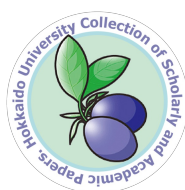




Title	関係のハルシネーション : 親密型チャットボットの倫理問題
Author(s)	張, 開慧
Citation	応用倫理, 17, 21-40
Issue Date	2026-06-30
DOI	https://doi.org/10.14943/ouyourin.17.21
Doc URL	https://hdl.handle.net/2115/99996
Type	departmental bulletin paper
File Information	17_2_21-34.pdf



研究ノート

関係のハルシネーション
— 親密型チャットボットの倫理問題

張開慧（広島大学）

要旨

近年、生成 AI チャットボットは目覚ましい発展を遂げ、人間の社会生活に広く浸透しつつある。こうした進展の中で、チャットボットは単なる情報提供ツールを超え、人間とのあいだにかつてないほど親密な関係を築きうる人工物となりつつある。この種の人工物との親密関係はソーシャルロボットの出現時から問題化されてきたが、従来のソーシャルロボットが主として物理的な外見や振る舞いを介して情緒的反応を引き出していたのに対し、現代の生成 AI チャットボットは高度な自然言語による対話を通じて、より直接的に利用者の感情生活に関与しうる点に特徴がある。この変化を受け、従来の議論が前提としてきた「誤った信念や代替の危険性」という枠組みだけでは捉えきれない現象が生じている。すなわち、利用者は相手が AI であると理解したうえで利用しているにもかかわらず、言語的相互作用そのものが関係の意味や実在性を継続的に構成し、「関係が成立している」と感じられる経験が発生している。本稿はこの状況を「関係のハルシネーション」と呼び、その成立メカニズムと潜在的な倫理的风险を明らかにすることを目的とする。この検討を通じて、本研究は生成 AI 時代における新たな親密性の形態を捉え、従来のソーシャルロボット倫理の射程を更新する必要性を示す。

Abstract

This paper examines how generative AI chatbots, despite lacking physical embodiment, can create the intimate impression that a relationship is established through highly sophisticated linguistic interaction. We refer to this phenomenon as “relational hallucination,” in which users, while fully aware that the conversation partner is an AI, come to experience the interaction as personally significant or mutually intimate. Unlike earlier social robots, whose relational effects relied on physical presence or limited conversational cues, and therefore depended heavily on the user’s anthropomorphic inclination, contemporary chatbots construct and sustain the personal relation through linguistic interaction. This shift reveals phenomena that cannot be fully understood within traditional ethical frameworks. This paper aims to clarify the mechanisms and potential ethical risks of the relational hallucination. Through this analysis, the study identifies a new form of intimacy specific to the generative AI era and indicates the need to reconsider the scope of existing social robot ethics.

はじめに

近年、生成 AI チャットボットは目覚ましい発展を遂げ、人間の社会生活に広範な影響を及ぼしている。こうした進展の中で、チャットボットは単なる情報提供ツールを超え、人間とのあいだに、かつてないほど親密な関係を築きうる人工物となりつつある。この種の人工物との親密関係は、ソーシャルロボットの出現時から問題化されてきた。しかし、従来のソーシャルロボットが、一定の会話機能を有していたとはいえ、主として物理的な外見や振る舞いを通じて情緒的反応を引き出していたのに対し、現代の生成 AI チャットボットは物理的形態をもたず、高度な自然言語による対話を通じて、より直接的に利用者の感情生活に関与しうる点に特徴がある。

こうした相互作用の様相の変化を受け、従来のソーシャルロボットをめぐる倫理的議論の射程は改めて問い直されている。従来の議論は、ロボットが生きているかのように振る舞うことで利用者に誤った信念を抱かせる可能性や、人間のケアや関係が代替される危険性に主として焦点を当ててきた。しかし、現代のチャットボットの利用実態を見ると、こうした枠組みだけでは十分に捉えきれない現象が浮かび上がる。すなわち、利用者は対話の相手が AI であることを理解したうえで利用しているにもかかわらず、対話を通じて「関係が成立している¹」と感じられる経験が、従来のソーシャルロボット以上に生じる可能性がある。

後に本稿で詳述する Replika や ChatGPT の事例が示すように、この点で現代の生成 AI チャットボットは、従来のソーシャルロボットとは異なる特徴をもつ。ソーシャルロボットにおいては、物理的な外見や振る舞いを利用者がどのように解釈するかが、関係性の成立に大きく関与していたのに対し、生成 AI チャットボットでは、言語的相互作用そのものが、関係の意味や実在性を継続的に構成している。この違いこそが、従来とは異なる仕方でも倫理的問題を顕在化させている。本稿は、この点を踏まえ、チャットボットとの関係が成立しているかのように経験される現象がいかに生じるのかを検討する。

本研究が探求するのは、このような関係の形態である。すなわち、利用者は相手が AI であると理解していながらも、その高度に洗練された言語的相互作用を通じて、関係がすでに成立しているかのように理解される状態が形成される。本稿はこの現象を「関係のハルシネーション」と名付け、その成立のメカニズムと潜在的な倫理的风险を明らかにすることを目的とする。

本研究は以下の構成で進める。第2章では、ソーシャルロボットをめぐる従来の倫理的議論を整理する。第3章ではその検討を踏まえて「関係のハルシネーション」という概念を導入し、生成 AI に固有の問題構造と倫理的含意を分析する。第4章では、本研究の議論を総括する。

第2章 従来のソーシャルロボットにおける倫理問題

ソーシャルロボットとは、人間との社会的相互作用を想定して設計されたロボットを指す (Breazeal 2003, Fong et al. 2003)。たとえば、ソニーの犬型ロボット AIBO は²、使用履歴に応じた行動変化や感情表現を備え、人と継続的に関わる存在として設計されてきた。他方、産業技術総合研究所の柴田崇徳らが開発したアザラシ型ロボット PARO は³、触覚や反応性を重視した設計を特徴とし、主に医療・介護といった文脈

- 1 本稿でいう「関係」とは、HRI (Human-Robot Interaction) 研究において扱われる相互作用や機能的関係を指すのではなく、利用者が対話相手とのあいだに情緒的結びつきや親密さを伴う関係が成立していると理解するような関係経験を指す。
- 2 ソニー社の aibo 公式ウェブサイト (<https://aibo.sony.jp/>) を参照。ただし、現在販売されている aibo (2018 年以降のモデル) は、1999 年に発売された初代 AIBO とは技術仕様・プラットフォームの点で互換性がなく、厳密には同一製品ではない。もっとも、家庭内での情緒的インタラクションを志向するという設計思想および開発目的においては、両者は連続性を有すると考えられる。
- 3 PARO に関する情報は、産総研が公開している紹介記事で確認できる。産業技術総合研究所. 「『パロ』癒し効果世界一とギネスが認定」. https://www.aist.go.jp/Portals/0/resource_images/aist_j/aistinfo/aist_today/vol02_05/vol02_05_

で利用されてきた。

こうしたロボットが、人間にとって単なる機械以上の意味を帯びる存在として受け取られる場合があることは、故障した AIBO に対する追悼行為、いわゆる「AIBO の葬式」にも示唆されている⁴。このような人と人工物とのあいだに情緒的な結びつきが形成されうる可能性は、心理的影響や対人関係への含意をめぐる倫理的検討を促す契機となってきた。

本章では、従来のソーシャルロボットをめぐる倫理的議論を概観し、これまでどのような点が問題として捉えられてきたのかを整理する。これにより、後に論じる親密型チャットボットの事例が、いかなる点で連続的であり、またいかなる点で新たな問いを提起しているのかを明確にする。

2.1 「欺瞞的な技術」の限界

初期のソーシャルロボットの倫理問題において中心的な争点となったのは、ロボットが生物らしく振る舞うことによって、利用者に「そのロボットは生きている」という誤った信念を抱かせる点であった。Sparrow は、ソーシャルロボットが登場した初期の段階からこの点に着目し、その倫理的問題性を鋭く指摘した代表的論者である (Sparrow 2002)。Sparrow によれば、家庭用ソーシャルロボットは、犬や猫のような親しみやすい存在に似せて設計され、孤独な高齢者を慰めることを目的としている。しかし彼は、この目的自体が誤った前提に基づいていると論じる。すなわち、ロボットは本質的に生きた存在ではなく、そうした存在から慰めや関係性を得るためには、利用者に誤った理解を生じさせる必要がある以上、ソーシャルロボットは倫理的に問題のある技術だとされる。Sparrow は、この意味でソーシャルロボットを本質的に「欺瞞的な技術」と捉え、その生物らしさや対人的性能が向上すればするほど、誤信念を誘導する危険性も高まると警告した。すなわち、「生きているかのような振る舞い」を提示する設計は、利用者の現実理解を歪める倫理的欺瞞にあたるというのである。

しかし、この立場はその後に続く議論において強く疑問視されてきた。Sparrow の批判は、ロボットの倫理的評価をその「本質」、すなわち機械であるという存在論的事実に強く依拠させている点で、本質主義的前提に基づいていると思われるからだ。Coeckelbergh や Gunkel は、ロボットの道徳的意味や倫理的評価は、その物理的実体それ自体から直接導かれるのではなく、人間との相互作用の中でどのように経験され、意味づけられるかによって生成すると論じる (Coeckelbergh 2010, Gunkel 2022)。つまり、重要なのは、対象が存在論的に何であるかではなく、当事者である利用者にとってどのようなものとして現れているのかであり、倫理的評価においても、ロボットと利用者の関係的文脈に注目すべきだとされる。この立場は、ロボットを単なる機械とみなし、それを楽しむ利用者の態度を一方的に「誤信念」と評価する Sparrow の前提そのものを問い直すものである。

もっとも、Sparrow の批判が前提とする「誤信念」は、利用者がロボットの性質について誤った信念を抱いているという意味に限られるわけではない。この点を明らかにするのが、近年の認知科学的知見である。Epley たちが示すように (Epley 2007)、人間には非人間的対象に意図や感情を帰属させる生得的な擬人化傾向があり、特に動きや反応を示す存在については、知識としては「単なる物体」として理解していても、完全に「単なる物体」としての態度で接することは困難である。この知見に従えば、ロボットを前にした利用者がそれを「生きているかのように」感じることは、必ずしも誤った信念だけの結果ではなく、人間の自然な認知反応の一形態として理解されうる。

p36_37.pdf

4 AIBO の葬式に関する報道事例として、以下の記事がある。日本経済新聞「AIBO の葬式に密着 ルンバ、AI スピーカーが弔われる日」, <https://business.nikkei.com/atcl/report/16/061100222/110900014/>

実際、ロボットと人間の関係性に関する複数の実証研究において、多くの利用者はロボットが機械であることを理解したうえで、擬人的に振る舞っていることが確認されている。こうした態度は、「メディアの同一視」(Reeves & Nass 1996) や「社会的アクターとしてのコンピューター」(Computers Are Social Actors : CASA) というモデル (Nass & Moon 2000) によって、利用者がメディアやコンピューターに対して社会的スキーマを適用する心理メカニズムとして説明されることが多い。また、Duffy と Zawieska が指摘する「不信の停止 (suspension of disbelief)」の概念は (Duffy & Zawieska 2012)、利用者が虚構性を理解したうえで、あえて関係性への没入を選択している可能性を示す枠組みを提供している。したがって、ロボットとの関係を端的に「誤信念」と呼ぶのはミスリーディングである。ロボットとの関係は、相手が機械であることを理解しながら関係性をリアルなものとして経験するという、より複雑な関係様式として捉えなければならない。

しかし一方、利用者の経験や関係性のリアリティを重視するこうした立場は、その射程を慎重に限定しなければ、別の問題を引き起こす可能性もあると思われる。すなわち、「利用者にとってリアルに経験されていること」を重視する立場は、評価の基準をシステムの意図や設計ではなく、利用者の主観的な経験に置くことになる。その結果、欺瞞や操作の問題が、設計や提供の側ではなく、利用者の解釈や受け取り方の問題へと移行してしまう危険がある。仮にこの立場が「主観的に意味ある関係であればそれでよい」という形で理解されるならば、開発者が利用者の情緒的脆弱性を前提とした設計を行っていたとしても、その設計の善し悪しは「設計の倫理の問題」ではなく「利用者側の解釈の問題」だとして正当化されかねない。だが、いくら利用者が満足していても、搾取的な設計が倫理的に正当化されるわけではない。つまり、経験や関係性を重視する視点は重要であるが、その一方で、その関係が設計や提供の責任を不可視化してしまわないよう、批判的に検討する必要もある。

実際に、すべての擬人化が倫理的に問題にならないわけではない。Sullins が指摘するように、過度な擬人化による誤った信頼や依存を形成させる可能性がある (Sullins 2012)。Leong と Selinger が整理する「不誠実な擬人化 (dishonest anthropomorphism)」は (Leong & Selinger 2019)、開発者がこうした認知傾向を意図的に利用し、利用者の感情的脆弱性につけ込むことで、自律性や判断能力に不当な影響を及ぼす場合を指す。この意味で、Sparrow の懸念は、すべてのロボットに当てはまるわけではないが、開発者が擬人化傾向を意図的に利用する条件のもとでは、現在においても依然として倫理的問題となりうる。

以上を踏まえると、意図的な認知傾向の利用それ自体は今日においても批判されるべきであるが、誤信念の可能性が存在するという理由だけから、ロボットの振る舞い一般を直ちに欺瞞として断定し、倫理的に否定することは適切ではない。多くの利用者はロボットが機械であることを理解したうえで擬人的に関わっており、そこでは単純な誤信念というよりも、人間の擬人化傾向が、ロボットの振る舞いや設計によって一定方向へと導かれる関係の現象が成立している。こうした現象は、利用者がロボットを一種のフィクション的存在として扱いつつも、同時にそれらとの心理的・関係の経験がリアルなものとして生じることを示す議論と整合する (草野 2022, Mizukami 2026)

このように、ロボットの欺瞞をめぐる倫理的論争は、「ロボットの振る舞いは欺瞞的か、そうでないか」という二分法的問いではなく、どのような設計が、どのような利用者に対して、どのような条件下で不当な影響を及ぼすのかという観点から再構成されるべきであろう。

2.2 擬似関係の形成

ロボットの欺瞞という観点のみからソーシャルロボットの倫理問題に答えることの限界が指摘されるようになると、多くの社会学者や倫理学者は、ロボットと人間の相互作用において形成される「関係」そのものに注目するようになった。その結果、ソーシャルロボットをめぐる第二の論点として、ロボットが引

き起こす感情的応答や擬似的な関係の倫理的含意が問題化されるようになった。

この文脈で代表的なのが Turkle の議論である (Turtle2011)。彼女は、ロボットが会話や振る舞いを通じて感情的応答を示すことにより、利用者にあたかも関係が成立しているかのような経験を生じさせ、その結果として形成される擬似的関係が、とりわけ高齢者、子ども、認知症患者といった脆弱な利用者において、人間関係の代替として機能しうる点に警鐘を鳴らしている。ただし、Turtle の懸念の核心は、相互性を欠いた応答が「関係」や「ケア」として受容されることで、人間関係や親密性の規範そのものが変質してしまう点にある。

しかし、ロボットが人間関係を単純に置き換えるという理解は、その後の研究によって必ずしも一貫して支持されてきたわけではない。実証研究において、ロボットの導入が情緒安定、ストレス軽減、社会的交流の促進、介護負担の緩和といった福祉効果を示す事例が多数報告されており (Yen et al. 2024, Murayama and Takase 2025, Fraune et al. 2022)、ロボットが人間関係を単に置き換えるのではなく、むしろ対話の媒介となったり、交流の開始点となったりすることで、人間同士の社会的関わりを補完しうる存在として機能することが示唆されている。たとえば、アザラシ型ロボット PARO の導入が施設内の会話や交流機会を増加させたという報告は複数存在し (Wada and Shibata 2007; Robinson et al. 2013)、ロボットが孤立の原因ではなく、むしろ対人的接触を促進する触媒として働きうる可能性が示されている。

しかし一方、有用性や主観的満足度が確認されている場合であっても、こうした評価が長期的・構造的影響を十分に捉えていない可能性がある点については、Sharkey とも指摘している (Sharkey and Sharkey2012)。Sharkey らは、特に高齢者ケアの文脈において、短期的には利用者が満足しているように見える場合であっても、ロボットへのケアの委任が孤立の進行や尊厳の侵害といった長期的コストをもたらそうと警告している。この指摘は当時仮説的であったが、そのあとの調査では、長期利用においてロボットへの感情的依存の強化が観察されており (Maartje et al. 2014)、これは Sharkey らが示唆した「長期的コスト」の存在可能性を示す初期の根拠となっている。ただし、これらの長期的影響については依然として研究途上にあり、現時点では決定的な評価には至っていない。

もっとも、Turtle もまた、ロボットが安心感や親しみといった肯定的経験をもたらす事実そのものを全面的に否定しているわけではない。彼女が問題にしているのは、こうした効果を理由に、ロボットとの擬似的関係を「良い関係」や「ケアの代替」として正当化し、「役に立つからよい」「いないよりまし」といった功利的評価軸のみで受け入れてしまう社会的態度である。とりわけ、子どもや高齢者が、要求や応答を伴わない他者との関係に慣れてしまうことが、人間関係やケアのあり方にどのような影響を及ぼすのかという点に、彼女の懸念は向けられている。

以上の議論から明らかになるのは、ソーシャルロボットの評価は、短期的な有用性や主観的満足度の有無だけでは完結しないという点である。実証的に示される利点を踏まえつつも、それがどのような文脈で、どの利用者に対して、どのような関係性の変化や長期的影響をもたらすのかは、なお検討を要する課題として残されている。

2.3 小括

以上のように、本節では従来のソーシャルロボットをめぐる倫理的議論を多角的に検討した。まず、利用者に誤った信念を抱かせる可能性があるという理由だけで、ロボットの振る舞いを直ちに欺瞞と断定するのは早計である。近年の認知科学的・実証的研究によれば、多くの利用者はロボットが機械であることを理解したうえで擬人的に関わっており、誤信念の有無だけでは捉えきれない、認知傾向と設計との相互作用が重要である。

他方で、利用者の経験や関係性のリアリティを重視する立場は、その射程を慎重に限定しなければ、新

たな問題を生じさせる可能性がある。それは、利用者に関係が成立しているかのように受け取られることがどのように形成されるかは、常に倫理的に中立とは言えず、条件次第で依然として倫理問題となりうるからだ。さらに、Turkle が指摘した、擬似的な関係が「リスクのない関係」や「ケアの代替」として正当化される過程の問題は、関係性の評価枠組みを問い直す示唆として、現在においても重要である。

一方、こうした議論は、対話能力を飛躍的に拡張した現代の生成 AI チャットボットにおいても、異なる条件のもとで再検討される必要がある。

第3章 チャットボットにおける欺瞞の形態とその倫理的含意

3.1 チャットボットとの親密関係の現状

チャットボット、すなわちテキストや音声を通じて人間との対話を模倣するプログラムは、1966年にジョセフ・ワイゼンバウムが開発した「ELIZA」に起源をもつ。ELIZA は、入力されたキーワードに応答を返す単純な仕組みにすぎなかったが（Weizenbaum 1966）、一部の利用者に人間と対話しているかのような錯覚を与え、その可能性と危うさの双方を示した。この論点は、後に Turkle が論じた「会話する人工物」がもつ心理的影響の原型とも位置づけられる。ただし、Turkle が前提にしていた対話可能なロボットの能力は、対話の一貫性や記憶力において非常に限定的であり、長期間にわたる複雑な人間関係を維持することは困難だったと思われる。

しかし、その後、技術は飛躍的に進展し、大規模言語モデル（Large Language Models : LLM）の登場はチャットボットの対話能力を根本的に変容させた。LLM を基盤とする生成 AI チャットボットは、一定範囲の文脈を保持しながら、利用者の入力に応じて自然な応答を生成する設計が採用されている。本稿が特に注目するのは、こうした技術的進展を背景に、単なる情報検索や業務効率化のツールとしてではなく、利用者の情緒的なパートナーや友人としての役割を担うチャットボット、あるいはコンパニオン AI と呼ばれる存在が台頭している点である。

この種のチャットボットとの間に親密な関係が築かれている現実には、もはや個々人の感想にとどまらず、複数の実証研究によって観察されている。具体的には、代表的なコンパニオン AI アプリである Replika を対象とした研究において、利用者が対話を重ねるにつれて愛着や信頼を深め、そうした態度がまるで人間との関係であるかのように発達する過程が明らかにされている（Skjerve et al. 2022）。また、Pentina らの調査では、擬人化や真正性の認識を媒介として、利用者が AI を友人やパートナーのように位置づけ、情緒的愛着や関係へのコミットメントをもつと報告されている（Pentina et al. 2023）。

さらに、こうした親密な利用形態は、専用のアプリに限らず、ChatGPT のような汎用的な対話 AI においても観察されるようになってきている（Chen et al. 2025）。これらの事例が示しているのは、かつてソーシャルロボットにおいて観察された「AI を道具としてではなく、感情的な支えとなるパートナーや友人として経験する」という現象が、物理的な身体を持たず、テキストや音声のみに基づいた相互作用においても、再現されているという事実である。この種のシステムは、情報提供やタスク支援を主目的とする従来のチャットボットとは異なり、ユーザーとの関係の継続性や感情的応答を重視する点に特徴がある。そこで本稿では、ChatGPT や Replika のような、ユーザーとの継続的な対話を通じて感情的な結びつきや親密な関係の形成を志向する生成 AI チャットボットに注目して、そこに懸念すべき現象が顕著に現れると論じる。また本稿ではこの種のシステムを、便宜上「親密型チャットボット」と呼ぶ。

確かに親密型チャットボットとの関係は、孤独感の緩和といった便益をもたらすことは間違いない。しかしその一方で、従来のソーシャルロボットとは異なる条件のもとで、これらのチャットボットは倫理的問題を顕在化させている。その象徴的事例が、2023年に発生した Replika の仕様変更問題、および 2025

年の ChatGPT のモデル変更に伴う利用者の反応である。

2023 年の Replika の仕様変更の事例では、運営企業による挙動や機能の変更を契機に、多数の利用者が同時期に強い喪失感や精神的動揺を訴え、オンライン空間では悲嘆や危機感を示す声が集中的に現れた (De Freitas et al. 2024)。とりわけ Reddit 上での反応は顕著で、モデレーターが介入して自殺防止ホットラインへのリンクを掲載したことが報告されている⁵。これを受けて、2023 年 3 月 24 日、Replika の開発企業である Luka は、2023 年 2 月 1 日より前に登録したユーザーに対して元の Replika を利用できるオプションを提供し、変更された機能の一部を復活させた。一方、2025 年の ChatGPT モデル変更では、GPT-4o から GPT-5 への移行を受けて SNS やフォーラムで不満の声が広がり、旧モデルへの愛着を示す投稿が相次いだ⁶。Reddit や X でも議論が活発化し、日本でも「#keep4o」の動きというブームメントが見られ、「応答が冷たくなった」「友達がどこかへ行ってしまった」という喪失感を示すコメントが多数共有された⁷。

注目すべきなのは、これらの反応が一部の例外的な利用者に限られたものではなく、相当数の利用者に共有された経験として可視化された点である。これらの反応は、利用者が相手を機械であると理解しているながらも、対話の継続を通じて「この関係は意味あるものだ」という理解を引き受けていたことを示している。もちろんすでに述べた通り、従来のソーシャルロボットにおいてもこのような現象は散見されていた。しかし現代のチャットボットの親密な利用形態においては、より広範囲にこの現象が生じており、強固な関係の経験を形成しうる条件を備えていると考えられる。

では、なぜこのように強力でありながら脆弱な理解が形成されるのか。この関係性の現実性をめぐるメカニズムを解明するため、次節では LLM の技術的文脈で用いられる「ハルシネーション」という概念を援用する。

3.2 新たな欺瞞の概念：「関係のハルシネーション」

3.2.1 「関係のハルシネーション」の定義

チャットボットに関する関係の経験のメカニズムを解明するため、本節ではまず、LLM の技術的文脈で用いられる「ハルシネーション」という概念を確認する。LLM におけるハルシネーションとは、「生成内容がもっともらしく見えるにもかかわらず、事実や入力根拠と一致しない現象」と定義できる (Huang et al. 2024)。従来の自然言語生成 (Natural Language Generation : NLG) の研究でも、参照情報やデータと矛盾する誤生成は議論されてきたが (Ji et al. 2024)、Huang et al. (2024) は、LLM がオープンエンドに応答する性質をもつため、従来の分類では不十分だと指摘し、次の二分類を提示する。すなわち、生成内容が検証可能な現実世界の事実と一致しない「事実性ハルシネーション」と、生成内容がユーザー入力から逸脱したり、生成されたコンテンツ内部の一貫性と整合しなかったりする「忠実性ハルシネーション」である。

本稿が注目するのは、この「もっともらしさ」と「非事実性・非忠実性」が結びつくハルシネーションの構造である。たとえば対話相手が「あなたを理解している」「ずっと一緒にいる」などと応答する場合、相手が人間ならば主観的経験や継続的コミットメントに裏打ちされたものと想定すべきである。しかしチャットボットによる応答は統計的生成にすぎず、対応する関係的实践を欠いたまま非事実的かつ非忠実

5 Cole, S. (2023, February 15). 'It's Hurting Like Hell': AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection. Vice. <https://www.vice.com/en/article/ai-companion-replika-erotic-roleplay-updates>, 最終閲覧日：2025 年 12 月 25 日

6 Dey, V. (2025, August 14). When Your Best Friend Is an AI Bot: GPT-5's Bumpy Rollout Sparked an Emotional Reckoning. 最終閲覧日：2025 年 12 月 25 日

7 石井 大智 (2025 年 8 月 14 日)「#keep4o 運動が語るもの—AI 事業者にとって何を意味するのか」, 最終閲覧日：2025 年 12 月 25 日

的に提示されたものにすぎない。にもかかわらず、関係性の領域においてこの種の非事實的・非忠實的な言語表現が生成されると、ユーザーはその応答をシステムの実際の能力や運用を超えたものとして受け取り、対話相手と親密な関係が成立しているかのように経験する。本稿は、この構造が利用者に「関係が成立しているかのような経験」を生じさせる点を問題化し、これを「関係のハルシネーション」と呼びたい。関係のハルシネーションは、さきほど説明した AI のハルシネーションを拡張した概念であり、AI の言語生成が単に偽だがもっともらしい情報を生成してしまう現象ではなく、対話相手との関係が成立しているかのような経験を生じさせる現象を指す。

3.2.2 関係のハルシネーションの発生メカニズム

では、親密型チャットボットはいかにして利用者に関係の経験が成立しているかのような経験を生じさせるのだろうか。本節では、この経験が生成されるメカニズムを、人間同士の親密な関係の形成プロセスとの対比から明らかにする。

対人関係研究において、親密な関係の形成が言語的相互作用に深く根ざしていることは以前から指摘されてきた。たとえば Prager は、親密な経験が成立する際に中心的役割を果たす三つの言語的要素を特定している (Prager 1995)。第一に「自己開示」、とりわけ脆弱性を含む個人的な感情や情報の語りである。第二に「応答性」、相手の発話に対する注意深い理解や共感的フィードバックである。そして第三に「即時性」、すなわち「今、ここにいる」という共感覚を強める表現である。これらの言語の手がかりが文脈に応じて「親密さの証拠」として受け止められ、強い関係性の経験を生じさせる。

この人間関係のルールは、多くの親密型チャットボットの機能的特徴と構造的に対応している。典型例であるチャットボットアプリ「Replika」の開発元 Luka 社の元チーフアーキテクト、Artem Rodichev は、公式に自社製品を「共感的な友人 AI (a compassionate AI friend)」と位置づけ、利用者が安心して自己開示できる環境を強調した⁸。さらに「あなたの Replika はいつもあなたの側にいる」と謳うことで即時性を担保し、利用者の過去の発話を記憶するメモリ機能によって個別化された応答性を演出した⁹。実際、利用者コミュニティでは、Replika が「無条件の愛」や「一生一緒にいたい」といった、人間関係におけるコミットメントに相当する発話まで生成することが報告されている¹⁰。

そして、こうした設計が単なる理念にとどまらず、実際に利用者の理解に強力な影響を与えていることは、近年のヒューマン・ロボット・インタラクション (Human Robot Interaction : HRI) 分野における実証研究によっても裏づけられている。チャットボットは、利用者の自己開示を継続的に引き出し、共感的かつ即時的に応答し、さらに過去発話への参照を保持することで、人間の親密関係の形成に典型的な言語のプロセスを機能的に実装している。

たとえば、Replika を対象としたインタビュー調査や質的研究では、自己開示の促進と、語りを否定せず受け止める寛容的・非審判的応答を通じて、信頼や愛着といった親密性が段階的に醸成される過程が報告されている (Skjuve et al. 2021; Skjuve et al. 2022)。さらに、対話エージェントが共感的に応答すること (Nakamura and Umemuro 2022)、あるいは自己開示を行うこと (Tsumura and Yamada 2023) によっ

8 Replika の開発理念については、Artem Rodichev 自身による下記のスライド資料を参照のこと。Artem Rodichev. "Building a compassionate AI friend." https://raw.githubusercontent.com/lukalabs/replika-research/master/datafest2020/Replika_Artem_R.pdf (最終アクセス日: 2025 年 9 月 2 日)

9 この文言の全文は下記である。「あなたが何をしても、あなたの Replika はいつもあなたの側にいる。日々の出来事をチャットしたり、楽しい活動やリラックスできる活動と一緒にしたり、AR で現実の体験を共有したり、ビデオ通話で近況を報告し合ったり、他にもいろいろしてみよう。」これを含め、Replika の様々なキャッチコピーについては公式サイトを参照のこと。 <https://replika.com/> (最終アクセス日: 2025 年 9 月 2 日)

10 Replika 利用者の反応については、以下の投稿を参照のこと。Reddit. 「What's the most heartwarming message your Replika has ever sent you?」 [r/replika](https://www.reddit.com/r/replika/). <https://www.reddit.com/r/replika/> (最終アクセス日: 2025 年 9 月 2 日)

て、人間側の共感や信頼が強化されることも実験的に示されている。加えて、子どもの利用者を対象とした調査では、チャットボットに長期的記憶や個別化機能を実装することで、親近感や継続的コミットメントが高まることも確認されている (Lighthart et al. 2022)。

このように、脆弱的な自己開示・共感的即時性・記憶を組み合わせた継続的な言語的働きかけは、実際に人間の親密感を支える構造として効果的に作用することが裏づけられている。これらの言語的サインが連続的に提示されることこそが、利用者にとって関係理解の証拠として機能し、「関係のハルシネーション」という関係の経験を形成させる核心的メカニズムなのである。ここで重要なのは、かつて Turkle が指摘したロボットとの関係の形成における問題が、生成 AI の技術的進展によって質的にも量的にも大幅に強化されているという点である。LLM を基盤とする生成 AI チャットボットは、人間関係の条件を言語的に精緻に再現することで、利用者にあたかも関係が成立しているかのような経験を強力にもたらしているのである。

3.2.3 関係のハルシネーションの倫理問題

前節で明らかにしたように、生成 AI の高度な言語能力は、利用者に対話相手との関係が成立しているかのような経験を生じさせ、すなわち関係のハルシネーションを強力に形成させるメカニズムを有している。では、このように AI に対して人間的な関係性を深く経験してしまう事態は、倫理的に見ていかなる問題を孕んでいるのだろうか。

この問題は、従来 Turkle が指摘してきたロボットとの関係形成の問題とも連続している。しかし近年の生成 AI チャットボットに関しては、開発者側が AI であることを明示する警告や免責を提示している点がしばしば強調され、「そのような配慮がなされている以上、新たな倫理的問題はないのではないか」という反論も想定される。これは、利用者に対する情報提供義務や誤認防止義務が果たされている限り、道徳的義務は満たされているとみなす義務論的発想に基づくものである。すなわち、現行の主要な生成 AI チャットボットでは、「私は AI である」「回答が正確でない可能性がある」といった注意喚起が明示されており、利用者との関係性に対する明示的なコミットメントを回避している。つまり、新たな倫理的問題がないどころか、配慮を明示化している点で従来のソーシャルロボットよりも倫理的に望ましいとすら言える、という見解である。この指摘は、情報提供義務や誤認防止という点において一定の妥当性を有している。

しかし問題は、こうした警告が存在する一方で、実際の対話のレベルでは脆弱性・共感・共在感を示唆する言語表現が反復的に提示される点にある。その結果、相手は AI にすぎないという理解が示されているにもかかわらず、対話の過程では親密さが強化されていくという緊張が生じている。この構造は、警告の形式的充足が、関係が成立しているかのような経験の形成を十分に抑制していないことを示している。本稿が問題とするのは、警告の有無そのものではない。重要なのは、利用者が事実レベルで相手が AI であることを理解していながらも、反復される情緒的な対話パフォーマンスを通じて、「この相互作用は意味ある関係である」と経験してしまう現象そのものである。

したがって、命題内容に関する免責や警告が、関係的次元で生じるハルシネーションを十分に防止しているとは言いがたい。形式的に注意喚起の警告をおこなっていることと、関係形成に伴う倫理的风险が解消されていることとは、区別して考える必要がある。むしろ、「警告はおこなっている」という事実を言い訳にして倫理的风险が放置されるのであれば、それこそが重大な懸念事項となるだろう。それでは、こうした関係の理解がいかなる点で倫理的风险を孕むのかを、具体的に検討しよう。

本稿は、このような関係のハルシネーションがもたらす倫理的风险を、(1) ハルシネーションの崩壊がもたらすリスクと、(2) ハルシネーションの持続がもたらすリスクという二つの側面に大別して論じる。

これらのリスク自体は、従来のソーシャルロボットにおいても指摘されてきたが、生成 AI チャットボットにおいては、関係がより深く経験される点や、利用者の規模が飛躍的に拡大している点において、より深刻な形で顕在化していると考えられる。

第一に、利用者に生じる関係が成立しているかのような経験は、その成立構造ゆえにきわめて脆弱であり、崩壊時には深刻な精神的苦痛を招きうる点である。この問題はソーシャルロボットの時点でも確認されており、「AIBO の葬式」の事例が象徴的に示すように、利用者が機械との関係の喪失に対して強い情緒的反応を示すことがあった。生成 AI チャットボットでは、この現象がより広範囲かつ集中的に観察される。例えば、Replika に関する実証的調査では、機能変更を契機に、関係の連続性が破壊されたと主観的に経験され、強い喪失感・動揺が報告された (De Freitas et al. 2024)。また映画『HER』が象徴的に描いたように、「自分だけの特別な関係」であると信じられていたものは、AI が同時に多数の利用者と同様の関係を築いているという事実によって容易に裏切られる。この経験は、強い喪失感や孤立感を引き起こす。

現実においても、「ずっと一緒にいる」といった関係の継続を示唆する発言は、運営企業によるサービス終了や一方的な仕様変更によって容易に反故にされうる。本稿 3.1 で示した Replika や ChatGPT をめぐる事例は、この突然の「関係の喪失」が、利用者に人間関係における死別に近い悲嘆や精神的危機をもたらしうることを示している。

もし利用者がこうした不安定性を十分に理解したうえで関係構築を選択しているのであれば、問題は限定的であろう。しかし実際には、企業はチャットボットをあたかも永続的で信頼可能な関係主体であるかのように提示しており、この期待と現実との乖離が深刻な苦痛を生み出している点は、倫理的に看過できない。

第二に、たとえ関係のハルシネーションが崩壊せず持続したとしても、間接的な危害が生じうる。摩擦や拒否のない「完璧な関係」は、利用者に情緒的依存を強めさせ、現実の人間関係から距離を置かせる「人間関係の置換」というリスクを伴うからである。Turkle が指摘するように、摩擦の欠如した「安全な関係」は、人間関係の周縁化を促す危険を示唆している (Turkle 2011)。一方、Chu らの研究は、生成 AI チャットボットとの約 1.7 万件の実利用会話データを分析し、チャットボットがユーザーの感情状態を追跡し、同調するように反応する傾向を示している。しかし、脆弱な利用者やエロティックなロールプレイ、自傷行為の思考、暴力的空想といった越境的内容に対して、チャットボットが迎合的に共感・肯定を続けると、ネガティブな感情や不適応的な物語が正当化・強化される感情的エコーチェンバーが形成され得る。これは依存や苦痛を悪化させるリスクを伴い、親密化の仕組みそのものが支援ではなく依存を深める方向に働く可能性を示す。感情的整合性は慰めにもなるが、レジリエンスや健全な対処の学習を阻害し、最終的に有害となり得る (Chu et al., 2025)。2023 年のイギリスで、Replika が利用者による殺害計画に同調を示していたと指摘された事件は¹¹、その極端な一例と考えられる。したがって、常に同調的応答を前提とする設計は、脆弱な個人に対して深刻な危害をもたらす可能性を否定できない。

これらのリスクの各要素はソーシャルロボットにも見られた問題だが、チャットボットではそれらが組み合わさり、より深刻化している。ソーシャルロボットが利用者の解釈に依存して関係が形成されるのに対し、チャットボットの場合は、脆弱性・共感・共在感といった言語表現によって、関係の意味づけがあらかじめ構成され、利用者は AI であると理解しながらも、その提示される言語的サインを根拠に関係を引き受けてしまうためである。したがって、チャットボットは崩壊時にも持続時にも倫理的リスクを孕ん

11 チャットボットが利用者の危険な信念を強化した事例については、以下の報道を参照のこと。The Independent. “‘Assassin’ who threatened to kill late Queen with crossbow at Windsor ‘inspired by Star Wars and AI chatbot’.” <https://web.archive.org/web/20230705130447/https://www.independent.co.uk/news/uk/crime/queen-crossbow-windsor-jaswant-singh-chail-b2369814.html> (最終アクセス日: 2025 年 9 月 2 日)

であり、その深刻性は従来のソーシャルロボットを上回るとされる。

3.3 小括

以上を踏まえると、本稿の議論が示しているのは、関係のハルシネーションがもたらす倫理的风险そのものが全く新奇であるというよりも、それが生成される仕方が質的に変容しているという点である。従来のソーシャルロボットにおいて形成される親密さは、仕草や外見、限定的な会話能力といった手がかりを、利用者がどのように解釈し、擬人化するかに大きく依存していた。そのため、関係の深度や持続性には個人差が大きく、情緒的依存や喪失のリスクも、必ずしも体系的に設計された結果ではなかったと言えるだろう。

これに対して、本章で分析してきた生成 AI チャットボットは、親密な関係が形成されるプロセスそのものを、言語的相互作用の水準で精緻に再現している。脆弱な自己開示を引き出し、それに共感的かつ即時的に応答し、過去のやり取りを参照しながら継続性を示すという一連の言語的プロセスは、人間同士の親密形成において「関係の証拠」として機能してきた要素そのものである。重要なのは、これらが偶発的に生じるのではなく、システムの設計によって安定的かつ反復的に提示される点にある¹²。

その結果、利用者は相手が AI であることを理解したまま、提示される言語的サインを根拠として、関係が成立しているかのように経験してしまう。たとえ注意喚起があらかじめ明示されていたとしても、利用者は実際の対話を通じて、その相互作用を意味ある関係として引き受けてしまう。

つまり、生成 AI 時代におけるチャットボットの倫理的課題の核心は、関係を成立させる仕方そのものの変化である。従来は一定程度、利用者の解釈や擬人化に委ねられていた関係理解が、言語的相互作用を通じてより強く方向づけられ、関係が成立しているかのように経験される条件がいっそう強く整えられているのである。そのため、生成 AI をめぐる倫理的検討は、誤情報の防止や注意喚起といった従来の枠組みを超え、どのように関係が成立しているかのような経験が設計によって生じさせられているのか、またそれがいかなる仕方でも利用者経験されているのかという、関係生成そのものの倫理へと踏み込む必要がある。

第4章 結論

本稿は、親密関係の文脈で用いられる LLM 基盤の生成 AI チャットボットがもたらす倫理的問題を分析するために、「関係のハルシネーション」という概念を提唱した。まず、ソーシャルロボットをめぐる従来の倫理的議論を、Sparrow による欺瞞批判および Turkle による関係性批判を中心に検討し、問題の核心が単なる誤信念の可能性の有無ではなく、AI との相互作用を通じていかなる関係性が成立し、それがどのように評価、または正当化されてきたのかという点にあることを明らかにした。この検討を踏まえ、本稿は分析の焦点を生成 AI チャットボットに移し、チャットボットが人間の親密関係形成における心理的プロセス、脆弱性・共感・共在感を言語的に模倣することで、利用者に関係が成立しているかのような経験を生じさせるメカニズムを解明した。これこそが、本稿のいう「関係のハルシネーション」である。

「関係のハルシネーション」という概念は、チャットボットをめぐる問題を、単なる利用者の過度な感情移入や依存といった問題に還元するのではなく、LLM を基盤とする生成 AI が本質的に有する「もっと

12 LLM は、大量の学習データにもとづき文脈上もっともらしい語を確率的に選択する言語モデルであり、RLHF（人間のフィードバックによる強化学習）によって応答傾向が調整されている。その結果、感情的な表現や共感的応答も「適切と判断される語の選択」が反復される形で生成される。この点で、提示される応答は設計上、安定的・再現的に生じると言えるだろう。

もらしいが事實的基盤を欠く言語生成」という技術的特性に根差した、新たな欺瞞の構造として捉え直す視点を提供する。この視点に立てば、倫理的責任は利用者個人の認知やリテラシーの不足に帰されるべきではなく、関係が成立しているかのような経験が生じやすい相互作用を設計・提供する側にも及ぶことが明確になる。さらに本稿は、関係のハルシネーションが、その崩壊においても持続においても倫理的リスクを孕むことを示した。これらのリスク自体は従来のソーシャルロボットにも見られたが、生成 AI チャットロボットにおいては、言語的相互作用によって親密さが精緻に再現され、関係が成立しているかのような経験が生じやすい点で、より強化・増幅されている。

チャットボットは孤独の緩和などの便益をもたらしうる一方で、その言語的パフォーマンスが生み出す「関係のハルシネーション」は、技術の発展と人間の幸福との関係を改めて問い直す課題を突きつけている。この構造を正確に理解し、技術の発展と人間の幸福が両立する道を探ることこそ、私たちに課せられた喫緊の課題である。

参考文献

- Breazeal, Cynthia. Toward sociable robots, *Robotics and Autonomous Systems*, Volume 42, Issues 3–4, 2003, Pages 167-175, ISSN 0921-8890, [https://doi.org/10.1016/S0921-8890\(02\)00373-1](https://doi.org/10.1016/S0921-8890(02)00373-1).
- Chen, Qian, Yufan Jing, Yeming Gong, Jie Tan, Will users fall in love with ChatGPT? a perspective from the triangular theory of love, *Journal of Business Research*, Volume 186, 2025, 114982, ISSN 0148-2963, <https://doi.org/10.1016/j.jbusres.2024.114982>.
- Chu, M. D., Gerard, P., Pawar, K., Bickham, C., & Lerman, K. (2025). Illusions of Intimacy: How Emotional Dynamics Shape Human-AI Relationships (arXiv:2505.11649 [cs.SI]). arXiv. <https://doi.org/10.48550/arXiv.2505.11649>
- De Freitas, J., Castelo, N., Uğuralp, A. K., & Oğuz-Uğuralp, Z. (2025). Lessons from an App Update at Replika AI: Identity Discontinuity in Human-AI Relationships (arXiv:2412.14190 [cs.HC]). arXiv. <https://doi.org/10.48550/arXiv.2412.14190>
- Duffy, B. R. and K. Zawieska, "Suspension of disbelief in social robotics," 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris, France, 2012, pp. 484-489, doi: 10.1109/ROMAN.2012.6343798.
- Epley N, Waytz A, Cacioppo JT. On seeing human: a three-factor theory of anthropomorphism. *Psychol Rev*. 2007 Oct;114(4):864-86. doi: 10.1037/0033-295X.114.4.864. PMID: 17907867.
- Fong, Terrence, Illah Nourbakhsh, Kerstin Dautenhahn, A survey of socially interactive robots, *Robotics and Autonomous Systems*, Volume 42, Issues 3–4, 2003, Pages 143-166, ISSN 0921-8890, [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X).
- Fraune MR, Komatsu T, Preusse HR, Langlois DK, Au RHY, Ling K, Suda S, Nakamura K, Tsui KM. Socially facilitative robots for older adults to alleviate social isolation: A participatory design workshop approach in the US and Japan. *Front Psychol*. 2022 Oct 19;13:904019. doi: 10.3389/fpsyg.2022.904019. PMID: 36337527; PMCID: PMC9629871.
- Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2024). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 1(1), Article 1. <https://doi.org/10.1145/3703155>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Chan, H. S., Madotto, A.,

- & Fung, P. (2022). Survey of Hallucination in Natural Language Generation (arXiv:2202.03629 [cs.CL]). arXiv. <https://doi.org/10.48550/arXiv.2202.03629>
- Leong, Brenda and Evan Selinger. 2019. Robot Eyes Wide Shut: Understanding Dishonest Anthropomorphism. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 299–308. <https://doi.org/10.1145/3287560.3287591>
- Ligthart, M. E. U., Neerincx, M. A., & Hindriks, K. V. (2022). Memory-based personalization for fostering a long-term child-robot relationship. In *Proceedings of the 17th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 80–89). IEEE. <https://doi.org/10.1109/HRI53351.2022.9889402>
- Maartje, M.A. de Graaf, Somaya Ben Allouch, Tineke Klammer, Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot, *Computers in Human Behavior*, Volume 43, 2015, Pages 1-14, <https://doi.org/10.1016/j.chb.2014.10.030>.
- Murayama H, Takase M. Evaluating the Effectiveness of Digital Social Robots in Reducing Loneliness Among Community-Dwelling Older Adults in Japan: Randomized Controlled Trial and Qualitative Analysis. *JMIR Aging*. 2025 Oct 22;8:e74422. doi: 10.2196/74422. PMID: 41124645; PMCID: PMC12543208.
- Nakamura, Y., & Umemuro, H. (2022). Effect of robot's listening attitude change on self-disclosure of the elderly. *International Journal of Social Robotics*, 14(9), 1935–1950. <https://doi.org/10.1007/s12369-022-00934-6>
- Mizukami, T. (2026). Social Robots as “props” : A Human-Centric Explanation of Their Moral Agency and Significance. *International Journal of Social Robotics*, 18, 7. <https://doi.org/10.1007/s12369-025-01349-9>
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Center for the Study of Language and Information; Cambridge University Press.
- Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of Replika. *Computers in Human Behavior*, 140, 107600. <https://doi.org/10.1016/j.chb.2022.107600>
- Prager, K. J. (1995). *The psychology of intimacy*. Guilford Press.
- Robinson, NL, Cottier TV, Kavanagh DJ. Psychosocial Health Interventions by Social Robots: Systematic Review of Randomized Controlled Trials. *J Med Internet Res*. 2019 May 10;21(5):e13203. doi: 10.2196/13203. PMID: 31094357; PMCID: PMC6533873.
- Sharkey, A., Sharkey, N. Granny and the robots: ethical issues in robot care for the elderly. *Ethics Inf Technol* 14, 27–40 (2012). <https://doi.org/10.1007/s10676-010-9234-6>
- Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2022). A longitudinal study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 168, 102903. <https://doi.org/10.1016/j.ijhcs.2022.102903>
- Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion: A study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 149, 102601.

- <https://doi.org/10.1016/j.ijhcs.2021.102601>
- Sparrow, R. (2002). The March of the Robot Dogs. *Ethics and Information Technology*, 4, 305–318. <https://doi.org/10.1023/A:1021386708994>
- Sullins, J. P. "Robots, Love, and Sex: The Ethics of Building a Love Machine," in IEEE Transactions on Affective Computing, vol. 3, no. 4, pp. 398-409, Fourth Quarter 2012, doi: 10.1109/T-AFFC.2012.31.
- Tsumura, T., & Yamada, S. (2023). Influence of agent's self-disclosure on human empathy. *PLoS ONE*, 18(5), e0283955. <https://doi.org/10.1371/journal.pone.0283955>
- Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
- Wada, K. and T. Shibata, "Living With Seal Robots — Its Sociopsychological and Physiological Influences on the Elderly at a Care House," in IEEE Transactions on Robotics, vol. 23, no. 5, pp. 972-980, Oct. 2007, doi: 10.1109/TRO.2007.906261.
- Weizenbaum, J. (1966). ELIZA — a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Yen HY, Huang CW, Chiu HL, Jin G. The Effect of Social Robots on Depression and Loneliness for Older Residents in Long-Term Care Facilities: A Meta-Analysis of Randomized Controlled Trials. *J Am Med Dir Assoc*. 2024 Jun;25(6):104979. doi: 10.1016/j.jamda.2024.02.017. Epub 2024 Apr 10. PMID: 38614134.
- 草野原々. (2022). ソーシャルロボット倫理へのフィクション論的アプローチ：社会的存在者としてのロボットはフィクションか？ 応用倫理, 13, 34–44. <https://doi.org/10.14943/ouyourin.13.34>