A Systematic Review of Generative AI and (English Medium Instruction) Higher Education

Peter Bannister  
Universidad Internacional de La Rioja (UNIR)  
mail: peter.bannister@unir.net  
ORCID: https://orcid.org/0000-0002-7216-3912

Alexandra Santamaría Urbieta  
Universidad Internacional de La Rioja (UNIR)  
mail: alexandra.santamaria@unir.net  
ORCID: https://orcid.org/0000-0003-0935-0616

Elena Alcalde Peñalver  
Universidad de Alcalá  
mail: e.alcalde@uah.es  
ORCID: https://orcid.org/0000-0002-1606-4792

ABSTRACT

This systematic review investigates the current state of research on Generative Artificial Intelligence (GenAI) and its implications for (EMI) Higher Education. The study employs a methodology based on an evidence-informed and theoretically credible framework to answer two research questions: (1) What studies of relevance to (EMI) Higher Education have been published thus far, considering the most recent developments of GenAI? and (2) Which key areas are currently lacking in extant literature and in need of further scholarly exploration in this regard in (EMI) Higher Education research? The results of the study reveal a limited number of pertinent publications, indicating a sparse scholarly landscape with a dearth of work on the implications of Generative AI in EMI Higher Education. Based on these findings, preliminary recommendations have been made to guide future research in this area. This study contributes to the literature by highlighting the need for further research on the potential of GenAI to enhance the teaching and learning experience in (EMI) Higher Education and provides a theoretical framework to guide future research. These findings may inform researchers and educators interested in exploring how GenAI may be leveraged from different educational perspectives.

Key words: English as a Medium of Instruction, generative artificial intelligence, Higher Education, qualitative research synthesis, systematic review.

Una revisión sistemática de la IA generativa y la educación superior (en inglés como medio de instrucción)

RESUMEN

Esta revisión sistemática examina la investigación más actual sobre inteligencia artificial (IA) generativa y sus implicaciones para la educación superior en contextos de inglés como medio de instrucción (EMI). El estudio aborda dos preguntas de investigación: (1) ¿Cuáles son los estudios relevantes para la educación superior (EMI) publicados hasta ahora, teniendo en cuenta los desarrollos más recientes de la IA generativa? y (2) ¿Qué áreas clave faltan por abordar actualmente en la literatura existente y necesitan una mayor exploración académica a este respecto en la investigación de la educación superior (EMI)? Los resultados revelan una falta de investigación sobre las implicaciones de la IA generativa en la educación superior, con un número limitado de trabajos relevantes. El estudio ofrece sugerencias preliminares para guiar la investigación futura en este campo. La revisión destaca la necesidad de más investigación sobre el potencial de la IA generativa para mejorar la experiencia de enseñanza y aprendizaje en contextos de educación superior en los que se emplea el inglés como medio de instrucción, y proporciona una base teórica para futuros estudios. Los resultados podrían ser de utilidad para académicos y educadores universitarios que busquen explorar el uso de la IA generativa desde diferentes perspectivas educativas.

Palabras clave: educación universitaria, inglés como medio de instrucción, inteligencia artificial generativa, revisión sistemática, síntesis de investigación cualitativa.
1. Introduction

The striking emanation of the generative artificial intelligence (GenAI) postchild, ChatGPT, has undoubtedly marked a transcendental historical occurrence on an unprecedented scale in modern times. The untold far-reaching ramifications of large language models (LLMs) are being subjected to embryonic and yet perpetual decipherment at a rate which pales in comparison to that of the prolific substantive advances in the field announced almost on a daily basis. Amidst a myriad of fledgling pavlovian and haphazard responses both on social media and in the international press, storm clouds have amassed on the future horizons of Higher Education, and the outlook would seemingly oscillate somewhere between impending doom and the apocalypse. Such stances may, in fact, be symptomatic of a humanity grappling to come to terms with the ricocheting debris protruding from the latest sledgehammer-like wallop taken to the ontological, epistemological, and axiological assumptions which have thus far enabled us to observe, comprehend, and interact with the world as we know it.

In the quicksand of the latest paradigmatic shift, a profound critical examination of every aspect of civilisation is not only an apposite by-product of such a process but is also a necessary rite of passage in order for a novel cultural transition in the strict anthropological definition of the term, to fully emerge from the ashes. International Higher Education (HE), therefore, has rightfully been highlighted as a domain susceptible to this phenomenon. Nonetheless, prior to endorsing or decrying the manifold proclamations of hyperbolic conclusions for all and sundry, and plotting a way forward, it is only right to adjourn momentarily to seek out evidence-informed scholarship and documented empirical expert consensus.

To narrow down the extremely broad and wide-ranging arena of operation, the authors set out to compose a bespoke systematic review of extant literature with a specific focus on English Medium Instruction (EMI) Assessment in Higher Education. According to Macaro (2018, p. 19), this refers to ‘the use of the English language to teach academic subjects (other than English itself) in countries or jurisdictions where the first language of the majority of the population is not English’. From the authors’ perspective, GenAI in EMI HE contexts may indeed offer a multitude of as-yet unexplored and continually increasing affordances for enhancing the teaching and learning experience, such as language practice, feedback and error correction, and bespoke learning materials creation.

Nevertheless, assessments have been previously described as ‘problematic’ in EMI HE contexts (Dearden, 2015, p. 17) without considering the further challenges posed by the emergence of GenAI. Therefore this context in particular continues to be undoubtedly a markedly vulnerable battle ground. This is owing to the potential impact LLMs may have on pertaining key pedagogical components, such as language proficiency evaluation, cross-cultural communication and academic writing skills development and enhancement, meaning that the linguistic and authorial implications of potential GenAI misuse represent an issue which is particularly acute to academic integrity in EMI HE (Lasagabaster, 2022). However, in practice, this was a fruitless pursuit as, at the time of writing, the authors were unable to locate neither scholarly literature nor expert consensus of such nature.

Therefore, this study aims to offer a comprehensive revision of research into GenAI and HE as is documented in scientific journal articles and research reports between 2022 and April 2023. Although the chronological limitations established seemingly go somewhat against the grain (Macaro, 2019), in light of the most ground-breaking recent developments within the field of GenAI that have taken place since 2022, marked by the launch of GPT 3.5 until the time of writing, this time period has been selected to focus on studies which specifically address these through the lens of EMI HE. Such an investigation in a distinct field or even into a different phenomenon may, in fact, seem to be implausible or perhaps redundant, given that international cognate studies from recent years have previously delved into this area of investigation (e.g., Chen et al., 2020; Chu et al., 2022; González-Calatayud et al., 2021; Ouyang et al., 2022; Zawacki-Richter et al., 2019). Nonetheless, even the most recent scholarly example has a scope which predates the monumental, rapid, and prolific singular developments in GenAI and HE that have occurred most recently.

Moreover, a correlation of this can also be found specifically within the field of EMI. This is a complex educational phenomenon in its own right and has generated a wealth of global research, as is documented in the seminal work of Macaro et al. (2018). However, even a timelier update (Macaro, 2022), whilst identifying a finite quantity of EMI assessment publications, notably neglects to signpost studies that deal with AI and EMI assessment.

Thus, in the spirit of these unprecedented advances since 2022 in the field, the authors have decided to impose quite particular chronological constraints in order to provide insight into the state-of-the-art in the present scholarly landscape and to tentatively offer a roadmap for further preliminary investigative exploration going forward.

2. Methodology

There is a multitude of established approaches to the research design of such studies varying from discipline to discipline (Miltra, 2023). Nevertheless, the authors of this study have opted to employ the evidence-informed and theoretically credible methodological framework crafted by Chong et al. (2022) using grounded theory (Charmaz, 1996). This decision is not solely motivated by the grounding of Chong et al.’s work in the field of Higher Education per se but also because the framework is both ‘principle driven’ and ‘captures the need to balance methodological heterogeneity and homogeneity’ (p. 10). Figure 1 below offers an overview of the research design architecture proposed by these authors:

![Figure 1. A Bottom-up Methodological Framework for Conducting Systematic Literature Reviews (Chong et al., 2022)](image)

The following lines of this section detail the bespoke application and adaptation of this methodological proposal.

2.1. Background Information

2.1.1 Research Questions

Building on earlier remarks with regard to the purpose, scope and remit of the present study and the theoretical definitory es-
establishment of EMI HE, the following research questions (RQs) were devised:

— RQ1: Considering the most recent developments of GenAI, what studies of relevance to (EMI) Higher Education have been published thus far?
— RQ2: Which are the key areas currently lacking in extant literature and in need of further scholarly exploration in this regard in (EMI) Higher Education research?

2.2. Search Strategy

2.2.1. Eligibility Criteria

As stated previously, this systematic review addresses publications focused on the use of GenAI in (EMI) HE, which have been published or have been made available as pre-print editions of texts in peer-reviewed journals between 2022, the year GPT 3.5 was launched, and April 2023. This decision, as previously stated, is due to the authors’ firm commitment to focus on the latest ground-breaking developments, which have taken place since 2022 and continue to occur up until the time of writing. Table 1 below has been populated with the full compendium of inclusion and exclusion criteria employed:

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published 2022- April 2023</td>
<td>Published before 2022</td>
</tr>
<tr>
<td>Written in English or Spanish</td>
<td>Not written in English or Spanish</td>
</tr>
<tr>
<td>Published in peer-reviewed journal</td>
<td>Not published in peer-reviewed journal</td>
</tr>
<tr>
<td>Has thematic focus of GenAI in (EMI) HE</td>
<td>Does not have thematic focus of GenAI in (EMI) HE</td>
</tr>
</tbody>
</table>

Note: Composed by the authors.

2.2.2. Consulted Sources

Taking into account the criteria outlined previously, the authors sought to carry out a search for eligible publications on electronic academic databases. Google Scholar, Scopus, Web of Science, and Education Resources Information Center (ERIC) were all accessed to this end, and articles of relevance were selected for analysis.

In accordance with the procedures established by Page et al. (2021), a total of 138 papers were yielded during the first stage of searching the aforementioned databases. Subsequently, on applying the inclusion and exclusion criteria, a total of 70 papers were excluded.

2.3. Screening and Selection

2.3.1. Data Collection

As per other cognate studies (Basilotta-Gómez-Pablos et al., 2022), the first step in the data collection process was to conduct a co-occurrence analysis to define the most frequent search terms and to subsequently increase reliability in the following publication data yielded (Galvez, 2018). The established terms used were “EMI”, “English as a Medium of Instruction”, “HE”, “Higher Education”, “GenAI”, “Generative Artificial Intelligence”, “AI”, and “Artificial Intelligence”. From these, the following search strings were used:

— (“EMI” OR “English as a Medium of Instruction”) AND (“HE” OR “Higher Education”) AND (“GenAI” OR “Generative Artificial Intelligence” OR “AI” OR “Artificial Intelligence”))
— (“HE” OR “Higher Education”) AND (“GenAI” OR “Generative Artificial Intelligence” OR “AI” OR “Artificial Intelligence”))

Furthermore, the search was carried out within all fields of potential publications. The articles were then assessed and chosen based on the aforementioned inclusion and exclusion criteria. A wide range of journals were identified, with those with the highest quantity of publications being International Journal of Educational Technology in Higher Education, Computers and Education: Artificial Intelligence, and Journal of University Teaching and Learning Practice.

2.3.2. Data Analysis

Two independent reviewers were then invited to evaluate the titles and abstracts of the studies selected. To guarantee a complete and fair evaluation process, these experts were carefully selected based on their experience, qualifications, and proven scholarly publication track record in the cognate research domain of Technology-Mediated Language Teaching and Learning. The chosen publications were then analysed through a comprehensive process of qualitative synthesis. In cases where discrepancies arose between the two reviewers, a third party was called upon to evaluate the submissions. To conduct the study analysis, we used the bespoke Rayyan tool, which facilitates the blind review process in order to reduce potential bias.

3. Results

Although exhaustive attempts were made to find publications that dealt with GenAI and any matter relating to EMI HE, locatable specific studies in this vein or in proximate fields such as CLIL HE were extremely scarce or non-existent at the time of writing. This is to be expected given the nature of the phenomenon of focus of the present study. Therefore, in order to obtain comparative insights and offer correlative findings with clear application on matters such as teaching and learning, assessment, and research, amongst others from the wider context in which EMI HE finds itself embedded, a number of studies have also been included which relate to the academic subjects in HE which form such a key component of the EMI educational phenomenon (Rose et al., 2021).

To this end, a synthesis of the main findings here highlights key themes and patterns that have been identified amongst the selected eligible studies together with recommendations for future avenues of scholarly exploration. In Table 2 below, an overall summary of the themes found together with the provenance of pertaining articles is provided and subsequently expanded upon in the ensuing sections:
of GenAI on the learner experience, with a prevailing notion of GenAI and student performance prediction (Jiao et al., 2022).

Several publications were identified that looked into student perceptions of GenAI, which included Abdelwahab et al. (2023), Banihashem et al. (2023), and Martin-Ramallal et al. (2022). Even though these studies are grounded in different contexts, i.e. business students, HE lecturers and students, and Spanish universities respectively, and approached their research objective in different ways, that is, through questionnaires, questionnaires and interviews and documentary analysis and questionnaires respectively, common themes manifested include student doubt towards HEI readiness and effectiveness to deal with present GenAI challenges and to successfully carry out a process of professionalisation to prepare them for AI-dominated labour market of tomorrow.

In addition to scholarly works which examine the potential impact on student attainment, such as Ramo et al. (2022), Saad and Tounkara (2023), and Urtasun (2022), other scholars have examined questions such as student motivation (Martin-Núñez et al., 2023), learner drop out (e.g., Bañeres et al., 2023; Nagy & Molontay, 2023), and student performance prediction (Jiao et al., 2022).

Works, such as those composed by Chen et al. (2023), Liu et al. (2022), and Odden et al. (2023), also examined the potential impact of GenAI on the learner experience, with a prevailing notion of GenAI.
nAI taking on the role of learning assistant as opposed to educator replacement found to be in common amongst findings.

3.2.3. The Role of GenAI in HE Assessment and Feedback

A number of publications have examined the way in which AI can be used to enhance test and assessment design (e.g., Yang, 2023; Yao, 2022; Yıldırım-Emilli & Bulut, 2023), and in a similar way, Chen (2022) explored its use for the evaluation of teaching in Mathematics HE. Scholars, such as Rahman (2022), have also investigated the efficacy of AI in HE online proctoring systems.

Marked concerns on the implications of GenAI apps and HE assessment were particularly underlined by Morreel et al. (2023), who reported that the answers given in a multiple-choice family medical exam were equivalent to a pass. This apprehension, in turn, was also echoed by Fergus et al. (2023), who sought to put ChatGPT to the test in answering the final assessments of modules from a Pharmaceutical HE programme. Cotton et al. (2023) highlighted the difficulties in AI-assisted misconduct detection, a task which Sadasivan et al. (2023) deemed to be theoretically and practically unreliable, considering the present software available.

Turning to the use of GenAI in feedback, Hooda et al. (2022) have penned an insightful review which emphasises the correlation between valid feedback immediacy and HE learning enhancement. Grounded in the field of pre-service teacher education, in their exploratory study, Sailer et al. (2023) examined the case for AI-assisted feedback in simulation-based learning.

3.2.4. GenAI Affordances for HE Research Activity

Green shoots in scholarly studies which methodologically employ GenAI are also beginning to be published, perhaps as an indication of a future direction of travel in this regard. Dahmen et al. (2023) offer an interesting overview of the potential affordances that GenAI use to this end may offer researchers. Both publications penned by Latif et al. (2022) and Subirats et al. (2023) are poignant, practical examples of this within the field of HE research. Whilst Farrokhnia et al. (2023), after conducting a fruitful SWOT analysis of ChatGPT, conclude with a valuable agenda of research recommendations to be investigated in HE.

3.2.5. GenAI, Academic Integrity, and HEI Policy

In his exploration of the implications of GenAI apps such as ChatGPT, King (2023) offers a further example of methodological innovation in the novel approach taken to address the question of academic integrity by reporting on responses from ChatGPT to his prompts on the matter of plagiarism. Perkins (2023) also explores student-used LLM-informed GenAI applications and academic integrity. Whilst his findings are in confirmatory concordance with those of King (2023) mentioned previously, he also highlights the lack of lecturer readiness and awareness for GenAI tool use detection. Lund et al. (2023), on the other hand, take a broader look at the area of GenAI in research authorship and publication whilst highlighting key ethical concerns.

On addressing ethics and GenAI, Crawford et al. (2023) highlight this as an additional area of substantial apprehension and make a potent call for global leadership to deal with a matter that has taken HE ‘by storm’ (p. 1). Graf and Bernadini (2023) echo ethical concerns of potential GenAI misuse within scientific research and call for greater discussion within the field on how best to deal with this. Zembylas (2023) identifies further issues and outlines a vision for how a decolonial approach may be implemented to address these. Highlighting further concerns, Henry and Oliver (2022, p. 330) address GenAI ethics through the prism of assessment proctoring and underline that ‘the ethics of using AI in education are political, involving the distribution of power, privilege and resources’. Both Koo (2023) and Masters (2023) emphasise the benefits of GenAI in Health Professions HE and offer insights for its ethical use in the field.

Finally, as HEIs begin to formulate and establish GenAI policy responses, Dwivedi et al. (2023) helpfully highlight the multidisciplinary complexity of the matter within different industries. Building on this, Fernández et al. (2023) report on findings from a multivocal literature review on the matter of the HEI status quo on digital transformation initiative uptake. Gellai (2022) offers insight into findings from an investigation which aimed to fill the void in institutional policy provision in British HE.

3.3. Future Research Recommendations

Addressing RQ2 and taking into consideration the current status quo within academia, the following section includes preliminary recommendations of areas for further scholarly exploration. This is by no means intended to be exhaustive nor prescriptive but merely an evidence-informed proposal to kickstart and shape much-needed investment on GenAI within the field of EMI HE.

3.3.1. GenAI and EMI HE Learning

— Student Attitudes, Perceptions, and Uses of GenAI
— Impact of GenAI on Student Agency and Self-efficacy
— GenAI as a Facilitator of Digital Literacy Enhancement
— Implications of GenAI on Student Cognitive Load and Working Memory
— GenAI as an Idea Generation Tool in EMI HE

3.3.2. GenAI and EMI HE Instruction

— The Role of GenAI in In-service and Pre-Service Teacher Development and Education
— GenAI for Pedagogical Content Creation
— Impact of GenAI on Academic Reading Instruction
— GenAI and Academic Writing Instruction
— The Role of GenAI in Academic Listening Instruction
— Use of GenAI in Presentation and Seminar Skills Development
— GenAI Efficacy in Critical Thinking Skills Enhancement

3.3.3. GenAI, Assessment and Feedback in EMI HE

— Educator GenAI Assessment Literacy
— AI-Assisted Academic Misconduct
— GenAI-Enhanced Assessment
— GenAI and Error Correction
— The Role of GenAI in In-House and Large-Scale Test Materials Development
— GenAI to improve Feedback Accuracy and Personalisation
— GenAI in Formative Peer Feedback
— AI-Facilitated Self-Assessment

3.3.4 GenAI in EMI HE Research Activity

— GenAI for Research Simulation in Complex Scenarios
— GenAI to Facilitate Data Collection
— The Role of GenAI to Enhance Data Analysis
— GenAI-Assisted Meta-Analysis
3.3.5. GenAI, EMI Academic Integrity and HEI Policy

— Ethical Considerations for AI-Assisted Authorship
— HEI Policy Development and Quality Assurance
— Stakeholder Training and Sensitisation
— Policy Infringement

The proposed future study areas include a thorough examination of the relationship between GenAI and EMI HE. These fields of study have major significance for furthering our knowledge of how GenAI might be effectively incorporated into EMI HE settings to improve teaching and learning experiences. Examining student attitudes, views, and usage of GenAI can give information on their acceptability and adoption, driving the creation of student-centred applications and support systems. Investigating the influence of GenAI on student agency, self-efficacy, and cognitive load might help to create instructional practices that empower students and improve their learning experience in this context.

Furthermore, research into the role of GenAI in teacher development, pedagogical content creation, and various aspects of instruction may help to design and develop effective GenAI-based tools and interventions tailored to the specific needs of EMI HE learners. In addition, investigating GenAI’s impact on assessment, feedback, research activity, and academic integrity can help to guide policy formulation, ethical concerns, and training programs to promote appropriate and successful GenAI adoption in EMI HE environments. Overall, these recommendations represent a roadmap for investigating the potential of GenAI in EMI HE, with the goal of improving all aspects of provision going forward.

4. Limitations

In such scholarly pursuits, there are a number of well-documented transdisciplinary limitations which ought to be addressed, e.g., those highlighted in full by Egger et al. (2008). From the authors’ perspective, whilst all such concerns may be applicable to the present study to a greater or lesser extent, publication lag is undoubtedly one of the principal shortcomings of note. The time associated with study conception, elaboration, and publication may be a limitation which means that the full extent of publications in elaboration and revision are generally not contemplated in studies of such nature. However, this issue is particularly acute in this instance as it starkly contrasts with the rapid and sweeping developments within the field of focus.

4.1. Broader Implications of Limitations Going Forward

Moreover, this juxtaposition arguably highlights a deeper underlying issue at play in academia in that, going forward, the possibility may exist of a potential paradigm shift in the advancement of society (Bornmann, 2013). However, this point should not be misconstrued with the somewhat hyperbolic speculation of cinematic proportions regarding AI takeover leading to a Post-Human society (Palatinitus, 2017). This apparent brewing impasse may perhaps be catalytically conducive towards a re-evaluation and realignment of often-maligned top-down research (Lubienski, 2019; Rycroft-Smith, 2022), which for Kris et al. (2021) does little to narrow the research-practice gap.

Addressing a research problem through scholarship may be characterised as a long-term developmental procedure that requires collective critical reflection, multifaceted dissent, and innovation in order for expert consensus to be confirmed (Elliot, 2022). This is undoubtedly a time-consuming endeavour to which international scholars can devote an entire academic career. However, the singular way in which seemingly constant ground-breaking technological developments in this particular area are occurring at breath-taking speed may, in fact, fundamentally challenge established research methods and praxis together with the role of the HE researcher at the apex of the hegemonic meaning and knowledge transfer cultural ecosystem (Dameri & Demartini, 2020).

5. Conclusion

This systematic review set out to provide insight into the state-of-the-art in the present scholarly landscape and to offer a roadmap for furtherly investigative exploration in GenAI and (EMI) HE. This review’s findings confirm that GenAI has the potential to dramatically alter all aspects of education in the (EMI) HE sector. Despite having yielded a limited number of publications, it would seem that GenAI may be beneficial in a number of educational settings, including language acquisition, content development, and evaluation. While GenAI research in (EMI) HE is still in its infancy, the results are encouraging and justify further scholarly examination. Previously, amidst the vast possibilities for future research, a compendium of preliminary avenues of investigation was presented, which, whilst by no means prescriptive in nature, does indeed intend to notionally signpost key domains to further enhance our theoretical and practical understanding of this highly complex phenomenon.

In sum, in these unprecedented times, there is undoubtedly a tremendous quantity of intellectual inquiry to be undertaken which certainly may be a daunting task from the offset. Nonetheless, this also represents an extraordinary solitary occurrence to leave no stone unturned in HE investigation. Thus, we now find ourselves at this fortuitous juncture where everything we have assumed to be unerring may be submitted to critical examination through the prism of the affordances and challenges posed by GenAI to determine its validity and legitimacy in this new light.

References


Emenike, M. E., & Emenike, B. U. (2023). Was This Title Generated by ChatGPT? Considerations for Artificial Intelligence Text-Generation Software Programs for Chemists and Chemistry Educators. Journal of Chemical Education. https://doi.org/10.1021/acs.jchemed.3c00087


