

## 社会技術研究開発事業

### 科学技術の倫理的・法制度的・社会的課題 (ELSI) への包括的実践研究開発プログラム 研究開発プロジェクト 事後評価報告書

「科学技術の倫理的・法制度的・社会的課題 (ELSI) への包括的実践研究開発プログラム」  
プログラム総括 唐沢 かおり

#### 1. 課題代表者

宇佐美 誠（京都大学 大学院地球環境学堂 教授）

#### 2. 課題名

人工知能の開発・利用をめぐる自律性および関係性の理論分析と社会実装

#### 3. 実施期間

2021(令和3)年10月1日～2025(令和7)年3月31日

#### 4. 事後評価結果

本プロジェクトは、法哲学・社会哲学・倫理学などと工学の緊密な協働の下、AI やそれを用いたビッグデータ分析の技術開発・利活用において生起する倫理的課題を発見・予見した上で、「個人の自律性および自律的 AI」と「人間-AI の関係性」という 2 つの観点から理論分析を実施し、その成果を基盤に、ステークホルダーとの連携の下、AI の開発・利用の現実やそれを取り巻く日本社会の特徴を踏まえつつ、開発者がとるべき行動や、開発・利用を行う企業が設けるべき制度に関して、技術開発者の行動指針 (CoC) を公表することで責任ある研究・イノベーションの定着を目指した。

また、本プログラムにおける研究開発として、研究開発において、「個人と AI」だけでなく、AI 同士や AI と社会との関係も考慮し、技術の枠組みで捉えられない部分を明示し、精緻な検討を進めることが期待した。

#### A. 総合評価

一定の成果が得られたと評価する。

研究開発における目標や構想が明確に設定されており、工学系研究者からなる開発グループと、人文社会系研究者の自律性グループおよび関係性グループそれぞれから、着実に成果を産出してきた。特に、自律性と関係性をキーワードにした理論的な分析においては、AI と人間、社会との関係について幅広いテーマで検討が行われ、学術的にも重要な成果が挙げられていることが評価できる。

生成 AI の登場などにより、AI を取り巻く環境が急速に変化する中、その急速な展開を踏まえた議論への対応として、AI を対象とする ELSI の本質を深掘りする嘗みは、プログラム全体にとっても重要な課題であった。

一方で、各グループの知見を統合し、ステークホルダーとの連携のもとで、策定と公表を期待していた「技術開発者の行動指針 (CoC)」については、課題が残った。まとまりのある成果導出が見込めるものの、第三者からの評価による妥当性の検討、具体的な活用方法や、

それによってもたらされる社会的インパクトについての構想については、明示的ではなく、今後のフォローが必要であると思われる。行動指針がどのような場面で誰によって活用されるのか、また、それが政策形成や産業界、あるいは市民社会にどのような影響を与えるのかを提示し、研究成果の実装の道筋をより明確化する取り組みに期待したい。また、AIと人間の関係性に基づき、それぞれの自律性を考察する営みについては、根源的問いの観点からも、さらに深堀りしていただきたい。

## **B. 項目評価**

### **(1) 研究開発プロジェクトの目標の達成度**

概ね達成したと評価する。

本プロジェクトは、技術開発や企業の開発者との共同探索的会合を通じてAIの開発・利用上の倫理的課題を把握し、人間とAIの関係性の多様な観点から新たな規範倫理の構築を行った。それらの理論分析結果を基に、ステークホルダーと連携しつつ、日本社会の特徴を踏まえた開発者や企業における行動規範を策定し公表した。また、各成果については一般向けの書籍や学術論文を通して発信し、人文社会系における高水準の議論を工学系の研究者へ届けており、個々の目標については一定の成果を挙げていると評価できる。国際的な倫理基準の比較を行い、日本社会の文脈における特徴や倫理的課題の抽出に取り組んだ点も評価できる。

一方で、AIの倫理課題の倫理分析に関しては、プロジェクト全体として体系的にどのような成果発信に繋がったのかが明確ではなかった。理論分析の枠組みや導出されたELSI課題に関する概要について、より精緻な取りまとめが必要であったと思われる。また、社会実装のための行動指針については、論文発表といった研究成果のみならず、社会実装に寄与するCoCの提言がなされているものの、その導出経緯と概要についてのより詳細な説明が望まれるところであった。今後発出されるCoCの持つインパクトに期待したい。

### **(2) 研究開発成果の創出状況**

一定の成果が得られたと評価する。

プロジェクトの体制が整っており、工学系研究者からなる開発グループと、人文社会系研究者の自律性グループ・関係性グループが有機的に繋がるためのマネジメントがなされていた。その結果、理論分析・技術開発における個別成果については、論文または学会における成果発表等、多くの成果が得られたと評価できる。

プロジェクトの実施を通じて複数の大学・研究機関とネットワークを構築し、協力体制を深化させていったこと、国内外の研究者との共同研究を進め、成果の国際的な発信に積極的に取り組んだ点も評価できる。

一方で、各グループにて進めてきた研究が、本プロジェクトの主要な主張をどう補強するのかが、わかりにくいという課題が残った。自律性と関係性というテーマの魅力が本プロジェクトの強みの一つであったが、この点について、統合的な成果として何が得られたのかをより明確にしたうえで、それをわかりやすく世に示すような活動が弱かった点が惜しまれる。

### **(3) 研究開発プログラムの目的達成への貢献度**

一定の成果が得られたと評価する。

人間とAIの自律性および人間とAIの関係性に関するELSI課題は、本プログラムだけではなく人類にとっても重要なテーマである。これに取り組み、代表者およびグループリーダーの専門を生かした学術的に高い水準の業績をあげていることは評価できる。また、技術者に

とっての行動指針を策定するという野心的な目標を掲げ、AIの倫理的課題に関する国際比較を実施し、CoCの提案を行ったことは、研究成果を技術開発の現場に還元するという、本プログラムの目的達成に貢献した点である。

一方で、成果の活用・展開状況においては、さらなる検討の余地があった。例えば、今後、産業界との連携をどのように強化するのか、研究成果を社会実装につなげる仕組みを具体化した構想が明示されることが望ましいと思われる。また、AIの利活用が広範になされている状況を踏まえ、ELSI/RRIのフレームワークをバイオテクノロジーやスマートシティ等の他分野に適用する可能性についても、より一層の貢献が期待できるだろう。AIは多様な技術開発の基盤的な技術でもあることを踏まえ、本プロジェクトの成果が他の科学技術に関するELSIの展開にもつながるような連携を今後期待したい。

本プロジェクトとして掲げた「我々がいかなる科学技術を手にしたいか」、「どのような人間でありたいか」という根源的な問い合わせについては、プロジェクトを通して暫定的な答えが見出されているが、当初掲げた自律性と関係性の枠組みについて改めて問い合わせ直すところにまでは至らなかった。さらに踏み込んだ考察と明確な答えを導き出していくことにも、引き続き取り組んでいただきたい。

以上

(別紙) 評価者一覧

〈プログラム総括〉

唐沢 かおり 東京大学 大学院人文社会系研究科 教授

〈プログラムアドバイザー〉

大屋 雄裕 慶應義塾大学 法学部 教授

四ノ宮 成祥 国立感染症研究所 客員研究員

中川 裕志 理化学研究所 革新知能統合研究センター  
社会における人工知能研究グループ チームリーダー

西川 信太郎 株式会社グローカリンク 取締役  
／日本たばこ産業株式会社 D-LAB シニアディレクター

納富 信留 東京大学 大学院人文社会系研究科 教授

野口 和彦 横浜国立大学 総合学術高等研究院 リスク共生社会創造センター 客員教授

原山 優子 東北大学 名誉教授

水野 祐 シティライツ法律事務所 弁護士

山口 富子 国際基督教大学 教養学部 教授

(2025年5月末時点)