# The Ghost in the Machine: Ethical Implications of AI in Moral Education

**Abstract**

The integration of Artificial Intelligence (AI) into the educational sphere represents an ontological shift in how knowledge is acquired and moral character is formed. This article interrogates the ethical dimensions of AI in the classroom through the lens of Aristotelian Virtue Ethics, specifically the concept of *phronesis* (practical wisdom). As AI moves from administrative support to "automated tutelage," we must ask whether the outsourcing of cognitive labor to Large Language Models (LLMs) stunts the development of moral agency. Furthermore, the study examines how algorithmic bias can reinforce systemic social and religious prejudices, threatening the pluralistic goals of modern education. By asserting the necessity of the "relational ethic" between teacher and student, the paper argues that while AI can provide information, it cannot cultivate virtue. The article concludes with a proposal for ethical guardrails designed to preserve human-centered moral growth in a digitized era.

---

## Introduction: The Rapid Ascent of the Automated Tutor

In less than a decade, Artificial Intelligence has transitioned from a speculative technological frontier to a ubiquitous presence in the global classroom. Initially relegated to administrative optimization—grading assistance, scheduling, and plagiarism detection—AI has now evolved into the "automated tutor." With the advent of Large Language Models (LLMs) such as ChatGPT and Claude, the educational landscape faces a "Ghost in the Machine": a non-sentient, algorithmic entity capable of simulating human dialogue, generating complex essays, and providing immediate answers to existential and ethical inquiries.

For the field of moral education, this ascent is particularly fraught. Unlike mathematics or the hard sciences, where AI might serve as a sophisticated calculator, moral education is predicated on the development of the "self." If education is, as the ancients believed, the cultivation of the soul, what happens when the primary interlocutor for a student is a statistical probability engine? As we inaugurate this volume of the *Journal of Religion, Ethics & Education (JREE)*, we must move beyond the utilitarian benefits of AI to address the deeper philosophical crisis it poses to human character formation.

---

## The Threat to Agency: *Phronesis* vs. Algorithmic Convenience

The core of Aristotelian Virtue Ethics is the development of **phronesis**, or practical wisdom. For Aristotle, virtue is not a theoretical set of rules to be memorized, but a "habit" (*hexis*) formed through repeated action and deliberate choice. It is the ability to navigate complex, real-world situations by finding the "mean" between extremes.

### The Stunting of Intellectual Muscle

The ethical risk of AI in the classroom is the erosion of the "struggle" necessary for *phronesis*. When a student relies on an algorithm to synthesize a moral argument or resolve a dilemma, they are bypassing the very process of deliberation that builds moral muscle.

- **Outsourcing Discernment:** *Phronesis* requires the agent to weigh competing values. If the "Machine" provides the synthesis, the student is no longer an agent but a consumer of a pre-packaged moral conclusion.
- **The Death of Inquiry:** Intellectual growth thrives on the "aporia"—the state of being at a loss. AI, designed to provide immediate, frictionless answers, removes the productive discomfort of not knowing, which is the starting point of true moral inquiry.

If virtue is a habit, then "algorithmic dependency" is a counter-habit. We risk producing a generation of students who possess "information" but lack the "wisdom" to apply it, effectively creating a void where the moral "self" should be.

---

## Bias and Inequality: Reinforcing the Secular and Sacred Prejudices

Algorithms are often presented as objective, but they are "ghosts" of the data they haunt. AI models are trained on massive datasets that reflect the historical, social, and religious biases of their creators and the internet at large.

### The Feedback Loop of Prejudice

In the context of religious and ethical education, these biases are particularly insidious:

1. **Western Hegemony:** Most LLMs are trained predominantly on Anglocentric data, often leading to an "Orientalist" or reductive view of non-Western religious traditions.
2. **Religious Erasure:** AI may struggle with the nuance of internal diversity within faiths, presenting a "sanitized" or "standardized" version of religion that ignores the lived experience of marginalized believers.
3. **Algorithmic Dogmatism:** Because AI predicts the "most likely" next word, it tends toward the middle-ground or the status quo. This "orthodoxy of the average" can stifle the radical or prophetic voices necessary for ethical progress.

The "Machine" does not merely reflect the world; it reinforces it. In an educational setting, this can inadvertently teach students that the "correct" moral or religious view is the one most commonly represented in the data, further marginalizing minority perspectives.

---

# The Relational Ethic: Why the Human Cannot Be Offloaded

At its heart, education is a relational endeavor. The "Teacher-Student" dynamic is not merely a transfer of data; it is an ethical encounter. Martin Buber's "I-Thou" philosophy reminds us that true learning happens in the "between"—the space where two human beings recognize each other's presence.

## The Limits of the LLM

An AI can simulate empathy, but it cannot "care." It can provide a script for moral behavior, but it cannot model virtue.

- **Modeling Virtue:** A student learns *phronesis* by observing it in a mentor. The way a teacher handles a classroom conflict, shows compassion to a struggling student, or admits their own moral failings provides a "living text" that an algorithm cannot replicate.
- **The Vulnerability of Learning:** Learning requires vulnerability. A student is more likely to engage in deep moral reflection when they feel "seen" by another human being. An AI provides an "I-It" relationship—a transactional interaction that lacks the moral weight of human accountability.

Offloading the teaching of ethics to an LLM is a categorical error. It treats morality as a series of logic puzzles to be solved rather than a way of being in the world to be lived.

---

# Conclusion: Proposing Ethical Guardrails for the JREE Vision

As the *Journal of Religion, Ethics & Education (JREE)* charts a course for the future, we must advocate for a "Hybrid Humanism." AI should be viewed not as a replacement for the teacher, but as a "sophisticated library" that requires a human librarian.

## The Guardrails

1. **Preservation of Deliberation:** Policies must ensure that AI is used to *scaffold* the deliberative process, not *replace* it. Students should be required to show the "work" of their moral reasoning.
2. **Algorithmic Transparency:** Educators must be trained to recognize and critique the biases inherent in AI outputs, using "Machine failure" as a teaching moment for media and religious literacy.
3. **Prioritization of the Relational:** School boards must protect the time and space for face-to-face dialogue, recognizing that the "human element" is the non-negotiable core of moral growth.

The "Ghost in the Machine" is only a threat if we forget that we are the ones who breathe life into education. By grounding our use of technology in the ancient wisdom of *phronesis*, we can ensure that while the tools of the classroom change, the soul of the student remains intact.

---

# References


1. **Aristotle.** (1999). *Nicomachean Ethics* (T. Irwin, Trans.; 2nd ed.). Hackett Publishing Company.
2. **Bostrom, N.** (2014). *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press.
3. **Buber, M.** (1970). *I and Thou* (W. Kaufmann, Trans.). Charles Scribner's Sons.
4. **Coeckelbergh, M.** (2020). *AI Ethics.* MIT Press.
5. **Dreyfus, H. L.** (1992). *What Computers Still Can't Do: A Critique of Artificial Reason.* MIT Press.
6. **Eubanks, V.** (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* St. Martin's Press.
7. **Floridi, L.** (2019). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities.* Oxford University Press.
8. **Gunkel, D. J.** (2012). *The Machine Question: Deconstructing the Relation between AI and Others.* MIT Press.
9. **Joshi, N. J.** (2021). *Artificial Intelligence and the Future of Education.* UNESCO.
10. **Kroger, J.** (2021). *Digital Virtues: A Philosophy of Technology for a Global Age.* Routledge.
11. **Noble, S. U.** (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism.* NYU Press.
12. **O'Neil, C.** (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown.
13. **Russell, S.** (2019). *Human Compatible: Artificial Intelligence and the Problem of Control.* Viking.
14. **Turkle, S.** (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other.* Basic Books.
15. **Vallor, S.** (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting.* Oxford University Press.
16. **Zuboff, S.** (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* PublicAffairs.


---

**Call for Manuscripts**

"Navigating the ethical frontiers of religion and technology in the classroom."

Join our inaugural cohort of researchers. We are currently seeking original research that investigates the intersection of AI, moral development, and religious education.

Formal Submission Portal: https://cjirs.com/index.php/cjcc/about/submissions

Peer-Reviewed • Open Access • Respectful Dialogue