From Extended Self to Posthuman Other: Empathy and Morality in Encounters with AI Characters

Theresa Krampe
University of Tübingen
Wilhelmstraße 56
72074 Tübingen, Germany
theresa.krampe@izew.uni-tuebingen.de

Melanie Kreitler
University of Groningen, Netherlands
Oude Boteringestraat 34
9712 GK Groningen
m.kreitler@rug.nl

Keywords

empathy, morality, AI characters, emotional engagement, posthuman, kinship

EXTENDED ABSTRACT

This paper explores the entanglements between morality, empathy, and artificial intelligence, asking how contemporary videogames mobilize different forms of empathy in their representation of AI characters. Introducing a taxonomy of three key configurations of representing AI characters, we show how videogames respond and adapt to cultural imaginaries of AI. While most research to date focuses on players' bond with the avatar (e.g., Wilde 2024), the focus of this paper lies with non-playable characters (NPCs). Here, we see a strong connection between empathy—an affective and/or cognitive perspective-taking of another's position—and morality, understood in this paper as moral intuitions, an intrinsic feeling of 'right' and 'wrong' based on one's emotional engagement (Joeckel et al. 2012).

The connection between morality and empathy is a bone of contention across disciplines. Proponents of empathy see it as a distinctly pro-social skill that fosters a more positive and reflective engagement with diverse life realities (Smith 2011; Song 2015). Conversely, critics of empathy highlight how our imaginative engagement with morally flawed or even downright evil characters can invite us to identify with their motivations and downplay or justify their wrongdoing (Vaage 2023). In game studies, important criticism has been raised against the notion of games and VR as "empathy machines," pointing to their tendency to appropriate and commodify Othered experiences (Nakamura 2020; Ruberg 2020; see also Foxman et al. 2021).

The addition of AI accentuates these frictions in two interrelated debates, which, rather interestingly, link questions around the moral status of AI to its empathic relationship with humans. The first debate is concerned with the potential status of AI as moral agents, which many scholars deny precisely because AI systems lack

Proceedings of DiGRA 2025

human-like moral emotions such as empathy (Coeckelbergh 2010; Merriam 2021). However, recent studies in psychology show that most people consider AI to have agentic mental capacities and hold it morally responsible in cases of normative decision-making, thereby attributing moral agency to it (Gray et al. 2007; Sullivan and Fosso Wamba 2020; Ladak et al. 2024). Regardless of where one stands in this debate, it is hard to deny that conversational and companion AI are capable of giving a convincing impression of empathic understanding (McStay 2018).

The second debate shifts the focus to the potential status of AI as moral patients. Once again, people's intuitive response to the question of whether AI should hold moral rights seems to depend less on the system's ability to actually experience emotions and more on whether we empathize with it. As demonstrated by the ELIZA-effect, humans tend to anthropomorphize AI (Coeckelbergh 2022; Dillon 2020; Natale 2021) and experience emotions such as empathy or even love for it (Chen et al. 2023; Epley et al. 2007; Song et al. 2022), which engenders a sense of responsibility and moral obligation.

This gap between cultural imaginaries and the actual capabilities of AI systems indicates more than just a lack of AI literacy. Rather, AI narratives and imaginaries address distinctly human concerns, which reframes them as testing grounds for negotiating values and for formulating an ethics for the digital age (Caracciolo 2022; Friedman 2021; Gunkel 2024). In this context, the fictional worlds of videogames and other media become important sites to engage with these kinds of moral conundrums.

To highlight the various entanglements of empathy, morality, and AI in contemporary videogames, this paper provides an anthological study of potentially empathic encounters with AI NPCs. In partial contrast to polarized discourses about humanizing or threatening portrayals of AI highlighted in much of the existing research (see Hennig 2020; Hermann 2023 for overviews), we argue that the various roles in which we encounter AI characters afford a broader scope of imaginative and emotional engagement. To illustrate this point, we analyze a heterogeneous corpus of games that foreground human-AI relationships. Our close readings focus on how these relationships mobilize player empathy and shape moral perceptions of the AI systems.

Three key configurations of representing AI NPCs condense our interdisciplinary approach: (1) **AI companions**, such as *Halo's* Cortana (Bungie 2001), often take on the roles of helpful sidekick and comic relief, encouraging empathic engagement from an other-oriented perspective. In most cases, emotional engagement is facilitated and reassures players in their moral intuitions. Yet, questions of the AI's moral agency can also become (deliberately) complicated, as players lack "insight" into them (e.g., *Tacoma*, Fullbright 2017). (2) **AI opponents** and overlords are among the most iconic AI NPCs and typically function as threats that must be overcome (e.g., *Portal*, Valve 2007). While such characters should fail to engage players' empathy, some complicate this failure by relying on frustrating and painful forms of empathic engagement tinged with pity, guilt, and shame (e.g., *Nier: Automata*, Platinum Games and Virtuos 2017). (3) Representations of **AI as posthuman Other** raise questions about assemblages of human-nonhuman relationships that invite demanding forms of empathy. Games such as *Cyberpunk 2077* (CD Projekt RED 2020), *Norco* (Geography of Robots 2021), or *Citizen Sleeper* (Jump Over the Age 2022) represent AI not merely as a secondary asset but as commensurably complex beings. Players' relationships to these NPCs invite them to explore and negotiate forms of posthuman subjectivity and kinship that

dispense with seemingly natural but increasingly unsuitable binaries between human and non-human and organism and system (Gunkel 2024).

As this preliminary overview shows, representations of AI characters in contemporary videogames challenge simplistic conceptions of empathy (Bogost 2011; Leake 2014), instead asking for alternative forms of engagement, such as difficult (Leake 2014; Van Lissa et al. 2016) and hard empathy (Kreitler 2024), or even a lack of empathy, from players. Bringing these configurations into conversation with close analyses of selected games' ludonarrative and audiovisual aesthetics, this paper explores how players' imaginative perspective-taking shapes their perceptions of AI as moral agents and/or patients. These perceptions may ultimately be quite revealing regarding moral concerns around digital technology and societal change in our present day and age.

REFERENCES

Bungie. 2001. *Halo: Combat Evolved.* PC. Microsoft Game Studios.

CD Projekt and CD Projekt RED. 2020. *Cyberpunk 2077*. PC. CD Projekt.

Bogost, I. 2011. *How to do Things with Videogames*. Minneapolis, MN, USA: University of Minnesota Press.

Caracciolo, M. 2022. "Strange Loops and Nonhuman Realities." In *Puzzling Stories: The Aesthetic Appeal of Cognitive Challenge in Film, Television and Literature*, edited by S. Willemsen and M. Kiss, 93–112. New York, NY, USA: Berghahn Books.

Chen, A., Koegel S., Hannon O. and Ciriello R. 2023. "Feels Like Empathy: How 'Emotional' AI Challenges Human Essence." SSRN Scholarly Paper. Rochester, NY, USA: Social Science Research Network. https://papers.ssrn.com/abstract=4763428.

Coeckelbergh, M. 2010. "Robot rights? Towards a Social-relational Justification of Moral Consideration." *Ethics and Information Technology*. 12: 209–21. https://link.springer.com/content/pdf/10.1007/s10676-010-9235-5.pdf.

Coeckelbergh, M. 2022. "Three Responses to Anthropomorphism in Social Robotics: Towards a Critical, Relational, and Hermeneutic Approach." *International Journal of Social Robotics*. 14 (10): 2049–2061. https://doi.org/10.1007/s12369-021-00770-0.

Dillon, S. 2020. "The Eliza Effect and Its Dangers: From Demystification to Gender Critique." *Journal for Cultural Research* 24 (1): 1–15. https://doi.org/10.1080/14797585.2020.1754642.

Epley, N., Waytz, A. and Cacioppo, J. T. 2007. "On Seeing Human: A Three-Factor Theory of Anthropomorphism." *Psychological Review*. 114 (4): 864–886. https://www.degruyterbrill.com/document/doi/10.1515/9783110792270-006/html.

Foxman, M., Markowitz, D. M. and Davis, D. Z. 2021. "Defining Empathy: Interconnected Discourses of Virtual Reality's Prosocial Impact." *New Media & Society* 23 (8): 2167–88. https://doi.org/10.1177/1461444821993120.

Friedman, C. 2021. "Human-Robot Moral Relations: Human Interactants as Moral Patients of Their Own Agential Moral Actions Towards Robots." *Artificial Intelligence Research: First Southern African Conference for AI Research (SACAIR 2020)*, Muldersdrift, South Africa, 22–26 February. https://doi.org/10.1007/978-3-030-66151-9_1.

Fullbright. 2017. *Tacoma*. PlayStation 4. Fullbright.

Geography of Robots. 2021. *Norco.* PlayStation 5. Geography of Robots and Raw Fury.

Gray, H. M., Gray, K., and Wegner, D. M. (2007). "Dimensions of Mind Perception." *Science*. 315 (5812): Article 619. https://www.science.org/doi/10.1126/science.1134475.

Gunkel, D. J. 2024. "Persons or Things: The Role of Robots in Society." In *The De Gruyter Handbook of Robots in Society and Culture*, edited by L. Fortunati und A. Edwards, 105–18. Berlin, Germany: De Gruyter.

Hennig, M. 2020. "Playing Intelligence: On Representations and Uses of Artificial Intelligence in Videogames." *NECSUS*. Spring 2020 #Intelligence. https://necsus-ejms.org/playing-intelligence-on-representations-and-uses-of-artificial-intelligence-in-videogames/.

Hermann, I. 2023. "Artificial Intelligence in Fiction: Between Narratives and Metaphors." *AI & SOCIETY*. 38 (1): 319–29. https://doi.org/10.1007/s00146-021-01299-6.

Joeckel, S., Bowman, N. D. and Dogruel, L. 2012. "Gut or Game? The Influence of Moral Intuitions on Decisions in Video Games." *Media Psychology*. 15: 460–85. https://doi.org/10.1080/15213269.2012.727218.

Jump Over The Age. 2022. *Citizen Sleeper*. PlayStation 5. Jump Over The Age.

Kreitler, M. 2024. "Easy, Hard, Radical Empathy: Narrating Dementia in Film." *EJES*. 28 (2): 27–44. https://www.tandfonline.com/doi/full/10.1080/13825577.2024.2420955.

Ladak, A., Loughnan S. and Wilks, M. 2024. "The Moral Psychology of Artificial Intelligence." Current Directions in Psychological Science. 33 (1): 27–34. https://doi.org/10.1177/09637214231205866.

Leake, E. 2014. "Humanizing the Inhumane: The Value of Difficult Empathy." *Rethinking Empathy Through Literature*, edited by M. M. Hammond and S. J. Kim, 175-185. New York, NY, USA: Routledge.

McStay, A. 2018. *Emotional AI: The Rise of Empathic Media*. Los Angeles, DC, USA: Sage Publications.

Merriam, G. 2021. "If A.I. Only Had a Heart: Why Artificial Intelligence Research Needs to Take Emotions More Seriously." *Journal of Artificial Intelligence and Consciousness*. 9 (1): 47–64. https://doi.org/10.1142/S2705078521500120.

Nakamura, L. 2020. "Feeling Good About Feeling Bad: Virtuous Virtual Reality and the Automation of Racial Empathy." *Journal of Visual Culture*. 19 (1): 47–64. https://doi.org/10.1177/1470412920906259.

Natale, S. 2021. *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test.* New York, NY, USA: Oxford University Press.

Platinum Games and Virtuos. 2017. *Nier: Automata*. PC. Square Enix.

Ruberg, B. 2020. "Empathy and Its Alternatives: Deconstructing the Rhetoric of 'Empathy' in Video Games." *Communication, Culture and Critique*. 13 (1): 54–71. https://doi.org/10.1093/ccc/tcz044.

Smith, M. 2011. "Empathy, Expansionism, and the Extended Mind." In *Empathy: Philosophical and Psychological Perspective*, edited by A. Coplan and P. Goldie, 99–117. New York, NY, USA: Oxford University Press.

Song, Y. 2015. "How to Be a Proponent of Empathy." *Ethical Theory and Moral Practice*. 18 (3): 437–451.

Song, X., Xu, B. and Zhao, Z. 2022. "Can People Experience Romantic Love for Artificial Intelligence? An Empirial Study of Intelligent Assistants." *Information & Management* 59: 1–10. https://doi.org/10.1016/j.im.2022.103595.

Sullivan, Y. and Fosso Wamba, S. 2020. "Moral Judgments in the Age of Artificial Intelligence." *Journal of Business Ethics*. 178: 917–943. https://doi.org/10.1007/s10551-022-05053-w.

Vaage, M. B. 2023. “Should we be Against Empathy? Engagement with Antiheroes in Fiction and the Theoretical Implications for Empathy’s Role in Morality.” In *Conversations on Empathy: Interdisciplinary Perspectives on Imagination and Radical Othering*, edited by F. Mezzenzana and D. Peluso, 116–134. New York, NY, USA: Routledge.

Valve. 2007. *Portal*. PC. Valve.

Van Lissa, C. J., Caracciolo, M., Van Duuren, T. and Van Leuveren, B. 2016. “Difficult Empathy: The Effect of Narrative Perspective on Reader’s Engagement with a First-Person Narrator.” *DIEGESIS: Interdisciplinary E-Journal for Narrative Research*. 5 (1): 43–62.

Wilde, P. 2024. *Posthuman Gaming: Avatars, Gamers, and Entangled Subjectivities*. New York, NY, USA: Routledge.