

課題設定による先導的人文学・社会科学研究推進事業
(実社会対応プログラム)

研 究 成 果 報 告 書

「工学・脳科学をエビデンスとした社会的基盤概念と価値の創生」

研究代表者： 松浦和也

(東洋大学 文学部 教授)

研 究 期 間： 平成30年度～令和3年度

1. 研究基本情報

課題名	世代間衡平性・持続可能性・社会安全性等の倫理的観点を考慮した政策設計のための実践的研究
研究テーマ名	工学・脳科学をエビデンスとした社会的基盤概念と価値の創生
責任機関名	東洋大学
研究代表者(氏名・所属・職)	松浦和也・文学部・教授
研究期間	平成30年度 ～ 令和3年度
委託費	平成30年度 2,925,000円
	令和元年度 4,270,500円
	令和2年度 4,270,500円
	令和3年度 2,135,000円

2. 研究の目的

世代間衡平性や持続可能性が現代の社会の問題として表面化した理由は多数挙げうるが、社会思想の観点から見れば、現行の社会制度の基盤となっている近代的「主体」概念の限界にあるように思われる。すなわち、近代的主体は合理的・理性的に振る舞う個人を指し、それを社会の最小単位の構成要素と見なす。だが、このような概念からは、次世代の人間たちは、未だ存在していないがゆえに、社会の構成要素とは見なされない。また、動物や植物、自然環境も、人間社会の維持存立の基盤のひとつでありながら、人間ではないがゆえに社会の外に置かれる。それゆえ、世代間衡平性を有する持続可能な社会を目指すのであれば、この種の主体概念を解体するか、主体概念以外の要素を社会思想の基盤に付加するといった理論的作業が必要となる。他方、科学技術の進展がもたらす脳科学や人工知能はこのような近代的主体概念に風穴を開ける。つまり、近代的主体概念が人間に期待する自律性について脳科学は懐疑を投げかけるし、合理的・理性的という観点から見れば、人工知能はそのような要素を正に有するものと考えられうるからである。そこで、現状の打開のためには、脳科学や人工知能がもたらす諸問題を目に見える事例として、近代的主体概念を中心に人間観や社会観の妥当性を原理的に検討しつつ、世代間衡平性や持続可能社会を実現するために必要な社会基盤概念を法曹界の実務家とも協議することで、実践性を伴うものとして社会に浮き彫りにするという手順が考えられる。

本研究においては具体的には次のような事例を扱う。レベル3の自動運転車やウェアラブルロボットは、通常は人間と協調しているように見えるが、事故時には使用者やメーカーの意思や意図とは異なる形で動作する。このような製品が起こしうる事故に関する考察は、国内外で盛んになってきている。ただし、その主要な関心事は、事件に関する責任の帰属先を被害者、使用者、製作者にどの程度分配するか、そしてその分配のための手続き—たとえば、法的承認を使用者に課す等—とはいかなるものか、という点に終始してしまっている。本研究はこの関心を一步踏み超える。すなわち、この種の議論の妥当性を一定程度認めつつも、それらの議論が前提している近代的主体およびその周辺にある政治学的・倫理学的概念に法学的議論や思想史的背景から切り込み、他方で上記に接したときに生じる人間の心理学的事実を基礎付けのひとつとすることで、近代的主体概念をはじめとした現行の社会の基盤となっている社会的概念とは異なる選択肢や、現在には忘れ去られてしまった社会的基盤概念や価値観の可能性を探り、それらの基盤概念を社会に導入するための方策を法学・実務的視点から検証することで、次世代に適した社会のあり方そのものを再編することを目指すものである。

3. 研究の概要

工学グループ内ロボット工学チームは、人間と協調するようなロボットの工学的設計の理念とその設計が人間にもたらす影響を分析する。すなわち、実機のコミュニケーションロボットを活用したヒューマンロボットインタラクション実験を通じ、一般市民が思うロボットに対する責任意識や法意識を調査する。たとえば、人と自律ロボットが共同で意思決定する状況を設定し、そのときに人が感じる行為主体性や、望まない結果に至ったときの責任主体性、人工物に対する倫理観・価値観を析出する。その際、特に若い頃から情報技術に親しんでいる若い世代の自立システムに対する責任意識に注目し、工学の視点からの今後の自律システムの発展性を考慮する。また、工学グループ内脳・認知科学チームは人間が主体的に振る舞う状態を脳科学的観点から分析することで、人間にとって主体性とはどのような状態に定位でき

るかを明らかにする。すなわち、ロボット工学チームと共同しながら、ヒトと協調動作する見かけ上自律的なシステム、例えば、半自動運転システムやウェアラブルパワーサポートロボット、バーチャルリアリティの義体を想定し、これらがヒトと一緒に特定の行為や行動を実施する際に、その行動がヒトからみて自分自身が引き起こしたものであり(行為主体性)、ロボットや人工知能側に強要されていないと感じる心理的・物理的要件(例えばエージェンシー知覚や時間応答特性)の許容範囲を心理物理学の測定方法及びオンラインアンケート調査を用いて検討する。これら一連の研究により、ヒトが自分自身の身体行動や思考を「自分自身が主体となって行った」と(時に誤って)見なし、責任を帰属する過程を推定する。同時に、協調動作しているシステム側の意思や意図がヒト質を、人間の発達段階によって生じる脳作用の差違にも考慮しながら、脳科学的に浮き彫りにする。具体的には、ヒトの行為主体感覚に関連する研究では、他者ないし非自己が影響する場合の行為や責任の帰属過程に関する先行研究調査とそれに基づく心理物理学の実験を考案し、人、ロボット、コンピュータとの対戦による結果の行為主体感覚を調査した。ヒューマンロボットインタラクションにおける責任意識と法意識の調査に関して、脳・心理物理実験を考案したが、新型コロナの影響で実施を諦めざるを得ず、オンラインアンケート調査に変更し、自律システムに対する責任意識と法意識の世代間比較を実施した。

哲学グループは本プロジェクトの全体に関わる哲学的考察を行った。具体的には、分担者が専門とする領域における思想モデルを明確にしたうえで、他のグループから提示される社会制度と価値観に関して原理的考察を行う。その中で、1)脳科学・人工知能の社会実装が生じさせる社会的諸問題の把握と近代的主体概念の関連性の分析、2)近代的主体概念の妥当性の検討、3)「責任」概念の理解および使用法の差異の明確化、4)近代的主体以外に社会を構成する要素の明確化と、脳科学や人工知能との関係の考察、といった課題を遂行した。また、法学グループの課題とも接点があるAI・ロボット倫理の調査と検討は、該当領域の本邦における調査状況の遅れを取り戻すだけではなく、欧米圏の考察が無批判に前提している社会観や世界観をあぶり出すことにも寄与した。

法学グループは、1)工学グループ及び哲学グループの研究成果に基づき、既存の法理論の基盤に関する修正を行った。すなわち、2)既存の法学が基盤としてきた近代的主体概念及び自由意志を基礎とする責任理論について、より関係的な主体概念に基づく全体論的考察に基づく責任分配という観点から批判的検討を行った。その結果として、3)複雑なシステムが引き起こすリスクをマネジメントするために必要とされる全体論的な方法論を採用するための、自然科学的・倫理学的基礎を明らかにした上で、4)応答的規制や修復的司法といったより未来志向的・全体論的な司法制度を基礎とする法制度の実装に向けた理論的基盤の整備と、5)具体的な制度設計のあり方に関する提案を行った。

4. 研究プロジェクトの体制

研究代表者等の別	氏名	所属機関・部局・職名	研究項目
研究代表者／哲学グループリーダー	松浦 和也	東洋大学・文学部・教授	研究の統括、古代ギリシア・ローマの政治哲学における主体概念の彫琢
分担者(哲学)	野村 智清	秀明大学・学校教師学部・講師	近代イングランドにおける主体概念の再検討
分担者(哲学)	相松 慎也	東京大学・大学院人文社会系研究科・助教	近代イングランドにおける主体概念の再検討
分担者(哲学)	山田 有希子	宇都宮大学・教育学部・准教授	近代ドイツにおける主体概念の再検討
分担者(哲学)	岡田 大助	江戸川大学・基礎・教養教育センター・准教授	日本思想における主体について
工学グループリーダー	河合 祐司	大阪大学・先導的学際研究機構附属共生知能システム研究センター・特任准教授	工学グループの研究計画全体の遂行。ロボット工学チームの総括。ヒューマンロボットインタラクション実験の実施と人工物に対する責任意識と法意識の検討
分担者(工学)	高橋 英之	大阪大学・大学院基礎工学研究科・特任講師	ヒューマンロボットインタラクションにおける責任意識の認知科学的検

分担者(工学)	島谷 二郎	大阪大学・大学院基礎工学研究科・特別研究員(DC2)	討 ヒューマンロボットインタラクションにおける自律的に振る舞うロボットの技術的検討
分担者(工学)	守田 知代	国立研究開発法人情報通信研究機構 未来ICT研究所 脳情報通信融合研究センター	脳・認知科学チームの総括。人間の発達段階を考慮した主観的主体と責任感に関する脳科学的検証と心理物理学実験を伴う調査
分担者(工学)	紀ノ定 保礼	静岡理科大学・講師	人間の主観的主体と責任感に関する認知科学知見の検証と調査
法学グループリーダー	稲谷 龍彦	京都大学・大学院法学研究科・教授	理論刑法の観点からの現行刑事司法制度の下における過失犯処罰の射程の検討と、その理論的問題点の明確化
分担者(法学)	坂下 陽輔	東北大学・大学院法学研究科・准教授	理論刑法の観点からの現行刑事司法制度の下における過失犯処罰の射程の検討と、その理論的問題点の明確化
分担者(法学)	福岡 真之介	西村あさひ法律事務所・パートナー・弁護士	法学者の理論的分析に対する、人工知能と法に関する法律実務の観点からの批判的検討
分担者(法学)	平尾 寛	西村あさひ法律事務所・パートナー・弁護士	理論的分析に対する、企業犯罪に関する法律実務からの批判的検討

5. 研究成果及びそれがもたらす波及効果

【工学グループ】

対人の帰属理論調査結果: 社会心理学において「帰属」とは、ある状況で生じた事象(出来事や行為)の原因を推論する心理過程であり[Heider, 1958]、ある人の行為の原因はその行為者の性格や意図(内的帰属)とその行為者のおかれていた環境や運(外的帰属)に分けられ、判断者はその行為を内的にまたは外的に帰属する。そして、多くの場合、この原因帰属に基づいて、その行為の結果に対する行為者の責任の大きさが判断される。責任帰属には、行為と結果の因果性だけでなく、行為者の意図性や予見可能性(行為に先立って、行為者が結果を知り得たかどうか)が考慮される[唐沢、松村、奥田、2018]。すなわち、行為者の意図しなかった行為や、結果を予見できなかった行為の結果については行為者は免責される。これに加え、行為者がその原因をコントロールできた(統御可能だった)場合に行為者の責任が大きく帰属される[Weiner, 1995]。これらの因果性に基づく責任帰属に加え、役割や地位に応じた期待や規範からの逸脱(すべきことをしなかった、または、すべきでないことをした)に基づく責任帰属もある[Hamilton, 1978]。本来、因果関係は曖昧であり、帰属には様々なバイアスがかかる。

ロボットへの責任帰属調査結果: Hinds *et al.* [2014] は、対ロボットの非難／賞賛は対人のものと同程度であることを実験により示した。課題の失敗における人らしいロボットへの非難が、機械らしいロボットへの非難に比べて有意に大きいことも示された。Moon and Nass [1998]は、成功は自分のおかげで失敗は自分以外のせいという自己奉仕バイアスが人工物に対しても存在することを実験で示した。ただし、参加者とコンピュータの性格傾向が類似している場合にはこのバイアスが緩和されることも示したが、これは対人でも観察される[Shaver, 1970]。Kim and Hinds [2006]は、自律性(自律的だと知覚されること)の程度が大きいことで、課題の失敗時に人工物に帰属される責任が大きくなることを示している。ロボットへの責任帰属におけるエージェンシーの影響に踏み込んだ実験として、van der Woerd and Haselager [2019] は、ロボット単体が課題を達成するだけの能力を有していながら、その課題に失敗する事態を設定し、それを観察した実験参加者は、その失敗の原因が能力不足ではなく、努力不足であると思われるロボットに、失敗の原因が能力不足である

と思われるロボットよりも、大きなエージェンシー（自己制御や意思決定の能力）と大きな責任を帰属することを明らかにした。この結果は、その原因が統御可能である場合に大きな責任が帰属されるというWeiner [1995] の対人の帰属モデルと一致する。

心の知覚と原因帰属: 上記のように、人工物に対する原因と責任帰属の多くは、対人の帰属理論で解釈可能であるといえる。ただし、人工物においては、その見かけの人らしさや知覚された自律性が重要な要因になる。その理由として、人工物への原因や責任の帰属には、判断者がエージェンシーといった人のような心や機能を推論する、いわゆる擬人化の過程があることが考えられる。そこで、本研究項目では、擬人化の種類と程度を定量化する「心の知覚尺度」を用いて、人工物への心の知覚と原因帰属の関係を調査した。まとめると以下になる。

- I. ゲームを通してエージェンシーのスコアが低下したエージェントには、大きく原因帰属。
- II. 能力についての参加者の事前の期待に反することが、原因帰属。
- III. 人は相手のことを考慮して、「相手に譲る」選択をすべきという道徳的な規範から逸脱したことが大きな帰属。
- IV. 人工物への期待や人工物に適用される規範が、対人のものと異なる可能性。

ロボットの事件事例映像を用いた責任帰属の調査: ロボットの故障や明らかな欠陥ではなく、社会的な相互作用の結果として発生したロボットの事故における責任帰属について、市民の意見を集めるためのワークショップとオンラインアンケートを実施した。別プロジェクトで作成したロボットによる事故の映像を題材とし、老人が発作を起こしたときに、その老人を助けるために住宅管理ロボットが見知らぬ他人（実は泥棒）を家に入れてしまい、泥棒の被害にあった時の責任問題を問うた。「ロボットに責任がないと思う理由」については市民から「ロボットはプログラムどおり人命を優先しただけだから」、「もしロボットが人であつたら、人道的な判断だったと思うから」、「ロボットに責任能力はないから」などのコメントが得られた。世代間では、そうは思わないという傾向が60代に強く、20代で弱かった。

【哲学グループ】

総論: 他グループとの研究活動を通じて明らかになったことは、当初想定していた以上に欧米圏における社会制度や法律の作成、運営、解釈において哲学的言説が役割を演じつづけている事実であった。たとえば、近代的主体を概念的に構成する理性や自律性といった発想は、AIやロボット等の人工物の倫理や法を扱うときにそのまま転用され、人工物が道徳的主体になりうるか否か、人工物に権利を与えるか否か、といった問題を論ずる際に導入される。しかし、哲学グループが見出した問題は、この種の議論に應用された哲学的言説や概念は断片的で、形骸的なものになりがちだということである。つまり、ある概念（たとえば自律性）は別の概念（たとえば責任）と接続されることがあるが、この接続が真に有効かつ有意義なのは、ある特定の思想体系（たとえばカント認識論）にコミットした場合に限られるはずである。だが、そのような思想体系はしばしば無視されるだけではなく、その議論体系や言語の用法から遊離した定義をその概念に与えた上で、接続関係だけが残される。このような形で構成されてしまった概念の接続関係は、その接続関係を提示した先行する哲学的言説を権威として用いながらも、その言説の内実からは乖離し、説明能力を欠くものとなる。人類による言説や思想の伝承の中で、このような事態がどの程度生じているかを計測することは難しいが、もし世代間衡平性をはじめとした倫理的・社会的問題を解決し、持続可能社会を実現するためには多くの市民による合意形成が求められるのであれば、形骸化した議論ではなく、内実と実感を伴う議論を形成することが求められる。

脳科学・人工知能の社会実装が生じさせる社会的諸問題の把握と近代的主体概念の関連性の分析: 工学的成果から社会的問題が生じる理由は多々あるが、概念使用と相互理解という側面に着目した場合、コミュニケーションの中で用いられる社会基盤概念に齟齬があることが挙げられる。主体概念ひとつをとってみても、日本思想の中ですら時代や思想的背景によって多様な内容を持ち、その多様性は明治維新期における西洋語の漢語訳を通じてさらに深化してしまった。近代的主体概念の中心である「理性」も同様である。理性を持つものを社会構成員の最小単位とするという発想は、第一に「理性」の理解が曖昧であるがゆえの混乱と、第二に「理性」の概念が希薄な文化圏における社会思想の排除という傾向を生む。たとえば、古代ギリシア的に理性を専ら計算能力ととらえた場合、計算能力において卓越している計算機はより人間であり、より主体として相応しいという発想を導出する。このような発想は、計算能力を人間の能力として重視しない文化圏から見れば、一笑に付されるだろう。

近代的主体概念の妥当性の検討: 近代的主体概念には様々な理解がありうるものの、およそ最大公約数的なところを挙げるならば、人間の中心は理性であり、理性に基づき行為選択をすること、その理性に自然権が紐づけられていること、そして社会を構成するものが人間である限り、社会の構成要素の最小単位はひとりひとりの理性であることである。このうち、理性を重視するという点は、歴史的には古くギリシア哲学に立ち戻れるものの、その時代の人間観においては必ずしも人間の本性を理性とは見なしてはいない。また、日本の思想史においては、明治時代以前、武士の独立の精神など、主体に近い考え方はあったものの、上述したような意味での主体概念は存在せず、明治時代以降に西洋から輸入した後は、大きな影響を受けたものの、和辻哲郎の共同主観的な主体論など、独自の変容を遂げた。このように、近代的主体概念は文化的で歴史的に形成さに応じて変容させる、あるいは却下する余地を十分に残すものであった。他方、その妥当性については、工学グループとの意見交換から疑義が呈された。たしかに、原因となる事象や状態と、結果となる事象や状態は因果関係で結ばれる。しかし、原因が人間内部の「自由意志」に求められることはない。というのは、工学において物理的因果関係とは無関係な「自由意志」を作成することは不可能だからである。このような事態が特に問題となるのは、これからの社会システムにおいて近代的主体概念の有効性が薄まる点にある。たとえば、近代的主体が先天的に有するとされる自然権、およびそれを基盤として構成される社会システムは、実際のところ様々なところで軋轢を起こしうる。「所有権」をとっても、ほとんどの倫理思想も宗教的戒律もそれを無制限に認めるようなことはなく、むしろ制限、ともすると放棄することを促すように働く。同様に、発達心理学で扱われる「所有意識」の発達とは、近代的主体に認められるような所有概念を適切に理解し、実践することではなく、むしろ他者がものを所有することを容認し、さらには自己と他者がものを共有することを目指すように発達する。

「責任」概念の理解および使用法の差異の明確化: 「責任」の概念については、実際のところは哲学領域においても判明ではない。そこで、responsibilityの語源であるラテン語のresponderelに一度立ち戻って整理すると、この概念は多層性を有する概念であることが明らかになった。すなわち、この概念は、少なくとも「誰かを責める」、「誰かが誰かに苦痛を与える」、「被害・損害に対して応報(補償)することができる」、「事件や事故に対して因果関係を有する」、といった要素に分解することができる。ただし、これらの要素はそれぞれ固有の哲学的問題を含んでいる。たとえば、ある自動車による事故の原因は何かと問うたとしても、その原因として挙げられることはあまりに多く挙げることができるし、究極的には「神」「カルマ」といった超越の対象に還元することすらできる。もちろん、因果的遡及は、近代的主体概念が想定する「自由意志」を認めさえすれば、ある特定の人間に定めることができるだろう。しかし、法学グループからは、このような解決は、自動運転車のように機械と人間との協調動作によって生じ、機械の制作にもたくさんの人間が参与するような状況では実務的に限界があることが指摘された。

近代的主体以外に社会を構成する要素の明確化と、脳科学・人工知能との関係の考察: 文化や歴史といった多様な背景に巻き込まれた人間、企業をはじめとした法人または集団、さらにロボットや人工知能も、社会を構成する主体と見なすためには、人間一人一人が持っていると指定された理性を中心とする近代的主体概念から一度離れる必要があるように思われる。そこで、哲学グループがその選択肢の一つとして着目したのが「性格」であった。Ethicsは「性格」を意味するギリシア語エートスに因む。その意味では、倫理とはどのような行為や規範が望ましいかを提示する学ではなく、善い性格を見定め、善い性格になるための条件や方法を探求する学であった。性格を持つものは、自然法則に即した物理的対象とは異なり、ある刺激や入力に対して、別個の動きや振る舞いを成す。そのような動きや振る舞いの傾向性によって、それが善い性格を持つか、悪い性格を持つかを、事実われわれは判断している。このような性格概念は個々の人間が有するものであるが、動物にも適用できるだけではなく、アルゴリズムの性格、ロボットの性格、企業の性格、などのように拡張できる。それゆえ、ひとつのアイディアは、社会を構成する主体を「性格を持つもの」として捉えなおす、というものである。このような考えは、近代的主体概念を必ずしも排除しない。というのは、義務論的規範や理性的判断に従う傾向性も性格の一種だからである。つまり、性格に基礎づけられた主体概念は近代的主体を包括することが可能である。他方で、本プロジェクトの出发点であった主体概念そのものの妥当性を検証する必要も生じた。この世界あるいは社会から独立した個人を「主体」として切り出すという発想は明治維新前の日本ではあまり見られるものではなく、アリステレス主義のある種の解釈に由来することが明らかになったからである。

【法学グループ】

総論: 法学においては、外的環境の影響から独立して、自由意志に基づいて意思決定を成すことができる主体としての個人と、その個人の自由意志に基づく行為と決定論的な因果関係(行為と結果とが必ず1対1で対応する、「あれなけ

ればこれなし」の関係)に立つ結果の存在とが指定され、この指定に基づいて望ましくない結果を引き起こした個人に対する法的責任を分配するという方法が一般に採られてきた。このような責任理論は、いうまでもなく、カントに代表されるような、西洋近代哲学の人間像に依拠しているものである。

しかし、哲学グループが指摘するように、このような人間像を指定することを正当化する概念体系を離れて、それ自体必ずしも反省的検討に基づくとはいえない「現実的必要性」の観点から、このような責任体系は拡張される傾向にあった。例えば、過失犯における、結果の予見可能性に基づく結果回避義務違反を過失行為の本質としつつ、この結果の予見可能性概念を相当程度抽象化することにより、処罰範囲を拡張する議論などがこれに該当するといえる。その結果として、人工知能のように統計的な学習に基づいて、一定の確率で望ましい挙動をする人工物を開発・流通させることに関する刑事責任や、人と自律性を備えた人工物との協調動作によって生じる事故に関する刑事責任などについて論じることが非常に難しい状況に至っている。あるいは、処罰範囲を限定するために、融通無碍に功利主義的考慮を導入することを試みることによって、概念体系の更なる混乱を招いている。

そこでグループとしては、このような問題状況を解決するために、工学・哲学グループの研究成果に依拠しながら、①「徳」のように蓋然性を伴う状況において機能しやすい概念を基礎とする責任理論を基礎とし、②責任の内容を応報性から応答性へと遷移させることにより、③人工知能やそれが構成する複雑なシステムの性質に則した、より全体論的なアプローチに基づく法制度に関して、実務的な観点を加味しつつ議論を重ねた。

人工知能と刑法理論の齟齬：人工知能がもたらす現実と既存の刑法理論との齟齬は、過失犯に著しい。通説に従って、結果の予見可能性に基づく結果回避義務違反として過失犯を理解し、予見可能性の内容を抽象化した場合、確率的に挙動する人工知能は事故を避けることが出来ないため、開発者等には常に刑事責任が問われうることになる。一方、過失の内容を具体化すると、人工知能の個々の異常挙動について予見できる者は存在しないため、開発者等は常に刑事責任を問われないことになる。この過剰処罰か過少処罰かという極端に触れる結論を回避するために、功利主義的観点から一定の安全配慮義務に止めると、「結果回避義務」の設定に見られるような、応報刑を基礎付けてきた義務論的な人間観との齟齬を生じることになりうる。このことは、功利主義からは、そもそも「結果回避義務」ではなく、費用便益計算に基づく最適ナリリスク低減措置が導かれることにも表れている。

実証研究と整合的な規範理論の必要性：上述の問題を解決するにあたっては、実証的な研究と整合性を保つことのできる規範理論が必要とされる。というのも、法は当為の問題を取り扱うが、事実のレベルと大きく食い違う規範理論に基づいて構築される法理論は、既に見たように実践的に役に立たないからである。この意味で、工学及び哲学グループとの対話を通じて特定される規範理論に依拠して法概念及び概念体系を再構築することは、必須といえる。

徳に基づく応答的責任理論：工学及び哲学グループとの対話を通じて見出された規範理論のうち、行為者の傾向性や性格という、徳を根拠として応答的に課される責任という議論には大きな可能性が見出された。というのも、既にアメリカにおいては、企業が引き起こす法益侵害を蓋然性のレベルで理解し、一定の水準に危険を保つよう努力する企業を「良き市民」として評価し、そうでない企業に対して制裁を加えるという法人処罰制度が運用されていたからである。このような法制度は、従来の応報的な法制度との乖離が指摘され、また、ほぼ専ら法と経済の観点から従来の教義学的な法理論を否定する形で主張されたものであったため、法教義学的な観点からは空隙が生じていたと評価できる。しかし、徳に着目し、義務論及び功利主義を包摂する理論を用いることにより、このような空隙を埋めつつ、法と経済の成果を生かして法教義学的な観点からも規範理論を再構築することができると可能性を手にしたといえる。このような応答的な責任理論は、人と人工知能が構成するシステムそのものの全体論的な改善に力点を置いて責任を分配することを正当化しやすいため、現実の問題解決という実践的観点からも優れているという評価を得た。

6. 今後の展開

工学グループが提示したように、人工物に対する一般人の責任帰属感がその人工物が人間らしく振る舞うか否かに依拠するのであれば、人工物が人間らしく振る舞うことをどこまで許容すべきか、という倫理的問いが生じる。この問いはAuthenticityの問題として一部で話題になっているが、綿密な考察や法制度への適用は未だ行われていない。

また、哲学グループが提案した「性格」の強調は、思想的には徳倫理への回帰でもある。それゆえ、徳倫理に向けられた疑義は、この提案にも同様に適用しうる。すなわち、人間の美德として挙げられる徳目は時代地域的にも多様であり、その徳目間の関係もさらに複雑である。このような歴史的背景の上に、これから何を美德として挙げるか、しかも、

ロボットの美德や企業的美徳として何を挙げればよいか、といった問いに回答せねばならない。おそらく、それは世代間衡平性や持続可能性を有することになるが、これらの理念を遵守するような適切な美德は、未だ名前さえも与えられていないだろう。また、どこからどこまでを性格と見なすか、適切な性格はどのようなものか、といった哲学的・倫理的課題、ロボットに求められる性格とはどのようなものか、それはアルゴリズム化できるか、といった技術的課題、仮にこのような主体概念を採用したとして、適切な法体系ができるか、悪しき性格であるというだけの根拠で処罰の対象としてもよいか、といった法学的課題なども検討しなければならない。とりわけ、「徳」を中心とする責任制度は、従来の行為責任を中心として構成されてきた近代的法制度の抜本的な再構築を必要とするため、そのような法制度の機能可能性や社会的受容性を確保するためには、従来の法制度のもとで蓄積されてきた議論との整合性をより精緻に検証しつつ、必要な概念設計を行っていく必要もあると考えられる。

とはいえ、このような課題に関する知見は各専門領域の独立した取り組みでは得られるようなものではない。本邦では複数の領域専門家が協働できる場がまだまだ限られている。そのような場を頻繁に設置し、市民も交えた協働の機会を増やすことが、実社会をより適切な形に発展させていくには不可欠だと思われた。

【研究成果の発表状況等】

○論文（計25件）うち査読付論文 計12件、うち国際共著論文 計2件、うちオープンアクセス 計14件

- ・「ロボットへの原因と責任の帰属」、河合祐司、『日本ロボット学会誌』第38号1巻、pp. 32-36、2020年1月。
- ・‘Exploring future rules for AIs with citizens using a fictitious case video: A workshop report’, Yuji Kawai, Tatsuhiko Inatani, Takako Yoshida, and Kazuya Matsuura, *Proceedings of the International Workshop on Envision of Acceptable Human Agent Interaction based on Science Fiction at HAI 2019*, pp. 1-5, October 2019.
- ・‘Mind perception and causal attribution for failure in a game with a robot’, Tomohito Miyake, Yuji Kawai, Jihoon Park, Jiro Shimaya, Hideyuki Takahashi, and Minoru Asada, *Proceeding of the 28th IEEE International Conference on Robot and Human Interactive Communication*, TuCT1. 2, pp. 1-6, October 2019
- ・「ロボットへの主観的な責任帰属とエージェンシー」、河合祐司、浅田稔、『第38回日本ロボット学会学術講演会予稿集』、2020年10月。
- ・「ロボットの行為」、松浦和也、『日本ロボット学会誌』第38巻1号、pp. 13-17、2020年1月、査読有
- ・「機械・設計・哲学」、松浦和也、『国際哲学研究別冊』第13号、pp. 49-56、2020年3月、査読有
- ・「呪われた孤児としての現存せざるロボット」、松浦和也、『日本ロボット学会誌』第38号1巻、pp. 95-97、2020年11月、査読有
- ・「向社会的行動と教育ロボット」、荒井明子、松浦和也、『国際哲学研究』第10号、pp. 23-29、2021年3月、査読有
- ・「生成消滅とヒュポケイメノン —アリストテレス『自然学』第5巻第1章における変化の分類—」、松浦和也、『白山哲学』、第56号、2022。3。10、pp. 65-84、査読有
- ・「人間並みだが知性から離れたところにある幸せ—古典期ギリシアのエウダイモニア観から—」、『国際哲学研究』、松浦和也、第11号、2022。2。2、pp. 144-122、査読有
- ・「ドイツ観念論およびヘーゲル哲学における「生命」概念—『生命とは何か』という問い それ自体を問いながら」、山田有希子、『哲学』第70号、pp. 106-122、2019年4月
- ・「『死生学』教育の試み —終末期医療およびがん診療の事例に基づいて—」、山田有希子、村上恵理、『宇都宮大学教育学部研究紀要第1部』第70号、pp. 153-166、2020年3月
- ・「ヘーゲル論理学における様相論—偶然性としての絶対的必然性または絶対的必然性としての偶然性—」、山田有希子、『宇都宮大学教育学部研究紀要第1部』第70号、pp. 171-187、2020年3月
- ・「西田のヘーゲル批判再批判に向けて —ヘーゲルの『概念(Begriff)』をてがかりに」、山田 有希子、『哲学雑誌』、第135巻第808号、2021年10月、pp. 112-133、依頼論文
- ・岡田大助、『親鸞「教行信証」における因果応報と縁起について 阿闍世救済の物語と三願転入の照応に注目して』、『国際哲学研究』、第11号、2022年2月、pp. 223-231、査読有
- ・「絵画化された認識論に抗して」、野村智清、『因果・動物・所有 —ノ瀬哲学をめぐる対話』（宮園・大谷・乗立編）、pp.

55-73、2020年1月

・「知識と常識という方法—アイルランドにおける一系譜」、野村智清、『「常識」によって新たな世界は切り拓けるか コモン・センスの哲学と思想史』(青木・大谷編)、pp. 29-52、2020年2月

・「Society 5.0とa nation of men blind from their infancy」、野村智清、『国際哲学研究別冊』第13号、pp. 57-71、2020年3月、査読有

・「来るべき教育に向けて陰謀論を無毒化する」、野村智清、『世界を可視化する哲学地図』、学芸みらい社、pp. 190-211、2021年4月

・「限りなく人に似た機械と井上円了という可能性」、野村智清、『国際哲学研究』第11号、2022. 2. 22、pp. 203-212、査読有

・「因果性と規範性:一ノ瀬化されたヒューム因果論」、相松慎也、『因果・動物・所有 一ノ瀬哲学をめぐる対話』、武蔵野大学出版会、pp. 333-356、2020年1月

・「コンヴェンション・言語・常識:ヒュームの道徳哲学における常識の生成」、相松慎也、『「常識」によって新たな世界は切り拓けるか コモン・センスの哲学と思想史』、武蔵野大学出版会、pp. 53-72、2020年2月

・「ヒュームにおける道徳性と合理性」、相松慎也、博士学位論文(東京大学)、2021年2月、査読有

・「ロボット事故の刑事責任」、稲谷龍彦、『日本ロボット学会誌』38巻1号、pp. 37-40、2020年、査読有

・「ポスト・ヒューマニズムにおける刑事責任」、稲谷龍彦、『AIで変わる法と社会: 近未来を深く考えるために』、岩波書店、pp.113-136、2020年9月

・「法をめぐる現代的課題とヘーゲル」、稲谷龍彦、『ヘーゲル哲学研究』27号、pp.41-53、2021年12月、査読有

○著作物 (計5件)

・『ロボットをソーシャル化する:「人新世の人文科学」10の論点』、松浦和也(編著)、学芸みらい社、2021年3月、256頁

・『現代哲学のキーコンセプト 因果性』、ダグラス・クッチ(著)、相松慎也(訳)、岩波書店、2019年12月、230頁

・『創発と危機のデッサン』、河本英夫、松浦和也(共著)、学芸みらい社、2022年4月、308頁

・『現代倫理学基本論文集III 規範倫理学篇2』、大庭健(編)、古田徹也、相松慎也(共訳)、2022年8月、344頁

・『ヘーゲル全集 第10巻2』、久保陽一(責任編集)、山田有希子(共訳)、知泉書館、2021年10月、360頁。

○講演 (計27件) うち招待講演 計14件、うち国際学会 計3件

・「運動創発から身体性認知への計算論的アプローチ」、河合祐司、JST/RISTEX/HITE研究開発プロジェクト「自律機械と市民をつなぐ責任概念の策定」第8回研究会、2019. 6. 2.

・「脳の構造と機能をつなぐ構成的神経科学:脳から身体、社会へ」、河合祐司、シンギュラリティサロン、2019. 10. 5.(大阪)・11.2.(東京)

・「機械学習の説明性と応用脳科学」、河合祐司、応用脳科学コンソーシアムニューロ/AI ELSIワークショップ、2019.10.29.

・'Self-organization of connectivity in spiking neural networks with balanced excitation and inhibition', Jihoon Park, Yuji Kawai, and Minoru Asada, The 29th Annual Computational Neuroscience Meeting, July 20, 2020.

・「ヒトとロボットが共に生きる社会を考える」、河合祐司、第一回大阪大学共生知能システム研究センター【おうちで】サイエンスカフェ、2020. 8. 10日

・Autistic children classification on MEG signal using deep learning and its explanation、河合祐司、第43回日本神経科学大会シンポジウム『人工知能とビッグデータは精神神経疾患の神経科学に何をもたらすか?』、2020年7月29日-8月1日

・「エピロスからの自由—原因と責任の交差をめぐる—」、松浦和也、第12回RISTEX浅田PJ研究会、2019. 4. 21、参加者数15名(うち研究者15名、一般15名)

・「動作・行為・解釈」、松浦和也、第13回RISTEX浅田PJ研究会、2019. 7. 21、参加者数13名(うち研究者13名、一般15名)

・「呪われた孤児としての現存せざるロボット」、松浦和也、第37回日本ロボット学会学術講演会、2019.9.5、参加者数40名(うち研究者30名、一般10名)

・「哲学研究とこれからの社会」、松浦和也、東洋大学国際哲学研究センターキックオフミーティング、2019.10.26、参加者

数20名(うち研究者15名、一般5名)

・「哲学カフェ・人工知能と暮らす世界を共に考える」、松浦和也、電気通信大学第13回サイエンスカフェchofu、2019.11.8、参加者数35名(うち研究者5名、一般30名)

・「責任概念の構造」、松浦和也、NTTデータ経営研究所第2回ニューロ/AI ELSI ワークショップ 2019年12月18日、参加者数20名(うち研究者5名、一般15名)

・「エートス・エコンタ ―人間・集団・人工物―」、松浦和也、人間と機械の協同を考える、2020年8月2日、参加者数15名(うち研究者15名)

・「機械と世界」、松浦和也、アカデミーヒルズ・JST/RISTEX/HITE「混沌(カオス)を生きる」第3回サイボーグと魂のつながり:インド・日本・ギリシアの哲学から考える機械と身体、2020.12.13、参加者数350名(うち研究者20名、一般330名)

・「二義的な、神々に愛されぬ者たちの幸せ」、松浦和也、ワークショップ:思考力とウェルビーイング、Zoom、2021.1.23、参加者数50名(うち研究者40名、一般10名)

・「authenticityと工学の倫理」、コンピュータ化によるパラダイム変化とその対応研究会、松浦和也、2022. 3.7、関西大学東京サテライトキャンパス、参加者数10名(うち研究者10名)

・「終末期医療における『全人的ケア』を考える」、山田有希子、日本大学医学部PreBSLシンポジウム、2021.10.21、日本大学医学部

・「Let others think for you. Berkeley and Assent」、野村智清、UK-Japan Special Conference: Aspects of Early Modern British Philosophy、参加者13名(うち研究者12名、一般1名)

・「機械と協働する教室」、野村智清、人間と機械の協同を考える、2020. 8. 2、参加者数15名(うち研究者15名)

・「ロボットと権利」、野村智清、第39回日本ロボット学会学術講演会オーガナイズドセッション、2021.9.10、Zoom開催。

・「ヒューム倫理学の概要:道徳の解剖学」、相松慎也、JST/RISTEX/HITE研究開発プロジェクト「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」および「自律機械と市民をつなぐ責任概念の策定」合同研究会、2019.11.30、参加者数8名(うち研究者8名)

・「自由意志なき「責任」:ヒュームの経験論哲学を手掛かりに」、相松慎也、人と機械の協働を考える、2020. 8. 2、参加者数15名(うち研究者15名)

・「道徳的配慮に値するロボット:道徳的被行為者性の可変性」、相松慎也、第39回日本ロボット学会学術講演会オーガナイズドセッション、2021.9.10、Zoom開催、参加者数25名(うち研究者25名)

・「因果性とは何か:言語・概念・実質の交錯」、相松慎也、令和三年度産業日本語研究会ワークショップ、2021.12.14、Zoom開催

・「Tesla死亡事故裁判と法的責任概念の行方」、稲谷龍彦、人と機械の協働を考える、2020.7.31、参加者数15名(うち研究者15名)

・「ロボットの法と倫理の現在」、稲谷龍彦、第38回日本ロボット学会学術講演会、2020.10.9.

・「法をめぐる現代的課題とヘーゲル」、稲谷龍彦、第31回日本ヘーゲル学会大会、2020.12.12.

○本事業で主催したシンポジウム等(計3件)うち国際研究集会 計0件

・一般公開ワークショップ「一緒にさがそう未来のルール ～ロボットの事故は誰かのせい?」、日本科学未来館、2018年8月21日、10月14日、参加者数合計125名以上。

・「人間と機械の協同を考える」、Zoom、2020年7月31日、8月2日、参加者数15名(うち研究者15名)

・「ワークショップ:ロボット倫理の最前線とその社会化に向けて」、日本ロボット学会第39回学術講演会(Zoom)、2021年9月10日、参加者数25名(うち研究者25名)