

リスク選好性と現在・将来バイアス性が不正の際の AI 信頼度に与える影響*

岡嶋美花^a 鎌田悠梨子^b 小林奈津実^c 中島愛衣^d

要約

本研究は、個人の意思決定に重要な影響を与えるとされる現在・将来バイアス性やリスク選好が、AI 不正を行う人にどのような影響を与えているのかについて検証する。近年の犯罪や不正の傾向として AI を使用したものが数多く上げられる。不正を防止する AI の発達が目覚ましい中でも横行する AI 不正を行う人にはどのような特徴があるのか興味を抱いたため本研究を行った。分析の結果、リスク選考は不正の際の AI 信頼度に影響を与えることが確認されたが、現在・将来バイアス性等の要素は統計的に有意な影響を示さなかった。これにより、不正の際に AI 技術を使用する理由には、他の外的要因が影響している可能性が示唆された。

JEL 分類番号： D81, D90, D91

キーワード：リスク選好性，時間選好性，AI 不正，意思決定

*なお、本論文に関して、開示すべき利益相反関連事項はない。

^a 岡嶋美花 同志社大学商学部 cgfh2091@mail3.doshisha.ac.jp

^b 鎌田悠梨子 同志社大学商学部 cgfh0191@mail3.doshisha.ac.jp

^c 小林奈津実 同志社大学商学部 cgfh0258@mail3.doshisha.ac.jp

^d 中島愛衣 同志社大学商学部 cgfh0445@mail3.doshisha.ac.jp

1. イントロダクション

1. 1. AI の現状

近年 AI 技術は目覚ましく発展し、さまざまな場面で利用されている。現代の AI には音声や画像をもとにつくられた情報から認識・予測・判断・提案などを行い問題解決するといった、自ら学習するという能力も備えられている。AI を取り入れることにより人件費の削減や事業の効率化が期待できる一方で、最近では AI を使用した不正行為が国内外で後を絶たない。東京新聞によると、実際に中国ではディープフェイクによる AI 詐欺事件が世間を騒がせた。このように AI は今後より多くの場面で活用されていくと同時に、AI を使った犯罪や不正がさらに拡大すると考える。

1. 2. 先行研究

先行研究によると、現在・将来バイアス性やリスク選好は、個人の意思決定に重要な影響を与えることが示されている。たとえば、Åkerlund et al. (2016)は、犯罪行動において個人の時間選好が重要であることを明らかにしている。これは、犯罪によって得られる即時の報酬と、将来的に負担する可能性がある逮捕や処罰などのコストとのバランスが、個人によって異なるためである。

また、リスク選好についても注目すべき研究がある。佐々木・平井・大竹(2017)によると、日本人女性の乳がん検診の受診行動に関する調査では、リスク回避的な人ほど受診率が低いことが分かった。さらに、時間割引率が高い人ほど検診を受診しにくい傾向も示された。これらの結果は、現在・将来バイアス性とリスク選好が、個人が将来的なリスクをどう評価し行動するかに関与していることを示している。

1. 3. 研究に至った考え

現在 AI 技術の発達によりインターネット上での不正が後を絶たない。他方で不正を防止する AI の発達も目覚ましく、機械学習のアプローチを用いて最先端の不正予測を行う技術精度の向上は進んでいる。(Bao et al. 2020)

その中でも AI 不正を積極的に行う人にはどのような特徴があるのか興味を抱き、AI 使用に関する意思決定において「現在・将来バイアス性」と「リスク選好性」の二つが大きく影響を及ぼしているのではないかと考えたため、本研究に至った。

AI を用いて不正をする際、どのような性質が最も人に影響を与える要因となりうるのかを明らかにすることを本研究の目的とする。そして、AI の適切な使用を促進する事に貢献したい。

以上を踏まえ、我々は以下の仮説を設定した。

仮説 1：リスク愛好的だと、不正行為を行う際 AI 技術をより積極的に頼る傾向がある

仮説 2：現在・将来バイアス性が高く近視眼的であると、不正行為を行う際 AI 技術をより積極的に頼る傾向がある

2. サーベイ実験

2. 1. 実験期間・参加者数

実験実施期間は、2024 年 8 月 22 日から 2024 年 8 月 24 日の 3 日間であり、最終的な有効回答数は 20 代から 70 代の 143 名（男性 111 名，女性 30 名，その他 2 名）となった。

2. 2. 手続き

本調査は Google Form を用いて作成し，Yahoo!クラウドソーシングにて配布を行った。参加者の不正する際の AI 技術の信頼度を求めるにあたり，ある商品の評価を上げるための不正レビュー（口コミ）を作成してもらうという場面を想定した。ある商品の現状での評判を，評判の良い商品を☆4 以上，悪い商品を☆2 以下として，現状での評価が良い商品，評価が悪い商品それぞれの評価をさらに上げる為の不正レビュー（口コミ）を作成する際に AI 技術に頼るかアナログで行うかを問う。またスピードくじに対する支払い意思額からリスク選好性（佐々木・平井・大竹，2017），報酬を得られる時間を仮定して現在・将来バイアス性を測り（盛本晶子，2018），これを説明変数として分析を行った。

3. 結果

「不正レビューを作成するために AI 技術を使用する際，どちらの商品のレビューに AI 技術を使用するのか？」という不正レビュー度合い（この数値が高ければ AI 技術への信頼度が高い，低ければ AI 技術への信頼度が低いとする）を従属変数とし，リスク選好性（この数値が高ければリスクを好む傾向がある，低ければリスクを嫌う傾向があるとする）を説明変数として単回帰分析を行った。

図 1 リスク選好性を説明変数とする単回帰分析

	係数	標準誤差	t	p-値
切片	3.055347142	0.162848438	18.7619063	2.84907E-40
リスク選好	1.28576E-05	4.80937E-06	2.673438586	0.008387505

結果として，説明変数の p 値=0.008 で係数がプラスなため，説明変数であるリスク選好性は従属変数である不正レビュー度合いに 1%水準で有意にプラスの影響を与えているこ

とがわかった。よってスピードくじに値段をかけるほど、不正レビューが増えるといえる。

また、不正レビュー度合いを従属変数とし、現在バイアス性（この値が高ければ未来の大きな利益よりも目の前の小さな利益を重視する傾向がある）と将来バイアス性（この値が高まれば現在からより遠い未来の利益よりもすこし先の未来の利益を重視する傾向がある）を説明変数としてそれぞれ単回帰分析を行った。（図2、図3）

図2 現在バイアス性を説明変数とする単回帰分析

	係数	標準偏差	t	p-値
切片	3.285976	0.468862	7.00840	8.95E-11
現在バイアス性	0.001098	0.086365	0.01271	0.98987

図3 将来バイアス性を説明変数とする単回帰分析

	係数	標準誤差	t	p-値
切片	3.44974	0.433505	7.957794	5.04E-13
将来バイアス性	-0.03157	0.081935	-0.38533	0.700568

結果として、説明変数の p 値=0.989 と、説明変数である現在のバイアス性は従属変数である不正レビュー度合いに有意にプラスの影響を与えていないとわかった。

また、将来のバイアス性を説明変数として単回帰分析を行った。結果として、説明変数の p 値=0.700 で係数がマイナスであった。よって、仮説2は棄却された。

図4 リスク選好性と AI 使用の際のモラルを従属変数とした重回帰分析

	係数	標準誤差	t	p-値
切片	2.1772	0.4964	4.3857	0.0000
リスク選好性	0.0000	0.0000	2.7746	0.0063
モラル	0.1960	0.1048	1.8707	0.0635

また、「AI 使用に倫理的な問題点を感じるか」という観点を説明変数とし、リスク選好性に加えて重回帰分析を行った。（図6）その結果、説明変数のリスク選好性の p 値は 0.0063、AI 使用に関するモラルの p 値は 0.0635 という値が算出された。よって、リスク選好性と倫理観の二つの説明変数はいずれも従属変数をほとんど説明していることがわかった。

図5 近視眼性と迅速性を説明変数とする重回帰分析

	係数	標準誤差	t	p-値
--	----	------	---	-----

切片	3.7633	0.550799	6.832439	2.40E-10
近視眼性	0.032975	0.108848	0.302943	0.762386
迅速性	-0.10524	0.121833	-0.86376	0.389206

さらに、現在・将来バイアス性から求めた近視眼的な傾向（近視眼性）と迅速性を説明変数に加えて重回帰分析を行った結果（図5）、近視眼性の係数は0.033（ $p=0.762$ ）、迅速性の係数は-0.105（ $p=0.389$ ）であり、いずれの説明変数も統計的に有意な影響を与えていないことが示された。

4. 考察

研究にあたって、「リスク愛好的だと、不正行為を行う際 AI 技術をより積極的に頼る傾向がある」、「現在・将来バイアス性が高く近視眼的であると、不正行為を行う際 AI 技術をより積極的に頼る傾向がある」と仮説1及び仮説2を設定した。

単回帰分析及び重回帰分析（図1、図2、図3、図4、図5、図6）を行い、リスク選好性及びリスク選好の際の倫理観はAI使用の際に影響を与えるという結果が得られたが、その他の現在バイアス性・将来バイアス性は目的変数をほとんど説明していないことが読み取れた。よって、AI技術を不正レビュー作成に使用する動機には、個人の時間に対する優先度や迅速性の必要性よりも、他の外的要因が強く影響していることが考えられる。

調査の結果、仮説1は確証を得られたが、現在バイアス性・将来バイアス性及び分析時に使用した、それらのバイアス性から求めた近視眼性や倫理観といった説明変数はAI技術を使用する際の意味決定時に関与しないという結果となり仮説2は棄却された。

仮説2が棄却された原因として、調査において考慮されなかった他の外的要因も、仮説2の説明力が不足した一因と推測される。

5. 研究の限界と今後の展望

本研究における限界として、スピードくじのデータ収集の時点で金額設定の幅が広すぎた為、くじの金額の上限下限を決めるべきだったと考えられる。

今後の研究では、AI技術の不正使用に対してより深めた観点から調査を進めることが重要である。本研究で有意な結果が得られたリスク選好性に加え、他の要因についても着目することで、AI不正の際の心理状態がより明らかになり、AIの不正使用を防ぐための有効な対策を見出すことにつながるだろう。

引用文献

芝正太郎, 2017. リスク選好と時間選好の統合に向けて. WINPEC Working Paper

Series No. J1701.

- 佐々木周作, 平井啓, 大竹文雄, 2017. リスク選好が乳がん検診の受診行動に及ぼす影響: プログレス・レポート. 行動経済学会第10回大会プロシーディングス 10, 1-4.
- Åkerlund, D., Golsteyn, B., Grönqvist, H., Lindahl, M., 2016. Time discounting and criminal behavior. *Proceedings of the National Academy of Sciences*, 113(22), 6160-6165. <https://www.pnas.org/doi/full/10.1073/pnas.1522445113>.
- Kipp, P. C., Curtis, M. B., Li, Z., 2020. The attenuating effect of intelligent agents and agent autonomy on managers' ability to diffuse responsibility for and engage in earnings management. *University of North Texas*.
- BAO, 2020. Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research*, Wiley Online Library.
- 東京新聞, 2023. <https://www.tokyo-np.co.jp/article/255568>.
- 盛本晶子, 2018. 時間選好率および現在バイアス性がオンラインゲーム内コンテンツへの課金行動に与える影響. 行動経済学 11, 1-13.