

ナフィールド生命倫理評議会『医療と研究における人工知性 (AI)』

(Artificial intelligence (AI) in healthcare and research) ブリーフィングノートの紹介

京都大学大学院文学研究科・研究員

大庭弘継

1 はじめに

『医療と研究における人工知能 (AI)』 (Artificial intelligence (AI) in healthcare and research) は、英国のナフィールド生命倫理評議会 (The Nuffield Council on Bioethics) が、「医療における AI の現在および将来の可能性、その限界、およびその使用に起因する倫理的問題について検討」¹したブリーフィングノート²であり、2018年5月に刊行された。このブリーフィングノートは <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> からダウンロードできる。なおナフィールド生命倫理評議会とは、ナフィールド財団によって設立された「生物学と医学における倫理的問題の調査研究を行う独立組織」である。

本報告の内容は以下のとおりである。

概要 (Overview)

AI とは何か? (What is AI?)

AI への近年の関心 (Recent interest in AI)

BOX: AI に関する調査とイニシアチブの事例 (BOX 1. Examples of Inquiries and Initiatives on AI)

医療と研究における AI の応用 (Applications of AI in healthcare and research)

AI の限界 (Limits of AI)

倫理的・社会的課題 (Ethical and social issues)

ガバナンスへの課題 (Challenges for governance)

AI の将来 (the Future of AI)

結論 (Conclusions)

ブリーフィングノート自体は註を含めて 8 頁と短い。本紹介は、各項目の重点を抄訳するとともに、「概要」、「倫理的・社会的課題」、「AI の将来」、「結論」については、すべて訳出した。なお、紹介者が最も重要と考える「倫理的・社会的課題」については、次節の「概要」でコンパクトにまとめているので、参照いただきたい。

訳語についてだが、healthcare については「医療」を、health については「健康」と「保健」を文脈で使い分け、intelligence については AI の場合は「知能」を、人間の場合は「知性」を、それぞれ用いている。本紹介の脚注については、ブリーフィングノートにリンクされている各種文書に加えて、一部、要約者が追加した情報もある。

なお本紹介作成に際し、児玉聡氏（京都大学大学院文学研究科・准教授）と高木裕貴氏（京都大学大学院文学研究科・博士後期課程）に貴重なコメントとアドバイスをいただいた。記して感謝したい。

2 概要（全訳）

AI の使用には、病気の発見、慢性疾患の管理、医療サービスの提供、創薬が含まれるなど、医療や研究目的の広い範囲で、その使用や試験が行われている。

AI は重要な健康上の課題の治療に役立つ可能性を持つが、利用できる健康データの質や、AI が人間のもつ諸特徴を体現できないことによって、AI は限界を抱えることになる。

AI の使用は、以下のような倫理的課題を生じさせる。

- ・ AI が誤った意思決定をする可能性。
- ・ AI が意思決定支援を行った場合に、誰が責任を負うかという問題。
- ・ AI システムによるアウトプットの検証が困難。
- ・ AI システムの訓練に用いるデータに内在するバイアス。
- ・ 潜在的にセンシティブなデータ保護の確保（ensuring）。
- ・ AI 技術の開発と使用についての国民の信頼の確保（securing）。
- ・ 介護の状況における人々の尊厳の感覚と社会的孤立に対する影響。
- ・ 医療専門職の役割と要求される能力に対する影響。
- ・ 悪意ある目的に AI が使用される潜在性。

主要な課題（key challenge）は、当該領域においてイノベーションを刺激し推進しながら、透明性と公共の利益と両立する方法で、AI の開発と使用を担保していくことである。

3 AI とは何か（抄訳）

AI は、万人が同意する普遍的な定義をもたないが、推論・学習・適応、感覚的理解（sensory understanding）と相互作用といったような人間の知性を連想させるプロセスと類似したコンピューターテクノロジーを一般的には指すと言える。機械学習（machine learning）³は、近年、最も成功している AI の形態であり、現在、応用が多くなされている特筆すべきアプローチである。

4 AI への近年の関心（抄訳）

多額の公的・私的投資および多大な関心のもと、AI の幅広い応用が探究されている。論議を呼びながらも、政府と企業、大学と企業とのあいだに、協働（collaboration）やパートナーシップ（partnership）が形成されつつある。

5 BOX：AI に関する調査とイニシアチブの例（抄訳）

- ・ 英国政府の「データ倫理とイノベーション・センター（CDEI）」：2018年1月、（センターの任務は）データドリブン（data-driven）技術の安全で倫理的で革新的な利用についての助言を行う、とし

た⁴。

・「エイダ・ラブレス研究所 (Ada Lovelace Institute)」⁵： ナフィールド財団は、2018 年末までに、データ・アルゴリズム・AI の利用に由来する、倫理的・社会的課題を検証し、社会的福利に活用することを確かにするため、この研究所の設立を発表した。

・「パートナーシップ・オン・AI (Partnership on AI)」： Amazon、Apple、DeepMind、Facebook、Google、IBM、Microsoft が設立した、AI に関わる議論とコミットメントのためのプラットフォーム。

・「IEEE 自律知能システムの倫理に関するグローバル・イニシアチブ (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems)」⁶： 2016 年に立ち上げ。

・国連地域間犯罪司法研究所 (United Nations Interregional Crime and Justice Research Institute)： 2015 年に人工知能とロボットに関するプログラム⁷を立ち上げ。

・「アシロマ・AI 原則 (Asilomar AI Principles)」⁸： 2017 年に Future of Life Institute (米国) が、AI の研究と応用の指針として開発し、世界中の AI やロボット工学に携わる 3,800 人以上の研究者が署名した。

・AI に関する報告は、英国上院の人工知能に関する特別委員会 (House of Lords Select Committee on Artificial Intelligence)⁹、王立協会 (the Royal Society)¹⁰、リフォーム¹¹、フューチャー・アドボカシーとウェルカム・トラスト (Future Advocacy and Wellcome)¹²、ネスタ (Nesta)¹³、「科学と新興技術の倫理に関する欧州グループ (EGE)」¹⁴などが刊行しており、また英国下院の科学技術特別委員会からの報告¹⁵が予定されている。

6 医療と研究における AI の応用 (Applications of AI in healthcare and research) (抄訳)

・医療機関

効率的なケアプランの作成に使用。IBM のワトソン・ケア・マネージャー (Watson Care Manager)¹⁶など。また、患者と AI が交流してケアを目指すコグニティブ・ホスピタル (Cognitive Hospital) も試験運用中¹⁷。

・医学研究

AI を利用した創薬、創薬に取り組むべき分野の予測、AI を使用した「ロボット科学者」であるイブ (Eve) の活用などがある¹⁸。

・臨床ケアでの活用

医療用画像をスキャンし疾患を検出する医療用イメージング (Medical imaging)。エコー画像から心疾患を AI¹⁹が診断する心エコー検査 (Echocardiography)。スピーチパターンを分析してパーキンソン病などの神経疾患を同定する、神経疾患のスクリーニング (Screening for neurological conditions)。AI 制御のロボットツールで手術後の縫合など特定の作業を行う、外科手術 (Surgery)。

・患者・消費者向けアプリケーションがいくつも市場化され、健康評価と在宅ケアアドバイスを提供。たとえば、IBM が開発した「関節炎仮想アシスタント (the Arthritis Virtual Assistant)」など。

・公衆衛生において、AI は、感染症のアウトブレイクなどの早期の検知に使用できる可能性がある。

7 AIの限界 (Limits of AI) (抄訳)

AIは、データの利用可能性と質 (the availability and quality of data) に制限される。また多大かつ複雑なデータセットを扱うため、高度な計算能力が必要である。そもそも医療記録がデジタル化されていないことや、記録の相互運用性や標準化の欠如も問題となる。臨床の現場では、複雑な判断や能力が必要とされるが、現状のAIは、思いやり (compassion) などの人間的諸特徴に欠け、文脈や社会的手がかり (cue) を読み取れず真似できない。

8 倫理的・社会的課題 (Ethical and social issues) (全訳)

AIが提起する倫理的・社会的課題の多くは、データ利用の問題、オートメーション (自動化)、広い意味での技術への依存 (reliance on technologies)、支援技術や遠隔医療 (テレヘルス) の利用で生じる課題とオーバーラップしている。

(1) 信頼性と安全性 (Reliability and safety)

信頼性と安全性は、機器の制御、治療の提供、医療における意思決定といったAIが利用される場面で、主要な課題である。AIがエラーを犯す可能性があり、そして、もしそのエラーの発見が困難か、波及効果を持つ場合、深刻な影響を持つことになる。たとえば、2015年のある臨床試験²⁰では、肺炎の後に、どの患者が合併症を併発するか、そして入院させるべきかを予測する、AIアプリを利用した。このアプリは、文脈的情報を考慮に入れる能力に欠けたため、ぜんそくのある患者を帰宅させるよう、医師に誤った指示を出した。

AIを利用した症状チェックのアプリの性能は、疑問視されている。たとえば、アプリによる推奨事項が過剰に警告的で、不要な検査や治療を増加させる可能性があることがわかっている。

(2) 透明性と説明責任 (Transparency and accountability)

AIが生成するアウトプットの根底にあるロジックを見出すことは、困難もしくは不可能である。多くのAIは、特許で守られており慎重に秘匿されているが、それ以外のAIであっても、人間が理解するには単純に複雑すぎるのだ。機械学習の技術は、学習するにつれて自身のパラメーターや規則を継続的に調節していくため、特に不透明である。このことは、AIシステムのアウトプットの検証、データにおけるエラーやバイアスの特定といった問題を生み出す。

EUの新しい一般データ保護規則 (GDPR)²¹は、データ主体は、法的もしくは同義の効果をもたらす自動処理のみに基づく決定の対象とはならない権利を有する、としている。さらに、自分自身に関するデータが使用される際、個々人に提供される情報には、「自動化判断 (automated decision-making) の有無、(…) [処理に利用する] ロジックに関する有意な情報や、データ主体に対する処理の意義や想定上の結果」²²が含まれるべきであるとする。しかしながら、この規制の射程と内実—例えば、AIが理解できるか、どう理解するか、どのように英国で適用されるか—は不明確であり論争となっている。

関連する問題には、AIによってなされた決定をだれが説明するのか、AIによって損害を被った人はどう補償を求めることができるのか、といったものも含まれる。

(3) データのバイアス、公平性、公正性 (Data bias, fairness, and equity)

AIアプリは、人間のバイアスやエラーを減少させる可能性もある一方で、AIを訓練するために用いるデータにあるバイアスを反映し強化することもありうる。ジェンダー、エスニシティ、障害、年齢といっ

法的に保護されている特徴とは異なる、隠れた特徴に基づく差別につながる可能性、への懸念が生じている。上院の人工知能に関する特別委員会は、AI システムの訓練に用いられるデータセットが広範な人々を代表するものではないことが多く、結果として、社会に含まれる大きなバイアスを反映した不公平な決定を下しかねないこと、を警告した。同委員会はまた、AI 開発者の信念や偏見が反映した、バイアスがアルゴリズム自体に埋め込まれる可能性も見出した。一部の識者（several commentators）は、この問題に取り組むために、開発者の多様性を増加させることを要求した。

医療における AI の恩恵は、平等に分配されないかもしれない。AI は、データが不足している場合や、デジタル化してのデータの収集や提供が難しい場合、うまく機能しないだろう。これは、病気が稀な人々や、黒人やアジア人のマイノリティの人々など、臨床試験や研究データのなかで少数しか扱われていない人々などに影響を与えることになるだろう。

（4）信頼（Trust）

（AI 開発企業の）DeepMind²³とロンドンにあるロイヤルフリー病院の協働は、営利企業に患者データへのアクセスを与えることについての国民的な議論を引き起こした。識者（commentators）は、技術が公共の利益のために開発されているということを人々が信頼することができなければ、AI に対する国民的な反発につながると警告している。

医療において、AI が成功裏に実装されるためには、実務レベルで、患者と医療専門職がともに、AI システムを信頼できるようにすることが必要である。IBM のワトソン（Watson）による腫瘍学の臨床試験では、がんの診断に AI ツールが使用されたが、米国外の医師は AI の勧告に自信を持たず、がん治療に対する米国特有のアプローチが反映されたモデルだと感じており、一部の診療所で中止されたと報告されている。

（5）患者への影響（Effects on patients）

AI 健康アプリは、可能な時に、自身の症状を評価し自分をケアできるよう、人々をエンパワーする可能性を持つ。慢性的な健康問題や障害を持つ人々をサポートすることを目指す AI システムは、人々の尊厳の感覚、独立性、QOL を高めることができる。そして、介護施設に入居せざるをえない人々も、長期間、自宅にとどまることを可能にする。しかしながら、もし AI 技術が、患者が医療スタッフや家族と（過ごす）時間にとって代わるようになるならば、人間とのつながりの喪失と、社会的孤立の増加につながる懸念が高まることになる。

AI システムは、個人の自律性に悪影響をもつこともありうる。たとえば、リスクや、何がユーザーにとっての最大の利益かの計算に基づいて、AI は選択肢を制限するかもしれない。もし AI システムが、診断作成や治療計画の策定に使用されるならば、医療専門職はどのようにそう結論付けたのか説明ができない。これは、自身の健康についての十分な情報を得たうえでの意思決定という、患者の自由の権利を抑制するものとみなすことができる。人間の同僚や介護者を模倣することを目指すアプリは、本当の人間とテクノロジーのどちらとコミュニケーションしているのか、利用者が判断できなくなる可能性がある。これは欺瞞や詐欺の一形態として経験されるだろう。

（6）医療専門職への影響（Effects on healthcare professionals）

医療専門職は、その専門性が AI の挑戦を受けたとき、自律性と権威が脅かされると感じるかもしれない。個々の患者に対する医療専門職の倫理的義務は、AI の意思決定支援の使用によって影響を受け、費

用効率や公衆衛生への幅広い関心などの、その他の優先順位や利益に左右されることになるかもしれない。多くの新技術と同様に、AI の導入は、医療専門職に必要とされる能力と専門性を変えてしまうことを意味しそうである。

多くの領域で、これまで人間によってなされてきた仕事の自動化を、AI は可能にするだろう。これは、医療専門職が患者に、より直に接する時間をもたらすことになる。しかしながら、AI システムの導入が、能力に欠けるスタッフの雇用を正当化するために使用されるとの懸念もある。もし技術が失敗し、スタッフがエラーを認識できず、コンピューターの手助けなしに必要な仕事を実行することができない場合、問題となるだろう。関連する懸念は、AI が医療専門職を無気力にし、結果のチェックやエラーの探知をしなくなる、というものもある。

(7) データのプライバシーと保全 (Data privacy and security)

医療における AI アプリは、多くの人々がセンシティブでプライベートだと考えるデータを使用する。これらは法の規制の対象である。しかしながら、ソーシャル・メディアでの活動履歴やインターネットの検索履歴といった、健康状態に明らかに該当しない他の種類のデータが、ユーザーとその周囲の健康状態についての情報を暴露することに使用することもできる。ナフィールド生命倫理評議会は、プライバシーへの懸念を引き起こすデータ使用についての取り組みが、データがどう使用されるかについての人々の期待を考慮に入れたうえで、(単なる) コンプライアンスの先を目指すべきだと、提案してきた。

AI は、サイバー攻撃を探知し、医療用コンピューターを防護することにも使用されるだろう。しかしながら、センシティブなデータにアクセスするためハッキングされ、また容易には探知できないよう、欺瞞や偏見のあるデータをまき散らされる潜在性がある。

(8) AI の悪意ある使用 (Malicious use of AI)

AI は善い目的のために使用される可能性がある一方で、悪意ある目的のためにも使用できる。たとえば、AI がひそかにサーベイランスやスクリーニングに使用されるという恐れがある。(キーボードのタイピングのような) 運動行動の分析、スマートフォンの追跡によって検知された行動パターンの分析といった AI 技術は、本人に知られることなく、その人の健康についての情報を明らかにすることができる。AI は、金銭的成本をかけずに、より大規模なサイバー攻撃を実行することにも使用できる。このことにより、政府、研究者、エンジニアたちは、AI のデュアルユース性を熟考すること、AI 技術の潜在的な悪用に対して備えることを要請するに至っている。

9 ガバナンスへの課題 (Challenges for Governance) (抄訳)

AI に適用される規制は多岐にわたり、また進展も急速であるため、確立した規制を脅かすことにもなる。また AI という特定の分野として規制するべきか、潜在的に AI の影響を受けうる様々な分野の規制を見直すべきかという問題もある。

更なる課題として、開発され使用される AI が、透明性を保ち、説明責任を担い、公共の利益 (public interest) と両立すること、また、英国のイノベーションを推進しようとする欲求とバランスを取る方法、これらを確実にする必要がある。さらに研究者、医療従事者、政策立案者は、AI を評価し活用する能力と知識を備える必要がある。

10 AIの将来 (The future of AI) (全訳)

将来的には、AIシステムはさらに進歩して、人間の制御やインプットなしに、幅広い範囲の業務(task)を遂行する能力を獲得する可能性がある。もしこうなった場合、AIシステムは「倫理的であること ('be ethical')」や倫理的な意思決定を行うことを学ぶ必要がある、と提案する人もいる。これは哲学的議論の対象であり、倫理的価値や原則をマシン (a machine) にコーディングしたり学習させたりすることが可能かどうか、またその方法はどのようなものかといった問題や、決めるとすれば誰がその価値観を決定すべきか、人間に適用される義務がマシンに適用できるかどうか、あるいは新たな倫理原則が必要かどうかなどの問題を惹起している。

11 結論 (Conclusions) (全訳)

AI技術は、疾病の検知、慢性疾患の管理、保健サービスの提供、創薬など、医療と研究分野における幅広い目的のために使用・試用されている。AI技術は、重要な健康課題への対処の手助けとなる可能性があるが、利用しうる健康データの質や、思いやり (compassion) のような、人間的特性を持たないことによって、制限される可能性がある。AIの利用は、多くの倫理的・社会的課題を生じさせるが、その多くは、データおよび医療技術全般の利用によって惹起された問題と、広く重なっている。AI技術に対する将来のガバナンスにおける重要な課題は、この分野のイノベーションを刺激し促進しつつも、AIが透明性を持ち公共の利益に合致する形で開発され使用されることを、確かなものにするることである。

¹ <http://nuffieldbioethics.org/project/briefing-notes/artificial-intelligence-ai-healthcare-research>

² 2016年までナフィールド生命倫理評議会は、バックグラウンドペーパーを刊行してきた。これに代えて導入されたのがブリーフィングノートであり、2018年度から始まった「アクティブ・レスポンス・プログラム ('active response' programme)」の一環として、ディスカッション・ワークショップとブログとともに挙げられている。ブリーフィングノートは迅速な発刊を重視し、「これらは我々の主要プロジェクトのような深さはなく、結論や政策提言を出さないが、国民に問題を周知し公共の議論を刺激する能力の向上をもたらす」とされている (<http://nuffieldbioethics.org/blog/scanning-the-horizon-one-song-at-a-time#more-16589>)。本ブリーフィングノートは、この制度導入以後の3本目 (the third in a new series) になる (<http://nuffieldbioethics.org/future-work/background-papers-resources>)。

³ 近年の日本においては、機械学習よりも、深層学習とも訳されるディープラーニング (deep learning) を頻繁に耳にする。この二つはどう異なるのか。端的には、ディープラーニングは、機械学習の一種である、といえる。AIに大量のデータを読み込ませる点では、一致している。ただし機械学習があらかじめプログラムされた条件での能力を高めるのに対して (例えば、リンゴの種類の違いを見分けるために色に着目せよ、など指示する)、ディープラーニングは条件そのもの (リンゴの種類の違い) を自ら見つけ出すという点で大きな違いがあるとされる。ただし、本報告における機械学習の語は、ディープラーニングも包摂するかたちで使用されている。以下を参照した。

<https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>、 <https://innovation.mufig.jp/detail/id=93>

⁴ CDEI (The Centre for Data Ethics and Innovation)、2018年11月20日に任務改定。なお、情報倫理学で著名な、オクスフォード大のフロリディ氏 (Professor Luciano Floridi) もボードメンバーである。 <https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation>

-
- ⁵ エイダ・ラブレスは、詩人バイロンの娘で、階差機関 (difference engine) を考案したチャールズ・バベッジとの交友で知られ、史上初のプログラマーとも呼ばれる。なお、プログラミング言語の Ada は、エイダ・ラブレスにちなんで命名された。この研究所は、2018 年 5 月 15 日に目的・手法等を公開、9 月 3 日に Alan Turing Institute の CEO であった Sir Alan Wilson がボードのチェアに就任、12 月 4 日に 4 名のボードメンバーを公表。詳細は、<https://www.adalovelaceinstitute.org/>
- ⁶ 以下を参照。 <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>
- ⁷ センターを設置。以下を参照、http://www.unicri.it/in_focus/on/UNICRI_Centre_Artificial_Robotics
- ⁸ 1975 年に遺伝子組み換え技術の規制を議論したアシロマ会議にちなんだ名称。なおこの原則は邦訳もされている。以下を参照、<https://futureoflife.org/ai-principles-japanese/>
- ⁹ <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>
- ¹⁰ <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>
- ¹¹ 英国の公共サービスの改革のための独立シンクタンク。報告書は、以下を参照、<https://reform.uk/research/thinking-its-own-ai-nhs>
- ¹² フューチャー・アドボカシーは英国のシンクタンクであり、医療系の研究資金を提供する財団ウェルカム (・トラスト) の援助で、下記の報告書を刊行。<https://wellcome.ac.uk/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf>
- ¹³ ネスタは英国の財団であり、ネスタ・ヘルス研究所 (Nesta Health Lab) が刊行した報告書は以下を参照。https://media.nesta.org.uk/documents/doctor_robot_summary.pdf
- ¹⁴ EGE (The European Group on Ethics in Science and New Technologies) は、EU 欧州委員会への諮問機関であり、下記の報告を刊行。https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf
- ¹⁵ 本ブリーフィングノートの刊行とほぼ同時期である、2018 年 5 月 23 日に以下を刊行。<https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf>
- ¹⁶ <https://www.ibm.com/us-en/marketplace/watson-care-manager>
- ¹⁷ <https://www.hartree.stfc.ac.uk/Pages/Cognitive-hospital.aspx>
- ¹⁸ イブ (Eve) については、以下を参照。<https://www.cam.ac.uk/research/news/artificially-intelligent-robot-scientist-eve-could-boost-search-for-new-drugs>。なおイブが抗マラリア薬を発見との報道もある。<https://futurism.com/robot-scientist-helps-discover-new-ingredient-antimalarial-drug>
- ¹⁹ <http://www.ultromics.com/>
- ²⁰ 以下を参照。Caruana R, et al. (2015) Intelligible models for healthcare, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp1721-30. <http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf>
- ²¹ EU の GDPR の導入は、日本のメディアも取り上げるなど、日本企業も対応を迫られた。そのため JETRO は、『「EU 一般データ保護規則 (GDPR)」に関わる実務ハンドブック (入門編)』と『「EU 一般データ保護規則 (GDPR)」に関わる実務ハンドブック (実践編)』を作成し、下記リンクで公開している。https://www.jetro.go.jp/ext_images/_Reports/01/dcfcebc8265a8943/20160084.pdf、https://www.jetro.go.jp/ext_images/_Reports/01/76b450c94650862a/20170058.pdf
- ²² 上記『入門編』24 頁の抄訳を引用した。なお、データ主体とは、「個人データが関連する当該個人」(同左、3 頁)を意味する。
- ²³ DeepMind は英国の AI 開発企業。開発した AlphaGo は、2017 年に囲碁の世界勝者を撃破した。