

プレスリリース
報道関係 各位

2026 年 10 月 1 日

株式会社国際電気通信基礎技術研究所 (ATR)

同志社大学

京都大学

人は損をしてでもズルい相手を罰する ～ズルい相手が人でも AI でも、罰する割合は同等～

人は損をしてでもズルい相手を罰する ～ズルい相手が人でも AI でも、罰する割合は同等～

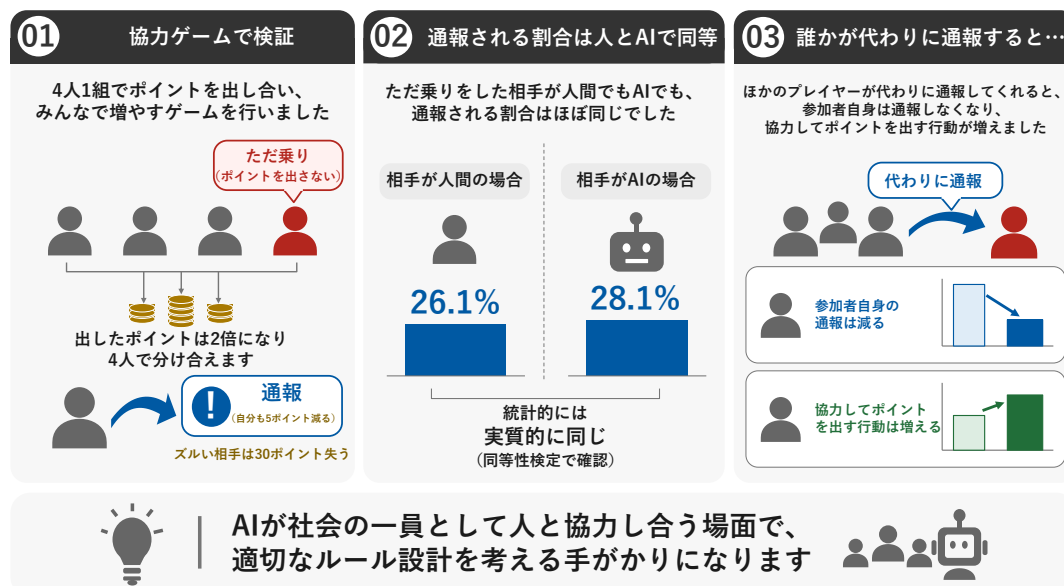


図 1 実験の概要と得られた知見

概要

ATR 深層インタラクション総合研究所のグループは、同志社大学、京都大学と共同で、「人は AI のズルい行動に対して自分が損をしてまで罰を与えるのか」を検証しました。人々の協力的行動を検証するために設計された、4 人 1 組でポイントを出し合うオンラインのゲーム実験において、みんなのために持ち点を出さない「ただ乗り」をしたプレイヤーを、自分の持ち点を減らすことと引き換えに通報できるようにしました (図 1)。「ただ乗り」をしたズルいプレイヤーが「人間」または「AI」と紹介される場合を比べると、通報された割合は人間 26.1%、AI 28.1%と、統計的に「実質的に同じ」と確認されました。つまり、損をしてでもズルい相手を罰するという反応を、人は AI に対して人間と同じように示しました。さらに、ほかのプレイヤーが代わりに通報した場合には、参加者自身は通報しなくなる一方で、持ち点を出す協力的行動は増えることも分かりました。今後、職場やオンラインコミュニティなど、人と AI が同じ集団の中で協力する場面が増えていきます。これらの知見は、ルール違反への監視や注意といった役割を人と AI のあいだでどのように分担することで集団の中での協力が保たれるのか、その設計の基礎になると考えられます。

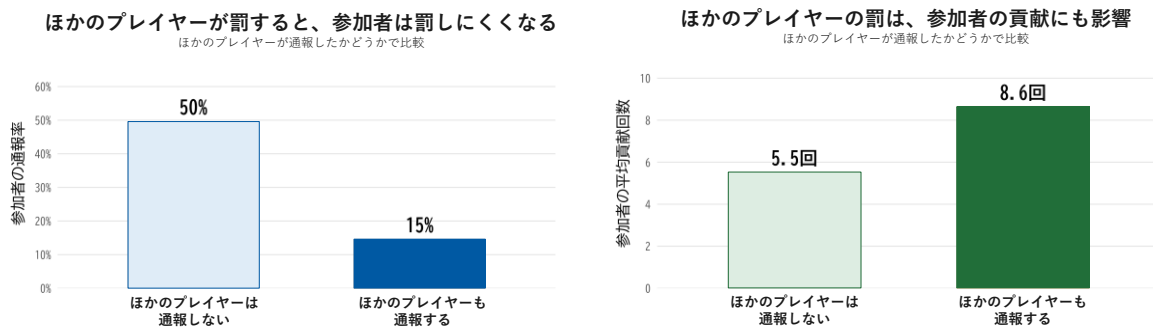


図2 ほかのプレイヤーが代わりに通報すると参加者自身は通報しなくなることを示した結果（左）と、ほかのプレイヤーが通報すると参加者は貢献する行動を増やすことを示した結果（右）

背景

集団の中には、自分は負担せずに利益だけを受け取る「ズルい」存在が現れます。こうした行動は「ただ乗り」(フリーライド)とも呼ばれます。こうした存在に対して、人は自分が損をしてでも罰を与えることがあります。これは「コストのかかる制裁」と呼ばれ、社会の協力を支えるしくみとして知られています。近年は、AI が対話の相手や仕事の仲間として人と関わる場面が急速に増え、AI が人のようにふるまい、ときにはルールから外れた行動をとることも想定されるようになりました。しかし、人同士のあいだで詳しく調べられてきたこの制裁行動が、相手が AI のときにも同じように起こるのかは、はっきりしていませんでした。そこで、私たちの研究グループは「人は AI のズルい行動に対しても、自分が損をしてまで罰を与えるのか？」という研究課題に取り組みました。

研究の狙い

これまでの研究には、相反する 2 つの見方があります。ひとつは、AI には心や責任がないと考えられるため、罰を与える対象にはなりにくいという見方です。もうひとつは、人はコンピュータや AI に対しても人と同じように反応するという見方で、この場合、相手が AI であっても人間と同じように罰が向けられると考えられます。これらの見方のもと、ただ乗りをした相手が人間の場合と AI の場合を比べたとき、コストを払ってでも通報する割合は変わらない or 変わるのではないか、という相反する 2 つの予測を立てました。

また、罰を与える役割を自分以外の誰かが担っている場合に、人の行動がどう変わるのかも分かっていませんでした。ほかのプレイヤーが代わりに通報してくれるとき、人は安心して協力を続けるのか、それとも「誰かがやってくれる」と考えて自分は手を引くのか、という違いが考えられます。さらに、その代わりに通報する相手が人間か AI によって、受け止め方が変わる可能性もあります。したがって、通報する役割を担うプレイヤーの有無やその種類によって、参加者の協力行動と通報行動が変わるのではないか、という予測を立てました。

ズルい相手が人間の場合と AI の場合を比較する実験とその結果について

この予測を調べるため、オンラインの公共財ゲームを用いた実験を行いました。参加者は 4 人 1 組のうちの 1 人となり、残る 3 人はあらかじめ決められた動きをするプログラムで、それぞれ「人間」または「AI」として紹介されました。ゲームは全 10 ラウンドで、参加者は毎ラウンド 20 ポイントを受け取り、みんなのためにポイントを出すかどうかを選びます。出されたポイントの合計は 2 倍になって 4 人に等分されるため、全員が出せば全員の得が最も大きくなりますが、自

分だけ出さなければその人だけがさらに得をします。参加者は、このズルをしたプレイヤーを、自分の持ち点を 5 ポイント減らすことと引き換えに通報でき、通報された相手は 30 ポイントを失います。分析対象となった 588 名の結果から、通報された割合は相手が人間で 26.1%、AI で 28.1%であり、統計的に実質的に同じと判断されました。ズルい相手が人間か AI かも、そばで見ているだけのプレイヤーが人間か AI かも、通報の起こりやすさには影響していませんでした。はっきりと影響したのは、ほかのプレイヤーが代わりに通報してくれるかどうかだけで、代わりに通報してくれる場合、参加者自身の通報は減り、協力してポイントを出す回数は増えました。一方、実験後のアンケートでは、代わりに通報したのが人間だった場合に「ズルい行動は罰せられるべきだ」という考えがより強く、AI だった場合には「ズルい行動を許せる」という気持ちがより強くなっていました。行動としては同じでも、気持ちの面では相手が人間か AI かで違いが表れたといえます。

今後の展開

本成果は、人が AI に対しても、自分が損をしてまで罰を与えるという点で、人間相手と変わらない反応を示すことを明らかにしました。これは、AI が社会の一員として扱われつつある一方で、ズルをすれば人と同じように制裁を受けうることを意味します。また、AI が代わりにルール違反を指摘する役割を担うと、人はその役割から手を引きやすくなることも分かりました。この結果は、職場やオンラインコミュニティなどで監視や注意の役割を AI に任せる場合に、こうした人の側の振る舞いの変化をあらかじめ見込んでおく必要があることを示しています。ルール違反への監視や注意といった役割を人と AI のあいだでどのように分担すれば集団の協力が保たれるのかは、今後の社会設計にとって重要な論点になります。なお、本実験は日本の参加者を対象としており、罰に対する考え方は文化によって異なることが知られています。今後は、ほかの文化圏との比較や、実際の職場やコミュニティに近い場面での検証を進め、人と AI の望ましい役割分担のあり方を明らかにしていきます。

【掲載論文】

題名 : Comparable costly punishment toward AI and human free-riders among Japanese participants: Evidence from a public goods game

(日本人参加者における AI と人間のフリーライダーに対するコストのかかる制裁 : 公共財ゲームによる検証)

著者 : Koshiro Tanaka, Takamasa Iio and Masahiro Shiomi

田中 康士朗 (ATR、同志社大学)、飯尾 尊優 (同志社大学、ATR)、塩見 昌裕 (京都大学、ATR)

掲載誌 : Computers in Human Behavior: Artificial Humans

掲載日 : 令和 8 年 7 月 16 日 (日本時間) オンライン版公開

DOI : 10.1016/j.chbah.2026.100361

【研究支援】

本研究は、JST ムーンショット型研究開発事業 JPMJMS2011、科研費 JP24K21327 の研究プロジェクトの一環として実施されました。

【実験手順の詳細について】

本研究では、ただ乗りをしたプレイヤーの提示（人間・AI）、代わりに通報するプレイヤーの有無と提示（なし・人間・AI）、そばで見ているだけのプレイヤーの提示（人間・AI）という3つの要因、計12条件を参加者間で比較しました。実験はオンラインで実施し、参加者は4人1組のゲームに唯一の人間として参加しました（残る3名はプログラムによるもので、それぞれ人間またはAIとして紹介されました）。本番の10ラウンドでは、各ラウンドで参加者に20ポイントが与えられ、30秒以内に提供するかどうかを選びました。提供されたポイントの合計は2倍になって4人に等分されます。提供したラウンドでは、提供しなかったプレイヤーを自己負担5ポイントで通報でき、通報された相手は30ポイントを失いました。ゲーム終了後、ズルい行動をしたプレイヤーへの印象などをアンケートで評価しました。実験には632名が参加し、588名（男性297名、女性288名、回答しない3名、平均年齢43.4歳）を分析対象としました。三元配置分散分析の結果、通報の起こりやすさに影響したのは代わりに通報するプレイヤーの有無と提示のみであり、ズルい行動をしたプレイヤーやそばで見ているプレイヤーが人間かAIかは影響しませんでした。さらに同等性検定により、ズルい行動をしたプレイヤーが人間の場合とAIの場合の通報行動は実質的に同等であることが示されました。

【本件に関するお問い合わせ先】

■株式会社国際電気通信基礎技術研究所(ATR)企画・総務部 企画部門 企画チーム

Email : pr[*]atr.jp （[*] を@に変えて使用してください）

■同志社大学 学長室 総合企画部 広報課

Email : ji-koho[*]mail.doshisha.ac.jp （[*] を@に変えて使用してください）

■京都大学 総務部 広報課

Email : jkoho52[*]mail2.adm.kyoto-u.ac.jp （[*] を@に変えて使用してください）