（別紙 1）

論文の内容の要旨

論文題目　人工知能に対する認知とその社会的影響
――適切に活用する社会の形成に向けて――

氏名　谷邊　哲史

1. 序論

本研究の目的

本研究の目的は、人工知能（artificial intelligence; AI）の開発が急速に進展する現代において、人が AI をどのような存在として認知し、AI の利用に対してどのような態度を示すのかを実証的に検討すること、そしてその検討を通じて AI の利用に関わる社会的課題を議論するための基礎となる実証的知見を蓄積することである。

AI は研究は近年、第 3 次 AI ブームと言われる発展を見せている。そして、過去 2 回のブームとの大きな違いは、AI が一部の専門家だけのものではなく、一般の人々が日常生活の中で利用できるようになりつつあることである。AI の利用される場面が急激に拡大する中で、AI の開発や利用について適切な指針を定めることが課題となっている。

AI の開発・利用は工学のほか、倫理学や法学などさまざまな研究領域に関わる問題があるが、本研究では社会心理学の手法を用いて、AI のユーザーとなる一般の人々の態度を実証的に解明することを目指す。

AI の実用化が期待される 2 つの分野と社会的な課題

本論文では AI の実用化が期待される分野を 2 つに分類し、それぞれの分野において検討すべき課題を整理する。

1 つ目は、従来は人間が行っていた判断を代替するための AI である。自動車の自動運転や自動翻訳など、AI が情報を処理した結果を利用することが直接の目的となるものがこの分類に含まれる。

この分野では、AI の高度化によって人間の操作によらず判断を行えるようになり、その判断の結果として損害が生じる場合もあることから、AI の判断に対する責任の所在など、新たに生じる倫理的問題について論点を整理し、研究や実用化を進めるための指針を策定することが社会的な課題となる。

2 つ目は日常生活の場で人と相互作用するソーシャルロボットである。この場合、AI はロボットを制御する手法の中の一つに過ぎず、ユーザーにとってはロボットの制御が AI によるものかどうかが重要ではない。ただ、現実に AI による制御がロボットの機能向上の要因として期待されていることから、AI の実用化の一形態として本論文で検討の対象とする。

この分野では、ユーザーがロボットに求める要素を探るという設計上の課題と、高度に自律的になったロボットに人間のような道徳的権利は認められるかという倫理的な課題が存在する。

AI の社会的受容の 2 つの側面

先端技術の普及に関する社会的課題を議論する際には、社会的受容という語がしばしば登場する。その意味合いは文脈によってさまざまだが、おおむね「購入したい、利用したい」という個人的な問題と、「その技術が社会に普及することに賛成する」という社会的な問題の 2 つの側面があるようである。

本論文でもこの 2 つの側面を扱い、AI や、AI によって制御されるソーシャルロボットに対するユーザーの利用意図と、社会的に適切な利用方法についての人々の態度を検討していく。

それぞれの実証研究の内容を整理すると図 1 のようになる。

	人間の判断を 自動化するAI	AIによって制御される ソーシャルロボット
個人的な 利用意図	**研究1:** AIの助言と自己決定 （AIを用いた就職活動支援）	**研究4:** 家族の役割を代替する ロボット利用 （コミュニケーション ロボット）
社会的に適切な 利用方法	**研究2:** AIの判断が生む バイアスに対する不公正認知 （AIを用いた融資審査） **研究3:** AIの判断と責任帰属 （自動運転車）	**研究5:** ロボットに対する 心の知覚と道徳的配慮 （ヒト型の家庭用ロボット）

図 1 実証研究の概要

2. 実証実験

研究 1　AI の助言と自己決定

研究 1 では、AI による判断が人間の判断とは異なるものとして受け取られるかを確かめるため、AI の助言を受けたユーザーの反応を検討した。実験では大学生の実験参加者を対象に、就職活動の場面を想定して「AI（または人間のアドバイザー）が勧める企業」と「自分では良いと思ったが AI（またはアドバイザー）の推薦度は低い会社」のどちらに入社したいかを尋ねた。助言者が AI か人間かは参加者ごとにランダムに変えていた。また、実験の最初に自己決定欲求（ものごとを自分の意思で決めたいと考える度合い）の個人差も測定しておいた。

実験の結果、人間が助言する条件では、自己決定欲求が高い人ほど助言に従わなかった。しかし AI の助言の場合には、自己決定欲求の高低によらず同程度に助言に従った。つまり AI の助言は人間の他者からの助言と違い、本人が自律的に意思決定したという感覚を損なわずに助言することができていた。

研究 2　AI の判断が生むバイアスと不公正認知

研究 2 では AI の判断が生む倫理的な問題を検討した。実験では銀行の融資審査において女性が男性よりも不利な評価を受けるという出来事を提示し、その出来事がどの程度不公正かを尋ねた。融資審査の判断を行う主体が AI か人間のどちらかで、ランダムに変えていた。

分析の結果、AI の判断で女性が不利に扱われる場合には、人間の判断で同じ不公平が生じた場合よりも、不公正さの認知が低くなっていた。つまり、人間の判断と同じく不公平が生じているにもかかわらず、AI が判断したというだけで妥当であるかのように判断されていた。

研究 3　AI の判断と責任帰属

研究 3 では自動運転車の事故で歩行者を死なせるというシナリオを提示し、自動運転車のユーザーやメーカーへの原因帰属、責任帰属を判断させた。その結果、AI に事故の原因を帰属する人ほど、ユーザーやメーカーにも原因を帰属し、さらに責任も帰属していた。つまり、AI が自律的に判断したとはいっても、それはユーザーやメーカーの免責にはつながらず、むしろ AI の動作に対して責任を負うべきだと判断されていた。

研究 4　ソーシャルロボットに対する心の知覚と利用意図

研究 4 では、高齢者介護の場面で話し相手となってくれるコミュニケーションロボットの利用意図を検討した。調査対象者は介護者の立場になると想定される若年層で、ロボットの写真を見て、ロボットに心があると感じる程度を評定した後、自分の家族の介護にロボットを使いたいかを回答した。また、介護に対する価値観の個人差として、介護は家族が担うべきであるという態度（家族介護意識）も測定した。

ロボットの利用意図に影響する要因を分析したところ、家族介護意識が高い人は基本的にロボットの利用に否定的であったが、ロボットに感情などの心の能力があると感じることで、利用意図が向上することが示された。機械であっても心があるように感じられることで、介護という場で利用することへの抵抗感を和らげる効果があることが確認された。

研究 5　ソーシャルロボットに対する心の知覚と道徳的配慮

研究 5 では、高度に自律的になったロボットが道徳的な配慮の対象と見なされるようになる可能性について、心の知覚との関連という観点から検討した。先

行研究では、加害行為を想起すると、その対象となるロボットに心があるという知覚が促進されることが示されている。この実験を参考にして、本研究では、①道徳的によい行為（故障したロボットを修理する）を想起しても心の知覚が促進される、②ロボットに心があるという知覚がロボットへの道徳的配慮を促進する、という 2 つの仮説を検証した。実験の結果、仮説はともに支持され、心の知覚と道徳判断が相互に影響しあうことを実証できた。

3. 総合考察

AI やロボットはどのような存在として認知されていたのか

研究 1 から 4 では一貫して、AI がどれだけ高度化し人間の操作によらず判断・行動していても、人間とは異質な存在として認知されていた。しかし研究 5 の結果からは、人間とは異質な存在でありながら、他の機械とも異なる新たな存在として受け入れられる可能性も示唆される。本論文ではこのような認知のあり方を正確に記述することはできていないが、ロボットと人の相互作用をより正確に理解するには、このような認知のあり方を記述する新たな枠組みが必要になるかもしれない。

AI の開発や関連する議論への示唆

ロボットの開発の現場では、人間や動物を模した外見・動作になるように工夫がなされることが多い。しかし研究 1、4 の結果からは、AI やロボットが人間とは異質な存在であることがむしろ利用意図を高めることが示唆される。単に人間に近づけるほどよいということではなく、AI やロボットに求められている人間らしさの要素をより精緻に解明していくことが必要である。

AI をめぐる倫理的な問題に関して、自律的な判断者としての AI に対する道徳判断を検討したところ、AI はどれだけ自律的であっても道徳的な責任の主体と見なされることはなかった。一方で人間が自律的なロボットをどう扱うかという点では、心の存在を知覚させるロボットが道徳的配慮の対象と認められる可能性が示された。

上記の研究結果は、AI のユーザーとなる人々の認知や態度に関する実証的な知見を蓄積するものであり、今後の AI の開発や、AI の利用をめぐる社会的課題を解決するための議論に際して、議論の基盤となる基礎的な資料を幅広く提供するものである。