「第 5 回『AI の哲学』会議」参加報告*

清水颯・宮原克典

概要

This is a short report on the 5th Philosophy of AI conference that took place in Erlangen, Germany, on 15-16 December, 2023. The conference featured four keynote lectures by leading researchers in the field, Sven Nyholm (LMU Munich), Marta Halina (Cambridge University), Herman Cappelen (University of Hong Kong), and Joanna Bryson (Hertie School), 29 oral presentations, and 19 poster presentations. The keynote lectures each addressed fundamental issues identified at the intersection of philosophy and AI, such as the meaning and authorship of AI-generated text (Nyholm), the role of AI in scientific explanations and scientific understanding (Halina), the reliability of introspective self-reports by AI (Cappelen), and the conditions of moral agency ascription in relation to AI (Bryson). In the first section, we briefly overview each keynote lecture and also our own presentation delivered on the second day of the conference. It explored whether we should give moral treatment to LLM-based chatbots drawing on the perspective of virtue ethics. We proposed a negative answer to the question, highlighting the notion of practical wisdom. In the second section, we report on the kickoff meeting of the newly founded Society for the Philosophy of Artificial Intelligence (SPAI), which also took place during the conference. SPAI aims to foster research and communication among researchers across the globe in the field of philosophy of AI. It has launched a new open-access journal focused on issues in philosophy of AI, Philosophy of AI. We conclude with an invitation to the readers to join in this emerging field of research.

Keywords: AI の哲学、国際学会、学会参加報告

1 はじめに

人工知能（AI）の哲学・倫理学は、現在、哲学・倫理学の分野で最も注目を集めているテーマの一つで

* *CAP* Vol. 15 (2024) pp.172-177 . Submitted: 2024.5.2. Accepted: 2024.9.11. Category: 研究ノート. Published: 2024.9.25.

ある。2010 年代以降、深層学習技術の飛躍的な発達によって、いわゆる「第 3 次人工知能ブーム」が到来した。近年では、大規模言語モデル（Large Language Model; LLM）の開発が加速的に進展し、2022 年の ChatGPT のリリースによって多くの人がそのインパクトを実感するようになった。このような技術の進展と並行して、AI の開発と普及がもたらす哲学的・倫理的な諸問題をめぐる研究への取り組みも世界各地で急激に進んでいる。

2022 年、ドイツのエアランゲン大学では、アレクサンダー・フォン・フンボルト財団の資金提供を受けて「哲学と AI 研究センター（Centre for Philosophy and AI Research, PAIR）」と呼ばれる研究所が設立された。創設者は、2000 年代初頭から AI の哲学・倫理学に関する研究にいち早く取り組み、この分野で著名なヴィンセント・C・ミュラー（Vincent C. Müller）である。2023 年 12 月には、本センターが主催する初めての国際学会として「第5回『AI の哲学』会議（5th Conference on “Philosophy of AI”, PhAI 2023）」が当地で開催された。筆者たちは、本会議で AI の哲学の国際的な研究動向の最前線に触れて大きな刺激を受けた。本稿では、本会議の様子を報告することで、AI の哲学・倫理学の最近の情勢を筆者たちの見聞した範囲で紹介したい。

2 PhAI 2023 の報告

「AI の哲学」会議は、これまでテッサロニキ（2011 年）、オックスフォード（2013 年）、リーズ（2017 年）、ヨーテボリ（2021 年）の欧州各地で数年おきに開催されてきた。筆者たちが参加した第5回大会は「PhAI2023」とも呼ばれていた。AI の哲学・倫理学への世界的な関心の高まりの影響で、今大会は過去の大会に比べて発表応募数が多かった（181 件）ようで、開会の挨拶のなかでミュラー教授は会場規模や日程の問題で採用率を「たったの 18%」にしなければならなかったことを残念そうに語っていた。（「たったの」という表現は公式サイトにも表記されている）。

会議の規模はそこまで大きいものではなく、二日間で基調講演 4 件、一般発表 29 件、ポスター発表 19 件がプログラムされていた。基調講演者として招待されたのは、スヴェン・ニホルム（Sven Nyholm)、ハーマン・カッペレン（Herman Cappelen）、マータ・ハリナ（Marta Halina）、ジョアンナ・ブライソン（Joanna Bryson）の 4 名であった。教会を改装したイベント施設「KREUZ+QUER」が会場として使用された。一般発表は、二つのパラレルセッション形式で実施され、それぞれ 120 人程度、80 人程度を収容できそうな中規模会場で行われた。参加者全体の人数は（おそらく）100 人ほどであり、講演やポスター発表の時間だけでなく、コーヒーブレイクやディナーの時間にも、連日活発な議論と社交が繰り広げられていた。ドイツで開催されているだけに欧州各地からの参加者が多く、そのほかにシンガポールや香港などの研究者も参加していたが、日本からの参加は筆者たちだけであった。

現在、AI の倫理的・社会的・法的諸問題（いわゆる ELSI）に関する研究に対する社会的な要請が急激に高まっているが、本学会は狭い意味での「AI 倫理」にテーマを限定していない。大会サイトでは、大会趣旨に合致した研究の具体例として、AI の限界やポテンシャルに関する研究、AI が人間社会にもたらす機会やリスクを明らかにする研究、AI を用いて非人工的な知能を理解する研究、科学や哲学の諸領域で

AI を事例や道具として用いる研究が挙げられていた。実際、AI 倫理の問題を論じる発表はどちらかといえば少数派であり、科学哲学、言語哲学、心の哲学などの従来の研究で培われてきた概念や方法論を用いて、AI の特性や AI を用いた実践を理論的に考察するタイプの発表が目立っていた。社会への悪影響や利益を考慮してAIを使用・設計するための倫理的方針の検討など、社会的要請に応じた研究を行うことはもちろん重要であるが、それればかりに気を取られるのではなく、このように必ずしもただちに社会から解決を要請されているわけではない根本的な理論的問題への取り組みも忘れてはならないとあらためて思わされた。

以下では、大会の基調講演の概要といくつかの印象に残った出来事を紹介したい。あくまでも筆者たちが聞いた限り、かつ覚えている限りでの紹介ではあるが、大会の様子がいくらかでも伝われば幸いである[1]。

一日目は、ミュラー教授の開会の挨拶に続いて、ニホルム教授（ルートヴィヒ・マクシミリアン大学ミュンヘン）による基調講演「生成 AI のギャップ：有意味性、著作者性、名誉と責任の非対称性（Generative AI's Gappiness: Meaningfulness, Authorship, and the Credit-Blame Asymmetry）」から始まった。ニホルム教授は、AI 倫理の分野で広範に活躍する哲学者・倫理学者であり、「自動運転車のトロリー問題」「責任ギャップ問題」「ロボットの行為者性の問題」などに関して顕著な業績を挙げている。この講演では、ChatGPT などの生成 AI による画像や文章などのアウトプットの「作者（author）」や「意味」をめぐる問題が論じられた。ユーザーと生成 AI のアウトプットは従来の作者と作品の関係とは違うものであり、その特性を正確に評価するためには、これらの伝統的な概念の刷新が必要となると主張された。

午後には、あいだにコーヒーブレイクを挟んで、二つのポスターセッション、二つの一般発表セッション、そして、「AI 哲学会（Society for the Philosophy of Artificial Intelligence; SPAI）」のキックオフミーティングが実施された。こちらのキックオフミーティングについては次節であらためて取り上げたい。

夕方には、ハリナ准教授（ケンブリッジ大学）の基調講演「AI の文脈における志向的説明と機械的説明（Intentional and mechanistic explanations in the context of AI）」が行われた。ハリナ准教授は、認知科学・心理学・生物学の哲学を専門とする科学哲学者である。動物の認知やコミュニケーション、AI の認知、機械論的説明などに関して幅広い研究を行っている。この講演では、科学的な説明に AI モデルを使用したときの AI モデルに対する理解可能性のあり方が論じられ、そのためには AI モデルに対する「志向的説明」だけではなく「機械論的説明」が必要であると主張された。近年、AI には科学研究の文脈でも大きな期待を寄せられているが、これが科学という実践のあり方を改めて問い直す契機でもあることを鮮やかに描き出す講演であった。

二日目は、カッペレン教授（香港大学）の基調講演「私たちは ChatGPT が自分について語ることを信頼すべきか（あるいは、気にすべきか）？（Should we trust (or care about) what ChatGPT tells us about itself?）」から始まった。カッペレン教授は、言語哲学や概念工学に関する多くの業績をあげている哲学者であり、近年は、AI の哲学や政治哲学に関する研究にも精力的に取り組んでいる。この講演では、大規

[1] 大会プログラムや発表要旨を確認したい場合は、次のサイトを参照いただきたい。
https://www.pt-ai.org/2023/programme

模言語モデル（LLM）が一人称的／内観的な自己報告に相当する文章を生成したとき、それを内観的な自己報告として信頼してよいのか、LLM を経験や知性をもつ存在として解釈することは妥当なのかを詳しく検討し、LLM の自己報告と人間の自己報告に根本的な差異はないという立場を擁護した。Bard[2]や ChatGPT との会話の実例をあげながら聴衆の心をつかむ生き生きとした講演が印象的であった。

この日も、午後には、ポスターセッションと一般発表セッションが二回ずつ行われた。筆者たちは、後半の一般発表セッションで「真正な社会性と人工的な社会性とを見分けること（Discerning genuine and artificial sociality）」と題した発表を行った。はじめに、2022 年 6 月、Google 社のエンジニアが大規模言語モデル「ラムダ（Language Models for Dialog Applications；LaMDA）」を人格（パーソン）として認めるべきだと主張して話題を呼んだ実例を紹介した。本発表では、このように大規模言語モデルを一人の主体として認めることの倫理的な是非を徳倫理の観点から検討した。一部の先行研究によると、徳倫理の観点は、人間は社会的なロボットに対して道徳的な配慮をする必要があることを含意する。そうだとすると、人間は（大規模言語モデルに基づく）チャットボットも一人の主体として扱ってしかるべきだと思われるかもしれない。それに対して、筆者たちは、実践知（フロネーシス）を備えた賢人であればチャットボットの「発言」に簡単に惑わされて、チャットボットと本物の主体を混同しないはずだと論じ、徳倫理はチャットボットを道徳的配慮の対象として扱うべきだということは含意しないと主張した。質疑応答の時間では、問題となる実践知の具体的な内容や主体性の意味に関する鋭い質問を受けた。今後の研究の進展に向けて大変有意義な機会であった。

二日目の最後には、ブライソン教授（ヘルティ・スクール）による基調講演「言語、意識、正義、そして、AI の規制（Language, Consciousness, Justice & AI regulation）」が行われた。ブライソン教授は AI を含めた知能研究で国際的な著名な研究者であり、企業、政府、国際機関、NGO などで AI 政策に関するアドバイザーも務めている。1990 年代終盤から世界に先駆けて擬人化された AI が生み出す混乱について考察を行ったり、世界初の国家レベルの AI 倫理政策である英国の「ロボティクス原則（Principles of Robotics）」（2011 年）の立案にも関わったり、名実ともに AI の哲学・倫理学の第一人者である。この講演では、人工知能は人間の道徳的主体性の条件を探究する機会を提供すると論じ、特に「意識をもつこと」は道徳的主体性の必要条件ではないと主張した。道徳的主体、正義、責任などの概念は、社会を規制し存続させるために構築された調整システムであり、そのシステムの構成員であるかどうかを決めるときに意識的な内面性があるかどうかは重要なポイントではないということであった。

3 AI 哲学会と国際誌「AI の哲学」

大会一日目にキックオフミーティングが行われた「AI 哲学会（Society for the Philosophy of Artificial Intelligence; SPAI）」について、少し詳しく紹介したい。発起人のヴィンセント・ミュラー（エアランゲン大学）、ギド・ロア（Guido Löhr）（アムステルダム自由大学）、カルロス・ゼドニク（Carlos Zednik）（アイントホーフェン工科大学）によると、これまで AI 哲学をメインテーマとする学協会が存在せず、AI の哲学や倫理学

[2] Google が開発した生成 AI である Bard は、現在 Gemini という名称で知られている。

に関する研究は、AI に関する人文社会的研究全般、技術哲学、技術倫理などに関する専門誌――たとえば『AI と社会（AI & Society）』『技術と哲学（Technology and Philosophy』『倫理と情報技術（Ethics and Information Technology）』『科学倫理と工学倫理（Science and Engineering Ethics）』など――に発表されてきた。しかし、AI に関する哲学的研究がその学術的・社会的な役割を有効に果たせるようになるためには AI 哲学の発展をミッションに掲げる新たな学協会が必要だと考え、本学会 SPAI の設立に動き出したということであった。

今後の SPAI の主な活動内容は、学術会議の開催、ジャーナルの創刊、会員向けメーリングリストの運営の三つだという。キックオフミーティングでは、2025 年にアムステルダムで「第 6 回『AI の哲学』会議」を開催する計画を進めていることが知らされた。現時点で詳細は未定だが、第 5 回の参加者数が予想を大きく上回っていたことをふまえて、次回はより大きな規模で実施予定だということであった。ジャーナルについては、ドイツ研究財団（Deutsche Forschungsgemeinschaft、DFG）の財政的支援を受けて、AI の哲学に関するオープンアクセスジャーナル『AI の哲学（Philosophy of AI）』を創刊予定であることが発表された。本稿執筆時点（2024 年 4 月 10 日）で、すでに一般公募論文の投稿受付が開始されており、「言語と AI」に関する特集号への論文の投稿受付も開始されている[3]。メーリングリストには、学会の公式サイト（https://philai.net/）から誰でも簡単に登録することができる。

ミュラー教授はこの分野で著名な研究者であるが、ロア博士とゼドニク博士はまだ若手～中堅の研究者である。そのようなキャリアステージの研究者が、AI 哲学という分野の将来と存在意義を見据えて、新たな学協会と国際ジャーナルを設立するという大仕事に意欲的に取り組んでいるのは非常に印象的であった。ドイツやオランダでは（AI を含めた）技術の哲学・倫理学に関する先進的な議論が活発に進んでおり、そのような環境だからこそ生まれてきた運動なのだろう。一方で、発起人たちは、SPAI がドイツやヨーロッパだけを対象にした地域学会ではなく、世界各地の研究者を対象にした AI の哲学に関する国際学会を志向していることも強調していた。大会後、ロア博士との相談の結果、AI 哲学分野における東アジア地域とヨーロッパ地域の研究交流を活発化させる一つのきっかけとして、筆者の一人（宮原）は領域編集者として『AI の哲学』誌の運営に参画することになった（AI の哲学に関する英語論文を準備されている方には、ぜひ投稿先として検討いただきたい）。

4 むすび

AI は、従来の人間社会を大きく変革させるだけではなく、AI をめぐる哲学的な問いに対する学術的な関心も急速に再燃させている。たとえば「AI は意味のあることを言えるのか」「AI が生む科学的知識は世界に対する理解を深めてくれるのか」「AI に意識や心を認められるのか」「AI に責任を問うことはできるのか」といった問題である。どの問いも明確な答えはなく、議論されなければならない課題が山積している。研究者にとって、AI の哲学は、新たな理論や概念を創造する余地が無限に広がる「ブルーオーシャン」だと言ってもいいだろう。

[3] 詳細はジャーナル公式サイトを参照いただきたい https://journals.ub.uni-koeln.de/index.php/poai/

最後に、こうした種明かしをするのも今となっては陳腐なやり口かもしれないが、本稿の執筆にあたっては、ChatGPT に下書きを依頼して筆者たちがそれを編集する、という手順をふんだ段落がいくつかある[4]。最初から最後まで人間が執筆した文章とのあいだに質的な違いはあっただろうか。AI に下書きを任せたのだとしたら、これは本当の意味で私たちが執筆した文章だと言えるのだろうか。学会報告ならまだしも、学術論文であったとしたら、このような執筆方法は倫理的に許容できるだろうか。AI をめぐる哲学的思考への扉は、いまや日常的な活動のいたるところで開かれている。

著者情報

清水颯（北海道大学）
宮原克典（北海道大学）

[4] 具体的には、4 名の基調講演者の紹介文を執筆する際に、ChatGPT（無料版）にインターネット上に公開されているプロフィールや業績の情報を入力して、下書きとなる文章を出力させた。しかし、冗長で長い文章が生成されることが多く、そのままではとても使うことができないので、その下書きを編集することで文章を作成した。当然ながら、内容の正しさ、表現の適切さについては著者が責任を負う。