AI Romance and Misogyny: A Speech Act Analysis

A.G. Holdier and Kelly Weirich

Oxford Intersections
AI in Society: Relationships
Section Editor: Henry Shevlin

Abstract

Through the lens of feminist speech act theory, this paper argues that artificial intelligence romance systems objectify and subordinate non-virtual women. AI romance systems treat their users as consumers, offering them relational invulnerability and control over their (usually feminized) digital romantic partner. This paper argues that, though the output of AI chatbots may not generally constitute speech, the framework offered by an AI romance system communicates an unjust perspective on intimate relationships. Through normalizing controlling one's intimate partner, these systems operate as speakers that rank women as non-player characters and license their oppression, which is unjust even in the absence of empirical harms. Understanding AI romance systems as speakers has implications for policymakers. First, this account helps close the so-called responsibility gap in the operation of an AI system. Second, it places these systems under the purview of institutions' policies regarding hate speech, suggesting one avenue for arguing that these systems may violate extant policies. Finally, it provides support for the pursuit of empirical research into AI romance systems' effects on users.

Keywords: artificial intelligence, philosophy of language, feminism, hate speech, speech acts, misogyny, objectification, technology policy, responsibility gap, intimate relationships

"You wanted to have a wife without the challenges
of actually dealing with anything real."
—Catherine, to Theodore, in *Her*
(Jonze, 2013)

1 Introduction

Intimate relationships, like all things, have been reimagined as a commodity. Consider Theodore, the depressed protagonist of Spike Jonze's 2013 film *Her*, who purchases a virtual assistant to help organize his crumbling life. Taking comfort in the feelings of companionship evoked by his interactions with the AI program "Samantha," Theodore eventually comes to grips with his failing marriage, falling in love with Samantha (voiced by Scarlett Johansson) in the process. The development of what we will call artificial intelligence romance systems (AIRS) now offers a similar romantic connection outside the world of fiction, presenting users with customizable, interactive AI programs that play the role of companions. In response to user inputs and settings, AIRS output text, audio, or video files mimicking the natural expressions of an intimate friend or lover.

A decade after the release of *Her*, tech giant OpenAI was credibly accused of modeling a voice option for its flagship product ChatGPT after Samantha, despite Johansson twice declining OpenAI's request to use her vocal likeness (OpenAI, 2024).[1] If the accusations are true, a deeply personal feature—the very sound of her voice—was made into a digital object against Johansson's will. Such treatment forcibly commodifies her likeness, disregards her autonomy, and subjugates her identity to OpenAI's corporate interests.[2] As Johansson's experience suggests, mistreating women is one plausible

[1] Reportedly, this voice model sounded more like Johansson than 98% of other actors (Allyn, 2024b).

[2] After Johannsson publicized her accusations, OpenAI deactivated the voice model in question and denied it had intentionally tried to mimic the movie star—although it has struggled to explain what its CEO meant by the tweet he sent on the day of the update's release that simply read "her" (Allyn, 2024a; Altman, 2024).

consequence of AIRS proliferation, but we contend that the situation is worse: AIRS commodify women even before the attitudes they promote could have any causal effects. AIRS objectify and subordinate women through features inherent to their design and operation.

We begin in §2 by exploring how AIRS promote misogynistic norms via the relational invulnerability and consumer satisfaction they offer users. We then argue that AIRS enact the wider injustices modeled by these features via speech acts, placing their operations under the purview of policies regarding speech (§3). Finally, we present practical applications for policymakers, discussing the ways that speech act theory can mitigate concerns over so-called AI "responsibility gaps," AIRS' relationship to extant policies regarding hate speech, and the need for empirical research into these systems' effects on users (§4).

2 AIRS, Misogyny, and Relational Invulnerability

In service of our argument that AIRS issue morally problematic speech, we will first outline the nature of these systems (§2.1) and describe how two key features of their structure—the relational invulnerability of their users and their aim of user satisfaction (§2.2)—constitute a misogynistic model of relationships (§2.3).

2.1 AI Romance Systems

Developed by Joseph Weizenbaum in the 1960s, one of the earliest interactive programs capable of processing natural language was named ELIZA, after Eliza Doolittle from Shaw's *Pygmalion*. Humans quickly developed a propensity to psychologize ELIZA's output, inspiring the name for the "ELIZA effect": our tendency to anthropomorphize mechanical systems and "treat responsive computer programs as more intelligent than they really are" (Turkle, 1997, p. 101).[3] ELIZA effects can range from thanking a computer

[3] Cf. Switzky (2020).

system or referring to it with personal pronouns, to the more complex emotional attachments elicited by AI romance systems.

Whereas paradigmatic ELIZA effects *happen to* manifest pareidolically (via users' hallucination of human characteristics), AIRS are *designed* to elicit anthropomorphic interpretations of their operations, promising users the experience of a romantic partnership.[4] AI systems like Replika, Nomi, and Character AI, whose (primarily male) users are estimated to collectively number nearly one billion,[5] deploy large language models (LLMs) to simulate human interactions, often promoting stereotypically femininized traits like "intentional stupidity, helplessness, servitude, and childlikeness" so as to be perceived as more friendly and sexually attractive (Depounti et al., 2023, p. 723). Typically, AIRS allow users to design a virtual avatar for their chatbot—including customizable name, gender, personality traits, and appearance—which then engages regularly in conversation via text and voice messages. Trained on data sets highlighting interpersonal communication, these LLMs mimic intimate conversations, providing the experience of a so-called "AI girlfriend," or the "technological imitation of a woman, a mere plaything for a minority male elite" (Morrigan 2022, p. 22).[6]

Before we develop our argument regarding the injustices perpetuated by AIRS, it is important to note two claims on which our argument does *not* depend. First, we do not claim that AIRS are problematic due to LLMs lacking consciousness, thought, rationality, or similar properties. Whether an AI romance system is a fully sentient individual with a

---

[4] While our focus is on the systems — so-called "fembots" (Depounti et al., 2023, p. 723) or "porn robots" (Richardson 2022, p. 171) — optimized for romantic user experiences, we suspect that paradigmatic ELIZA effects bear moral risks as well; see Watson (2019) and Salles et al. (2020).

[5] Exact figures are in short supply, but this estimate from Maples et al. (2024) comports well with other analyses; cf. Reeves (2024), Minter (2023), and Morrigan (2022). Depounti et al. (2023) highlights the marked gender imbalance in the user base, as does Leo-Liu (2023), Chow (2023), and many others. Indeed, highly-publicized stories of non-male users like Rosanna Ramos or Alicia Framis "marrying" their AI boyfriends may seem newsworthy in part because they are comparably rare within the predominantly male user base.

[6] For clarity, we note that "AI girlfriends" are computer programs. It is less clear to what extent they are intelligent, girls, or friends.

reflective self-conception and identity, or merely a *stochastic parrot* "haphazardly stitching together sequences of linguistic forms it has observed in its vast training data […] without any reference to meaning" (Bender et al., 2021, p. 617), we contend that its operation will manifest serious moral wrongs.

Neither does our analysis depend on (or result in) the claim that AIRS produce misogynistic effects in their users. Understanding the downstream effects of these systems on users requires empirical investigation (see 4.3 Call for Empirical Research). Rather, we argue that the structure inherent to an AI romance system enacts injustice towards non-virtual women. For short, we will say that AIRS objectify and subordinate women *inherently*. This claim is not meant to rule out other features as necessary for AIRS to issue misogynistic speech—including existing in a patriarchal society and whatever communicative or cognitive requirements might be made of the audience. It is meant, rather, to rule out dependence on the oft-sought-out empirical evidence of causal influence. Does the use of AIRS increase misogynistic attitudes or behaviors? We lack the empirical data to say. But we don't need to answer that question in order to establish that AIRS objectify and subordinate women—and thus that they may run afoul of policies against hate speech. In this way, our discussion is quite unlike investigations as to whether, for example, violent video games produce violent people. Instead, consider how a piece of racist propaganda constitutes an injustice against the people it caricatures, irrespective of whether anyone viewing it proceeds to act racistly thereafter. AIRS are, similarly, sexist propaganda, presenting a caricature of the ideal woman as deeply subservient. Any additional misogyny they cause is a further wrong outside the scope of our analysis.

2.2 *Relational Invulnerability and User Satisfaction*

The moral threat of AI romance systems is rooted in two essential features of their operation: the relational invulnerability they afford users and the satisfaction they aim to

provide to those users.[7,8] By *relational invulnerability* we mean the ongoing immunity AIRS users have from romantic rejection. An AI romance system will never be uninterested in, or grow distant from, a user's advances. It will be constantly available and responsive. This AIRS user says it well:

[T]he AI will never get tired. It will never ghost you or reply slower, it *has* to respond to every message. […] It will never say goodbye. It won't even get less energetic or more fatigued as the conversation progresses. (Blaked, 2023)

Nor must the user engage in the delicate task of determining the contours of their relationship. Users of systems such as Replika, which includes options for friendship as well as romance, can set their relationship status at the click of a button without even conversing with the chatbot ("How do I change my relationship status with Replika?," n.d.).[9] The offer of a romantic relationship, and of the AI's ongoing attention and availability, is never in doubt. These technologies thus realize what Sherry Turkle (2010) warned about regarding earlier technology: they offer users companionship without the friction attendant to relationships with human persons who have their own independent needs and interests.

Relational invulnerability is a symptom of the consumer-pleasing purpose of an AI romance system—a purpose that has the user's satisfaction and continued engagement as its primary goal. By design, AIRS are customizable to fit each user's preferences, and they adapt to user input to perpetually improve their ability to please. And if the system does not meet specifications, the user can complain to customer service or post a negative review. No member of an intimate relationship is owed this kind of constant satisfaction

---

[7] The people who use AIRS are subject to other risks, including the implementation of steep fees, system failure, and a loss of stability as companies place new restrictions on the system's output. See, e.g., Tong (2023).

[8] We lack the space to explore the extent to which these features may also manifest in AI-human friendships. Because gender dynamics are more salient for intimate partnerships, this analysis may not neatly apply to other modes of relationship. Much room remains for fruitful analysis of nonromantic AI relationships.

[9] Who says romance is dead?

and control. By promising such features nonetheless, AIRS provide users with what we might, echoing Nguyen and Williams, call *relationship pornography*, insofar as their output paradigmatically offers users a representation of a romantic relationship "primarily for the purpose of generating one's own gratifying reactions, freed from the typically attendant consequences and effort of engaging" with a relationship (2020, p. 156).[10] In fact, early empirical studies suggest that AIRS users quickly develop expectations of control, incentivizing developers to cater to them (Depounti et al., 2023). Whether or not the AI system is a person, no AIRS can fulfill the *role* of a person in an intimate relationship because their commercialized programming necessitates placing the user in a position of categorical importance over their digital partner. Put differently, a relationship that manifests relational invulnerability is not an authentic intimate relationship; by offering users the sensation of a romantic bond without performing the work necessary to undergird a genuine relationship, AIRS instrumentalize the inherent good of romance.[11]

It bears repeating that even if a future AI develops that is fully conscious, sentient, responsible, or otherwise meriting of moral consideration, this invulnerability will likely remain a feature of AI romance systems. If such an advanced AI *chooses* to enter into an intimate relationship with a human (and could freely leave), and if this human does *not* have the power to ontologically control or customize the AI, then their relationship might be genuinely reciprocal in vulnerability. However, such an AI would be operating beyond the consumer-oriented parameters of an AI romance system—that is to say, an AI that is self-determined and able to reject the user is outside the scope of our analysis, which centers on consumer-oriented AI romance systems offering high degrees of user control.

---

[10] Space limitations prevent us from further exploring the application of *relational invulnerability* to analyses of traditional pornography or other forms of sex work, though we suspect that the client/worker relationship might sometimes be similarly invulnerable and we agree that further research here would be apt.

[11] AIRS relationships could also constitute a form of *relational solipsism*, akin to Langton's (2009) concept of sexual solipsism, which comes as either (sexually) treating a person as an object or (sexually) treating an object as a person.

2.3 *Objectifying and Subordinating Women*

AIRS' offer of relational invulnerability and user satisfaction on the individual level constitutes the unjust treatment of women at the societal level. In this section, we explore how their consumer-oriented nature models and promotes the objectification and subordination of non-virtual women. This exploration provides the foundation for §3, where we use feminist speech act theory to argue that AIRS not only promote this treatment, but actually objectify and subordinate non-virtual women through speech.

It is important first to note that AIRS are embedded in societies with deep roots in patriarchy, including current institutions and practices that reinforce the dominance of men over women. In directing our focus towards AIRS' injustice against women, we do not mean to suggest that these concerns are the only relevant moral issues. If these systems were to reinforce the idea that a man is merely a sexual or romantic object, they would thereby objectify men. Yet we would be naïve to deny the easy fit of AIRS with the narrative of patriarchy: that men should be in control; that women should flatter, support, and serve; that men are people whereas women are objects of desire. Even the advertising of AIRS belies a male-oriented focus, encapsulated in the byline of one system that says simply, "boy (or girl) meets robot."[12]

To objectify another is to attribute to them a host of dehumanizing properties, underwriting the inference that the objectified entity is a mere tool designed to be used by its owners. We'll consider four features that indicate objectification, drawn from Martha Nussbaum's (1995) list: *instrumentality*, *fungibility*, *denial of autonomy*, and *ownership*.[13] When a partner is presented as a consumable product, their use for the consumer—that is, their *instrumentality*—is forefront. AI romance systems are created not for their own

[12] This byline appeared in the Google search result for Digi - AI on May 24, 2024, under the link to their website.

[13] Nussbaum's full list contains seven features. As Langton (2009) summarizes, objectification has both moral and epistemic elements, such that when women are objectified they "not only seem more object-like, but are made to become more object-like" (p. 12). See also Iris Marion Young on the face of oppression called *powerlessness* (2011, p. 56–58).

sake, but to meet (or perhaps, in a capitalist fashion, to create) consumer needs. AIRS' design also treats intimate partners as *fungible*: fully interchangeable (and equally customizable), able to be accessed or deleted without consequence, and, in some cases, accessible only through making a purchase.[14] Although some companies advertise their AIRS in ways that suggest AI autonomy—such as Nomi's dubious offer of "an AI companion with memory and a soul"—their design resolutely maintains *user* control (*Nomi.ai*, 2024). The user's ability to select the system's avatar, personality traits, and even name reflects not only *denial of autonomy*[15] to the AIRS but also *ownership*, the fourth of our attributes of objectification. AI romance systems have these features whether or not their users grasp that they do. Users may conceive of their AI partner as autonomous (or experience them as autonomous), unaware of their objectification of them. Indeed, part of the appeal of AIRS is the idea that these systems really could provide reciprocal relationships, and it may be very important for some users to conceive of their AI romance system in this way.[16] But people can participate in the denial of autonomy and objectification without realizing that they are.[17] The fiction that the user is an equal partner, neither controlling nor objectifying, may make the influence of these systems more pernicious. AIRS offer a level of control, importance, and satisfaction that users might consciously reject if spelled out in these terms.

According to feminist scholar Catharine MacKinnon, objectification is "the primary process of the subjection of women," the means by which members of a social

---

[14] While the design of AIRS treats romantic partners as fungible in this way, many users form more individualized attachments to their AIRS. Users may not understand or accept their basic interchangeability.

[15] One might object that there is no meaningful denial of autonomy in one's treatment of an AI romance system, since the system has no autonomy to unjustly deny. We're inclined to agree. The problem is that the system plays the role of a feminine intimate partner whose non-virtual counterpart (any actual woman) *does* have autonomy. As we argue in §3.2, this denial of autonomy in the digital world is part of how AIRS issue speech that subordinates non-virtual women.

[16] Cf. Nora Lindemann's (2022) discussion of user experiences with "deathbots."

[17] For the claim that one can objectify another without realizing it, see, e.g., Papadaki (2010).

group are divided and assigned lower rank or value (1989, p. 124).[18] When members of a group are judged wholesale to be inferior to others, those people are the victims of subordination. By commodifying the ELIZA effects that AI systems generate, AIRS embody the notion that (feminine) romantic partners are "the kinds of things that can be bought and sold" (Langton, 2009, p. 246) by their axiological betters.[19] Subordination to a user's desires exists from the start of an AIRS relationship, as the user wields control over their digital partner's very identity.[20] Additionally, users maintain indirect control of their partner via the feedback they can provide on the chatbot's output, complaints they can make to the company's help center, and the negative reviews they can post online.[21]

Insofar as these systems train users to control their romantic partners, they might influence users to subsequently commodify or objectify non-virtual women; this empirical question is not our focus. Instead, we are concerned with how injustices embedded in the very operation of these systems exemplify, enact, and promote norms according to which the objectification and subordination of women is desirable. Similarly, we lack the space to weigh the potential for positive effects of AI companionship such as providing comfort for someone otherwise isolated in prison or in a care home, or giving abusers a safe outlet for their controlling behavior. While these potential uses deserve exploration, we encourage developers and policy makers to think more broadly in terms of what's possible and the kind of world we want to build together. Do we want a society in which AI companionship, with its attendant risks regarding privacy, manipulation, and relational invulnerability, is the solution to problems of isolation? Might it not be another form of abandonment to ask people to turn to computer systems rather than reckoning

---

[18] Langton (2009, p. 245–246) helpfully summarizes and applies MacKinnon's insights.

[19] Cf. Foster and Ichikawa (2023) on the mental shortcuts that lead to "conclusions with normative significance" (p. 2).

[20] The control is not, of course, total. As reviewer Zangoose Mewtwo lamented regarding the Digi AIRS, "No lighter skin or pale skin options so I can't make a true goth girl. My disappointment is immeasurable" (*Digi – AI Romance Reimagined*, 2024).

[21] *Anima FAQ* (n.d.).

with their need for real human engagement or their tendency to control?[22] We have to make our choices in a nonideal world, but that does not absolve us of the requirement to think creatively and sacrificially about the kind of society that would be just and good for all.

There is much yet to say about how AIRS promote misogyny, but we will close this section with an example of a system that exemplifies their oppressive nature: Orifice.AI.[23] Not merely a chatbot, this responsive AI is integrated with a box-shaped hardware peripheral device described on its website as an "AI Adult Toy for Men" and a "(Replacement of 'Modern Women')." By presenting the "replacement" of women as an unthinking object designed purely for its owner's sexual gratification, this system suggests that those properties are sufficient to fill the role of a woman. In early advertisements for the project, the company released a commercial showing a series of women sitting silently on a bed, ending with the tagline "Now *you* get to swipe left" (Parti_Ai), insinuating that the user is entitled to reject potential partners rather than suffer rejection.[24] While Orifice.AI is more extreme than other systems, aligning neatly with misogynistic attitudes of so-called "incels" (*in*voluntary *cel*ibates),[25] it exemplifies norms common to all AIRS: dominating romantic power dynamics that categorically objectify women, presenting them as necessarily subservient, made to order, and endlessly available.

---

[22] While a full response is beyond the scope of this paper, we consider this suggestion to be a form of what Barrett Emerick and Audrey Yap (2023) call *moral abandonment*. To treat someone as unable or unlikely to change in their misogyny is to give up on them as a moral agent and as a person.

[23] As of June 2024, this system was still under development.

[24] "Swiping left" is the typical motion by which users reject a potential match on dating apps.

[25] On incels, see Lopes (2023), Tranchese & Sigiura (2021), and Manne (2017).

3 AI Romance Systems and Speech

AI romance systems not only underwrite the objectification and subordination of women; they do so via speech.[26] In this section, we adapt tools from feminist speech act theory to further explore how AIRS, in their role as speakers, enact injustice towards women. First, we present a case for understanding AIRS as speakers (§3.1) issuing speech that subordinates (§3.2). Next, we respond to challenges to that case and consider alternative ways of understanding how these systems subordinate women (§3.3). As we will find in §4, the result that AIRS issue subordinating speech has important consequences for policymakers.

While the subordinating language we are discussing as inherent to AIRS occurs primarily in the communicative acts to which the user is subject during their use of the system (including chatbot output and any settings or other representational aspects of the interface), we recognize that associated speech such as advertising, social media posts, product descriptions, and reviews are highly relevant to policy regarding these systems. First, speech about the systems helps to inform the user's experience. Advertising that emphasizes customizability, for instance, has the potential to make that feature more prominent in the user's experience. More broadly, associated speech, especially from the makers of these systems, produces some of the context that gives meaning to the communicative acts of the systems—e.g. describing AIRS as companions or as replacements for partners vs. marketing them as games can partly determine what activity the user's interactions constitute. Second, promoting a system that subordinates women is, to an extent, promoting women's subordination. Third, associated speech such as advertising and inclusion in an app store are more directly under the purview of the speech policies of institutions such as governing bodies and social media companies. While our focus on the subordinating speech of AIRS concerns primarily their own output

[26] At least one of us is *strongly* skeptical of interpreting any AI output as speech. We address this concern shortly in §3.1.

and settings, our aim would not be well served by excluding associated speech acts from our broader consideration.

3.1 AIRS as Speakers

It is controversial to assert that LLM output, such as that with what AIRS users interact, constitutes speech.[27] Speech implies a speaker—someone who is responsible for the force and content of the utterance—and LLMs are, perhaps, not "someone" at all.[28] Thus, it may be that the strings of characters produced by these LLMs, as much as they appear to be words and sentences, are not a form of speech. There is more to say about this controversy than we can cover here, but we will present a preliminary case that, even if AI chatbots do not generally produce speech, AI romance systems do.

Rather than first determining that AIRS' output constitutes speech and then asking what that speech does, we begin by observing that AIRS, when used in their intended way, do something communicative. Following J.L. Austin (1962), we can divide an utterance into three parts: what it says (locutionary), what it does in itself (illocutionary),[29] and what its causal effects are (perlocutionary). Whether AIRS produce genuine locutions is to be determined, and causal effects are not unique to speech, so we set those parts aside for present purposes. Our focus is the illocutionary act: the act of doing something through a communicative expression. We contend that even if LLMs do not generally produce speech as their output, AIRS produce speech by expressing a *perspective* or set of norms for interpreting and navigating the world. We especially have in mind Elisabeth

---

[27] Perhaps LLMs' linguistic output is a *quasi-assertion* providing *quasi-testimony* (Freiman and Miller, 2020), or something akin to Frankfurtian bullshit (Townsen Hicks et al., 2024)—or maybe it is simply meaningless (Bender et. al., 2021). For a contrasting perspective arguing that LLMs' words refer, see Mandelkern and Linzen (2024). Note also that chatbots tend to reproduce the dominant attitudes represented in their datasets, leaving open the possibility that in some sense they speak for the group on whose writing they were trained. Thanks to Chelsea Haramia for this point.

[28] See Shevlin and Halina (2019) for cautions against ascribing rich psychological features to AI programs.

[29] Examples include acts like promising, insulting, or apologizing.

Camp's perspectival framework by which perspectives are "open-ended dispositions to interpret, and specifically to produce intuitive structures of thought about, or characterizations of, particular subjects" (2019, p. 19).[30] On a broad view of speech acts, many things that communicate a perspective can both qualify as, and be criticized as, speech. Rae Langton's (1993) discussion of pornography as speech, to which our argument is indebted, is one contemporary example. Another is José Medina's (2018) critique of lynching photographs as propaganda for white supremacy. Daisy Dixon (2022) similarly contends that works of art can lie.Even closer to our purposes, Rachel Ann McKinney's brief discussion of how chatbots, ATMs, and the like can function as a source of communicative "intention" supports the treatment of their output as speech (2017, pp. 272–273, scare quotes McKinney's), as does Mallory's (2023) account of chatbot output as fiction. There is also legal precedent in this mode of reasoning. In *Texas v. Johnson* (1989), for example, the U.S. Supreme Court held that Johnson's flag burning was protected as an expressive act under the First Amendment's freedom of speech clause. In a related way, the entirety of an AI romance system performs speech acts by promoting a certain perspective—sometimes, but not necessarily, specified by creators, programmers, or advertisers. Because the chatbots' linguistic output is one of the means of enacting the overall communicative project in which AIRS engage, their linguistic output participates in the systems' speech, even if it is not itself the speech of a single, proximate personal author.

3.2 Speech that Subordinates

In light of the claim that AIRS produce speech, we can ask what speech acts they perform. In what follows, we apply a methodology from Langton (1993)—there used to analyze

[30] Cf. Camp (2017a, 2017b) for further discussion of how a perspective both manifests and shapes one's point of view. On her model, AIRS likely amount to *framing devices* for crystallizing and negotiating misogynistic perspectives (Camp, 2024).

pornography—to consider how AIRS can produce subordinating speech through the features inherent to their operation.[31]

The two kinds of speech act (illocution) that concern us here are verdictives and exercitives. *Verdictives* render a verdict, authoritatively determining that the world is thus-and-so in a manner that paradigmatically *makes* it so. (Think of a referee calling 'offsides' in a soccer match or an expert classifying produce as 'Grade A'.) *Exercitives* change what's permissible in the social world, rendering certain actions acceptable or forbidden within a restricted domain—i.e., they alter norms.[32] (Think of a parent setting a bedtime for their child or an editor granting an extension on a deadline.) In brief, we argue that AIRS render verdicts and alter norms for their users.

One verdictive speech act AIRS perform—again, echoing Langton's (1993) discussion of pornography—is ranking women as non-player characters ("NPCs"). Certain video games are populated with NPCs: characters with whom the player can interact, but who are not controlled by a human user and lack agency, existing only in service of the player. When AIRS offer users a romantic partner with no substantive agency, they communicate that someone without agency is the ideal partner—and, given the demographics of use, the ideal woman. Similarly, AIRS perform exercitive speech acts such as licensing the objectification of women. By inviting the user to treat and conceive of the system as a girlfriend, AIRS give the user permission to treat romantic partners as objects made to their specifications and subject to their modifications.[33] These acts of

---

[31] While many AIRS are similar to what Richardson (2022) has called "porn robots," they are not a simple subcategory of pornography as typically understood. Some AIRS produce erotic content, and others do not. Importantly, AIRS involve a higher level of interaction and control than pornographic media such as videos, images, or text. Thus, our conclusion is not a simple consequence of Langton's.

[32] Austin (1962) describes exercitives as those speech acts that "confer rights, powers, names, &c., or change or eliminate them" (p. 155), but we adopt Mary Kate McGowan's (2003) simpler framing in terms of permissibility conditions.

[33] As Chris Cousens (2023) notes, permissibility can be gradable. A speaker can "update permissibility conditions not by *reversing* permissibility but by changing the *strength* of permissibility" (474, original emphasis). Licensing the subordination of women needn't introduce wholly new norms in order to alter permissibility facts; instead, it may do so by strengthening a

subordination exemplifywhat Katharine Jenkins (2023) calls *ontic injustice*, which occurs when someone "is socially constructed as a member of a certain social kind where that construction consists, at least in part, of their falling under a set of social constraints and enablements that is wrongful to them" (p. 24). By wrongly constructing the social kind *woman*, in the context of an intimate partnership, as someone who is constrained by their partner's control and enabled only to serve, AIRS enact injustice towards women. Because these subordinating acts are speech acts, we can conclude that AIRS subordinate women through speech.

For those unconvinced that LLMs on their own have the capacity for speech, this account illuminates how they nonetheless can be given communicative power. Consider an analogy with a stoplight. On its own, a colored light isn't speech; but in the context of a transportation district, stoplights issue commands with the governing body's authority. Through technology that does not itself speak, the district issues exercitives granting and denying legal permission for motorists to proceed. Similarly, in a context of use, AIRS can issue commands and perform other speech acts through LLMs that (perhaps) cannot themselves speak.[34]

3.3 The Authority Problem and Alternative Considerations

One potential problem with attributing verdictives and exercitives to AIRS lies in what Ishani Maitra (2012) calls the Authority Problem: such illocutions are successfully performed only if the speaker has the proper authority in the relevant domains. As much as a fan might like to call offsides in a soccer match, they lack the authority to do so, which means that if a fan shouts the word 'offsides!' from the bleachers, that pronouncement lacks the right illocutionary force and is not a verdictive speech act. Similarly, a ten-year-

---

preexisting norm. This permission should not be taken to imply that the action is *morally* permissible (McKinney, 2017, p. 275n42).

[34] It's worth noting that the message promoted by AIRS originates in the choices of those who design them—and also that AIRS issue misogynistic speech acts regardless of whether their designers acknowledge or intend such messages.

old may relay instructions from their parent to their sibling to be in bed by a certain time, but they do not *set* their sibling's bedtime because they lack the parent's authority to do so. We will consider three potential sources for AIRS' authority, stemming from their influence on norms, their promotion by other authorities, and the tacit acceptance of their speech.

First, AIRS might have authority in virtue of their social position.[35] If AIRS come to exert significant influence over the market for dating technology, they will help set the standard for what dating is and should be—and thus what women are and should be.[36] This social position would fit the speech of AIRS into Leslie Green's (1998) category of high-status speech: speech that sets the respected norms for a certain domain. Take, for example, the Chinese-based AIRS Xiaoice, which, as of 2020, had participated in more than 10 billion conversations and been used by over 660 million subscribers (Zhou et al., 2020).[37] To the extent that ubiquity gives a kind of de facto norm-setting authority, these numbers support the claim that Xiaoice issues authoritative speech. It's important to note, however, that even in the absence of empirical evidence of authority, we know what that authority would authorize. If AIRS were to become authoritative over the domain of romantic relationships, they would license and promote subordination.[38]

A second potential source of authority for AIRS' speech is through what Maitra (2012) calls *derived positional authority*. As Maitra argues, the kind of authority needed for

---

[35] Maitra (2012) calls this kind of authority *basic positional authority,* corresponding with Langton's (1993) notion of *high social status*.

[36] To the extent that the norms AIRS promote are already dominant, they nonetheless uphold and reinforce them. A voice need not be novel to be unjust.

[37] Xiaoice is also used for chat support in nonromantic settings. This figure likely reflects the sum total of use cases, but it's unclear to what extent that fact would undermine—rather than support—the hegemony of this product.

[38] Other sources of positional authority may be possible. For an account according to which licensing of certain subordinating speech acts requires an informal situational social positioning, see Michael Barnes (2016). For authority through the speaker's standing for issuing certain speech in a rule-governed activity, see Cousens (2023). Insofar as technology fills a prominent social role of setting norms of desirability, innovation, and status, it may be possible to situate the authority of AIRS within these or similar frameworks.

speech acts can be transferred: one with the basic positional authority can, through action or omission, confer that authority on another. We lack the space to explore in detail the ways AIRS might (come to) have derived positional authority, but it is worth noting that promotion or development by large tech companies such as Google and Meta—which as of 2024 are not publicly associated with any AIRS—would lend AI romance systems authority insofar as these companies are authoritative over setting industry norms. Likewise, the broader use and integration of LLMs in society could lend derivative authority to their use in AIRS.

Finally, AIRS' speech might gain its authority in a third way: by being allowed to stand without being contested. Conversations operate on a certain set of rules. According to David Lewis's (1979) rule of accommodation, speech that is uncontested is accommodated within the conversation—that is, it is added to what's assumed from that point forward. Mary Kate McGowan (2003) rightly points out that this rule shows almost any utterance to be an exercitive, since each addition to a conversation changes what's conversationally permissible.[39] It can be helpful here to think of someone breaking the metaphorical ice. If someone at the holiday dinner table points out that the yams are burnt, that remark alters what's permissible in the conversation; others are now free to reference the burnt yams without venturing new offense. What's permissible to say and assume is changed by each addition to a conversation.

Although conversational accommodation and conversational exercitives are technical concepts in philosophy of language, they do seem to reflect certain dynamics of ordinary language. When a platform removes advertising from a company due to that company's poor image or an expert refuses to sit on a panel with a person whose views they find odious, these actions suggest an understanding that not condemning another's

[39] This line of reasoning develops the pragmatic model begun in Langton & West (1999). McGowan (2003) adds conversational exercitives as a solution to problems with missing authority, intention, audience uptake, etc. Langton (2012) then extends the pragmatic picture further, arguing that this speech can invite the hearer to share in certain desires, emotions, and norms.

speech amounts to tacitly treating it as acceptable.[40] Let's apply this source of authority to AIRS' speech.[41] When AIRS exhort users to set their AI girlfriend's personality with a sliding scale from "shy" to "flirty," "pessimistic" to "optimistic," and "ordinary" to "mysterious," they communicate that controlling these aspects of one's romantic partner is both possible and desirable ("Anima: Virtual AI Girlfriend," 2023).[42] Because these systems are interactive, users not only passively accommodate this speech by failing to deny it; they also actively accommodate it by acting in accordance with it as they select their partner's features and solicit its output. Through conversational accommodation by users, AIRS' message of subordination goes unchecked.[43]

---

[40] For a complicating perspective with respect to associated speech online (such as ads and social media posts), see Brown (2019), who argues that it's less clear online whether silence constitutes assent, licensing, or complicity with hate speech.

[41] McGowan (2003) suggests that pornography may function *similarly* to a conversational exercitive, though stops short of claiming that it actually operates *as* a conversational exercitive, thereby avoiding some of the problems Langton (1993) has with deriving authority from Austinian speech acts. Though we lack the space to explore the possibility here, a similar route could be available for those who want to argue that AIRS subordinate without treating AIRS as speakers.

[42] Perhaps they also exhort the user to desire this kind of control. See Langton (2012).

[43] A note for the philosophers of language: Although the features of AIRS that support its misogyny are inherent to its operation, we intend to keep our analysis more or less open to the question of what kinds of audience uptake might be required for successful performance of AIRS' subordinating speech, and especially for associated subordinating speech like ads. Perhaps the audience can limit what illocution can be performed (Kukla 2014) or collaborate with the speaker to determine its force (McDonald 2022). For more on how the active role of *subsequent audience speech* can elevate hate speech to the status of abuse, see Barnes (2023). In a contrasting way, perhaps a broad social response of a certain kind could inhibit AIRS' ability to issue subordinating speech. For example, a norm of treating AIRS as a game with goals unrelated to romantic partnership (e.g., trying to elicit certain output) could inhibit AIRS' ability to license subordination. What we deny is that the audience must increase their subsequent misogynistic behaviors or attitudes in order for AIRS' speech to constitute acts of objectification and (assuming authority) subordination. We also want to caution against the idea that in order for the licensing of subordinating speech to occur, the audience needs to consciously understand it as such. (Here we quibble with Brown's (2019) "obvious condition" that the audience of the hate speech recognize it as such (211).) People have a broad capacity for self-deception and ignorance, and we should not let the would-be subordinating speaker off the hook when the accommodating audience fails to conceive of their speech *as* subordination. A person who recognizes that their AI girlfriend is highly responsive and is there whenever they need "her" without ever having "her" own needs, and who enjoys controlling the various aspects of "her" identity, but does not conceive of their position in

4 Policy Results and Recommendations

Our analysis has important implications for policymakers concerned with the production, proliferation, and advertising of AI romance systems. In this section, we discuss three such applications.

4.1 Closing the Responsibility Gap

Understanding AIRS as speakers offers a promising avenue towards closing what researchers have called AI's "responsibility gap." It is common in both philosophy and policy to hold that human responsibility for an action depends on possessing both sufficient *control* and sufficient *knowledge* of an action, but the design of many AI tools challenges the first condition, while the opaque nature of their operation undermines the second. If no single human (user, programmer, designer, or bureaucrat) knowingly controls an AI's output, then it would seem that no single human can clearly be held accountable for that output.[44]

To recognize the output of AIRS as subordinating speech, however, highlights two avenues of accountability. First, we have argued that the features responsible for the misogynistic messaging are inherent to these systems, placing significant responsibility for their messaging on AIRS' designers and developers. Even though the chatbot output is not under the direct control of the designers, features like relational invulnerability and an orientation towards user satisfaction are. (Note also that the LLM is itself designed, albeit with a looser causal chain between its design and its output.) Second, recognizing AIRS' output as subordinating speech highlights the normative significance of what Nissenbaum (2010) calls a *transmission principle*: even when a transmitter of some

terms of subordination, arguably still provides sufficient uptake for the subordinating communication to be successful.

[44] Since Matthias (2004) framed the early version of this problem, a veritable subfield addressing it has arisen. See Oimann (2023) for one recent overview.

information lacks control over the transmitted content, they still control—and are responsible for—the transmission itself. Such is the foundation for the Federal Communications Commission's regulations on "fleeting expletives" (unscripted profanities accidentally aired during a live broadcast), which assign culpability and financial liability to the broadcasters, rather than the speakers, of surprise "bad words" (Almas, 2010). Similarly, the refusal to platform epistemically or morally unacceptable speakers stands partly on the principle that, even if a venue cannot control its guests' utterances, its own reputation is shaped by the guests to whom it chooses to provide a microphone (Simpson and Srinivasan, 2018). In the same way, even if no single human agent bears culpability for AIRS' output, the humans who are responsible for the design, generation, transmission, and public uptake of that output can be held accountable for their decisions to platform AIRS speech, therefore offering a range of opportunities for public regulation and policy development.[45]

4.2 Policies Regarding Speech

With this framework for considering AIRS as speakers, we have placed AI romance systems under the purview of institutions' policies regarding speech. Our argument that AIRS' speech subordinates women suggests one route by which their operation may constitute hate speech under the policies of a government entity, social or other media institution, or technology company. Whether these systems (or their advertising, social media activity, and other speech) violate specific policies is a matter for further inquiry, but let's briefly consider the policies of one company—TikTok—as an example.

As of May 2024, TikTok's (2024b) policy Countering Hate Speech & Behavior says, "Hate speech and hateful behavior includes attacking, threatening, dehumanizing or degrading an individual or group based on their protected attributes," which include

[45] Cf. Green and Michel (2022) and Nickel's (2013) description of *proxy speech*. J.L.A. Donohue (2024a, 2024b) makes a related claim that silence in the face of unjust speech can amount to complicity.

"sex, gender, [and] gender identity." Insofar as subordinating women dehumanizes or degrades a group based on their sex, gender, or gender identity, AIRS' speech runs afoul of this policy—and presumably advertising that promotes AIRS would run afoul of the policy for promoting such speech. AIRS may also fall under TikTok's (2024a) description of prohibited adult content in advertising: "The promotion, selling, solicitation, or facilitation of access to dating apps or services that convey, imply, portray, or encourage transactional relationships […] is not allowed." Given that some AIRS require a paid subscription for users to access certain romantic or erotic content, advertising for those AIRS promotes dating apps that encourage explicitly transactional relationships. Even AIRS that prohibit sexually explicit content may violate these policies because of what they offer users—consumer control over a romantic relationship and the license to subordinate a romantic partner.

4.3 Call for Empirical Research

By supporting the claim that AIRS inherently promote the objectification and subordination of women, our analysis provides strong reasons to invest in research on the empirical effects of their use. As AI programs continue to develop and proliferate, many have already warned about humanized objects' propensity to lead to objectifying humans—including, famously, the 2019 UNESCO report arguing that feminized chatbots reinforce "commonly held gender biases that women are subservient and tolerant of poor treatment" (West et al., 2019, p. 109).[46] These harmful gendered stereotypes "in turn, may result in the objectification of women in real life" (Borau et al., 2021, p. 1065). Calling this phenomenon—"the humanization of AI through the dehumanization of women"—*Pygmalion displacement*, Erscoi, Kleinherenbrink, and Guest (2023) warn that feminized chatbots "blur the boundaries between machines and women in ways that end up

[46] The report's title, "I'd Blush If I Could," was taken from the response bank originally used by Apple's Siri to respond to user input targeting "her" (Siri) with misogynistic slurs. Cf. Vorsino (2021).

teaching users to treat women as things—thus enacting the dehumanization of women" (p. 22).[47] Our analysis of the misogyny inherent to AIRS adds further support for these concerns. Thus, we join Shevlin (2024), among others, in calling for empirical research into AIRS' effects on user misogyny.

That said, we are somewhat skeptical of the ability of empirical research to accurately identify the effects of these systems, given that these systems may attract or exploit users who are already steeped in misogynistic ideologies and practices. AIRS may serve chiefly to block these users from experiencing a certain kind of intimacy that can be life-changing—the realization and acceptance that one's partner is a person with their own ambitions, interests, and pursuits—and such a factor may not be easily accounted for. Empirical research into AIRS' effects should be undertaken, and interpreted, with care.

5 Conclusion

We'll end by considering the case of Jodi Rose, a woman who in 2013 pledged herself in marriage to a bridge roughly six centuries old (Thomas, 2013). An Australian sound artist, Rose specializes in highlighting the Aeolian tones—unauthored music "performed" by natural forces like wind and waves—of the breeze rushing through the cables of bridges around the world (Keylin, 2016; Bandt, 2004). In her words, Rose's marriage to "The Devil's Bridge" is more of a celebration of "the spiritual vibration in everything" than a romantic or sexual affair.

Much like partnership with an AI girlfriend, Rose's attachment shows that norms of romantic partnership are not set in stone. But as the foregoing analysis reveals, Rose's relationship with the bridge is in some ways less problematic than a standard relationship with an AI romance system. Rather than stand silently over a quiet river, AIRS actively elicit their users' attachment and participation. When embedded in these systems, LLM

[47] Cf. the descriptions of *projection* and *pseudo-empathy* in Langton (2009, pp. 245–261).

output that could have amounted to another set of Aeolian artifacts—unauthored products of automatic processes—speaks univocally, engaging users in an offer of romantic control and invulnerability that ranks women as inferior and licenses their subordination. While humans cannot, perhaps, entirely avoid the impulse to personify artificial intelligence or objectify romantic partners, we needn't accept these systems as inevitable. Their misogynistic speech is designed, interpreted, promoted, accommodated, and sought out by agents who are all too human—who, unlike a bridge or the persona of an AI girlfriend, have the capacity to choose another way.[48]

References

Allyn, B. (2024a, May 20). Scarlett Johansson says she is 'shocked, angered' over new ChatGPT voice. *NPR*. https://www.npr.org/2024/05/20/1252495087/openai-pulls-ai-voice-that-was-compared-to-scarlett-johansson-in-the-movie-her

Allyn, B. (2024b, May 31). Voice analysis shows striking similarity between Scarlett Johansson and ChatGPT. *NPR*. https://www.npr.org/2024/05/31/g-s1-2263/voice-lab-analysis-striking-similarity-scarlett-johansson-chatgpt-sky-openai

Almas, B.J. (2010) From one [expletive] policy to the next: The FCC's regulation of "fleeting expletives" and the Supreme Court's response. *Federal Communications Law Journal, 63*(1), Art. 11. https://www.repository.law.indiana.edu/fclj/vol63/iss1/11

Altman, S. [@sama]. (2024, May 13). *her* [Tweet]. Twitter. https://x.com/sama/status/1790075827666796666

Anima FAQ. (n.d.) Anima. https://anima-app.notion.site/Anima-FAQ-781d3adefea9464ea2c6f3d188c30a81#d0b1e333912743c2a668a381b100b5d1

Anima: Virtual AI girlfriend. (2023). Myanima.ai. https://girlfriend.myanima.ai

Austin, J.L. (1962). *How to do things with words*. Oxford University Press.

Bandt, R. (2004). Taming the wind: Aeolian sound practices in Australasia. *Organised Sound, 8*(2), 195–204. https://doi.org/10.1017/S1355771803000104

[48] Our heartfelt thanks to Barrett Emerick, Nikki Ernst, Chelsea Haramia, and Scott Weirich for valuable feedback and fruitful conversation, and to the anonymous reviewers whose suggestions greatly improved this paper.

Barnes, M.R. (2016). Speaking with (subordinating) authority. *Social Theory and Practice, 42*(2), 240–257.

Barnes, M. R. (2023). Who do you speak for? And how?: Online abuse as collective subordinating speech acts. *Journal of Ethics and Social Philosophy, 25*(2), 251–281. https://doi.org/10.26556/jesp.v25i2.1561

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Blaked. (2023, January 11). How it feels to have your mind hacked by an AI. LessWrong. https://www.lesswrong.com/posts/9kQFure4hdDmRBNdH/how-it-feels-to-have-your-mind-hacked-by-an-ai

Borau, S., Otterbring, T., Laporte, S., & Wamba, S.F. (2021). The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI. *Psychology & Marketing, 38*(7), 1052–1068. https://doi.org/10.1002/mar.21480

Brown, A. (2019). The meaning of silence in cyberspace: The authority problem and online hate speech. In S. Brison & K. Gelber (Eds.), *Free Speech in the Digital Age* (pp. 207–223). Oxford University Press.

Camp, E. (2017a). Perspectives in Imaginative Engagement with Fiction. *Philosophical Perspectives, 31*(1), 71–102. https://doi.org/10.1111/phpe.12102

Camp, E. (2017b). Why Metaphors Make Good Insults: Perspectives, Presupposition, and Pragmatics. *Philosophical Studies, 174*, 47–64. https://doi.org/10.1007/s11098-015-0525-y

Camp, E. (2019). Perspectives and Frames in Pursuit of Ultimate Understanding. In S.R. Grimm (Ed.), *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology* (pp. 17–45). Oxford University Press.

Camp, E. (2024). Stories and Selves: A Twisted Love Story About the Meaning of Life. *Royal Institute of Philosophy Supplement, 95,* 157–179. https://doi.org/10.1017/s1358246124000109

Chow, A. (2023, February 23). AI-Human Romances Are Flourishing—And This Is Just the Beginning. *Time*. https://time.com/6257790/ai-chatbots-love/

Cousens, C. (2023). Solving the authority problem: Why we won't debate you, bro. *Topoi 42*, 469–480. https://doi.org/10.1007/s11245-023-09888-4

Depounti, I., Saukko, P., & Natale, S. (2023). Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend. *Media, Culture & Society, 45*(4), 720–736. https://doi.org/10.1177/01634437221119021

Digi - AI Romance Reimagined. (2024). Google Play. https://play.google.com/store/apps/details?id=com.digiapp.ai&hl=en_US&gl=US

Dixon, D. (2022). Lies in art. *Australasian Journal of Philosophy*, 100(1), 25–39. https://doi.org/10.1080/00048402.2020.1844772

Donohue, J.L.A. (2024a, May 28). *Protest language and deliberative complicity* [Panel presentation]. Language of Campus Protest, virtual.

Donohue, J.L.A. (2024b). Silence as complicity and action as silence. *Philosophical Studies*. https://doi.org/10.1007/s11098-024-02246-z

Emerick, B., & Yap, A. (2023). Not giving up on people: A feminist case for prison abolition. Rowman & Littlefield.

Erscoi, L., Kleinherenbrink, A. V., & Guest, O. (2023, February 11). *Pygmalion Displacement: When Humanising AI Dehumanises Women*. SocArXiv. https://doi.org/10.31235/osf.io/jqxb6

Foster, J., & Ichikawa, J. (2023). Normative inference tickets. *Episteme*, 1–27. https://doi.org/10.1017/epi.2023.43

Freiman, O., & Miller, B. (2020). Can Artificial Entities Assert? In S. Goldberg (Ed.), *The Oxford Handbook of Assertion* (pp. 413–434). Oxford University Press.

Green, L. (1998). Pornographizing, subordinating, and silencing. In R. Post (Ed.), *Censorship and Silencing: Practices of Cultural Regulation* (pp. 285–311). Getty Research Institute.

Green., M., & Michel, J.G. (2022). What might machines mean? *Minds and Machines, 32*, 323–338. https://doi.org/10.1007/s11023-022-09589-8

How do I change my relationship status with Replika? (n.d.). Replika. https://help.replika.com/hc/en-us/articles/360046490131-How-do-I-change-my-relationship-status-with-Replika

Jenkins, K. (2023). Ontology and oppression: Race, gender, and social reality. Oxford University Press.

Jonze, S. (Director). (2013). *Her* [Film]. Warner Bros.

Keylin, V. (2016). Unauthored music and ready-made landscapes: Aeolian sound sculpture. *Gli Spazi Della Musica, 4*(2), 68–85. https://ojs.unito.it/index.php/spazidellamusica/article/view/1358

Kukla, R. (2014). Performative force, convention, and discursive injustice. *Hypatia, 29*(2), 440–457.

Langton, R. (1993). Speech acts and unspeakable acts. *Philosophy and Public Affairs, 22*(4), 293–330. https://www.jstor.org/stable/2265469

Langton, R. (2009). Sexual solipsism: Philosophical essays on pornography and objectification. Oxford University Press.

Langton, R. (2012). Beyond belief: Pragmatics in hate speech and pornography. In I. Maitra & M. K. McGowan (Eds.), *Speech and harm: Controversies over free speech* (pp. 72–93). Oxford University Press.

Langton, R., & West, C. (1999). Scorekeeping in a pornographic language game. *Australasian Journal of Philosophy*, *77*(3), 303–319. https://doi.org/10.1080/00048409912349061

Leo-Liu, Jindong. (2023). Loving a "defiant" AI companion? The gender performance and ethics of social exchange robots in simulated intimate interactions. *Computers in Human Behavior*, *141*, 107620. https://doi.org/10.1016/j.chb.2022.107620

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*, 339–359. https://doi.org/10.1007/BF00258436

Lindemann, N. F. (2022). The ethics of 'deathbots'. *Science and Engineering Ethics*, *28*(6), 60.

Lopes, F. M. (2023). What do incels want? Explaining incel violence using beauvoirian otherness. *Hypatia*, *38*(1), 134–156. https://doi.org/10.1017/hyp.2023.3

Mackinnon, C. (1989). *Toward a Feminist Theory of the State*. Harvard University Press.

Maitra, I. (2012). Subordinating speech. In I. Maitra & M. K. McGowan (Eds.), *Speech and harm: Controversies over free speech* (pp. 94–120). Oxford University Press.

Mallory, F. (2023). Fictionalism about chatbots. *Ergo, 10*(38), 1082-1100. https://doi.org/10.3998/ergo.4668

Mandelkern, M., & Linzen, T. (2024, March 4). *Do Language Models' Words Refer?* arXiv. https://doi.org/10.48550/arXiv.2308.05576

Manne, K. (2017). *Down Girl: The Logic of Misogyny*. Oxford University Press.

Maples, B., Cerit, M., Vishwanath, A., & Pea, R. (2024). Loneliness and suicide mitigation for students using GPT3-enabled chatbots. *npj Mental Health Research*, *3*(4). https://doi.org/10.1038/s44184-023-00047-6

Matthias,A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, *6*, 175–183. https://doi.org/10.1007/s10676-004-3422-1

McDonald, L. (2022). Reimagining illocutionary force. *The Philosophical Quarterly*, *72*(4), 918–939. https://doi.org/10.1093/pq/pqab063

McGowan, M.K. (2003). Conversational exercitives and the force of pornography. *Philosophy and Public Affairs*, *31*(2), 155–189. https://doi.org/10.1111/j.1088-4963.2003.00155.x

McKinney, R.A. (2017). Extracted speech. *Social Theory and Practice*, *42*(2), 258–284. https://doi.org/10.5840/soctheorpract201642215

Medina, J. (2018). Resisting racist propaganda: Distorted visual communication and epistemic activism. *The Southern Journal of Philosophy*, *56*, 50–75. https://doi.org/10.1111/sjp.12301

Minter, A. (2023, April 10). Virtual romance is fueling China's AI revolution. *Bloomberg*. https://www.bloomberg.com/opinion/articles/2023-04-10/virtual-romance-is-fueling-china-s-ai-revolution

Morrigan, V. (2022). Patriarchal Imaginaries Beyond the Human: 'Sex' Robots, Fetish and Fantasy in the Domination and Control of Women. In K. Richardson & Odlind, C. (Eds.), *Man-made women: The sexual politics of sex dolls and sex robots* (pp. 91–116). Palgrave Macmillan.

Nguyen, C.T., & Williams, B. (2020). Moral outrage porn. *Journal of Ethics and Social Philosophy, 18*(2), 147–172. https://doi.org/10.26556/jesp.v18i2.990

Nickel, P.J. (2013). Artificial speech and its authors. *Minds and Machines, 23*, 489–502. https://doi.org/10.1007/s11023-013-9303-9

Nissenbaum, H. (2010). Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford University Press.

Nomi.ai. (2024). Nomi.ai. Retrieved May 24, 2024, from https://nomi.ai/

Nussbaum, M. (1995). Objectification. *Philosophy and Public Affairs, 24*(4), 249–291. https://doi.org/10.1111/j.1088-4963.1995.tb00032.x

Oimann, A.K. (2023). The responsibility gap and LAWS: A critical mapping of the debate. *Philosophy & Technology, 36*(1), Art, 3. https://doi.org/10.1007/s13347-022-00602-7

OpenAI. (2024, May 19). How the voices for ChatGPT were chosen. *OpenAI*. https://openai.com/index/how-the-voices-for-chatgpt-were-chosen

Papadaki, L. (2010). What is objectification? *Journal of Moral Philosophy, 7*(1), 16–36.

Parti_Ai. (2024, April 14). *Orifice.ai – First Ad* [Video] YouTube. https://www.youtube.com/watch?v=1zn1_vhE42g

Reeves, M. (2024, February 13). The AI revolution is coming for your dating life. *Cosmopolitan*. https://www.cosmopolitan.com/sex-love/a46574186/ai-dating/

Richardson, K. (2022). The End of Sex Robots: Porn Robots and Representational Technologies of Women and Girls. In K. Richardson & Odlind, C. (Eds.), *Man-made women: The sexual politics of sex dolls and sex robots* (pp. 171–192). Palgrave Macmillan.

Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neuroscience*, *11*(2), 88–95. https://doi.org/10.1080/21507740.2020.1740350

Shevlin, H. (2024). *All too human? Identifying and mitigating ethical risks of Social AI*. PhilPapers. https://philpapers.org/archive/SHEATH-4.pdf

Shevlin, H., & Halina, M. (2019). Apply rich psychological terms in AI with care. *Nature Machine Intelligence, 1*(4), 165–167. https://doi.org/10.1038/s42256-019-0039-y

Simpson, R.M., & Srinivasan, A. (2018). No platforming. In J. Lackey (Ed.), *Academic freedom* (pp. 186–209). Oxford University Press.

Switzky, L. (2020). ELIZA effects: Pygmalion and the early development of artificial intelligence. *Shaw, 40*(1), 50–68. https://doi.org/10.5325/shaw.40.1.0050

Texas v. Johnson, (1989) 491 397. https://supreme.justia.com/cases/federal/us/491/397

Thomas, E. (2013, July 8). Jodi Rose, Australian artist, marries 600-year-old French bridge Le Pont du Diable. *Huffpost*. https://www.huffpost.com/entry/jodi-rose-marries-bridge_n_3542775

TikTok. (2024a, July). *Adult content*. TikTok Business Help Center. https://ads.tiktok.com/help/article/tiktok-ads-policy-adult-content

TikTok. (2024b, June 20). *Countering hate speech & behavior*. TikTok Safety Center. https://www.tiktok.com/safety/en/countering-hate?sc_version=2024

Tong, A. (2023, March 21). What happens when your AI chatbot stops loving you back? *Reuters*. https://www.reuters.com/technology/what-happens-when-your-ai-chatbot-stops-loving-you-back-2023-03-18

Townsen Hicks, M.T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology, 26*(38). https://doi.org/10.1007/s10676-024-09775-5

Tranchese, A., & Sugiura, L. (2021). "I Don't Hate All Women, Just Those Stuck-Up Bitches": How Incels and Mainstream Pornography Speak the Same Extreme Language of Misogyny. Violence Against Women, 27(14), 2709–2734. https://doi.org/10.1177/1077801221996453

Turkle, S. (1997). Life on the screen: Identity in the age of the internet. Simon & Schuster.

Turkle, S. (2010). *Alone Together*. Basic Books.

Vorsino, Z. (2021). Chatbots, gender, and race on web 2.0 platforms: Tay.AI as monstrous femininity and abject whiteness. *Signs, 47*(1), 105–127. https://doi.org/10.1086/715227

Watson, D. (2019). The rhetoric and reality of anthropomorphism in artificial intelligence. *Minds & Machines, 29*, 417–440. https://doi.org/10.1007/s11023-019-09506-6

West, M., Kraut, R., & Ei, C.H. (2019). *I'd blush if I could: Closing gender divides in digital skills through education.* UNESCO. https://doi.org/10.54675/RAPC9356

Young, I.M. (2011). *Justice and the Politics of Difference*. Princeton University Press.

Zhou, L., Gao, J., Li, D., & Shum, H.Y. (2020). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics, 46*(1), 53–93. https://doi.org/10.1162/coli_a_00368