そのようなステークホルダー連携の目的や体制もまた、変更を迫られる可能性があるように思われる。これは大変なことではあるが、立法府がさまざまな理由で即応性が弱くなりがちであることや、情報の収集や検討に時間がかかってしまいがちなことを考えると、ステークホルダー参加型のアジャイルガバナンスは相対的に優れており、長期的に起こる可能性がある変化の幅を考えても、維持する意義があるアプローチであるということになると思われる。

註

＊1　２０２３年２月に小説投稿の受付停止については各所で報じられた。たとえば特に参考になるものとして、Sato, M. (2023) "AI-generated fiction is flooding literary magazines – but not fooling anyone," *The Verge*, Feb 26, 2023. https://www.theverge.com/2023/2/25/23613752/ai-generated-short-stories-literary-magazines-clarkesworld-science-fiction (最終閲覧 2026/5/10)
また、星新一賞について
鏡明（２０２６）「第13回星新一賞総評」https://hoshiaward.nikkei.co.jp/result.html（最終閲覧 2026/5/10）

＊2　法的対応の例として、
「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」第７条（広島ＡＩプロセスに関するＧ７首脳声明別添）https://www.mofa.go.jp/mofaj/ecm/ec/page5_000483.html（最終閲覧2026/5/10）
EU AI Act Article 50(2). https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng（最終閲覧2026/5/10）
SB 942: California AI Transparency Act (2023-2024). https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942（最終閲覧2026/5/10）
など。なお、これらはいずれもＡＩ開発者や提供者向けに義務を課すものになっている。

＊3　技術的対応の例として、動画生成サービスＳｏｒａにおける透かしの自動埋め込み、Ｃ２ＰＡ（Coalition for Content Provenance and Authenticity）https://c2pa.org/（最終閲覧2026/5/10）、オリジネーター・プロファイル https://originator-profile.org/（最終閲覧2026/5/10）などを挙げることができる。

＊4　電子機器類や通信設備などが電磁パルス（ＥＭＰ＝electromagnetic pulse）によって広域的に故障する事態を想定する場合、デジタル技術なしで社会が機能する可能性の担保は重要になるだろう。ＥＭＰ概説として、
一政祐行（２０１６）「ブラックアウト事態に至る電磁パルス（ＥＭＰ）脅威の諸相とその展望」『防衛研究所紀要』18（2）：1-21 https://www.nids.mod.go.jp/publication/kiyo/pdf/bulletin_j18_2_1.pdf（最終閲覧2026/5/10）
ただし、被害想定面ではやや古くなっていると思われる。

＊5　この問題について論じたものとして、
渡辺智暁・小林心（２０１９）「クリエイティブ・コモンズ ── オープンソース、パブリック・ドメインとの関係からの考察」『パテント』72（9）：34-47　https://jpaa-patent.info/patent/viewPdf/3373（最終閲覧2026/5/10）

＊6　Europeanaにおけるライツステートメンツについては、
Europeana (undated) "Understanding the rights statements used by Europeana." https://pro.europeana.eu/page/available-rights-statements（最終閲覧2026/5/10）

＊7　Ｍ・フーコー著、渡辺一民・佐々木明訳（１９７４）『言葉と物 ── 人文科学の考古学』新潮社
また本書を参照しつつＡＩガバナンスにおける人間観について批判的に検討した論考として

Ryan, M. (2025) "We're only human after all: a critique of human-centered AI," *AI& Society*, 40: 1303-1319. https://doi.org/10.1007/s00146-024-01976-2

GLOCOMウェブサイトにて、全文ダウンロードいただけます。

印刷版は、Amazon.co.jpにてご購入いただけます。

https://www.glocom.ac.jp/publicity/chijo/11795

智場 #125
情報社会研究のポリローグ ──AI、ガバナンス、智の実装

責任編集	伊藤 将人、小林 奈穂
監修	砂田 薫
編集・制作進行	武田 友希
発行人	松山 良一
発行日	2026年7月9日
発行所	国際大学グローバル・コミュニケーション・センター（GLOCOM）
	〒106-0032 東京都港区六本木6-15-21 ハークス六本木ビル 2F
	URL https://www.glocom.ac.jp/
	TEL 03-5411-6677
	E-mail inquiry-glocom@glocom.ac.jp
印刷・製本	株式会社 紙藤原
校閲・校正	濱田 美智子
表紙・装丁	株式会社 コンセント

『智場』は、国際大学グローバル・コミュニケーション・センター（GLOCOM）が発行している機関誌です。
「智場」は、「知識や意見の交換と流通の場」を意味する言葉です。
1995年に創刊され、情報社会学のフロンティアに挑み続けています。