

教育 AI に関する倫理的課題

—「教育的判断」を中心に—

張 建*

Ethical issues related to educational AI

—Centered on the concept of Pedagogical Judgment—

ZHANG Jian*

Abstract

With the rapid development of artificial intelligence, ethical research on its development and utilization has become popular. A number of proposals have been published by academia, business organizations and governments, advocating the importance of setting ethical guidelines for AI. On the other hand, it is difficult to say that the use of artificial intelligence (Educational AI) in school education is being tackled with full awareness of its peculiarities. This paper discusses the ethical issues of educational AI, focusing on the concept of "Pedagogical Judgment".

キーワード：教育 AI, 倫理, 教育的関係, 教育的判断

Keywords: Educational AI, Ethic, Educational relationship, Pedagogical judgment

1. はじめに

機械学習技術の飛躍的な発展を契機に、人工知能 (Artificial Intelligence) は第3のブームを迎えているといわれ、自動運転、医療など産業領域への応用はすさまじい速さで進行している。AI 技術の進展とその応用は、「広範な産業領域や社会インフラなどに大きな影響を与えている」¹とされている。

人工知能技術は、その発展とともに、教育への応用も本格化してきている。アメリカでは、KNEWTON や IBM 社などの AI 技術に基づく教育サービスが多くの教育機関に採用され、学校教育に強いインパクトを与えている。日本においても、教育関連 AI 技術の開発と、学校教育への積極的な利活用を推進する政府の方針が出され、AI 技術に

よる教育や学びの変革が提唱されている。

「教育用 AI が発達し普及していくことにより、AI が個人のスタディ・ログ (学習履歴・学習評価・学習到達度など) や健康状況等の情報を把握・分析し、一人一人に対応した学習計画や学習コンテンツを提示することや、スタディ・ログを蓄積していくことで、個人の特性や発達段階に応じた支援や、学習者と学習の場のマッチングをより高い精度で行うことなどが可能となる」²とされている。さらに内閣府統合イノベーション戦略推進会議は、「AI が広く浸透した社会において、教育環境は、一方的かつ均一的に教える教育の在り方から、個々人の持つ関心や力を活かす在り方へと変化すると考えられる。そのため、社会は、これまでの教育環境における成功体験に拘ることなく、常に最適な形へと柔軟に変化し続ける意識を全体として共有」³すべきと

* 理工学部共通教育群特任教授 Professor, Division of Liberal Arts, Natural Social and Health Sciences, School of Science and Engineering

の見解を示し、AIの学校教育への利活用の必要性説をうたえている。

しかし、学校教育におけるAIの利活用は、現在の学校教育のパラダイムを大きく転換させる可能性をはらんでいる。「AIは単なる教育方法や教育技術ではない。むしろ人間に求められる資質や能力とは何か、教育の役割とはなにか、どこまでをAIに委ねるのか、等に係る倫理的検討」⁴が重要な課題となってくる。つまり、教育におけるAIの利活用において最大の課題は、単に技術のことでなく、十分な倫理的検討が不可欠であるということである。

これまで、国際的にも多様なセクターにおいてAIの倫理的課題についての検討が行われ、多くのAI倫理の原則指針やガイドライン案が発表されてきた。本稿は、AI倫理ガイドラインなどを分析整理したうえで、AI技術の利活用に関して、教育における特有な倫理的課題を提示する。

2. 教育AIの開発

AIは教育に浸透し、「教育におけるAIの利活用」(AI in education)という段階から、「教育AI」(Educational AI)の段階へと進んでいる。前者はAIが学習者側のデータを分析し、分析結果を教育者側に提示する技術が中心であるが、後者はAI自体が直接に判断し、教育的行為を行う技術である。アメリカの学校教育では、すでに以下の4つの領域において、教育AIが活用されている。

チュートリアル:一般的に知的指導システムと呼ばれるAIプログラムは、生徒を対話に参加させ、質問に答え、フィードバックを提供する。

個人学習:ITSおよび適応チューターは、学習教材・ペース・順序および難易度を各生徒のニーズに合わせて調整する。AIはまた、例えば自閉症の子どもたちの表情を識別することで、特別なニーズを持つ生徒を支援することもできる。

テスト:コンピュータ適応評価は、生徒の回答の正確さに基づいて連続する質問の難易度を調整し、生徒の習熟度をより正確に識別する。

自動化タスク:AIは、出欠取り・課題採点・テスト問題の作成などの日常的なタスクを実行で

きる。⁵

さらに、音声認識、画像認識、言語翻訳等の分野で、AIは人間の能力を超える段階に達している。教育AIでは「子どもを教える」ことが現実になりつつある。

一方で、現段階では人工知能の教育への応用は、企業などの産業セクターが主体となり、人工知能の技術を商品として教育の市場に提供するというビジネスモデルが主流である。このような状況の中で、教育AIの開発と利活用に関しては、教育学からの視点が必要不可欠であり、その倫理的な側面の検討および倫理ガイドラインの策定が喫緊な課題であると考えられる。

3. AIと倫理

人工知能が人間社会を大きく変える可能性を持つようになるにつれて、人間と人工知能との関係は大きく注目されるようになり、AI倫理に関する研究が多く展開されている。「進化し続ける人工知能開発段階では、倫理という社会規範(Social Norm)に、開発方向の判断において依拠するところが多く、開発方向性を誤らないためにも、倫理という規範(Norm)の意義と限界点を明らかにする必要がある」。⁶各国政府や学術研究機関など各セクターによる、AIに関する倫理ガイドラインの研究と策定は急ピッチで推進されている。

多くのAI倫理の原則と指針が、幅広い分野の専門家の意見を反映して特定され整理されてきている。そのうち、もっとも代表的なAI倫理案が以下の3つである。

i. 『アシロマの原則』⁷ (米国 Future of Life Institute, 2017.2)

ii. 『信頼できるAIの倫理ガイドライン』⁸ (欧州委員会, 2019.3)

iii. 『AI利活用ガイドライン』⁹ (日本AIネットワーク社会推進会議, 2019.8)

ここでは、これらの原則、指針およびガイドラインを分析し、AI倫理の一般的な課題を整理する。3つのAI倫理の原則案から、AIの倫理的課題を以下の6つのカテゴリーに分類することができる。「安全性」「アカウントビリティ」「人間の尊

「人権・権利と自由」「プライバシー保護」「公平性」「人間による制御」である。

I. 安全性の原則

① リスクへの配慮

○利用者は、AI システム又は AI サービスの利活用により、アクチュエータ等を通じて、利用者等及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。(『AI 利活用ガイドライン』)

② 「耐障害性」

○人工知能システムは、運用寿命を通じて安全かつロバストであるべきで、適用可能かつ現実的な範囲で検証されるべきである。(『アシロマの原則』)

○AI システムには耐障害性と安全性が求められる。何か問題が発生した場合のフォールバック計画を確実にし、正確性・信頼性・再現性を確保し、安全である必要がある。これは、意図しない危害も最小化し、防止できることを保証する唯一の方法である。(『信頼できる AI の倫理ガイドライン』)

③ 「障害の透明性」

○人工知能システムが何らかの被害を生じさせた場合に、その理由を確認できるべきである。(『アシロマの原則』)

④ 「セキュリティ」

○利用者及びデータ提供者は、AI システム又は AI サービスのセキュリティに留意する。(『AI 利活用ガイドライン』)

II. アカウンタビリティの原則

○責任：高度な人工知能システムの設計者および構築者は、その利用、悪用、結果がもたらす道徳的影響に責任を負い、かつ、そうした影響の形成に関わるステークホルダーである。(『アシロマの原則』)

○説明責任：AI システムとその成果に対する責任と説明責任を確保するための仕組みを整備すべきである。アルゴリズム・データおよび設計プロセスの評価を可能にする監査機能は、特に重要なアプリケーションで重要な役割を果たす。さらに、十分な接近可能な補償が保障されるべきである。(『信頼できる AI の倫理ガイドライン』)

○透明性の原則：AI サービスプロバイダ及びビジネス利用者は、AI システム又は AI サービスの入

出力の検証可能性及び判断結果の説明可能性に留意する。(『信頼できる AI の倫理ガイドライン』)

○アカウンタビリティの原則：AI サービスプロバイダ及びビジネス利用者は、消費者的利用者及び間接利用者を含むステークホルダーに対しアカウンタビリティを果たすよう努める。(『AI 利活用ガイドライン』)

III. 人間の尊厳・権利と自由の原則

○人間の価値観：人工知能システムは、人間の尊厳・権利・自由、そして文化的多様性に適合するように設計され、運用されるべきである。(『アシロマの原則』)

○自由とプライバシー：個人のデータに対する人工知能の適用を通じて、個人が本来持つ、または持つはずの自由を不合理に侵害してはならない。(『アシロマの原則』)

○非破壊：高度な人工知能システムがもたらす制御の力は、既存の健全な社会の基盤となっている社会的および市民的プロセスを尊重した形での改善に資するべきであり、既存のプロセスを覆すものであってはならない。(『アシロマの原則』)

○尊厳・自律の原則：利用者は、AI システム又は AI サービスの利活用において、人間の尊厳と個人の自律を尊重する。(『AI 利活用ガイドライン』)

IV. プライバシーの保護の原則

○個人のプライバシー：人々は、人工知能システムが個人のデータを分析し利用して生み出したデータに対し、自らアクセスし、管理し、制御する権利を持つべきである。(『アシロマの原則』)

○プライバシーとデータガバナンス：プライバシーとデータ保護の完全な尊重を保証するだけでなく、データの品質と完全性を考慮し、データへの正当なアクセスを保証する適切なデータガバナンスメカニズムも保証しなければならない。(『信頼できる AI の倫理ガイドライン』)

V. 公平性の原則

○利益の共有：人工知能技術は、できる限り多くの人々に利益をもたらす、また力を与えるべきである。(『アシロマの原則』)

○繁栄の共有：人工知能によって作り出される経済的繁栄は、広く共有され、人類すべての利益となるべきである。(『アシロマの原則』)

○多様性・無差別・公平性：不公平なバイアスは、脆弱なグループの周縁化から偏見や差別に至るまで、複数の負の影響を与える可能性があるため、避けなければならない。多様性を育むためには、障害の有無にかかわらず、すべての人が AI システムにアクセスできるようにし、関連する利害関係者を生涯にわたって関与させる必要がある。(『信頼できる AI の倫理ガイドライン』)

○社会的・環境的幸福：人工知能システムは、将来の世代を含むすべての人間に恩恵をもたらすべきである。したがって、彼らが持続可能で環境にやさしいことを確実にしなければならない。また、他の生物を含む環境にも配慮し、社会的・反社会的影響にも十分配慮する必要がある。(『信頼できる AI の倫理ガイドライン』)

VI. 人間による制御の原則

○人間による制御：人間が実現しようとする目的の達成を人工知能システムに任せようとする場合は、その方法と、それ以前に、判断を委ねるか否かについての判断を人間が行うべきである。(『アシロマの原則』)

○人工知能軍拡競争：自律型致死兵器の軍拡競争は避けるべきである。(『アシロマの原則』)

○人間の機関と監視：人工知能システムは人間に権限を与え、十分な情報に基づいた意思決定を可能にし、基本的権利を促進するものでなければならない。同時に、適切な監視メカニズムを確保する必要がある。これは、ヒューマン・イン・ザ・ループ、ヒューマン・イン・コマンドの各アプローチを通じて実現できる。(『信頼できる AI の倫理ガイドライン』)

○司法の透明性：司法の場においては、意思決定における自律システムのいかなる関与についても、権限を持つ人間によって監査を可能としうる十分な説明を提供すべきである。(『アシロマの原則』)

○適正利用の原則：利用者は、人間と AI システム間、及び利用者間における適切な役割分担のもと、適正な範囲及び方法で AI システム又は AI サービスを利用するよう努める。(『AI 利活用ガイドライ

ン』)

○適正学習の原則：利用者及びデータ提供者は、AI システムの学習等に用いるデータの質に留意する。(『AI 利活用ガイドライン』)

以上の AI 倫理指針には、AI 以外の技術にも共通する部分が多く含まれている。安全性、プライバシーの保護、公平性などは、すべての技術や製品にも通用すると言える。一方、AI の技術的特徴から、「責任」「人間の尊厳・権利と自由」「人間による制御」の3つの領域が AI 特有の倫理問題であると考えられる。つまり、AI の自律性が人間と技術とのこれまでの関係性を根底から変えるという事象に対処することが、AI 倫理の本質であると考えられる。

「責任」については、たとえば自動運転のように、AI が独自の判断で車を運転することによって発生した事故に対しては、誰が責任を負うのか、あるいはどの程度責任を負うのかを明確にする必要が出てくる。

「人間の尊厳・権利と自由」に関しては、アルゴリズムの研究開発において、ビッグデータなどに由来するバイアスの問題のほかに、人間的価値の定義問題が存在する。また、「人間による制御」とは、人間が AI の「技術の限界」を設定することである。つまり、どこまで AI 技術の利活用を容認するのかである。

4. AI による「教育的判断」とその倫理的課題

教育 AI の開発とその利活用の範囲が広がっている。アフェクティブ・コンピューティング (Affective Computing) の分野では、AI などで「感情を計測・分析し、また、人の感情を喚起したり、感情に働きかけたりする」¹⁰などの技術が進んでいる。表情認識、音声認識などの AI 技術によって、児童生徒の感情を分析し、授業中の学習状況を判断する実証実験がすでに日本の学校でも行われている。また、AI 技術を活用した「バーチャル・ティーチング・アシスタント」は、日本の大学でも導入されている。さらに、クラス分けや就職指導などでも、AI が活用されている。

これらの教育 AI の開発によって、AI は学校教育のより深い領域に入り込んでいる。これまで人間の教師に独占されてきた「教育的判断」の領域への進出により、特に重要な課題が浮き彫りになってくる。AI の一般的倫理原則は、教育 AI にも指針を与える。「責任」「人間の尊厳・権利・自由」「人間による制御」の3つの原則は、特に検討されるべき課題である。また、学校教育とその制度の特性から、AI 利活用の倫理的問題点の特定も必要であると考えられる。

現在の学校制度は、「教育的関係」を基礎としている。つまり、「教育的実践は、人間を人間にまで作りあげること志向するという点において、ほかの社会的実践とは区別される。ここでは人間の行為が人間にかかわって問題となる。つまり、実践の科学的規定や技術的規定では十分ではないのである」¹¹ この教育的関係は、これまで人間に特有なものであると考えられていた。

教育学者の田中正浩は、「教育的関係」の特質を以下のように指摘する。

第一に、教育的関係のきっかけとして人間的諸特質について考えることは、一方でその関係を一对一の個人と個人の関係としてしまい、実際には両者の関係が文化的、社会的規定を受けていることを見逃すことになってしまう。

第二に、教育的関係を築くきっかけとして抽出された人間的諸特質は、抽象的で非合理的な性質を帯びているために、科学的に捉えることがきわめて困難である。

第三に、実際のところ曖昧で客観的に把握することが難しい教育的関係は、そのきっかけとしての人間的諸特質に関する経験的な研究を重ねたとしても、その成果を集積し、再現することは不可能に近いといえる。¹²

このような教育的関係の中で、学校教育の実践は「教育的瞬間」の連続であり、「教師は授業進行中に生起する出来事を把握し、判断し、その情報を活用しながら、授業を運営していく」¹³ 教師の教育実践には、常に意味解釈が含まれている。この教師による「意味解釈」とは、「教育的判断」である。学校教育における教育的判断は、それぞれの教師の「意味解釈」によって構成され、社会全体の「複

合体としての人間価値判断」であるとの性質を持っている。

これまでの教育的判断は、人と人との関係についての判断であった。しかし、人工知能が「正しい」と「間違い」の価値判断ができるようになるにつれて、教育的判断は、人と、人以外の「自律的主体」との関係において行われることとなる。強化学習技術によって自動的な「副目標設定」を実現することはすでに可能であり、AI はすべての子どものデータ分析から、それぞれの子どもの「最適な教育判断」を行うことができるとされている。

しかし、「AI の判断の価値基準（教育の目標）」自体は、データから抽出するものであり、AI のアルゴリズムによる「最善」の判断である。つまり、AI の「価値基準」は人間のデータを「純化」するものであり、目標達成の効率性が唯一の「基準」となる。教育的判断の領域への AI 進出によって、学校教育の「教育的関係」が根本から変えられ、「教育的判断」の単純化がもたらされる可能性は大いにあり得る。

「社会における人工知能の広範な利用は、組織内の人間の自律性を低下させ、創造的な感情の管理の在り方を置き換えることによって、組織をより脆弱にする可能性がある」¹⁴ と同時に、「人工知能によって促進される我々の社会の他のメンバーに関する個人情報へのアクセスの増加は、基本的に人間の感情的な経験を変える可能性があり、潜在的に個人の自律性において重大でおそらく急速な損失をもたらす」¹⁵ との指摘があるように、「効率」の高い教育 AI の普及は、人間の教育の深層を変えてしまう危険性がある。

“教育は人間を人間まで作り上げる”という命題は、教育 AI の開発および利活用の倫理的な課題を内包する。AI 技術のさらなる発展に伴い、「教育的判断」の問題は、教育 AI の中心的な倫理的課題となるだろう。

5. むすび

久木田¹⁶ は、「AI は近代以降の社会価値に大きな影響を与え、人間の自律性を減じ、人間同士の感情的なつながりを変化させる」など、人工知能特有の

倫理問題を指摘する。人間の個性の核である「心」は、AI という外部の介入者に操られる。

学校教育特有の「倫理問題」が「教育的判断」であるとすれば、はたして「AI 教師」に「教育的判断」を任せられるのかという問題は、教育 AI の開発をする前に、「人間による制御」という一般的倫理原則の範囲を超えて、最優先に検討すべき倫理課題であるといえよう。

これまでに多くの AI 倫理の原則指針が構築されてきた。なかでも『アシロマの原則』に見られる「既存の健全な人間社会のプロセスを覆すものであってはならない」との指針は、教育、とくに学校教育における教育 AI の開発に重要な視点を与えるものとする。

注

¹ 統合イノベーション戦略推進会議, 2019, 「AI 戦略 2019 ～人・産業・地域・政府全てに AI～」

² Society5.0 に向けた人材育成に係る大臣懇談会, 2018, 「Society5.0 に向けた人材育成 ～社会が変わる、学びが変わる～」

³ 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」

⁴ 斎藤里美, 2016, 「人工知能は教師の役割をどう変えるか—教師に求められる役割と倫理—」, 『音楽教育学』(46), p.33

⁵ The Congressional Research Service (CRS), 2018, Artificial Intelligence (AI) and Education

⁶ 伊博文, 2018, 「人工知能と倫理」, 愛知大学情報メディアセンター紀要『COM』Vol.28/No.1 第 43 号, p.13

⁷ <https://futureoflife.org/ai-principles-japanese/>

⁸ The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)

⁹ http://www.soumu.go.jp/main_content/000637097.pdf

¹⁰ 江間有沙 長倉克枝, 2018, 「倫理的に調和した設計」の論点整理——異分野・異業種によるワークショップからの示唆——, 『情報法制研究』(4), p.7

¹¹ 山口匡, 2014 「教育の概念と教育的判断力: ハイトガーのコメント解釈に依拠して」『愛知教育大学教育創造開発機構紀要』(4), p.77

¹² 田中正浩, 2010, 「教育者と子どもの教育的関係について考える—「教育愛」を手がかりに—」, 『日本教材文化研究財団研究紀要』(40) 58～61 頁

¹³ 鹿毛雅治, 2012, 「教師による「教育的瞬間」の把握・判断に関わる心理的メカニズム」, 『科学研究費補助金研究成果報告書 (課題番号: 20330138)』

¹⁴ IEEE Ethically Aligned Design version2, Workshop series (<https://sites.google.com/view/ethically-aligned-design-2-workshop>)

¹⁵ Ibid.

¹⁶ 久木田水生, 2017, 「人工知能の倫理: 何が問題なのか」, AI ネットワーク社会推進会議資料