DOI https://doi.org/10.30525/978-9934-26-708-6-12

LLM AGENTS IN LEARNING MANAGEMENT SYSTEMS: ETHICAL DIMENSIONS AT THE LEARNING–RESEARCH INTERFACE

Puhach R. S.

INTRODUCTION

Digital technology has reshaped higher education from the inside. What began as a way to modernise a few processes now affects how universities teach, assess, and build knowledge. Students and staff mostly welcome the change, since it widens access, adds flexibility, and lets decisions rest on better data – but those gains are not automatic. They depend on whether an institution is ready, whether its systems work, and whether its people have the skills to use the tools well.[1] Recent work on the field takes a sharper view of where this is going. Studies that once treated artificial intelligence (AI) as one classroom tool among many now read it as a force reshaping the whole system, and since 2022 three concerns have come to the fore: generative AI, academic integrity, and education policy.[2] The turning point was the arrival of large language models (LLMs) that can write fluent text on demand. Almost at once, universities faced questions about authorship, assessment, and the trustworthiness of knowledge that their existing rules were never built to answer[3].

The learning management system (LMS) sits at the centre of this change. For years, platforms like Moodle, Canvas, and Blackboard mostly held things in place: they kept the course materials, set out the activities, took in assignments, and recorded grades, but they did little on their own. That is now changing. A new wave of tools builds LLM-driven agents straight into these platforms, so the LMS can tutor a student, deliver a lesson, give feedback, and help with assessment. Intelligent tutoring systems had already

[1] Nazyrova A., Miłosz M., Bekmanova G. et al. The digital transformation of higher education in the context of an AI-driven future. *Sustainability*. 2025. Vol. 17, № 22. Art. 9927. P. 1. DOI: 10.3390/su17229927.

[2] Hong T. T. M., Tung N. T. T., Thanh N. T. P. Mapping artificial intelligence research in higher education toward sustainable development. *Discover Sustainability*. 2025. Vol. 6, № 1. Art. 1240. P. 5. DOI: 10.1007/s43621-025-02162-0.

[3] Slimi Z. A systematic critical review of generative AI's impact on authorship, pedagogy, and integrity (2023–2025). *Frontiers in Education*. 2026. Vol. 11. Art. 1769680. P. 1. DOI: 10.3389/feduc.2026.1769680.

shown what adaptive, responsive teaching could achieve[4], and generative agents take that further by bringing open-ended conversation into the platform itself. This shift is not ethically neutral. Once an LLM agent is built into the LMS, it reaches into two overlapping parts of academic life – what students do as learners, and what students and staff do as researchers – and it carries the ethical risks of both. On the teaching-and-learning side, researchers warn about over-reliance, the loss of skills, unequal access, and the weakening of student agency;[5] on the research side, the same technology raises worries about who counts as an author, whether work is original, how contributions are credited, and the invention of false citations[6]. The LMS is exactly where these two worlds meet. It is both the place where students learn and the channel through which their work – research projects, theses, scholarly writing – is submitted, marked, and assessed. An agent living there straddles the line between learning integrity and research integrity, yet the two bodies of work on these subjects have grown up largely apart.

This study takes up that gap. It asks four questions: how LLM agents are being built into the LMS, what ethical issues arise across the learning and research they touch, what governance and design responses people have proposed, and what we still do not know. In answering them, it draws a scattered literature together, reads it through a single lens – integrity – so that the teaching side and the research side are weighed together, and stays honest about the limits of the current evidence. The argument runs in two parts of roughly equal weight. The first sets out how the agents are integrated and maps the ethical terrain of teaching, learning, and research. The second turns to the governance these risks call for and builds the paper's central proposal: integrity-by-design at the learning–research interface.

The review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement; the protocol was specified in advance, and prospective registration, for example in PROSPERO or the Open Science Framework, is recommended to strengthen transparency. The Scopus and Web of Science (Core Collection) databases were searched as the primary sources, supplemented by Google Scholar to broaden coverage and capture grey

[4] Lin C.-C., Huang A. Y. Q., Lu O. H. T. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learning Environments*. 2023. Vol. 10, № 1. Art. 41. P. 1. DOI: 10.1186/s40561-023-00260-y.

[5] Madleňák R., Madleňáková L., Cvacho V., Gachulinec D. Ethical challenges of artificial intelligence in higher education: a four-pillar student-activity framework for institutional governance. *Education Sciences*. 2026. Vol. 16, № 4. Art. 555. P. 2. DOI: 10.3390/educsci16040555.

[6] Nguyen T. H. Ethical challenges in the era of generative AI: insights from a practice-informed rapid review. *International Journal of Research and Innovation in Social Science*. 2025. Vol. 9, № 10. P. 8146–8162. DOI: 10.47772/IJRISS.2025.910000666.

literature and by arXiv for recent preprints, which are flagged as such throughout. A two-stream search design was used to accommodate the narrow conceptual anchor of the review without losing sensitivity to its ethical dimensions: an integration-core stream combined technology terms (large language model, LLM, generative AI, AI agent, conversational agent) with platform terms (learning management system, LMS, Moodle, Canvas, Blackboard) and higher-education terms, screened afterwards for ethical content; an ethics-core stream combined the same technology terms with ethics terms (ethics, integrity, governance, privacy, bias, fairness, accountability, transparency) and higher-education terms, screened afterwards for LMS relevance. Searches were limited to English-language records published between January 2022 and 2026. Records were screened in two stages – title-and-abstract, then full text – against pre-specified inclusion criteria, with dual independent screening recommended to reduce selection bias; a structured form then extracted bibliographic, integration, ethical, and governance characteristics, coded separately for learning-related and research-related activity where studies distinguished them. Given the heterogeneity of the evidence, a narrative thematic synthesis was adopted, organised throughout by the learning–research interface, with the recurrent constraints of short study horizons, small samples, and a bias toward positive findings treated as cross-cutting threats to certainty. Because the precise intersection that defines the review – LLM agents integrated into the learning management system, examined through an ethical lens – is both narrow and very recent, the literature addressing it directly is still small, and the searches returned a correspondingly modest pool rather than the large volume that a broader query on generative AI in education would yield. This narrow, well-specified scope, which mirrors the focus of the broader doctoral research from which the review arises – the ethical dimensions of scientific research in the digital transformation of higher education and the development of AI – meant that a relatively high proportion of the identified records proved on-topic. The records excluded were those addressing generative AI in higher education only in general terms, or drifting into adjacent technical literatures concerned with the software engineering of agent systems or with natural-language-processing methods, both lacking the educational-and-ethical anchor that defines the review. The proportion of records retained therefore reflects the precision of the search together with the small, still-emerging state of this specific literature, rather than permissive screening. The study flow is summarised in Figure 1.

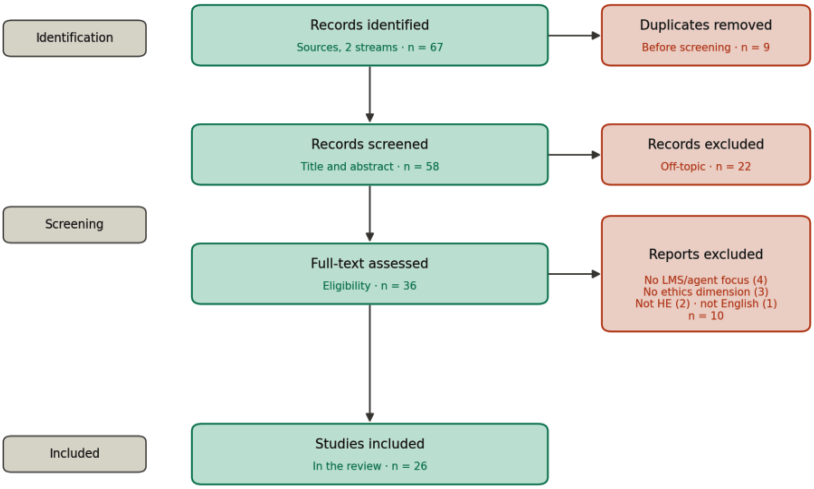

Fig. 1. PRISMA 2020 flow diagram of the study selection process

1. Integration of LLM agents into the LMS and the ethics of teaching, learning, and research

This first part sets out what is being integrated and what can go wrong. It describes how LLM agents are built into the learning management system and how good the evidence on them actually is, then works through the ethical concerns that keep coming up across the learning–research interface: the teaching-and-learning side, the research side, and the questions of privacy, bias, transparency, and accountability that run through both. The aim is to give the governance discussion in the second part a clear picture of the risks it has to answer to.

It helps to start with what a learning management system is. It is the software a university uses to run its courses online – to organise material, deliver it, and keep track of who is learning what. For about twenty years these systems mostly stored and sorted. They held the readings, laid out the activities, collected assignments, returned grades, and logged progress, but they did not think about a student's answer or change how they behaved in response. An LLM agent is a different kind of thing. It puts a large language model at its core and uses it to read a situation, make a plan, and act on a goal, instead of simply replying once to a prompt. It can break a task into steps, keep track of what has been said earlier in a conversation, and call on other tools or sources to get something done. Putting an agent like this inside the LMS does more than add a feature – it changes what the platform is,

turning a passive store into an active participant in teaching. Researchers have settled on a few standard ways of doing this. The main one is retrieval-augmented generation: the agent is tied to the course's own materials – its textbooks, lecture notes, and recordings – and answers from those rather than from the broader, less reliable knowledge built into the model. This matters in teaching, because it keeps the agent on topic and cuts down on the things models tend to invent. The open-source LAMB framework is a good example, pairing ordinary software engineering with generative AI and retrieval-augmented generation to create assistants meant to live inside an LMS[7]. Designs vary in scale. Some use a single agent for one job, such as answering questions about a course; others coordinate several specialised agents – one to retrieve material, one to assess, one to manage the conversation – and the tasks they take on range across tutoring, delivering content, marking and feedback, designing curricula, and tailoring instruction to the individual[8].

Empirical designs at the level of specific learning activities are emerging alongside these architectural surveys, and they are instructive precisely because they are mixed. On the encouraging side, a generative-AI-based conversational agent built to provide cognitive, metacognitive, and social scaffolding during online collaborative writing was found to have a positive effect on students' cognitive engagement, illustrating how a purpose-built agent situated within a concrete, LMS-mediated task can support learning in a way that a general-purpose chatbot does not[9]. On the cautionary side, evaluations of agent-generated feedback reveal sharp limits: in a university course on concurrent programming, neither of two leading chatbots could accurately assess student exercises for characteristic errors such as starvation, deadlocks, and race conditions when their judgements were measured against expert-teacher ground truth, despite the generally positive reception such tools enjoy[10]. The juxtaposition is important, because it shows that an embedded agent may perform well at scaffolding a learner's

[7] Alier M., Pereira J., García-Peñalvo F. J., Casañ M. J., Cabré J. LAMB: an open-source software framework to create artificial intelligence assistants deployed and integrated into learning management systems. *Computer Standards & Interfaces*. 2025. Vol. 92. Art. 103940. P. 1. DOI: 10.1016/j.csi.2024.103940.

[8] Córdova-Esparza D.-M. et al. AI-powered educational agents: opportunities, innovations, and ethical challenges. *Information*. 2025. Vol. 16, № 6. Art. 469. P. 12. DOI: 10.3390/info16060469.

[9] Hu W., Tian J., Li Y. Enhancing student engagement in online collaborative writing through a generative AI-based conversational agent. *The Internet and Higher Education*. 2025. Vol. 65. Art. 100979. P. 1. DOI: 10.1016/j.iheduc.2024.100979.

[10] Estévez-Ayres I., Callejo P., Hombrados-Herrera M. Á., Alario-Hoyos C., Delgado Kloos C. Evaluation of LLM tools for feedback generation in a course on concurrent programming. *International Journal of Artificial Intelligence in Education*. 2025. Vol. 35, № 2. P. 774–790. DOI: 10.1007/s40593-024-00406-0.

process and yet fail at evaluating a learner's product – two functions that are easily conflated in deployment. At the most consequential end of the spectrum, field deployments instantiate an LLM-based instructor agent within a live university course and link it to the LMS as a primary instructional actor, with a human instructor retaining responsibility for course structure and for the question-and-answer sessions where accountability matters most[11]. Two findings recur across this body of work and frame everything that follows. First, the evidence consistently favours hybrid human–AI configurations over fully autonomous ones; a synthesis of eighty-two studies reports that workflows in which educators curate and moderate agent output combine scalable automation with pedagogical expertise more effectively than autonomous tutors, and the feedback-evaluation evidence explains why, by showing where unsupervised agent judgement breaks down. Second, the empirical foundation remains immature, and the synthesis itself cautions that limitations in the current evidence base – characteristically short study horizons, small samples, and a tendency toward positive reporting – temper the generalisability of the reported gains[12]. Integration, in short, is technically demonstrated but not yet robustly evaluated – and this evidential immaturity is itself ethically significant, because the pedagogical benefits invoked to justify accepting the ethical risks examined below are themselves not yet firmly established.

On the learning side, the concern with the most evidence behind it is over-reliance and what it does to thinking. One randomised study compared students helped by ChatGPT, by a human expert, by writing-analytics tools, and by nothing extra. AI help raised the quality of the immediate work, but it also brought on what the authors call "metacognitive laziness": students handed the planning, monitoring, and judging that learning depends on over to the tool, and the better-looking final product did not turn into lasting knowledge or carry over to new tasks[13]. That is what makes the risk easy to miss. The work is not worse; the learning is, and the gap hides behind a polished result. Survey and interview data point the same way on a larger scale, finding that heavier AI use goes with weaker critical thinking, with cognitive offloading as the link[14]. Findings like these put evidence behind an old worry: a tool that hands over answers instead of helping a student work

[11] Towards AI agents for course instruction in higher education: early experiences from the field: preprint. *arXiv*. 2025. P. 1. DOI: 10.48550/arXiv.2510.20255.

[12] Córdova-Esparza D.-M. et al. Op. cit. P. 23.

[13] Fan Y., Tang L., Le H. et al. Beware of metacognitive laziness: effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*. 2025. Vol. 56, № 2. P. 489–530. DOI: 10.1111/bjet.13544.

[14] Gerlich M. AI tools in society: impacts on cognitive offloading and the future of critical thinking. *Societies*. 2025. Vol. 15, № 1. Art. 6. P. 13–14. DOI: 10.3390/soc15010006.

things out can erode the productive struggle that understanding is built from. An LMS tutoring agent, on call at any hour to every student, is in a perfect position to create that situation at scale. The problem deepens because unsupervised agents are unreliable graders in hard subjects, so an agent can both take over a student's effort and misjudge the result, leaving them under-stretched and wrongly marked at once[15]. Past the cognitive side, the literature keeps returning to de-skilling, loss of student agency, and unequal access, and groups them with integrity and governance as the field's main worries, noting that they show up differently across the different things students do[16]. The shared recommendation is to change the design, not to ban the tool: because fully autonomous tutoring does worse than hybrid setups, and because the cognitive harm clusters where students lean on the tool unguided, the fix is to build agents that keep the teacher in the loop and the student mentally active. This is the first time a single remedy – human oversight – appears; it returns on the research side and runs through the rest of the review.

On the research side, the literature rests on one settled point and a set of growing worries. The settled point is that an AI system cannot be an author, because authorship means taking responsibility for the work and a system cannot do that. Institutional rules for responsible AI use in research start from this, and it is the fixed point the rest of the debate works out from[17]. The growing worries gather around who counts as an author and whether the resulting knowledge can be trusted. Critical reviews track how generative tools are unsettling authorship and credit, and warn that the usual reactive policies cannot keep up[18]; studies of the research workflow flag risks to authorship, scientific integrity, and transparency as AI moves into writing, interpreting data, and generating ideas[19]; and many note the technology cuts both ways, since the same features that promise speed and access also threaten authenticity[20]. One fast-growing strand is about peer review, where trust matters most. Using a language model in a review without saying so can undercut transparency, blur the line between a reviewer's judgement and the machine's, and complicate the editorial oversight the published record

[15] Estévez-Ayres I., Callejo P., Hombrados-Herrera M. Á., Alario-Hoyos C., Delgado Kloos C. Op. cit. P. 774.

[16] Madleňák R. et al. Op. cit. P. 4–5.

[17] Smith S., Tate M., Freeman K. et al. A university framework for the responsible use of generative AI in research: preprint. *arXiv*. 2024. P. 4. DOI: 10.48550/arXiv.2404.19244.

[18] Slimi Z. Op. cit. P. 1.

[19] Sousa N. Ethical and practical implications of AI in academic library research. *IFLA Journal*. 2025. P. 1. DOI: 10.1177/03400352251391753.

[20] Haider E., Irshad N. N., Imran S. B. Negotiating the ethical frontier: the double-edged role of generative AI in scientific research. *Annals of Medicine & Surgery*. 2025. P. 1. DOI: 10.1097/MS9.0000000000004351.

depends on – even when the model improves clarity and speed[21]; and other work turns these worries into concrete warning signs and recommendations for keeping review honest[22]. Rapid reviews grounded in research-integrity codes add further harms, including made-up citations and the watering down of real authorship, and they too land on the same remedies: human oversight, accountability, and disclosure[23]. None of this is a loose analogy for the present review. A student writing a thesis or a paper through an LMS agent is working inside exactly the research-integrity regime these studies describe – while, on the same platform, their coursework falls under a separate academic-integrity regime. That is the heart of the learning–research interface: one agent, one platform, two regimes. And the literature covers each regime far more closely than it covers the join between them.

Two more issues cut across both sides of the seam. The first is privacy, where the sheer amount of data an LLM agent handles matters most. Because these systems take in large volumes of often-sensitive personal information, reviews regularly put privacy and data protection among the top ethical risks of educational AI. The recommended answer is privacy-by-design – building data minimisation and protection in from the start instead of bolting compliance on later – together with disclosure that lets the people affected see what the system does with their data[24]. At the policy level, the first global guidance on the technology makes protecting user data a core duty and points out that the tools have come out faster than national regulation can keep up, leaving user data poorly protected and institutions unready to vet what they buy.[25] The interface sharpens the question in a way the literature has not really taken in: one agent in the LMS gathers data from a student's day-to-day learning and from their research in the same place, and those two streams answer to different rules – ordinary data protection on one side, research-data and human-subjects rules on the other – yet a single mechanism collects, stores, and logs them without telling them apart. The second cross-cutting issue is bias. The foundational account of bias in education shows that algorithms can disadvantage particular groups, that

[21] Hosseini M., Horbach S. P. J. M. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity and Peer Review*. 2023. Vol. 8, № 1. Art. 4. P. 1. DOI: 10.1186/s41073-023-00133-5.

[22] Kocak B., Onur M. R., Park S. H., Baltzer P., Dietzel M. Ensuring peer review integrity in the era of large language models: a critical stocktaking of challenges, red flags, and recommendations. *European Journal of Radiology Artificial Intelligence*. 2025. Vol. 2. Art. 100018. P. 1. DOI: 10.1016/j.ejrai.2025.100018.

[23] Nguyen T. H. Op. cit. P. 8146.

[24] Madleňák R. et al. Op. cit. P. 4.

[25] Miao F., Holmes W. Guidance for generative AI in education and research. Paris : UNESCO, 2023. P. 14. URL: https://discovery.ucl.ac.uk/id/eprint/10176438/ (date of access: 25.06.2026).

bias can creep in at many points in how a system is built and used, and that some students – those with disabilities, international students, those from poorer backgrounds – are under-represented in the training data, so the systems work worst for those already worst served.[26] Generative AI carries this forward rather than fixing it. A clear and directly relevant case is the finding that widely used GPT detectors wrongly flag much non-native English writing as AI-generated while getting native writing almost always right.[27] That one result hits both sides of the interface at once: the same detector acts as an academic-integrity tool on a coursework essay and a research-integrity tool on a manuscript, and in both it penalises the same non-native writers unfairly. Across all of this, transparency and accountability are what bias makes non-negotiable, because when a system is opaque it is hard to say who is answerable for the decisions it shapes. That is exactly why the same prescription keeps coming up – disclose that AI was used and keep process evidence such as prompt and version logs – so those decisions can be checked afterwards instead of taken on faith.

Taken together, these threads point to one observation that drives the second part. Across integration, the ethics of teaching and learning, research integrity, privacy, and bias, the safeguard that comes up most often – keep a person accountable and keep the process checkable – is the same on both sides of the learning–research interface. Yet it is put into practice through instruments that stay split between an academic-integrity regime and a research-integrity regime. The second part turns to those instruments, to the wider governance around them, and to the proposal that follows from the split.

2. Governance and integrity-by-design at the learning–research interface

This second part turns to the governance response. It lays out the instruments meant to deal with the risks above – at the international, regulatory, and institutional levels – and then makes the study's main argument: that the real weakness today is not a lack of safeguards but the way they are split between two regimes, and that integrity-by-design at the learning–research interface is a coherent way to close that split. It ends with the tensions the interface brings out, what they mean for institutions in practice, and the gaps and limits that qualify the analysis.

[26] Baker R. S., Hawn A. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*. 2022. Vol. 32, № 4. P. 1052–1092. DOI: 10.1007/s40593-021-00285-9.

[27] Liang W., Yuksekgonul M., Mao Y., Wu E., Zou J. GPT detectors are biased against non-native English writers. *Patterns*. 2023. Vol. 4, № 7. Art. 100779. P. 1. DOI: 10.1016/j.patter.2023.100779.

Governance works at three levels, and they read best as a step down from broad principle to everyday practice. At the international level, UNESCO's first global guidance on generative AI in education and research sets out a human-centred approach, makes protecting user data a core duty, proposes concrete steps including a minimum age for using such tools on one's own, and urges governments to build the regulatory capacity the technology has outrun.[28] At the regulatory level, the European Union's Artificial Intelligence Act lays down binding, risk-based rules: it sorts systems by how much risk they pose and raises the obligations as the risk rises. Many serious educational uses count as high-risk, and for those the duties – data governance, technical documentation, record-keeping, and real human oversight – bear directly on how an LMS agent has to be built and run.[29] Sector studies are starting to work this general framework down into the specifics of higher education, spelling out what it means for universities day to day.[30] For research in particular, the European Research Area Forum's living guidelines tell researchers to disclose their use of AI, check AI-generated text for bias, and keep AI out of sensitive jobs such as peer review – a rare case of governance that already, if quietly, covers both the writing and the reviewing of scholarship[31]. At the institutional and design levels, finally, the literature settles on principles turned into working safeguards. Universities are advised to boil the tangled, shifting rules down into a short statement of principles that then steers training, communications, infrastructure, and process change, so that everyday decisions follow from settled commitments rather than improvisation[32]; named frameworks give researchers step-by-step ethical guidance for AI-assisted work[33]; and the three cross-cutting principles drawn from the most directly relevant of these frameworks are disclosure backed by process evidence such as prompt and version logs, privacy-by-design, and fairness scaffolds such as guaranteed access and bias audits[34]. Alongside these structural measures, a growing body of work treats AI literacy not as a nice-to-have but as a first-

[28] Miao F., Holmes W. Op. cit. P. 18.

[29] Smuha N. A. Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 (EU Artificial Intelligence Act). *International Legal Materials*. 2025. P. 1–148. DOI: 10.1017/ilm.2024.46.

[30] Tabacu A. E. Regulatory framework for the use of artificial intelligence in European higher education. *Jurídicas CUC*. 2025. Vol. 21, № 1. P. 124–152. DOI: 10.17981/jurideuc.21.1.2025.07.

[31] European Research Area Forum. Living guidelines on the responsible use of generative AI in research. Brussels : European Commission, Directorate-General for Research and Innovation, 2024. P. 6–7.

[32] Smith S., Tate M., Freeman K. et al. Op. cit. P. 2.

[33] Eacersall D., Pretorius L., Smirnov I. et al. Navigating ethical challenges in generative AI-enhanced research: the ETHICAL framework for responsible generative AI use: preprint. *arXiv*. 2024. P. 1. DOI: 10.48550/arXiv.2501.09021.

[34] Madleňák R. et al. Op. cit. P. 6.

order safeguard, setting out competency frameworks that help students and staff understand what the tools do, where they fall short, and what they mean ethically[35].

Read together rather than separately, these responses reveal something striking. The strongest, most consistent safeguard in the whole literature – keep a person accountable and keep the process checkable – is the same on both sides of the learning–research interface, yet the instruments that put it into practice stay split into an academic-integrity regime and a research-integrity regime, each with its own policies, language, and enforcement. The result is a mismatch built into the structure: one source of risk handled by divided tools. The mismatch is not abstract. The LMS is now where a university's learning and research meet, and an agent embedded there works across both. To return to the earlier example, a biased AI-detection tool penalises a non-native writer in just the same way whether the text is a first-year essay under academic-integrity policy or a thesis chapter under research-integrity policy[36]. The remedies on offer already sense this overlap, since they prescribe the same things – human oversight, disclosure, process logging – on both sides. What they do not yet do is set those remedies up in a way that fits the single object they are meant to govern. Part of why the split has held is organisational rather than intellectual. In most universities, academic integrity sits with teaching-and-learning offices and student-conduct procedures, while research integrity sits with research offices, ethics committees, and graduate schools. The two seldom share staff, systems, or even vocabulary, so a tool that moves between them has no natural home: each side can assume the other is watching, and the seam goes ungoverned. So this study proposes *integrity-by-design at the learning–research interface* as a unifying frame. The point is not to fold academic and research integrity into one flat standard, because the two really do govern different things – honest learning on one side, trustworthy knowledge on the other – and merging them would lose distinctions that matter. The point is that when the same agent on the same platform handles both, the design and governance response should hold together across them. In practice, the disclosure and the prompt-and-version logging that the literature recommends separately for marking and for research auditing are, at the level of the LMS agent, the very same mechanism. They could be specified, bought, and governed once instead of twice – designed into the agent as a single audit trail that serves whichever regime a given use falls under. Such a trail would record, for each real interaction, the prompts a student gave, the model and version used, the

[35] Chiu T. K. F., Ahmad Z., Ismailov M., Sanusi I. T. What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*. 2024. Vol. 6. Art. 100171. P. 1. DOI: 10.1016/j.caeo.2024.100171.

[36] Liang W., Yuksekgonul M., Mao Y., Wu E., Zou J. Op. cit. P. 1.

course materials or outside sources the agent pulled in, and the points where a human checked or overrode it. Captured once, at the platform level, that one record answers both the academic-integrity question of how a student reached a submitted assignment and the research-integrity question of how a draft was produced – with no need to build or reconcile two separate systems. Designing the mechanism once, rather than bolting parallel logs onto two regimes after the fact, is what integrity-by-design at the interface actually asks for.

A short example shows what this buys. Picture a student who, in a single term, uses the LMS agent both to draft a coursework essay and to develop a chapter of their thesis. Under the split arrangement, the essay is checked by an academic-integrity tool and the chapter, if it is checked at all, by a separate research-integrity process – each with its own record, or none. Under a single audit trail, both rest on one record: the prompts the student gave, the model and version used, the sources the agent retrieved, and the points where the student revised or a supervisor stepped in. The same evidence answers either question, and the student is never asked to account for the same work twice in two different vocabularies. The trail does not judge the work; it makes the work legible – which is what each regime was trying to do on its own.

There is a tension inside the proposal itself, and it should be named. An audit trail that records prompts, sources, and revisions is also a detailed record of how a particular student thinks and works, and a system that logs everything in the name of integrity can shade into surveillance. The same privacy-by-design that governs the agent's other data has to govern the trail: it should capture what is needed to answer the integrity question and no more, be readable only by those with a defined reason, and be kept only as long as that reason lasts. Disclosure cuts both ways here – students should know the trail exists, what it holds, and who can see it, just as they are told when an agent is in use. Handled carelessly, the mechanism meant to protect integrity becomes one more store of sensitive data of exactly the kind the privacy literature warns against; handled well, it can be auditable without being intrusive. The point is not that logging is unsafe but that it inherits the duties of any sensitive record, and integrity-by-design at the interface has to carry those duties with it.

Reading the evidence through the interface also surfaces tensions that single-domain studies tend to keep apart, and naming them is part of the contribution. The first is between automation and integrity: the same power that lets an agent give help at scale can, when it hands over answers instead of guiding, bring on metacognitive laziness in learning and push aside

authorship in research – one mechanism doing harm in two settings[37]. The second is between personalisation and privacy: the data that lets an agent adapt to a particular student is exactly what raises the privacy stakes, and an agent that stores both learning and research data piles that risk into one place. The third is between efficiency and accountability: the time an agent saves is only real if a person stays genuinely answerable for the result, which is why the safeguard the literature leans on hardest is also the hardest to put into practice – accountability cannot be automated. The fourth is between innovation and regulation, made concrete by the EU AI Act's risk-based duties, which deliberately rein in high-risk uses through documentation, record-keeping, and human oversight[38]. For institutions, the upshot of the whole analysis is to treat an LMS agent as one governance object that spans teaching, learning, and research, rather than a teaching tool covered by academic-integrity policy with research left to a separate code. A statement of principles of the kind recommended for research can be widened to cover the agent's teaching role, so that one coherent set – human accountability, transparency, privacy-by-design, and fairness – governs the agent wherever it works; disclosure and logging can be set once and applied to coursework and scholarship alike; AI-literacy provision, too often built separately for students or for researchers, can cover both roles, since today's student usually fills both at once; and procurement and design can favour ethics-by-design and the hybrid, human-in-the-loop setups that the evidence finds both more effective and less risky than fully autonomous ones[39]. In practice this hands institutions a concrete checklist to put to the vendors and developers of LMS agents: whether the agent gives an exportable audit trail of the kind above, whether its points of human review and override can be configured rather than fixed, whether its retrieval can be limited to institution-approved materials, where and under whose laws the data it gathers are stored, and whether its behaviour can be tested for bias across the language and demographic groups the institution serves. Asking these questions at procurement, before an agent is embedded, is itself integrity-by-design, since it is far easier to build governability in than to retrofit it once the agent is already running across teaching, learning, and research. None of this requires an institution to wait for a complete framework before acting. The most consequential first step is also the most modest: switch on disclosure and process-logging by default, so that every substantive use of the agent leaves a record, and decide who may read it. With that record in place, the rest – shared principles, a position statement that covers teaching as well as research, AI-literacy provision for students

[37] Fan Y., Tang L., Le H. et al. Op. cit. P. 489.
[38] Smuha N. A. Op. cit. P. 1.
[39] Córdova-Esparza D.-M. et al. Op. cit. P. 22–23.

who occupy both roles – can be built up gradually on evidence the institution actually holds, rather than designed in the abstract. The partition closes not in one reform but in a sequence that begins with making the agent's work visible. The theoretical contribution of the review is to reframe a debate that has run in two separate literatures: by putting the LMS at the centre, where learning and research meet, it shows that the field's apparent fragmentation hides a deeper agreement at the level of remedy, and a matching, fixable split at the level of governance.

The review also turns up several gaps that mark the limits of what can be claimed now and set the agenda for what comes next. The most pervasive is the state of the evidence: across both integration studies and learning-outcome studies, the designs tend to be short and small, with a lean toward positive results, so robust, long-run, and where possible causal evaluations of LLM agents in real LMS settings are needed before anyone can speak confidently about benefit or harm. A second gap is the interface itself: very little work separates the learning-data and research-data streams that an embedded agent gathers, or studies how to govern an agent that works across both – the very situation this review puts at the centre. A third is the shortage of studies that look at agents actually built into the LMS, rather than generic AI tools used beside it, with several of the most relevant integration studies still circulating as preprints awaiting peer review. A fourth is representation: the ethical analyses themselves note patchy global coverage and the under-representation of some student groups in the data behind educational AI, which limits how far any fairness claim can reach. These gaps sit alongside the limits of this review. Sticking to English-language sources, chosen for precise terminology, brings a language bias and may under-represent work and policy from non-English settings; leaning on two databases plus Google Scholar and a preprint source, though broad, cannot promise complete coverage of so fast-moving a field, and the 2022–2026 window leaves out both earlier groundwork and anything published after the search; including preprints improves currency at the cost of material that has not been peer-reviewed; and the thematic synthesis, while suited to mixed evidence, rests on judgement, as does the learning–research interface itself, which organises the analysis well but is a chosen lens that brings some relationships forward and pushes others back. Future work should therefore prioritise governance frameworks that span the interface, privacy-by-design built and tested in practice, equity audits of how agents behave across language and demographic groups, AI-literacy provision designed for the double role a single student now fills, and above all the move of the current preprint evidence into peer-reviewed findings.

CONCLUSIONS

Building LLM agents into learning management systems is one of the most far-reaching parts of the digital change in higher education, because it puts generative, goal-directed capability at the exact point where a university's learning and research meet. The first part of this review showed that the ethical risks of that move – to integrity, privacy, fairness, student agency, and accountability – are well documented on each side of the learning–research line, and that the safeguards proposed across the literature share a common core: human accountability backed by a record that can be checked. The second part showed that, although a single agent on a single platform works across that line, the governance meant to handle it stays split between an academic-integrity regime and a research-integrity regime. The study has argued that integrity-by-design at the learning–research interface is a coherent way to close that split – not by erasing the real difference between honest learning and trustworthy knowledge, but by governing the agent that handles both as the single thing it is. For now, the technology is further along than our ability to evaluate or govern it; closing that gap, in a way that spans the interface rather than splitting it, is the task this review sets for institutions, for designers, and for the long-run, interface-aware, and globally representative research the field still lacks.

Treating the LMS agent as the single object this review has argued it to be carries two wider lessons. The first is that technical performance alone cannot settle the ethical questions such an agent raises. An agent may support learning well and answer faithfully from the course material and still leave open problems of authorship, privacy, and fairness – problems that come not from how well it works but from where it sits and what it touches. The ethical difficulty lies in the agent's position at the meeting point of learning and research, and it is that position, not some flaw in the model, that governance has to address. The second lesson is that the most durable safeguard the literature offers – human accountability backed by a checkable record – is also the hardest to automate, and so the most dependent on institutional will. Building the audit trail, setting the points of human review, and writing the one set of principles that governs the agent across teaching, learning, and research are choices an institution has to make on purpose and early, because the alternative is not a neutral absence of governance but the quiet spread of two split regimes over a technology that has already erased the line between them. What this review contributes is to make that choice visible and to give it a name.

SUMMARY

This article looks at the ethics of building large language model (LLM) agents into learning management systems (LMS) in higher education.

It argues that the LMS has become the place where students' learning and the research of students and staff meet, so that an embedded agent acts on both at once. Drawing on a systematic review carried out under PRISMA 2020 – covering peer-reviewed and selected preprint work from 2022 to 2026 across Scopus, Web of Science, Google Scholar, and arXiv – the study identifies familiar integration patterns, namely retrieval-augmented grounding, single- and multi-agent designs, and instructor agents, alongside five recurring ethical areas: academic and research integrity, privacy and data protection, bias and fairness, student agency, and governance. Reading each area across the learning–research interface, it finds that the evidence, though growing, is still mostly short and small-sample work, and that an unsupervised agent's judgement can be unreliable in hard subjects. It shows that the strongest safeguard proposed across the literature is the same on both sides of the interface: human accountability backed by a record that can be checked. The central finding is that, although one agent works across the learning–research line, the governance for each side stays split. The study proposes integrity-by-design at the learning–research interface as a unifying response, treating the LMS agent as a single governance object. It sets out what this means for institutional policy and points to the long-run, interface-aware, and globally representative evidence that future research will need to supply.

REFERENCES

1. Alier M., Pereira J., García-Peñalvo F. J., Casañ M. J., Cabré J. LAMB: an open-source software framework to create artificial intelligence assistants deployed and integrated into learning management systems. *Computer Standards & Interfaces*. 2025. Vol. 92. Art. 103940. DOI: 10.1016/j.csi.2024.103940.

2. Baker R. S., Hawn A. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*. 2022. Vol. 32, № 4. P. 1052–1092. DOI: 10.1007/s40593-021-00285-9.

3. Chiu T. K. F., Ahmad Z., Ismailov M., Sanusi I. T. What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*. 2024. Vol. 6. Art. 100171. DOI: 10.1016/j.caeo.2024.100171.

4. Córdova-Esparza D.-M. et al. AI-powered educational agents: opportunities, innovations, and ethical challenges. *Information*. 2025. Vol. 16, № 6. Art. 469. DOI: 10.3390/info16060469.

5. Eacersall D., Pretorius L., Smirnov I. et al. Navigating ethical challenges in generative AI-enhanced research: the ETHICAL framework for responsible generative AI use: preprint. *arXiv*. 2024. DOI: 10.48550/arXiv.2501.09021.

6. Estévez-Ayres I., Callejo P., Hombrados-Herrera M. Á., Alario-Hoyos C., Delgado Kloos C. Evaluation of LLM tools for feedback generation in a course on concurrent programming. *International Journal of Artificial Intelligence in Education*. 2025. Vol. 35, № 2. P. 774–790. DOI: 10.1007/s40593-024-00406-0.

7. European Research Area Forum. Living guidelines on the responsible use of generative AI in research. Brussels : European Commission, Directorate-General for Research and Innovation, 2024.

8. Fan Y., Tang L., Le H. et al. Beware of metacognitive laziness: effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*. 2025. Vol. 56, № 2. P. 489–530. DOI: 10.1111/bjet.13544.

9. Gerlich M. AI tools in society: impacts on cognitive offloading and the future of critical thinking. *Societies*. 2025. Vol. 15, № 1. Art. 6. DOI: 10.3390/soc15010006.

10. Haider E., Irshad N. N., Imran S. B. Negotiating the ethical frontier: the double-edged role of generative AI in scientific research. *Annals of Medicine & Surgery*. 2025. DOI: 10.1097/MS9.0000000000004351.

11. Hong T. T. M., Tung N. T. T., Thanh N. T. P. Mapping artificial intelligence research in higher education toward sustainable development. *Discover Sustainability*. 2025. Vol. 6, № 1. Art. 1240. DOI: 10.1007/s43621-025-02162-0.

12. Hosseini M., Horbach S. P. J. M. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity and Peer Review*. 2023. Vol. 8, № 1. Art. 4. DOI: 10.1186/s41073-023-00133-5.

13. Hu W., Tian J., Li Y. Enhancing student engagement in online collaborative writing through a generative AI-based conversational agent. *The Internet and Higher Education*. 2025. Vol. 65. Art. 100979. DOI: 10.1016/j.iheduc.2024.100979.

14. Kocak B., Onur M. R., Park S. H., Baltzer P., Dietzel M. Ensuring peer review integrity in the era of large language models: a critical stocktaking of challenges, red flags, and recommendations. *European Journal of Radiology Artificial Intelligence*. 2025. Vol. 2. Art. 100018. DOI: 10.1016/j.ejrai.2025.100018.

15. Liang W., Yuksekgonul M., Mao Y., Wu E., Zou J. GPT detectors are biased against non-native English writers. *Patterns*. 2023. Vol. 4, № 7. Art. 100779. DOI: 10.1016/j.patter.2023.100779.

16. Lin C.-C., Huang A. Y. Q., Lu O. H. T. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic

review. *Smart Learning Environments*. 2023. Vol. 10, № 1. Art. 41. DOI: 10.1186/s40561-023-00260-y.

17. Madleňák R., Madleňáková L., Cvacho V., Gachulinec D. Ethical challenges of artificial intelligence in higher education: a four-pillar student-activity framework for institutional governance. *Education Sciences*. 2026. Vol. 16, № 4. Art. 555. DOI: 10.3390/educsci16040555.

18. Miao F., Holmes W. Guidance for generative AI in education and research. Paris : UNESCO, 2023. URL: https://discovery.ucl.ac.uk/id/eprint/10176438/ (date of access: 25.06.2026).

19. Nazyrova A., Miłosz M., Bekmanova G. et al. The digital transformation of higher education in the context of an AI-driven future. *Sustainability*. 2025. Vol. 17, № 22. Art. 9927. DOI: 10.3390/su17229927.

20. Nguyen T. H. Ethical challenges in the era of generative AI: insights from a practice-informed rapid review. *International Journal of Research and Innovation in Social Science*. 2025. Vol. 9, № 10. P. 8146–8162. DOI: 10.47772/IJRISS.2025.910000666.

21. Slimi Z. A systematic critical review of generative AI's impact on authorship, pedagogy, and integrity (2023–2025). *Frontiers in Education*. 2026. Vol. 11. Art. 1769680. DOI: 10.3389/feduc.2026.1769680.

22. Smith S., Tate M., Freeman K. et al. A university framework for the responsible use of generative AI in research: preprint. *arXiv*. 2024. DOI: 10.48550/arXiv.2404.19244.

23. Smuha N. A. Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 (EU Artificial Intelligence Act). *International Legal Materials*. 2025. P. 1–148. DOI: 10.1017/ilm.2024.46.

24. Sousa N. Ethical and practical implications of AI in academic library research. *IFLA Journal*. 2025. DOI: 10.1177/03400352251391753.

25. Tabacu A. E. Regulatory framework for the use of artificial intelligence in European higher education. *Jurídicas CUC*. 2025. Vol. 21, № 1. P. 124–152. DOI: 10.17981/jurideuc.21.1.2025.07.

26. Towards AI agents for course instruction in higher education: early experiences from the field: preprint. *arXiv*. 2025. DOI: 10.48550/arXiv.2510.20255.

Information about the author:
Puhach Roman Serhiyovych,
Post-Graduate Student
Dragomanov Ukrainian State University
9 Pyrohova str., Kyiv, Ukraine, 01601