

Design Conditions for Educator Adoption of LLM-Based Ethics Scaffolding in Higher Education

Stefania Zourlidou, University of Koblenz, Germany

The European Conference on Education 2026
Official Conference Proceedings

Abstract

This paper examines what has to align before educators can integrate LLM-based ethics scaffolding into ordinary higher-education teaching without increasing marking burden, rule ambiguity, or inappropriate reliance on automated output. The paper develops an educator-facing framework from a bounded cross-case synthesis of recent higher-education literature and two prior design cases, one addressing classroom use of an LLM-supported ethics activity, and one addressing assessment redesign for keeping student reasoning inspectable under AI use. The paper proposes six design conditions for course integration: pedagogical positioning, workflow fit, visible reasoning, governance clarity, infrastructural reliability, and educator capability-building. These conditions suggest that adoption is less a matter of individual willingness to use AI than a matter of coordinated design across pedagogy, assessment, policy, infrastructure, and staff development. The paper contributes a practical framework, a temporal reading of how the conditions surface across the life of a course, and a worked illustration showing how the framework travels beyond computing to other judgment-intensive disciplines. Its central argument is that responsible adoption becomes more plausible when AI is bounded to a clearly defined pedagogical role and when student deliberation remains interpretable, assessable, and governable in everyday teaching conditions.

Keywords: generative AI, higher education, educator adoption, ethics education, assessment design

iafor

The International Academic Forum
www.iafor.org

Introduction

Generative AI has not eliminated the educational problem of ethics teaching, instead, it has made course-design choices more complicated. In many higher-education settings, instructors are under pressure to respond to tools that can produce fluent analyses on demand, although the educational target in ethics-related work is rarely fluent text alone. It is the capacity to frame a dilemma, weigh competing considerations, justify a position under uncertainty, and revise that position when better evidence or criticism appears. When AI tools enter this space, the practical question for educators is not simply whether a model can be useful, but whether the surrounding activity still keeps judgment visible, responsibilities clear, and marking manageable. That question extends well beyond computing courses, arising anywhere students are asked to reason through contested cases, professional responsibilities, or uncertain evidence. This paper addresses that problem by developing an educator-facing framework for course integration. It draws on a bounded cross-case synthesis of recent post-2022 literature on AI in higher education together with two prior design cases from the same broader research programme. The first is an exploratory classroom pilot of EthicsAdvisor (Oloke & Zourlidou, 2026), a structured LLM-based scaffold for ethical reflection in data-science projects. The second is an assessment-design study that introduced an Ethics Decision Log (EDL), a compact reasoning rubric, and a critique-and-revision workflow for AI-rich project courses (Zourlidou et al., 2026). The present paper does not re-evaluate EthicsAdvisor, reanalyse the pilot participants, or present the EDL and rubric as new contributions. Instead, the two prior cases serve as bounded analytic inputs to a different question: what has to align before an educator can adopt this kind of scaffold as a credible part of ordinary teaching practice? The resulting framework is aimed at educators and programme teams who need a practical way to think about adoption across pedagogy, assessment, governance, infrastructure, and professional learning.

Related Work

Recent reviews make clear that AI in higher education is already a large and fast-moving field, while also showing why an educator-facing synthesis remains useful. Bond et al. (2024), in a meta-review of secondary research on AI in higher education, identify rapid growth alongside uneven ethical, methodological, and contextual grounding. Their synthesis also shows substantial attention to broad AI-in-higher-education questions and to profiling and prediction. Crompton and Burke (2023), reviewing primary research from 2016 to 2022, identify assessment and evaluation, prediction, AI assistants, intelligent tutoring systems, and the management of student learning among prominent applications, while students receive more research attention than instructors. Read together, these reviews show a field rich in accounts of AI applications but less developed in educator-centred accounts of how such tools can be integrated responsibly into ordinary teaching practice. The present paper responds to that gap by shifting the level of analysis from technical possibility and student reception to adoption conditions for instructors.

A second body of work shows why educator adoption cannot be reduced to individual openness or technical skill. Chan (2023) proposes an AI Ecological Education Policy Framework organized across pedagogical, governance, and operational dimensions, explicitly linking teaching and learning to privacy, security, accountability, infrastructure, and training. Dotan et al. (2024) argue that higher education's traditions of collaborative faculty governance and academic freedom can sit uneasily with centralized, top-down approaches to AI governance. Jin et al. (2025), analysing policies and guidelines from 40 universities across six global regions, find proactive attention to academic integrity, teaching and learning, and equity, while

identifying remaining gaps particularly in data privacy and equitable access. They also emphasize clear communication channels, stakeholder collaboration, and defined roles and responsibilities. Miao and Holmes (2023) similarly place GenAI policy within a human-centred approach that foregrounds data privacy, human agency, and ethical and pedagogical validation. Therefore these findings suggest that responsible integration depends not only on educator willingness but also on policy clarity, institutional support, and workable governance arrangements.

The assessment literature sharpens the problem by showing that adoption is inseparable from what educators are being asked to grade. Luo (2024), analysing GenAI policies at 20 universities, found that the dominant policy framing centred on the originality of students' submitted work, even though other issues such as reliability and equal access were also present. Smolansky et al. (2023) and Kizilcec et al. (2024) report that educators strongly favour assessment approaches adapted to assume AI use and to encourage critical thinking, whereas student responses are more mixed. Khlaif et al. (2025) identify faculty motivations for assessment redesign that include academic integrity, preparation for future careers, technological change, and institutional policy, while also reporting professional-development, equity, and accessibility challenges. Xia et al. (2024) conclude that assessment transformation should cultivate self-regulated learning, responsible learning, and integrity, and should be accompanied by teacher professional development and more innovative, holistic teaching. This literature supports a design requirement central to the present synthesis, i.e., if an AI-related scaffold is to become part of routine teaching, the assessed work must leave interpretable evidence of student thinking and not rely on polished output alone.

A final strand of research brings educator capability and professional learning into view. Mah and Groß (2024) identify distinct faculty profiles associated with differences in AI use, self-efficacy, perceived benefits and challenges, and professional-development needs, with AI literacy emerging as an important challenge. Shahzad et al. (2025), in qualitative interviews with 29 management educators, report concerns including ethics, data privacy, academic integrity, institutional readiness, and the need to balance automation with human-centred teaching. Miao and Cukurova (2024) organize teacher AI competence across five dimensions: human-centred mindset, ethics of AI, AI foundations and applications, AI pedagogy, and AI for professional learning. This literature helps distinguish between visible reasoning, which concerns what an assignment makes observable in student work, and capability-building, which concerns what educators need to understand and be able to do to interpret that work, set boundaries for tool use, and teach students to evaluate model outputs critically.

Methodology

Methodologically, this paper is an integrative design synthesis, not a systematic review or a new empirical instructor-readiness study. The goal is to build a practically usable framework for one specific problem: what conditions make educator adoption of LLM-based ethics scaffolding plausible in ordinary course settings. This orientation follows integrative-review methodology, which uses existing literature to create new knowledge through synthesis (Torraco, 2016). The external literature therefore functions here as framework-building evidence. The corpus was limited to recent sources from 2023 onward that speak directly to higher-education AI adoption, governance, assessment redesign, educator perspectives, or professional capability. The two prior cases were deliberately read at the level of function and not treated as new empirical data. The EthicsAdvisor case was an exploratory, single-site classroom pilot with 20 data-science students that reported perceptions of usefulness, clarity,

applicability, and related classroom experience (Oloke & Zourlidou, 2026). In the present synthesis, that published case is used only to surface implementation questions such as pedagogical role framing, low-friction access, response behaviour, and the practical consequences of service latency or verbosity. No participant-level data were reanalysed, and the pilot's student-perception results are not presented as findings about educator adoption. The second case developed an assessment package consisting of the EDL, a 0-2 reasoning rubric, critique prompts, and worked exemplars (Zourlidou et al., 2026). Here, that case is used to examine the assessment implications of keeping student judgment inspectable under AI use. The present paper does not reproduce the EDL template or rubric as a new artifact, and it makes no new claim about their effectiveness.

The synthesis proceeded in four passes. First, the two prior cases were compared for the educator-facing problems they exposed at different points in a course. Second, each external source was coded for the educator-facing issue it illuminated, including workload, policy ambiguity, integrity, assessment redesign, staff capability, and institutional support. Third, recurrent concerns were clustered across the literature and tested against the two cases to determine whether they appeared across different parts of the teaching process rather than as isolated themes. Fourth, the clusters were refined into six design conditions, each expressed as an educator-facing question and linked to a course-integration implication (Table 1).

This procedure has clear limitations. The synthesis is deliberately bounded, the two prior cases are close to one another in topic and institutional context, and one is exploratory (not large-scale). The paper therefore does not claim that the six conditions are exhaustive, that all instructors experience them in the same way, or that the framework has already been validated across disciplines. Its more modest claim is that when recent literature is read alongside bounded course-design cases, a coherent set of educator-facing alignment problems becomes visible. The framework organizes those problems for further testing and for practical course-design discussion.

Table 1
Six Design Conditions for Educator Adoption

Condition	Educator-facing question	Course integration implication
Pedagogical positioning	What precise teaching job is the scaffold being asked to do?	Define the scaffold as a bounded aid for deliberation, not a substitute for judgment.
Workflow fit	Can the activity fit ordinary teaching time and grading time?	Use low-friction access, short task cycles, and explicit contingency steps.
Visible reasoning	Can student judgment be seen without intrusive monitoring?	Assess claims, warrants, revisions, and evidence checks rather than polished prose alone.
Governance clarity	Do rules for use, disclosure, privacy, and retention make sense to everyone?	Provide concise, transparent expectations tied to pedagogical purpose.
Infrastructural reliability	Will the tool behave predictably enough for live coursework?	Prepare for service instability, response-length control, and non-AI continuation routes.
Educator capability-building	Does the instructor know how to teach with the scaffold and interpret student work?	Support staff with calibration opportunities and professional learning on ethics, equity, and critique.

Framework and Synthesis Findings

The synthesis identifies pedagogical positioning as the first condition. Before a tool is evaluated for convenience or novelty, the educator has to decide what educational role it is allowed to play. Across the reviewed literature and the two prior cases, integration appears more defensible when the scaffold is bounded to a narrow judgment-supporting function: helping students surface relevant considerations, test assumptions, or widen a line of inquiry without claiming authority over the conclusion itself. This positioning is consistent with the EthicsAdvisor case, where model output was framed as scaffolding rather than as a decision or authoritative answer (Oloke & Zourlidou, 2026). The precise domain may differ, fairness in computing, confidentiality in nursing, professional responsibility in law, or stakeholder accountability in business, but the pedagogical issue is constant: the tool should support deliberation without displacing responsibility.

The second condition is workflow fit. A practice may be conceptually attractive and still be unsuitable for routine course use if it cannot live inside ordinary teaching and grading time. The synthesis repeatedly points to low-friction access, manageable task length, predictable review demands, and explicit contingency steps as relevant integration concerns. The earlier classroom pilot also documented first-call latency and occasional verbosity as practical limitations, illustrating how apparently minor interaction costs can shape the feasibility of live use (Oloke & Zourlidou, 2026). Workflow fit therefore concerns both classroom activity and marking time: a sound pedagogical idea can still fail as a course practice if it is cumbersome to run, supervise, or assess at scale.

The third condition is visible reasoning, the point at which pedagogical aims and assessment requirements meet. For adoption to be credible, the activity needs to leave an interpretable trail of how students framed a case, weighed competing considerations, checked claims, and revised their position. The prior assessment-design study operationalized this concern through an EDL and rubric that grade decision-specific reasoning and revision rather than polished prose (Zourlidou et al., 2026). The present framework generalizes the design requirement rather than reintroducing those artifacts: visible reasoning can be implemented in different ways, but the work submitted for assessment should make claims, warrants, revisions, evidence checks, and uncertainty inspectable. When such a trail is absent, fluent AI-assisted prose can obscure shallow engagement; when it is present, educators have a basis for judging deliberative quality without relying on intrusive monitoring.

The fourth condition is governance clarity. Educators can often work around imperfect tools more easily than they can work around ambiguous rules. The reviewed literature points to governance questions concerning acceptable use, disclosure, privacy, equitable access, communication, and responsibility for AI-supported teaching and assessment (Dotan et al., 2024; Jin et al., 2025; Miao & Holmes, 2023; Shahzad et al., 2025). The two prior cases reinforce the practical importance of making expectations explicit. For course integration, clarity need not mean an elaborate policy; it can mean a concise explanation of why the scaffold is being used, what students must disclose, what data are retained, and what continuation route exists when AI is unavailable or inappropriate. Without such clarity, otherwise well-designed activities are harder to justify and harder for students to interpret consistently.

The fifth condition is infrastructural reliability. Although this condition may look technical, its consequences are pedagogical. If access is unstable, response length is unpredictable, or live sessions stall, the activity can move away from deliberate reasoning and toward frustration or

hasty acceptance. The classroom-use case provides a bounded illustration of this problem through observed latency and verbosity issues, while also showing that explicit limitations and verification advice can help keep model output in a supporting role (Oloke & Zourlidou, 2026). The synthesis therefore treats reliability broadly: service stability, predictable interaction cost, scope control, and a clear non-AI continuation route all matter. A tool does not need to be perfect to be teachable, but it does need to be dependable enough that an educator can plan around it.

The sixth condition is educator capability-building. This condition is distinct from visible reasoning even though the two interact closely. Visible reasoning concerns what the activity makes available for inspection; capability-building concerns what the instructor must know and be able to do in order to design the task, explain boundaries, interpret student work, and challenge inappropriate reliance on model output. Recent studies identify self-efficacy, AI literacy, ethical and privacy concerns, institutional readiness, and professional-development needs as salient factors in educators' experiences of AI adoption (Mah & Groß, 2024; Shahzad et al., 2025). The teacher competency framework of Miao and Cukurova (2024) further situates capability across human-centred, ethical, technical, pedagogical, and professional-learning dimensions. In practical terms, capability-building means that instructors need examples, calibration opportunities, and a language for explaining both the value and the limits of AI-supported critique. Without that support, even a well-designed activity can fail because the educator has been asked to improvise too much.

A brief course illustration shows how the six conditions can operate together. Consider a third-year course in which teams must analyse the ethical implications of a proposed AI-enabled service. The same underlying structure could be adapted for a computing lab, a nursing case seminar, a law-and-technology elective, or a business analytics capstone. The instructor introduces the AI scaffold as a bounded aid for testing assumptions and broadening analysis, not as a source of final answers. Students submit a short reasoning record at an intermediate checkpoint and a revised version later, explaining how their position changed, what evidence or critique informed that change, and where uncertainty remains. This checkpoint-and-revision pattern is informed by the earlier assessment-design case (Zourlidou et al., 2026), but it is used here only as an illustration of how the six adoption conditions can be coordinated, not as a newly evaluated intervention. The assessment focuses on the quality of the deliberative trail rather than stylistic polish; students receive clear guidance about acceptable use and disclosure; a non-AI fallback is available; and the instructor has seen calibrated examples before grading begins.

There is also a temporal logic worth making explicit. The six conditions do not merely coexist. Instructors encounter them at different points across the life of a course. Before teaching begins, pedagogical positioning and governance clarity dominate, because the educator has to decide what the scaffold is for, whether it is permitted, and how expectations will be communicated. When the task is introduced, workflow fit and capability-building become more visible, because students need manageable instructions and the instructor needs enough fluency to explain examples, boundaries, and likely failure modes. During active use, infrastructural reliability and visible reasoning come to the foreground, because access problems or overfluent output can immediately distort participation and judgment. At grading time, visible reasoning, workflow fit, and capability-building converge again, since the educator must interpret evidence efficiently and distinguish substantive revision from cosmetic editing. This time-based reading strengthens the framework by treating adoption not as a one-time decision but as a chain of design commitments that must survive contact with teaching practice.

Discussion

The main implication of the synthesis is that educator adoption is best understood as a coordination problem and not as a question of individual attitude or technical enthusiasm. The framework is not intended to prescribe a single artifact, interface, or workflow but its purpose is to help educators judge whether local course conditions are sufficient for responsible integration. An instructor may recognize pedagogical value in AI-supported ethics work and still avoid adoption if the workflow is too cumbersome, if institutional rules are vague, or if the platform cannot be relied on during class. Conversely, a technically smooth tool may still fail educationally if it obscures the deliberative work that assessment needs to see or shifts authority toward model fluency. This is why the six conditions should not be read as an additive checklist in which strength in one area easily compensates for weakness in another. Pedagogical positioning gives the activity a legitimate teaching purpose. Workflow fit makes it livable, visible reasoning makes it reviewable, governance clarity makes it justifiable, infrastructural reliability makes it dependable in teaching time, and capability-building makes it teachable.

For institutions and programme teams, the framework suggests a distributed model of support. Adoption should not be delegated entirely to the individual educator, but neither can it be solved by generic top-down policy statements alone. Programme leaders can identify where judgment-heavy tasks already appear in capstones, case seminars, simulations, or reflective placements, learning designers can help reshape those moments into assessable activities, IT teams can support low-friction access and continuation routes, and academic development units can provide exemplars, calibration routines, and professional learning that address ethics and equity as well as tool operation. This institutional view matters beyond computing. A business lecturer redesigning a consultancy brief, a nursing educator structuring case-based reflection, or a law instructor using AI-supported critique in a professional-responsibility seminar may not share the same disciplinary standards, but they face similar adoption questions about clarity, trust, workload, and student accountability. The framework is therefore best read as discipline-sensitive but structurally portable.

The limitations of the paper should remain visible. The framework is grounded in a bounded corpus and two prior design cases that are close to one another in topic and institutional context, and one of those cases is exploratory rather than large-scale. The synthesis is therefore better at identifying plausible design conditions than at ranking them by causal importance. It also does not settle discipline-specific questions about what counts as sufficient evidence, feasible mitigation, acceptable disclosure, or sustainable workload in every field. Those judgments still require local interpretation. Future work should test the six conditions across a broader disciplinary range, examine how they interact with institutional policy maturity, and explore how staff capability-building and visible reasoning reinforce one another over time.

Conclusion

This paper proposes six aligned design conditions for educator adoption of LLM-based ethics scaffolding in higher education: pedagogical positioning, workflow fit, visible reasoning, governance clarity, infrastructural reliability, and educator capability-building. The article develops an educator-facing framework from a bounded cross-case synthesis of recent literature and two prior design cases. In doing so, it shifts attention from whether a model or artifact works in isolation to the educational and institutional alignments required for credible course integration. The broader lesson is that responsible adoption should make human judgment more interpretable, not less. In AI-rich classrooms, educators need activities,

interfaces, and support structures that keep deliberation open to review, revision meaningful, and accountability human. When an AI scaffold is given a bounded pedagogical role, embedded in a manageable workflow, supported by clear rules, delivered through dependable infrastructure, accompanied by assessable evidence of student deliberation, and matched by adequate educator learning, it becomes more plausible as a support for ethical reflection without displacing responsibility. Although the argument begins with ethics-related teaching, its relevance extends to other higher-education settings in which students are asked to reason through contested cases, professional obligations, and uncertain evidence.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

The author declares that OpenAI ChatGPT (GPT-5.6 Sol) was used to assist with language refinement, structural editing, and manuscript formatting. The tool was not used to generate, alter, or analyse study data. The author reviewed the resulting text and retains responsibility for the argument, interpretation, citations, and final submitted version.

References

- Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, 21, Article 4. <https://doi.org/10.1186/s41239-023-00436-z>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20, Article 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, Article 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Dotan, R., Parker, L. S., & Radzilowicz, J. (2024). Responsible adoption of generative AI in higher education: Developing a “points-to-consider” approach based on faculty perspectives. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2033–2046). Association for Computing Machinery. <https://doi.org/10.1145/3630106.3659023>
- Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, 8, 100348. <https://doi.org/10.1016/j.caeai.2024.100348>
- Khlaif, Z. N., Alkouk, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning assessments for AI-enhanced learning: A framework for educators in the generative AI era. *Education Sciences*, 15(2), 174. <https://doi.org/10.3390/educsci15020174>
- Kizilcec, R. F., Huber, E., Papanastasiou, E. C., Cram, A., Makridis, C. A., Smolansky, A., Zeivots, S., & Radulescu, C. (2024). Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. *Computers and Education: Artificial Intelligence*, 7, 100269. <https://doi.org/10.1016/j.caeai.2024.100269>
- Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the “originality” of students’ work. *Assessment & Evaluation in Higher Education*, 49(5), 651–664. <https://doi.org/10.1080/02602938.2024.2309963>
- Mah, D. K., & Groß, N. (2024). Artificial intelligence in higher education: Exploring faculty use, self-efficacy, distinct profiles, and professional development needs. *International Journal of Educational Technology in Higher Education*, 21, Article 58. <https://doi.org/10.1186/s41239-024-00490-1>
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. <https://doi.org/10.54675/ZJTE2084>

- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Oloke, P., & Zourlidou, S. (2026). EthicsAdvisor: A pilot LLM-based tool for ethical reflection in data science education. In *2026 22nd International Conference on Information Technology Based Higher Education and Training (ITHET)*, Lillehammer, Norway, 2026, pp. 1–8, IEEE. <https://doi.org/10.1109/ITHET69978.2026.11584968>
- Shahzad, F., Avunduk, Z. B., Ullah, Z., & Arslan, A. (2025). Understanding management educators' perspectives on AI adoption: A qualitative study. *The International Journal of Management Education*, 23(3), 101274. <https://doi.org/10.1016/j.ijme.2025.101274>
- Smolansky, A., Cram, A., Radulescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and student perspectives on the impact of generative AI on assessments in higher education. In *Proceedings of the Tenth ACM Conference on Learning @ Scale* (pp. 378–382). Association for Computing Machinery. <https://doi.org/10.1145/3573051.3596191>
- Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, 15(4), 404–428. <https://doi.org/10.1177/1534484316671606>
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21, Article 40. <https://doi.org/10.1186/s41239-024-00468-z>
- Zourlidou, S., Oloke, P., & Hopfgartner, F. (2026). Grading the reasoning, not the answer: Assessable artifacts for LLM-assisted ethics in data science projects. In *Proceedings of the 18th International Conference on Computer Supported Education (CSEDU 2026)* (Vol. 2, pp. 1578–1589). SciTePress. <https://doi.org/10.5220/0014826600004021>

Contact email: zourlidou@uni-koblenz.de