JSiSE Research Report
vol.33,no.7(2019-3)

教育システムと「倫理的に配慮されたデザイン」

武田俊之、 高等教育推進センター
関西学院大学

Issues for Educational Technologies in IEEE Ethically Aligned Design

Toshiyuki Takeda
Kwansei Gakuin University

IEEE Ethically Aligned Design(「倫理的に配慮されたデザイン」)は、知的で自律的なシステム（人工知能）が人間や社会の価値観や倫理原則と適合するよう、潜在的な有害性と望ましさに関する対話、議論、ポリシーを促進するために作成された報告書である。この報告では教育システムや教育データに関する「倫理的に配慮されたデザイン」の論点について整理、検討をおこなう。

キーワード: 人工知能、 自律知能システム、 Ethics by Design、 パーソナルデータ

1. はじめに

ビッグデータと人工知能の進展の社会への影響に関する議論が、アシロマ会議や人工知能学会など組織横断的な会議体においておこなわれている(1)(2)。

そのうち、The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems（The IEEE Global Initiative）は、数百人の学術研究、産業、市民、行政などの領域の関係者をグローバルに集めて、IEEE Ethically Aligned Design(EAD、倫理的に配慮されたデザイン)を作成した(3)。これは人工知能に関連する技術のうちロボット、IoT、エージェントのような自律知能システム（Autonomous and Intelligent System、以下 A/IS）の設計、技術などについての課題と対応策を整理したものである。日本でも EAD の論点について異分野・異業種によるワークショップがおこなわれている(4)。

教育分野では早くから人工知能が使われており、技術の発展の影響は大きいと思われるが、ヘルスケアなど他領域とくらべて技術の適用への倫理的検討が十分であるとはいえない。この報告では EAD を元に教育システムや教育データに関する「倫理的に配慮された設計」の論点について整理、検討をおこなう。

2. IEEE Ethically Aligned Design

2.1 IEEE Ethically Aligned Design の概要

EAD は、人工知能に代表される知的で自律的なシステム（Autonomous and Intelligent System=A/IS）が人間や社会の価値観や倫理原則と適合するよう、潜在的な有害性と望ましさに関する対話、議論、ポリシーを促進するために作成された報告書である。EAD の第 1 版（EADv1）は 2016 年 12 月に、第 2 版（EADv2）は 2017 年 12 月に公開された。

EADv2 では v1 からの継続を含めて、以下の 13 のトピックについての議論が委員会を組織して進められた。

1. 一般原則（General Principles
2. 自律知能システムへの価値観の埋め込み（Embedding Values into Autonomous Intelligent Systems）
3. 倫理的研究と設計をガイドする方法論（Methodologies to Guide Ethical Research and Design）
4. 汎用人工知能の安全性と便益（Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)

5. パーソナルデータとアクセスコントロール Personal Data and Individual Access Control）
6. 自律兵器システムのりフレーム（Reframing Autonomous Weapons Systems）
7. 経済、人道上の課題（Economics/Humanitarian Issues）
8. 法律（Law）
9. Affective Computing
10. 政策（Policy）
11. A/IS における伝統的倫理（Classical Ethics in A/IS）
12. 複合現実（Mixed Reality in ICT）
13. ウェルビーイング（Well-being）

EAD には多分野から産官学民メンバーの参加者内での共有と、今後の議論の多様性の確保や社会的な展開のために、使用するキーワードの定義した用語集（Glossary）が作成された。用語集においては、各キーワードについて、一般用語、コンピューター関連領域、工学領域、政策・社会科学領域、倫理・哲学で使用されている定義を整理している。

2.2 標準化

IEEE は EADv1 および EADv2 の議論を元に P7000 シリーズとして標準化を目指している。P7000 シリーズは最終版の EAD に反映される予定である。現在、現在以下のオープンなワーキング・グループが P7000 に存在する。

1. システムデザインにおいて倫理的問題を取りあつかうモデルプロセス（P7000）
2. 自律システムの透明性（P7001）
3. データプライバシープロセス（P7002）
4. アルゴリズムによるバイアスの考慮（P7003）
5. 子どもと学生のデータガバナンスに関する標準規格（P7004）
6. 透明性のある雇用者のデータガバナンスに関する標準規格（P7005）
7. パーソナルデータ人工知能に関する標準規格（P7006）
8. 倫理的に動作するロボットおよび自動システムに関するオントロジー的標準規格（P7007）
9. ロボット Standard for Ethically Driven Nudging for Robotic、 Intelligent、 and Automation Systems (P7008)
10. 自律的および半自律的システムのフェイルセーフなデザインに関する標準規格（P7009）
11. 倫理的な AI と知能システムのためのウェルビーイング・メトリクスの標準規格（P7010）
12. ニュースソースの信頼性を特定し評価するプロセス標準(P7011)
13. 機械可読なプライバシー条項の標準規格（P7012）
14. 自動化された顔分析技術の組み込みと適用の基準（P7013）

3. 教育システムに関連した EADv2 の論点

教育システムにおける AI の研究開発は、ICAI（Intelligent Computer Assisted Instruction）や ITS（Intelligent Tutoring System）として活発におこなわれてきた。これらの自律知能システムは、学生-教師のインタラクション（の一部）を、事前の知識を使って代替または強化するものである。また、これらの研究と関連して（あるいは並行して）、テストの結果から推定した学生の能力にもとづいて、個々の学習者に応じたインストラクションやテストを提示するパーソナライズド・ラーニングや適応的テストの研究も盛んである。このようなパーソナライズド・ラーニングは、学術研究以外の教育産業等に多数の事例がある。

このように教育分野においても自律知能システム（A/IS）の研究は豊富である。また、教育において知識とデータは本質的に重要であり、教室等の教育場面に導入されるロボット、ソフトウェアエージェント、IoT デバイス、VR デバイスなど A/IS のもたらすインパクトは大きいであろう。しかし、そのインパクトの評価や倫理上の課題について十分に検討されているとはいえない。EADv2 においても A/IS 開発者の教育は言及されているが、教育そのものへの A/IS の影響については取りあげられていない。

教育システムにおける倫理的に配慮されたデザインを検討する手がかりとするために、EADv2 において、教育システムに関連するトピックとその論点を表 1 に

整理した。以下では教育分野で特に独自の検討を要するトピックと思われる「5. パーソナルデータ」と「4. 汎用人工知能」「9. Affective Computing」「12. 複合現実」について述べる。

パーソナルデータ: 教育においてさまざまなシステムが生成するデータは教育の改善や学生自身の学習のリフレクションに有用である。その一方で、データの取り扱いにはさまざまな課題が存在しており(5)、今後教室等での A/IS の実用とともにさらに検討課題が増加するであろう。学生（教師も）は A/IS によって自動的に個人のデータが収集されていることをおそらく明確に理解していない。もし、データの利用目的や分析方法、分析結果の利用者について同意があったとしても、それが抽象的で包括的な形式的なものである可能性も高いであろう。また、学校内の関係性や教育への影響によって事実上同意を拒否できない問題もある。未成年、高齢者、障害者の場合の同意の方法も課題である。

個人データに関する議論では、個人がパーソナルデータの利用管理、制限をおこなう方向で望ましいという意見がある。しかし学生にそれが可能であるか。また、過去のパフォーマンスのデータからのプロファイリングの共有と誤る可能性や、そしてプロファイリングから選択された教授法の適切性など、今後の研究が必要であろう。

汎用人工知能／Affective Computing／複合現実: 教育分野において、これらの開発は学術研究よりも産業が先行している。複合現実への没入は学習を向上させる可能性はあるものの、健康上の懸念や仮想的な経験の現実への影響は未知である。また、このようなシステムが教員を代替する、あるいは教員が利用、連携する際の方法については研究が必要である。

4. 今後の課題

この報告では EADv2 の論点の中で、教育システムに特に関連したトピックと論点を抽出した。しかし、これは人工知能の教育への応用における倫理上の課題の一部にすぎない。

EADv2 は自律知能システムに関する報告書であって、人工知能領域すべてをあつかうものではない。教育分野に影響の大きい機械学習やディープラーニングなどの問題はパーソナルデータの論点として触れられるだけである。教育におけるデータからの個人のプロファイリングは、評価・選別、人権、差別、プライバシーなどへの長期的な影響がある。

また、教育は公教育から生涯教育、さらにフォーマル、インフォーマルを含む長期的な営みである。影響が長期に渡ることは、ヘルスケア等と比べたときの教育データの特徴であり、教育研究を踏まえたうえでのデータの倫理的な取り扱いが必要であろう。

これらのトピックを含めた教育の目的、方法、システムなどへの AI の影響は、AI に関する諸議論を踏まえて、広範囲なステークホルダーのさまざまな立場から議論が必要であろう。EADv2 は議論のための基礎資料として有用である。

謝辞

本研究は JSPS 科研費 16K12564 の一部の助成を受けている。

参 考 文 献

(1) 村上祐子: “人工知能の倫理の現在”、 IEICE Fundamentals Review、 Vol. 11、 No. 3、 pp. 155–163 (2018)

(2) AI ネットワーク社会推進会議: “報告書 2018—ＡＩの利活用の促進及びＡＩネットワーク化の健全な進展に向けて—” (2018) http://www.soumu.go.jp/main_content/000564147.pdf (2019 年 2 月 7 日確認)

(3) The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: “Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems Version 2”, http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf (2019 年 2 月 7 日確認)

(4) 江間有沙, 長倉克枝: 「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、第 4 号、pp. 3-14 (2018)

(5) Ho, A.: “Advancing Educational Research and Student Privacy in the ‘Big Data’ Era” Washington、DC: National Academy of Education (2017)

表 1 教育システムに関連した IEEE Ethically Aligned Design の論点と課題

トピック	論点（issue）
1. 一般原則	人権侵害の防止、ウェルビーイング、アカウンタビリティ、透明性、技術の誤用の認識、規範
2. 価値観	規範の識別、更新、コンフリクトとその解消、多種多様な規範の実装、実装と展開の透明性、失敗の可能性、規範が実装されないコミュニティの存在、特定のグループに不利なバイアス、第三者による評価
3. 研究と設計の方法論	研究者への倫理教育、倫理的課題に関する学際的な共同研究、ステークホルダーの関与、倫理委員会のための組織内リソース、ドキュメンテーション、アルゴリズムの一貫性、監視の欠如、独立した審査機関の不在、ブラックボックスコンポーネントの使用
4. 汎用人工知能	汎用人工知能の領域横断性と予期しない動作の危険性、安全性の設計を組み込む困難さ、倫理的技術的な安全上の課題の複雑化、世界的で大規模なインパクト
5. パーソナルデータ	デジタルと現実の人格の差異、パーソナルデータの定義とその識別、法律と個人の価値の矛盾、個人を識別する情報（PII）と個人データのコントロールの定義、収集された情報へのアクセス、訂正、修正、管理、プライバシーインパクト評価の作成、政府による情報の収集、パーソナライズ化された AI によるプライバシー保護、同意の再定義、共有を望まない情報のプロファイリング、利用していない A/IS による個人情報の取得、同意のベストプラクティス、個人情報の同意を理解できない場合
7. 経済	SDGs、発展途上国における A/IS、雇用構造と A/IS、市場以外への自動化の影響、従業員訓練の技術的変化への対応、個人情報の理解の欠如、A/IS 分野の大学教育におけるグローバルな側面、AI と自律技術の国際格差
8. 法律	法的地位または法的分析の枠組み、行政による権利侵害の可能性、A/IS の法的責任を保証する設計、アカウンタビリティと検証可能性の改善
9. Affective Computing	適切な規範、長期利用の影響、文化・社会・宗教的価値へのマイナスの影響、ユーザーとの親密な関係の道徳的倫理的影響、ナッジの利用の倫理的懸念と意図しない結果、アフェクティブ・システムの欺瞞の可能性、個人の自律性への影響
10. 政策	A/IS が法的基本を理解することの保証、A/IS 人材の育成、AI 産業の育成と公共の安全と責任のための規制の両立、A/IS の利益とリスクに関する国民の理解の醸成
11. 伝統的倫理	道徳、自律、知性についての基本的知識、agent と patient の区別、伝統的な倫理学の語彙、開発者への倫理の提示、企業による倫理学へのアクセス、職場でのインパクト、人間の自律性の維持、プラグラミングのためのルールベース倫理学の要件
12. 複合現実	フィジカルな現実への影響、社会との関係、社会規範、健康への影響およびその評価、仮想の経験の問題、トレーニングや業務の有効性とそれがもたらす変化、子どもや未成年者の利用上の問題、専門的仕事の自動化への影響、データの収集と制御の法的倫理的課題、公共の場での AR/VR に関連したデータとプライバシーの法律や規制
13. ウェルビーイング	ウェルビーイングのメトリクスと A/IS がもたらすシナリオやインパクトのモデル化、評価の A システムへの組み込み、 ウェルビーイングが人権とコンフリクトする可能性