Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : **2591-801X**

Content Available at : **http://journals.sfu.ca/jalt/index.php/jalt/index**

Artificial intelligence in higher education. A protocol paper for a systematic literature review

Fadhil Ismail[A] — A — *Lecturer, Kaplan Singapore*

Eunice Tan[B] — B — *Head of Programme, Academic Division, Singapore Institute of Management, Singapore*

Jürgen Rudolph[C] — C — *Director of Research, Kaplan Singapore*

Joseph Crawford[E] — D — *Senior Lecturer in Management, University of Tasmania, Australia*

Shannon Tan[D] — E — *Research Executive, Kaplan Singapore*

Keywords

Artificial Intelligence (AI);
assessments;
chatbots;
ChatGPT;
conversational agents;
curriculum design;
generative AI;
GPT (Generative Pretrained Transformer);
higher education;
large language models (LLMs);
systematic literature review.

Correspondence

Fadhil.md@gmail.com [A]

Article Info

Received 18 September 2023
Received in revised form 23 October 2023
Accepted 23 October 2023
Available online 23 October 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.34

Abstract

Higher education continues to be confronted with significant learning and teaching challenges. Still reeling from the effects of the pandemic, the sector has grappled for the past year with the advent and impact of generative artificial intelligence (AI). Since the introduction of ChatGPT by OpenAI in November 2022, a growing number of studies have discussed AI models and their impacts and influence on higher education. However, the novelty of what we aim to do in a future paper, outlined in the current one, lies in the systematicity of our approach. There is yet to be a study in which a systematic search strategy is developed to critically review extant research longitudinally across all available generative AI chatbot models within higher education. This protocol paper identifies a prospective systematic approach to reviewing the emergent literature. In addition, this protocol paper documents the structural approach to facilitate a systematic literature review. We seek to offer a systematic approach to create an open-access resource to support future learning and teaching scholars to gain timely access to pre-examined literature on different forms of generative AI and their impact on higher education. This protocol paper, as such, offers an approach that can be used to initiate closer scrutiny of the metadata of articles published on AI models in higher education since its initiation in November 2022. We also suggest that the protocol presented in this paper be considered a relevant and rigorous approach for conducting systematic literature reviews in other domains.

Introduction

ChatGPT-3.5 became publicly available in November 2022 (Haleem et al., 2022). This iteration became known for its quick and comprehensive responses to queries in various domains. However, upon further review, it turned out that these responses, despite their refined articulations and structured presentations, were, in many parts, inaccurate and unsubstantiated (Kleesiek et al., 2023; Sallam, 2023). Subsequently, there have been accelerated developments, discourses and predictions offered about generative AI chatbots from various scholars and organisations within the teaching and learning spaces. The rise of ChatGPT and other competing generative AI models is expected to transform teaching and learning journeys in academia (Rudolph et al., 2023a, 2023b). AI chatbots have evolved from being topics in intellectual discussions to challenging realities confronting every higher education stakeholder, including universities, institutional policymakers, lecturers, curriculum and assessment developers, and students.

The rapid and transformational effect the COVID-19 pandemic had on higher education learning and teaching left scholars expected to innovate and respond to the emergent challenges of continuity of learning. Journals with an average CiteScore of 7.3 saw hundreds of article retractions for poor-quality research rushed to print (Taros et al., 2023). The challenges are still being uncovered today (e.g., Glushenkova & Zagato, 2023). While the human cost of artificial intelligence is likely lower than that of the COVID-19 pandemic, there is a need for clear and quality approaches to consider the rapid and transformative effects of AI, particularly generative AI, on higher education learning and teaching.

This protocol paper documents the method for creating a systematic literature review to facilitate an analysis of studies on generative AI and higher education. The main inspiration for this effort is to provide an open-access resource to support and facilitate academics and stakeholders in higher education to gain timely access to the research literature on generative AI and higher education. Using a rigorous approach, this protocol paper proposes a systematic approach to craft out and analyse the metadata of articles published on specific types of generative AI and higher education one year after their release to consider the impact it has on shaping the future of higher education. By providing an open-access database, we aim to facilitate future research in AI chatbots and their global impact on the higher education space.

The introduction of ChatGPT-3.5 in late 2022 instigated an array of research studies in relation to ChatGPT, generative AI and higher education being published in various higher education learning and teaching journals and databases (e.g., Adarkwah et al., 2023; Chaka, 2023; Crawford et al., 2023; Eager & Brunton, 2023; Firat, 2023; Gamage et al., 2023; Hassoulas et al., 2023; Ifelebuegu, 2023; Ifelebuegu et al., 2023; Kelly et al., 2023; Khademi, 2023; Limna et al., 2023; Mills et al., 2023; Mohammadkarimi, 2023; O'Dea & O'Dea, 2023; Perkins, 2023; Popenici, 2023; Popenici et al., 2023; Rasul et al., 2023; Rudolph et al., 2023a, 2023b; Calonge et al., 2023; Sullivan et al., 2023; Xames & Shefa, 2023). These publications, while numerous, only paint a partial picture of an area of rapidly growing knowledge.

Given the exponential growth of publications, it is valuable to take stock of the findings in these papers and their respective quality. More importantly, this will enable the synthesis of the findings to understand the functions and implications of AI applications (Rudolph et al., 2023b). Specifically, it will allow us to determine the opportunities and threats to learning and teaching in higher education. Rasul et al. (2023) highlighted that ChatGPT and other generative AI chatbots can potentially enhance and augment learning outcomes and experiences in higher education. However, there is a need to investigate its potential benefits and challenges to ensure its ethical, effective, and responsible use. With the increase of publications doing this, it becomes critical to not only synthesise the information through comprehensive literature reviews but to also conduct meta-analyses to understand the implications of this increase in academic literature in different contexts. The novelty of what we aim to do in a future paper, as outlined by our recommended approach, lies in the systematicity of our approach.

Since the introduction of ChatGPT-3.5 by OpenAI in November 2022, a growing number of studies have discussed AI models and their impacts and influence on higher education. Although there is an excellent paper by Tlili et al. (2023) on transparent AI literature reviews, there is yet to be a study in which a systematic search strategy is developed to review extant research longitudinally across all available generative AI chatbot models within higher education. Moreover, current publications on AI applications in relation to higher education still tend to be in their infancy. Efforts to establish coherence among these publications are rather disjointed and, more often than not, conducted at a granular level. The dearth of systematic and macro-level research inspired our research team (based in Australia and Singapore) to create a rigorous research protocol to examine research on AI applications and higher education. A rigorous systematic review is sound when designed and administered effectively and aligned to core research thematic dimensions within a structured methodology (Crawford & Cifuentes-Faura, 2022). Toward this end, we crafted the following research objective:

> To design a rigorous research protocol to curate and conduct a systematic literature review on the first year of published literature on AI applications (e.g., Bard, Bing Chat, ChatGPT, and Ernie) to support policymakers, educators, and researchers in higher education.

This protocol paper, as such, offers an approach that can be used to initiate closer scrutiny of the metadata and findings of articles published on AI applications in the higher education space, such as ChatGPT. Moreover, we suggest that the protocol presented in this paper be considered a relevant and rigorous approach for conducting systematic literature reviews in other domains as well.

Method

Systematic reviews aim to collate and synthesise the extant state of knowledge on a particular area of research via a systematic, structured analysis of aggregated findings from research outputs based on prespecified criteria (Higgins et al., 2011; Motyka, 2018). Research metrics are useful instruments to assess the quality and impact of research outputs (Moed & Halevi, 2015). However, it is important to note that each metric only measures a particular aspect and has limitations (Nestor et al., 2020). Therefore, it is critical not to consider any particular metric in isolation but instead to consider a series of metric measurements to evaluate the quality of the database or journal. Consequently, the databases chosen for this systematic review were selected based on known metrics such as Journal Impact Factor, h-index, g-index, Eigenfactor score, and Altmetrics (alternative metrics).

Search strategy

This protocol paper suggests a systematic approach for article selection guided by PRISMA – the Preferred Reporting Items for Systematic Reviews and Meta-analyses (Moher et al., 2009; Page et al., 2021). Specifically, it outlines the reporting recommendations for systematic reviews suggested in the PRISMA 2020 guidelines to reflect recent developments and protocol suggestions in systematic review methodologies (Page et al., 2021). This approach to a systematic review is commonly used in educational and sustainability research and has been previously described by Butler-Henderson et al. (2020a, 2021a) and explored by Bearman et al. (2012). Following PRISMA search guidelines, the proposed systematic review will conduct a database search of all published journal articles (including those published online first) and preprints that relate to the topic of generative AI and teaching and learning in higher education. Special consideration will be paid to preprint articles for quality and those articles that are yet to undergo peer review. All research outputs published between 30 November 2022 and 31 December 2023 in the following sources will be considered: (1) Academic Search Ultimate, IEEE Xplore, Informit Online, Ovid, Proquest, ScienceDirect, Scopus, and Web of Science; (2) Google Scholar (the first ten pages for each search string will be reviewed). A snowball reference analysis will also be conducted based on the articles extracted first.

A comprehensive and rigorous search strategy requires clearly aligning the search phrases (search terms, keywords and Boolean Operators) to the thematic dimensions relevant to the study's research objectives. Within the context of this study, Boolean Operators and key search terms were generated based on definitions and thematic dimensions derived from the preliminary literature search, which are congruent to the study's focus on generative AI in higher education. The search strings suggested to be searched in the title, abstract, or keywords in the proposed protocol include the following: For each search, the first core strings (higher education, artificial intelligence, and the focal artificial intelligence) will be paired with one of the other strings to complete five strings. Where possible, we relied on existing reviews that included one of these framed, noting some required adaption to this context. For example, Spelt et al. (2009) use interdisciplinary in all searches (e.g., interdisciplinary curriculum), and this review splits concepts to refer to them separately. For the focal AI, it could include reviews on diverse generative AI chatbots (e.g., ChatGPT, GPT-4, Bard, Bing Chat, Claude, or Ernie) and generative non-chatbot AI (e.g. DALL-E, GitHub Copilot, GPT-4 plugins, Midjourney, Runway, or Synthesia).

Table 1: Concepts, search strings and reviews guiding frames.

Concept	Search string	Review that guided this frame.
1. **Higher education**	"Higher education" OR university* OR college OR tertiary OR undergrad* OR graduate OR postgrad*	Butler-Henderson et al. (2022); Zawacki-Richter et al. (2019)
2. **Artificial intelligence**	"artificial intelligence" OR "machine intelligence" OR "intelligent support" OR "intelligent virtual reality" OR "chat bot*" OR "machine learning" OR "automated tutor" OR "personal tutor*" OR "intelligent agent*" OR "expert system" OR "neural network" OR "natural language processing"	Zawacki-Richter et al. (2019)
3. Focal artificial intelligence	ChatGPT* OR "Chat Generative Pre-trained Transformer"	*Use specific tool related text.*
4. **Learning Setting**	Curricul* OR learn* OR student*	Zawacki-Richter et al. (2019)
5. **Education policies**	Polic*	Aikens et al. (2016)
6. **Assessment**	Assess*	Struyven et al. (2005)
7. **Teachers and lecturers**	Teach* OR Lectur*	Spelt et al. (2009)
8. **Pedagogical Approaches**	Pegagog*	Spelt et al. (2009)

Eligibility criteria, selection procedure, and quality assessment

In the search, only English-language academic journals and pre-prints are planned to be included, with a time-based limit of 12 months following release. Given ChatGPT was released towards the end of 2022, the first planned review will include papers up until 31 December 2023. The eligibility criteria for inclusion are as follows: articles related to teaching, curriculum, education, and students, including assessments, teaching practice, and course design in higher education, and relate to the specific artificial intelligence tool of reference. Excluded articles included those concerning university administrative processes unrelated to teaching and learning. In the instance where articles were about students but not related to teaching or learning, they were also excluded. For example, if the article discusses an issue that does not include any connection to higher education, that study will be excluded.

A double screening procedure will be adopted in the planned systematic review during the verification process across initial title and abstract screening and full-text screening to determine the final selection of sources of evidence for analysis. An appropriate reliability check (e.g., Cohen's Kappa) will be conducted with at least fair agreement between all pairs required prior to progression. The quality of the evidence gathered in the planned systematic review will be evaluated using the Cochrane Collaboration's tool for risk of bias assessment (Higgins et al., 2011; Page et al., 2021; Zeng et al., 2020) to minimise bias. The flow of information through this systematic review and aggregated findings based on the prespecified criteria will be subsequently reported through a PRISMA Statement (Figure 1). The quality assessment tool and PRISMA Checklist to appraise the study validity are discussed in the next section.

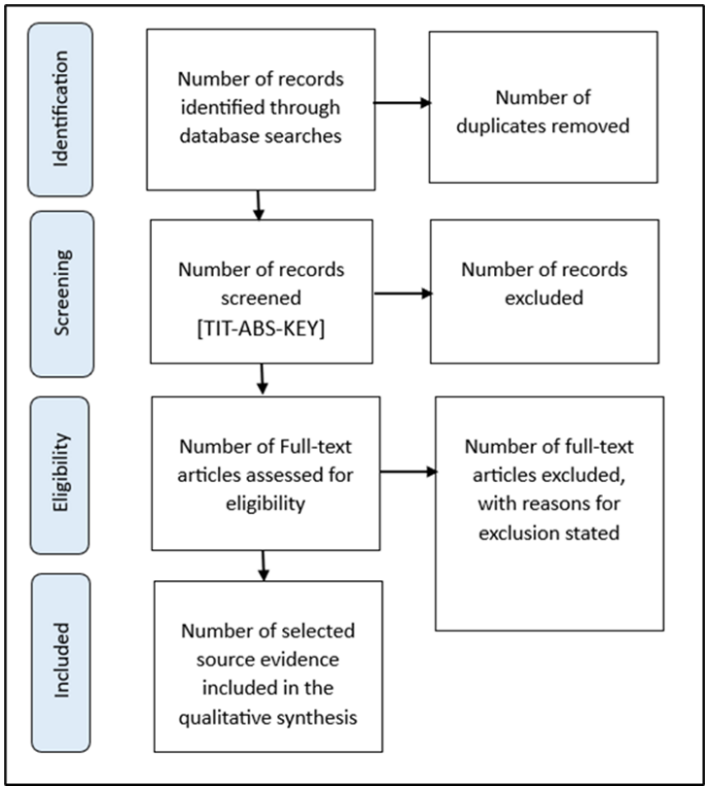

Figure 1: Proposed PRISMA statement.

Study validity assessment

We will use the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist and critical appraisal tools suited to the methods of the included studies to appraise and critically assess the validity of studies. The PRISMA checklist is a document that guides reporting systematic reviews and meta-analyses clearly and transparently. It ensures that the systematic review is written comprehensively and transparently so that readers can assess the quality and validity of the evaluation (PRISMA, 2020).

Critical Appraisal Tool

A critical appraisal tool assesses the dependability, significance, and practical relevance of evidence (Centre for Evidence-Based Medicine, 2023). In cases where any of these inquiries yields a negative response, it may be judicious to abstain from further engagement with the material (see Crowe & Sheppard, 2011; Zeng et al., 2015). An appropriate critical appraisal tool (or suite) will be selected and used for manuscripts included in the review based on the methods of included manuscripts.

Data coding and extraction strategy

Our data coding and extraction strategy includes the production of a detailed spreadsheet that will be available as an open-access database for scholarly reuse (similar to Butler-Henderson et al., 2021a). In constructing the database, we will incorporate certain theoretical assumptions detailed in Table 1. These are shared to present our reflexivity as researchers and to help others understand the adaptability of the data for their respective contexts. Although many data elements are clear and can be readily used in future research (like DOI, journal metadata, and country of origin), others, like the quality assessment score, study type, and participant type, necessitate further explanation.

The discipline and sub-discipline categories require some elaboration. The discipline category is grouped in four ways: health science, humanities and social science, STEM (science, technology, engineering and mathematics), and 'others'. In addition, we categorise the subdisciplines in the same way the researchers did -- for example, 'chemistry' or 'journalism' (see Butler-Henderson et al., 2020a).

The type of study is defined as theoretical, quantitative, qualitative, or mixed methods. Should the method remain unclear, the field will be blank. For participants, possible categories are academic, professional or management, undergraduate student, postgraduate student, and doctoral student, mixed staff (including two or more categories of staff), mixed students (including two or more categories of students), and mixed staff and students (for samples comprising both students and staff) (see Butler-Henderson et al., 2020a).

Table 2: Description of data elements.

No.	Data element	Field type	Description
1	Year of publication	Numeric	the year of the manuscript publication
2	Month of publication	Numeric	the month of the manuscript publication
3	First author's last name	Alphabetic	last name of first author of the study
4	Generative chatbots written about	Alphabetic	the names of the generative chatbots that the study focuses on (e.g. ChatGPT-3.5, GPT-4, Bard, Bing Chat, Claude, Ernie)
5	Non-chatbot generative AI	Alphabetic	the names of the generative non-chatbot AI that the study focuses on (e.g. DALL-E, GitHub Copilot, GPT-4 plugins, Midjourney, Runway, Synthesia)
6	Quality assessment score	Numeric	score derived from the quality tool (QATTL) used for the study: poor, low, medium, or high
7	Country of origin	Alphabetic	the country or countries of origin of the authors
8	Article type	Alphabetic	type of article (e.g. case study, commentary, research study, review article, opinion piece, etc.)
9	Type of study	Alphabetic	type of study (qualitative, quantitative, mixed methods, or theoretical)
10	Participant type	Alphabetic	type of participants (academic, professional or management, undergraduate student, postgraduate student, and doctoral student, mixed staff (including two or more categories of staff), mixed students (including two or more categories of students), and mixed staff and students (for samples comprising both students and staff)

11	Discipline	Alphabetic	broad higher education discipline groupings (e.g. health science, humanities, STEM)
12	Sub-discipline	Alphabetic	the higher education sub-discipline if available (e.g. chemistry, dentistry, teacher education)
13	Authors (listing all authors)	Alphabetic	names of all authors of study (APA7 style)
14	Title	Alphabetic	study title (APA7 style)
15	Journal	Alphabetic	name of journal where study is published
16	Volume	Numeric	volume number of journal where study is published
17	Issue	Numeric	issue number of journal where study is published
18	Pages	Numeric	page numbers in journal where study is published
19	DOI	URL	digital object identifier of the study

To test the repeatability of our process, the description of the above data elements will be executed with different researchers. The outcomes from each repetition will be recorded and compared for consistency using the metrics described in Table 1. In instances of missing or unclear data, authors will be contacted via email. They will be given a window of 14 calendar days to provide the necessary clarifications. Once received, the clarified data will be cross-referenced with the original submission to ensure consistency.

To ensure intercoder reliability, all coders will undergo standardised training using Table 1 as a shared coding manual. Their outputs will be periodically cross-checked against one another to assess consistency. Reliability will be statistically measured using Cohen's Kappa, with a threshold set at 0.80 (Warrens, 2015). Should reliability metrics fall below this, coders will undergo refresher training sessions, and the problematic data will be recoded.

Data synthesis and presentation

The process of writing a narrative synthesis can be particularly challenging, especially if the review includes a large number of different types of studies. Approaches like Braun and Clarke's (2006) thematic analysis can be helpful in developing initial themes for presentation. Complexities can also arise from examining a variety of complex interventions and outcomes. However, it is important to note that adopting a systematic approach to synthesis is key to making sense of the results in these different studies. The research question that underpins the review will determine the type of approach chosen to synthesise and present the findings of the review. The process of synthesising the data must also be rigorous and transparent, completely aligned with the methods specified in the protocol. These methods should be justified and followed systematically.

A narrative synthesis can be a useful first step in analysing and organising the data extracted from the review systematically and presenting the data in a coherent structure that can inform readers (Popay et al., 2006). Synthesising the results of different studies in a review in a narrative form is not simply describing or summarising the main features of each study, although doing this can be a useful start when writing systematic literature reviews. It would be practical, for example, to describe and comment on the methodological quality of each study that may provide significant insights for readers to become acquainted with the data presented in them.

One way to approach a narrative synthesis is by combining and evaluating data from several different studies. This step is taken to draw insights and conclusions about outcomes, effects, limitations of the studies and the applicability of findings in these studies. A narrative synthesis includes an examination of the similarities and differences between the findings of different studies, as well as an exploration of patterns in the data presented in these studies. For example, the similarities and differences in study design, populations, interventions or other aspects of the study can be examined and presented. This could include examining related factors and associations between research study designs and the findings. Some examples of synthesising the findings from different studies could be comparing the different research designs (e.g., RCT or mixed-methods approach) or with possible explanations to account for the pattern of results. Another way of organising narrative synthesis could also be to look at the different interventions or implementation strategies in the studies. This might involve examining associations between their research purpose, the manner in which the findings will be applied, and any other factors influencing the design and conduct of the research study. Studies with incomplete or missing information or an ambiguous description of the data will not be included in this synthesis. In addition, quantitative information in the form of tables, graphs, and figures will be summarised and presented in table form in the narrative synthesis to compare the different findings of the literature examined.

The next steps

This protocol has designed and outlined a rigorous systematic review method to ensure the maximal utility of the information and metadata in the databases mentioned in this paper. This includes an approach that can be applied to conduct an extensive search across the literature, databases, and online sources to ensure coverage of publications for the curation of a database. This approach resource will be critical in supporting researchers, educators in higher education, curriculum designers, assessors and policymakers in learning and teaching and providing them with a guide to navigate the AI space. We aim to update this approach longitudinally, with additional time periods to refine and revise coding rules and include other relevant databases over the coming years to make this resource robust and relevant. This will provide us with an impetus to mitigate and manage the impacts of AI and other EdTech technologies. It will facilitate the shift for the global higher education community towards new insights in learning and teaching as we navigate an evolving realm challenged and transformed by AI applications.

References

Adarkwah, M. A., Amponsah, S., van Wyk, M. M., Huang, R., Tlili, A., Shehata, B., Metwally, A. H. S., & Wang, H. (2023). Awareness and acceptance of ChatGPT as a generative conversational AI for transforming education by Ghanaian academics: A two-phase study. *Journal of Applied Learning and Teaching, 6*(2), 1-16. https://doi.org/10.37074/jalt.2023.6.2.26

Aikens, K., McKenzie, M., & Vaughter, P. (2016). Environmental and sustainability education policy research: A systematic review of methodological and thematic trends. *Environmental Education Research, 22*(3), 333-359. https://doi.org/10.1080/13504622.2015.1135418

Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development, 31*(5), 625-640. https://doi.org/10.1080/07294360.2012.702735

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Butler-Henderson, K., Crawford, J., Rudolph, J., Lalani, K., K. M. Sabu. (2020a). COVID-19 in Higher Education Literature Database (CHELD V1): An open access systematic literature review database with coding rules. *Journal of Applied Learning & Teaching, 3*(2), 11-16. https://doi.org/10.37074/jalt.2020.3.2.11.

Butler-Henderson, K., Crawford, J., Rudolph, J., Lalani, K. & Sahu, K. M, (2020b). COVID-19 in Higher Education Literature Database (CHELD V1). *Institute of Research Innovation*. https://doi.org/10.37074/jalt.2020.3.2.11d

Butler-Henderson, K., Tan, S., Lalani, K., Mandapam, S. K., Kemp, T., Rudolph, J., & Crawford, J. (2021a). Update of the COVID-19 Higher Education literature database (CHELD v2). *Journal of Applied Learning & Teaching, 4*(1), 134-137. https://doi.org/10.37074/jalt.2021.4.1.22

Butler-Henderson, K., Tan, S., Lalani, K., Karakka Mandapam, S., Kemp, T., Rudolph, J., Crawford, J. (2021b). COVID-19 in Higher Education Literature Database (CHELD). Version 2. *Institute of Research Innovation*. https://doi.org/10.37074/jalt.2021.4.1.22d

Calonge, D. S., Smail, L., & Kamalov, F. (2023). Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics. *Journal of Applied Learning and Teaching, 6*(2), 1-15. https://doi.org/10.37074/jalt.2023.6.2.20

Centre for Evidence-Based Medicine. (CEBM). (2023). *Critical appraisal tools*. University of Oxford. https://www.cebm.ox.ac.uk/resources/ebm-tools/critical-appraisal-tools

Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching, 6*(2), 1-11. https://doi.org/10.37074/jalt.2023.6.2.12

Crawford, J., & Cifuentes-Faura, J. (2022). Sustainability in higher education during the COVID-19 pandemic: A systematic review. *Sustainability, 14*(3), 1879. https://doi.org/10.3390/su14031879

Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice, 20*(3), 02. https://doi.org/10.53761/1.20.3.02

Crowe, M., & Sheppard, L. (2011). A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. *Journal of Clinical Epidemiology, 64*(1), 79-89. https://doi.org/10.1016/j.jclinepi.2010.02.008

Eager, B., & Brunton, R. (2023). Prompting higher education towards AI-augmented teaching and learning practice. *Journal of University Teaching & Learning Practice, 20*(5), 02. https://doi.org/10.53761/1.20.5.02

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching, 6*(1), 57-63. https://doi.org/10.37074/jalt.2023.6.1.22

Gamage, K. A., Dehideniya, S. C., Xu, Z., & Tang, X. (2023). ChatGPT and higher education assessments: More opportunities than concerns?. *Journal of Applied Learning and Teaching, 6*(2), 1-12. https://doi.org/10.37074/jalt.2023.6.2.32

Glushenkova, M., & Zagato, M. (2023). Effect of COVID-19 on digitalisation of higher education. A tale of one business school. *Journal of University Teaching & Learning Practice, 20*(6). https://doi.org/10.53761/1.20.6.10

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil transactions on Benchmarks, Standards and Evaluations, 2*(4), 100089. https://doi.org/10.1016/j.tbench.2023.100089

Hassoulas, A., Powell, N., Roberts, L., Umla-Runge, K., Gray, L., & Coffey, M. (2023). Investigating marker accuracy in differentiating between university scripts written by students and those produced using ChatGPT. *Journal of Applied Learning & Teaching, 6*(2), 1-7. https://doi.org/10.37074/jalt.2023.6.2.13

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., ... & Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ, 343*, 1-9. https://doi.org/10.1136/bmj.d5928

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning and Teaching, 6*(2), 1-8. https://doi.org/10.37074/jalt.2023.6.2.2

Ifelebuegu, A. O., Kulume, P., & Cherukut, P. (2023). Chatbots

and AI in Education (AIEd) tools: The good, the bad, and the ugly. *Journal of Applied Learning and Teaching, 6*(2), 1-14. https://doi.org/10.37074/jalt.2023.6.2.29

Kelly, A., Sullivan, M., & Strampel, K. (2023). Generative artificial intelligence: University student awareness, experience, and confidence in use across disciplines. *Journal of University Teaching & Learning Practice, 20*(6), 12. https://doi.org/10.53761/1.20.6.12

Khademi, A. (2023). Can ChatGPT and bard generate aligned assessment items? A reliability analysis against human performance. *Journal of Applied Learning & Teaching, 6*(1), 75-80. https://doi.org/10.37074/jalt.2023.6.1.28

Kleesiek, J., Wu, Y., Stiglic, G., Egger, J., & Bian, J. (2023). An opinion on ChatGPT in health care—written by humans only. *Journal of Nuclear Medicine, 64*(5), 701-703. https://doi.org/10.2967/jnumed.123.265687

Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching, 6*(1), 64-74. https://doi.org/10.37074/jalt.2023.6.1.32

Mills, A., Bali, M., & Eaton, L. (2023). How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *Journal of Applied Learning and Teaching, 6*(1), 16-30. https://doi.org/10.37074/jalt.2023.6.1.34

Moed, H. F., & Halevi, G. (2015). Multidimensional assessment of scholarly research impact. *Journal of the Association for Information Science and Technology, 66*(10), 1988-2002. https://doi.org/10.1002/asi.23314

Motyka, B. (2018). Employee engagement and performance: A systematic literature review. *International Journal of Management and Economics, 54*(3), 227-244. https://doi.org/10.2478/ijme-2018-0018

Mohammadkarimi, E. (2023). Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. *Journal of Applied Learning and Teaching, 6*(2), 1-9. https://doi.org/10.37074/jalt.2023.6.2.10

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264-269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Nestor, M. S., Fischer, D. L., Arnold, D., Berman, B., & Del Rosso, J. Q. (2020). Rethinking the journal impact factor and publishing in the digital age. *The Journal of Clinical and Aesthetic Dermatology, 13*(1), 12. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7028381/pdf/jcad_13_1_12.pdf

O'Dea, X. C., & O'Dea, M. (2023). Is Artificial Intelligence really the next big thing in learning and teaching in higher education? A conceptual paper. *Journal of University Teaching and Learning Practice, 20*(5), 05. https://doi.org/10.53761/1.20.5.05

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ, 372*. http://dx.doi.org/10.1136/bmj.n160

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice, 20*(2), 07. https://doi.org/10.53761/1.20.02.07

Popenici, S. (2023). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching, 6*(2), 1-7. https://doi.org/10.37074/jalt.2023.6.2.4

Popenici, S., Rudolph, J., Tan, S., & Tan, S. (2023). A critical perspective on generative AI and learning futures. An interview with Stefan Popenici. *Journal of Applied Learning and Teaching, 6*(2), 1-21. https://doi.org/10.37074/jalt.2023.6.2.5

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *A product from the ESRC methods programme Version, 1*(1), b92. https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf

PRISMA. (2020). *PRISMA checklist.* http://prisma-statement.org/prismastatement/checklist.aspx?AspxAutoDetectCookieSupport=1

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://doi.org/10.37074/jalt.2023.6.1.29

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Sallam, M. (2023). ChatGPT utility in healthcare, education, research and practice. Systematic review on the promising perspectives and valid concerns. *Healthcare, 11*(1), 887. https://doi.org/10.3390/healthcare11060887

Spelt, E. J., Biemans, H. J., Tobi, H., Luning, P. A., & Mulder, M. (2009). Teaching and learning in interdisciplinary higher education: A systematic review. *Educational Psychology Review, 21*, 365-378. https://doi.org/10.1007/s10648-009-9113-z

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education, 30*(4), 325-341. https://doi.org/10.1080/02602930500099102

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching, 6*(1), 31-40. https://doi.org/10.37074/jalt.2023.6.1.17

Taros, T., Zoppo, C., Yee, N., Hanna, J., & MacGinnis, C. (2023). Retracted Covid-19 articles: significantly more cited than other articles within their journal of origin. *Scientometrics, 128*(5), 2935-2943. https://doi.org/10.1007/s11192-023-04707-4

Tlili, A., Huang, R., Mustafa, M. Y., Zhao, J., Bozkurt, A., Xu, L., ... & Burgos, D. Speaking of transparency: Are all Artificial Intelligence (AI) literature reviews in education transparent? *Journal of Applied Learning and Teaching, 6*(2), 1-12. https://doi.org/10.37074/jalt.2023.6.2.15

Warrens, M. J. (2015). Five ways to look at Cohen's kappa. *Journal of Psychology & Psychotherapy, 5*. https://doi.org/10.4172/2161-0487.1000197

Xames, M. D., & Shefa, J. (2023). ChatGPT for research and publication: Opportunities and challenges. *Journal of Applied Learning and Teaching, 6*(1), 390-395. https://doi.org/10.37074/jalt.2023.6.1.20

Zawacki Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International Journal of Educational Technology in Higher Education, 16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zeng, X., Zhang, Y., Kwong, J. S., Zhang, C., Li, S., Sun, F., ... & Du, L. (2015). The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: A systematic review. *Journal of Evidence-based Medicine, 8*(1), 2-10. https://doi.org/10.1111/jebm.12141