A Review of AI-Powered Writing Tools and Their Implications for Academic Integrity in the Language Classroom

Jasper Roe
James Cook University Singapore, jasper.roe@jcu.edu.au

Willy A. Renandya
Nanyang Technological University, willy.renandya@nie.edu.sg

George M. Jacobs
Universiti Malaya, george.jacobs@gmail.com

Follow this and additional works at: https://animorepository.dlsu.edu.ph/jeal

Part of the Educational Assessment, Evaluation, and Research Commons, Educational Technology Commons, and the Scholarship of Teaching and Learning Commons

Recommended Citation
DOI: https://doi.org/10.59588/2961-3094.1035
Available at: https://animorepository.dlsu.edu.ph/jeal/vol2/iss1/3

This Article is brought to you for free and open access by the DLSU Publications at Animo Repository. It has been accepted for inclusion in Journal of English and Applied Linguistics by an authorized editor of Animo Repository.
A Review of AI-Powered Writing Tools and Their Implications for Academic Integrity in the Language Classroom

Jasper Roe
James Cook University, Singapore
jasper.roe@jcu.edu.au

George M Jacobs
Universiti Malaya, Kuala Lumpur, Malaysia

Willy A Renandya
Nanyang Technological University, Singapore

Abstract
Writers have many digital tools available to help them with the creation of text. In some cases, these tools have been in existence for a long time, such as spellcheckers and basic grammar checkers that are available on word processing software. Today, new and increasingly more advanced tools are in use, and the ramifications of their use are not yet fully understood, particularly in the language classroom. Public interest in such tools has reached new levels with the release of artificially intelligent tools such as ChatGPT. In addition to this, the speed at which assistive writing technologies are developing may outpace that of the creation of institutional academic integrity policies and guidelines for their use. This results in grey areas, confusion, and a lack of awareness of such tools and their uses among educators, students, and administrators. This conceptual article seeks to systematically review and categorize these tools drawing on the research literature and the authors’ personal experience in the classroom. From this inductive analysis, three categories of digital writing tools aside from ChatGPT and large language model (LLM)-driven text creation tools are described. These are machine translators (MTs), digital writing assistants (DWAs), and automated paraphrasing tools (APTs). The key contribution of this article is the development of these three categories, which can form a basis for crafting comprehensive pedagogical approaches and academic integrity policies that focus on a broader range of tools than ChatGPT and LLMs alone.

Keywords: LLMs, ChatGPT, digital writing assistants, AI writing tools, language learning
Introduction

This article seeks to develop a thorough review and systematic categorization of the artificial intelligence (AI)-powered writing tools used in language education aside from large language models (LLMs) such as ChatGPT. The development and increasing use of these tools must be viewed as a part of a broader shift in higher education (HE), which has been undergoing a process of digitalization for the past few decades, and has been rapidly accelerated by the onset of the COVID-19 pandemic in December 2019 (Bygstad et al., 2022). This acceleration took place as a result of higher education institutions (HEIs), students, and staff having to cope with a sudden switch to online learning to allow for physical distancing and reduce the spread of the SARS-CoV-2 virus (Adedoyin & Soykan, 2020). The world has now begun returning to a “new normal” with the lifting of social distancing restrictions; it is likely that an increased focus on online learning will remain a component of HE in the postpandemic landscape (Turnbull et al., 2021). As a result, it is possible that assessments and learning activities will more frequently require digital writing. With this comes the potential for the use of digital writing tools. However, there is little research on the academic integrity implications of digital writing tools, which is concerning as the COVID-19 pandemic led to rising cases of academic integrity violations (Dey, 2021).

As a result of these changes to academic life, there has been significant attention given to some of the more well-known forms of academic misconduct, such as contract cheating, in which a student outsources their work to another party and then submits it as their own (Clarke & Lancaster, 2006). Lancaster and Cotarlan (2021) found an increase in searches for contract cheating sites during the early stages of the pandemic (Roe, 2022). HE regulatory bodies took action to combat contract cheating during the pandemic era. In Australia, the Tertiary Education Quality and Standards Agency (TEQSA) was instrumental in proposing the “Prohibiting Academic Cheating Services Bill 2019” in the Australian Parliament (TEQSA, 2020), and similar legal actions have been taken to outlaw “essay mills” in the United Kingdom (gov.uk, 2022). Less attention has been given to other forms of academic integrity violations, including both unintentional and intentional violations through the use of digital writing tools, meaning that this is an understudied area of academic integrity research.

Digital Writing and Academic Integrity

The main violation of academic integrity principles that can occur through the use of digital writing tools is textual plagiarism. Plagiarism in this case is defined as when an intellectual product that does not belong to the author is passed off or presented as their own (Helgesson & Eriksson, 2015) or misrepresenting authorship by submitting information that belongs to another without an accurate reference (Perkins et al., 2018). In HE, plagiarism detection software has been studied for decades (Decoo & Colpaert, 2010). To date, the most popular software for detecting textual plagiarism is Turnitin, which is used by 15,000 HEIs in 140 countries (Mphahlele & McKenna, 2019). However, Mphahlele and McKenna (2019) argued that Turnitin is not an effective plagiarism detection tool but is more suitable as an instructional tool. Turnitin can be used to show students how sources are used to construct text. The idea of Turnitin as solely a plagiarism detection tool is a “myth.” The new categories of digital writing tools discussed here (machine translators [MTs], digital writing assistants [DWAs], and automated paraphrasing tools [APTs]) can also effectively bypass such systems.

Formulating a categorical system to describe these different writing tools simplifies the job of addressing HE policy on their use. Prior to the pandemic, university academic integrity policies had been the subject of criticism. Abasi and Graves (2008) highlight that these policies can be opaque and only serve to mystify academic writing, while Pecorari (2022) equally claims that there existed little specificity in many university codes of academic conduct. For English-as-a-second-language (ESL) students enrolled in HEIs with English as a medium of instruction (EMI), the challenges are greater still. There is a growing consensus that often, especially among ESL students, inappropriate borrowing of text is more often a learning issue than a case of intentional deception (Shi, 2012), and the same may apply when learners use tools to assist their writing. As a result, clarity is needed on what these tools are and how they can and should be used in postpandemic digital writing. An ideal situation should find common ground on what plagiarism is and include policies that help to form an educative rather
than a punitive reaction (Bacha & Bahous, 2010).

The conversation regarding academic integrity and digital writing tools has recently moved to that of LLMs. LLMs are tools that use AI to generate text and include the now well-known ChatGPT, a chatbot built on the Generative Pre-Transformer 3 (GPT-3) model belonging to the OpenAI company (Dale, 2021). GPT-3 and ChatGPT are now attracting significant attention in the media, and academic studies regarding their impact on academic integrity are ongoing (Dehouche, 2021; Eaton et al., 2021; Kumar, Mindzak, Eaton, et al., 2022; Kumar, Mindzak, & Racz, 2022; Wilder et al., 2021). Perkins (2023) offered a broad overview of LLMs and their implications for academic integrity on a holistic level, finding that even trained faculty cannot distinguish between human-produced and LLM-produced text. Artificial-intelligence-produced text is an existential issue that is affecting academia. However, when having discussions about academic integrity policy and acceptability in the language classroom, it is not only these LLMs which educators should be aware of. Indeed, it is useful to understand the broad range of tools that are currently used by students. In this article, we review and categorize AI-powered writing tools that are not solely limited to LLMs such as ChatGPT that produce original text.

In order to develop a system of categories of digital writing tools that may result in violations of academic integrity aside from LLMs that produce original text, the authors of the current article used an inductive approach building on personal experience and secondary online research. Following this, the different tools were separated into three categories based on similarity of function. However, many of these tools can be used in conjunction with one another or at the same time within a single online interface and there is significant overlap between them. Furthermore, each of these tools are easily accessible online, often without charge.

**Machine Translators**

The first category of online writing tools is that of MTs. Translation between languages is vital for the sharing of knowledge across borders. The development of technology has brought with it the potential to translate from one language to another at the click of a button. One of the most well-known translation tools, Google Translate, serves 500 million users daily and allows for the translation of 133 languages, using a neural-machine translation model that translates entire sentences at a single time (Google, 2016). A similar translator, launched in 2017, DeepL, uses neural networks to translate between 29 languages, claiming a level of nuance and accuracy 10 times higher than any other machine translation (DeepL, 2022). This increasing accuracy and ease of access increase the potential for translation plagiarism.

Translation plagiarism is a form of “hidden” plagiarism that has been noted in academia, as the detection of cases can often take years or even decades to surface (Dougherty, 2019). Such cases commonly occur when the author takes the credit for the translation of source material and passes it off as their own. As it is translated, text-matching software cannot detect such cases. Even manually, detecting translated plagiarism is complex, making it a “subtle” form of plagiarism (Dougherty, 2019). Consequently, tools to detect cross-language and translation plagiarism are being developed, such as that described by Potthast et al. (2011), who found near matches across six major European languages using cross-language plagiarism detection. Therefore, although there is the possibility for detection of translation plagiarism using advanced computer science methods, this is not yet widespread (Dougherty, 2019).

Dougherty’s (2019) work on translation plagiarism applies to professional researchers who are authoring academic works. However, the same principles can be applied when students complete academic assignments in their course of study. Highly accurate MTs such as DeepL can be used to commit translation plagiarism on student essays. However, whether translation plagiarism has occurred depends on the extent to which original material has been translated. If a student from a non-English-speaking background needs to submit an assignment in English but is not sure how to say what they want to say in English, then using an MT is a helpful pedagogical tool. If a student wishes to clarify their understanding of a text in English, then translating it back to the first language may equally be helpful. If, on the other hand, a student writes an assignment entirely in their first language and then translates it to the target language, it is less clear whether this is academically acceptable. Furthermore, when a student begins to translate entire sentences or paragraphs, the potential for translation plagiarism increases.

Another case of translation plagiarism that can occur among students using MTs is that of back-translation. 
Back-translation is when text is “re-translated” or passed through a machine translation tool multiple times to sufficiently change the language to avoid detection by text-matching antiplagiarism software (Jones, 2009) and effectively thwart tools like Turnitin (Jones & Sheridan, 2015). Both back-translation and straightforward translation plagiarism can be achieved using MTs and are generally difficult to detect, but such cases are made even more difficult to detect when MTs are used in conjunction with students’ original work. This can occur when students take a passage of writing in a foreign language, translate it to the target language, and then add to it, edit it, and refine it using their own resources. Or more simply, students may use an MT and then edit as appropriate for style and content or to fix any errant or nonsensical mistakes that have occurred during the translation process.

The growing effectiveness and accuracy of MTs means that this is an important area of attention in HEI policy and in the postpandemic digital writing landscape. While an outright ban on the use of MTs in academic work may in theory solve the problem, this would ignore the huge benefits that MTs can provide to students from ESL backgrounds. MTs have been used effectively in language teaching (Niño, 2009), and given that lower language ability has been shown to relate to higher incidences of academic misconduct (Bretag, 2007; Perkins et al., 2018), reducing the options for students may inversely lead to an increase in other forms of academic cheating. In their discussion of policy implications of MT in English for Academic Purposes (EAP) teaching, Mundt and Groves (2016) argued that technological help in academic work is the norm, and MTs should be accepted on this basis, while also proposing two strategies for dealing with these tools. The first of these is that HEIs must be prepared to create clear legislation on how MTs can be used to avoid a lack of clarity. The second is that a research agenda should be formalized for how students use MTs. By combining these approaches, it is possible to retain the positives of MTs as an educational tool and a valuable aid to study. It is also important to recognize that MTs may lead to greater equality and social justice outside of the language classroom, given their ability to ensure that all students will be assessed solely on content, rather than language use (in non-language-based assessments).

In terms of MTs, the implications for HEIs’ policy response must consider to what extent back-translation and other forms of MT use represent academic integrity violations if the content is from another author and not acknowledged and whether a student writing in a language other than the target language then translating it automatically constitutes their own original work.

**Digital Writing Assistants**

A further AI-powered assistive writing tool that is an area for consideration in terms of academic integrity is DWAs. These differ from LLMs such as ChatGPT, in that they run alongside originally written content and do not have the ability to produce original content in response to a prompt. DWAs can be defined as software packages that are either browser or device based and make use of AI to assist with the writing process. DWAs, in this discussion, include not simple grammar and spell checkers but tools that have the ability to achieve complete transposition and paraphrase of text at the sentence level and which provide macrostructural feedback. One of the most common DWAs currently in use is Grammarly. Grammarly operates by using AI to improve writing and provide suggestions for text (Grammarly, 2022). The program runs alongside and complementary to word processing software and has been shown to be effective when used in a pedagogical setting for instructing students on English writing. O’Neill and Russell (2019) found that Grammarly may help provide input to students to improve their grammatical accuracy, encouraging self-confidence and autonomy in the editing stages of writing. Aside from grammar, Grammarly improves stylistic expression and locates lexical errors, while also offering alternative phrasing for written sentences (Barrot, 2020).

Tools such as Grammarly can also be used by instructors to lessen the burden of assessment marking. Thi and Nikolov (2021) found that automated feedback from the Grammarly tool can be complementary to traditional feedback from a teacher and that this may allow instructors to dedicate more time to feedback on higher order elements of writing. Further to this, some parts of Grammarly’s product line incorporate a plagiarism checker, allowing users to reduce the possibility that they are accidentally plagiarizing using text-matching software (Grammarly, 2022). Consequently, DWAs of this nature can help learners to become more autonomous and self-efficacious and increase self-confidence (Nazari et al., 2021). The
effectiveness of Grammarly as a tool to help students learn how to avoid plagiarism and improve their writing is demonstrated by its popularity, even among schools and HEIs. Over 3,000 institutions, among them world-class universities, currently use Grammarly for both faculty and students (Grammarly, 2022).

Aside from Grammarly, tools such as Wordtune fulfill a similar function. Wordtune is a DWA that uses AI to help alter tone, length, and structure. This tool is used primarily for ESL writers (Zhao, 2022). Other DWAs like ColloCaid have been developed to assist EAP learners in choosing the correct language (Frankenberg-Garcia et al., 2019), and corpus-based natural language processing (NLP) models of DWAs have been developed in Taiwan to further aid ESL learners (Chang et al., 2008). Relating DWAs to the question of academic integrity, the issue is whether extensive use of such tools truly constitutes students’ own work. If it can be argued that using AI to improve the quality, structure, and content of writing does not constitute a student’s own work, then perhaps it can be seen as a violation of academic integrity. Grammarly’s website, however, identifies in the FAQ section that text coming from Grammarly will not “show up as plagiarized” as it is not indexed on the internet (Grammarly, 2022). This answer indicates mainly that the text will avoid detection, rather than addressing the core issue of whether the text has truly been written by the author. That said, DWAs do not give ideas, and they are not capable of writing entire texts. They may misrepresent a student’s language ability and create an unfair advantage in stylistic expression, but alone it is difficult to make a case that their use could constitute textual plagiarism. On the contrary, they may improve students’ understanding of what constitutes plagiarism through on-the-fly checking tools such as those in Grammarly.

The main area of concern in the use of DWAs is one of equity. In assessment of language especially, DWAs may offer an unfair advantage. This is evidenced by the actions of the South Australian regulatory board, which has considered banning the use of Grammarly in English examinations as it would privilege those students who used it over others (Shepherd, 2022). As a result, if university-level assessments offer marks for style and expression, grammar, and language use within their rubrics, provision must be made for an equitable approach so that DWAs are either permitted or not permitted. If marks are available for the quality of writing rather than content of ideas, then the use of DWAs by some students means a less level playing field that disadvantages those without access. Institutions may consider taking a whole-of-university approach and providing access to DWAs by default to promote equity of assessment.

**Automated Paraphrasing Tools**

The final category of tools used by students in digital writing that present a risk to academic integrity is that of APTs. APTs are a growing phenomenon, with Rogerson and McCarthy (2017) among the first to recognize their use. APTs were first developed to help avoid antiplagiarism software based on text-matching technology. They function by substituting words and phrases for others (Prentice & Kinden, 2018).

Illicit APT use is generally conducted by taking a source text and passing it through the tool on an online platform. The tool, with varying levels of sophistication, produces a paraphrased version through synonym substitution. This can at times result in incomprehensible “word-salad” (Rogerson & McCarthy, 2017), or it can produce an output that retains the same ideas but in a different text, evading text-matching plagiarism detection tools and passing off the source text as an original work. The individual submitting the work will then claim original ownership and authorship of the work, having done little more than press a button. The ways in which use of APTs can sabotage academic integrity are numerous. APTs can be used in self-plagiarism, intentional paraphrase plagiarism, and in cases of collusion and can also be used alongside MT tools. In self-plagiarism, an APT can be used to change the wording of previously written text by the author themselves. In intentional paraphrase plagiarism, the author can take source material from elsewhere (e.g., a website or a journal) and rephrase it automatically, while in cases of collusion, students may work together to automatically paraphrase parts of different texts (such as their own essays or previously submitted assignments) into an entirely new one. Another instance involving MT may be to take a foreign-language text and translate it to English through an MT tool, then provide an additional layer of alteration by using an APT to change the wording or refine the structure and meaning. With each iteration of this process, the likelihood of detecting plagiarism becomes smaller and smaller.
However, individuals can also unintentionally violate academic integrity policies by using an APT inappropriately. Appropriate use may include using an APT to give suggestions or to show possible ways to recast text to express an idea. However, using unedited output without any author input without declaring it may be considered a violation of academic integrity. Indeed, the use of a thesaurus to explore and understand synonyms is of benefit in the language classroom and beyond, and the argument can be made that using a DWA is not too dissimilar from the use of the thesaurus function in a word processor. The relevant detail again comes down to a matter of quantity; a single word or cluster of words may be acceptable, but a sentence, paragraph, or more may not be. A further complicating factor is that institutions may not have any provision in their policy documents on the use of APTs.

Dinneen (n.d.) described a case study that exemplifies this lack of clarity in HEI policy. In this case, a student’s submitted assignment showed hallmarks of academic misconduct. The student was not able to articulate the content of her submitted work but voluntarily showed her search history on the internet, which led to the discovery of her use of several paraphrasing and translation tools in the writing process. Less than a quarter of the original work was independently authored by the student, yet the student claimed she had operated within the letter of the policies of the institution. Dinneen (n.d.) highlighted that the student had provided citations and altered the wording of the source text, and the policy documents in question made no mention of the tools that she had used, meaning that she had not broken any of the prescribed rules given to her.

There are several implications for HE writing in a postpandemic world when it comes to APTs. Firstly, further research is required to understand how and why students are using APTs and whether they are aware of the risk of such use occasioning textual plagiarism. Secondly, the “silence” (Dinneen, n.d.) on the allowable use of writing tools in policy must be addressed at an institutional level. Thirdly, HEIs must prepare to rapidly respond in a clear and consistent manner as new technologies continue to become available, which may pose a threat to academic integrity. New technologies such as LLMs and artificially intelligent text producers such as OpenAI’s GPT-3 are one such example (Thunström, n.d.). Pedagogically speaking, instructors should focus on the “why” of paraphrase rather than the “how” to help students understand why the use of APTs is not necessarily helpful to learning. This can be achieved by communicating that paraphrasing is not merely transposition of words or substitutions of synonyms but should be additive, as in capture the ideas of the source text while also adding an aspect of inferential thinking, which Yamada (2003) describes as a “good” paraphrase.

**Implications for Academic Integrity Policies**

These three categories of digital writing tools each need addressing within HE. Against a backdrop of an accelerating digitalization of HE in a postpandemic world, educators must actively seek to learn about the tools that their students may use, and this is not limited to the conversations regarding artificially intelligent text-producing tools such as ChatGPT and LLMs. It is also important to note that MTs, DWAs, and APTs each have legitimate pedagogical uses, particularly in the language classroom. MTs and DWAs have shown their value for ESL learners and can assist all students to write in line with academic integrity principles. On the other hand, if used incorrectly or dishonestly, these tools can pose a significant risk to digital writing.

**Implication One—Translation Plagiarism Policies Must Be Comprehensive**

MTs can be used for translation plagiarism and back-translation, which is unidentifiable to text-matching software commonly used as a first-line defense against academic misconduct. Policies must focus on giving accurate and fair uses cases and the thresholds of machine translation constituting translation plagiarism.

**Implication Two—DWA Use Should Be Approached With Caution**

Policies must recognize that the use of DWAs can lead to an inequitable situation if only some students use it and others do not, particularly if there are marks or credits available for structure, language, style, and expression. DWAs are less likely to produce textual plagiarism as they provide limited suggestions for text improvement, but it is likely that more sophisticated DWAs will emerge in the future, which could result in academic integrity becoming more of a relevant concern.
Implication Three—Acceptable APT Use Must Be Specified

APTs are comparatively more likely to be used for engaging in paraphrase plagiarism, but this is not necessarily intentional. The case study from Dinneen (n.d.) exposed the deficiencies in institutional policy, which led a student to believe that submitting an assignment that was 75% unoriginal work is acceptable. Thresholds must be given as well as when and where APTs may be used, if at all.

Other Considerations for Academic Integrity Policies

There are several different approaches that HEIs can take towards policy creation to deal with the categories of writing tools described here. A clear and comprehensive institutional policy would highlight the risks and benefits of such tools and their appropriate use through examples. A disadvantage of this approach is that it may require frequent updates and become rather long and unwieldy. A second approach that can be considered is that of a policy regarding cognitive offloading. The strategy of cognitive offloading described by Dawson (2020) means that use of some tools is allowed to assist with writing, but that this is clearly acknowledged and credited by the author. This would mean students may be able to use some digital writing tools without fear of repercussions. However, in cases where little to no work is completed by the submitting author, such as those who use APTs extensively, this may not be suitable.

Within the classroom, instructors should focus on ensuring that students understand why they engage in techniques such as paraphrasing or writing in the target language of assessment and demonstrate the risks of inadvertently or deliberately using tools such as APTs to disguise authorship or gain advantage without the author’s “true” writing, for example, the output of “word salad” (Rogerson & McCarthy, 2017), which can form a useful and entertaining classroom activity.

Conclusion

Despite academic integrity featuring in the media throughout the pandemic (Dey, 2021) and an increasing focus on the use of ChatGPT and LLMs and their potential for authoring academic writing in the classroom and beyond, there is still little said in academic integrity policies on where currently used tools such as MTs, DWAs, and APTs described in this paper sit on the spectrum of acceptability (Dinneen, n.d.). Research has demonstrated that students are often concerned about breaching academic rules and regulations but fear that they cannot comply with the rules and do not know how to effectively cite or reference source material, leading them to the conclusion of it being unfair that they receive a punishment or sanction (East & Donnelly, 2012). As technology continues to accelerate, the rate of development in advanced writing tools that manipulate language for a variety of purposes, including to aid academic work both legitimately and illicitly, will continue to grow (Roe & Perkins, 2022). For these reasons, institutions must work towards clear guidelines that evolve with technology. By understanding the situation fully, training can be implemented. Training is essential and has shown to be an effective intervention in reducing academic integrity violations (Perkins et al., 2020), so presents one of the most important interventions available.

References


Bretag, T. (2007). The emperor’s new clothes: Yes, there is a link between English language competence and academic standards. *People and Place, 15*(1), 13–21.


of its advantages and disadvantages. *ReCALL*, 21(2), 241–258. https://doi.org/10.1017/S0958344009000172


Zhao, X. (2022). Leveraging artificial intelligence (AI) technology for English writing: Introducing Wordtune as a digital writing assistant for EFL writers. *REL: Journal 0(0).* https://doi.org/10.1177/00336882221094089