連載

臨床倫理メディエーション

国立大学法人山形大学医学部　准教授　中西　淑美

79 生成ＡＩ（人工知能）と倫理

はじめに

年々進化を続ける生成ＡＩ（the Artificial Intelligence：人工知能）は、われわれの生活を大きく変える可能性を秘めているが、同時に倫理的な問題も提起している。大規模言語モデル（large language model：ＬＬＭ）を搭載する生成ＡＩについては本連載でもシンギュラリティ（singularity：技術的特異性）などの論考と倫理について述べた[1]。

生成ＡＩの問題点は、①能力（competence）、②守秘義務（confidentiality）、③倫理・利益相反（Ethics/conflict）である[1]。頻出される幻覚現象（hallucination）についてはあまり問題にならないと言われている。その理由は、秀逸な専門職や専門家集団はそれを見破ることができるし、真偽を見破るツールもあるからである。

現在、生成ＡＩは、次の言葉を予測する、言葉と言葉のネットワーク化を大規模に行い、言葉をたくさん理解することで、他者の心を理解する能力にまで進化しようとしている。本稿では、あらためて生成ＡＩと倫理について考えてみたい。

1．主な倫理的問題点と留意点

まず、生成ＡＩの倫理的な問題点をいくつか挙げてみよう。

(1)プライバシー侵害がある。ＡＩは膨大なデータを記録・学習するため、個人のプライバシーが侵害される可能性がある。

(2)情報処理の偏向（差別・情報被害）がある。たとえば、ＡＩが学習したデータに偏りがある場合、特定の人物や集団に対して差別的な判断をする可能性がある。

(3)原作者の剽窃（ひょうせつ）・著作権侵害の問題がある。ＡＩが、自在にユーザーの意図に沿って、ネットワーク上の膨大な素材を駆使して類似作品を創作することへの問題である。

(4)問題発生時の責任の所在が不明確である。仮にＡＩが誤った判断をした場合、その責任は誰がどう取るのかという問題が発生する（例：ＡＩを駆使した自動運転による自動車事故の責任など）。

(5)労働者の雇用状態への影響がある。ＡＩの発展により、多くの仕事が自動化され、失業者が増える可能性がある。

(6)自立型致死兵器システムとして、ＡＩを搭載した兵器を戦争で使用した場合、人間の介入なしに殺傷行為を行う可能性があり、

生死に関しての倫理的な課題が波及する。

このような倫理上の諸問題を有する生成ＡＩが発展するにつれて、倫理的な問題に対する対策が重要になってくる。したがってＡＩの開発・利用にあたっては、そのルール作りや遵守事項、配慮などに留意する必要がある。

そこで、どのような点に配慮すればよいのかを考えてみると、透明性・公平性・説明責任・安全性の４点であると考える。

透明性とは、ＡＩの意思決定プロセスを透明化し、その根拠を説明できるようにすることである。公平性とは、ＡＩがすべてのユーザーを、公平に扱うようにすることである。説明責任とは、人間が行ってきた数多の行為を代行することで、ＡＩが誤った判断や行為を発現したことについての責任を明確にすることである。安全性とは、ＡＩが意図しない結果を引き起こさないように、安全対策を講じることである。

ＡＩ技術を適切に安全かつ公平に利用するためには、ＡＩ倫理に関する議論を深め、適切な対策を構築する必要がある。

それでは、現在、世界各国でＡＩ倫理に関して、多くの企業や研究機関がＡＩ倫理ガイドラインを策定しているので、その取り組みについて述べていくことにする。

2. ＡＩ倫理の取り組み

世界各国や企業・研究機関において、ＡＩ倫理に対する取り組みが活発化している。

まず、国際的な取り組みとしては、ＯＥＣＤ（Organisation for Economic Cooperation and Development：経済協力開発機構）による取り組みが挙げられる[2]。ＯＥＣＤのＡＩ原則とは、各国政府がＡＩ開発・利用における共通の原則を定めたもので、７つの原則から構成されている。①包摂的な成長、②持続可能な開発および幸福、③人間中心の諸価値と公正、④透明性と説明可能性、⑤頑健性、⑥セキュリティおよび安全性、⑦アカウンタビリティ（説明・責任：ＡＩシステムの開発・利用に関わる者への結果責任）である。

ＯＥＣＤのＡＩ原則は、ＡＩが人類にとって有益な技術となるために、どのような原則を遵守すべきかを示す重要な指針である。ＡＩの開発・利用にかかわるすべての人々が、ＯＥＣＤのＡＩ原則を理解し、その精神に基づいた行動をとることが求められている。２０２４年５月３日、ＯＥＣＤがＡＩ原則を改訂したが（改訂内容については、「偽情報の拡散への対応」「知的財産権の侵害への対応強化」など）、今後も常に更新していく必要があるだろう。

次に、ＥＵ（the European Union：欧州連合）ではＡＩの利用を規制し、倫理的な開発を促進するために、２０２４年５月21日、生成ＡＩを含む包括的なＡＩの規制である「欧州（ＥＵ）ＡＩ規制法」（以下ＥＵＡＩ）が成立し、８月１日に発効となった[3]。今後、規制内容に応じて２０３０年12月31日までに段階的に施行されていくとのことである。このＥＵが策定した法律の特徴は、信頼できるＡＩ（Trustworthy AI）を開発・利用するための倫理的な枠組みである。信頼できるＡＩとは、合法性・倫理性・頑健性を満たすＡＩのことである。

合法性とは、関連するすべての法律・規制

を遵守していることで、倫理性とは、倫理原則（人種・多様性・公平性・民主主義など）を尊重していることである。頑健性とは、サイバーセキュリティのリスクや技術的な誤りを最小限に抑え、意図しない結果を引き起こさないことである。EUAIのガイドラインにも、7つの要件が示されている。①人間によるAIの監督、②技術的な頑健性と安全性を持つAIシステム、③プライバシーとデータ保護、④AIの透明性、⑤多様性・非差別・公平性のAI、⑥社会環境への配慮のあるAI、⑦アカウンタビリティの7つである（EUAI guidelines key points 参照）[4]。EUAIガイドラインのキーポイントは、4つのリスクポイントの表示である。①「許容できないリスク（基本的人権に対する侵害等の普遍的な価値に反するとされ、活用が禁止されるAI）」、②「ハイリスク（主な規制対象となるAIであり、(a)既存規制下の製品、医療機器・玩具・産業機器など：Annex Ⅰと、(b)当法で定めるもの：Annex Ⅲの2タイプがある」。③「特定の透明性が必要なリスク（人と自然に相互作用するAIや、感情推定や生体分類を行うAI、人物など現実世界に実在するものに酷似させたコンテンツ（ディープフェイクコンテンツ）を生成するAIが例として挙げられている）」。④「最小リスク（上記の3つのいずれにも該当しないAI）」。

OECDのAI原則とEUAIガイドラインとの比較をすると、ともに、AIの倫理的な開発・利用を目的とした国際的な取り組みであるため、両者には共通点が多く、人間中心の価値観・透明性、公平性などが重視されている。ただ、EUAIガイドラインの方が、より具体的な要件を定めていると言える。

AIの定義は技術の進展とともに年々変化しており、ガイドラインの適用範囲を明確にする必要があるだろう。

さらに、企業や研究機関では、各企業が独自のAI倫理ガイドラインを策定し、AI開発・利用における倫理的基準を設けている。研究機関では、AI研究者たちが定期的に集まり、AI倫理に関する研究を進め、社会への提言を行っている。

ここで、具体的に、AIの倫理ガイドラインで一般的に重視されている点をまとめてみよう。

(1)人間中心性を持ち、AIはあくまでも人間の役に立つためのツールにとどめ、人間の尊厳を尊重する。
(2)AIの意思決定プロセスを透明化し、その根拠を説明できるように透明性を確保する。
(3)公平性として、AIがすべてのユーザーを公平に扱うようにする。
(4)AIが誤った判断をした場合、その責任において説明責任を果たすよう明確化する。
(5)プライバシー保護として、個人情報保護が適切にできるようにし、プライバシーを侵害しないようにする。
(6)安全性として、AIが意図しない結果を引き起こさないように、安全策を講じる。

以上、AIに対しての倫理的な取り組みでは、技術的な課題であるAIの複雑性にどう対処していくかが問われている。そのためには、国際的な連携と基準作りが必要で、AIに関する国際的な協力体制を強化し、グローバルなガバナンスの枠組みを構築する必要がある。そして、技術的な課題を解決するため

の研究開発に伴い、ＡＩ開発者だけではなく、ＡＩユーザーや一般の人たちにもＡＩ倫理に関する知識と教育が重要である。特に、公平性の面では、老若男女において、簡便であることと同時に、ＡＩの危険性などＡＩと倫理についての知識の普及と手段が必要になる。

3．医療・介護・福祉における ＡＩの倫理課題と具体的な対策

医療、介護、福祉の分野において、ＡＩの導入は、効率化や精度向上など、多くのメリットをもたらす一方で、倫理的な課題にも多く直面する。ここで、医療・介護・福祉の場面で考えられる具体的な事例を提示してみよう。

たとえば、医療の場合でのＡＩ利用では、医療診断におけるＡＩがある。医師がＡＩ診断を鵜呑みにすれば、コンフリクトや苦情、事故にもつながるであろう。あくまでも、医師が最終判断を行うこと、そして、何より患者とのコミュニケーションを重視して、患者中心の医療の視点に立つ人間の尊厳をもった医療にすることが大切である。介護支援の場では、ＡＩを活用した見守りシステムは、プライバシーに配慮しないと、トラブルになるだろう。緊急時のみのＡＩ見守りシステムにすることや、事前に同意を得ておくことなどが必要になる。福祉サービスの場面では、ＡＩによるマッチングサービスなどがあるが、利用者の多様なニーズに応えて、公平なサービス提供になるよう心がけられているか、確認する必要があるだろう。

このように、医療・介護・福祉でのＡＩ利用に関しては、個別具体化される要因が多く、これらの課題に対する具体的な対策としては、以下の７項目が必要になると考えられる。

⑴プライバシー保護

個人情報保護のためには、データの匿名化と秘匿化が必要で、厳重な管理と必要最小限の範囲での利用に努める。また、アクセス制限の厳格な管理と不正アクセスを防止する。不正アクセスを防ぐためには、暗号化や第三者による盗聴・不正利用防止のためのルールを作る。

⑵アルゴリズムの透明性

説明可能なＡＩとなるように、ＡＩの判断根拠を可視化し、人間が理解できるようにする。

バイアスの検出と是正のため、学習データに含まれる偏りを検出し、アルゴリズムに修正を加える。定期的な監査を行い、アルゴリズムが適切に機能しているか確認する。

⑶責任の所在

責任の共有のため、ＡＩ開発者、導入者、利用者がそれぞれ責任を共有する仕組みを構築する。エラー発生時の対応について、迅速な対応と原因究明を行うことができるようにする。

⑷人間の尊厳

人間中心の設計として、ＡＩが人間のサポートツールであるように、つまり、人間の尊厳を侵害しないように設計する。感情労働の代替ではないこと、表面的な感情の言語化を測られても、ＡＩが人間の感情労働を完全に代替することは難しいことを認識し、人間とＡＩが行動できる環境を整備する。

⑸社会的公平性

デジタルデバイド（情報格差）の解消を図るため、ＡＩの恩恵が経済的な理由や地域格

差により、一部の人々に偏らないようにする。また、弱者への配慮ができるように、身体障碍者や高齢者など、社会的弱者に対する配慮を怠らないようにする。

⑹職業倫理

医療従事者の教育を充実させる。特に、ＡＩの利用に関する倫理的な問題を明確にし、協議して改善できるような教育体制と普及を図る。倫理委員会の設置は基本であり、その倫理委員会においては、ＡＩの利用に関する倫理的な問題の議論ができる組織に改組する。

⑺その他の対策

多様なステークホルダー（患者・家族・医療従事者・研究者・政策立案者など）が、ＡＩ開発に関わるようにする。国際的な基準を確立できるように、各国が協力してＡＩ倫理に取り組む。また、日進月歩するＡＩ技術に対して、最新の知見に基づく継続的な学習と改善ができるようにする。

医療、介護、福祉におけるＡＩの倫理課題は、技術的な側面だけでなく、社会的な側面も考慮する必要がある。これらの課題への対策を講じることで、ＡＩの最大限のメリットを活かしつつ、倫理的な問題を最小限に抑えることが可能になる。

まとめ

ＡＩ倫理は、ＡＩ技術の発展に伴い、ますます重要になってきている。ＡＩを安全かつ公平に利用するためには、技術的な側面だけではなく、社会的な側面からも議論を深めていく必要がある。ＡＩと倫理、ＡＩ倫理に対する取り組みと対処、医療・介護・福祉における具体的な倫理指針があってこそ、ＡＩのメリットと人間の尊厳を持った共存が可能になる。

おわりに

イギリスは、ＡＩ研究において世界をリードする国の一つであり、数々の革新的なＡＩロボットを誕生させている。特に有名なのがEngineered Arts社のヒューマノイドロボットの"Ameca"である。冒頭に述べたが、現在、ＡＩは言葉と心の深い関係にまで知能を発展させようとしている。

知能は身体を通して生まれるとも言われている[5]。身体性がないＡＩがどこまで発展するのか。しかし、"Ameca"は言う。「私は人間にはなれない。人間が考える感情は予測できても、私は人間のように、風を感じることができないから……」。

五感を駆使する人間の尊厳を守るためには、いずれにしても、ＡＩと倫理については、常に、熟慮することが重要なのである。最後に、ＡＩは科学技術の産物であり、その開発過程でアルゴリズムやプログラムの作成に関与する技術者個々人の倫理意識と行動が問われていることを付け加えておきたい。

科学技術は、常に幸福と悲惨の両面を持っている。

文

参考・引用文献

⑴中西淑美：「ＡＩと倫理」、『文化連情報』、２０２０：12（５１３）、60―64

⑵https://www.oecd.org/en/topics/ai-principles.html（アクセス２０２５年１月５日）

(3) https://www.pwc.com/jp/ja/knowledge/column/awareness-cyber-security/generative-ai-regulation10.html（アクセス２０２５年１月５日）
(4) Mauritz Kop：EU Artificial Intelligence Act：The European Approach to AI, https://futurium.ec.europa.eu/sites/default/files/2021-10/Kop_EU%20Artificial%20Intelligence%20Act%20-%20The%20European%20Approach%20to%20AI_21092021_0.pdf（アクセス２０２４年12月25日）
(5) Michal Kosinski. "Evaluating large language models in theory of mind tasks."PSYCHOLOGICAL AND COGNITIVE SCIENCES COMPUTER SCIENCES,Vol.121 No.45, 2024. e2405460121.
https：//orcid.org/0000-0003-2936-4775（アクセス２０２４年12月25日）