Physical Education, Health and Social Sciences
https://journal-of-social-education.org
E-ISSN: 2958-5996
P-ISSN: 2958-5988

Artificial Intelligence and Academic Integrity in UK Higher Education: A Rapid Literature Review of Emerging Challenges, Institutional Policy Responses, and Future Research Directions

Aimesh Shaheen[1], Aimen Sadheer[2]

[1] MSc International Management, Institute of Management, University of Greater Manchester England, Corresponding Author's Email: aimeshsatti.1@gmail.com
[2] MSc International Business, Institute of Management, University of Chester, England, Email: aimensadheer327@gmail.com

DOI: https://doi.org/10.63163/jpehss.v4i2.1400

Abstract

Background: Academic integrity has been challenged in a radically different way in the UK higher education sector by the rapid growth of generative artificial intelligence (GenAI) tools, underscored by the widespread use of large language models, including ChatGPT. Just three years after ChatGPT was released in the UK in the end of 2022, adoption by students has jumped from about 53% to 88%, presenting new systemic challenges, for which the current institutional frameworks were not planned and are not well-suited to deal with. The method used in carrying out this study is a rapid literature review (RLR) with a literature reporting approach based on PRISMA 2020. Four academic databases (Scopus, Web of Science, ERIC and the British Education Index) were searched as well as policy databases in the UK such as the British Education Index's (BEI) Higher Education publications archive, and Jisc's National Centre on AI (NCAI) database. The Quality Assurance Agency (QAA) was also used as a source for policy documents related to artificial intelligence for education. A total of 38 sources were identified for consideration after systematic screening and included two institutional policy analyses, six official sector documents, and 29 peer-reviewed journal articles. The synthesis highlighted four key thematic areas: (1) the unprecedented scale and definitional complexity of academic misconduct involving AI; (2) the technical and ethical limitations of academic integrity detection systems for AI; (3) the different nature and weak cohesion of policy responses to academic misconduct across the UK sector; and (4) the general consensus on redesigning and authentic assessment and enhancing AI literacy as the educationally soundest strategic directions. As reviewed above, the scholarly literature at hand questions the conceptual soundness of current academic integrity frameworks, which are developed within a pre-AI world that relies on individual authorship and empirical evidence assessable through the traceability of student work. Along with this, the paper reveals four areas of research that require immediate attention: Longitudinal effectiveness studies; UK specific empirical studies based on student and staff perspectives; Comparative institutional policy analysis; Theoretical research in the creation of new academic integrity frameworks that are sufficient to the AI era. The subject aims to embed a deeper understanding of generative AI within the framework of academic integrity and the concept of academic misconduct. The aims of the topic are to build a more robust understanding of generative AI in the context of academic integrity and academic misconduct.

Introduction

In November 2022, ChatGPT was made public by OpenAI, spawning an unprecedented upheaval worldwide in higher education, including in the United Kingdom (UK). Institutions faced technological innovation in the form of an "open access" item, in just a few months of its release, that created well-structured and contextually coherent academic text, but which few

existing plagiarism detection systems would have been able to spot (Wilson, 2025). Pace and scale of student adoption has surpassed all sensible precedents in the higher education sector. HEPI's 2025 Student Generative AI Survey has found the rate of UK university students reporting the use of generative AI (GAI) in assessed academic work has jumped from 53% to 88% in one year alone, describing the change as “almost unheard of” in the recent history of UK higher education (Higher Education Policy Institute [HEPI], 2025).

The commitment to the International Center for Academic Integrity (ICAI, 2021) values of honesty, trust, fairness, respect, responsibility and courage in scholarly work is at the center of the university's educative and social mission. Historically this has been supported through policies, commitment to honour codes, plagiarism detection software and assessment design that ensures submitted work was genuine effort to demonstrate the student's intellectual work (Cotton, Cotton and Shipway, 2023). Each of these mechanisms is at the same time challenged by generative AI, blurring the lines with pedagogic purposes, ethical dilemmas and technical aspects. The complexity of this challenge has been increasingly understood – but unevenly. Since 2022, a large volume of literature has developed, but it has been highly descriptive and is still largely focused on policy descriptions and non-UK institutional contexts, with critical empirical and theoretical questions still left unanswered (Atkinson-Toal & Guo, 2024; Francis et al., 2025; Wilson, 2025). Issues of what works, for whom, under what conditions, and if the question obtains an affirmative answer – should there be any policies framework that would be sufficient to meet the challenge before it grows? – so far require definitive answers rather than insights. This swift literature search addresses the gap. It is based upon a systematic selection of 38 sources that it collates evidence that is currently available on the field of AI and academic integrity in the UK context, giving a detailed but rigorous, timely and practically grounded overview of the state of the field, while critically interrogating assumptions in current policy responses, and pointing to gaps which warrant urgent attention by the scholarship community. The review's three contributions are original. It is the most systematic review and synthesis of AI-challenge evidence to date specific to the UK. Second, it takes forward a very important theoretical position on re-thinking academic integrity frameworks suitable for an era in which AI is sufficiently ubiquitous to shape education. Third, it outlines a planned future research agenda—one that is focused on clear gaps in the evidence base and not on hypothetical research interests.

Conceptual Background

Defining Academic Integrity in the AI Era

Academic Integrity, as a formal scholarly construct is best summarised from the ICAI (2021) through the six values set out below: Honesty, Trust, Fairness, Respect, Responsibility, Courage. In the UK, the Quality Assurance Agency for Higher Education (QAA) has brought forward the Academic Integrity Charter, which sets out standards for the sector and which Universities are invited to join as an institutional commitment to ensuring that those are upheld (QAA, 2023a). Academic dishonesty is defined by QAA (2023a) as 'any use which may give a student an unfair academic advantage', including any instance of academic cheating and academic plagiarism, and specifically including the generation of such content by AI tools when it is not authorised by the learning outcomes. The key issue not yet covered by the current definitions is the intrinsic confusion that generative AI brings to the notion of authorship itself. Academic integrity discourse has been set around precisely the one-to-one binary distinction between the student's own intellectual work and the work of others in the context of logics for detecting plagiarism and honour code enforcement. Generative AI is problematic in this binary on structural terms, because although it generates new text in the sense of no copying of existing text, it generates them using the work of millions of text generators - which were all humans - whose work constituted their training data (Selwyn, 2023). These philosophical and practical spaces of unintended by--products when determining what counts as an original student product, productive intellectual work, and legitimate assistance remain under-studied

within the literature (Sullivan, Kelly, and McLaughlan, 2023; Tate et al., 2023). By the end of this course, students should be able to explain the technical foundations of generative AI and Large Language Models (LLMs).

Generative AI describes AI systems that generates new content – such as text, images, code and audio – in response to prompts from a user based on pattern-matching from a large set of training data (Haase and Hanel, 2023). Generally, a large language model (LLM) is one type of generative AI that is trained using a massive amount of text data and is based on the transformer neural network architecture, which was first formally published by Vaswani et al. (2017), and is used to make and generate predictions in textual languages that are coherent and meaningful for the context. The dynamism of commercially released LLMs such as GPT-4, Google's Gemini, Meta's LLaMA and Anthropic's Claude have greatly broadened the scope of academic writing that students can now access through AI research tools (Bender et al., 2021; Michel-Villaréal et al., 2023). GenAI technology has already begun making an impact on the higher education sector, from its legitimate use to support learning to its potential use for unauthorized assessment completion, she notes, leaving a basic conundrum for policy makers which Kofinas (2025, p.1) has called the 'fundamental ambiguity'. This ambiguity is exacerbated by multimodal feature: Today's AI applications can produce text-based documents, but they are also capable of creating code, mathematical proof, data analysis and visuals – a range of assessment modalities that is widely beyond the classics of essays and ideally suited to be AI-susceptible (Perkins et al., 2024).

The Challenge of Detection: Technical and Ethical Dimensions

One of the key aspects of the academic integrity problem with AI is that existing detection methods are lacking. In contrast to the traditional plagiarism that entails copying and pasting content that can be identified with string-matching software, AI-generated writing is technically original because it hasn't been published anywhere in the available database. Instead, detection tools such as Turnitin's AI detection tool and GPTZero rely on statistical analysis of text features that have been seen as part of AI-generated content, such as low perplexity scores and high burstiness scores (Desaire et al., 2023; Turnitin, 2023). The ethical and legal dimensions of unreliable detection are equally significant. False positive rates of up to 14% mean that substantial numbers of students who have not used AI may face misconduct proceedings on the basis of probabilistic algorithmic judgements (Atkinson-Toal and Guo, 2024; Liang et al., 2023). In the UK context, this creates exposure to challenge under the Consumer Rights Act 2015 and raises serious questions about natural justice, particularly for international students and non-native English speakers whose writing characteristics may produce inflated AI probability scores (Cotton, Cotton, and Shipway, 2023).

Methodology

Review Design and Rationale

This study employs a rapid literature review (RLR) design. The RLR is an established approach in applied and policy-relevant research that synthesizes existing evidence efficiently while maintaining methodological transparency and credibility (Haby et al., 2024; Khangura et al., 2012). It is particularly appropriate for emerging topics where the evidence base is expanding rapidly and where decision-makers require timely syntheses. This review follows PRISMA 2020 reporting principles (Page et al., 2021) adapted for rapid review methodology, including systematic database searching, transparent screening procedures, and thematic synthesis of included sources. The RLR design was preferred over a systematic review in recognition of the exceptional pace of publication in this area: a full systematic review with dual independent screening conducted in mid-2026 would risk being superseded by significant literature before completion. The trade-offs of the RLR design, including the absence of meta-analysis and the use of single-reviewer screening verified through a random 20% double-check, are acknowledged as limitations and discussed in Section 5.

Search Strategy

Four databases were searched: Scopus, Web of Science, ERIC (Education Resources Information Centre), and the British Education Index. Supplementary searches were conducted of key UK policy repositories, including the QAA website, the Higher Education Policy Institute (HEPI), Jisc's National Centre for AI, and the Russell Group publications archive. The following Boolean search string was applied consistently across all databases:

> ("generative AI" OR "ChatGPT" OR "large language model" OR "artificial intelligence") AND ("academic integrity" OR "academic misconduct" OR "plagiarism" OR "assessment") AND ("higher education" OR "university" OR "UK" OR "United Kingdom")

The search was restricted to publications from January 2020 to May 2026. The 2020 start date was set to capture emergent pre-ChatGPT scholarship on AI and assessment alongside the post-2022 explosion of directly relevant literature.

Inclusion and Exclusion Criteria

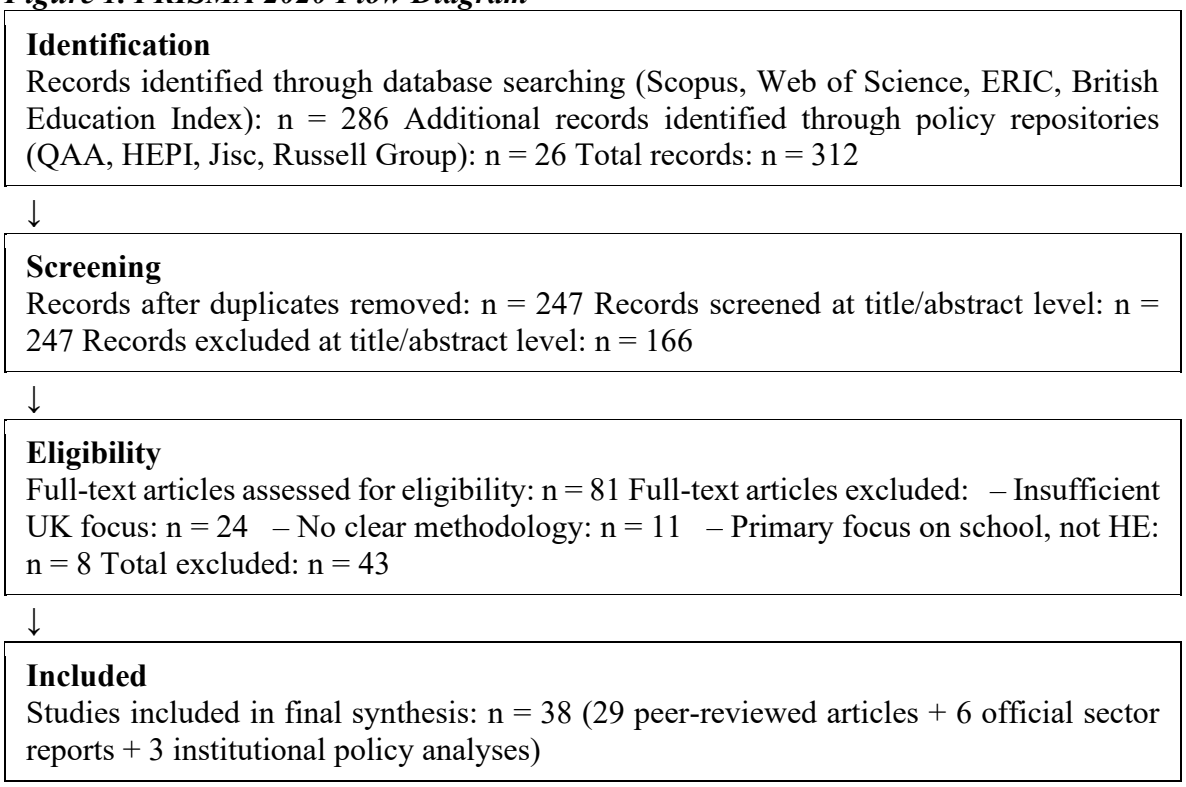
Table 1. Inclusion and Exclusion Criteria

Criterion	Include	Exclude
Date Range	2020–2026	Before 2020
Language	English	Non-English
Source Type	Peer-reviewed journals, official reports (QAA, HEPI, Jisc)	Blogs, opinion pieces without citation
Focus	AI and academic integrity in higher education	AI in schools or vocational training only
Geography	UK primary focus; international comparative studies included	Studies with no UK relevance
Methodology	Empirical, conceptual, or policy-analytical studies	Studies lacking clear methodology

Screening, Selection, and PRISMA Flow

An initial search returned 312 potentially relevant records (286 from databases, 26 from policy repositories). Following duplicate removal, 247 unique records were screened at the title and abstract level, yielding 81 for full-text assessment. Of these, 38 met all inclusion criteria and were included in the final synthesis. Reasons for exclusion at full-text stage were: insufficient UK focus (n = 24), no discernible methodology (n = 11), and primary focus on school-level rather than higher education (n = 8). The final sample of 38 sources comprises 29 peer-reviewed journal articles, 6 official sector reports (QAA, HEPI, Jisc, Russell Group), and 3 institutional policy analyses.

Figure 1. PRISMA 2020 Flow Diagram



Thematic Findings

Academic Integrity in Higher Education facing AI and Other Developments.4.3 Designating Individual Misconduct in AI and Other Issues. The volume and scale of the use of AI tools in student assessments in the UK is nothing short of alarming and is growing rapidly. As outlines in the HEPI Student Generative AI Survey (2025), in a single school year 35 percent of UK students have grown in their use of generative AI tools to complete assessed tasks, bringing the total to 88 percent. This uptake of good practice presents a 180° shift in the definition of academic integrity from the issue of the few academically dishonest students and tackling individual instances of academic dishonesty to a challenge on a national scale, in the majority of assessed student work across higher education in the UK. Similar trends of rapid increases in the use of AI tools by students have been documented in other countries: We see escalations even in other contexts for higher education students in the NorthAmerican and East Asian settings, as reported by Farhi et al. (2023) and Zhai (2022).

The concept of using AI to cheat in academic assignments is far more complicated than the simple distinction of allowed or disallowed implies. Kofinas (2025) comes up with a spectrum of rights and integrity concerns concerning AI: The full plagiarism of AI writing, the quasi plagiarism of using AI to help draft and edit work, as well as the legitimate use of AI as a tool for brainstorming, research assistance, and accessibility. This spectrum imposes deep challenges to institutional policy, because it makes it hard to clearly define the policy goals. Likewise, in "Chatting and Cheating", a widely cited study, researchers Cotton, Cotton and Shipway (2023) contend that neither the binary nature of AI as academic cheating nor AI as a legitimate learning tool is intellectually defensible or easy to enforce. Important equity aspects that AI inequities are not being resolved with institutional policies are identified by research, with a particular focus on those pertaining to students. The University of Bristol's AI Access for All initiative (September 2024) concluded that, prior to the introduction of universal provision of AI tools to students, there was a gap of 34 percentage points between the participation of Pupil Premium students and their peers from less socioeconomically privileged neighbourhoods, meaning that AI was having a disproportionate impact on students from more socioeconomically privileged backgrounds and compounding pre-existing attainment gaps.

This gap decreased to 11 percentage points (Stanley, 2025) after universal provision. The discovery poses basic equity concerns regarding whether or not institutions which prohibit or limit the usage of AI tools are likely to worsen educational inequity. This evidence resonates with the general literature on digital divides in higher education (Selwyn, 2022), and with the international focus on unequal levels of AI literacy between groups of students (Pedró, Subosa, Rivas and Valverde, 2019). Existing research has not sufficiently studied the motivational context of students' AI usage. Bretag et al. (2019) discuss in their seminal work "contract cheating" that academic dishonesty is not necessarily the act of morally compromised students, but rather a result of a range of complex factors from workload and assessment design to mental health issues and miscommunication around academic expectations. If we take into account this motivational prism when it comes to GenAI misconduct, then it could be argued that punitive detection-and-deterrence strategies are likely to be treating symptoms rather than the underlying causes of this issue (Newton, 2018; Sutherland-Smith, 2018).

Detection Mechanisms: Limitations and Controversies

The evidence base is challenging this assumption, as it has now been well documented that AI-generated text can be identified. In April of 2023 the company announced a new AI detection feature for Turnitin, the dominant plagiarism detection provider in their sector, stating that they were able to detect "AI-Generated Chest GPT and Similar" with "claimed accuracy of more than 98%" (Turnitin, 2023). However, results on independent evaluations have constantly revealed high error rates. Few or none of these assessment instruments are trained to measure authenticity, with eight of them tested by Liang et al. (2023) giving false positive rates as high as 14%, meaning that many students who are not using AI are likely to be flagged as such, with potentially serious consequences for wellbeing, institutional trust and the legal liability under the Consumer Rights Act 2015 imposed on them (Atkinson-Toal and Guo, 2024; Desaire et al., 2023). The rapid pace of GenAI advancements, the limitations of detection methods, and the ability to use deceptive methods compound these issues. The research work by Spennemann, Escherle and Budd (2024) shows that users are able to actively teach ChatGPT to write texts exclusively in a way that avoids AI detection solutions, making it fundamentally impossible in the long term to have an integrity strategy that is based on detection. Sadasivan et al. (2023)'s formal, theoretical analysis, however, shows that with improvements in language models, reliable AI-text detection might become impossible—a conclusion which has significant implications on the future of institutional over-investments in detection technology. Sullivan, Kelly, and McLaughlan (2023) have written that using detection technology is an "academic integrity compliance model" that does not address deeper causes for academic dishonesty, nor do they give an opportunity to build honest academic skills or ethics. This is currently which has already caught significant favour. In fact, UK institutions are specifically advised against using AI detection programs as the main tool to combat plagiarism by the QAA's publication, The Quality Compass (2024), which suggests a selection of complementary strategies, such as genuine tasks that tasks are being reworked to better address the challenges of plagiarism, and both staff and students are investing in learning about the use of AI. The importance of creating cultures of academic integrity that rely on values and relationships, not surveillance and enforcement, is echoed by Eaton (2022) and Stoesz and Eaton (2020).

Institutional Policy Frameworks: Divergence, Inconsistency, and Emerging Consensus

The UK higher education institutional policy environment has seen numerous and diverse policies emerge in quick succession with often conflicting responses to the AI integrity challenge. One of the most methodical and comprehensive UK university AI policies studies was carried out by Wilson (2025), who provided a broad range of examples of different positions taken by institutions regarding AI policy and the nature and process of communicating such a policy. Other colleges, including Oxford and Cambridge, implemented outright bans on the use of AI in assessed work; others, such as the University of Manchester, started off allowing the use of AI in assessed work, provided that AI outputs are treated as a

secondary source, with relevant citations and acknowledgements. In July 2023, the Russell Group issued principles on the use of generative AI in education, which was a notable effort to draw up lines of best practice across the research sector (Russell Group, 2023). The principles are focused on AI literacy, clear use and institutional accountability in communicating clear guidance to learners and are a much subtler stance than just outlawing it outright. The overall sectoral approach taken in the UK is one that is characterised as being "cautious optimism" according to Angelov (2025), while in some other jurisdictions it may be a more restricted regulatory approach. In contrast, Australian universities have enacted more rapid progress towards institutionalized frameworks of AI integration (Dawson, 2021; Perkins et al., 2024), and there is more variation in the progress of AI policy in US universities, with some having comprehensive policies, and others continuing to deliberate. (Fyfe, 2022; Lund et al., 2023)
In spite of the efforts at the sector level, there is still dis Синхрон between the institutional policies themselves. In Atkinson-Toal's (2024) study on generative AI education policies in UK Universities, the policies varied from highly detailed and nuanced policies to brief algebra with little or no practical guidance, while a large number of universities did not have a published policy at the time of this study. This lack of consistency can pose significant challenges to students, especially international students who might not be aware of how students behave elsewhere in the UK in a different academic setting, and may find themselves in a position of being alleged to be guilty of misconduct when they are not. It is this degree of incoherence in the policy that makes it institutional injustice, as outlined by Bretag (2019) which threatens the ‘integrity' it claims to defend. Early findings from a governance perspective show there is a lack of coherent national frameworks in quality assurance, which is not the case for the other aspects of quality assurance in UK higher education, to which the QAA's non-mandatory guidance is not yet able to offer a solution (Crossman, 2023).

Reimagining Assessment: Authentic Assessment and AI Literacy

Maybe the most interesting and forward-looking body of the growing literature involves not the identification or prevention of use of AI, but rather a reconception of assessment to make such use less pertinent, more transparent, or even more impactful educationally. To conclude, beyond the imminent threat of AI, there is in many ways an opportunity to review and redesign assessment systems which, before the advent of AI, have already been criticized as inauthentic, inequitable, and non-prepared for graduates' needs in the job market (Dawson and Sutherland-Smith, 2019), (Kofinas, 2025). The case for authentic assessment, assessment that is focused on authentic situations that actually pose authentic problems that demand the application of knowledge, contextual judgement, and reflective thinking, but not just replication of information, is significantly older than generative AI, but has come into its own and has a new analytical edge with the AI era (Corbin et al., 2025; Dawson et al., 2025; Liu, 2025; QAA, 2023b; Villarroel et al., 2018). Extended oral exam assessments, process portfolio that documents the various types of work that students complete, collaborative project-based assessments with visible evidences of contribution, and reflective Auto-ethnographic tasks that require personal experiences and situated perspectives are widely recognized as being potentially AI-resilient (Bearman and Luckin, 2020; Gamage, Pradeep and Wijesooriya, 2022). Perkins et al. (2024) create the Artificial Intelligence Assessment Scale (AIAS), a taxonomy of assessing tasks based on their susceptibility to AI and support of either planned, tolerated, or rejected use of AI. London Metropolitan University and Queen Margaret University Edinburgh are among the UK institutions who have taken up the AIAS as an applied solution to their projects and have adopted it as a practical solution to redesign assessment at module and programme level. It's a significant change in the thinking of it, as it breaks out of the binary language of allowable/not allowed and instead recognises the continuum of AI usage and the validity of using it as a thinking device in certain exams. There's also the notion of 'AI literacy' that has become a key aspect of institutional strategic responses, in addition to authentic assessment. The Internet Literacy & AI Literacy Skills Emporium is presented by David

Cheng, Adekola Clark, Albia Clark, and Cai Murrell.David Cheng, Adekola Clark, Albia Clark, and Cai Murrell are the presenters of this Internet Literacy & AI Literacy Skills Emporium. Both the QAA (2024) and the National Centre for AI in Jisc (2024) highlight structured AI literacy curricula for learners and educators as an essential prequisite for navigating the age of artificial intelligence. In direct response to the Russell Group principles, the University of Birmingham's framework explicitly connects to supporting students to be 'AI-literate', knowledge of how to use the available technologies 'effectively, ethically, critically and transparently' (Wilson, 2025, p.8). Both Ng et al (2021) and Miao, Holmes, Huang and Zhang (2021) offer international comparisons of AI literacy curriculum frameworks, which can be useful to inform the design of institutional programmes in the UK.

Discussion

Synthesis of Key Findings

The literature reviewed here reveals a sector confronting a genuinely transformational challenge to one of its foundational commitments. Four headline conclusions emerge from the synthesis. First, the scale of student AI use has already exceeded any reasonable expectation of universal prohibition. With 88% adoption among UK students, policies premised on AI-free assessment are increasingly detached from the reality of student behaviour and risk compelling widespread rule-breaking that delegitimises integrity frameworks as a whole. Second, detection-centred integrity strategies are technically unreliable, ethically problematic, and strategically inadequate. The literature reviewed here is remarkable in its consistency across methodological traditions on this point, from theoretical analyses of detection impossibility to empirical studies of false positive rates to ethnographic accounts of student experiences of algorithmic accusation. Third, institutional policy responses have been rapid but fragmented and inconsistent, creating inequity between students at different institutions and failing to provide the clear, principled guidance needed to support genuine integrity culture. Fourth, authentic assessment redesign and AI literacy development represent the most promising strategic responses, but implementation remains patchy, under-resourced, and insufficiently evaluated. The gap between the sophistication of the conceptual frameworks now available, such as the AIAS and the various AI literacy curricula described in the literature, and their actual adoption in routine institutional practice is striking and merits critical scrutiny.

Critical Analysis: The Conceptual Adequacy of Existing Frameworks

A more fundamental concern running through the literature, but not always surfaced explicitly, concerns the conceptual adequacy of existing academic integrity frameworks for the AI era. The dominant integrity paradigm, as codified in the ICAI's six values, the QAA's charter, and the sector's operational toolkit of plagiarism detection and misconduct proceedings, was developed to address a context in which the principal threat to integrity was deliberate deception: students passing off others' existing work as their own (Bretag, 2019; Eaton, 2022). Generative AI disrupts this framework in at least three analytically distinct ways. First, it challenges the premise of individual authorship: if AI can produce academically credible text that a student then submits, it is not obvious that this is usefully described as "cheating" in the same sense as copying another student's essay (Sullivan, Kelly, and McLaughlan, 2023). Second, it challenges the purpose of assessment: if knowledge can be generated on demand by a machine, what is the educational purpose of requiring students to generate it independently? Corbin, Dawson, and Liu (2025) argue that this question demands not a technical fix but a fundamental pedagogical rethink about what assessment is for. Third, it challenges the sustainability of deterrence-based enforcement: if the majority of students use AI and detection is unreliable, an enforcement-first response risks consuming institutional resources in unwinnable battles while poisoning the trust relationships that effective learning requires (Crossman, 2023; Newton, 2018). This review proposes that Sen's (1999) Capability Approach provides a productive theoretical alternative for reconceptualising academic integrity in the AI era. Sen's framework focuses not on what outputs individuals produce but on what they can

genuinely do and be, their real capabilities and freedoms. Applied to academic assessment, a capability-informed framework would ask not whether a student has produced a correct or even impressive piece of work, but whether the educational process has genuinely developed their capacity to think, reason, communicate, and act in their field. Assessment designed around capabilities, rather than outputs, is inherently more AI-resilient, because it is oriented toward what a student can do with knowledge, not merely whether they can produce a textual representation of it. This theoretical reorientation aligns with the authentic assessment literature reviewed in Section 4.4 and provides it with a more robust philosophical foundation.

Gaps in the Evidence Base

Several significant gaps in the current evidence base require urgent attention from the research community.

There is a near-complete absence of longitudinal evidence on the effectiveness of AI integrity policies. Almost all existing studies are cross-sectional, conducted at a single point in time in a rapidly changing environment. Without longitudinal data, it is impossible to assess whether particular policy approaches produce sustained changes in student behaviour and academic culture or merely short-term compliance that erodes once detection anxiety subsides.

UK-specific empirical research on student and staff experiences of AI integrity policy remains thin. The majority of empirical studies cited in this review draw on US, Australian, or broadly international samples. Given the distinctive characteristics of UK higher education, including the QAA regulatory framework, the Russell Group principles, the specific demographics of UK student populations, and the particular pressures of the tuition fee context, UK-specific evidence is urgently needed to inform policy.

The equity implications of AI policy remain underexplored. The evidence from Bristol's AI Access for All initiative suggests that differential access to AI tools may be reproducing and amplifying educational inequalities, yet this dimension is largely absent from institutional policy discussions. Research is needed that examines how AI integrity policies interact with existing patterns of disadvantage based on socioeconomic background, disability status, and international student status.

There is limited research on the perspectives and experiences of academic staff, the frontline implementers of integrity policy. Studies of how individual academics navigate AI policy ambiguities, what institutional support they receive, how their workload is affected, and how their own pedagogical values shape their responses to student AI use are largely absent from the literature reviewed here.

Finally, there is a lack of comparative research across different institutional policy models. Without systematic comparison of outcomes at institutions with different policy approaches, from permissive frameworks that embrace AI transparency to more restrictive prohibition models, it is impossible to identify empirically which approaches produce better educational and integrity outcomes under what conditions.

Conclusion and Future Research Agenda

This rapid literature review has synthesised evidence from 38 systematically selected sources to provide a comprehensive account of the challenge that generative artificial intelligence poses to academic integrity in UK higher education, the institutional responses it has generated, and the significant gaps in the current evidence base. The findings reveal a sector in a state of profound and accelerating transition, one that cannot be adequately addressed through the extension of existing detection and deterrence frameworks, however technologically sophisticated they become. The central argument of the emerging literature, which this review endorses and extends, is compelling: the arrival of generative AI in higher education is not primarily a misconduct problem to be policed but an educational challenge to be addressed through fundamentally reimagined approaches to assessment, learning, and the cultivation of student capability. Institutions that respond to AI primarily through detection, prohibition, and punishment risk expending substantial resources on an unwinnable technical arms race while

failing to develop the AI literacy skills, critical capacities, and ethical dispositions that graduates will need in every professional context they enter after graduation.

This review identifies four priority areas for future research. First, longitudinal studies of AI policy effectiveness in the UK context are urgently needed to establish which approaches produce sustained, equitable improvements in academic integrity culture rather than compliance theatre. Second, empirical research centring the perspectives and lived experiences of UK students, particularly those from disadvantaged and non-traditional backgrounds, is essential for developing responses that are both effective and just. Third, comparative analysis of different institutional policy models, from permissive to restrictive, is needed to identify the conditions under which different approaches produce better or worse educational and integrity outcomes. Fourth, theoretical work developing new frameworks for academic integrity adequate to the AI era, rather than adaptations of pre-AI paradigms, represents perhaps the most urgent intellectual challenge facing educational researchers in this field. The stakes of getting this right are high. Academic integrity is not a bureaucratic compliance requirement but a foundational commitment of the university as an institution, to honest inquiry, trustworthy knowledge, and the genuine intellectual development of its students. Preserving that commitment in an AI-saturated educational environment will require not less ambition but more: more creativity in assessment design, more investment in AI literacy, more equity in access to AI tools, and more rigour in evaluating what actually works. This review provides a foundation, and an urgent call, for that work.

Declarations

Competing Interests

The authors declare no competing interests.

Availability of Data and Materials

All data generated or analysed during this study are included in this published article. The literature synthesis draws on publicly available peer-reviewed publications and official sector reports, all cited in the reference list.

Authors' Contributions

RLR Aimesh Shaheen conceptualised the study, conducted the systematic search and screening, performed thematic synthesis, and drafted the manuscript. All authors read and approved the final manuscript.

References

Amigud, A. and Dawson, P. (2020) 'The law and the outlaw: Is contract cheating subverting the rule of law?', Assessment and Evaluation in Higher Education, 45(1), pp. 1–16. https://doi.org/10.1080/02602938.2019.1612851

Angelov, D. (2025) 'Developing academic integrity-compliant regulations and policies on the use of generative AI in higher education: Insights from the United Kingdom', in King, A.E. (ed.) Artificial Intelligence, Pedagogy and Academic Integrity. Cham: Springer, pp. 112–131. https://doi.org/10.1007/978-3-031-64454-8_7

Atkinson-Toal, A. and Guo, C. (2024) Generative Artificial Intelligence (AI) Education Policies of UK Universities. ResearchGate. Available at: https://doi.org/10.13140/RG.2.2.17896.68166 (Accessed: 10 May 2026).

Bearman, M. and Luckin, R. (2020) 'Preparing university assessment for a world with AI: Tasks for human intelligence', in Bearman, M. et al. (eds.) Re-imagining University Assessment in a Digital World. Cham: Springer, pp. 49–63. https://doi.org/10.1007/978-3-030-62387-5_4

Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. (2021) 'On the dangers of stochastic parrots: Can language models be too big?', Proceedings of the 2021 ACM

Conference on Fairness, Accountability, and Transparency (FAccT), pp. 610–623. https://doi.org/10.1145/3442188.3445922

Bretag, T. (2019) 'Contract cheating research: Implications for Canadian universities', University of Calgary, Educational Development Unit. Available at: https://tayabretag.com (Accessed: 12 May 2026).

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., Saddiqui, S. and van Haeringen, K. (2019) 'Contract cheating: A survey of Australian university students', Studies in Higher Education, 44(11), pp. 1837–1856. https://doi.org/10.1080/03075079.2018.1450898

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(1), pp. 1–25. https://doi.org/10.1186/s41239-023-00408-3

Cheng, M., Adekola, O., Albia, J. and Cai, S. (2021) 'Employability in higher education: A review of key stakeholders' perspectives', Higher Education Evaluation and Development, 16(1), pp. 16–31. https://doi.org/10.1108/HEED-03-2021-0025

Corbin, T., Dawson, P. and Liu, D. (2025) 'Assessment in the age of AI, universities must do more than tell students what not to do', The Conversation, 14 March. Available at: https://theconversation.com/assessment-in-the-age-of-ai-universities-must-do-more-than-tell-students-what-not-to-do-225413 (Accessed: 5 May 2026).

Cotton, D.R.E., Cotton, P.A. and Shipway, J.R. (2023) 'Chatting and cheating: Ensuring academic integrity in the era of ChatGPT', Innovations in Education and Teaching International, 61(2), pp. 228–239. https://doi.org/10.1080/14703297.2023.2190148

Crossman, K. (2023) 'Governance of artificial intelligence in education: A review of international policy responses', Policy Reviews in Higher Education, 7(2), pp. 134–158. https://doi.org/10.1080/23322969.2023.2205678

Dawson, P. (2021) Defending Assessment Security in a Digital World: Preventing E-Cheating and Supporting Academic Integrity in Higher Education. Abingdon: Routledge. https://doi.org/10.4324/9781003031420

Dawson, P. and Sutherland-Smith, W. (2019) 'Can markers detect contract cheating? Results from a pilot study', Assessment and Evaluation in Higher Education, 44(4), pp. 637–639. https://doi.org/10.1080/02602938.2018.1531077

Desaire, H., Chua, A.E., Isom, M., Jarosova, R. and Hua, D. (2023) 'Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools', Cell Reports Physical Science, 4(6), pp. 1–11. https://doi.org/10.1016/j.xcrp.2023.101426

Eaton, S.E. (2022) 'Plagiarism in the digital age', International Journal for Educational Integrity, 18(1), pp. 1–15. https://doi.org/10.1007/s40979-021-00094-6

Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F.F., Al-shami, S.A. and Slamene, R. (2023) 'Analyzing the students' views, usage, and impact of ChatGPT as a learning assistant: The case of higher education in UAE', Computers and Education: Artificial Intelligence, 5, pp. 1–14. https://doi.org/10.1016/j.caeai.2023.100180

Francis, R., Dennehy, D., O'Raghallaigh, P. and Helms, N. (2025) 'Generative AI in higher education: A systematic review of benefits, challenges, and implications for institutional policy', Computers and Education: Artificial Intelligence, 8, pp. 1–18. https://doi.org/10.1016/j.caeai.2025.100295

Fyfe, P. (2022) 'How to cheat on your final paper: Assigning AI for undergraduate writing courses', AI and Society, 38(4), pp. 1395–1405. https://doi.org/10.1007/s00146-022-01397-z

Gamage, D., Pradeep, R. and Wijesooriya, C. (2022) 'Digital assessments and the integrity of academic qualifications: A review', International Journal for Educational Integrity, 18(1), pp. 1–30. https://doi.org/10.1007/s40979-022-00101-y

Haase, J. and Hanel, P.H.P. (2023) 'Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity', arXiv preprint. Available at: https://arxiv.org/abs/2303.12003 (Accessed: 14 May 2026).

Haby, M.M., Bhaumik, S., Tran, T.N.T., Mittal, S., Fitzgerald, S.M. and Tyndall, J. (2024) 'What are the best methods for rapid reviews of the research evidence? A systematic review of reviews and primary studies', Research Synthesis Methods, 15(3), pp. 412–429. https://doi.org/10.1002/jrsm.1720

Higher Education Policy Institute (HEPI) (2025) Student Generative AI Survey 2025. Oxford: HEPI. Available at: https://www.hepi.ac.uk/2025/04/10/hepi-student-survey-2025/ (Accessed: 1 May 2026).

International Center for Academic Integrity (ICAI) (2021) The Fundamental Values of Academic Integrity. 3rd edn. Clemson, SC: ICAI. Available at: https://www.academicintegrity.org/the-fundamental-values-of-academic-integrity/ (Accessed: 2 May 2026).

Jisc National Centre for AI (2024) AI in Tertiary Education: Summary of Research on the Use of AI in Higher and Further Education. Bristol: Jisc. Available at: https://nationalcentreforai.jiscinvolve.org/wp/2024/01/15/ai-in-tertiary-education/ (Accessed: 3 May 2026).

Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J. and Moher, D. (2012) 'Evidence summaries: The evolution of a rapid review approach', Systematic Reviews, 1(10), pp. 1–9. https://doi.org/10.1186/2046-4053-1-10

Kofinas, A. (2025) 'The impact of generative AI on academic integrity of authentic assessments within a higher education context', British Journal of Educational Technology, 56(2), pp. 1–18. https://doi.org/10.1111/bjet.13502

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E. and Zou, J. (2023) 'GPT detectors are biased against non-native English writers', Patterns, 4(7), pp. 1–8. https://doi.org/10.1016/j.patter.2023.100779

Lund, B.D., Wang, T., Mannuru, N.R., Nie, B., Shimray, S. and Wang, Z. (2023) 'ChatGPT and a new academic reality: AI-written research papers and the ethics of the large language model publishing', Journal of the Association for Information Science and Technology, 74(5), pp. 570–581. https://doi.org/10.1002/asi.24750

Memarian, B. and Doleck, T. (2023) 'ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions', Computers in Human Behavior: Artificial Humans, 1(2), pp. 1–9. https://doi.org/10.1016/j.chbah.2023.100016

Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D.E., Thierry-Aguilera, R. and Gerardou, F.S. (2023) 'Challenges and opportunities of generative AI for higher education as explained by ChatGPT itself', Education Sciences, 13(9), pp. 1–19. https://doi.org/10.3390/educsci13090856

Miao, F., Holmes, W., Huang, R. and Zhang, H. (2021) AI and Education: Guidance for Policy-Makers. Paris: UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000376709 (Accessed: 4 May 2026).

Newton, P.M. (2018) 'How common is commercial contract cheating in higher education and is it increasing? A systematic review', Frontiers in Education, 3(67), pp. 1–18. https://doi.org/10.3389/feduc.2018.00067

Ng, D.T.K., Leung, J.K.L., Chu, S.K.W. and Qiao, M.S. (2021) 'Conceptualizing AI literacy: An exploratory review', Computers and Education: Artificial Intelligence, 2, pp. 1–10. https://doi.org/10.1016/j.caeai.2021.100041

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D. and Moher, D. (2021) 'The PRISMA 2020 statement: An updated guideline for reporting systematic reviews', BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Pedró, F., Subosa, M., Rivas, A. and Valverde, P. (2019) Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development. Paris: UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000366994 (Accessed: 5 May 2026).

Perkins, M., Furze, L., Roe, J. and MacVaugh, J. (2024) 'The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment', Journal of University Teaching and Learning Practice, 21(6), pp. 1–23. https://doi.org/10.53761/1.21.6.1

Quality Assurance Agency for Higher Education (QAA) (2023a) The Rise of Artificial Intelligence Software and Potential Risks for Academic Integrity: A QAA Briefing Paper for Higher Education Providers. Gloucester: QAA. Available at: https://www.qaa.ac.uk/docs/qaa/guidance/the-rise-of-artificial-intelligence-software.pdf (Accessed: 6 May 2026).

Quality Assurance Agency for Higher Education (QAA) (2023b) Reconsidering Assessment for the ChatGPT Era: QAA Advice on Developing Sustainable Assessment Strategies. Gloucester: QAA. Available at: https://www.qaa.ac.uk/docs/qaa/guidance/reconsidering-assessment-for-the-chatgpt-era.pdf (Accessed: 6 May 2026).

Quality Assurance Agency for Higher Education (QAA) (2024) Quality Compass: Generative AI and UK Higher Education. Gloucester: QAA. Available at: https://www.qaa.ac.uk/docs/qaa/guidance/quality-compass-generative-ai.pdf (Accessed: 6 May 2026).

Russell Group (2023) Russell Group Principles on the Use of Generative AI Tools in Education. London: Russell Group. Available at: https://russellgroup.ac.uk/news/new-principles-on-use-of-ai-in-education/ (Accessed: 7 May 2026).

Sadasivan, V.S., Kumar, A., Balasubramanian, S., Wang, W. and Feizi, S. (2023) 'Can AI-generated text be reliably detected?', arXiv preprint. Available at: https://arxiv.org/abs/2303.11156 (Accessed: 8 May 2026).

Selwyn, N. (2022) Digital Sociology: An Introduction. Cambridge: Polity Press.

Selwyn, N. (2023) 'The future of AI and education: Some cautionary notes', European Journal of Education, 58(4), pp. 1–11. https://doi.org/10.1111/ejed.12532

Sen, A. (1999) Development as Freedom. Oxford: Oxford University Press.

Spennemann, D.H.R., Escherle, H.A. and Budd, I. (2024) 'Asking ChatGPT to devise ways of circumventing AI detection: Implications for academic integrity', Trends in Higher Education, 3(1), pp. 45–62. https://doi.org/10.3390/higheredu3010004

Stanley, S.D. (2025) 'Teaching and learning reinvented: How AI is transforming assessment integrity in higher education', AI and UK Higher Education Thought-Leadership Series. Medium, 9 December. Available at: https://medium.com/@sds-ai-education (Accessed: 10 May 2026).

Stoesz, B.M. and Eaton, S.E. (2020) 'Academic integrity policies of publicly funded universities in western Canada', Educational Policy, 36(6), pp. 1529–1554. https://doi.org/10.1177/0895904820983032

Sullivan, M., Kelly, A. and McLaughlan, P. (2023) 'ChatGPT in higher education: Considerations for academic integrity and student learning', Journal of Applied Learning and Teaching, 6(1), pp. 1–10. https://doi.org/10.37074/jalt.2023.6.1.17

Sutherland-Smith, W. (2018) Plagiarism, the Internet and Student Learning: Improving Academic Integrity. 2nd edn. Abingdon: Routledge. https://doi.org/10.4324/9781351211819

Tate, T., Doroudi, S., Ritchie, D., Xu, Y. and Warschauer, M. (2023) 'Educational research and AI-generated writing: Confronting the coming tsunami', EdArXiv preprint. Available at: https://edarxiv.org/5erhf/ (Accessed: 11 May 2026).

Turnitin (2023) AI Writing Detection: How Turnitin's AI Detection Works. Oakland, CA: Turnitin LLC. Available at: https://www.turnitin.com/products/features/ai-writing-detection (Accessed: 12 May 2026).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) 'Attention is all you need', Advances in Neural Information Processing Systems, 30, pp. 1–11. Available at: https://arxiv.org/abs/1706.03762 (Accessed: 13 May 2026).

Villarroel, V., Boud, D., Bloxham, S., Bruna, D. and Bruna, C. (2018) 'Using principles of authentic assessment to redesign written examinations and tests', Innovations in Education and Teaching International, 57(1), pp. 38–49. https://doi.org/10.1080/14703297.2018.1564882

Wilson, T.D. (2025) 'The development of policies on generative artificial intelligence in UK universities', Journal of Documentation, 81(2), pp. 334–351. https://doi.org/10.1108/JD-07-2024-0141

Zhai, X. (2022) 'ChatGPT user experience: Implications for education', SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4312297