WIP Post-Assessment Processes Given the Rise of Generative AI: Findings from the Literature

Crista Mohammed
The Department of Electrical and Computer Engineering
The University of the West Indies
St.Augustine Trinidad and Tobago
Crista.Mohammed@sta.uwi.edu

Sean Rocke
The Department of Electrical and Computer Engineering
The University of the West Indies
St.Augustine Trinidad and Tobago
Sean.Rocke@sta.uwi.edu

***Abstract*— Free-to-use generative AI (GAI) threatens assessment integrity. The scholarship establishes that GAI can produce passing to highly sophisticated responses to a range of assessment items. And detecting AI-generated output is fraught. Human detection is spotty and there are no proven software solutions at the time of writing. Even where there are promising detection solutions, these are likely to become obsolete as GAI evolves. The challenge for instructors is reliably and consistently establishing authorship of student submissions. There are two main perspectives on academic cheating. Proactive approaches appeal to students' honor and precede submission. Whereas punitive strategies are employed after submission, intending to detect and punish dishonesty. This paper focuses on post-assessment regimes where academic integrity is checked after submission. This work-in-progress, research-to-practice paper collates post-assessment strategies from the literature for detecting unsanctioned use of GAI. Two sets of scholarship were synthesized to answer one question: What post-assessment strategies are there for detecting unsanctioned use of GAI? The first set of literature focusses on post-assessment strategies in general. The second set addresses post-assessment strategies given the proliferation of free-to-use GAI. The Education Resources Information Center (ERIC) database was searched for general post-assessment strategies, yielding five tools. This search of general, non-discipline specific scholarship returns instructors to practices that have been tried and tested before the advent of GAI. It is possible that they can form part of a larger strategy of reducing GAI misuse; but this requires further study. The second set of tools, extracted from scholarship published by IEEE Xplore, totals six. These are post-assessment tools specific to detecting GAI-generated text. The study reveals that as scholars discuss the nature of GAI misuse, there is need to problematize definitions of that misuse. There are several tools available to instructors for detecting dishonest content; no single tool is infallible. Instructors need to think not in terms of a one-off detection solution but a regimen of post-assessment checks to compensate for this fallibility of tools. And there is the need to test tools that work best given teaching and learning contexts.**

Keywords*— *academic honesty, assessment, assessment integrity, engineering education, generative AI

I. Introduction

The proliferation of free-to-use generative AI (GAI) has given pause to the entire higher education (HE) sector [1-6]. Of particular and repeated concern is the threat that these all-pervading technologies pose to assessment integrity [7-10]. The research establishes that GAI can produce passing to highly sophisticated responses to a range of assessment items [3, 4, 11, 12], including coding tasks [4]. And detecting output generated by AI is fraught. Human detection is spotty [3, 12, 13] and there are no proven software solutions as at the time of writing [3, 4, 12]. Even where there are promising detection solutions, these are likely to become obsolete as generative AI is evolving [3, 12]. The threat from generative AI is that its outputs are indistinguishable from human outputs [3, 5, 12, 13]. The challenge for instructors is establishing authorship of student submissions reliably and consistently.

Ensuring academic integrity is decidedly the business of higher education institutions (HEIs) and its stakeholders. In an environment of increased accountability [14], HEIs are hard-pressed to demonstrate, in palpable terms, how they are ensuring quality in their programs and operations. While this demand for accountability has been recent for some programs, it has always been the case for accredited engineering programs. This paper provides engineering faculty with a cache of post-assessment tools, from which they can chose based on available resources and suitability to their teaching contexts. The intent is to strengthen post-assessment processes so that student misuse of GAI can be reduced.

Two sets of scholarship were synthesized to answer one question: What post-assessment tools are there for detecting unsanctioned use of GAI? The first set of literature focusses on post-assessment tools in general. This search of general, non-discipline specific scholarship returns instructors to practices that have been tried and tested before the advent of GAI. It is possible that they can form part of a larger strategy of reducing GAI misuse; but this requires further study. The second set addresses post-assessment tools given the proliferation of free-to-use GAI.

Overall, the study is part of a larger, urgent discussion on how to deploy GAI responsibly and productively in our classrooms, while discouraging dishonest use of the technology.

II. Addressing Academic Honesty

The scholarship reflects two attitudes on academic cheating from the point of view of the instructor [15, 16]. These attitudes in turn inform approaches to treating with academic dishonesty. Proactive approaches appeal to students' sense of honor and seek to encourage and cultivate integrity [17]. Proactive strategies are necessarily before the fact; that is the intervention—whether an awareness campaign on academic honesty; guidance on referencing—precedes the student submission. Whereas punitive approaches happen after the fact, after the student submits, and the instructor's intent is to detect and punish academic dishonesty. Both approaches have their merits, and in practice it is hardly ever that one approach supplants the other.

This paper examines one dimension of ensuring academic honesty given the proliferation of GAI—detecting misuse of GAI as part of post-assessment processes.

III. Method

The study draws on two sets of scholarship: scholarship that provides general post-assessment tools and scholarship which provides post-assessment tools given the massification of GAI.

For general post-assessment tools, the Education Resources Information Center (ERIC) database was searched. This database was chosen because of its focus on educational sources, irrespective of discipline. Given the nascence of GAI research, particularly investigating its impacts on classrooms, this database was deemed extremely useful as it provides, inter alia, peer-reviewed journal and conference articles, guides, reports and case studies.

The search was limited to the period 2015 to May 07, 2024. After experimenting with several search strings, the database was searched using *assessment integrity*, as this resulted in the largest number of relevant hits. This yielded 40 peer-reviewed, full-text papers in the field of higher education. Of these, nine provide post-assessment methods for detecting academic cheating in general. Of these nine, were five studies [7, 8, 10, 12, 18] that address cheating given the advent of GAI. These were added to the cache of scholarship on detecting misuse of GAI.

To collate post-assessment tools for detecting unauthorized use of GAI, IEEE Xplore was searched. Besides providing peer-reviewed scholarship, it was anticipated that Xplore would provide empirical studies specific to engineering education. This contrasts with the search of ERIC, where the aim was to map the general terrain of post-assessment procedures and to revisit detection solutions that are established.

The search of IEEE Xplore was limited to 2022 to May 07, 2024. November 2022 was used as the start because this marks the public release of free-to-use ChatGPT, arguably the most used GAI. After trial and error with several search terms, we eventually queried all metadata using the string *generative AI* AND *assessment integrity* OR *academic honesty* OR *plagiarism*, as this extracted more papers than other search strings. Results were then filtered by selecting only those articles that appeared in the subject fields of plagiarism detection; academic integrity; academic dishonesty; artificial intelligence tools. This yielded 175 papers. Abstracts were read to select those papers which provide post-assessment tools for detecting GAI generated text. The vast majority of papers dealt with the engineering behind artificial intelligence, like developing large language models. We found nine articles that address, in the main or tangentially, detecting AI generated content in HE teaching and learning contexts.

IV. Findings

The scholarship on general detection tools offers up five (see Table 1). And the scholarship on detecting GAI content proposes six tools (see Table 2). The findings can be grouped into three categories: policy frameworks; manual checks; and technological solutions. Policy frameworks cannot be considered a tool so much as an approach and therefore was not entered into Tables 1 and 2.

TABLE I General Detection Tools

Tools	***Source***
Moderate assessments.	[15, 19]
Use ipsative assessments.	[20]
Use learning analytics.	[21, 22]
Use text matching software.	[15, 22]
Use predictive analytics.	[16]

TABLE I Tools Specific for Detecting GAI

Tools	***Source***
Use detection tools for machine generated content.	[8, 10-12, 23-27]
Check references as GAI is tending to get these wrong.	[26, 27]
Use learning analytics.	[26, 28, 29]
Use plagiarism tools.	[11, 26, 27]
Look for telltale signs of GAI use, like errors of fact	[27]
Manual reviews of code, looking for unusual patterns, style inconsistencies; suspicious comments	[30]

A. *Policy Approaches*

Some scholars address GAI misuse detection as part of larger policy frameworks [10, 21, 31]. One study, for example, [10] recommends a broad multi-pronged strategy to treating with academic honesty. Predictably this includes both proactive and punitive measures. But among the punitive measures is recognition of the need to train instructors in the use of AI detection tools and invest in sourcing suitable detection tools. This call for educator training in AI use and detection of AI-generated content has been repeated by others [7, 18]. The implication is that having the detection tool is not sufficient in of itself. Rather, the efficacy of the tools depends largely on how proficient users are in deploying the tool and the fitness of the tool to the teaching and learning context.

B. *Manual Checks*

Some manual checks include checking for signposts of GAI use, like incorrect referencing, and incorrect information [26, 27]. Also, some scholars recommend moderation schemes [15, 19], where the goal is not just ensuring fairness of marking but looking for similarities across student submissions. Such a strategy needs to be tested in the context of GAI use—will, for example, the AI generate the same or nearly the same text for different users who are making similar but not the same queries? Another recommendation that may be useful for detecting GAI-generated content is using ipsative assessments [20]. Like learning analytics, the goal is to compare submissions by a single student with a view to detecting anomalies that might indicate cheating. Those anomalies may be a marked improvement in performance or difference in writing style, for example.

Three sources recommend the use of learning analytics [21, 22, 28], which is typically automated to some degree. However, one study [28] proposes a manual process for checking students' language proficiency—specifically their mastery of language frames—as a way of establishing authorship. Their process has nine steps, making it like the other manual checks described here, time and resource intensive.

C. *Technological Solutions*

It is necessary to distinguish between plagiarism detection tools and detection tools for machine generated language. We regard plagiarism tools as those that aim to detect similar text across artefacts. While detection tools for machine-generated content aim to identify text that has likely been created by artificial intelligence.

Sometimes the distinction was not clear in the scholarship. Consider here this description of ZeroGPT which obfuscates the difference in the aims of the technology: "ZeroGPT is a *plagiarism detection* (our emphasis) tool that differentiates content created by AI tools from human-authored content"[25]. The root of the confusion lies in a failure to problematize what plagiarism is in the context of GAI. If lifting text from a third-party constitutes plagiarism [32] , then there is no identifiable third-party; GAI generates a veritable mash-up of work belonging to others. As noted by publishers [32, 33], GAI can hardly enjoy the status of an author, nor can the technology bear the responsibilities that come with authorship. Instead of considering such material as plagiarized it is best to speak in terms of dishonest content or unsanctioned use of GAI. We concede that working through the semantics of this issue is beyond the scope of the paper. However, larger, more rigorous, and urgent discussion is needed; how are we to formulate policies and strategies, where we are mislabeling the problem? As it relates to describing tools like ZeroGPT, it might be more accurate to say that such tools aim to detect content that was likely created by AI, without invoking ideas of plagiarism, until the nature of cheating through GAI misuse is better ventilated.

The jury is very much out on the effectiveness of current detection tools to reliably and consistently detect AI generated text [3, 7, 12, 13]. One study [7] found that Crossplag, OpenAI Detector, GPTZero, and Turnitin were fairly successful in detecting GAI text, where the entire artefact was generated by GAI. However, these detection tools had spotty performance where elements of the text were written by humans. More alarmingly, human evaluators fared worse in detecting GAI use [7]. And where there are promising tools, these are likely to grow obsolete as GAI technologies evolve [12].

A few studies recommend using learning analytics which has been automated to some degree [21, 22, 29]. Recall here the manual use of learning analytics in the previous section, where the goal was the same, but the method of deployment was different. An analytics-based approach attempts to match learners to their work by elucidating patterns in their writing. Where textual features do not match with the student's previous patterns of language use, then claims at authorship can be challenged. However the efficacy of this strategy depends on a number of factors, not just the algorithms used to classify the data; factors include, among others, the size of the data set, the normalization techniques used, and classifier parameters [21]. And, the accuracy of matching authors to text can be highly spotty, with one study reporting a success rate of between 18% −100% [21].

Another solution is to use the very technologies upon which GAI draw , such as machine learning models, to detect GAI-generated content [23]. However, these detection tools like learning analytics and machine learning models require instructors to have specific domain knowledge—they must know how to devise and use algorithms; train the model; and interpret the outputs of the system, for example. And where instructors do not themselves have this capacity, there is a cost to hiring support for this type of post-assessment process. These methods present a huge, perhaps insurmountable, learning curve for some instructors or where additional staff is needed there is a cost burden attached.

Another technological solution is predicative analytics [16]. When students work online, a slew of data related to their physical work context and learning behaviors can be collected. These data can be used in integrity checks. Unusual or suspicious activity can trigger investigations into possible dishonesty. One such scenario is where a student has scored well on assessments, but data show low course engagement [16].

V. CONCLUSION

The scholarship establishes that there is no one-off strategy that can be used consistently and reliably to detect unauthorized use of GAI. In fact, instructors are advised to think in terms of post-assessment regimens drawn from 'old' ways of checking for cheating and new emerging ways of looking for AI-generated content. For example some studies recommend specific combinations of detection tools, like using in tandem AI detection and similarity matching software [11]. But careful thought is needed: highly complex regimens have cost and time implications [21]. And some tools can be invasive [21], threatening students' privacy, for example.

Moreover, instructors are encouraged to construct their post-assessment regimens so that they are defensible given prevailing good assessment practices. But then, the question arises as to what constitutes good practice. This paper contributes to that discussion. It focusses on a contemporary, vital issue: what cheating detection tools are there, both pre-GAI and in the face of GAI? And how can engineering faculty deploy these to strengthen their post-assessment processes so that student misuse of GAI can be detected and ultimately reduced? In this work we have also taken a generalist approach, all the while appreciating that different teaching and learning contexts require specific post-assessment treatments. This no doubt will be an ongoing project for instructors like us—figuring out what works best given our teaching needs and resources.

What is unequivocal is that HEIs need to invest in frequently upgrading their GAI-detection systems to keep apace of AI advances [12, 26]. In these increasingly litigious times, the need to ensure fitness of solutions cannot be overstated. Indeed, it will not just be reputationally damaging to accuse a student of dishonesty based on the outputs of unreliable detection systems, but there can be legal consequences. And again, having robust systems are not enough, staff must be trained [7, 10, 18] and that training must be ongoing.

One limitation is that this study investigates post-assessment process for detecting GAI-generated text, failing to account for the many other artefacts that can be produced by GAI. This limitation is mirrored in the literature where there is understandably an instructor pre-occupation with text. But this single-minded focus ought not to be encouraged in engineering, where students are asked to produce more than text as evidence of learning achievement—like creating code and visual representations of data.

To sum, the HEI community needs to have a shared working understanding of the nature of GAI-enabled cheating. There is need to think not in terms of a single post-assessment detection tool but several, as no one tool is infallible. Engineering instructors need to think beyond GAI-generated text, considering all the other GAI outputs. HEIs are to invest in searching for and testing tools that work best in their teaching and learning contexts. And there is need for ongoing staff training.

REFERENCES

[1] Y. K. Dwivedi et al., ""So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," International Journal of Information Management, vol. 71, p. 102642, 2023.

[2] D. O. Eke, "ChatGPT and the rise of generative AI: Threat to academic integrity?," 2023.

[3] C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," Education Sciences, vol. 13, no. 4, p. 410, 2023.

[4] K. Malinka, M. Peresíni, A. Firc, O. Hujnák, and F. Janus, "On the educational impact of chatgpt: Is artificial intelligence ready to obtain a university degree?," in Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1, 2023, pp. 47-53.

[5] J. Rudolph, S. Tan, and S. Tan, "ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?," Journal of Applied Learning and Teaching, vol. 6, no. 1, 2023.

[6] D. Sardana, T. R. Fagan, and J. T. Wright, "ChatGPT: A disruptive innovation or disrupting innovation in academia?," The Journal of the American Dental Association, vol. 154, no. 5, pp. 361-364, 2023.

[7] K. Alexander, C. Savvidou, and C. Alexander, "Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education," Teaching English with Technology, vol. 23, no. 2, pp. 25-43, 2023.

[8] N. Cong-Lem, T. N. Tran, and T. T. Nguyen, "Academic Integrity in the Age of Generative AI: Perceptions and Responses of Vietnamese EFL Teachers," Teaching English with Technology, vol. 24, no. 1, pp. 28-48, 2024.

[9] D. R. Cotton, P. A. Cotton, and J. R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT," Innovations in Education and Teaching International, pp. 1-12, 2023.

[10] P. Bannister, A. Santamaría-Urbieta, and E. Alcalde-Peñalver, "A Delphi Study on Generative Artificial Intelligence and English Medium Instruction Assessment: Implications for Social Justice," Iranian Journal of Language Teaching Research, vol. 11, no. 3, pp. 53-80, 2023.

[11] M. Neumann, M. Rauschenberger, and E.-M. Schön, ""We Need To Talk About ChatGPT: The Future of AI and Higher Education"," in 2023 IEEE/ACM 5th International Workshop on Software Engineering Education for the Next Generation (SEENG), 2023.

[12] M. Perkins, "Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond," Journal of University Teaching & Learning Practice, vol. 20, no. 2, p. 07, 2023.

[13] J. Crawford, M. Cowling, and K.-A. Allen, "Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI)," Journal of University Teaching & Learning Practice, vol. 20, no. 3, p. 02, 2023.

[14] N. Macheridis and A. Paulsson, "Tracing accountability in higher education," Research in education, vol. 110, no. 1, pp. 78-97, 2021.

[15] L. Amrane-Cooper, S. Hatzipanagos, and A. Tait, "Developing student behaviours that support academic integrity in distance learning," Open Praxis, vol. 13, no. 4, pp. 378-384, 2021.

[16] A. Lee-Post and H. Hapke, "Online learning integrity approaches: Current practices and future solutions," Online Learning, vol. 21, no. 1, pp. 135-145, 2017.

[17] S. E. Eaton, M. Guglielmin, and B. K. Otoo, "Plagiarism: Moving from Punitive to Proactive Approaches," in Selected Proceedings of the IDEAS Conference 2017: Leading Educational Change Conference, School of Education, University of Calgary, Calgary, 2017, pp. 28–36.

[18] M. Imran and N. Almusharraf, "Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature," Contemporary Educational Technology, vol. 15, no. 4, p. ep464, 2023.

[19] E. Polisca, S. Stollhans, R. Bardot, and C. Rollet, "How Covid-19 has changed language assessments in higher education: a practitioners' view," Innovative language teaching and learning at university: facilitating transition from and to higher education, pp. 81-91, 2022.

[20] S. S. Vellanki, S. Mond, and Z. K. Khan, "Promoting Academic Integrity in Remote/Online Assessment-EFL Teachers' Perspectives," TESL-EJ, vol. 26, no. 4, 2023.

[21] A. Amigud, J. Arnedo-Moreno, T. Daradoumis, and A.-E. Guerrero-Roldan, "Using learning analytics for preserving academic integrity," International Review of Research in Open and Distributed Learning, vol. 18, no. 5, pp. 192-210, 2017.

[22] K. A. Gamage, E. K. d. Silva, and N. Gunawardhana, "Online delivery and assessment during COVID-19: Safeguarding academic integrity," Education Sciences, vol. 10, no. 11, p. 301, 2020.

[23] H. Alamleh, A. A. S. AlQahtani, and A. ElSaid, "Distinguishing human-written and ChatGPT-generated text using machine learning," in 2023 Systems and Information Engineering Design Symposium (SIEDS), 2023, pp. 154-158: IEEE.

[24] A. Singh, "A comparison study on AI language detector," in 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), 2023, pp. 0489-0493: IEEE.

[25] M. Čavojský, G. Bugár, T. Kormaník, and M. Hasin, "Exploring the Capabilities and Possible Applications of Large Language Models for Education," in 2023 21st International Conference on Emerging eLearning Technologies and Applications (ICETA), 2023, pp. 91-98: IEEE.

[26] Y. Liu, "Leveraging the Power of AI in Undergraduate Computer Science Education: Opportunities and Challenges," in 2023 IEEE Frontiers in Education Conference (FIE), 2023, pp. 1-5: IEEE.

[27] J. Mrabet and R. Studholme, "ChatGPT: A friend or a foe?," in 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2023, pp. 269-274: IEEE.

[28] P. Ilic and N. Carr, "Work in Progress: Safeguarding Authenticity: Strategies for Combating AI-Generated Plagiarism in Academia," in 2023 IEEE Frontiers in Education Conference (FIE), 2023, pp. 1-5: IEEE.

[29] D. Jagtap, S. Ambekar, H. Singh, and N. Sharma, "An Approach to Detecting Writing Styles Based on Clustering Technique," in 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 2024, pp. 1-7: IEEE.

[30] J. Berrezueta-Guzman and S. Krusche, "Recommendations to create programming exercises to overcome ChatGPT," in 2023 IEEE 35th International Conference on Software Engineering Education and Training (CSEE&T), 2023, pp. 147-151: IEEE.

[31] A. H. Verhoef and Y. M. Coetser, "Academic integrity of university students during emergency remote online assessment: An exploration of student voices," 2021.

[32] Elsevier. (2024). Publishing ethics. Available: https://www.elsevier.com/about/policies-and-standards/publishing-ethics

[33] Springer Nature (2004, April 23 2024). Editorial policies: our editorial policies (A -Z). Available: https://www.springernature.com/gp/policies/editorial-policies