Az-Zubayr Harrison

## Balancing Assistance and Integrity: The Role of Large Language Models in Academic Writing and Learning Support

Large Language Models (LLMs) have entered higher education at a time of large cohorts, limited staff capacity, and rising expectations for personalised support. This review draws on fifteen peer reviewed sources published between 2023 and 2025 that focus on generative AI or LLMs in higher education. Studies were included if they addressed higher education, examined LLMs or related generative tools, and engaged with at least one of three domains, academic writing and feedback, academic integrity, or governance and policy. The analysis focuses on findings that speak directly to learning support and integrity, and especially on how these domains interact. Older sources are used more selectively to frame constructs such as academic integrity and plagiarism, although even these definitions may need revision as generative AI alters the pedagogical landscape.

Following Kasneci et al. (2023), LLMs are understood as large neural language models that generate, transform, and evaluate natural language in response to prompts. Learning support refers to formative assistance that helps students plan, draft, revise, or understand work for which they remain responsible, rather than automated grading. Academic integrity is understood, following Bretag and Mahmud (2016), as a commitment to values such as honesty, trust, fairness, respect, and responsibility, and is operationalised as honest authorship and transparent use of sources and tools (Pagaling et al., 2024; Chan, 2025). Both Pagaling et al. and Chan show that these concepts become unstable when generative AI is involved. Pagaling et al. (2024) describe tensions between formal integrity rhetoric and everyday misconduct procedures, while Chan (2025) introduces AI-giarism to capture new ambiguities around plagiarism, authorship, and AI assistance. Earlier work on plagiarism and patchwriting already questioned simple copy versus originality binaries (Howard, 1995), so LLMs are best seen as intensifying long standing grey areas in how plagiarism, authorship, and legitimate support are defined and policed.

Across the selected literature, there is broad agreement that LLMs are neither inherently beneficial nor inherently harmful. Fluency, accessibility, and immediacy make them powerful learning tools, yet these same properties can blur authorship and enable undetectable ghost-written work if used uncritically. Bittle and El-Gayar (2025) describe generative AI as having a dual capacity, supporting learning while destabilising assumptions about assessment and detection. Ramírez Montoya and Lugo Ocando (2024) add that ethical and pedagogical frameworks risk becoming outdated even as they are implemented, so any balanced approach must be treated as provisional. Taken together, this literature suggests that the main risk is not that LLMs suddenly destroy academic integrity, but that institutions respond with technically weak, inequitable fixes instead of revising assessment and support.

**LLMs as learning support**

On the assistance side, three main points of convergence appear. First, several studies report that LLMs and AI writing tools can support surface level features of academic writing, especially for EFL and ESL learners. Xu and Zhou (2023) find perceived gains in coherence, vocabulary, and argumentation when students use AI writing tools. Tai et al. (2024) show that first year students who used ChatGPT for self and peer feedback achieved higher rubric scores in content, organisation, grammar, and vocabulary than control groups. Survey work by Scholars (2024) adds that regular users tend to view these tools as helpful for productivity and confidence, while non users are more likely to emphasise risks to standards.

Second, LLM based feedback can be both rubric aligned and scalable. Fan et al. (2024) report that human comments remain more accurate and better prioritised, but ChatGPT's feedback is closely aligned with marking criteria and delivered instantly. Dai et al. (2023) find no significant differences in learning gains between AI and human feedback, yet students welcome AI's detailed, rubric linked comments even when conceptual gains are modest. These studies support a view of LLMs as complements that widen access to formative feedback in crowded systems, rather than as replacements for human markers.

Third, LLMs can provide adaptive support beyond writing. They generate personalised explanations, summaries, and practice questions, and can respond in multiple languages (Kasneci et al., 2023; Shen, 2023). Ramírez Montoya and Lugo Ocando (2024) emphasise benefits for under represented and multilingual learners, who gain access to explanations that would otherwise require staff time and specialist language support.

These positive findings are tempered by disagreements over the depth and durability of learning gains. Almoraie and Alhejaili (2025) review more than one hundred studies and conclude that improvements in fluency and organisation are consistent, whereas evidence for gains in critical thinking or originality is weak. Tai et al. (2024) present relatively optimistic findings on sustained writing improvement, while Dai et al. (2023) find no advantage of AI feedback over human feedback. Current empirical work therefore does not yet justify common claims that LLMs enhance higher order skills; it mainly evidences short term gains in surface features and engagement.

A further tension concerns how students interpret support. Chan (2025) finds that students widely regard AI use for idea generation, paraphrasing, and language refinement as helpful, yet they are unsure when these practices cross into misconduct. Perceptions of acceptable help do not map neatly onto institutional policies. This is not a minor implementation detail; it shows that the same LLM output can be framed as formative feedback or as outsourced authorship, so learning support and integrity cannot be analysed in isolation.

**Integrity risks and AI-giarism**

On the risk side, three themes recur. LLMs make it easy to generate fluent, seemingly original text that can conceal a student's real contribution, blurring the line between support and ghost-writing (Almoraie and Alhejaili, 2025; Bittle and El-Gayar, 2025). At the same time, detection is unreliable, since AI text detectors generate both false positives and false negatives (Susnjak and McIntosh, 2024). Tools such as Turnitin struggle with AI assisted work and are sometimes treated as a proxy for integrity rather than one source of evidence (Bittle and El-Gayar, 2025; Pagaling et al., 2024), so a detection led strategy offers only limited control. Finally, LLM related risks intersect with pre-existing inequities, as misconduct referrals fall unevenly on certain groups (Pagaling et al., 2024), norms of plagiarism vary across cultures and disciplines (Chan, 2025), and text matching regimes can misread novice practices such as patchwriting (Howard, 1995). This increases the danger that some students are sanctioned for practices that others frame as legitimate support.

Within this shared risk narrative, there are disagreements about how novel these threats really are. Chan's (2025) concept of AI-giarism frames generative AI as complicating, but not completely transforming, longstanding issues of authorship and effort. In her account, AI-giarism is best treated as one category within academic misconduct rather than as a wholly separate phenomenon. By contrast, interpretations in Bittle and El-Gayar (2025) and Zawacki Richter and Marín (2024) sometimes imply that LLM driven misconduct is qualitatively different, given the speed, scale, and difficulty of detection. This disagreement shapes whether institutions prioritise incremental adaptation of existing policies, for example clearer disclosure rules and redesigned tasks, or more radical changes to assessment and integrity frameworks.

Assessment design is a further concern and a potential point of leverage. Susnjak and McIntosh (2024) show that GPT 4 with vision can solve complex multimodal exam questions using iterative, self reflective prompting, which casts doubt on the idea that LLM proof question formats will secure online exams. In response, Bittle and El-Gayar (2025) argue for rethinking how authenticity and effort are evidenced, including greater emphasis on process, drafts, and oral defence rather than reliance on final written products alone.

**Governance, literacy, and equity**

On governance, the literature converges on three broad points. First, blanket bans on LLMs are widely seen as impractical and counterproductive. Zawacki Richter and Marín (2024) argue that LLMs should be treated as supplements to human teaching, with assessments redesigned to foreground process and oral explanation. Second, policy and professional development are seen as lagging behind everyday classroom practice. Almoraie and Alhejaili (2025) note that teachers often recognise pedagogical benefits but lack clear frameworks and assessment redesigns, while Bittle and El-Gayar (2025) contend that institutional policies are reactive rather than strategic. Third, multiple authors emphasise that AI literacy and ethics education must sit at the heart of any sustainable response (Kasneci et al., 2023; Lund et al., 2026).

Within this broad convergence, several tensions remain. One concerns the relationship between policy awareness and behaviour. Lund et al. (2026) find that awareness of institutional AI policies does not significantly predict ethical judgments or use of generative AI, whereas students' own ethical beliefs do. This challenges approaches that treat policy dissemination as the primary lever of change and reinforces Chan's (2025) finding that students are often left to interpret grey areas themselves. Another tension lies between detection oriented and developmental approaches. Pagaling et al. (2024), drawing on the UK QAA charter and wider academic integrity scholarship (for example Bretag and Mahmud, 2016), argue that integrity should be understood as alignment with ethical and professional values, and that misconduct procedures must distinguish deliberate cheating from developmental issues in academic writing. A detection dominated response to LLMs collapses this distinction in practice and risks driving problematic use underground rather than encouraging disclosure and reflection.

Equity and diversity create further pressure points that cut across governance and pedagogy. Pagaling et al. (2024) call for audits of integrity systems to identify disproportionate referrals and outcomes, while Chan (2025) emphasises that language and sociocultural factors shape how students interpret both plagiarism and AI-giarism. Without such audits, institutions risk implementing AI policies that formally apply to all students but in practice fall hardest on those who are already marginalised.

Despite these tensions, there is strong agreement on the importance of AI literacy as a practical bridge between learning support and integrity. Kasneci et al. (2023) and Dai et al. (2023) call for explicit teaching of verification, critical reading, and proper citation of AI assisted work. Pagaling et al. (2024) advocate embedded, developmental integrity education that moves beyond punitive messaging. Chan (2025) argues that teaching should address both traditional plagiarism and AI-giarism together, so that students can transfer norms and recognise when AI use must be disclosed. Lund et al. (2026) suggest focusing on overreliance and the ethical status of AI assisted authorship rather than treating all AI uses as equivalent. Empirical evaluations of such interventions are, however, still rare, which limits institutions' ability to choose between competing models of ethical AI use.

**Gaps and research priorities**

Methodologically, the field is still at an early stage. Most studies are classroom interventions, perception surveys, or conceptual analyses, with limited longitudinal coverage. Quantitative work tends to report short term improvements in structure or fluency but rarely tracks independence, originality, or long-term dependence on AI tools. Yet these outcomes are central to debates about skill erosion.

Across the reviewed work, three priority gaps emerge. First, there is a need for longitudinal studies that track how sustained exposure to LLM based support influences not only performance but also independence, critical capacity, and patterns of AI reliance. Without such evidence, claims that LLMs either erode or enhance higher order skills remain largely speculative, and institutions are effectively designing policy blind to long-term skill effects.

Second, there is a lack of robust evaluations of AI literacy and ethics interventions. Authors commonly recommend integrated ethics and AI literacy education, yet there is little causal evidence on which designs change behaviour without simply shifting misconduct into harder to detect forms. Without comparative evaluations of approaches such as mandatory declarations, structured drafting with AI, or reflective commentaries, AI literacy policy risks remaining symbolic or even counterproductive.

Third, equity effects of LLM regulation are under explored. Pagaling et al. (2024) and Chan (2025) raise concerns about disproportionate referrals and cultural variation in norms, but systematic studies of how AI policies affect different groups over time are still missing. Without such work, institutions risk entrenching inequities under the banner of integrity, for example by allocating suspicion and sanction unevenly across language groups.

Overall, the selected literature portrays LLMs as amplifiers of existing strengths and weaknesses in higher education. On the assistance side, they help students generate ideas, organise arguments, refine language, and access instant feedback at a scale that human educators cannot match (Xu and Zhou, 2023; Tai et al., 2024). On the integrity side, they blur authorship, risk overreliance, and operate within systems that already struggle with conceptual clarity, equity, and stable ethical guidance (Pagaling et al., 2024; Chan, 2025). The most defensible path from current research is structured integration. This involves using LLMs for formative learning and feedback within transparent governance frameworks, requiring disclosure of AI assistance, embedding AI literacy that cultivates verification and ethical judgment, and auditing integrity processes for fairness as well as deterrence. Until stronger longitudinal and equity focused evidence emerges, the balance between assistance and integrity will remain a negotiated practice, driven less by the capabilities of LLMs than by institutional choices about pedagogy, assessment, ethics education, and support.

## References

Almoraie, A. and Alhejaili, M. (2025) 'A systematic literature review to implement large language model in education', *Education and Information Technologies* (online). Available at: https://link.springer.com/article/10.1007/s44217-025-00424-7 (Accessed 29 November 2025).

Bittle, K. and El-Gayar, O. (2025) 'Generative AI and academic integrity in higher education: A systematic review and research agenda', *Information*, 16(4), p. 296. doi:10.3390/info16040296. Available at: https://www.mdpi.com/2078-2489/16/4/296 (Accessed 6 December 2025).

Bretag, T. and Mahmud, S. (2016) 'A conceptual framework for implementing exemplary academic integrity policy in Australian higher education', in Bretag, T. (ed.) *Handbook of Academic Integrity*. Singapore: Springer, pp. 463-480. doi:10.1007/978-981-287-098-8_24.

Chan, C.K.Y. (2025) 'Students' perceptions of "AI-giarism": investigating changes in understandings of academic misconduct', *Education and Information Technologies*, 30, pp. 8087-8108. Available at: https://link.springer.com/article/10.1007/s10639-024-13151-7 (Accessed 6 December 2025).

Dai, H. *et al.* (2023) 'AI-generated feedback on writing: insights into efficacy and ENL student writing', *International Journal of Educational Technology in Higher Education*, 20(1), p. 58. Available at: https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-023-00425-2 (Accessed 25 November 2025).

Fan, X., Li, R. *et al.* (2024) 'Comparing the quality of human and ChatGPT feedback of students' writing', *Journal of Writing Research* (online). Available at: https://www.sciencedirect.com/science/article/pii/S0959475224000215 (Accessed 27 November 2025).

Howard, R.M. (1995) 'Plagiarisms, authorships, and the academic death penalty', *College English*, 57(7), pp. 788-806.

Kasneci, E. *et al.* (2023) 'ChatGPT for good? On opportunities and challenges of large language models for education', *Learning and Individual Differences*, 104, 102274. Available at: https://www.sciencedirect.com/science/article/pii/S1041608023000195 (Accessed 24 November 2025).

Lund, B., Mannuru, N.R., Teel, Z.A., Lee, T.H. and Ortega, N.J. (2026) 'Student perceptions of AI assisted writing and academic integrity, ethical concerns, academic misconduct, and use of generative AI in higher education', *AI in Education*, 1(1), 2. Available at: https://www.mdpi.com/3042-8130/1/1/2 (Accessed 6 December 2025).

Pagaling, H., Cochrane, A. and Philp, J. (2024) 'Academic integrity or academic misconduct? Conceptual and practical tensions in higher education', *Higher Education Research & Development*, 43(8), pp. 1564-1584. Available at: https://doi.org/10.1080/07294360.2024.2339833 (Accessed 6 December 2025).

Ramírez-Montoya, M.S. and Lugo-Ocando, D. (2024) 'The impact of large language models on higher education', *Frontiers in Education*, 9, 1392091. Available at: https://www.frontiersin.org/articles/10.3389/feduc.2024.1392091 (Accessed 28 November 2025).

Scholars (2024) 'Student perceptions of ChatGPT: benefits, costs, and attitudinal differences between users and non-users toward AI integration in higher education', *Education and Information Technologies* (online). Available at: https://link.springer.com/article/10.1007/s10639-025-13575-9 (Accessed 23 November 2025).

Shen, S. (2023) 'Application of large language models in the field of education', *International Journal of Emerging Technologies in Learning*, 18(4), pp. 24-35. Available at: https://www.researchgate.net/publication/380177498_Application_of_large_language_models_in_the_field_of_education (Accessed 25 November 2025).

Susnjak, T. and McIntosh, T.R. (2024) 'ChatGPT: The end of online exam integrity?', *Education Sciences*, 14(6), 656. Available at: https://doi.org/10.3390/educsci14060656 (Accessed 27 November 2025).

Tai, M. *et al.* (2024) 'Impact of ChatGPT on ESL students' academic writing skills', *Smart Learning Environments*, 11, 2. Available at: https://slejournal.springeropen.com/articles/10.1186/s40561-024-00295-9 (Accessed 29 November 2025).

Xu, J. and Zhou, Y. (2023) 'The impact of AI writing tools on the content and organisation of EFL student writing', *Cogent Education*, 10(1), 2236469. Available at: https://www.tandfonline.com/doi/full/10.1080/2331186X.2023.2236469 (Accessed 26 November 2025).

Zawacki-Richter, O. and Marín, V.I. (2024) 'ChatGPT in higher education: a synthesis of the literature and a bibliometric analysis', *Educational Technology Research and Development* (online). Available at: https://link.springer.com/article/10.1007/s10639-024-12723-x (Accessed 30 November 2025).