AI 時代の研究管理に関する原則の宣言（和文仮訳）

[Statement of Principles: Research Management in the Era of Artificial Intelligence]

前　文[1]

世界の研究資金提供機関は、優先すべき研究テーマの選定、研究資金提供プログラムの設計、研究提案募集の定式化、審査員データベースの構築、堅牢なピア／メリットレビュー方式の確立等を通じて、研究活動の舵取り、方向付けにおいて重要な役割を担っている。また、採択及び交付決定に係る一連の手続きを管理し、定期的な研究報告評価や資金提供した研究のインパクト測定、研究資金に関する方針及びガイドラインの策定も行っている。AI は、これらのプロセスの質を向上させ、処理の迅速化に役立つ可能性があり、適切なガバナンスの下で運用されるならば、倫理的・法的枠組を遵守しながら、これらプロセスの一部を効率的に自動化できる潜在性を有している。

AI は、過去5年間で著しい進展を遂げ、様々な側面において人間が生み出したものと見分けがつかないようなコンテンツを生成できるシステムが構築されるようになった。現在の AI システム、特に基盤モデル（Foundation Models）[2]は、膨大な量の人類の知識をコード化し、人間と同じレベルの推論を必要とする多くのタスクを自動化するために広く使われている。

グローバルリサーチカウンシル（GRC）は、研究資金提供プロセスにおける卓越性と公平性を支援するため、厳格で透明性の高い審査システムに必要な、中核的且つハイレベルな原則に対する世界的な合意をもたらし、国境を越えて協力する研究資金提供機関の信頼関係構築を目指してきた[3]。また、本合意はピア／メリットレビューシステムや研究コミュニティの差異に対する寛容さの基盤ともなっている。以降に示す研究管理における AI 導入に関する原則は、GRC のメリットレビューの原則に沿ったものである。

GRC 参加機関は、研究資金提供機関が以下の原則を採用すべきことを提言する。

---

[1] グローバルリサーチカウンシルは、任意且つ参加機関主体の組織であり、各国の研究エコシステムの中で参加機関がそれぞれ異なる使命、任務、責任を負っていると認識している。GRC の立場、決定、声明は、参加機関に対して拘束力を持たない。このような声明等への支持は、GRC 参加機関が各国の政策や優先事項に沿った形で声明を採用する可能性があることを表すものである。

[2] Bommasani, R., Hudson, D., Adeli, E., Altman, R., Arora, R., Arx, S., Bernstein, M., Bohg, J., et al. (2022). On the Opportunities and Risks of Foundation Models, ArXiv.

[3] GRC, "Statement of Principles on Peer/Merit Review," 2018. [Online]. Available: https://globalresearchcouncil.org/fileadmin/documents/GRC_Publications/Statement_of_Principles_on_Peer-Merit_Review_2018.pdf.

AI の導入

AI は、様々な領域において導入が進んでおり、効率を格段に高め、プロセスを最適化する可能性を秘めている。AI の導入は、研究の作業効率を上げ、意思決定を合理化し、イノベーションを推進する貴重な機会を提供する。研究資金提供機関は AI の重要性を理解すべきであるが、リスクを最小限に抑えながらその恩恵を最大化するため、AI の導入には規制を設け、ベストプラクティスに沿って実施するべきである。適正に管理された AI の導入は研究管理における AI の効率的かつ責任ある利用を保証するものである。

意思決定

研究提案に関する最終決定は人間が行うべきである。AI システムは、予測不能な動作をするなどの恐れがあり、AI システムのみで意思決定を行うべきではなく、人間の監視が必要である。AI システムは人間をサポートするものであり、人間に取って代わるものであってはならない。

デジタル・ディバイドの解消

研究管理における AI の導入は、既存のデジタル不平等 [digital inequalities] を拡大するものであってはならない。AI の開発と配置には、相当なインフラと専門的なスキルが必要であり、特に発展途上地域では大きな課題となる。AI 技術への公平なアクセスを確保するためには、こうした不均衡に対処することが不可欠である。デジタル・ディバイドを緩和するためには、先進的なインフラ、リソース、AI 技術へのアクセスを拡大し、リソースに乏しい地域を支援すべきである。オープンソースの AI システムの採用を奨励し、包摂的で世界的に利用可能な AI トレーニングプログラムを開発することは、地域の能力を高め、研究管理における AI 利用への幅広い参画を可能にする。さらに、AI サービスやツールへの確実なアクセスを保証するためには、リソースに乏しい地域におけるアクセシビリティを改善することが極めて重要である。

国際協力

国際協力の促進は、研究管理における AI 導入を進める上で不可欠である。GRC 参加機関は、国境を越えてベストプラクティスやリソース、専門知識を共有することにより、特に技術力、リソースが限られている地域において、広く AI 採用を促進することができる。国際協力の強化は、AI ツールへのより公平なアクセスを確保し、能力構築・向上の取組を支援し、世界中で研究管理における AI の効果的かつ責任ある利用を促進する。

バイアスと公平性

AI を活用した研究管理において公平性を確保することは、意思決定プロセスにおける公平性、包摂性、透明性を維持するうえで極めて重要である。AI は人間の持つバイアスを低減する可能性を秘めているが、AI が偏ったデータに基づき訓練された場合、低代表研究コミュニティ [underrepresented

research communities] の疎外化を招くような、他の新たなバイアスを導入する恐れもある。公平性を支援するためには、AI システムが、差別的なバイアスを回避しつつ、（マイノリティであるコミュニティを含む）すべてのコミュニティに関わりのある重要な要因を適切に捉え、多様で、包摂的であるという目的に適ったデータセットにより訓練される必要がある。ピアレビューや研究評価において AI を使用する際は、AI を既存の中立的な審査方針やガイドラインに適合させ、資金提供の決定や研究の優先事項等への不当な影響を防止する必要がある。さらに、文献計量データに埋め込まれた言語的バイアスは、研究のインパクトや分野固有の進展に対する誤った認知を形づくる可能性があり、慎重に評価する必要がある。AI 主導の評価は、周縁的な研究コミュニティの貢献が可視化され、尊重され、研究コミュニティが繁栄するための適正なリソースが確実に提供されるよう、その貢献を正しく認識し、支援するよう設計されなければならない。これらの課題に対応するためには、研究資金提供機関が、研究管理において、公平で包摂的な責任ある AI ガバナンスに積極的に取り組むことが不可欠である。

透明性と説明責任

透明性と説明責任は、研究管理における AI の利用において、リスクを軽減し、責任ある展開を確実にするために極めて重要である。また、AI システムは誤解を与えかねない、予期せぬアウトプットを生成する可能性があり、時に予測できない動作をすることや、多くの AI モデル、特に基礎モデルの内部動作はまだ十分に理解されていないこと、AI システムは意図的な操作や悪意のある干渉を受けやすいことを認識することが不可欠である。透明性を高めるためには、AI 利用ガイドラインにおいて、AI システムを利用することができる者・時間・場所を明確に定義すべきであり、AI ポリシーに、欺瞞的・操作的な AI の使用を抑止するための方策を組み込むべきである。さらに、AI によるすべての決定は説明可能な状態、検証の対象とするべきであり、非合理的な、予期せぬアウトプットについては慎重に評価し、信頼できない場合は利用せず、説明責任を確保するため公式に報告すべきである。

プライバシー、データセキュリティ、知的財産、知識

研究管理のあらゆるプロセスにおいて AI を利用するには、機密データや独自のアイデアを不正使用・不正アクセスから確実に保護する必要がある。研究提案書の保管は、著者の同意に基づき、機密性、完全性、可用性を保証するセキュリティ基準を遵守しなければならない。著者の同意を得る際には、データの保存方法、場所、期間、およびデータが閲覧可能な者について明記すべきである。研究提案書の取り扱いは、著者に明らかにされる必要があり、AI システムの訓練等、AI を更に向上させるために研究提案書を使用する場合には、著者の明確な同意を得たうえで実施しなければならない。さらに、研究提案書には、研究者によって発案された今までにない技術、アイデア、知識が含まれており、研究管理における AI の導入が、知的財産の侵害、知識の誤った帰属、知的財産管理の阻害を引き起こしてはならない。また、提出された研究提案書をもとに新たなコンテンツを作るなど、

生成系 AI を悪用したり、デザイン、図表、マルチメディアコンテンツ、クリエイティブテキストを含むその他発明を意図的に操作してはならない。

AI リテラシー

AI システムとその限界についての認識を高め、AI システムを利用する人々をトレーニングすることは、AI システムの恩恵を最大化し、規則・方針等に定められた原則の誤用や違反を減らすために不可欠である。AI システムを操作または使用する者は、完全なコンプライアンスを確保し、ツールの限界を理解するために必要なすべてのトレーニングを受けなければならない。

持続可能な開発

一部の AI システム、特に基盤モデルは、膨大なエネルギー、水、そしておそらくその他の貴重な天然資源を消費する特殊なインフラを使用している。AI の使用は持続可能な開発を支援すべきであり、一般的に、AI の導入が、環境や生態系に害や悪影響を及ぼすものであってはならない。

行 動

GRC 参加機関は、研究資金提供機関が以下の行動をとるべきであると提言する。

1. 自由に利用可能なオープンソース AI モデルを推進し、他の研究資金提供機関がそれらを利用できるようにする。さらに、GRC とその参加機関は、必要な計算能力がより少ない AI モデルの構築に関する研究を支援すべきである。これにより、国際協力を拡大し、デジタル・ディバイドの解消を支援し、持続可能な発展を維持することができる。

2. 導入された AI システムがどのように訓練されたものであり、どのように使用されているか、また訓練に使用されるユーザーデータや研究提案書がどのように保存・処理されているかを公表し明らかにすることで、社会的信頼を維持する。これにより、透明性を高め、人間の監視を確保し、バイアスを軽減し、データのプライバシーとセキュリティを保護する。

3. プライバシー及びセキュリティ保護に関するすべての合意に準拠した公開訓練データの公表を促進する。ただし、訓練データの共有は推奨されるものの、共有されるデータがユーザーのプライバシーや知的財産権を侵害したり、個人または組織のセキュリティに危害を及ぼすものであってはならない。公開訓練データの共有は、国際協力を促進し、デジタル・ディバイドの是正を支援し、研究管理における AI 利用を拡大するものである。

4. 研究管理に利用されている AI システムの継続的な評価を実施し、そのパフォーマンスを公表する。これにより、AI システムの限界に対する認識が深まり、これらのシステムを意思決定に活用する際の判断材料ともなる。また、社会的信頼の維持にも寄与する。

さらに、GRC 参加機関は以下の点に合意する。

1. AI に関する作業部会設立の可能性について調査・検討する。参加機関が作業部会の設立に合意した場合、その作業部会のリーダーは、AI の導入に高い成熟度を有する参加機関から選出されることが提案される。