

令和 5 年 6 月 19 日現在

機関番号：33302

研究種目：基盤研究(C) (一般)

研究期間：2018～2022

課題番号：18K00027

研究課題名(和文)人工知能を利用したロボット金融トレーダー等に対する責任帰属の可能性に関する研究

研究課題名(英文)Research on the Possibility of Responsibility Ascription to AI Traders

研究代表者

増淵 隆史(Masubuchi, Takashi)

金沢工業大学・基礎教育部・准教授

研究者番号：60528248

交付決定額(研究期間全体):(直接経費) 900,000円

研究成果の概要(和文)：(1)人間の法的機能の一部を機械が代替することをもって、人工知能に一定の法的地位を認めうるという先行研究から、法以外の分野でも人工知能に限定的な人格性を認めうる余地があることを論証した。(2)AI金融取引の普及は金融市場の存立自体を危うくする可能性があることを論証した。(3)説明可能なAIの開発に際しては、哲学・社会科学分野における「説明」に関する研究成果を反映することが実用化に不可欠であることを論証した。(4)説明は人間の自律性の尊重といった哲学的な価値の実現と関係するため、AI取引においては投資判断における投資家の自律的意思決定を可能とするような説明が必要とされることを論証した。

研究成果の学術的意義や社会的意義

(1)法的分野における人工知能の地位に関する議論を、哲学倫理学分野における人格性の議論に敷衍した点に新奇性がある。(2)AI取引の普及の問題の核心がAIアルゴリズム取引の倫理性ではなく、金融市場の存立を脅かす可能性がある点にあることが新たに示された点に意義がある。(3)諸外国における哲学・社会科学分野の説明に関する研究をサーベイしたことにより、国内で展開されている「説明可能な人工知能」研究に寄与する点に意義がある。(4)今後法制化が予想されるアルゴリズム・AI取引に関する規制法の検討において、「投資家の自律的意思決定」のような哲学的観点からの問題意識を反映させる可能性を開いた点に意義がある。

研究成果の概要(英文)：(1) According to the previous research that claims it is possible to allow legal status to AI, this research demonstrates that there is room to admit having a certain degree of personality to AI.(2) This research demonstrates that it is expected that the spread of AI trading not only destabilizes the financial markets but also jeopardizes them.(3) This research demonstrates that it is indispensable to the development and implementation of explainable AI to reflect the research findings in the realm of philosophy and social sciences.(4) This research demonstrates that the explanations which enable the autonomous decision making are required in AI trading because the explanation involves the realization of philosophical values like 'respect for autonomy'.

研究分野：哲学・倫理学

キーワード：人工知能 説明責任 AI・アルゴリズム取引 金融倫理 自律的意思決定

1. 研究開始当初の背景

金融業は20世紀後半以降、世界経済の中で製造業をしのぐ主要な産業に成長した。またインターネット等の通信技術やコンピュータ技術の発展を受けて、金融取引の電子化・コンピュータ化が進んだ。これに伴い、多数の一般投資家が市場に容易に参入できるようになったが、証券取引の専門家とこれらの投資家間のスキルやノウハウ、あるいは技術革新への対応度の差は大きく、これを原因とした倫理的問題が発生している。

金融市場におけるこれらの倫理的問題に対しては、取引条件の公正の確保や個人投資家の保護などの対応が必要となる。特にテクノロジーの進歩は、最新技術の恩恵に与らない個人投資家等に不利益をもたらす可能性があるため、技術的欠陥や情報格差による一般投資家の損害を防止する責任が証券会社等に生じる。この責任を果たすためには、科学技術利用に起因する損害に対する責任の帰属や内容の明確化が必要であり、このためには規範倫理学や科学技術倫理、金融倫理等の倫理的知見が必要となる。これが金融市場の技術的革新とそれに起因する問題に対して倫理学が関与することになる学術的背景であった。

2. 研究の目的

本研究の目的は、金融業における人工知能テクノロジーの応用に関する実態把握を基に、伝統的な責任概念との乖離を明確化し、伝統的責任概念の修正の可否もしくは要否の検討を行ったうえで、金融業における人工知能等のテクノロジーを用いたいいわゆる「ロボ・トレーダー」や「ロボット投資アドバイザー」等の人工物に倫理的責任を帰属させることが可能か否か、またその責任の内容は何かということをも明らかにすることにある。

3. 研究の方法

最新の金融テクノロジーの内容の調査に関しては、まず外国語・邦語論文・著作やジャーナルなどの文献調査を行い、先行研究の把握や人工知能の金融業への応用の現状の確認を行う。さらに最新かつ詳細な研究や実用化の動向については、人工知能や金融工学関連の研究発表会等に参加して調査・研究を行う。また人工知能や情報通信テクノロジーに関する哲学・倫理学研究の調査に関しては、同じく外国語・邦語論文・著作やジャーナルなどの文献調査により先行研究を把握するとともに、応用哲学や企業倫理学に関する各種学会等の研究発表会等に参加して、最新の動向および研究内容について情報収集と分析を行う。最終的にはそれらの知見と申請者の先行研究などを総合的に検討し、責任の帰属可能性と内容の問題について結論を導き出す。

4. 研究成果

(1) 法律上の人工知能の地位に関する先行研究の知見の整理

人工知能に責任を帰属させるか否かという問題は、人工知能に人間と同等の倫理的地位すなわち人格性を認めうるか否かという問題として捉えることが一般的である。人工知能に人格性を認めうるかという問題に関しては、法学の分野において人工知能に法的な人格性を認めうるかという問題の研究が進んでおり、この研究の現状をサーベイすることが哲学・倫理的な領域における人工知能の人格性の議論に有効であるため、海外の先行研究をサーベイし、その議論を明確化した。

海外の先行研究をサーベイして明らかになったのは、海外の法的分野での人工知能の法的地位に関する研究においては、人工知能を契約における代理人と同様に扱うという考え方が提起されていることである。この考え方を提唱する Chopra と White によれば、人工知能は完全な自由意志を持つ人間ほどの自律的な能力はないが、契約にあたり契約者本人の意思の制約のもと、一定の範囲内での判断能力を行使しうる代理人と類似の自律性は持ちうる。したがって、法的領域において人工知能には代理人と同等の地位を与えることが可能であり、これによって契約上発生する人工知能と人間との間の責任負担の問題を適切に解決しうると主張している。

本研究ではこの先行研究について、人間の法的機能の一部を機械が代替することをもって、人工知能に一定の法的地位を認めるという点を評価し、これをもって法以外の分野でも人工知能に限定的な人格性を認めうる余地があると論じ、平成30年12月22日開催の北海道大学哲学学会研究発表会にて発表を行った(発表題目: 契約行為における人工知能の地位 - 「道具」からの脱却は可能か)。

本研究の国内での位置づけは、法的分野における人工知能の地位に関する議論を哲学倫理学分野における人格性の議論に敷衍した点で新奇性がある研究という位置づけとなると考える。

また国外に関しては、同様の新奇性があると評価される可能性があると考えられる。また今後人工知能による人間の機能代替が進むにつれて、人工知能の人格性を判断するための一つの考え方として重要なものとなる可能性を示すものとなると考えられる。

<参考文献> Chopra, S. C. and White, L. F. (2011), A Legal Theory for Autonomous Artificial Agent, The University of Michigan Press.

(2)人工知能を用いた金融取引が金融市場に及ぼす影響の仮説の構築

金融分野における人工知能の説明責任を考えるためには、人工知能を利用した機械取引が、金融市場に対してどのようなインパクトを与えるのかを明らかにする必要がある。

この問題について本研究では、金融市場においてAIを利用したいわゆる「アルゴリズム取引」が株価の乱高下などの一因となっているという見解に注目し、このような人工知能の金融市場への影響は、「市場の公正」という倫理的観点から正当な機能として説明可能であるかという観点からの研究を行った。

本研究ではこの問題について、政治哲学の分野における代表的な社会正義論の理論であるロールズの正義論、ノージックの権原理論そしてハイエクの社会的正義論という3つの学説を分析枠組みとして用い、分析を行った。その結果、本研究で得られた結論は、人工知能を用いた金融取引がもたらす金融市場の不安定化は、この3つの学説がよって立つ市場の公正の考え方からは不公正とは言えないが、金融市場が成立する条件である「多数の市場参加者間での知識の分散」を減少させる点で、市場構造の変容をもたらす恐れがあるというものである。金融市場におけるアルゴリズム・AI取引が普及することは、金融取引に利用されるデータの同質化が進むことという事態を伴い、多様な価値観をもつ市場参加者の減少をもたらすため、市場は構造的に不安定化を強いられることになる。これは人工知能の機能がもたらす不正というよりは、人工知能の普及による金融市場の構造的な脆弱化という別の問題を起こすと考えられる。

本研究ではこの見解について、平成30年度第9回応用倫理・応用哲学研究会において研究発表を行った(研究題目:AIトレードと金融市場の倫理 AIがもたらす市場の混乱とモラル・ディレンマ)。

本研究は、もともと金融市場におけるAIアルゴリズム取引の公正性とその説明可能性を解明することを目的として行ったものだが、研究が進むにつれ、この問題の核心がAIアルゴリズム取引の倫理性ではなく、金融市場の存立を脅かす可能性がある点にあることが表面化してきた。これは当初予期しなかった新たな知見につながるものと考えることができる。

また、この論点はAIアルゴリズム取引が引き起こす問題点としては、これまであまり学術的観点からは解明が行われていない論点であり、今後のAIアルゴリズム研究に新たな展望をもたらすことが期待される。

(3) AI技術者による人工知能の説明責任研究に関する哲学社会科学からの分析の整理

人工知能の説明責任を研究する際に、説明の仕方としてどのような方法が可能であるか、言いかえればどのような説明をすれば説明をしたことになるのかを明らかにすることは重要である。本研究ではこの点について、以下の、の2つの人工知能研究者の先行研究のサーベイを行うとともに、その研究をその後の哲学倫理学分野での人工知能の説明責任に生かす場合の課題について研究を行った。

法的領域における人工知能による説明に関する研究のサーベイ

Dashi-VelezとKortzは人間が行う「説明」という行動を分析し、それが人工知能によっても実行可能かどうかについて考察した。

Dashi-Velezらの考えでは、これらのケースで人間が行う説明とは、各々の説明要求に対してその理由または正当性を示すという形で与えられる。そして彼らは法の領域における説明という局面では、人工知能が人間同様の説明をすることは可能であると述べる。

この法的領域における人工知能の説明が、倫理学の領域における説明にそのような影響を与えるかについて、本研究では以下のように総括した。法的領域における人工知能の説明においては、判断の基準として人種差別等の違法性が前提とされていることが、倫理的な意味での説明となると、確固とした権威を持つ判断基準として明文化された法律がある法の領域における判断とは異なり、判断の基準となる倫理的判断基準の正当性自体が問われる場合もあるため、説明となりうる理由や正当性の根拠についてははもう一段上のレベルの回答を要求されると結論付けた。

本研究ではこの見解を、平成30年12月22日開催の北海道大学哲学会研究発表会における発表の一部として組み込み(発表題目:契約行為における人工知能の地位 - 「道具」からの脱却は可能か) その部分を「発表要旨『人工知能と説明責任』」として北海道哲学会会誌『哲学年報』65号(令和元年9月)に投稿、掲載した。

参考文献 Doshi-Velez, F and Kortz, M. (2017), "Accountability of AI Under the Law: The Role of Explanation", Cornell University Library (Submitted on 3 Nov 2017 (v1), last revised 21 Nov 2017 (this version, v2), <https://arxiv.org/pdf/1711.01134.pdf>)

人工知能研究者による哲学・社会科学分野の「説明」に関する研究のサーベイ

Millerは、「AIにおける説明と解釈可能性は、近年再注目されているにもかかわらず、この分野の研究と実践活動においては、何が十分な説明を構成するかについて、研究者の直観が使われているように見える」と述べ、より信頼性のある研究の基礎として、哲学、認知心理学/科学、社会心理学の分野のこれらのトピックスについての膨大な研究業績を取り入れるため、哲学・社会科学における「説明」に関する研究をサーベイし、それらの研究が、説明可能な人工知能(XAI)開発にどう寄与しうるかを考察した。

このサーベイの結果、Millerは「説明は「である」ではなく「××ではなく、であ

る」といった対照的 (contrastive) なものであること、説明は (バイアスがかかった形で) 選択的 (selective) であること、説明において確率 (probabilities) は重要ではないこと、そして説明とは社会的であることの4点を、哲学・社会科学から得られた説明の理解を成果として提示した。

Miller は、説明として人工知能にどのようなことができるようになる必要があるかを考えるうえで、説明の内容として上記のような要素を反映させることによって、責任を単に因果帰属を提示する説明より被説明者の理解可能性が高まると述べ、これが哲学・社会科学の研究からの寄与であると結論した。

このサーベイ研究については、2021年7月25日開催の「北海道哲学会シンポジウム『AIの倫理学 その問題圏の検討』」において、「AI技術者が説明について社会科学から学ぶこと」という題目で発表を行った。

本研究の成果は、第一に人工知能研究者の間で進む「説明可能な人工知能」の開発において、「説明」という活動をどのように理解して開発すべきかという点に関する研究の動向を紹介することができたことである。第二に、説明可能な人工知能の開発に際し、哲学・社会科学の「説明」に関する先行研究の紹介とその重要性を提示できたことである。

本研究は国内的には、国内で展開されている「説明可能な人工知能」研究に対し、これらの研究を取り入れることに寄与できる点に本研究の有用性があると位置づけられる。今後の展望については、これらのサーベイ研究を受け、さらに哲学・倫理学分野における説明に関する先行研究を洗練させていくことにより、より充実した人工知能の説明機能の発展に資することが期待できる。

参考・引用文献 Miller, T. (2019), 'Explanation in artificial intelligence: Insights from the social sciences', *Artificial Intelligence* :267(2019) p.2.

(4) 現行の技術水準下でのアルゴリズム・AI取引における説明責任の内容について

本研究では、現在の金融商品におけるAIアルゴリズム取引の技術水準下では説明責任としてどのような責任を果たすことが求められているかをサーベイし、それが説明責任として十分であるかについて分析を行った。

金融分野におけるAI利用の問題点については、日本銀行金融研究所から「投資判断におけるアルゴリズム・AIの利用と法的責任」というタイトルの報告書(以下「報告書」)が発行されており、そこでは解決の方向性が提示されている。その中にはブラックボックス化が懸念される投資判断基準の説明責任の問題が含まれている。

現行の技術水準下では、AI自身に投資判断の説明を行わせることは困難なので、説明責任の主体は人間となるが、そこではアルゴリズム・AIを用いた投資で投資判断基準に関し、原理的に説明不可能であるということが想定される場合、説明が不能であることに起因する投資リスクがあるという説明と、それに対する投資家との合意があれば、説明責任として十分であるという見解が示されている。

これに対し、本研究では、上記(3)のサーベイ研究の結果を踏まえ、「報告書」の示した説明責任は、哲学・社会科学的見地から見て適切であるかを考察した。その結果、説明は自律的で社会的な人間の尊重といった倫理的・哲学的な価値の実現と密接に関係しているが、「報告書」の説明責任の考え方はこの意味での説明の重要性を看過しており、説明責任の内容を矮小化することになるため、投資判断における投資家の自律的意思決定の可能性を損なうような非倫理性を是らむことになると結論付けた。

この研究については、2022年12月10日に開催された北海道哲学会・北海道大学哲学会共催研究発表大会において、「金融AIを用いた投資判断において説明責任は可能か」という題目で研究発表を行った。

本研究の成果は、現在主として金融取引法等の法的分野で研究が進められているアルゴリズム・AI取引における説明責任の問題について、国内外で進められてきた哲学・社会科学分野における説明に関する研究の観点から考察を行うことによって、新たな説明責任の考え方を提示することができたことである。

本研究は国内的には、今後法制化が予想されるアルゴリズム・AI取引に関する規制法の検討において、哲学・社会科学の観点からの問題意識を反映させる可能性を開いた点で、有益なものであると位置づけられる。

今後の展望については、今後哲学・社会学的見地からの説明責任に関する研究を進めることで、より投資家の安全に資するアルゴリズム・AI取引に関する規範形成に貢献することが期待できる。具体的には、この研究からは、アルゴリズム・AI取引に使用される人工知能に説明責任の主体性を認めるか否かという点については、投資判断基準に関する投資家の知る権利を満たすような説明が可能か人工知能の開発が実現した時点で、人工知能に行為主体性を認める法的倫理規範の制定を検討すべきであるといった考えを打ち出すことができると思われる。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 増淵隆史
2. 発表標題 AI技術者が説明について社会科学から学ぶこと
3. 学会等名 北海道哲学会・北海道大学哲学会
4. 発表年 2021年

1. 発表者名 増淵隆史
2. 発表標題 AIトレードと金融市場の倫理-AIがもたらす市場の混乱とモラル・ディレンマ-
3. 学会等名 日本経営倫理学会
4. 発表年 2019年

1. 発表者名 増淵隆史
2. 発表標題 契約行為における人工知能の地位 - 「道具」からの脱却は可能か
3. 学会等名 北海道大学哲学会
4. 発表年 2018年

1. 発表者名 増淵隆史
2. 発表標題 金融AIを用いた投資判断において説明責任は可能か
3. 学会等名 北海道哲学会
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------