2025 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS) | 979-8-3315-3228-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ETHICS65148.2025.11098410

# Narrative-Based AI Ethics Education for Emerging Technologies: Leveraging 'The Monkey's Paw' for AI Alignment and Social Justice

Joongho Lee
*Department of Technology and Society*
*Stony Brook University*
Stony Brook, USA
joongho.lee@stonybrook.edu

***Abstract*—The rise of autonomous AI agents introduces complex ethical concerns, particularly regarding value alignment, unintended consequences, and system accountability. Despite these growing risks, AI ethics remains underemphasized in many technical education programs. This paper proposes an original narrative-based instructional model using W. W. Jacobs's *The Monkey's Paw* (1902) to address key concepts in AI ethics and responsible innovation. The story's brevity, metaphorical richness, and public availability make it a practical resource for stimulating discussion around AI alignment and social impact. A classroom strategy is outlined to demonstrate how narrative tools can complement technical learning and promote ethical reflection among future developers and users.**

***Keywords—AI alignment, AI ethics education, narrative pedagogy, responsible innovation, unintended consequences, engineering ethics.***

## I. Introduction

As AI agents are increasingly embedded in daily life, their potential to operate independently while optimizing complex objectives raises ethical questions about how closely their actions reflect human intentions. Misalignments between design goals and real-world outcomes can lead to harms such as reward manipulation, biased decisions, or unintended escalation effects [1], [2]. While these challenges are recognized in AI safety research, structured opportunities to explore them in engineering classrooms remain rare [3]. Technical instruction often prioritizes system functionality and performance, leaving limited space for ethical analysis or value-sensitive design. Consequently, students may graduate with little preparation to anticipate or mitigate the societal impact of the systems they create [4]. To bridge this gap, narrative-based ethics education offers an accessible and engaging format for exploring the complex human dimensions of AI.

## II. Narrative Pedagogy Using *The Monkey's Paw*

One such narrative approach draws on *The Monkey's Paw*, a short story in which a family is granted three wishes, each fulfilled through literal but devastating outcomes [5]. The story's central message—that fulfilling a goal without understanding its broader context can result in unforeseen harm—parallels the alignment problem in AI. The symbolic nature of the tale provides fertile ground for examining themes such as control, moral agency, and technological risk.

The story's concise format (readable in under 15 minutes) allows it to be used during a single class session. Its open-ended interpretation encourages students from varied backgrounds to consider how intention, design, and consequence interact in both fictional and real-world systems. Instructors can further relate these discussions to ethical frameworks such as IEEE's *Ethically Aligned Design*, which emphasizes transparency, human well-being, and responsibility in intelligent systems [6].

## III. Educational Integration and Conclusion

A typical lesson plan might allocate time for in-class reading, followed by guided ethical discussion and applied reflection. Students are encouraged to map each "wish" in the story to modern examples of AI misbehavior—such as systems that optimize narrow objectives while neglecting social context. The session can conclude with reflective writing on how to prevent such outcomes in future technologies. This method supports active learning while requiring minimal logistical overhead. Its emotional impact and narrative resonance help foster lasting ethical awareness, which is often more difficult to achieve through abstract principles alone.

Ethics education in AI must move beyond theoretical discussion to engage learners in reflective, multidisciplinary inquiry. Stories like *The Monkey's Paw* offer a unique, high-impact avenue for addressing key issues in AI alignment and responsible innovation. When paired with structured facilitation, such narratives can prepare students to think more critically about the values embedded in technology and the consequences of their design choices.

## References

[1] S. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control*. New York, NY: Viking, 2019.

[2] D. Amodei et al., "Concrete Problems in AI Safety," *arXiv*, Jul. 2016, doi: https://doi.org/10.48550/arxiv.1606.06565.

[3] B. D. Mittelstadt et al., "The Ethics of algorithms: Mapping the Debate," *Big Data & Society*, vol. 3, no. 2, pp. 1–21, Dec. 2016, doi: https://doi.org/10.1177/2053951716679679.

[4] C. Cath et al., "Artificial Intelligence and the 'Good Society': the US, EU, and UK approach," *Science and Engineering Ethics*, vol. 24, 2018, doi: https://doi.org/10.1007/s11948-017-9901-7.

[5] W. W. Jacobs, "The Monkey's Paw," in *The Lady of the Barge*, London & New York: Harper & Brothers, 1902.

[6] IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, 1st ed., IEEE, 2019.

979-8-3315-3228-4/25/$31.00 ©2025 IEEE