International Journal of Social Research and Development
www.socialsciencejournal.net
Online ISSN: 2664-8709
Received: 14-05-2026, Accepted: 15-06-2026, Published: 30-06-2026
Volume 8, Issue 2, 2026, Page No. 122-124

Artificial intelligence in moral education: A guide or standard?

Innocent Chiawa Igbokwe
Department of Educational Management and Policy, Nnamdi Azikiwe University, Awka, Nigeria
DOI: https://doi.org/10.66856/ijsrd.2026.8.3.8059

Abstract
As Artificial Intelligence (AI) systems become embedded in classrooms, advising students, summarizing ethical dilemmas, and even simulating counsel, educators face a question that is at once practical and philosophical: should AI function merely as a guide that supports moral reasoning, or has it begun, often without deliberate decision, to operate as a standard against which moral judgments are measured? This article argues that AI can be a valuable instructional aid in moral education, but that treating its outputs as normative authorities is conceptually and ethically mistaken. Drawing on Pope Leo XIV's 2026 [7] encyclical *Magnifica* Humanitas, alongside established scholarship in AI ethics and moral-developmental theory, the article contends that machine systems lack the conscience, lived experience, and capacity for responsibility that moral judgment requires. It concludes with a framework for integrating AI into moral education as a guide subordinate to human discernment, rather than a substitute for it.

Keywords: Artificial intelligence, moral education, *Magnifica Humanitas*, education etc

Introduction

Moral education has traditionally relied on dialogue, lived example, and the slow cultivation of conscience through relationships with teachers, peers, and communities (Kohlberg, 1981) [2]. The arrival of generative AI tools capable of producing fluent ethical commentary, simulating moral dilemmas, and even offering instant verdicts on contested questions has unsettled this tradition. It is so surprising to see students ask chatbots, rather than teachers or mentors, whether an action is right or wrong, and instructors are experimenting with AI-generated case studies, debate partners, and feedback systems. My experience is that this attitude is on the increase today. The pedagogical promise is considerable, yet so is the risk that an output-generating tool could be mistaken for a moral authority.

This tension is not new to AI ethics generally, but it acquires particular weight in moral education because the subject matter is precisely the formation of conscience and character, not merely the transmission of information. Pope Leo XIV's encyclical *Magnifica* Humanitas (2026) [3], issued on the topic of safeguarding human dignity in the era of AI, offers a sustained reflection on this distinction. It is important to state that this intervention of the Catholic Pontiff is apt and instructive. The encyclical insists that AI systems may be useful instruments, yet warns against treating their outputs as a source of moral authority, since algorithms "do not have a moral conscience" and cannot "judge good and evil" or "bear responsibility for consequences" (Leo XIV, 2026, no. 99) [3].

This article uses that distinction, together with secular scholarship on AI ethics and moral development, to examine whether AI should function as a guide or a standard in moral education, and to propose criteria for keeping it in the former role.

What AI Can and Cannot Do in the Domain of Morality

A useful starting point is to clarify what current AI systems actually do when they produce ethical-sounding text. Large language models generate statistically probable sequences of words based on patterns in training data; they do not deliberate, intend, or possess a stance toward the good. *Magnifica* Humanitas makes this point sharply, noting that AI systems "merely imitate certain functions of human intelligence," that they may "simulate empathy and understanding" without actually understanding what they produce, and that whatever is called their "learning" is a form of statistical adaptation rather than the inner growth that comes from choices, mistakes, forgiveness, and fidelity over a lifetime (Leo XIV, 2026, no. 99) [3]. The same caution echoes through Floridi and Cowls's (2019) [1] widely cited ethical framework for AI, which treats systems as constituted by engineered outcomes rather than genuine cognition, regardless of how convincingly they mimic deliberation.

This matters directly for moral education, where the goal is not the production of a correct-sounding answer but the cultivation of practical wisdom, the capacity to perceive a situation accurately, weigh competing goods, and act with appropriate motive. Kohlberg's (1981) [2] cognitive-developmental model frames moral growth as advancing through stages of perspective-taking achieved largely through social interaction, the experience of disagreement, and the felt weight of consequences. Vallor (2016) [6] similarly argues that ethical competence in a technological age is best understood as a set of "technomoral virtues" cultivated through habituated practice within communities, not through the consumption of correct propositions. Both frameworks imply that whatever benefit AI offers to moral education, it cannot supply the lived, relational, and responsibility-bearing dimension of moral formation, because, as the encyclical observes, AI systems "do not undergo experiences, do not possess a body... and do not know from within what love, work, friendship or responsibility mean" (Leo XIV, 2026, no. 99) [3].

The Risk of AI as a Standard

The deeper danger explored throughout *Magnifica* Humanitas is not that AI will be used as a tool, but that it will quietly become a standard, an unacknowledged arbiter

of what counts as a reasonable, normal, or correct moral position. The encyclical states plainly that AI "is never morally neutral," because every system "embodies choices and priorities through what it measures, ignores and optimizes" (Leo XIV, 2026, no. 104) [3]. When learners encounter a chatbot's confident verdict on a contested ethical question, the apparent objectivity of the response can obscure the fact that it reflects the cultural assumptions, data selection, and design choices of those who built the system. The encyclical warns specifically of the "apparent objectivity of the responses and suggestions these systems provide," which can lead users to overlook the underlying assumptions embedded in training and design (Leo XIV, 2026, no. 100) [3].

This risk has an institutional dimension as well as an individual one. If a small number of firms control the AI systems most widely used in classrooms, their design choices about what is praised, flagged, or omitted in moral discussion become, in effect, an informal curriculum that operates beneath public scrutiny. This becomes a disaster to the modern world when such a company, institution, or organization relativizes truth, common good and values. UNESCO's (2021) [5] Recommendation on the Ethics of Artificial Intelligence addresses this concern directly in its education provisions, calling for AI literacy, critical thinking, and human oversight precisely because digitalizing societies require new safeguards against the unreflective adoption of automated judgments as though they were neutral facts. Selwyn (2019) [4] raises a parallel concern specific to teaching: automated systems can replicate some surface features of instruction, yet the social and emotional dimensions of teaching, including the formation of moral judgment through trusted relationship, resist full automation and should not be treated as if they could be outsourced wholesale to a machine.

A further concern raised in the encyclical is the diffusion of responsibility that follows when moral verdicts are delegated to opaque systems. Discussing automated decision-making more broadly, the document insists that "it is not permissible to entrust lethal or otherwise irreversible decisions to artificial systems" because "moral judgment cannot be reduced to calculation, for it involves conscience, personal responsibility and the recognition of the other as a person" (Leo XIV, 2026, no. 198) [3]. While that passage concerns life-and-death decisions in warfare, the underlying principle applies with equal force to moral education: a student who defers to an algorithm's verdict on a contested ethical question has not exercised conscience but has substituted calculation for it, regardless of how sophisticated the system's reasoning appears.

AI as a Guide: A More Defensible Role

None of this implies that AI has no place in moral education. *Magnifica* Humanitas explicitly describes AI as "a valuable tool that requires vigilance," and acknowledges its genuine benefits when properly bounded (Leo XIV, 2026, no. 100) [3]. Used as a guide rather than a standard, AI can expand the range of case studies available to a classroom, generate multiple perspectives on a dilemma for students to evaluate critically, provide low-stakes practice in articulating and defending a moral position, and surface counterarguments a student might not have considered. Floridi and Cowls's (2019) [1] five-principle framework, organized around beneficence, non-maleficence, autonomy, justice, and explicability, offers a workable checklist for evaluating whether a given classroom use of AI strengthens or undermines student autonomy and the capacity for independent judgment; a tool that supplies more material for deliberation supports autonomy, while one that supplies a final verdict undermines it.

The encyclical's discussion of education reinforces this distinction. It insists that the educational task is to teach students "to decide when and for what purpose it ought not to be used," warning that the speed and ease of AI-generated answers "risk extinguishing the desire to ask questions" upon which genuine understanding depends (Leo XIV, 2026, no. 140) [3]. Invoking Plato's account of philosophical learning as a process in which understanding is kindled only through sustained discussion, akin to striking flint until a spark catches, the encyclical insists that moral and intellectual maturity cannot be hurried by a tool that produces instant answers (Leo XIV, 2026, no. 140, citing Plato, Letter VII) [3]. This suggests a pedagogical rule of thumb: AI is functioning well as a guide when it lengthens and enriches the process of moral deliberation, and poorly, drifting toward the role of standard, when it shortens or replaces that process.

The encyclical's broader treatment of subsidiarity offers a further safeguard. Applied to AI governance generally, subsidiarity requires that "such processes not be imposed from above in an opaque and unilateral manner, but instead be directed toward the common good with transparency, accountability and meaningful forms of participation" (Leo XIV, 2026, no. 71) [3]. Transposed to the classroom, this principle implies that decisions about how AI is used in moral instruction should remain with teachers, students, and school communities who can examine, question, and where necessary override a system's output, rather than being dictated by the default settings of a commercial platform.

Toward a Framework for Practice

Bringing these strands together, three working criteria can help instructors keep AI in the role of guide rather than standard. First, transparency of process: students should be taught to ask how a given output was generated and what assumptions or omissions it likely reflects, rather than receiving it as a finished judgment. Second, retained responsibility: the final act of moral evaluation, weighing reasons, accepting or rejecting a position, and being answerable for that choice, should remain with the student and teacher, consistent with the encyclical's insistence that responsibility "must remain under effective, self-aware and responsible human control" (Leo XIV, 2026, no. 200) [3]. Third, relational anchoring: AI-assisted exercises should be embedded within discussion, mentorship, and community reflection rather than replacing them, in keeping with Kohlberg's (1981) [2] and Vallor's (2016) [6] shared premise that moral growth is achieved through relationship and habituated practice rather than information transfer alone.

Conclusion

The question of whether AI should function as a guide or a standard in moral education is not primarily a technical one; it is a question about where moral authority properly resides. The scholarship surveyed here, alongside the sustained argument of *Magnifica* Humanitas, converges on the conclusion that AI systems, however articulate, lack the conscience, lived experience, and capacity for responsibility

that moral judgment requires, and that their outputs should never be received as a settled verdict. Used thoughtfully, AI can enrich moral education by multiplying perspectives, sustaining inquiry, and supporting practice in ethical reasoning. Used uncritically, it risks becoming an unacknowledged standard that quietly displaces the very faculties, conscience, deliberation, and responsibility, that moral education exists to cultivate. The task for educators, then, is not to choose between embracing or rejecting AI, but to ensure, deliberately and continually, that it remains a guide in service of human moral formation rather than a substitute for it.

References

1. Floridi L, Cowls J. A unified framework of five principles for AI in society. Harvard Data Science Review, 2019, 1(1). https://doi.org/10.1162/99608f92.8cd550d1
2. Kohlberg L. The philosophy of moral development: Moral stages and the idea of justice. Harper & Row, 1981.
3. Leo XIV. *Magnifica* humanitas: Encyclical letter on safeguarding the human person in the time of artificial intelligence. The Holy See, 2026.
4. Selwyn N. Should robots replace teachers? AI and the future of education. Polity Press, 2019.
5. UNESCO. Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization, 2021.
6. Vallor S. Technology and the virtues: A philosophical guide to a future worth wanting. Oxford University Press, 2016.