

AI 導入に伴う人の利他性と責任感に関するサーベイ実験

柴田惇平^a 中辻早紀^b 濱中杏香^c 坪田凜々花^d 磯川雄大^e

要約

本稿は、将来的に人間が人工知能と共に働くような状況になった場合、人間の他者への利他性及び責任の感じ方、罪の意識が変化するかについて実験を用いて分析することを目的とするものである。研究では場面想定法を用いて、4人1チームで「自分の周りが人間の場合」と「自分の周りがAIの場合」を想定し、2×1の被験者間実験を実施し、相手への協力度の差をwebアンケートでの公共財ゲームによって明らかにした。加えて責任の感じ方や罪の意識を7段階で回答してもらい、状況によって回答に差が生じるかを分析した。結果的に、投資額には両方で差が生じたが、責任の感じ方や罪の意識には統計的に有意な差は見られなかった。

JEL 分類番号 : D91, J24, O33

キーワード : AI, 利他性, 責任, 罪の意識, 公共財供給ゲーム

a 柴田惇平 同志社大学商学部 tagulabo@gmail.com
b 中辻早紀 同志社大学商学部
c 濱中杏香 同志社大学商学部
d 坪田凜々花 同志社大学商学部
e 磯川雄大 同志社大学商学部

1. イントロダクション

1. 1. 研究の背景と目的

本稿は、AI (Artificial Intelligence) と人間が共存していく社会において、人間の他者への利他性および責任の感じ方、罪の意識が変化するのかどうかを分析することを目的とするものである。近年、新しいテクノロジーは著しく発達している。とりわけ AI は我々の日常生活においても非常に身近な存在となっており、今後も人間の社会や経済に大きな影響を与えることが考えられる。もっとも少子高齢化に伴う生産労働人口の減少が見込まれる日本では、AI は業務の効率化や労働不足を補う解決策の一つとして期待されており¹、多くの分野で AI の導入が進んでいる²。また、AI やロボット等の自動化によって、多くの職業が機械に代替されるという推計もなされている。米国では 10-20 年の間に労働人口の 47%が、我が国では将来的に労働人口の 49%が自動化されるといわれる (Frey and Osborne 2013 ; 野村総合研究所 2015)。このように、将来的には人間と AI が同じ環境で仕事をするという局面も増えてくるだろうし、自分の周りで働く「同僚」がすべて AI ということが起こるかもしれない。そうした状況が現実となるかどうかは議論の余地があるが、仮にそのような未来が到来しうるとすれば、社会としてどのように向き合っていくべきか検討することは重要であると考えられる。

社会心理学や行動経済学の分野では、人は周りの評判を求めて利他行動をとるとということが知られている。では周りが AI の場合、AI に対する人間の利他行動は変化するのかわという疑問が生じる。また近年では、AI と倫理の観点から、AI に責任を負わせるべきかといった議論が注目を浴びている。このような新しいテクノロジーの進展を分析するうえで、有効なツールが実験である。実験は、データのハンドリングが容易かつ現実にはまだ存在しない制度や仕組みをも検討することが可能であるという強みを持っており、AI と人とが共存する未来社会の分析と親和性が高い(田口 2015 ; 田口 2018)。

そこで本研究では、従来のような人間が働く状況と AI と共に働くこととなった仮想的な状況において、他者に対する利他性や責任の感じ方、罪の意識が異なってくるのかわ、公共財供給ゲームを用いた調査を行うことで明らかにする。

¹ 総務省(2016)では、職場に導入されている(される)AI が果たす役割・機能について日米の就労者に対してアンケートを実施しており、我が国では「既存の労働力を省力化する」(41.0%)が最も高く、次いで「不足している労働力を補完する」「既存の業務効率・生産性を高める」(ともに 35.0%)であった。

² 特に、医療、介護、芸術・デザイン、教育、接客サービス、輸送・移動、農業、治安・セキュリティの分野において AI・ロボットの導入が進んでいる。国立国会図書館調査及び立法考査局(2018)第 2 部を参照。

1. 2. 先行研究の整理と仮説

利己的な個人が自分だけの利益を追求すると、社会的に望ましい結果は得られない。伝統的経済学では、合理的経済人の存在を前提としていたため、このような社会的ジレンマが生じる。ゲーム理論における均衡もプレイヤーは自身の利得を最大化することから、社会的ジレンマが発生することを予測する。しかし、多くの実験研究から理論の予測に反し、完全利己的な結果と異なり、人間は利他行動をとることが確認されている³。

AI に対する人間の責任や罪悪感を扱った先行研究は数多く存在する。例えば、de Melo et al. (2016) では、コンピュータを相手とした最後通牒ゲームを実施している。ここでは、人間が配分者の場合、自身の利得を大きくすることが観察されている。すなわち、人間は機械に対して低い利他性を示すといえる。その他、Kirchkamp and Strobel (2019) は、2 人の人間と 1 人の人間と 1 台のコンピュータの共同意思決定者によるバイナリー独裁者ゲームを実施し、パートナーが人間か機械かによって被験者に知覚された責任や罪悪感に違いが生じるのか調べている。実験からは、パートナーが機械であっても統計的に有意な差は見られなかった。

ここで、AI と共に働く社会が到来したと仮定したとき、先行研究を踏まえると、人間はパートナーである AI に対して協力的でない行動をとることが予想される。また、その場合の人間の責任や罪悪感も変わらないことと考えられる。

以上から、本研究では、以下の 2 つの仮説をたて、実験を用いて検証する。

仮説 1 : AI と活動する時、AI に対する人間の協力率は低下する。

仮説 2 : 周りが人間でも AI でも、責任感や罪の意識は大きく変わらない。

2. 方法

2. 1. 実験内容

実験内容は以下の通りである。はじめに 1, 000 円を初期保有額として、4 人プレイヤーで合計の投資額を 2 倍した金額を等しく 4 人に分配する公共財供給ゲームを実施した。

場面想定法を用いて、4 人 1 チームで「自分の周りが人間の場合」と「自分の周りが AI の場合」を想定し、相手への協力度の差を調べた。なお、この実験で想定する AI とは、人間の業務の遂行を支援ないし業務の一部を代替する能力を備える人型のアンドロイドで、

³ 例えば、Zelmer(2003)などの研究がある。

仕事のパートナーである⁴。その後、それぞれのパターンで、被験者自身の意思決定の責任の重さ、他のメンバーの意思決定(想定)の責任の重さ、被験者の罪の意識を7段階(0:まったく責任がない(罪悪感がない), 6:非常に責任がある(罪悪感がある))で評価してもらった。⁵

2. 2 実験参加者

実験は2019年8月16日~8月28日にかけて、「自分の周りが人間の場合」と「自分の周りがAIの場合」の2×1の被験者間実験に分けて実施した。実験参加者は、Google Formによるwebアンケートに回答した161名(男性81名, 女性80名)である。被験者は主に学生である。

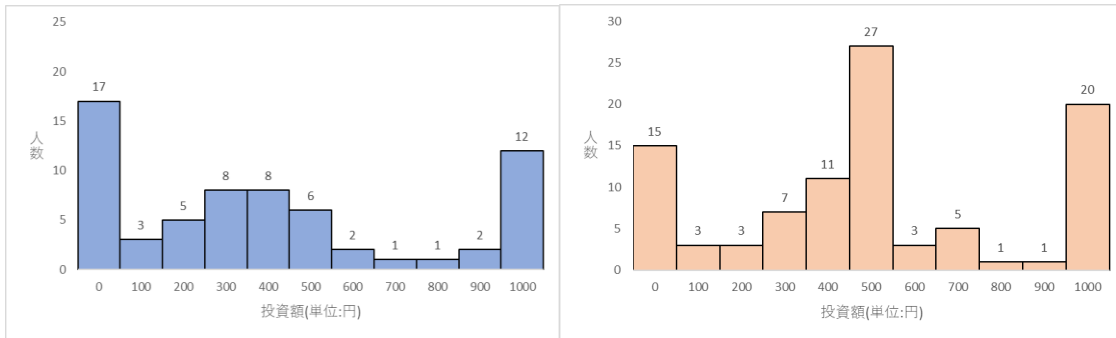
3. 結果

図1, 2, 3, 4にはそれぞれAIを想定した場合の投資額のヒストグラム, 人を想定した場合の投資額のヒストグラム, 各条件における平均投資額, 「自分の投資額への責任」に関する質問項目での平均ポイントを示している。アンケートの結果, それぞれの条件での平均投資額に92.8円の差が生じた。U検定の結果, 統計的にこの平均投資額の差は, 統計的に有意であり($U = 2517.5$, $Z = 2.104$, $P = 0.035$, $0.01 < P < 0.05$), 帰無仮説が棄却されたことにより仮説が立証された。つまり, AIと活動することで人は協力率が落ちることが明らかとなった。

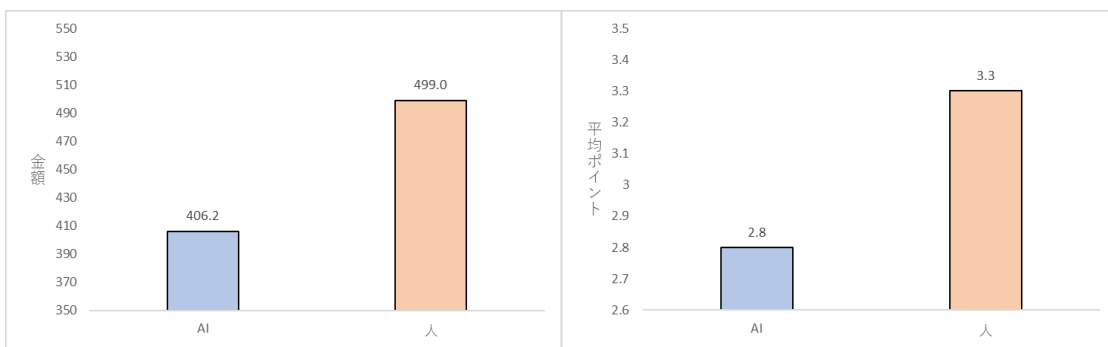
また, 「自分の投資額に対する責任」についての質問項目で, AIを想定したモデルでは7段階中平均2.8ポイントに対し, 人を想定したモデルでは平均3.3ポイントと統計的な有意差は見られなかったが, AIを想定したモデルの方が自分の投資額への責任を低く考える傾向にあることが明らかとなった。

⁴ ここで人型のアンドロイドとした理由は, 我が国の一般の生活者がロボットをイメージする時, 相対的に人型ロボットである割合が高いためである。野村総合研究所(2016)を参照。

⁵ 責任の重さ・罪の意識の評価に関しては, Kirchkamp and Strobel(2019 前掲)を参考にしている。



【図1 投資額のヒストグラム (AI モデル)】 【図2 投資額のヒストグラム (人モデル)】



【図3 各条件における平均投資額】 【図4 自身の投資額に対する責任 平均ポイント】

4. 終わりに

4. 1. まとめ

本研究をまとめると以下ようになる

- AI と共に活動することで、人は協力率が低下する。
- 責任の感じ方・罪の意識に関して統計的な有意差は見られなかったが、「人と活動する場合」と「AI と活動する場合」で、後者の場合に自身の責任の程度を低く見積もる傾向にあることが明らかとなった。

4. 2. 研究の限界と今後の展望

本研究での研究の限界は以下である。実験室実験ではなくアンケート調査での実験であったという点だ。公共財供給ゲームは本来実験室で参加者にゲームの結果に応じて謝礼を支払う形をとるが、今回はアンケートでの実験を行ったため、実験室実験に比べデータの正確性が劣っている。また、プロジェクト等は1回きりではなく複数に渡って行われることがほとんどであるため、アンケートでのゲームの繰り返しを行わない実験では我々の想定した状況をうまくゲームに落とし込むことができなかった。

今後の展望としては、実験室実験を行うことにより、内的妥当性を高める必要がある。また、今回はゲームを繰り返し行わないアンケート調査のため、本研究では扱わなかったが、例えば、人間の場合と同様に戦略的互酬性が働くのか等、繰り返しゲームを行うことで、生じる可能性について明らかにしたい。

引用文献

- de Melo C., Marsella S, Gratch J, 2016. People do not feel guilty about exploiting machines. *ACM Trans Comput Hum Interact* 23(2), 1-18
- Frey, C. B. and Osborne, M. A., 2013. The future of employment : How susceptible are jobs to computerization? Working paper on Oxford Martin Programme on Technology and Employment.
- Kirchkamp, O., and Strobel, C., 2019. Sharing responsibility with a machine. *Journal of Behavioral and Experimental Economics* 80, 25-33.
- 国立国会図書館調査及び立法考査局, 2018. 人工知能・ロボットと労働・雇用をめぐる視点：科学技術に関する調査プロジェクト報告書. http://dl.ndl.go.jp/view/download/digidepo_11065186_po_20180405.pdf?contentNo=1
- 野村総合研究所, 2015. 日本におけるコンピューター化と仕事の未来. <https://www.nri.com/-/media/Corporate/jp/Files/PDF/journal/2017/05/01J.pdf?la=ja-JP&hash=6B537BB1EB48465D0AF4A3EA1B1138809F916683>
- 野村総合研究所, 2016. ロボット・AIに関する日・米・独インターネット調査. 知的資産造 2016年5月号, 108-125 <https://www.nri.com/-/media/Corporate/jp/Files/PDF/knowledge/publication/chitekishisan/2016/05/cs20160509.pdf>
- 総務省, 2016. ICTの進化が雇用と働き方に及ぼす影響に関する調査研究.
- 田口聡志, 2015. 実験制度会計論：未来の会計をデザインする. 中央経済社. 東京.
- 田口聡志, 2018. 人間とAIとが共存する未来社会のデザイン：実験社会科学, *トランス・サイエンス*, *フューチャー・デザインの融合へ向けて*. *同志社商学* 69(6), 177-202
- Zelmer, J. 2003. Linear Public Goods Experiments: A Meta-Analysis. *Experimental Economics* 6, 299-310