

IEEE「倫理的に調和した設計」を用いた議論の場とコミュニティの設計

Design dialogue and community to discuss IEEE “Ethically Aligned Design”

江間有沙*¹
Arisa Ema

長倉克枝*²
Katsue Nagakura

工藤郁子*³
Fumiko Kudo

*¹ 東京大学
The University of Tokyo

*² 科学ライター
Science Writer

*³ マカイラ株式会社
Makaira KK

This document reports the methodology and process of inter-disciplinary workshops involving various stakeholders to discuss IEEE *Ethically Aligned Design (EAD)* report. *EAD* report aims to move beyond the excessive fear and expectations associated with AI and increase innovation by creating ethically aligned designed AI. Authors organized the report items into 6 workshop series. The goal of our workshops is to understand the content of *EAD* and create a network among AI/IT researchers, social science and humanities researchers, industries, policy makers and other stakeholders.

1. はじめに

「今までは情報技術で倫理と言われていなかったのに、なぜ今、人工知能でこれほど倫理の議論がされているのか」

このような質問を受けることが多いが、情報技術倫理の研究は 1980 年代から存在する。それにもかかわらず「今更」と思われてしまうことからは、多くの方にとって情報技術の倫理や価値の議論が身近ではないこと、そして倫理の議論は技術開発の制約と見なされがちであることが再確認される。

技術の設計段階からその倫理的、法的、社会的な観点について考えていくためには、技術者にとっても倫理や価値の議論が自分事として考えられるような場や環境づくりが重要となる。そのため、技術に関する倫理や価値の「中身」の議論だけではなく、多様な人々を巻き込むという「プロセス」や、コミュニティ形成の方法についても実践と記録を残していくことが、重要である。

筆者らは、これまでも「人工知能と社会について考える場作り」として、様々な分野・業種の専門家を対象とした企画を行ってきた[江間 17a]。本稿では、IEEE が公開している報告書に関するワークショップを事例として、どのように異分野・異業種の専門家による企画を組織、運営しているかを紹介する。

2. IEEE「倫理的に調和した設計」とは

2016 年 12 月に The IEEE Global Initiative (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems) が IEEE *Ethically Aligned Design* (「倫理的に調和した設計」) の第一版(以下 EADv1)を公開した。この報告書は知的な機械システム(Autonomous and Intelligent System)に対する恐怖や過度な期待を払拭すること、倫理的に調和や配慮された技術をつくることによってイノベーションを促進することが目的である。倫理そのものではなく、設計論や設計思想、それをどのように技術に落とし込めるかといった論点が整理されており、人工知能、法、倫理、哲学など他領域にわたる研究者、政策関係者、企業関係者の意見を集約して作成された。作成に関わった人は全世界で 250 人以上に上る。

続く 2017 年 12 月には第二版(EADv2)が公開され、以下の 13 項目から構成される[IEEE 17]。

1. 倫理的に調和したデザインをするための一般原則
2. 自律型知的システムに価値観を組み込む
3. 倫理的な研究や設計のための方法論やガイド

4. 汎用人工知能や人工超知能の安全性や便益
5. 個人データとアクセス制御
6. 自律型兵器システムの再構築
7. 経済 / 人道的課題
8. 法律
9. アフェクティブコンピューティング
10. 政策
11. 情報通信技術における伝統的倫理観
12. 複合現実
13. ウェルビーイング

EADv2 の公開と同時に、「自律的で知的なシステム」に関連する用語集(Glossary)委員会が組織された。分野も業種も違う人々が集まって作成している報告書だからこそ、定義を明確化する必要がある、というニーズから派生した委員会である。委員会では、アカウントビリティ、データ、人権などの多義的なキーワードに対し、それぞれ(1)一般用語、(2)情報系、(3)工学系、(4)政策や社会科学系、(5)倫理や哲学など、他領域での定義の整理が行われている。

今後、IEEE は約 1 年間のフィードバック期間を経て、2019 年に最終版となる第三版を公開する予定である。また、この EAD 報告書執筆陣を中心として、標準化を目指すプロジェクトが IEEE-SA (Standard Association: 標準規格) に承認されている[江間 17b]。2018 年 3 月現在、IEEE-SA の P7000 シリーズとして 13 ワーキング・グループが標準化を目指しドラフトを作成中である。

3. ワークショップ概要

3.1 EADv1 ワークショップの概要

第 1 回のワークショップを筆者らは 2017 年の 4 月から 5 月にかけて名古屋と京都で 1 回ずつ、東京で 2 日間連続開催し、総勢 44 名の方にご参加いただいた。特設ウェブサイトを作成し(<https://sites.google.com/view/ethically-aligned-design-ws/home>)、案内を各種関連学会などのメーリングリストに流した。ウェブページでは報告書の日本語版概要と本文がダウンロードでき、参加者は報告書に目を通していただくことを条件とした。各回では毎回、報告書の概要や今後の方針についての概説があり、回によっては Skype で Executive Director の John C. Havens 氏や

Vice-Chair の Kay Firth-Butterfield 氏と繋いで直接、意見交換を行った。

ワークショップの目的は多様な意見をまとめてコンセンサスを得るのではなく、様々な視点を紹介することであった。そのため、ワークショップ報告書の執筆にあたっては、各項目に対して一文字ずつ追って精査していくというよりは、EADv1 全体やその背景に対する意見(総論)と、各項目に対する意見(各論)を整理し、提案を行った。ワークショップ参加者全員から修正や加筆をいただくプロセスを経た報告書は、現在ウェブ上に公開されている。

なお、各論点の発言者名を特定はしていないが、どのような属性の人の発言かは情報として重要であると考えたため、報告書では発言者コメントに属性を示した。属性の種類とワークショップ参加者の内訳は、アカデミアの情報系(11名)アカデミアの人文・社会科学系(12名)、パブリックセクター(6名)、インダストリー(5名)、メディア(5名)、その他(5名)であった。報告書は2017年6月に米オースティンで開催された「自律システムの倫理に関するシンポジウム」(Symposium on Ethics of Autonomous Systems: 以下 SEAS)に提出され、各項目の担当者へフィードバックされた。



図1 EADv1 ワークショップの様子

3.2 EADv2 ワークショップの企画概要

EADv1 に対して全世界から寄せられたコメントを受けて、2017年の12月にEADの第二版が公開された。第一版から5項目新たに追加された13項目からなり、ページも第一版の138ページから約2倍の266ページとなった。

EADv1 ワークショップの反省として、(1) 少ない時間ですべての項目を扱うのが難しく表面的な議論になりがちであったこと、(2) 日本の具体的な事例などを埋め込むことができなかったことが挙げられた。そのため第2回ワークショップを開催するにあたっては、EADv2の全13項目を読んだ上で、扱っているテーマが近い項目をオーガナイザ側で6つに再分類し(表1)、それぞれの項目についてEADv2を参照して話題提供をいただけるゲスト講師をお招きするという形式を採用した。

ゲスト講師は、各テーマに対し研究者やジャーナリスト、企業の方で、(1) 技術的なアプローチをされている方と、(2) 人文社会科学的なアプローチをされている方をそれぞれ招きした。またEADv2を読み込んで解説を行っていただくのではなく、EADv2を1つの指標(インデックス)と見なし、それに対して各話題提供者が知っている情報や研究について具体的な事例を紹介していただくようお願いした。各回ともオーガナイザ側と話題提供者とで事前打ち合わせを行い、EADv2に関する論点と日本の事例についてクローズドで議論も行った。

全6回の順番は、まずは自律型兵器や汎用人工知能(第1回)、アフェクティブコンピューティングや複合現実(第2回)など技術的にイメージしやすい項目から開始し、そこから社会や倫

理について議論するための方法論(第3回)、法(第4回)、基盤(第5回)、最後に倫理・原則(第6回)といった抽象度の高いテーマへ進めていくこととした。

表1 各回のテーマと項目

第1回 リスク	4. 汎用人工知能や人工超知能の安全性や便益 6. 自律型兵器システムの再構築
第2回 感情	9. アフェクティブコンピューティング 12. 複合現実
第3回 社会	3. 倫理的な研究や設計のための方法論やガイド 7. 経済 / 人道的課題 10. 政策
第4回 法	5. 個人データとアクセス制御 8. 法律
第5回 基盤	11. 情報通信技術における伝統的倫理観 13. ウェルビーイング
第6回 倫理・原則	1. 倫理的に調和したデザインをするための一般原則 2. 自律型知的システムに価値観を組み込む

また、本ワークショップは多様な視点を持つ人に参加してもらうべく、オーガナイザや話題提供者が所属している組織だけではなく、各種関連学会にもパートナーとして後援いただいているほか、それぞれのメーリングリストでも告知を行った。



図2 EADv2 ワークショップの様子

4. おわりに

本稿を執筆している2018年3月上旬では第3回までが終了しており、今まで情報系と人文・社会科学系の研究者や企業関係者、政策関係者、学生などが参加している。また、各回の話題提供の内容とスライドはウェブに報告書が掲載されている。

全6回のワークショップ終了後には、全体を通しての報告書を日本語と英語で執筆する予定である。また、本ワークショップシリーズを通してどのようなコミュニティが形成され議論が継続されているのかについては、当日会場で議論を行いたい。

参考文献

- [江間 17a] 江間有沙 他: 人工知能と社会について考える場づくりの実践, 人工知能学会第31回全国大会予稿, pp. 4G2-OS-14b-3, 2017.
- [江間 17b] 江間有沙: 倫理的に調和した場の設計: 責任ある研究・イノベーション実践例として, 人工知能, Vol. 32, No. 5, pp. 694-700, 2017.
- [IEEE 17] The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems: Ethically aligned design: a vision for prioritizing human wellbeing with artificial intelligence and autonomous systems, <https://ethicsinaction.ieee.org/> (2017)