様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　１５　日現在

機関番号：　１１３０１
研究種目：　基盤研究(C)（一般）
研究期間：　2015～2017
課題番号：　１５Ｋ０１９７８

研究課題名（和文）シンギュラリティと責任の論理

研究課題名（英文）Singularity and logic of responsibility

研究代表者
村上　祐子（Murakami, Yuko）

東北大学・文学研究科・准教授

研究者番号：８０４３５５０２

交付決定額（研究期間全体）：（直接経費）　　3,700,000円

研究成果の概要（和文）：課題期間中に当初想定の人工知能が人類を滅ぼすというシンギュラリティ論の指摘するフェーズは過ぎたと判断し、電子的法人格、人間性、責任など哲学的概念の変容に注目して分析を進めており、欧州において実務提案が行われたことから方向性は正当化された。論理的記述については今後も進展させたい。学術・実務・教育というそれぞれの現場における人工知能の利用について調査した。人工知能の学術利用の可能性と研究者の労働代替のリスクについて議論を行った．教育現場における人工知能利用に関して現場のニーズについてケーススタディを行ったが、道徳的ジレンマの詳細分析は今後の課題である。

研究成果の概要（英文）："The singularity argument," or the alleged claim that AI will annihilate human beings, was viral at the moment of the project proposal, although it has little evidence via immature argument. We pointed out the gaps and also necessities of revising the fundamental notions of personality, humanity, and responsibility. Our claims are getting vindicated via proposals by German government and EU congress toward considering e-personality.
Logical investigations are still on the way and we wish to develop the theory more deeply in the future.
We also focus on much details of academic usage of AI in research and K-12 education. In academic research substitution of human beings with AI will rather help advanced research by human. In K-12, more detailed analysis of contents of moral dilemmas should be arranged for each cluster of learners. We hope to classify cases of moral dilemmas for educational levels in the future.

研究分野：　哲学

キーワード：　情報哲学　情報教育　情報倫理

１．研究開始当初の背景

自律的AIの能力が人間を上回るとともにロボット技術・ナノ技術・生体素子によるエンハンスメントによる因果系列への介入が行われるという「シンギュラリティ」現象がビッグデータと高速計算により現実となる可能性が語られるようになった。当研究を始めるにあたって、実はこのような「シンギュラリティ論」には概念仕様の混乱が含まれており、議論を整理することで真の問題の所在を解明することが可能になると考えた。

２．研究の目的

当プロジェクトでは、論理学と STS の観点から**強い AI の可能性**を改めて検討するとともに、仮に実現した場合に発生が不可避である**「自由意思」「人格の同一性」「プライバシー」**といった哲学的概念の変容について検討する。また、論理的にこれらの問題を解決する可能性についても議論し、AI への実装を許す**論理体系構築を**めざした。

３．研究の方法

研究動向調査は文献調査と人工知能研究者および情報教育実務担当者からの聞き取り調査を中心に進めた。
また、論理体系構築については個別に開発を進めた。

文献調査：行為論理の各理論の展開と AI への実装状況を調べたが、国内外の動向として実質的進展がなかった。
また、道徳判断 AI が実装されたロボットの理論的・技術的基盤と応用についての文献調査を継続的に行った。

聞き取り調査：アメリカ・レンスラー工科大学・Bringsjord 教授を含む人工知能研究者との対話を通して、人工知能研究者の関心及び哲学的問題の理解度を明らかにする方法をとった。また、情報教育実務担当者からのききとりでは、一般市民、とりわけ年少者とその保護者がコンピュータについてどのようなたいどを取りがちであるのか、さまざまなケースから示唆を得た。

４．研究成果

初年度ではさまざまなシンギュラリティ論を分類した上で、「シンギュラリティはすでに到来している」と結論づけた。ここで到来済としたシンギュラリティは局所シンギュラリティであり、特定の分野について既存の方法論（統計的手法を含む）でプログラム可能なタイプのシンギュラリティである。

たとえば囲碁将棋を含むゲームについては、すでにシンギュラリティは到来していると考えることができる。

	ユビキタス、自律	ユビキタス、人間補助	局所
汎用 AI	強いシンギュラリティ：人間殲滅	強いシンギュラリティ：人間の奴隷化	
目的別 AI			ゲーム AI
AI 制御なし	（自然）	機械的制御による機械	工業ロボ

表 1：シンギュラリティの分類

2 年目にはシンギュラリティ問題はシステム側の問題ではなく、システムの社会受容にあるという仮説のもと、局所シンギュラリティ状況に直面した人間のコミュニティの対応手法や次世代の育成における計算機の利用についての考察を進めた。

この際に判明したのは、道徳的人工知能開発に向けて用いられるビッグデータ＋帰納的・統計的手法には問題があり、道徳概念を表現する語彙が追加された演繹的論理体系を用いる必要性だった。現実が道徳的理想世界ではない以上、人間の挙動・判断を範として帰納的推論を行っても道徳的人工知能に到達することは不可能であるからである。

とくに問題となるのは異なる価値観が衝突するケースであるため、2017 年度は道徳的ジレンマを多層価値観論理モデルによる分析をすすめ、価値観そのものに半順序関係を導入したモデルでは道徳的ジレンマを記述できることをしめした。さらに具体的な道徳的ジレンマの教育方法の検討を進め、情報倫理教材へ間接的に反映させていった。

同時に最終年度では社会的にもシンギュラリティ論の問題が受容される方向に進んだため、研究成果としてシンギュラリティ論そのものの問題点を指摘するフェーズは過ぎたと判断することとなった。実際、シンギュラリティという事態が実現するかどうかが問題であるにもかかわらず、人工知能の脅威を強調することで、脅迫ビジネスを進めていたという妥当な指摘が各組織に向けられるに至っている。そのうえで、人工知能の脅威が何らかの形で存在するにしても、社会的にコントロールする方法を 探るべきだという論調が主流となり、当研究課題の社会的方向性は正当化されたとみなせる。
この状況を踏まえ、さらに議論を精密化するために、学術・実務・教育というそれぞれの現場における人工知能の利用についての議論を進めた。学術の現場に関しては、

数学・天文学・地球惑星科学の研究者と人工 知能の学術利用の可能性と研究者の労働代替のリスクについて議論を行った。実務では、法人格を人工知能に付与する社会的要請に対し、責任分担が十全に行えるのか検討を加えた。さらにトロッコ問題が代表的な例と挙げられる道徳的ジレンマの議論は二つの選択肢に限るという点で問題設定そのものに問題があり、選択肢および損害・利得についてさらに精緻な設定を行うべきだという結論に達した。

教育現場における人工知能及び情報システムの利用に関しては、研究・開発 の方向性と現場のニーズのずれについてケーススタディを行った。成果発表については、2018年 9 月の ETHICOMP2018 での講演（採択済）を予定している。

また、道徳判断 AI 実装の文献調査及び聞き取り調査の結果、研究者のコミュニティにおいても道徳判断 AI への過信が観察された。この点に関してはプロジェクト終了後も継続して中止するとともに問題点の主張と解決案の提示を行っていく必要がある。とりわけ、人工知能による犯罪予測システムがすでに導入されている現実に加えて、犯罪性が予測される時点での逮捕が法制化されようとしており、人工知能を含む社会システムに道徳的判断、とりわけ価値観の変更・過誤による冤罪への説明責任を実装する方法論を構築していかなければならない。

いっぽうで、人格概念の変容については一方で人工知能単体が法人格付与等の社会合意の上で問題発生時に責任を負う可能性を開くべきであるという主張を研究代表者は展開することとなった。この意見は、情報の哲学者や法哲学者の一部と一致する一方で、法曹実務者や法哲学者の主流には強い反対意見が見られる。後者の意見が主流である内は社会実装にはいたらないという観察を得た。現実論においては、人間の個別の活動には人間の複数の機能が含まれており、人工知能は個別の機能の代替を進めていくのであって、人間総体としての置き換えは開発の方向性として誤っていることとなる。

だが、ドイツ運輸省及び EU 会議で、電子的法人格の導入検討が提案されるにいたり、結果として研究代表者の提案の方向性の正しさが示された。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 11 件）

(1) 村上祐子、人工知能の倫理の現在--研究開発における技術哲学・倫理の意義 電気情報通信学会 基礎・境界ソサイエティ Fundamental Review, pp.155-163, 2018. 査読あり

(2) Yuko Murakami, Philosophy and Higher Education in Japan. *Philosophy*, 1、pp.184-196. 2017. 査読あり

(3) Murakami, Y., Tatsumi, T., Otani, T., & Harada, Y. Ethics of Information Education for Living with Robots.ORBIT Journal, 2017. https://doi.org/10.29297/orbit.v1i1.21 査読あり

(4) 塩野康徳, 辰己丈夫, 西村佳隆, 徐浩源, 田名部元成「大学における２１世紀型情報リテラシー教育デザインのための実態調査」コンピュータと教育研究会研究報告 14、pp.1-8. 2016. 査読あり

(5) 辰己丈夫、村上祐子、大谷卓志「未来の情報倫理教育」情報教育シンポジウム 2015 論文集 pp.45-52. 2015. 査読あり

〔学会発表〕（計 34 件）

(1) Yuko Murakami, Current situations in Japan under privacy concerns on household robots. ETHICOMP2018 （採択済）

(2) 阿部敬一郎、辰己 丈夫、村上 祐子、中谷 多哉子「道徳教育支援システム化に向けたモラルジレンマの試行実験」電気情報通信学会知能ソフトウェア工学研究会、2017.

(3) 村上祐子、アルゴリズム的偏見：行為論理によるデータ選別の可能性と限界 東北大学材料科学研究所数学連携グループセミナー（招待講演）2017.

(4) 村上祐子「人工知能と共存する人間」人工知能学会（招待講演）2016.

(5) Yuko Murakami, Cultural Impacts on/of Artificial intelligence. World Humanity Forum, （招待講演） 2016.

(6) 村上祐子「強いシンギュラリティ、弱いシンギュラリティ」電気情報通信学会 SITE 研究会、2016.

〔図書〕（計 6 件）

(1) 山田恒夫、辰己丈夫「情報セキュリティと情報倫理」放送大学教育振興会、2018 総ページ数 259 ページ。

６．研究組織

(1)研究代表者 村上祐子（MURAKAMI, Yuko）

(東北大学・大学院文学研究科・准教授)
研究者番号：80435502

(2)研究分担者　辰己丈夫
（TATSUMI, Takeo）（放送大学・教養学部・教授）研究者番号： 70257195

(3)連携研究者　なし

(4)研究協力者　大谷卓志（OTANI, Takushi　）
(吉備国際大学・アニメーション文化学部・教授)研究者番号：50389003