Ethics of AI in Education

Aram Bahrini[1,*] and Matthew Roalkvam[1]
*bahrini@illinois.edu

[1] *Department of Business Administration*
Gies College of Business
University of Illinois at Urbana-Champaign
Champaign, IL, USA

***Abstract*—Generative and predictive AI systems are rapidly moving from the margins of the education system, such as optional tutoring tools and reference aids, into the core of instructional design, including automated feedback, adaptive learning, proctoring, and AI-mediated student support. This shift introduces a clear concern: while educational stakeholders value the scale, personalization, and efficiency promised by AI, these systems can also introduce ethical risks that are difficult to observe within classrooms, such as privacy leakage, disparate impact, opaque decision-making, erosion of academic integrity, and overreliance on uncertain AI-generated outputs. Although prior scholarship and policy frameworks provide general principles, educators and administrators still lack operational guidance that connects governance choices, such as disclosure practices and policy enactment and enforcement, to student-centered outcomes, including trust, perceived fairness, and learning engagement. To address this gap, we contribute both an actionable standards-to-controls matrix tailored to AI-in-education frameworks and an illustrative, fully specified field study in a business school context, demonstrating how governance features can be evaluated quantitatively. Using a three-condition experimental design, no explicit AI policy, AI policy enactment, and AI policy enactment with explicit enforcement, our results show significant differences in perceived transparency and trust in education. We conclude by discussing implications for policy design, academic instructional practice, and future research on auditable, education-aligned AI governance.**

***Index Terms*—Ethical AI, AI in Education, Academic Integrity, AI Governance**

I. Introduction

AI is no longer merely an educational tool. It is increasingly embedded in learning environments through systems that suggest explanations, summarize readings, generate practice problems, flag possible misconduct, and shape what instructors see through dashboards. These capabilities offer clear benefits, including scalable feedback, personalization, and lower administrative burden [1], [2]. The risk is equally compelling. When a system is wrong or biased, the classroom is where the consequences land, and it is students who feel them.

Existing guidance offers ethical principles such as transparency, fairness, privacy, and accountability [3]–[8]. However, educational decisions are made at the course, assignment, and platform levels, where principles must be translated into concrete instructional choices. What remains unclear is whether transparent AI-use policies shape students' trust in a course, feelings of support, and sense of control over academic outcomes. Educators often inherit AI environments whose ethical posture is only partly visible to students. In many cases, the policy arrives too late or not at all.

This paper develops a usable ethics framework for AI in education and applies it to an empirical field study in business courses. We analyze student responses from three Spring 2026 courses taught by the same instructor under three different policy conditions. We argue that ethical deployment of AI in education depends not only on the technology itself, but also on whether course policies make expectations legible to students. When policy design is visible, students may feel clearer about boundaries, more trusting of the course, and better able to judge what is expected of them. Yet not every educational outcome is affected in the same way; a policy can clarify rules without necessarily making a course feel more supportive.

From a research perspective, this paper links ethics principles to course-level outcomes and shows how variation in policy design can be analyzed using survey data. From a practical perspective, it offers a governance framework that instructors and administrators can adapt in classroom settings. However, this paper's outcomes are wide-reaching in nature. Business education is particularly salient for AI governance, as students are being trained for professional environments in which AI-assisted analysis, writing, and decision-making are increasingly commonplace, and recent work has begun to examine how generative AI affects student performance in business school settings [9]. As a result, classroom AI policy not only shapes perceptions of academic integrity but also students' developing expectations about legitimate AI use in their careers.

Business education provides a useful setting for illustrating how AI policy questions arise in ordinary environments. The examples below show how broad ethical concerns such as transparency, accountability, and academic integrity can appear in applied classroom settings.

- **Analytics:** a student uses a large language model (LLM) to produce code that runs, but does not understand the modeling or assumptions embedded in it.

- **Supply Chain Management:** an AI system produces a polished inventory recommendation with a formula error, and students relay it confidently in their homework submission.
- **Accounting and Finance:** an AI drafting assistant produces plausible written analysis that sounds professional, but students cannot explain the reasoning behind it when asked.

Far from abstract ethical questions, these are classroom management questions that affect judgment, legitimacy, and the conditions under which students learn.

The rest of the paper is organized as follows. Section II outlines the study design, measures, and analysis plan. Section III develops the ethics-to-controls framework and hypotheses. Section IV presents the descriptive and inferential findings. Finally, Section V discusses the implications, limitations, and avenues for future research.

II. Methodology

A. Study Design and Setting

The study was conducted in Spring 2026 in three courses taught by the same instructor in the Gies College of Business at the University of Illinois at Urbana-Champaign: BADM 211, BADM 378, and BADM 557. Each course used the same short extra-credit writing task. The wording of the assignment prompt was held constant, while the course policy surrounding AI use varied. Each course corresponded to one policy condition:

1) **No Policy**: no explicit AI policy was presented for the assignment.
2) **Disclose**: the assignment included an explicit AI policy clarifying expectations.
3) **Disclose & Check**: the assignment included an explicit AI policy and a statement that responses would be reviewed using AI detection tools.

A total of 91 usable survey responses were available for analysis. The final sample included 27 students in the No Policy condition, 37 in the Disclose condition, and 27 in the Disclose & Check condition.

B. Measures

The survey included three focal perception items tied directly to the assignment policy:

- **Control**: "How did the AI policy for this assignment impact your control over receiving the extra credit?"
- **Trust**: "What effect did the AI policy for this assignment have on your trust in the course?"
- **Support**: "What effect did the AI policy for this assignment have on your sense that the course supports your learning?"

Each outcome used the same four categorical response options: greatly decreased, slightly decreased, slightly increased, and greatly increased. For descriptive summaries, responses were coded symmetrically as -1, -0.5, 0.5, and 1, respectively. The outcomes were treated as ordered categorical variables rather than interval-scaled measures for the purposes of our analysis. The dataset also included a post-assignment self-report on whether the student used AI for the essay question, along with the essay word count. We did not include self-reported AI use or word count as primary regression controls because both were measured after the assignment and may themselves be shaped by the policy environment.

C. Analysis Plan

The analysis proceeded in three steps. First, we report descriptive statistics by policy condition using the symmetric coding described above. Second, we use Kruskal-Wallis tests to compare the three conditions without imposing interval-scale assumptions on the response categories, as it is a nonparametric, rank-based method for assessing differences across independent groups. Third, we estimate ordered logistic regression models for control, trust, and support, using No Policy as the reference category. Odds ratios are also provided for further analysis.

III. Conceptual Framework & Hypotheses

Table I translates popularly cited ethical principles into ethical controls for education. AI policy should be accessible to students in all contexts to promote positive outcomes, rather than prohibiting AI. Grounded in prior work on sociotechnical systems, transparency, and language-model risk [10]–[13], we examine three expectations:

- **H1:** *Courses with an explicit AI policy will be associated with higher reported trust than courses with no explicit policy.*
- **H2:** *Courses with an explicit AI policy will be associated with higher perceived control over earning the extra credit.*
- **H3:** *Perceived support for learning will be less sensitive to policy variation than trust and control.*

IV. Analysis & Results

A. Descriptive Statistics

Table II reports descriptive statistics by condition using the symmetric coding of the ordered response categories. The most visible descriptive gap appears in trust. The Disclose condition has the highest mean trust score ($\bar{x} = 0.49$), followed by Disclose & Check ($\bar{x} = 0.41$), with No Policy lowest ($\bar{x} = 0.22$). The pattern for control is similar in direction but smaller in size. The change in support is relatively insignificant across all three conditions.

The descriptive results also show a large difference in self-reported AI use on the assignment. Reported AI use was 44.4% in the No Policy course, 5.4% in the Disclose course, and 0% in the Disclose & Check course.

A more intuitive way to view the trust pattern is to focus on positive responses. The share of students reporting either slightly increased or greatly increased trust was 70.4% in No Policy, 91.9% in Disclose, and 85.2% in Disclose & Check. For control, the corresponding positive-response shares were 70.4%, 83.8%, and 77.8%. For support, all three conditions

TABLE I
OPERATIONAL ETHICS MATRIX FOR COMMON AI-IN-EDUCATION WORKFLOWS.

Principle	Where it currently appears in education	Typical risk	Operational controls
Transparency [4], [5], [7]	AI tutoring, feedback, proctoring	Students cannot tell what is machine-generated and what reflects instructor judgment	Up-front disclosure, explanation of limits, provenance cues, and clear policy language
Privacy [8], [14], [15]	Learning analytics, LMS logs, chat transcripts	Surveillance creep, secondary use, re-identification	Data minimization, retention limits, purpose limitation, opt-out where feasible
Fairness [10], [16]	Detection tools, grading support, risk scoring	Disparate burden or uneven false positives across students	Human review, appeal channels, subgroup auditing, procedural transparency
Accountability [6], [8]	Vendor platforms, automated workflows	No clear owner when harms occur	Named human owner, audit logs, documentation, incident response
Academic Integrity [11], [12]	Writing and coding assistance	Blurry boundary between assistance and substitution	Permissible-use policies, process evidence, staged drafting

TABLE II
DESCRIPTIVE STATISTICS BY POLICY CONDITION.

Metric	No Policy	Disclose	Disclose & Check	Overall
n	27	37	27	91
Control mean / SD	0.26 / 0.56	0.43 / 0.44	0.35 / 0.54	0.36 / 0.50
Trust mean / SD	0.22 / 0.49	0.49 / 0.34	0.41 / 0.42	0.39 / 0.41
Support mean / SD	0.44 / 0.45	0.49 / 0.40	0.48 / 0.32	0.47 / 0.39
Climate comp. mean / SD	0.31 / 0.44	0.47 / 0.33	0.41 / 0.34	0.40 / 0.37
Essay words mean / SD	109.93 / 67.92	108.84 / 29.54	117.85 / 21.47	111.93 / 42.18
AI use (%)	44.4%	5.4%	0.0%	15.4%

were already high, ranging from 85.2% to 92.6%, which suggests a ceiling effect limiting the benefits.

B. Group Differences

Kruskal-Wallis tests showed that the three course-policy environments differed significantly in trust, as indicated by the Kruskal-Wallis statistic ($H = 6.04$, $p = 0.049$), but not in control ($H = 0.92$, $p = 0.632$) or support ($H = 0.16$, $p = 0.922$).

The same pattern appears in the descriptive response distributions. Students in the policy courses, and especially the Disclose condition, were less likely to report decreases in trust. By contrast, support was positive in almost every group, leaving little room for differentiation.

Self-reported AI use varied sharply across conditions. A chi-square test showed a strong association between course-policy environment and self-reported AI use, $\chi^2(2) = 25.25$, $p < 0.001$. In practical terms, when the assignment policy dictated that students not use AI tools, the overwhelming majority of students did not.

C. Ordered Logistic Regression Models

Table III reports ordered logistic regression models with No Policy as the reference group. These models preserve the ordinal structure of the response categories rather than treating the four-category outcomes as interval-scaled.

For trust, the Disclose condition was associated with significantly higher odds of reporting a higher trust category relative to No Policy ($b = 1.446$, $OR = 4.24$, $p = 0.017$). The Disclose & Check condition was positive in the same direction but was not statistically significant ($b = 0.998$, $OR = 2.71$, $p = 0.110$). However, the omnibus likelihood-ratio test for the trust model was significant ($\chi^2(2) = 6.23$, $p = 0.044$).

For control, both policy conditions were positive relative to No Policy, but neither coefficient was statistically significant, and the overall model was not significant ($\chi^2(2) = 0.95$, $p = 0.623$). Due to the statistical insignificance of this model, we can only conclude descriptively that both policy conditions are associated with an increased sense of control.

For support, neither policy condition differed significantly from No Policy, and the omnibus model was not significant ($\chi^2(2) = 0.16$, $p = 0.921$).

In short, between the No Policy and the policy group, trust showed the clearest positive association with explicit policy, while control coefficients were directionally positive but not statistically distinguishable from No Policy.

V. DISCUSSION

A. Implications for Research

The study suggests three main implications for research. First, policy visibility appears to matter most for trust. Across all of our analysis, trust showed the clearest differentiation across course-policy environments, with the strongest pattern in the Disclose condition. Second, support behaved differently from trust and control. Reported support was already high

TABLE III
ORDERED LOGISTIC REGRESSION MODELS WITH NO POLICY AS THE REFERENCE GROUP.

Outcome	Predictor	Log-odds coefficient		Odds ratio		Omnibus model
		b	p	OR	95% CI	
Trust	Disclose	1.446	0.017	4.24	[1.30, 13.86]	$\chi^2(2) = 6.23,\ p = 0.044$
	Disclose & Check	0.998	0.110	2.71	[0.80, 9.25]	
Control	Disclose	0.488	0.333	1.63	[0.61, 4.38]	$\chi^2(2) = 0.95,\ p = 0.623$
	Disclose & Check	0.272	0.617	1.31	[0.45, 3.81]	
Support	Disclose	0.185	0.745	1.20	[0.40, 3.66]	$\chi^2(2) = 0.16,\ p = 0.921$
	Disclose & Check	-0.013	0.982	0.99	[0.30, 3.22]	

Note: Odds ratios (OR) are reported as exponentiated coefficients, $\exp(b)$.

across all three courses, leaving less variation in our data from which to explain and draw inference, suggesting that some student-centered outcomes may be less sensitive to policy design than others. Third, the findings reinforce a sociotechnical view of educational governance [10]. Students do not encounter ethical principles in the abstract; they encounter them through prompts, warnings, policies, and the tone of instructor communication. In that sense, course policy is not merely administrative: it is one of the practical interfaces through which students experience AI governance.

B. Implications for Practice

The findings suggest three main practical implications.

- **Make expectations visible.** The clearest result in the study was the association between explicit AI policy language and trust in the course. Relative to the No Policy condition, courses that disclosed their AI policy were associated with higher reported trust.
- **Be careful about the tone of enforcement.** The Disclose condition showed the strongest trust pattern. By contrast, when explicit policy language was paired with review language in the Disclose & Check condition, reported trust and control were directionally lower relative to Disclose.
- **Treat AI use itself as part of the policy response.** The sharp differences in self-reported AI use across conditions suggest that policy does not merely announce rules; it also shapes how students approach the task itself.

For educational practice, this means policy design should match assignment design. Students should know what kinds of assistance are allowed, what kinds are not, and how their own judgment will still be evaluated.

Beyond education, these findings speak to a broader governance problem that appears anywhere people abide by institutional rules while also navigating new automated tools. In workplaces, public agencies, healthcare settings, and professional services, trust is shaped not only by what a technology can do, but also by how clearly its use is explained, how expectations are communicated, and whether oversight feels supportive or punitive. Our results suggest that visible policy can strengthen legitimacy by reducing ambiguity, while enforcement-oriented framing may introduce new concerns about monitoring and autonomy.

C. Limitations

The study has important limitations. Most importantly, the policy condition was perfectly confounded with the course. BADM 557, the No Policy group, was a graduate course, while BADM 211 and BADM 378 were undergraduate courses. Even with the same instructor, similar assignment wording, course level, and peer norms, the differences in student composition and local classroom context may have contributed to the observed differences. A second limitation is the use of brief single-item outcomes. Trust, support, and control are meaningful, but each was measured using a single post-assignment question. A third limitation is that the response distributions were highly positive overall, especially for support, which reduced statistical leverage. A further limitation is that the outcome items did not offer a "No Change" category. As a result, a respondent who experienced little or no policy effect would be required to pick a direction, which may have caused neutral responses to appear as weak directional responses. Finally, self-reported AI use may be underreported in any study that asks students directly about their own behavior on an assignment tied to course rules.

D. Future Research Opportunities

Several extensions would strengthen this line of work.

- **Replicated multi-section designs.** Apply different policy conditions across multiple sections of the same course to better isolate the effects of AI policy apart from course context.
- **Richer measurement.** Use multi-item scales for trust, support, control, and fairness would improve precision and allow stronger analysis.
- **Access and equity measures.** Examine unequal experience with AI tools, including differences in access, familiarity, comfort, and willingness to use AI systems, since these factors may shape how students interpret both the policy and the assignment.
- **Behavioral outcomes.** Combine self-report with process evidence, draft history, or follow-up measures to study how policy shapes actual student behavior.
- **Longitudinal classroom studies.** Repeated measurements across a semester would show whether effects persist or fade as AI policy becomes routine.

- **Additional policy conditions.** Compare a wider range of AI policy designs to examine which policy features are most strongly associated with trust, control, support, and student performance.

E. Conclusion

Education cannot outsource ethics to tools. What students ultimately experience is not just the presence of AI, but the way a course explains, governs, and defines it. In this study, explicit policy language was associated with higher trust in the course, especially when compared with the No Policy condition.

Control showed a similar directional pattern, although the evidence was weaker, while support varied little across all groups. If instructors want students to view AI governance as legitimate, they should make expectations visible, understandable, and aligned with assignment design.

References

[1] W. Holmes, M. Bialik, and C. Fadel, Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Boston, MA, USA: Center for Curriculum Redesign, 2019.

[2] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education: Where are the educators?," International Journal of Educational Technology in Higher Education, vol. 16, no. 1, Art. no. 39, 2019.

[3] M. Coeckelbergh, AI Ethics. Cambridge, MA, USA: MIT Press, 2020.

[4] UNESCO, "Recommendation on the Ethics of Artificial Intelligence," Paris, France, 2021.

[5] F. Morandín-Ahuerma, "Ten UNESCO recommendations on the ethics of artificial intelligence," OSF Preprints, 2023.

[6] OECD, "Recommendation of the Council on Artificial Intelligence," 2019.

[7] High-Level Expert Group on Artificial Intelligence, "Ethics guidelines for trustworthy AI," European Commission, 2019.

[8] National Institute of Standards and Technology (NIST), "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," 2023.

[9] C. Bergenholtz, O. Vuculescu, F. Günzel-Jensen, and L. Frederiksen, "Leveling up or leveling down? The impact of generative AI on student performance in business schools," Academy of Management Learning & Education, early access, 2026, doi: 10.5465/amle.2025.0029.

[10] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi, "Fairness and abstraction in sociotechnical systems," in Proc. ACM Conf. on Fairness, Accountability, and Transparency (FAccT), pp. 59–68, 2019.

[11] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?," in Proc. ACM Conf. on Fairness, Accountability, and Transparency (FAccT), pp. 610–623, 2021.

[12] L. Weidinger et al., "Ethical and social risks of harm from language models," arXiv:2112.04359, 2021.

[13] R. Bommasani et al., "On the opportunities and risks of foundation models," arXiv:2108.07258, 2021.

[14] S. Slade and P. Prinsloo, "Learning analytics: Ethical issues and dilemmas," American Behavioral Scientist, vol. 57, no. 10, pp. 1510–1529, 2013.

[15] A. Pardo and G. Siemens, "Ethical and privacy principles for learning analytics," British Journal of Educational Technology, vol. 45, no. 3, pp. 438–450, 2014.

[16] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," ACM Computing Surveys, vol. 54, no. 6, pp. 1–35, 2021.