

人と情報のエコシステム」研究開発領域
研究開発プロジェクト事後評価報告書

令和3年4月

研究開発プロジェクト名：自律機械と市民をつなぐ責任概念の策定
研究代表者：松浦和也（東洋大学文学部 准教授）
実施期間：2017年10月～2021年3月

A. 総合評価

成果は得られたが限定的である。

本プロジェクトでは、自律機械が出現しつつある現状についての科学技術者と市民との感覚的な受け取り方のギャップを分析しつつ、責任概念の歴史・文化相対性を倫理的観点のみならず歴史的観点・文化的観点から明らかにし、現状そして今後の状況にふさわしい責任概念の再検討を試みた。その結果、既存の人間観をダウングレードし、個人の自由意志ではなく性格に基づいて責任概念を再構築するという新たな方向性が示された。この知見は、その壮大な構想ゆえに明確な結論を出すというところまで至ってはいないものの、大きな発展の余地を残しており学術的パラダイムの転換をも促すような展開が今後期待できると考えられる。また実務的観点からは、本プロジェクトの議論をそのまま直接的に損害賠償などの形の自律機械の責任概念に結びつけることはなかなか難しいと思われるものの、本領域内の浅田プロジェクトや稲谷プロジェクトと連携しつつ議論の進め方を提示することは可能であり、今後の自律機械の開発および法制度の整備に対しても、本プロジェクトの知見は広く活用されるものと期待される。さらに、本プロジェクトの様々な活動を通じて、これまでこうしたトランスディスプリナリー研究にはあまり参加することのなかった科学技術の倫理を専門とするわけではない古代ギリシア哲学・日本倫理思想史・インド哲学・フランス哲学などの人文研究者が本領域のプラットフォームに参加し、既存のAI倫理が拠って立つ西洋近代哲学の前提を反省的に捉え新たな枠組みの視点を提示することが可能となったことは、学術的にもそしてネットワーク形成という点においても特筆に値すると考えられる。

しかしながら、本プロジェクトが示した責任概念の再構築という試みは重要な方向性であると考えられるものの、現状ではプラットフォームの形成に役立つ概念としてはまだ十分な具体性・詳細さを備えていないため、本プロジェクトの成果は今後の検証的議論を必要とする基礎的な段階に留まっていると判断せざるをえない。今後は研究から得られた成果を言語化・外化していくプロセスにさらに注力するためにも、稲谷プロジェクトでの研究の継続も含めてプロジェクトの基盤を再度検討し、他の研究資金を獲得することなどで本プロジェクトを継続・展開させていくことを期待したい。また、科学技術者や市民に本プロジェクトが主題とする基本的問題をどのように伝えるのかについては、さらなる工夫が必要だと考えられる。研究開発実施終了報告書や発表では、他で多数実践されているAI倫理と本プロジェクトの違いやなぜ人間観をダウングレードする必要があるのかなどについて平易な説明が十分なされていないため、特に哲学の専門外の評価者には本プロジェクトの主要な価値が伝わっていない可能性があると思われる。本プロジェクトが非専門家である一般市民にも違和感なく納得できる新たな責任概念を提示することを目指している以上、今後は多様な人々を巻き込むためのコミュニケーション手法をさらに洗練させていくことが期待される。さらに、上流に人文学者を入れるという研究開発体制の提案は興味深いものの、それに対応できるような人材の育成が現時点では不十分であるとも考えられるため、科学技術者や政策担当者に説得力のある説明をすることは容易ではない。それゆえ、今後はこうした研究開発体制を実験的に実施して方法論を提示するプロセスの機会があるならば積極的に参加し、成功事例を積み上げていくことが重要であると考えられる。

B. 項目評価

I. 研究開発プロジェクトの研究開発内容とその成果について

1. 目標の妥当性

妥当であったと評価する。

先端技術の進化に伴う人間社会を覆う漠然とした不安感に対して、「自律機械や AI にいかなる能力や機能が備われば、人間と対等の権利や責任を付与しうるか」という問いを立て、責任主体性や自律機械の社会性などに関して人文学の観点から検討を進め、それを市民に納得感をもって受け入れてもらうことを目標とした本プロジェクトのテーマは、科学技術政策の中に人文科学も含まれるとする改正法が通過した時勢を踏まえても、社会全体から期待されている命題であった。特に、古代ギリシア哲学・日本倫理思想史・インド哲学・フランス哲学などの専門家を集め倫理的観点のみならず歴史的観点・文化的観点から考察を深めるという方向性は、既存の AI 倫理がほぼ西洋近代哲学の前提を採用していることから独創性が高く、哲学の原理的立場と現場における議論を接続する試みとして有意義な取り組みであったと評価する。

2. 研究開発プロジェクトの運営・活動状況

ある程度適切になされた。

本プロジェクトは、文献調査の実施だけにとどまらず、情報技術研究者や浅田プロジェクト、葭田プロジェクトとの交流・哲学カフェにおける市民との対話の場の形成・書籍の発刊など、哲学の原理的立場と現場の議論の接続を試みる活動を積極的に行った。また、これまでこうしたトランスディスプリナリー研究にはあまり参加することのなかった科学技術の倫理を専門とするわけではない人文研究者を多数呼び込み、自律機械の社会的定着に向けて人文学研究の必要性を内外に示した。これらの精力的な活動により、自律機械が起こした事故の責任問題について倫理的観点のみならず歴史的観点・文化的観点から考察を深め、新たな責任概念を暫定的ではあるものの明確にしたことは高く評価できる。

しかしながら、結論として出された提案が問いに対する十分な解になりうるか、少なくともプロジェクト内で検証的議論が十分に行われたのかどうか不明であるため、提案された責任概念について第三者視点から検証するという項目は、研究期間中には必ずしも十分な活動はなされなかったと考えられる。

3. 研究開発プロジェクトの目標の達成状況および研究開発成果（アウトプット・アウトカム）

成果は得られたが限定的である。

全体を通じて、その壮大な構想ゆえに明確な結論を出すというところまで至ってはいないものの、考える問題の本質は明らかにされたと考えられる。

本プロジェクトでは、自律機械が出現しつつある現状についての科学技術者と市民との感覚的な受け取り方のギャップを冷静に分析しつつ、責任概念の歴史・文化相対性を根本的な哲学問題から明らかにし、現状そして今後の状況にふさわしい責任概念を暫定的ではあるもののある程度明確化した。この成果が社会で活用されるためには様々な観点からのさらなる検討が必要ではあるものの、個人の自由意志ではなく性格に基づく責任概念の構築という大きな流れを示したことは、学術的にも社会的にも大きなインパクトを持ちうる可能性を秘めていると評価する。このような壮大な構想が立ち上がった背景には、本プロジェクトには倫理的観点のみならず歴史的観点、文化的観点からも自律機械について検討を深めることを可能にした人文学の分厚い蓄積があったからであり、その蓄積は既存の AI 倫理が拠って立つ視点を反省的に捉え新たな枠組みを提示することを可能にしている。既存の AI 倫理がほぼ西洋近代哲学の前提を採用していることから自由意志のある個人が責任主体であるという定理から逃れることができず、また、責任を基礎付ける概念間の関係性が地域や時

代によって変化しているにも関わらず、その変化を見ることなく概念間の関係を維持した議論を続けていることなどから、自律機械の責任問題に対してこれまで適切な回答がなされていないのではないかという指摘は、人文学の議論の蓄積のある本プロジェクトの際立った独自性を示していると考えられる。また、本プロジェクトの様々な取り組みを通じて、新たな技術を解釈し市民レベルでの価値付けを果たすという新たな人文学の役割を提示し、さらには上流に人文学者を入れるという研究開発体制の提案まで至った経緯は、総合知の創出にいかにして人文学者が貢献するかという政策的な観点からも特筆に値すると考えられる。

しかしながら、責任を自由意志の観点からではなく性格と傾向の観点から再構築するという試みは重要な方向性であると考えられるものの、現状ではプラットフォームの形成に役立つ概念としてはまだ十分な具体性・詳細さを備えていないため、本プロジェクトの成果は今後の検証的議論を必要とする基礎的な段階に留まっていると判断せざるをえないと考えられる。今後は研究から得られた成果を言語化・外化していくプロセスにさらに注力するためにも、稲谷プロジェクトでの研究の継続も含めてプロジェクトの基盤を再度検討し、他の研究資金を獲得することなどで本プロジェクトを継続・展開させていくことを期待したい。また、科学技術者や市民に本プロジェクトが主題とする基本的問題をどのように伝えるのかについては、さらなる工夫が必要だと考えられる。研究開発実施終了報告書や発表では、他で多数実践されているAI倫理と本プロジェクトの違いやなぜ人間観をダウングレードする必要があるのかなどについて平易な説明が十分なされていないため、特に哲学の専門外の評価者には本プロジェクトの主要な価値が伝わっていない可能性があると思われる。本プロジェクトが非専門家である一般市民にも違和感なく納得できる新たな責任概念を提示することを目指している以上、今後は多様な人々を巻き込むためのコミュニケーション手法をさらに洗練させていくことが期待される。さらに、上流に人文学者を入れるという研究開発体制の提案は興味深いものの、それに対応できるような人材の育成が現時点では不十分であるとも考えられるため、科学技術者や政策担当者に説得力のある説明をすることは容易ではない。それゆえ、今後はこうした研究開発体制を実験的に実施して方法論を提示するプロセスの機会があるならば積極的に参加し、成功事例を積み上げていくことが重要であると考えられる。

4. 研究開発成果の活用・展開の可能性

一定の期待ができると評価する。

責任の歴史・文化相対性および自由意志に基づかない責任概念の構築は、その壮大な構想ゆえにまだまだ大きな発展の余地を残しており、学術的パラダイムの転換をも促すような展開が期待できる。また実務的観点からは、本プロジェクトの議論をそのまま直接的に損害賠償などの形の自律機械の責任概念に結びつけることはなかなか難しいと思われるものの、浅田プロジェクトや稲谷プロジェクトと連携しつつ議論の進め方を提示することは可能であり、今後の自律機械の開発および法制度の整備に対しても、本プロジェクトの知見は広く活用されるものと期待される。

しかしながら、工学者や政策担当者など哲学の専門家以外の人々の成果を活用するためには彼らがその成果を十分に理解する必要があるものの、上述したように現時点では多様な人々を巻き込むためのコミュニケーションが適切になされているとは必ずしもいえないため、今後は本プロジェクトの成果に対してさらに理解が深まるような情報発信手法を検討することが必要だと考えられる。

II. 研究開発プロジェクトの領域への貢献

研究開発プロジェクトの運営と活動、および得られた研究開発成果は領域の目標達成に大いに貢献できたと評価する。

自律機械をめぐる責任の問題は、人工知能を社会に定着させるうえで確信となる問題であり、それへの直接的貢献が本プロジェクトの実施を通じて大いになされたと評価する。また、浅田プロジェクトや葎田プロジェクトとの連携を通じて、哲学の原理的立場と現場における議論を領域内で活性化したことは、領域のプラットフォーム形成に大いに貢献したと考えられる。HITE 冊子上にて、同時期に採択された浅田プロジェクトの稲谷教授と葎田プロジェクトの葎田准教授と本プロジェク

トの松浦准教授との鼎談を実施したことをきっかけに
(<https://www.jst.go.jp/ristex/hite/topics/311.html>)、3 プロジェクトが連携して研究開発を実施するという本領域の広報上の先行事例として機能したことは特筆に値する。この鼎談時に問題提起がなされた「人工知能は近代的人間観では統制できない」という視座は、本領域が主催したシンポジウムや合宿、日英連携の募集要項などの設計にも大きく影響を与えており、これをきっかけに人工知能などの人工物のエージェンシー性を媒介として人間の在り方に関する根本課題を提起することが人と情報のエコシステムの中核的な要素となったといっても過言ではない。また、本領域は、人と情報技術が共進化していくためのプラットフォームを構築していくことを目標に掲げているが、本プロジェクトはこれまでこうしたトランスディスプリナリー研究にはあまり参加することのなかった科学技術の倫理を専門とするわけではない古代ギリシア哲学・日本倫理思想史・インド哲学・フランス哲学などの人文研究者がこのプラットフォームに参加することに直接的に貢献しており、ネットワーク形成という点においても大いに評価できると考える。

以上