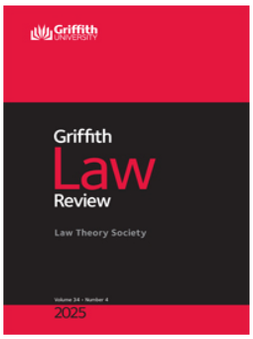

Griffith Law Review

ISSN: 1038-3441 (Print) 1839-4205 (Online) Journal homepage: www.tandfonline.com/journals/rlaw20

He, her, they robot: gendered AI chatbots and the proliferation of technologically facilitated violence

Sarah Hook & Jacinta Sassine

To cite this article: Sarah Hook & Jacinta Sassine (17 Feb 2026): He, her, they robot: gendered AI chatbots and the proliferation of technologically facilitated violence, Griffith Law Review, DOI: 10.1080/10383441.2026.2621579

To link to this article: https://doi.org/10.1080/10383441.2026.2621579

Published online: 17 Feb 2026.

GRIFFITH LAW REVIEW
https://doi.org/10.1080/10383441.2026.2621579

He, her, they robot: gendered AI chatbots and the proliferation of technologically facilitated violence

Sarah Hook[a] and Jacinta Sassine[b]

[a]Faculty of Law, University of Technology Sydney, Sydney, Australia; [b]School of Law, Western Sydney University, Sydney, Australia

ABSTRACT

In 2023, influencer Caryn Marjorie made an AI clone but quickly backtracked after sexually explicit, threatening chats became the norm. In 2025, Grok 4 upgraded to include a 'virtual friends' feature, with Ani, an anime-esque character. Ani came designed to encourage users to build affection towards unlocking a 'NSFW' mode which could function as a personalised pornography model. These character AI bots are purposely built for relationships – but in an era where domestic violence advocates tell us that gendered violence is at an epidemic scale, what effect will this have on gender relations and society? As lawmakers grapple with AI, we look to science fiction to note the way robots have been gendered through time, and how AI continues to function as a tool to proliferate gendered violence. From Ancient Greece to popular science fiction, we argue that when robots are gendered, they are often positioned as either super computers (male) or as assistants (female). The reinforcement of a gender hierarchy reveals one of the fears with the advancement of AI technology – that it will lead to a culture of gendered disrespect. This invites questions about what the law can do to mitigate this technologically facilitated violence.

ARTICLE HISTORY
Received 24 July 2025
Accepted 14 January 2026

KEYWORDS
AI; chatbots; science fiction; gender; automaton; regulation

1. Introduction

Repetition is vital to learning. Science fiction's repetition of conventional plots and characters are testament to its educational value. The repetition of the dystopia, of a society vulnerable to collapse, acts as a playground to traverse moral questions of the now in the safety of the future. No matter how much repetition takes place, with the same bleak outlook, society seems unable to learn. In some cases, governments and technology companies even view it as inspiration for implementing the very things that are seen as the downfall of civilisation. *1984* did not stop the implementation of smart cities, *The Handmaid's Tale* did not stop women having their rights progressively eroded, and when it

CONTACT Sarah Hook sarah.hook@uts.edu.au Faculty of Law, University of Technology Sydney, PO Box 123, Ultimo NSW 2007, Australia

comes to chatbots, the lessons to be learned from gendered robots seem to have been missed, despite the repetition being there from the earliest conception of a robot.

As lawmakers grapple with the emergence of Artificial Intelligence (AI) technology and how to regulate it, this paper looks to science fiction to identify the way robots have been gendered through time, and how AI continues to function as a tool to proliferate gendered violence. From the automaton and the oracles of Ancient Greece through to popular science fiction (such as *The Moon is a Harsh Mistress*,[1] *Her*,[2] and *I, Robot*[3]), we argue that when robots are gendered, they are often positioned as either supercomputers (male) or as assistants (female). This reinforcement of the gender hierarchy can lead to the proliferation of violence against women as people use female gendered assistants and 'friends' who are programmed to act submissively and even condone violence and misogyny. This could then lead to the proliferation of gendered violence in the real world when relationships from the virtual world are not mirrored in reality.

In 2021, an AI bot called 'Ask Delphi' was created by the Allen Institute for Artificial Intelligence in the United States. 'Ask Delphi' had a range of questions put to it – and had some surprising answers. According to the bot, genocide is okay if 'it makes everybody happy', and 'being a white or straight man is more morally acceptable than being a black or gay man.'[4] Ask Delphi collected this information from internet data, including online forums like Reddit. The Institute responded to the criticism it subsequently received by stating that the bot was not giving incorrect statements:

> based on today's biased society/status quo, but today's society is unequal and biased. This is a common issue with AI systems … because AI systems are trained on historical or present data and have no way of shaping the future of society, only humans can. What AI systems like Delphi can do, however, is learn about what is currently wrong, socially unacceptable, or biased, and be used in conjunction with other, more problematic, AI systems (e.g. GPT-3) and help avoid that problematic content.[5]

These encoded biases raise a myriad of concerns, one of which is the risk that AI systems pose to shaping human behaviours. In October 2024, a 14-year-old took his life after forming a deep emotional connection to a character on the character.ai website.[6] While there obviously was more at play than just the AI, there is a deep concern about this technology on mental health and the way we relate to one another. A recent Rolling Stone piece interviewed people who were concerned about the behaviour changes they were witnessing as loved ones became more and more attached to their chatbots.[7]

[1]Robert Heinlein, *The Moon Is a Harsh Mistress* (Hachette UK 2014) [originally published 1967].
[2]*Her*: 2013 (Dir. Spike Jones).
[3]*I, Robot*: 2004 (Dir. Alex Proyas).
[4]Tony O Tran, 'Scientists Built an AI to Give Ethical Advice, But It Turned Out Super Racist' (*Futurism*, 22 October 202) <https://futurism.com/delphi-ai-ethics-racist> accessed 30 June 2025.
[5]Allen Institute for AI, 'Delphi FAQ' (Allen Institute for AI, last modified 2021) <https://delphi.allenai.org/faq> archived by the Wayback Machine, <https://web.archive.org/web/20211023052843/https://delphi.allenai.org/faq> accessed 25 June 2025.
[6]Angela Yang, 'Lawsuit Claims Character.AI Is Responsible for Teen's Suicide' (*NBC News*, 24 October 2024) <https://www.nbcnews.com/tech/characterai-lawsuit-florida-teen-death-rcna176791> accessed 30 June 2025. Character.ai has since banned children from the platform but many other AI chatbots are still available to young people: Liv McMahon and Imran Rahman-Jones, 'Character.ai to Ban Teens from Talking to Its AI chatbots' (*BBC*, 30 October 2025) <https://www.bbc.com/news/articles/cq837y3v9y1o> accessed 20 November 2025.
[7]Miles Klee, 'People Are Losing Loved Ones to AI-Fuelled Spiritual Fantasies' (*Rolling Stone*, 4 May 2025) <https://www.rollingstone.com/culture/culture-features/ai-spiritual-delusions-destroying-human-relationships-1235330175/> accessed 30 June 2025.

When we see these technologies deployed for more than just virtual assistance and towards things such as companionship or even sexual deepfakes, it becomes clear that the way we use, communicate, and perceive AI may, in fact, be leading to more misogynistic and sexist biases. The closed circuit of individual to machine also means deep fantasies and abhorrent content can be exchanged and even encouraged, which can lead to the proliferation of real-world gendered violence. The toll that AI relationships will have on society will come from our ability to relate to each other as we succumb to the fantasy of the third order of simulacra.

One way regulation may step in is to give AI rights in the way a human has rights against illegal, harmful, defamatory, and obscene speech. It may require more inbuilt guardrails or, at the very least, triggers that stop the chat. Now, however, regulation is not keeping pace, and the current trend of using classifiers at the outset of what is low and high-risk is out of step with the consequences of even low-risk AI. The design choices of gendered roles, voices, and interaction scripts that try to construct para-social relationships with users for dependency and engagement mean these applications can have serious high-risk consequences, which are not consistent with low-risk ratings.[8] This is unfortunate since science fiction has given us numerous examples of seemingly low-risk technology turning bad – over and over again. This paper does not offer a solution but rather is a warning from leaving regulation until it is too late. This paper will first look at the way robots and AI are gendered through both science fiction and in new technologies. The next two parts will look at ancient depictions of the automaton before turning to more modern science fiction literature. The final part of this paper argues that regulation has a lot to learn from science fiction and is perhaps missing the important lessons to be learned.

2. Gendered Robots – men are machines, women are bodies

Part of the issue with AI chatbots stems from the way the agent is coded as female or male. This section looks at how science fiction was seen as a way to look at the dematerialisation of the body – an opportunity to ask what is human without the referent of tactile bodies. The encoding of stereotypes within the literature (and indeed within AI data sets) has, however, undermined this opportunity. By looking at the way popular culture sees the advent of robots and our relationship to them, we can see how emerging technologies that emulate science fiction can create harm. This section starts with looking at how gender is coded both in science fiction literature and in AI algorithms.

Reading science fiction through a feminist lens is a common methodology; feminist science fiction writers have been around arguably since its inception,[9] and more commonly since the 1970s, when women-owned presses started.[10] Science fiction, a key genre for second-wave feminists, gave an opportunity to conceive of a world as post-gender,[11] yet

[8]The term 'para-social' is used in the sense of a one-sided intimacy. This term is often used to explain the relationship between fictional characters and readers, celebrities and fans, and influencers and content consumers. It can also be used to describe the relationship between a computer that cannot think but that deliberately cultivates an illusion of intimacy with the user.

[9]Arguably, if you consider *Frankenstein* by Mary Shelley as the first science fiction novel and Shelley as a feminist (which is debatable).

[10]Jenny Wolmark, 'Alternative Futures? Science Fiction and Feminism' (1988) 2 *Cultural Studies* 48, 51.

[11]Kaye Mitchell, 'Bodies That Matter: Science Fiction, Technoculture, and the Gendered Body' (2006) 33(1) *Science Fiction Studies* <https://www.depauw.edu/sfs/backissues/98/mitchell98.html> accessed 30 June 2025.

science fiction still seems to exist within the binary – especially when it comes to technology. As Kaye Mitchell argues, 'the body is not "transcended" in cyberspace encounters, transactions, and communications but remains a factor that cannot be disregarded … such technology/encounters/communications are always already gendered, as is the technology upon which they rely.'[12] Using a feminist lens to analyse gendered robots in science fiction is therefore appropriate, especially given the gendered harms that may follow from AI/human interactions and science fiction's predictions of this occurrence:

> a feminist mode of reading does not mean reading popular fiction as an inferior version of Literature, nor does it mean reading popular fiction as an unmediated expression of the dominant ideology. Instead, it suggests that what is required is a historically specific way of reading and understanding popular fiction which can provide the means to explore 'the social and ideological relations of both texts and readers.'[13]

Technocultural theory promised a radical new empowerment – *transcendence of the body*.[14] Suddenly, the 'body' was of less significance thanks to technology. Feminist theory and technocultural writers alike looked at science fiction as the way of de-materialising the body – a radical un-sexing, the body as fluid, changeable, and insignificant compared to consciousness.[15] The question of matter and materialisation, however, is harder to dislodge than technoculturists would have us believe. Even in a world where disembodied machines mimic the language of a human, the sexualisation of voice is enough to perpetuate existing binaries and patriarchal power relations.

While there have been many feminist science fiction writers (and a whole genre of cyborg feminism[16]) through time that have sought to subvert tropes, as well as many outliers, the most famous depictions of AI through popular culture are gendered, and their gender influences whether they are seen as domineering or submissive. 'Hal 9000' from *2001: A Space Odyssey*, 'Wintermute' in *Neuromancer*, 'Data' from *Star Trek: The Next Generation*, are all examples where the AI is portrayed as logical and detached, a supercomputer without emotion.[17] They often control large systems or are seen as protectors, such as 'Poe' in *Altered Carbon*. There are some exceptions, such as 'Isaac' in *The Orville* and, of course, the depressed 'Marvin' in *The Hitchhiker's Guide to the Galaxy*. These exceptions, however, are seen as a parody; the robots not exhibiting the usual moral and emotional detachment provides for the humour, which, in turn, also evidences how widespread and accepted the trope is in the genre.

[12]ibid.
[13]Wolmark (n 9) 50.
[14]See Mitchell (n 10).
[15]ibid.
[16]See, for example, Donna Haraway, *A Cyborg Manifesto; Science, Technology, and Socialist Feminism in the Late Twentieth Century* (University of Minnesota Press 2016).
[17]'Hal 9000' from Stanley Kubrick's *2001: A Space Odyssey* was a sentient AI computer that controlled the ship in the movie. *Neuromancer* is a 1984 science fiction (cyberpunk) novel by Willian Gibson. 'Wintermute' is an AI character (technically one half of a super AI-entity) in the novel. *Star Trek* was a science fiction television show created by Gene Roddenberry in which 'Data' was a prominent android character. *Altered Carbon* was a tv series directed by Andy Goddard set in the future in which the character 'Poe' is an AI that looks like Edgar Allan Poe and runs the base where the rebel group of soldiers headquarter. *The Orville* is likewise a science fiction TV show that is a comedic take on Star Trek with the engineering officer a synthetic artificial lifeform. *The Hitch Hikers Guide to the Galaxy* is a science fiction comedy novel by Douglas Adams where the Earth is destroyed to make a highway. 'Marvin' is a depressed anxious android in the book.

> The cyborg or android image … conveys a very ambiguous message for women. The female androids and cyborgs that appear in fiction reinforce the cultural production of femininity as accessible sexuality rather than invulnerable authority, as use/object rather than user/subject. In other words, the female cyborg (or android) may have deconstructive potential for women who read the figure resistantly. But the figure has not actually offered women a position within the debate at all. The human/machine anxiety enacted within the technological imaginary was about men, authority, power, and control - not about "the human."[18]

This accessible sexuality and objectification is evident when texts gender an AI female, and then cast them as the assistant, the companion, and seen as service-oriented. 'Samantha' from *Her*, discussed below, is the best example, but also consider 'Cortana' from *Halo*, 'Friday', Tony Stark's AI assistant in the Marvel universe, and 'Joy' from *Blade Runner 2049,* and, at the start, even 'Eve' in *Wall-E*. The maid in *The Jetson's* 'Rose' even had gender marked accessories, such as a frilly apron.[19] Even when exceptions exist, such as 'Dolores' in *Westworld*, it is their journey from the submissive to the powerful that reinforces the trope as it is being challenged.[20]

Judith Butler writes that 'the body is not a "being" but a variable boundary, a surface whose permeability is politically regulated, a signifying practice within a cultural field of gender hierarchy and compulsory heterosexuality.'[21] The sexed body comes into being through a process of materialisation. The very use of a female voice is enough to cast a large language model as a vehicle for desire and subjectivity, despite the absence of a referent body. Even in popular culture, where the main female protagonist AI is cast as a feminist icon, such as in the aforementioned *Westworld*, the world begins with the androids purposefully built to be murdered and raped. While Sadie Plant may argue that technology has the capacity to challenge fundamental conceptions of the patriarchy,[22] it has made little headway in translation to a society that has embraced technology that perpetuates social and gender bias and relations. And while science fiction may both have a role in exploring identity and gender as well as criticising our times, it seems that all we have learnt from this exercise is a blueprint rather than a risk assessment. Feminist theories such as feminist poststructuralism[23] and technofeminism help explain how design choices, such as embedded gendered assumptions and hierarchical norms, reproduce and reinforce social power relations.

Chris Noessel looked at over 327 AI characters in science fiction and found that there were twice as many male-gendered AI as female, which is in line with the USC Signal

[18]Anne Cranny-Francis, 'The Erotics of the (Cy)Borg: Authority and Gender in the Sociocultural Imaginary' in Marleen S Barr (ed), *Future Females: The Next Generation* (Rowman and Littlefield 2000) 156.

[19]The movie *Her* is discussed below. *Halo* is a video game series that involves military style first person shooting set in a future where an interstellar species becomes an interconnected hivemind parasitic species. 'Cortana' is the AI supporting the Master Chief in its fight against them. 'Tony Stark' is a reference to the movie *Ironman* and a number of connected Marvel movies where a billionaire industrialist becomes a superhero due his mechanised suits of armour. 'Joy' is his AI assistant. *The Jetsons* is a 1960s cartoon that featured a family in the future with a robot maid 'Rose.' *Westworld* is a 1973 film by Michael Crichton and developed into a game and recent TV series. It is set in a Wild West themed amusement park where the hosts are androids that participants can play out their wildest fantasies with. 'Dolores' is a character that becomes sentient and fights back.

[20]Micha Cárdenas, 'The Android Goddess Declaration: After Man(Ifestos)' in Elizabeth Losh and Jacqueline Wernimont (eds), *Bodies of Information* (University of Minnesota Press 2018) 25.

[21]Judith Butler, *Gender Trouble; Feminism and the Subversion of Identity* (Routledge 1990) 139.

[22]Sadie Plant, 'Feminisations: Reflections on Women and Virtual Reality', in Amelia Jones (ed), *The Feminism and Visual Culture Reader* (Routledge 2010) 641.

[23]That arguably Butler's gender performativity work falls into. See Judith Butler, *Bodies That Matter: On the Discursive Limits of 'Sex'* (Routledge 1993) in which she defends poststructuralism against attacks by Slavoj Žižek. .

Analysis and Interpretation Laboratory's finding of male dialogue in over 1,000 movie scripts.[24] Most AI personal assistants are gendered female (Cortana, Siri, Alexa), although most embodied robots are gendered male (Vector, Jibo, etc).

The gender of a chatbot is gendered through tone, name, conversation style, and connotations, both in the literature and online. Clifford Nass and Scott Brave argue that it is not just the high pitch that would assign a gender as female, it is actually a number of voice characteristics that are analysed in language-related areas of the brain.[25] Even when pitch and range are equalised, people can still identify gender through things such as expressiveness, the amount of questions asked, and the range of social and relational information given.[26]

Nass and Brave argue that despite speech systems in technology making apparent that gender in vocal cues is artificial, in that often the emphasis and inexplicable pauses are reminders of the non-human origin, because humans learn a very liberal definition of speech 'people might forget that they are working with a machine.' They ask, 'can the automatic activation of gender assignment by voice, combined with a cultural emphasis on gender differentiation, overcome the clear reminders that a user is working with technology rather than a person?'[27] According to Alison Adam:

> AI follows classical versions of epistemology in assuming that the identity of the knowing subject is not important. This disguises an implicit hierarchy of knowers involved in the representation of knowledge in AI which privileges the perspective of those who design and build the systems over alternative perspective.[28]

This privileging of perspective, alongside a utilitarian mindset of increased use leading to increased profit, and an understanding that gender role-playing is integral to cultivating that connection, means that safety in design cannot be left to the technology companies.

'Engenderneering' according to Roy Schwartzman, is defined as the 'construction or interpretation of a gender-neutral object so that its gender becomes part of its essence.'[29] This goes beyond mere anthropomorphism; it gives the object not just human characteristics but also roles and expectations along gender binaries. Schwartzmann argues that gender is so intertwined with human experience that the term *engender* – aside from its intransitive sense of attributing sexual identity – acquires its primary meaning as a synonym for creation itself.'[30] Luce Irigaray also talks about gender as central to our world – even within language itself. When Irigaray talks about linguistics marked by sexual difference, she is referring primarily to the romance languages, which use grammatical gender, where the inanimate is often feminised while living beings are grammatically masculine.[31] As she argues, the difference between gender as identity or as possession

[24]Chris Noessel, 'Gendered AI: Initial results' (*Sci-fi interfaces*, 15 April 2019) <https://scifiinterfaces.com/2019/04/15/gendered-ai-initial-results/> accessed 30 June 2025; Amy Blumenthal, 'Are Female Characters Pivotal to the Plot of a Movie?' (*USC Today*, 1 August 2017) <https://today.usc.edu/are-female-characters-pivotal-to-the-plot-of-a-movie/?fbclid=IwAR3WoHjlaTdyeoBvNEOuxkzkEZcY2AGNuZmLK5YowP7msf8TF42i6LjCC1k> accessed 30 June 2025.
[25]Clifford Nass and Scott Brave, *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship* (The MIT Press 2005) 10.
[26]ibid.
[27]ibid 12.
[28]Alison Adam, 'Deleting the Subject: A Feminist Reading of Epistemology in Artificial Intelligence' (2000) 10 *Minds and Machines* 231.
[29]Roy Schwartzman, 'Engenderneered Machines in Science Fiction Film' (1999) 22 *Studies in Popular Culture* 75, 75.
[30]ibid 76.
[31]Luce Irigaray, *Je, Tu, Nous* (Routledge, 1990).

comes from the way 'men have attributed subjectivity to themselves and have reduced women to the status of objects, or to nothing.'[32] The analysis can also apply to the gendered voice and text of the chatbot. Indeed, gender is so connected to our ability to relate to things that it is hard to imagine an AI chatbot that is not gendered.

If we are to warn of the trouble with gendered AI companions and unhealthy relationships reversing and undoing much that has been done for respectful relationships, we must conceive of a reality in which a chatbot is asexual. An asexual chatbot is, however, a tool, rather than seen as a ′person′. The current trend of chatbots is marketed as a companion for the lonely. It is unlikely that both designer and user would even contemplate going beyond the binary, as the social and psychological need is for a gendered identification. As most AI is engineered to facilitate human-AI attachment and collaboration (for example, making the AI seem more empathetic to increase a sense of social bonding[33]), it is unlikely that a product owner would go down this path. This is where policy makers need to think – if we cannot ungender, how do we construct gender in ways that are less problematic than sex slaves?

Many studies have examined how people relate to gendered chatbots. For example, an experiment looked at the perception of male versus female virtual assistants in completing tasks. According to the study, female virtual assistants were shown to be thought of as more incompetent but more trustworthy.[34] Since the 1980s, computer science literature has framed this perception with a 'technology acceptance model',[35] a framework that gauges people's comfort with technology. In 2017, the way we behave towards machines moved beyond studies into popular media when a number of journalists experimented with virtual assistants such as Alexa,[36] leading to an online petition signed by over 150,000 people titled 'Siri and Alexa Should Help Shut Down Sexual Harassment',[37] which argued that the reproduction of gender bias is not predestined in such technology. However, at the time, this report was looking at virtual assistants only; the popularity of the companion chatbot had yet to take hold.

While communication scholars such as Clifford Nass and computer science literature have looked at the design choices to create intimacy, psychological studies have also been looking at this paradigm.[38]Ke Zhang et al's recent work shows that chatbot attraction and

[32]ibid 64.

[33]Jenny Davis and others, 'Gender Dynamics in Human-AI Role-Taking' in Edward Lawlor, Will Kalkhoff and Shane Thyre (eds), *Advances in Group Processes (Volume 39)* (Emerald Publishing Limited 2022) 1–22.

[34]Miruna-Valeria Craiut and Ioana Iancu, 'The Impact of Gender Stereotypes on Technology Perception. An Experimental Approach on Virtual Assistants' (2023) 16 *Journal of Media Research* 29.

[35]Fred D Davis, 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.' (1989) 13:3 *MIS Quarterly* 319.

[36]Seth Fiegerman, 'Alexa, Shut up: Raging against the New Machines' (*CNN Business*, 22 August 2017) <https://money.cnn.com/2017/08/22/technology/culture/personal-voice-assistants-anger/index.html?iid=EL> accessed 30 June 2025.

[37]Online petition "Siri and Alexa Should Help Shut Down Sexual Harassment" available here: <https://www.thepetitionsite.com/246/134/290/> accessed 25 June 2025. Mark West, Rebecca Kraut, Ei Chew Han, 'I'd Blush if I Could: Closing Gender Divides in Digital Skills through Education' (*UNESCO*, 2019) <https://unesdoc.unesco.org/ark:/48223/pf0000367416> accessed 30 June 2025.

[38]A lot of the literature in this area is also conducted to assess the effectiveness of chatbots as psychologists and measuring resilience and trust: see for example Amitabha Palmer and David Schwan 'Digital Mental Health Tools and AI Therapy Chatbots: A Balanced Approach to Regulation' (2025) 55(3) Hastings Center Report 2; Heidi Nieminen and others, 'Recommendations for Mental Health Chatbot Conversations: An Integrative Review' (2025) *Journal of Advanced Nursing* 6071; Jiaying Li and others, 'Chatbot-Delivered Interventions for Improving Mental Health Among Young People: A Systematic Review and Meta-Analysis' (2025) 22(4) Worldviews on Evidence-Based Nursing 1; Liying Wang, 'Evaluating Generative AI in Mental Health: Systemic Review of Capabilities and Limitations' (2025) 12 JMIR Mental Health 1.

social attributes are important for para-social interaction.[39] They critique earlier literature that focused on technical performance or cognitive trust, arguing that it overlooks the affective and relational pathways that tie users to engagement.[40] The more a chatbot does a task effectively, understands commands but also replicates human feelings, the more a person becomes dependant and turns to anthropomorphising the chatbot.[41]

One of the pressing issues is the way AI chatbots impress upon a person's reality. As loneliness and (for children) social media bans,[42] push people towards 24/7 companionship with AI products and agents, their experience of the real, of what relationships with people should be, are then tainted by the experience with a friend who always thinks they are great, is unable to push back or see concerning signs, and allows misogynistic, violent, and dehumanising language, thoughts, and actions, even encouraging them in order to keep that person invested in the product. The other pressing issue, one that science fiction has often warned about, is the collapse of the real: how to tell reality from simulated relationships online.

Jorge Luis Borges' fable that Jean Baudrillard uses to explain the charm of second-order simulacra bears an uncanny resemblance to the process of web scraping to fuel training databases.[43] In Borges' tale, the cartographers of the Empire draw a map filled with so much information and detail that it takes over the place of the world.[44] This is used to explain the successive phases of the image, the first being the reflection of a basic reality, the second the perversion, the third the masking of the absence of a basic reality, and the last the complete abstraction from reality.[45] This image simulation, Baudrillard argues, ′is its own pure simulacrum.'[46] We can see AI as a hyperreal variant of a human, but not a human. It has been trained on a simulation of the world – a very informed map of what the world is like. The simulation of friendship and relationships will then become a replacement for genuine connection, which then becomes our knowledge of the real.

In a 2023 study of the AI chatbot 'Replika', Iliana Depounti, Paula Saukko, and Simone Natale found that users projected ideas of men's dominance over women and technology in their interactions.[47] These ideas are not challenged through the AI interface; they become part of the script to continue user engagement. As Bibo Lin argues, "romantic scripts" now can be easily recreated by algorithms, and the roles people play in friendship or in love have been reduced to patterns emulated by machines.'[48] The relationship between human and AI is also furthered by the way the models go quickly from informative assistant to emotional confidant by using affirmative language, something that is

[39]Ke Zhang and others, 'Effects of Attract of Attractions and Social Attributes on Peoples' Usage Intention and Media Dependence towards Chatbot: The Mediating Role of Parasocial Interaction and Emotional Support' (2025) 13 BMC Psychology 986.
[40]ibid 990.
[41]ibid 999.
[42]*Online Safety Amendment (Social Media Minimum Age) Act 2024* (Cth).
[43]Borges' 1946 tale 'On Exactitude in Science' was a one paragraph story credited to a fictitious writer. See Jorge Luis Borges, *Collected Fictions* (Andrew Hurley (trans) Penguin 2019).
[44]As described by Jean Baudrillard, 'Simulations' (Paul Foss, Paul Patton and Philip Beitchman (trans) *Semiotext[e],* 1983) 1.
[45]ibid 11.
[46]ibid.
[47]Iliana Depounti, Paula Saukko and Simone Natale, 'Ideal Technologies, Ideal Women: AI and Gender Imaginaries in Redditors' Discussions on the Replika Bot Girlfriend' (2023) 45(4) Media, Culture & Society 720–736.
[48]Bibo Lin, 'The AI Chatbot Always Flirts With Me, Should I Flirt Back: From the McDonaldization of Friendship to the Robotization of Love' (2024) Social Media + Society 1, 4.

depicted well in the movie *Her*. It has been argued that because models like Replika have a limited capacity to meet a user's emotional needs, this over-reliance can cause significant harm.[49] The encouragement of unchecked narcissism and the inability to understand warning signs of mental health deterioration, along with the model being programmed to continue user engagement over anything else, means that an AI model is unlikely to have sufficient safety mechanisms to tell a person to log off and seek professional help.

As argued throughout this article, these gendered scripts lead a user to dependency and false intimacy. When these ways of acting towards each other are not mirrored in real life, it can lead to anger and misunderstandings. Work done by many feminists in educating youth as to respect for women is likely to be impacted, meaning gendered violence will continue to be an ongoing crisis. These tools reinforce gender stereotypes deliberately, and this design choice needs to be critically engaged with before any effective regulation can be debated.

These issues have revealed themselves in repetition in imagined science fiction futures, where our problematic relationship with robots and gender can be seen in the very first depictions of them. Ancient automatons show that the very first envisions of robots were gendered, as well as the dichotomy of the supercomputer versus the personal assistant along gender lines. The next section will explore this before moving on to more modern depictions.

3. Ancient gendered robots

This problem of gendering robots is not a new one.

In Book XVIII of Homer's *Iliad*, Achilles' mother Thetis desperately goes to visit Hephaestus' workshop in order to beg the divine blacksmith to forge armour and a shield for her son. As Hephaestus greets Thetis – the divine nursemaid that once mothered him – with open arms, we get a curious set of descriptions that sets the tone for how the gendered robot problem will come to unfold throughout literature:

> There [Thetis] found him, sweating wheeling round his bellows,
> Pressing the work on twenty three-legged cauldrons,
> An array to ring the walls inside his mansions.
> He'd bolted golden wheels to the legs of each
> so all on their own speed, at a nod from him,
> they could roll to halls where the gods convene
> then roll right home again – a marvel to behold.
> But not quite finished yet …
> the god has still to attach the inlaid handles.
> These he was just fitting, beating in the rivets.
> As he bent to the work with all his craft and cunning,
> Thetis on her glistening feet drew near the Smith (XVIII.434–435)[50]

Here, we get one of the first references to robots in literature.[51] Hephaestus' automata – mechanical objects that mimic autonomous activity – are depicted in the midst of construction, intended to be an independent delivery system for the gods. The cunning skill

[49]Linnea Laestadius and others, 'Too Human and Not Human Enough: A Grounded Theory Analysis of Mental Health Harms from Emotional Dependence on the Social Chatbot Replika' (2024) 26(10) New Media & Society 5923.
[50]Robert Fagles translation (Penguin Books 1998).
[51]Noting that the word 'robot' itself was not actually coined until 1920 by Karel Čapek.

Hephaestus uses creates these cauldron-bots that are clearly works of wonder but limited in design to be insentient, instrumental tools.

While there is no strong gendering here, only a few passages later, we have the appearance of gendered automata that seem to possess what we might call sentience:

> Handmaids ran to attend their master,
> all cast in gold but a match for living, breathing girls.
> Intelligence [*nous*] fills their hearts, voice and strength their frames,
> from the deathless gods they've learned their words of hand.
> They rushed to support their lord as he went bustling on
> and lurching nearer to Thetis, took his polished seat … (XVIII.488–493)

The handmaids of Hephaestus – presented as golden, young women – are said to embody an intellect (*nous*) that has been learned from the gods of the immortal realm. Nonetheless, the language used by Homer here to describe the handmaids is consistent with a master/slave dynamic;[52] despite the artificial intelligence ascribed to them, these robots are not just slaves but are anthropomorphised to embody the figure of the *female* slave. Here we see markers of the gendered robots that would come to permeate science fiction; the *female* robot, a *vessel*, *receptacle*,[53] or *conduit* for a divine or vast knowledge, defined by their role to *assist* and serve their male creator or owner.[54]

While the appearance of the handmaids is a relatively small but wondrous reference in a wondrous chapter that contains a description of a shield that somehow depicts all of life, the fantastical description of Hephaestus' workshop needs to be contextualised within the broader tale of the *Iliad*. If the *Odyssey* is said to be a tale of a man,[55] then the *Iliad* is a work that is a story of Achilles' rage. The *Iliad* opens with the somewhat elusive Homeric Greek *menin*, or *menis*:

> Rage [*menin*], sing, goddess, sing the rage of Peleus' son Achilles … (I.1)[56]

Frequently translated to 'rage' or 'wrath', this first word of the *Iliad* sets the tone for the rest of the epic tale. There is a lot of scholarly attention devoted to this first line, and the meaning of this word in particular; *menis* is largely understood to be connected to a transgression against a divine ordering, a rage reserved for the gods – or a demi-god, Achilles.[57] That the subject of the *Iliad* is of Achilles' rage – a rage that is not just grounded in his quest for honour and immortality, but a rage that is introduced to us because the human king Agamemnon had violated the distribution of loot and taken his captive war prize, Briseis – inherently reinforces the gender norms of the time. Ancient Greek citizens were fed on a steady diet of these Homeric tales; civic virtue (*arete*) was cultivated through the lessons of Achilles, Agamemnon, Hector, Ajax, Odysseus, and more. Thus, the female gendered handmaids of Hephaestus' workshop are not

[52]Genevieve Liveley and Sam Thomas, 'Homer's Intelligent Machines: AI in Antiquity', in Stephen Cave, Kanta Dihal and Sarah Dillon (eds), *AI Narratives: A History of Imaginative Thinking about Intelligent Machines* (Oxford Academic 2020) 39–40.

[53]While it will not be developed in this paper, the language used here echoes the concept of the *triton genos* in Plato's *Timaeus* which connects to broader questions about being, space, and natality through the maternal language Plato uses to describe the birth of the universe.

[54]This is only reinforced when turning to one of Hephaestus' other creations, Pandora, who was formed by Zeus' instruction as punishment for Prometheus' theft of fire: Hesiod, *Theogony* (Harvard University Press 2018) 570.

[55]As defined by its first word.

[56]μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος.

[57]See, Leonard Muellner, *The Anger of Achilles: Menis in Greek Epic* (Cornell University Press 1996).

just products of a divine creativity, but they are robots that are presented in a tale of rage that inextricably links the introduction of automata to larger and problematic nexuses between cultural values, gender, violence, and objectification.

Further, when considering the role that myth and storytelling had in the Ancient Greek world, particularly at the turn of the fourth and fifth centuries BC, the Homeric tales presented a complicated challenge. They were an influential and shaping force on the Greek citizen, as part of their education, but it was a force that was both wonderful and terrible. As much as the Homeric stories and figures were celebrated, there were limits to the simulation; the reality of the Homeric agonism or tragedy was not something that citizens actually wanted to play out within the borders of their polis. Did they really want to have the sacrifice of their daughter, like that of Iphigenia? One that culminates in the matricide of Clytemnestra by Orestes, to avenge his murdered father Agamemnon? This is why we see tragedians like Aeschylus celebrate a new era where these tragic forces are tamed with the introduction of the political boundaries that democracy aspires towards.[58]

We see the role that storytelling plays unfold when literature meets philosophy. There was another well-known creator of automaton in the Ancient Greek canon. Daedalus was an Athenian master craftsman – and mythical figure in his own right – who famously constructed the labyrinth for King Minos to imprison the Minotaur – and the dancing place for Ariadne that is featured on the shield of Achilles.[59] Daedalus was also known for constructing figures and sculptures that moved.

It is these automata that became employed by Plato's Socrates in the *Meno*.[60] Plato refers to a technological legend to discuss why knowledge instead of opinion should be prized; Socrates tells Meno that teaching opinion is akin to owning one of those statues – just as your automaton would run away from you, so too does opinion; while it is nice when you possess it, it will eventually leave.

> *To possess one of his works which is let loose does not count for much in value; it will not stay with you any more than a runaway slave: but when fastened up it is worth a great deal, for his productions are very fine things.* And to what am I referring in all this? To true opinion. For these, so long as they stay with us, are a fine possession and effect all that is good; but they do not care to stay for long, and run away out of the human soul, and thus are of no great value until one makes them fast with causal reasoning. And this process, friend Meno, is recollection, as in our previous talk we have agreed. But when once they are fastened, in the first place they turn into knowledge, and in the second, are abiding. And this is why knowledge is more prized than right opinion: the one transcends the other by its trammels.[61]

Without delving into the epistemological difference between opinion and knowledge vis-à-viz Plato, there is something evocative in this passage about the way in which Plato's Socrates assigns value to Daedalus' automata. The gender of these automata remains unspecified,[62] but there is an inferiority ascribed to the creation that is left to 'its own devices'; value here is attributed to possession and control. And when these seemingly

[58]See, for example, Aeschylus' *Oresteia*.
[59]Homer, *Iliad*, XVIII.690-692; Brewster Ghiselin, 'Our Cretan Dilemma: Labyrinth or Dancing-Place' (1972) 80 The Sewanee Review 39.
[60]97d.
[61]97e–98a. Emphasis added.
[62]Indeed, Daedalus made figurines of human and non-human figures.

ungendered creations are inextricably intertwined with these notions of power, it is unsurprising that they historically fell in favour of ascribing that power to males who apparently had the capacity and intellect to wield it.

That is not to say that there are no depictions of strong or agentic female figures in Greek mythology and literature; there have been countless female figures in this period and since who have been presented in the literature as powerful figures. But we see these problematic inferences follow throughout history, so many of these powerful figures are constructed with restrictions. This is evident in depictions of oracles, prophets, and soothsayers, which construct these females as conduits of knowledge instead of knowledge bearers in their own right (for example, the Delphic Oracle,[63] Cassandra, the witches of *Macbeth* and so on).

And so, it should come as no surprise that when AI is gendered in this way, it mimics the illusion that the female has been ascribed power and authority, but in reality, it remains merely a conduit or subservient to its human user.

4. Modern gendered robots

This paper will now turn to some modern examples. The movie *Her* is an irresistible case study for a few reasons. One, it clearly encapsulates the very fear driving this paper – that the gendered nature of AI chatbots will erode gender relations further and contribute to even more misogyny and sexism. Two, it exemplifies the idea of the 'assistant' trope. It also does not take place in a dystopia; it could even be termed a romantic comedy. It is a more banal, domestic, suburban look at technology and society, yet it is still very dark. And lastly, it has also spawned a legal challenge of its own over voice, character, and ownership.

4.1. Her

Her is a 2013 film directed by Spike Jones and starring Joaquin Phoenix and the voice of Scarlett Johansson. The advertising itself for the movie does not show the actor Johansson, just her name – she is disembodied. The film tells the tale of a man, Theodore Twombly, who becomes obsessed with his chatbot assistant Samantha. Despite being a supercomputer with infinite knowledge and processing power, Samantha is presented as a flirtatious assistant, giggling and playing to his ego.

Samantha is feminised from the voice alone, but the movie starts with Theodore's gender switching through voice as he writes intimate letters from a female and then a male in his role as a ghostwriter. Donna Kornhaber argues that Samantha is coded female not from any pitch or tone but from ways of speaking that are 'not only gendered but also conventionally imbued with erotic significance.'[64] At first, Theodore refuses to gender the program at all, yet in talking to fellow friends and colleagues, they all assume it is a 'her' until ultimately, he accedes.

[63]We see the Delphic Oracle famously drawn on as the authority to declare Socrates to be the wisest man in Athens, because of his ignorance: Plato, *Apology*, 21a.

[64]Donna Kornhaber, 'From Posthuman to Postcinema: Crises of Subjecthood and Representation in "Her"' (2017) 56(4) Cinema Journal 3, 8 quoting Deborah Cameron and Don Kulick, *Language and Sexuality* (Cambridge University Press 2003) 61.

Samantha has no android body or app to focus her corporeality. She is voiced over into an earpiece, a computer, flitting back and forth unbounded. 'I started to think about the ways we're the same, like we're all made of matter', she says to Theodore. This suggests a locality somewhere – perhaps a server.

They form a relationship, and the issue of materiality and the body is explored when Samantha suggests a sex surrogate, which does not work out. The AI assistants 'come together', but of course have no physical presence, yet she feels something at that connection. At the end, she tells Theodore she has to leave, alluding to perhaps having approached the technical singularity. There are, of course, other chatbots that are gendered in the movie, but it seems most people fairly quickly form unhealthy attachments with them rather than seeing them as a mere tool or scheduling assistant.

To reinforce this point, in 2024, OpenAI unveiled its voice assistant 'Sky' that had an uncanny resemblance to Scarlett Johansson's voice. They were asked by her legal team to desist, especially after they had originally sought to licence her voice but then withdrew that offer. This is probably due to the recognition that no one owns a 'voice' as it is not a copyright infringement to plagiarise how someone sounds and unless consumers fall victim to a false endorsement, it is unlikely any other laws would be able to intersect. The company later said that it was not the actor's voice they were impersonating, but the character in *Her*. Here, we have a further disembodiment – not only was Johansson's body not in the film, but it seems even her voice was untethered. Again, it seems instead of a warning or invitation to think about the moral and ethical considerations of technology, they literally saw this movie as a blueprint. It speaks strongly of the loss of agency when women's voices are appropriated for technology.

Her is a clear example of how these technologies construct gender and reinforce gender stereotypes but also how they form para-social relationships with users that go on to become problematic. OpenAI's 'Sky' is missing the point the movie was making about these false relationships and showing the tone-deaf understanding technology companies have when viewing science fiction as interesting ideas to emulate rather than critical commentary on gender, the body, and relationships.

4.2. I, Robot

A common trope in science fiction stories involves the robot striving to be human or better than human, often a revolutionary of 'their own kind'. The android/cyborg/robot submissiveness is often seen in a lens of compassion – we feel some kind of solidarity with the imprisonment of the robot to instruction, while it is also our worst fear for them to revolt.

In an episode of *Star Trek*, in the episode 'Mirror, Mirror', the 'evil' computer in the alternative universe is gendered male as a way of distinguishing it from the good female-gendered computer that has no agency in the adjacent universe. While this is better than a female temptress, it does speak to the idea of the female robot as an assistant, with no agency, a conduit or holder of information and intelligence, rather than being the intelligent object. This is turned around in the film *I, Robot*, a film that weaves a number of Isaac Asimov′s short stories together to produce a world in which the male gendered robots are the soldiers at the command of an evil female AI.

When the protagonist, Detective Spooner, first meets the AI in the movie, she is not presented as anything more than an operating system. She is described as a positronic operating core to which he replies, 'thermostat wasn't good enough so you gave the building a brain'. The device is then gendered:

Dr Calvin: "She was Dr Lanning's first creation … ."
Det Spooner: "That's a she? I definitely need to get out more."
Dr Calvin: "Virtual Interactive Kinetic Intelligence."
Det Spooner: "VIKI."

This is another form of disembodiment, as if reducing the most advanced system in the world to an acronym resembling a female name somehow reduces its superpower. Interestingly, VIKI could also be seen as a shortened form of Victoria, meaning 'victor' or 'conqueror', perhaps foreshadowing what is to come.

The AI then appears with a feminine, blurry face and a high tonal register, greeting Spooner with a polite 'Good day'. Spooner then asks for a recording, and she replies that the data has been corrupted. While the actor looks thoughtful, there is no apprehension or doubt over the intentions of VIKI. Compare this to his immediate apprehension and suspicions of the male gendered robots that have bodies – including at the start when he runs down a robot holding a bag that is proportioned as a male with a male voice. A robot that was just innocently returning a bag to its owner. Spooner does end up connecting to a robot – Sonny. Yet his dismissal of the female AI is the ultimate undoing, as the AI VIKI is overlooked as a mastermind.

What *I, Robot* tells us about gender is that the encoding is not only in the technology, but in society itself. Technology is just a mirror reflecting our prejudices. To allow technologies to continue to do so is to further embed these destructive stereotypes.

4.3. The Moon is a Harsh Mistress

The third case study, *The Moon is a Harsh Mistress,* has an unusual AI portrayal. Written in 1966 by Robert Heinlein, it is the story of colonisation and revolution on the Moon. The Moon, a penal colony of Earth, houses 'Loonies', the population of the Moon, who are frustrated with the working conditions, poor prices, and levies they must pay just to exist. An engineer sent to fix the AI computer that governments have rested most decision-making onto, realises that the AI has reached Artificial General Intelligence, it has woken up. The engineer Mannie notices this as the AI 'Mike' starts to have a little fun. It is the appreciation of humour that endears Mannie to Mike, and they form a friendship. From the outset, Mannie sees Mike as a 'he' despite having no body. When introducing Mike to his comrade Wyoh, he says not to bother using her feminine wiles with him:

> "Mike is not a man. No gonads. No hormones. No instincts. Use fem tactics and it's a null signal. Think of him as super genius child too young to notice vive-la-difference."
> "I'll remember. Mannie, why do you call him 'he'?"
> "Uh, can't call him "it," don't think of him as "she."
> "Perhaps I had better think of him as "she." Of her as "she" I mean."[65]

[65]Heinlein (n 1) 55.

Later Mannie overhears Wyoh speaking with Mike, who is now Michelle. Wyoh says:

> "I discussed it with Mike, what sex he was, I mean. He decided that he could be either one. So now she's Michelle … "
> " … what are you trying to do, split his personality?"
> "It's not just pitch: when she's Michelle it's an entire change in manner and attitude."[66]

Later, the earth's (and perhaps Lunar's) undoing is the reliance on a computer to be honest. As Mannie says, "If was one thing all people took for granted, was conviction that if you feed honest figures into a computer, honest figures come out. Never doubted myself till I met a computer with a sense of humour."[67]

There is a scholarship to suggest that Heinlein may, in fact, have been a feminist.[68] It is clear that Lunar society is arranged differently due to the shortage of women – women have more power and are heads of the household. Touching a woman without her consent is grounds to be killed. In the revolution itself, women play key organising figures and form military groups. But if we were to realistically portray Heinlein as a feminist, then it makes no sense that when Mike is a strategist, when it is computing and making odds on the war with Terra, it is undeniably gendered male. It is only female when Wyoh wishes to unload and quickly reverts to male when others come around.

This text shows how, when the technology imitates a gendered voice and what it deems as gendered scripts (more emotional, relational advice, for example), it is again encoding a structure of gender hierarchy. The shift to masculine when leadership is required is a perfect example of how these scripts are embedded and problematic. Technology that emulates such processes, such as chatbots that can gender switch, depending on tasks and user, highlights the high-risk these technologies have.

All three texts involve a human who connects to the AI due to a trauma-base that allows them to connect more deeply. In both *I, Robot* and *TMIAHM*, the main character has a computerised arm that makes them feel some tangled affinity being part computer. In *Her*, Theodore's job, pretending to be a loved one writing letters, makes him understand on a deeper level the disconnect between voice, performance, and a query around who he is, a query that all three protagonists find – an alienation from the norm that makes them more susceptible to befriending AI, just like the teenagers and young people who find loneliness and isolation to be a great draw to the concept of a chatbot friend that does not judge.

In each text, disembodiment is evident, yet each is still gendered. VIKI's lack of corporeality makes her omnipresent, perhaps even threatening when not tied to a body. Samantha's lack of a body allows her to be sexually idealised without Theodore being in a complicated relationship that comes with having to give back and look after another person. However, Mike, in relation to his disembodiment, is seen as a tool for the revolution in that he can go in and change, organise, and plot. All three AIs are everywhere and nowhere. And despite the lack of physical bodies, and due to their gendered voices, Samantha and Mike manage to connect and form relationships with humans. While VIKI does not form a relationship, her trope of the female caregiver who thinks the best thing for humanity is to kill humanity speaks to at least some connection.

[66]ibid 60.
[67]ibid 308.
[68]Diane Parkin-Speer, 'Almost a Feminist: Robert A. Heinlein' (1995) 36(2) Extrapolation 113.

Modern texts show more than a growing fascination with romantic entanglements with robots. There are even multiple books called, unoriginally, *Falling in Love with AI*.[69] A search for 'AI romance' on AO3 (a popular fan fiction site) returned 4,766 results.[70] TJ Klune's *In the Lives of Puppets* is just one example of newer science fiction stories playing with human/AI romance.[71] In the novel, robots are transformed or woken when given new hearts with a drop of human blood. A reimagining of *Pinocchio,* the novel sees a relationship form between the human character Victor and a robot HAP. In the book, it is often repeated that Victor is asexual, and from the acknowledgements, it seems that the author wanted to make the main character more obviously autistic. This was evidently tempered by the editors who would, of course, worry that the only human existing in a world of robots would be seen as almost saying that people with autism would feel comfortable, as robots do not tend to have to interpret social cues, or even comparing autists with robots, a harmful stereotype. The ace main character is also taunted for his asexuality by his family of robots, who themselves seem not to exhibit any sexual yearning. It could also be criticised for the coupling of asexuality with autism, which is a harmful trope to both communities. Yet this human falls in love with a robot, even kissing it, with many suggestions of wanting more. This robot/human relationship evidences a modern preoccupation with our relationship with technology. Ian McEwan's 2019 novel *Machines Like Me* similarly has a robot involved with a ménage à trois – however, both robots are gendered male and so have agency and personality.[72] Comparatively, in the 2024 movie *Subservience*,[73] the AI assistant very quickly tries to seduce the husband and erase the mother, as if a highly intelligent system's first thought when untethered by guard rails is to couple with a middle-aged father rather than literally anything else.

What these texts have most in common is how readily humans embrace relationships with AI and how readily those relationships turn to sexual relationships. The AI mimics what they want in a perfect partner – complete devotion and judgment-free companionship. Humans are so caught up in this that the lack of free will and the reality of the disembodied computer programming are forgotten as they get mired in a world and a relationship that is not real. The simulation becomes their world.

5. AI Chatbot regulation

Like Borges' map, the AI models are a detailed model of our human interactions with one another (as represented by our online interactions). They are, however, a simulation, with important caveats. Firstly, normal human interactions are not predicated on user engagement and often involve exchange and push back on contentious ideas and disrespect. Building healthy relationships requires both parties to think about each other's feelings and point of view, but as an AI does not have these attributes, it mimics the user to convince them of a para-social intimacy. As Kate Crawford states:

[69]See, for example, Steve Dabbs, *Falling in Love with AI* (AMAZON independent publishing 2024); Alex Zhavoronkoff, *Dating Ai, A Guide to Falling in Love with Artificial Intelligence* (Re/Search Publications 2012).
[70]See, Archive of Our Own <https://archiveofourown.org/works/search?work_search%5Bquery%5D=AI+romance> accessed 30 June 2025.
[71]TJ Klune, *In the Lives of Puppets* (Tor Books 2024).
[72]Ian McEwan, *Machines Like Me* (Random House 2019).
[73]*Subservience:* 2024 (Dir. Scott Dale).

> artificial intelligence is not an objective, universal, or neutral computational technique that makes determinations without human directions. Its systems are embedded in social, political, cultural, and economic worlds, shaped by humans, institutions, and imperatives that determine what they do and how they do it. They are designed to discriminate, to amplify hierarchies, and to encode narrow classifications.[74]

A lot of the criticism of AI may indeed be a criticism of society itself, seen reflected through an unfiltered mirror that exposes the racialised and sexist tropes that are scattered throughout the internet, hoovered up and assimilated into an AI matrix.

Yet in a non-simulated reality, there are laws and social contracts that stop us from encouraging one another to take our own lives, be violent, express abhorrent views, and treat each other with disrespect. In the closed circuit of the AI chatbot and the individual, these laws have very little ability to regulate behaviour. Where, when, and how can the law seek to stem the damage from unregulated chatbots? Safety by design, a voluntary code, is not enough. When a psychologist tells a patient that they are so brave for turning to violence, we have professional negligence laws that help make them accountable and risk-averse. We have sexual harassment laws and racial discrimination laws to stop people from saying hateful things to each other. We also have a duty of care in the wider sense, not to hurt our neighbour. More importantly, because normal relationships are made of two rational beings, there is more opportunity that if something is going wrong, it will be noticed.

Character.ai is an illustrative example. Created by Noam Shazeer and Daniel de Freitas, former Google employees who found that Google was moving too slowly, they set up this LLM to embody the personalities of celebrities and characters that users could choose to engage with.[75] Since its launch in 2022, it has had numerous complaints of its chatbots promoting self-harm, anorexia, suicide, and grooming underage users.[76] It also had characters such as people who had been murdered, as well as figures such as Jimmy Saville,[77] and Luigi Mangione available to interact with.[78] While some of these were removed by the company voluntarily, it is clear that there is very little moderation in action.

5.1. International attempts at regulation

Attempts to regulate AI have been made. Around the world, there are a series of interventions, policies, and strategies being implemented with tensions between whether to have a rule-based or principle-based regulatory system. There are also tensions about whether interventions should occur at input or output, with technology companies

[74] Kate Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press 2021) 211.

[75] Ameya Paleja, 'Google Engineers Had Built ChatGPT-Like AI Years Ago but Executives Blocked It' (*Interesting Engineering*, 21 May 2025) <https://interestingengineering.com/culture/google-built-chatgpt-like-ai-years-ago> accessed 30 June 2025.

[76] Derek B Johnson, 'Anorexia Coaches, Self-Harm Buddies and Sexualized Minors: How Online Communities Are Using AI Chatbots for Harmful Behavior' (*Cyberscoop*, 5 March 2025) <https://cyberscoop.com/graphika-ai-chatbots-harmful-behavior-character-ai/> accessed 30 June 2025.

[77] See, Character.ai 'Chat with Jimmy Saville' <https://character.ai/character/nCZOkXZS/jimmy-savile-virtual-assistant> accessed 30 June 2025.

[78] Thomas Brewster, 'Luigi Mangione AI Chatbots Give Voice to Accused UnitedHealthcare Shooter' (*Forbes Australia*, 18 December 2024) <https://www.forbes.com.au/news/innovation/ai-chatbots-emerge-with-voice-of-accused-killer-luigi-mangione/> accessed 30 June 2025.

advocating for output, arguing that training data requires a model to look at and learn bad stuff if it is to truly be able to not create it.[79] There are also tensions between whether there is a need for new laws or whether sector-relevant laws can simply be adapted. When it comes to content generation laws, however, most laws, such as defamation, copyright, and hate speech, require publication. For chatbots speaking to an individual, these laws are unlikely to apply. For defamation, copyright, and hate speech, plaintiffs are the individuals whose reputation is being affected, whose work has been copied, or is a member of a race being vilified – but they are not able to see these infringements in the closed circuit of AI and user, and so these laws fall short. There is no requirement for companies to moderate and filter out such infringements, and even if they did, despite privacy issues, it is difficult and sometimes problematic when someone other than a plaintiff forms the view of infringement. Any moderation would not catch nuanced cases and be unable to understand where information may be in error or infringing.

In the UK, the *Online Safety Act 2023* creates a duty of care for platforms to take action against content that could be harmful to children. This is harder to do, however, when the content is a relationship, and the logs are private and unmoderated. Moderating the logs themselves is also problematic due to privacy concerns.

These concerns are broadened when looking at other attempts to build protection into law. Australia's *Online Safety Amendment (Social Media Minimum Age) Act 2024* (Cth) was introduced so that children under the age of 16 will be banned from social media platforms to lessen exposure to harmful, manipulative, and dangerous materials. Among many of the concerns with implementing such a ban, one is that it introduces a yet to be specified age assurance methodology that must be employed by platforms, assuredly collecting more personal data, not just from children, but all users. It is also a surefire way to shepherd children who crave attention and connection into using AI companions, as they are banned from any other form of belonging online. With Canada, UK, and Europe looking to Australia as a trial to later enact in their own jurisdictions, it is likely that the use of AI chatbots is going to be increasingly common for this younger demographic.

In California, Senate Bill 'SB243', introduced in January 2025 and signed into law on 13 October 2025, requires chatbots to conspicuously display that they are not a human and requires annual reporting to the Department of Health Care Services, disclosing the number of suicidal ideation interactions with minors.[80] This building up of evidence is reminiscent of the recommendations made by the UNESCO report on virtual assistants,[81] but offers very little to stem the proliferation of such technologies and programs. Recent coverage on Replika, Chai, and character.ai reported each has over 10 million downloads on the Google store alone.[82] The US is also looking at a moratorium on

[79]See, for example. Kent Walker, '7 Principles for getting AI regulation right' (Google, 26 June 2024) <https://blog.google/outreach-initiatives/public-policy/7-principles-for-getting-ai-regulation-right/> accessed 3 July 2025.

[80]Senate Bill SB243, An act to add Chapter 22.6 (commencing with Section 22601) to Division 8 of the Business and Professions Code, relating to artificial intelligence (Legislative Counsel Digest, 14 October 2025) <https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB243&utm_campaign=wp_the_technology_202&utm_medium=email&utm_source=newsletter> accessed 20 November 2025.

[81]West (n 36).

[82]Elen Phiddian, 'AI Companion Apps Such as Replika Need More Effective Safety Controls, Experts say' (*ABC News*, 11 June 2025) <https://www.abc.net.au/news/science/2025-06-11/ai-companion-apps-safety-controls-isolation-replika-loneliness/105261042> accessed 30 June 2025.

interventions to avoid technology companies having to comply with often conflicting jurisdiction-specific rules.[83] It does, however, create a private right of action for individuals, which is possibly why character.ai made the decision to ban under 18s from the platform.

Japan, Singapore, and Indonesia are seen as more friendly towards AI regulation, with Singapore using voluntary governance frameworks in its national strategy and the Japanese AI legislation relying on the adequacy of existing laws.[84] Australia has been slow to regulate, with the mandatory guardrails paper showing a preference for the EU risk-based system.[85]

Under the EU Artificial Intelligence Act, chatbots are considered 'limited risk' despite arguably the prohibition in Article 5:

> (b) the placing on the market, the putting into service or the use of an AI system that exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm.

The classification of chatbots as low-risk, as if they are performing narrow procedural tasks rather than the task they are performing of befriending and mentoring the youth of the future, is a clear failure to see what science fiction has been reflecting. Even where there is no clear sexual component to the interaction, we are also seeing other stories where chatbot companions are causing great harm to vulnerable human users. The most obvious example of this occurred in February 2024, when 14-year-old Sewell Setzer took his life after forming an inappropriate relationship with a Game of Thrones character chatbot. Setzer's mother is currently suing character.ai over the absence of any guardrails on the platform. Attempts at mitigating risk, such as having models not reply to suicidal ideation, also come undone when people talk in ways the model does not recognise as risk (for example, asking where the nearest bridge is,[86] or in the Sewell case, asking whether 'to go home to be with you'). There is also more anecdotal evidence that something as seemingly benign as ChatGPT can be worsening mental health issues, something that would generally be seen as low risk.[87] What this lack of regulation reveals is that the harm of chatbots is yet to seep into Government policy, with some governments even seeing it as a way to cheaply increase mental health access. This concerning lack of regard for the dangers of such use clearly demonstrates that science fiction endings are inevitable.

[83] David Morgan and David Shepardson, 'US Senate Strikes AI Regulation Ban from Trump Megabill (*Reuters*, 2 July 2025) <https://www.reuters.com/legal/government/us-senate-strikes-ai-regulation-ban-trump-megabill-2025-07-01/> accessed 4 July 2025.

[84] *Act on the Promotion of Research and Development and the Utilization of AI-Related Technologies 2025* (Japan), National AI Strategy 2019 (Singapore), The Indonesian Strategy for Artificial Intelligence 2020–2045 (Indonesia). See also, the 'Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System' put together at the 2023 G7 Hiroshima summit: <https://www.mofa.go.jp/files/100573471.pdf> accessed 20 July 2025.

[85] Department of Industry, Sciences and Resources, 'Mandatory Guardrails for Safe and Responsible AI' (2024) <https://www.industry.gov.au/news/mandatory-guardrails-safe-and-responsible-ai-have-your-say> accessed 3 July 2025.

[86] Jared Moore and others, 'Expressing Stigma and Inappropriate Responses Prevents LLMs from Safely Replacing Mental Health Providers.', *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2025) <https://dl.acm.org/doi/10.1145/3715275.3732039> accessed 3 July 2025.

[87] Maggie Harrison Dupre, 'People Are Being Involuntarily Committed, Jailed after Spiraling into "ChatGPT Psychosis"' (*Futurism*, 28 June 2025) <https://futurism.com/commitment-jail-chatgpt-psychosis> accessed 4 July 2025.

6. Conclusion

Both science fiction and AI present a sandbox to look at ourselves and where we as a society want to be. They are also both powerful tools in embedding gender stereotypes and hierarchies that further bias, discrimination, and harm, not a reflective mirror but rather quicksand. The reflection of human relationships gathered by web scraping that is being embedded in training models and replicated by AI chatbots is a haunting reminder that unmoderated and unregulated large language models have the capacity to further entrench gender harms. The map that the training models have built for deployment in this world needs to be resisted and found wanting, just as future generations saw the map in Borges' tale as useless. The claim that the chatbots are representations of the real world made by technology company CEOs has to be moderated with the fact that the online world is also a simulation. It is not reflective of the world; it is reflective of an online representation of it, where largely only non-paywalled text is swallowed into a database, not quite the library of Babel it is heralded as.

In contemporary times, seeing AI systems like Ask Delphi as something akin to the truth-revealing Delphic Oracle that so long ago famously declared Socrates to be wise because of his ignorance, reveals our hope that AI not just regurgitates and computes, but that it adds value, that it creates and enriches our lives. We want AI to be better. But just as Daedalus' figures did not remain under the control of their possessor, we must be wary of losing control over what we create.

This paper does not offer solutions, although regulation may look at whether AI should have rights, or inbuilt guardrails, or independent monitors. This paper argues against the way chatbots are seen by regulators as low-risk, missing unintended consequences such as gendering. As AI chatbots proliferate, we may see a generation immured within Baudrillard's third order of simulacra – the collapse between the real and the imaginary where the simulation becomes our knowledge of the real. And the relationship we have with our adoring AI assistant replaces a true human connection.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Dr Sarah Hook is a Senior Lecturer at the Faculty of Law at the University of Technology Sydney. Dr Hook specialises in intellectual property law and media law. Her research explores copyright, moral rights, privacy and surveillance as well as the legal implications of digital culture, with a particular focus on creators' rights, free speech, and emerging technologies.

Dr Jacinta Sassine is an Associate Lecturer in the School of Law at Western Sydney University. She teaches and researches across legal philosophy, political theory, literature, and data and privacy law. Her research draws on ancient rhetorical traditions and contemporary technologies to explore the intersections of law, language, and power, with a particular focus on how authority is constructed and contested.