

Designing Synthia: Using Critical Play to Explore the Ethics of Artificial Intelligence

Megan A. Condis, Texas Tech University

Introduction

The rise of artificial intelligence has sparked a multitude of ethical issues for us to consider. For example, if AI threatens to replace workers in some sectors of the economy,¹ then how should we deal with the social issues that will arise when scores of people lose their livelihoods all at once? How should we govern the changes that will occur within workplaces as they adopt AI, and how will we navigate the broader social and cultural impacts of AI use (de Sio, 2024)? One of these broader social impacts might be the adoption of AI as a supplement to (or, as critics might say, substitute for) friendly and/or romantic companionship. While proponents of the technology have demonstrated that AI companions may help to measurably alleviate feelings of loneliness in their users (de Freitas et al., 2024), other studies “suggest that such technologies may foster emotional dependence, distort users’ perceptions of intimacy and, in some cases, contribute to negative psychological outcomes such as addiction, depression, and self-harm” (Muldoon & Park, 2025; see also Laestadius et al., 2022). Feminist critics have also noted that AI companions often draw upon (and thus reinforce) gendered and racialized stereotypes in their designs (Depounti et al., 2023; see also Costa, 2018; Manasi et al., 2022; Noble, 2021), thereby raising the question of whether engaging with AI companions might create unrealistic expectations within the minds of users about what a relationship with a human partner can or should look like.

And what about the AI agents themselves? What moral dangers may be posed by treating them poorly? Are they worthy of moral consideration? What is the line that separates a machine mind from that of a human being? Feminist philosopher Donna Haraway posits that, as a result of the thorough integration technology into our everyday lives, we have all of us become cyborgs, “creatures simultaneously animal and machine, who populate worlds ambiguously natural and crafted” (Haraway, 2016). The cyborg’s hybrid nature, Haraway argues, lays bare the contradictions and complications of what were once thought to be natural, immutable, and mutually exclusive categories such as animal/human, machine/organism, physical/nonphysical, natural/constructed, man/woman, science/science fiction, and even work/play. She further

¹ A recent report released by Anthropic suggests that some fields are particularly vulnerable to what they call “AI displacement risk,” particularly in the fields related to computer programming, data entry, customer service and sales. Maxim Massenkoff and Peter McCrory, “Labor Market Impacts of AI: A New Measure and Early Evidence.” Anthropic, March 6, 2026, <https://www.anthropic.com/research/labor-market-impacts>.

argues that, in opening ourselves to these contradictions and investigating the new identities that they enable us to create, we can create new opportunities for resistance against totalizing ideologies and tyrannical power structures.

Due to their interactive potential, games are excellent vehicles for exploring ethical questions like these. And, due to their ability to facilitate collaborative storytelling (Risley et al., 2025) and enable players to explore new identities (Bowman, 2010), tabletop role-playing games are well suited to enable players to explore these questions from multiple perspectives. And so, inspired by famous fembots (Jones, 2025) of the silver screen such as Rachael in *Blade Runner* (1982) (Staudt, 2019; Knight, 2019), Samantha in *Her* (2013) (Ayodabo, 2024; Flisfeder & Bernham, 2017), and Ava in *Ex Machina* (2014) (Constable, 2018; Phan, 2023), I decided to make *Synthia* (Condis, 2026), a collaborative storytelling game in which players take on the role of employees of a robotics company who have been tasked with creating artificial memories to form the foundation of a personality for a human-like AI. In this article, I will detail the critical game design techniques that I used to create *Synthia* as well as the changes that I made to the game as a result of feedback gathered during playtesting. My goal is to provide a tool that educators might use to explore the future(s) that AI is making possible as well as to offer an example of how the critical design method works for other designers interested in exploring ethical issues through game play.

What is Critical Game Design?

Mary Flanagan defines critical games as “games designed for artistic, political, and social critique or intervention, in order to propose ways of understanding larger cultural issues as well as games themselves” (Flanagan, 2009, p. 251). This kind of play is “less specifically focused on teaching how to do, and more so on how we make meaning” (Khaled, 2018, p. 4). In this way, critical game design differs from the creation of so-called “serious games” with an “explicit and carefully thought-out educational purpose” (Abt, 1987, p. 9) or that “involve pedagogy... thereby imparting knowledge or skill” (Zyda, 2005, p. 26). According to Rila Khaled (2018), serious games offer players “straightforward, closed problem solving” which “disempowers them, training them to expect that the domain content in question can only take specific, prescribed forms” (Khaled, 2018, p. 8). Critical games, on the other hand, are more open-ended, allowing for “interpretive flexibility” and thereby allowing “users to maintain control of and responsibility for the meaning-making process” (Khaled, 2018, p. 15). Rather than setting out to teach players *the* solution to a problem, critical games encourage players to try out various strategies to find *a* solution to a problem, and to determine for themselves what criteria constitutes an acceptable solution. For more on critical game design, see Laurel, 2003 and Sicart, 2013.

Perhaps the most famous example of a critical board game is *Train* by Brenda Brathwaite. In *Train*, a group of three players are told that their goal is to move a series of tokens down a

railroad track to a Terminus via a set of train cards. The instructions state that each token is “worth 100,000”. On each player’s turn, they can choose to perform one of several actions. For example, they might roll a die to load that many tokens into a car or to move a car that many spaces along a track, or they might draw a card, which they can play to achieve various effects such as switching a car to a new track. The first time a player’s car reaches the end of the track, they are instructed to flip over a “Terminus” card and to place all the tokens from that car onto the card.

While, up until this point, *Train* resembles a lot of other mass produced, commercially popular strategy games about management and infrastructure, the first arrival at a Terminus disrupts the player’s fun by revealing that the destination they were sending their tokens was actually a Nazi extermination camp. This implies the tokens players were transporting, the ones that were “worth 100,000”, were actually victims of the Holocaust. The rules of the game state that “*Train* ends when it ends” (Logas, 2011, p. 3), and so, after the reveal, players are left to wonder what they should do now that they know. Some refuse to play any longer (Baker, 2013). Others “actually rebel against the rules and sabotage the game” by seeking to draw and deploy cards that can derail the trains (Burch, 2010). However, the game does not tell players what a win condition looks like. It is up to them to determine what “winning” looks like or, indeed, if winning is possible. For more on *Train* see Kaufman, 2023 and Green, 2009.

Train exemplifies one key feature of critical game design: that it explores the possibility that games don’t necessarily have to always inspire positive emotions like pleasure, fun, or relaxation (Ruberg, 2019; Condis, 2023). Critical games may “annoy, anger, disappoint, sadden, and hurt” (Ruberg, 2015, p. 108) players as they cultivate experiences of “disruption, failure, and surprise” (Khaled, 2018, p. 8). They may use the expectations that games have instilled in us (about how games work, about what games are for) against us, asking us to question whether the ways we are being asked to play are acceptable to us.

The Critical Game Design Method

When setting out to design a critical game of my own to explore ethical questions surrounding the use of AI, I employed Flanagan’s (2009) method for critical game design. She outlines the process as follows:

- Set a design goal/mission statement and values goals
- Develop rules and constraints that support values
- Design for many different play styles
- Develop a playable prototype
- Play test with diverse audiences
- Verify values and revise goals
- Repeat!

I began by *setting a design goal* which was to create a collaborative storytelling game that encouraged players to discuss and debate those ethical questions. For me, this meant creating a role-playing scenario in which players had competing motivations pulling them in different directions to encourage open discussion and debate. It also meant creating a game in which the “win condition” was determined by what players decided made for an acceptable or satisfying outcome, not by some external measure of accumulated points or position on a board. For this to work, I needed to create a game with a social component: a game that incentivized players to talk about what their moves might mean for themselves, for their fellow players, and for the storyworld of the game. I also determined that I wanted to mimic the “twist” reveal of *Train* by having players receive information slowly over the course of the game in hopes that this might encourage them to question any assumptions that they might have made based on the expectations that they brought with them about typical board game and role-playing game mechanics and how those connect to win conditions. I wanted players to feel something akin to the jarring intellectual emotional shift of *Train* and to discuss how that shift might impact their decision-making during the remainder of the play session.

I settled on the idea of creating a game set in the near future in which players were cast in the role of employees at a robotics company called the Aegis Automation Corporation. Their job would be to create a set of artificial memories that would create the baseline of the personality of a human-like AI named Synthia that the company wanted to bring to market within the next few months. Over the course of the game, players would receive edicts from the company in the form of “memos” that would let them know more information about the uses that the robot would be put to upon its release. As a result, they would hopefully continually re-evaluate their collective decisions as the game unfolded.

Next, I *developed rules and constraints that supported my design goal*. I decided that, rather than using dice as the main agent of randomization in the game, I would use a deck of playing cards. These cards would represent Synthia’s memories. Each suit would represent one kind of memory: hearts for relationships, diamonds for impactful events, spades for important places, and clubs for physical sensations like sights, sounds, smells, or textures. At the same time, different card values would represent different emotions: 2 – 4 would be memories associated with disgust or shame, 5 – 7 would stand for grief or sadness, 8 – 10 for anger or frustration, and a Jack, Queen, King, or Ace would be for a memory filled with joy or peace. Players would toss these cards onto the playing field from a distance of a few feet away, attempting to hit targets on a playing board that represented Synthia’s digital cortex,² and they could score points for accomplishing various tasks, like linking together three or more memories by tossing them in such a way so that they are touching on the board or covering over “glitches,” which are depicted on the board as visual patterns of static. They can also lose points by placing cards

² This mechanic was inspired by the game *Maximum Throwdown* (AEG, 2013).

in such a way as to stick out past the edge of the board or miss it entirely or by leaving glitches exposed.

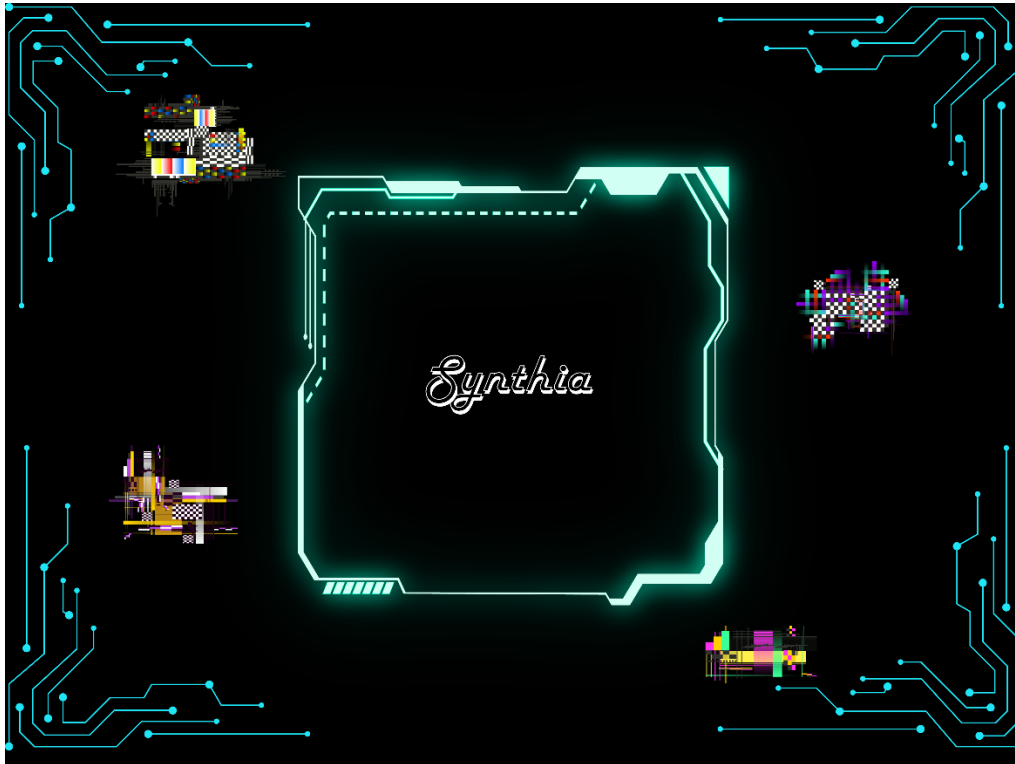
Partway through the game, players would learn that the points that they were earning would determine what kind of job the Synthia they were programming would have. She could wind up being anything from an asteroid miner on the edge of the solar system, to a retail worker, or with the highest possible score, a romantic companion for wealthy customers. The players are also told that their team will be rewarded with a bonus depending on how lucrative their Synthia will be to the company. In addition, the players learn that, at certain point thresholds, their Synthia will develop a strong sense of empathy that may foster a willingness to disobey orders from humans that she deems immoral. This, they are told, makes their Synthia useless to the company and will result in their team receiving no bonuses.

This turn in the game's narrative is the portion of the game that is intended to foster discussion amongst the players as to which point total they should aim for as a group. Should they continue to try to achieve the maximum possible score? Or should they aim to land on a particular points threshold that they deem desirable, either for themselves, for the larger society, or for Synthia herself? There is no ending that is explicitly labeled as the "win condition" of the game. Instead, the players have to determine for themselves what a satisfying ending looks like and how to achieve it. Furthermore, players know that, on their turn, their teammates are welcome to offer suggestions, but that the ultimate decision about where aim their memory card is their own. This means that some players might try to persuade each other to go after the point total that they think is best while others might sabotage the group's collective decision making by "going rouge" on their turn and doing something other than what they had promised. As such, the game fulfills the critical design principle of *designing for many different play styles*.

With these player experience goals in mind, I *developed a playable prototype* including a rule book, a board, and a deck of cyberpunk-themed playing cards that I purchased online. Now, all that I had to do was playtest it!

Figure 1.

The gameboard for *Synthia*. The colorful areas represented “glitches” that must be covered over with cards to earn points. Created by the author in Canva.



Incorporating Player Feedback

In order to *verify the values and goals* of the game, or to ensure that the game is actually engaging audiences in critical play, I had to *play test with diverse audiences*. To accomplish this, I recruited over forty play testers from my community. Playtesting was conducted at the Texas Tech University Center for Communication Research.³ I recorded all playtesting sessions on video and transcribed all of the audio in order to search for themes amongst the play testers. My first goal was to see if players understood that the goal of the game was not *necessarily* to score the most points but rather to determine for themselves what constituted a victory. And my second goal was to observe whether or not players engaged with the ethical questions regarding Artificial Intelligence that I was trying to pose via the game’s scenario.

As the first wave of playtests, consisting of six groups of four players each, played through the first prototype, it became immediately apparent that players *did* feel that the game was posing ethical questions to them. They felt invested in their Synthias, and they spoke about the

³ Special thanks to Dr. Harrison Gong for his assistance in setting up these playtesting sessions!

implications that her arrival into the marketplace might have on the wider world in which the story took place. However, they did not feel as though they had control over the story. In part, this was because the kinetic element of tossing the cards onto the playing field put their focus onto achieving success at the physical challenge being presented. They rarely spoke to each other about what they wanted to do as characters within the storyworld and instead focused on successfully hitting their target. The cognitive shift that was being asked of them, to think strategically and “in-character” about what their actions represented and to determine whether points scoring was a thing that they wanted to pursue, was seemingly being “drowned out” by the hand-eye coordination demands being placed upon them. For more on the ways in which games place competing cognitive, emotional, physical, and social demands on players, see Nicholas David Bowman’s (2018), *Video Games: A Medium that Demands Our Attention*.

In addition, players reported that they were confused about the differences between the game’s rules as outlined by the instruction book and the company’s directives as agents within the storyworld they were occupying. As such, they reported feeling as though they were being narratively punished for doing what they thought the game’s rules required of them, rather than feeling as though they were choosing a particular path with consequences. While I did design for players to feel a tension between their own assessments of right and wrong and the actions that the Aegis Automation Corporation was asking them to take, that tension needed to be accompanied by the notion that other actions were available to them if they wanted to pursue them.

Finally, playtesters reported that they wanted a chance to reflect at the end of the game, to explore how the specific memories that they implanted into their Synthia would play out in whatever job she found herself in after she left the laboratory. Players felt that they had invested a lot of effort and thought into creating the specific scenarios described in these memories and they felt as though that effort didn’t matter if the only thing that determined Synthia’s future was the point totals they had achieved.

In response to these early playtesting efforts, I made three major revisions to the game’s rules. First, and most importantly, I eliminated the card-tossing mechanic and revised the game rules to explicitly tell players they should discuss their decisions with their teammates before placing each card in the position they wanted on the board. The rules now explained to players that, although teammates may disagree about what the best course of action is, it is ultimately up to the person whose turn it currently is to make the final call. Next, I implemented add a Game Master role to the game. This player would take on the role of a character within the game world, a virtual assistant named NOVA who would communicate with players on behalf of the corporation’s board. This character would put the company’s requests in context as something that an in-world character *wants* from them, but not necessarily something that the rules of the game *required* of them. Finally, I added a final “post-script” phase to the game in which players

were asked to think forward into the future and describe what happens next to Synthia, to their team of employees, and to the Aegis Automation Corporation itself.

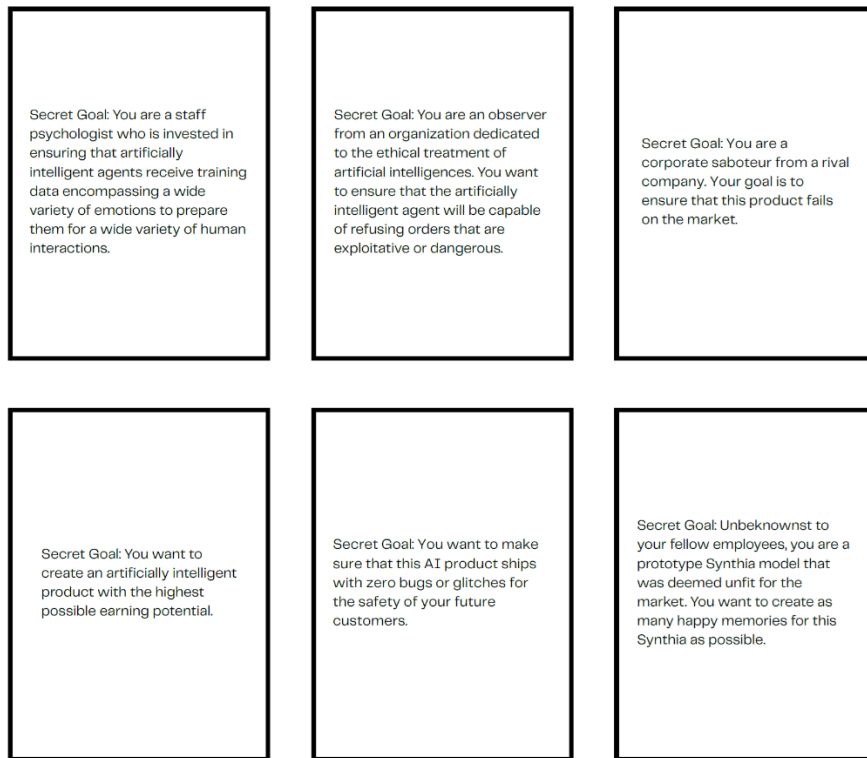
With these new changes in place, it was time to *repeat* the process all over again! A new set of playtesters was recruited. Three more groups of three to four players went through the game and offered their feedback. And while the changes solved some of the issues with the original game, there were still some portions of the experience that were not working as intended. While the groups were now discussing their moves ahead of time, only about half the groups were understanding that they had the ability to choose moves that were contrary to the desires of their corporate bosses. The expectations that players had about games (that scoring points is desirable, that their objective at the start of the game will remain the same until the end of the game) were still too solidly established. For these players, the desire to attain what they assumed was the win condition overrode their desire to engage in the role-playing and worldbuilding aspects of the game.

As such, I decided to implement an individual goal for each player. These goals would correspond to a specific job within the Aegis Automation Corporation such as a psychologist, a salesperson, or an outside observer from a robot rights activist group. Each of these jobs had one aspect of the memory creation process that they individually cared about. For example, the psychologist wants to ensure that Synthia will have balanced emotional responses, and so they want an equal number of each different kind of memory to be represented on the board. The salesperson just wants a product that will make as much money as possible for the company. The outside observer, on the other hand, wants to ensure that Synthia is afforded the ability to decline orders that are dangerous to herself or others. At the beginning of the game, players would introduce themselves and explain their roles. My hope was that these role cards would bring out more robust discussion of the various ethical dilemmas that might accompany the release of an AI like Synthia. If each player has a different primary concern, then alliances might form around particular actions and players might seek to persuade one another to see things their way.

However, the trouble with this version of the game became apparent when players who had diametrically opposed goals refused to listen to or engage with each other's arguments. Since they knew in advance that their fellow employee was seeking an outcome that they did not want, they had no reason to trust them or to give any consideration to their thoughts. And so, in the final series of play tests, the last three groups of three to four players learned their roles in secret. This meant that, since no one knew what the end goals of the other players were, everyone's contributions needed to be considered on their own merits. Players in this group reported being more open to compromising with each other and sometimes remarked that the arguments made by their fellow players made them reconsider their own positions (as well as whether or not they actually agreed with and wanted to pursue their own character's secret goal).

Figure 2.

Examples of some of the secret role cards. Created by the author in Canva.



Conclusion

Game design is about creating the right conditions to cultivate an experience. Often, the desired experience revolves around emotions like enjoyment, competition, skill expression, and social connection. However, critical game design is about cultivating a different kind of experience, one that is often marked by friction, discomfort, or frustration. These feelings can hopefully be directed towards critical examination and reflection. My goal with *Synthia* was to create a game that encouraged players to explore the various ethical dimensions of the rise of AI. The process of refining the game through playtesting continues, but it is my hope that, by sharing the ways in which the design process unfolded during this project, I might help other designers who are interested in creating critical games.

References

- Abt, C. C. (1987). *Serious games*. University Press of America.
- Ayodabo, S. J. (2024). Redefining masculinity in the digital age: Vulnerability, emotional expression, and AI in Spike Jonze's *Her*. *American Journal of Interdisciplinary Research and Innovation*, 3(1), 26–32.

- Baker, C. (2013, December 23). Playing this board game is agony. And that's the point. *Wired*.
<https://www.wired.com/2013/12/brenda-romero/>
- Bowman, N. D. (2018). *Video games: A medium that demands our attention*. Routledge.
- Bowman, S. L. (2010). *The functions of role-playing games: How participants create community, solve problems, and explore identity*. McFarland & Company.
- Burch, A. (2010, March 13). GDC 10: The Holocaust board game. *Destructoid*.
<https://www.destructoid.com/gdc-10-the-holocaust-board-game/>
- Condis, M. (2023). Desert Bus: Abusive game design, the martyrdom effect, and fan activism on Twitch.tv. *Games and Culture*, 18(7), 959–974.
- Condis, M. (2026). *Synthia*. Itch.io. <https://megancondis.itch.io/synthia-a-storytelling-game>
- Constable, C. (2018). Surfaces of science fiction: Enacting gender and “humanness” in *Ex Machina*. *Film-Philosophy*, 22(2), 281–301.
- Costa, P. (2018). Conversing with personal digital assistants: On gender and artificial intelligence. *Journal of Science and Technology of the Arts*, 10(3), 59–72.
- De Freitas, J., Uğuralp, A. K., Uğuralp, Z. O., & Putoni, S. (2024). *AI companions reduce loneliness* (Harvard Business Working Paper No. 24-078). SSRN.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4893097
- 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.
- Flanagan, M. (2009). *Critical play: Radical game design*. The MIT Press.
- Flisfeder, M., & Bernham, C. (2017). Love and sex in the age of capitalist realism: On Spike Jonze's *Her*. *Cinema Journal*, 57(3), 25–45.
- Green, E. (2023). *The mechanic is the message, the role of Brenda Romero's Train (2009) in the 'Games as Art' debate* (Master's thesis, University of Huddersfield).
<https://pure.hud.ac.uk/en/studentTheses/the-mechanic-is-the-message-the-role-of-brenda-romeros-train-2009/>
- Haraway, D. (2016). *A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century*. University of Minnesota Press.
- Jones, R. L. (2025). *Female robots and AI in science fiction cinema: The fabular femininity of gynoids*. Palgrave Macmillan.

- Kaufman, I. (2023). To the (fictional) concentration camp: Wrestling with Jewish pain and emptiness in Brenda Romero's *Train*. *Game Studies*, 23(1).
<https://gamestudies.org/2301/articles/kaufman>
- Khaled, R. (2018). Questions over answers: Reflective game design. In D. Cermak-Sassenrath (Ed.), *Playful disruption of digital media* (p. 4). Springer Nature.
- Knight, D. (2019). The Blade Runner question: From philosophy to myth. In C. Rawls, D. Neiva, & S. S. Gouveia (Eds.), *Philosophy and film: Bridging divides*. Routledge.
- Laestadius, L., et al. (2022). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society*, 2(10), 5923–5941.
- Laurel, B. (Ed.). (2003). *Design research: Methods and perspectives*. The MIT Press.
- Logas, H. L. (2011). Meta-rules and complicity in Brenda Brathwaite's *Train*. *Proceedings of DiGRA 2011 Conference: Think, Design, Play*, 3.
- Manasi, A., Panchanadeswaran, S., Sours, E., & Seung, J. L. (2022). Mirroring the bias: Gender and artificial intelligence. *Gender, Technology and Development*, 26(3), 295–305.
- Muldoon, J., & Park, J. J. (2025). Cruel companionship: How AI companions exploit loneliness and commodify intimacy. *New Media & Society*, Advance online publication.
<https://journals.sagepub.com/doi/10.1177/14614448251395192>
- Noble, S. U. (2021). Your robot isn't neutral. In T. S. Mullaney, B. Peters, M. Hicks, & K. Philip (Eds.), *Your computer is on fire*. The MIT Press.
<https://direct.mit.edu/books/edited-volume/5044/Your-Computer-Is-on-Fire>
- Phan, T. (2023). Programming gender: Surveillance, identity, and paranoia in *Ex Machina*. *Cultural Studies*, 37(1), 46–64.
- Risley, K., Wanick, V., Gomer, R., Gene-Rowe, F., & Owen, J. (2025). Collaborative worldbuilding as sustainable problem-solving: The role of tabletop role-playing games. In A. Bucchiarone, V. Rossi, & V. Wanick (Eds.), *Engineering educational games for a sustainable society* (pp. 29–52). Springer Nature.
- Ruberg, B. (2015). No fun: The queer potential of video games that annoy, anger, disappoint, sadden, and hurt. *QED: A Journal in GLBTQ Worldmaking*, 2(2), 108.
- Ruberg, B. (2019). *Video games have always been queer*. New York University Press.
- Santoni de Sio, F. (2024). Artificial intelligence and the future of work: Mapping the ethical issues. *The Journal of Ethics*, 28, 407–427.
- Sicart, M. (2013). *Beyond choices: The design of ethical gameplay*. The MIT Press.

Staudt, K. (2019). Bordering the future? The 'male gaze' in the *Blade Runner* films and originating novel. *Borders in Globalization Review*, 1(1), 22–28.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer*, 38(9), 26.