

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：自律性の検討に基づくなじみ社会における人工知能の法的電子人格
研究代表者：浅田稔（大阪大学大学 特任教授）
実施期間：2017年10月～2021年3月

A. 総合評価

十分な成果が得られたと評価する。

2010年代初頭のディープラーニングの発展により人工知能技術の社会的影響力が急速に高まる中、本プロジェクトはなじみ社会の実現に必要な「自律性」を有する人工知能の性質を明らかにした上で、人工知能の開発・利用によって生じうる事故等の法的責任を適切に取り扱うための制度のあり方について工学・法学・社会心理などの立場から検討を進めた。その結果、人工知能に法人格を付与するという議論を展開することは現時点で時期尚早であるものの、統計的・確率的に振舞う人工知能はイノベーション萎縮の観点からも安全性の確保という観点からも現行法の適応では課題が残るため、人と機械の合成状態を考慮した新たな人間観に基づき制度を修正する必要があるという結論を導き出した。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新たに提起したことも含めて、刑事法の規範的基礎について反省的考察を加えたこうした知見は、国内外に類似研究が少ない学術的新規性が高いものであったと高く評価する。また、実務的観点からも本プロジェクトの知見は特筆に値する。これまで先端的人工知能技術が引き起こす事故の取り扱いに関する責任法制度について現実的な回答が国内外問わず提言されてこなかったこともあり、本プロジェクトの提言が政策立案の場における今後の議論の基盤を提供する可能性が高いと考えられる。実際に、経済産業省の「Governance Innovation」報告書でも、人工知能の特性を踏まえた責任制度の必要性を提言する、本プロジェクトにおける研究の成果が反映されている。さらに、「自律性」に関する工学的探求や人と機械のインタラクション時の社会心理的分析からも暫定的ではあるものの新たな知見が生み出され、今後の研究開発の新たなアジェンダを発掘したことは強調されるべき成果である。

本プロジェクト実施期間中に、上記の学術的・政策的活動を継続的に実施していくための体制作りとして、後継の稲谷プロジェクトが日英共同研究として開始された他、研究代表者が会長を務める日本ロボット学会においては、ロボットの法及び倫理に関する研究専門委員会を2020年4月に立ち上げ、様々に提起された議題について新たな論者を加えてさらなる議論を展開しはじめている。本プロジェクトを実施することで形成されたこうしたネットワークは、それぞれの分野において第一線で活躍するメンバーであることから、人工知能と人間が共進化する社会における現在そして今後の議論の中核を占めていくことが予想されうるし、現時点においても国内外からこの分野の代表的ステークホルダーとして見做されるようになったといえるのではないだろうか。

一方で、研究開発実施終了報告書の記載方法や発表の仕方においてはやや不十分な点があると指摘しなければならない。特に、法人格に代わる人工知能の法的責任を適切に取り扱うための制度のあり方についての方法論の具体的な説明がなされていないため、報告書からだけでは本プロジェクトの成果を十分に把握することは難しいと考えられる。経済産業省の報告書の該当部分のポイントだけでも記載する必要があったと思われる。また、それぞれのテーマの整合性が分かりにくい部分が

見受けられたが、それはプロジェクトの成果を統合的にまとめるという視点で報告書の記載や発表がなされていない部分があったのではないかと推測される。

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

十分妥当であったと評価する。

2010年代初頭のディープラーニングの発展により人工知能技術の社会的影響力が急速に高まる中、人工知能の開発・利用によって生じる事故等の法的責任を適切に取り扱うための制度のあり方について、人工知能に法人格を付与するという方法論を含め工学・法学・社会心理などの立場から明らかにするという本プロジェクトのテーマは、社会全体から期待されている命題であった。また、なじみ社会の実現に必要な「自律性」を有する人工知能の性質を明らかにし技術側にフィードバックするというテーマは、シンギュラリティという言葉が流布していた研究期間中において誰かが取り組まなければならなかった本質的かつ不可避的な議論であり、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新たに提起したことも含めて、本プロジェクトから得られた知見は本領域の中核的な要素となりえる。それゆえ、人工知能が浸透する社会において社会的期待の高いテーマに正面から挑んだ本プロジェクトの目標設定は十分妥当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

十分適切になされたと評価する。

ロボット工学・社会心理学・法学・哲学・脳神経科学という実に多様な分野の専門家との共同研究を実施し、有意義な成果を創出するために適切な実施体制での運営がなされたと評価できる。途中でプロジェクトのメンバーの交代もあったものの他グループに吸収し実施、また日本科学未来館の公募にも積極的にトライ、本領域内の松浦プロジェクトや葭田プロジェクトとの連携など状況に応じて臨機応変に体制を変えるダイナミックなプロジェクト運営であった。各チームの部分目標がそれぞれ適切なメンバーで実行され、ロボットへの責任帰属に関する意識調査においては大規模な市民参加を実施するなど関与者の巻き込みも適切になされたため、現場で多くの情報やデータが収集・検討された。さらには統合的な議論の場として、ロボット学会特集号の刊行やロボット学会におけるロボットの法及び倫理に関する研究専門委員会の発足、また関連する国際シンポジウムの開催などを設定し、成果を積極的に発信した。

それぞれのテーマの整合性がやや弱い部分もあったものの、これだけの多様なディシプリンをまとめあげることが容易ではないと考えられる中、経験豊富な代表者の下で各々の強みを発揮しながら研究することを可能にした今回の運営スタイルは、今後のトランスディシプリナリー研究の模範となりうると考えられる。

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）

一定の成果が得られた。

全体を通じて、中長期の本質的論点と現実における対応の両面から検討を進めることができた。問題の難しさのために、全ての項目において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確化されたと考えられる。

本プロジェクトの様々な取り組みにより明らかになったことは、現時点において「自律性」をもった人工知能・ロボットは作られておらず且つ人工知能に事故の責任を負わせることは現状では市民感情的に受け入れられていない傾向があるがゆえに、人工知能に法人格を付与するという議論を展開することは将来的にはありえても現時点で時期尚早であるものの、統計的・確率的に振舞う人

工知能は現行の法制度内で処理することが難しいため、新たな人間観に基づき現行の法制度を修正する必要があるということであった。こうした結論は、現状において学術的にも政策的にも重要で新規性の高い知見であり、人工知能が浸透する今後の社会においても極めて有用性の高い研究を実施したと評価する。また、本プロジェクトで構築した国内外のネットワークは、それぞれの分野において第一線で活躍するメンバーであることから、人工知能と人間が共進化する社会における現在そして今後の議論の中核を占めていくことが予想される。

個々の項目の成果を記すと、成果①「人工知能の自律性検討に基づく自律モデルの構築」では、人工知能の自律性に関する痛覚を出発点とする概念モデルを構築した。そのモデルより、人工知能には自律性・共感能力・道徳的行動の学習などが備えられる必要があるものの、人工知能がそれを獲得するためには脳や身体的设计において現時点での技術レベルからのさらなるブレークスルーが必要との結論を導き出しており、今後の人工知能開発において注力すべき方向性を提示した。成果②「人工知能の責任の心理学的・社会的認識の調査」では、①で提案されたモデルをアルゴリズム化し実装したロボットと人とのインタラクション時における責任の帰属をゲーム化した実験を実施した。その結果、ロボットに責任が帰属すると思う程度はロボットの感情的能力のスコアと正の相関があるとの暫定的な結論が導き出され、ロボット設計者が配慮すべき点を提示した。また、人とロボットとの共同作業において生じた損害に対しての責任の帰属について、製作した動画をもとに評価するアンケート調査も実施した。その結果、ロボットを法的主体として責任を問うことは一般市民感覚として許容されにくい可能性や事故を起こしたメーカーに対して刑罰を与えるべきではないと思われる傾向があるという暫定的な結論が導き出され、事故時の責任主体における一般市民の感覚を可視化したという点で、今後の研究につながる重要な成果を創出した。成果③「既存の法的責任制度の限界と対応策の明確化」では、憲法・ローマ法といった人格性について専門知識をもった専門家も交えて検討した結果、現時点では人工知能に法人格を付与することはデメリットの方が大きいと判断した。一方で、現行の法制度の下で統計的・確率的に振舞う人工知能を活用しようとすると、イノベーション萎縮の観点からも安全性の確保という観点からも問題が残るため、新たな人間観に基づいた現行の法制度を修正する必要性が提言された。先端的人工知能技術が引き起こす事故の取り扱いに関する責任法制度については、現実的な回答がこれまで提言されてこなかったこともあり、本プロジェクトの提言が政策立案の場における議論の基盤を提供する可能性が高く、実際に経済産業省の「Governance Innovation」報告書でも、人工知能の特性を踏まえた責任制度の必要性が提言さえるなど、その知見が反映された。

上記のように本プロジェクトでは実に多彩な成果が創出され、その一部は既に現実の政策にも取り入れられ始めており、アウトプット・アウトカム両面からも大きな成果を創出したと評価できる一方で、研究開発実施終了報告書の記載方法や発表の仕方においてやや不十分な点があると指摘しなければならない。特に、法人格に代わる人工知能の法的責任を適切に取り扱うための制度のあり方についての方法論の具体的な説明がなされていないため、報告書からだけでは本プロジェクトの成果を十分に把握することは難しいと考えられる。経済産業省の報告書の該当部分のポイントだけでも記載する必要があったと思われる。また、既に指摘したようにそれぞれのテーマの整合性が分かりにくい部分が見受けられたが、それはプロジェクトの成果を統合的にまとめるという視点で報告書の記載や発表がなされていない部分があったのではないかと推測される。

4. 研究開発成果の活用・展開の可能性

大いに期待ができると評価する。

本プロジェクトから創出された成果は、社会的ニーズの高い分野でありながらも新規性が高く国内外に類似研究が少ないが故に様々に展開が可能であるし、実際に次の取り組みに向けて既に布石は打たれている。

2020年1月に発足した稲谷プロジェクトは、本プロジェクト及びHITE内松浦プロジェクトの後継プロジェクトとして位置付けられるものとして発足し、英国と緊密に連携することによって本プロジェクトの成果を国際的に展開しうるポテンシャルを持ちあわせている。研究代表者が会長を務め

る日本ロボット学会においては、ロボットの法及び倫理に関する研究専門委員会が 2020 年 4 月に立ち上がり、本プロジェクトにおいて提起された様々な議題についてさらなる議論が展開されている。本プロジェクトを実施することで形成されたこれらのネットワークは、国内外からこの分野の代表的ステークホルダーとして見做されるようになったといえるのではないだろうか。また、報告書に記載されている「人と機械の協調動作時の責任配分」と「人と機械の合成状態」というアジェンダは、本プロジェクトを実施することで見えてきた新規性の高いものであり、今後学術的に探求する価値が高いと考えられる。さらに、本プロジェクトで導き出した最先端人工知能技術に関する現行法制度の限界の指摘と今後の方向性の提案は、現時点でも経済産業省の委員会が継続されていることなどからも、政策立案の現場において現在・将来的な議論の基盤を提供する可能性が高く、実務的な観点からも本プロジェクトの成果の展開性への期待は高いと考えられる。

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。

機械の自律性をめぐる技術的・法的・道徳的問題の解明は、本領域の中心的課題であり、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。この議論を通じて、現時点での人工知能の可能性と限界を現実的に見極めることが可能となり、それに即した法制度の提案が実際に政策の現場に取り入れられるなど、社会実装という観点からの貢献も大きいと考えられる。

また、HITE 冊子上にて、同時期に採択された松浦プロジェクトの松浦准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稲谷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 HITE 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ト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。

さらに、本プロジェクトを実施することで獲得した国内外のネットワークは、人工知能と人間をめぐる議論を展開していく上で今後も中心的役割を担っていくことが想定されるため、本プロジェクト終了後も本領域と継続的に関係しながら、領域全体に対して議論の場や研究者間の連携の場の設定、国内外への知見の発信という機能を提供していくことを期待する。

以上