

## Writing assessment in the new landscape of AI writing software – How do we respond?

**Cathrine-Mette Mork**

Natural Language Generators or AI (Artificial Intelligence) Writers came onto the scene at the onset of the Coronavirus pandemic and have since been developing at an incredible pace. They're already used in multiple industries - from journalism to marketing to creating web content. On June 11, 2020, OpenAI released a significant upgrade to its Generative Pre-trained Transformer (GPT), an autoregressive language model that uses deep learning to produce human-like text. There have been subsequent upgrades since that third-generation release. At the time of this writing, the most recent was on March 3, 2023, with the GPT 3.5 Turbo model, which was ten times cheaper than the earlier version, and GPT3.5 Whisper, an upgrade to its 2022 speech-to-text model. Language instructors need not understand all the intricacies of how the model works, but what they need to know is this: robots are now capable of writing original prose with the fluency equivalent to that of a human, as noted in an April 2022 piece by Johnson and Iziev in the *New York Times*. They are also capable of paraphrasing, summarizing, and changing the style and tone of any text fed into them, as well as suggesting changes to grammar, punctuation, and soon much more.

The imposition of new technology on education is nothing new. There was a time when calculators were not well received, as they provided us with a way to bypass the need for mental computations – seen as a form of cheating. Most of us would never think that way about calculators now. Likewise, the use of spell checkers, grammar checkers, and thesauri in word processors are accepted today without much thought, and more recently tools like Grammarly.com are often accepted for student use, albeit sometimes begrudgingly. Many of us make use of time-saving predictive text when writing messages in the Gmail app and appreciate auto-correct functions on our phones (those times when it does accurately read our minds).

At this point, it is not known to what degree or how English as a foreign language (EFL) student in Japan are currently using AI writing tools. Anecdotally, I know that many

of them have discovered machine translation (although they may not know it is AI). I believe that thanks to heavy marketing in Japan, Grammarly.com is a tool that students may reach for to improve their writing, but how many have discovered one or two among the sudden and massive influx of online AI writing systems now on offer? If they're already using one, what do they think of it? How do we predict they might make use of them moving forward? There are many questions that beg for answers and there is much research to be done, but in what follows I hope to offer food for thought when it comes to the issue of non-native writing assessment.

ESL/EFL writing instructors who find themselves in the thick of it, before any institutional or broader policies regarding AI writing tools have been made, likely fall into three camps. The first camp includes those who chose to ignore the technology and carry on as usual, possibly making use of AI detection tools at most. There are already AI tools emerging to check if work has been done by AI (see <https://openai-openai-detector.hf.space/>, <https://gptzero.me/>, <https://writer.com/ai-content-detector>, or <https://openai-openai-detector.hf.space/>) that they could make use of. Second, rather than focus on detection, many instructors will try to circumvent AI use altogether. The final group is made up of those educators who will choose to embrace AI. In this paper I am more interested in sharing ways that instructors can either circumvent or embrace the use of AI writing tools.

### **Circumventing student use of AI**

Erikson (2022) outlines three solutions to at least reduce or discourage student use of AI. All come with downsides, however. First is to require students to always cite their sources (which instructors should be doing anyway). Students must prove where their research comes from, learn good research skills and practices such as being able to differentiate between reliable and non-reliable sources, and through independent research learn something from their reading. Most AI-produced content does not cite reliable sources yet, but it is likely that many soon will. Regardless, students can search for facts on the Internet from their AI-written work and pull sources from their search results after the fact.

A second solution Erikson (2022) suggests is for instructors to make assessments

in person, and/or purely exam based. Students can be monitored throughout their writing processes and can be prevented from using third-party assistance. Original work can be guaranteed. However, this is far from ideal. As many students get test anxiety, making assessments completely test-based is arguably unfair. The in-class-only assessment model also promotes a Big Brother mindset where students aren't trusted and cheating by students is assumed, given the opportunity. Furthermore, restricting writing assessment to the classroom, even going as far as only using pen and paper, does not match up with the realities of real-world writing. Longer-form or academic writing is particularly messy and non-linear. In practice, academic writing is rarely a one-and-done activity that neatly follows a process approach to writing (see below), and it typically takes more time than the class can afford.

The third solution with pros and cons offered by Erikson (2022) is to require students to use tailor-made online platforms in which they can write and submit work. This could ensure students do not copy and paste, since instructors can see what students are doing in real-time. Also, with screen recording, instructors can see if students are using “unauthorized” software. However, such software might be expensive. Also, screen recording is invasive. It also assumes that the written work must be done in one session, teacher present, without distractions or multi-focus. This does not fit all learning types. Students can still use AI writing programs without being detected if they have other devices. Indeed, this has proven to be an issue for the author even before 2019: I have caught students checking their cell phones under the table while taking a (real-time, in-class) reading assessment on Moodle equipped with a script preventing users from wandering off the site.

My approach to teaching writing in the EFL classroom has always been process-oriented, a model that encourages selecting, reselecting, and narrowing writing topics, brainstorming, organizing ideas, outlining, drafting, redrafting, and finally, proofreading. The process approach requires learners to manage their writing by allowing them a chance to think as they write (Brown, 2020). If writing instructors focus on this approach, it is inherently difficult for students to abuse AI writing tools. The focus on process over (or at least equal to) product requires a certain amount of student reflection and requires instructors to grade on progress steps and thinking processes (journaling can be added to

the mix so that students can share their reflections). McKnight (September 15, 2022), too, argues for a process approach that fosters planning, editing, and reediting. For this reason, writing assessment that includes time-limited, exam-style writing for no audience is not a solution because it does not embrace a process approach.

During the recent pandemic, instructors were faced with an array of integrity issues, and we rose to the challenge by coming up with a variety of solutions centered on building communities around academic integrity, upholding academic integrity in course and assessment design, communicating assessment purposes with students and parents, and following up with assessment (Lee, 2020). We are surely also capable of structuring our writing assignments in ways that are more resistant to shortcut-taking using AI, such as creating writing assignments with in-class portions (Bailey, 2022). In this new age of AI, in which it can be difficult to ascertain the degree to which the student is the real writer, another approach to assessment is to focus the grading elsewhere, such as oral discussions, oral presentations, oral paper defenses, and the like. Teachers could still require students to write research papers, but they could lower the stakes. Although time-consuming, oral assessment might be appealing in situations where oral outputs are often not on par with written outputs (like in Japan).

As one can imagine, clearly there are many ways instructors can adapt to circumvent the shortcut-taking temptations that accompany AI technology: They can ask students to cite course materials, notes, lectures, and sources behind firewalls. In addition to more oral presentations and in-class writing, they can create group-based assignments that require members to integrate the whole together. They can require student work be delivered in formats that AI cannot yet produce, such as infographics, slides, posters, timelines, etc. They can get students to write about “dark” topics that AI currently refuses to do. They can prompt students to write about more recent topics that AI will not yet have full access to. They can ask students to write highly personal perspectives of highly specific writing prompts (prompts that AI does not yet excel at). McKnight (September 15, 2022) says that students need to practice “writing in which they are invested, that they care about and that they hope will effect change in the world as well as in their genuine, known readers.” Such writing would be difficult to produce via AI writing tools as the main author, and therefore is easier for the teacher to evaluate. Of course, this approach

and mindset does not come as easily to the ESL/EFL context, but it is not impossible.

It could be argued that student use of AI writing tools is not always a case of shortcut-taking and indeed might be beneficial to learning. Kellogg (2008) has argued that non-native writers allocate excesses of cognitive (working memory) resources to low-level writing tasks (word production, translation) at the expense of time being allocated to higher-level writing tasks such as organization and revision, which are essential to developing writing proficiency and producing higher-level output. Preliminary results from a study by Gayed, Carlon, Oriola, and Cross (2022) found that non-native speakers tasked with writing in English significantly benefited from a structured, researcher-developed AI writing tool called “AI KAKU.” It presumably was able to reduce the cognitive barriers EFL writers face when producing written text in English.

### **Embracing student use of AI**

There are instructors who might elect to become more progressive in their mindset towards the new writing landscape. Rather than try to circumvent the problems of tools such as ChatGPT, instead they might embrace AI and encourage its incorporation into assignments. They could choose to assign work that requires learners to analyse, review, revise, or refine AI-produced texts. For example, they could have students get the AI to write an assignment and then critique it. Learners might also compare/contrast AI and human writing and make presentations about it. McKnight (2022, September 15) agrees with this type of approach, writing that creativity and co-creativity (with machines) should be fostered in learners. This is because in her view, literacy today means being able to interact with AI as well as critically evaluate it. She argues that writers will increasingly use AI as researchers, to produce text for critique, to produce text that can be compared and evaluated, to produce and attribute routine text (such as blog content), and much more. Hellmich and Vinall (2021) assert that AI developments will lead to improved performance in writing tasks and that we must live with the reality that students are going to access AI for help with their writing. Considering this, educators can act as mediators or guides, helping students discover the strengths and limitations of these tools and better use them.

In an interview with Turnitin (Integrity Matters by Turnitin, 2022), McKnight claims we are heading in the direction of a posthuman epistemology, in which the lines between humans, nature, and technology are blurring. In line with this transition, McKnight believes that we need to rethink and reconstruct the entire competitive assessment model that we use in education, which focuses on avoiding plagiarism. She argues that we must figure out how to make fair the high-stakes assessment that often helps to determine a young person's future in an environment where students have access to all these technologies that can augment, perhaps dramatically, what the students can do. She also feels that in the future, students will probably have to determine provenance and document their processes to share with instructors as well as develop their skills of critical evaluation. They will need to be able to evaluate the quality of what AI does for them as well as be able to prime the AI to do what they hope it can do. They'll also be required to think algorithmically so that they can conceive of the limitations of AI, and then what humans can do to improve AI outputs (Integrity Matters by Turnitin, 2021). Educators would have to focus more about the quality and originality of ideas presented through the writing rather than the writing itself.

Machines are trained to predict and replicate based on finite data, but humans are much better at innovating with intent. McKnight (2022, September 15) claims that students of writing in any language will now have to go beyond basic, formulaic writing, because the machines have apparently already mastered that. Now they need to additionally learn to write purposefully and masterfully to best meet their communication goals. They will still need to learn the rules, but now more than ever they will have to learn them "so that they can break them," showcasing their problem-solving skills, judgement, creativity, empathy with readers, and social and intercultural intelligence, providing insight with humor and satire (something AI cannot readily do).

Learners are going to find the plethora of AI tools exploding onto the Internet with or without the guidance of educators, so we are probably all going to have to address it at some point very soon as individual teachers in the classroom in addition to at the institutional level. Bailey (2022) recommends that we talk to our students about these tools, asking them to reflect on how they are using them. He argues that we should get students to think about the degree to which they are in control when writing their essays;

to question whether AI is crossing that important line from editor to writer. His example questions, “Who is responsible if your grammar checker makes a mistake?” and “If your word processor changes a sentence, did you write that sentence or did it?” resemble questions posed by Mindzak and Eaton (2021), who claim that “the traditional landscape surrounding academic integrity and authorship is being rapidly reshaped by technological developments. Educational stakeholders at all levels will be required to respond and rethink definitions as well as values surrounding plagiarism, originality, academic ethics, and academic labour in the very near future.” Specific points to ponder they mention include:

- “If a piece of writing was 49 percent written by AI, with the remaining 51 percent written by a human, is this considered original work?”
- “What if an essay was 100 percent written by AI, but a student did some of the coding themselves?”
- “What qualifies as ‘AI assistance’ as opposed to ‘academic cheating’?”
- “Do the same rules apply to students as they would to academics and researchers?”

As AI technologies continue to improve, educators are going to be faced with these difficult questions and more. Since AI is going to make authorship integrity a bigger aspect of academic institutions’ plans and policies for academic integrity, plagiarism checker tools that can detect changes in a student’s writing in addition to traditional plagiarism are going to become even more important for institutions to consider adopting if the focus continues to be on plagiarism avoidance.

Whether we focus on process approaches and/or oral evaluation, or something completely different, our assessment methods need reform for yet another reason: the unfair wealth gap that this new technology exacerbates. Most of the platforms that offer AI writing services operate on a freemium system, where you get a taste of the software’s offerings, but nowhere near the quality and functionality of the offerings once you start to pay for a subscription plan. Unless schools pay for student access (unlikely in most cases), the wealth gap will manifest in academic achievement and the problem will continue to grow.

### **Moving forward**

Ultimately, we are going to have to reassess how we define writing, how we evaluate the quality and effectiveness of writing, what the steps to achieving a good piece of writing are, what skills we need to develop good writing, and probably more. AI writing technology affords all writers room for experimentation and even a bit of fun, so bringing that into the classroom could be a positive.

Goshen postulates that in a perfect world, human writers and AI would collaborate harmoniously:

*As the machine learning tech develops, writers will be able to spend less time on the writing aspect of the process and more on deep thinking – honing their creativity and crafting unique perspectives. Writers with innovative ideas in content writing will be the ones who will benefit from this inevitable shift in the industry. And the world will benefit from their ideas — expressed with clarity and concision (2022).*

The above is indeed utopian, and since we ultimately have no direct control over the progress of AI, use of an optimistic lens through which to look at AI writing tools is tempting. It is probably prudent to take off the glasses, though. Late last year there were already reports of and confessions from students who had discovered AI writing technology and started using it to do homework for other students and make money in the process (Dysart, 2022,). And future potential problems indirectly related to education are already predicted. Goshen also observes, for example, that although AI won't replace writers, the writers who use AI will replace the writers who don't. A two-year study by Manyika et al. (2022) claims that intelligent agents and robots could replace up to 30% of the world's current workforce by 2030. More ominously, OpenAI's managers were concerned about the generator being used to create fake material, such as reviews of products or election-related commentary. So much so, in fact, that they had originally refused to release GPT-3 (Mak, 2019). While the creation of fake media and its repercussions are outside the scope of this paper, Mondschein claims that "professors have nothing to worry about: artificial 'intelligence' is, in fact, quite stupid." While he admits that AI can produce a paper "about on par with an uninterested first-year student

who went to a good school system and turned out an essay by rote, maybe looking up the assigned reading on CourseHero or something similar,” it falls short when asked to generate anything more specialized. It makes mistakes, makes things (facts) up, seems relentlessly politically correct and optimistic, and writes robotically (ha!) and repetitively, without wit or humor (2023). For how much longer will that be the case? And is not the fact that it at present makes things up and is “overly politically correct” disconcerting? And what about ESL/EFL teaching and learning? I personally *am* bothered that AI can write a paper “about on par with an uninterested first-year student,” because those students might be like many of my 4<sup>th</sup> year, senior thesis-writing, EFL students.

I decided it would be fun to consult an alternate perspective, or more accurately, an alternate *source*, on the future of writing assessment. Wouldn't it be interesting to read what OpenAI's ChatGPT-3 itself had to say? I plugged the following into <https://chat.openai.com/> (this was in January 2023):

*What changes should university educators make regarding writing assessment in the new landscape of advanced AI writing software?*

Here is its initial response (ChatGPT-3 users can always refine by asking for length, style, and content changes later), produced within seconds (in stark contrast to the time I have spent writing so far):

*There are a few key changes that university educators may want to consider regarding writing assessment in the face of advanced AI writing software:*

1. *Emphasize the importance of originality: As AI writing software becomes more advanced, it is increasingly important for students to understand the importance of producing original work that is not simply a reassembled version of existing content. This may involve emphasizing the importance of proper citation and avoiding plagiarism.*
2. *Focus on higher-level writing skills: While AI writing software can be very useful for proofreading and correcting basic grammatical errors, it is not as effective at helping students develop more advanced writing skills such as*

*critical thinking, analysis, and argumentation. Educators may want to focus on these higher-level skills in their writing assessments to help students develop more robust writing abilities.*

3. *Encourage the use of AI writing software as a tool: While it is important for students to understand the limitations of AI writing software and to focus on developing their own writing skills, it can still be a useful tool for improving the quality and clarity of their writing. Educators may want to encourage students to use AI writing software as a way to identify and correct basic errors, but to also be mindful of its limitations and not rely on it too heavily.*
4. *Consider alternative assessment methods: In addition to traditional written assignments, educators may want to consider using alternative assessment methods such as oral presentations, group projects, or online discussions to help students demonstrate their writing skills in different contexts.*

At the time of writing this paper, OpenAI does not have access to data more recent than 2021(ChatGPT FAQ); it cannot release digital spiders out to probe the more recent publications I discovered when writing this paper. Of course, that is one of the many things that will likely change with the release of GPT-4, which has already come out in March 2023 for beta testing. But one thing is certain: it appears from the above that I am not alone in trying to keep up with the literature. ChatGPT answered adeptly, all things considered.

I believe this new AI is indeed disruptive to writing assessment; we will have to change at least some of our teaching and evaluation methods, whether we try to prevent student use of AI tools or encourage them to embrace it. There will always be free riders in all aspects of any given society, and shortcut-taking students are no exception. Older types of services exemplified by companies such as Chegg, Course Hero, and Coles Notes offer students, for a price, the temptation to bypass the challenging processes of reading, understanding, thinking, and writing. Such companies now have stiff competition that can produce work for students at lightning speed, and at least for the moment, at no cost. The new AI competition simply makes the issue harder to ignore, pushing educators once more to refocus on the ultimate goals of student writing – to demonstrate understanding

and judgement.

When I worked as a level one EIKEN (*Jitsuyo Eigo Gino Kentei*; Test in Practical English Proficiency) examiner, I noted how the rubric was constructed in a way that a test-taker who didn't quite have the ideal proficiency levels in vocabulary, grammar, and pronunciation could still pass, so long as they had something of value to say, whereas a test-taker with flawless skills who produced nothing of substance could still fail. I feel this is as it should be. Work produced by AI tends to produce essays that may show a lack of regard for the truth and/or demonstrate a lack of understanding. Moreover, AI writing, unsurprisingly but importantly, simply does not care (White & Skorburg, 2023). Even as ESL/EFL educators, for whom the mechanics of writing may seem more important, we must reflect on why we are teaching writing skills to begin with, and address forebodings about fairness and integrity. We should concern ourselves with *how*: How we can encourage and maintain academic integrity alongside the proliferation of fleets of intelligent robot writers, how we can best help students to work with AI writing tools to help them write better and produce better work, how we can properly credit the software when used, and finally, how we are going to be able to work with AI writing tools in the future, as the technology continues to further improve, in a way that maintains authenticity and legitimacy.

## References

- Bailey, J. (2022, December 2). *How teachers can prepare for AI-based writing*. Turnitin. <https://www.turnitin.com/blog/how-teachers-can-prepare-for-ai-based-writing>
- Brown, H. D., & Lee, H. (2020). *Teaching by principles: An interactive approach to language pedagogy* (7th ed.). Pearson Education.
- ChatGPT FAQ*. OpenAI Help Center. (n.d.). Retrieved January 6, 2023, from <https://help.openai.com/en/articles/6783457-chatgpt-faq>
- Dysart, J. (2022, October 9). *Good at ai writing: Will cheat for cash*. Robot Writers AI. <https://robotwritersai.com/2022/10/10/good-at-ai-writing-will-cheat-for-cash/>

- Erickson, S. (2022, August 12). *Ai writing software will destroy the current school system*. Medium. <https://medium.com/illumination/ai-writing-software-will-destroy-the-current-school-system-8c8596e679ed>
- Gayed, J. M., Carlon, M. K., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, 3, 100055. <https://doi.org/10.1016/j.caeai.2022.100055>
- Goshen, O. (2022, October 25). *Ai's role in the future of content writing*. AiThORITY. <https://aithority.com/content/ais-role-in-the-future-of-content-writing/>
- GPT-4. (n.d.). <https://openai.com/research/gpt-4>
- Hellmich, E., & Vinall, K. (2021). FL Instructor Beliefs About Machine Translation: Ecological Insights to Guide Research and Practice. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 11(4), 1-18. <http://doi.org/10.4018/IJCALLT.2021100101>
- Integrity Matters by Turnitin (2021, June 2). *Anticipating the impact of AI-based writing on education and assessment* [Video]. YouTube. <https://www.youtube.com/watch?v=oEvvi-0Wggc>
- Johnson, S., & Iziev, N. (2022, April 15). *A.I. is mastering language. should we trust what it says?* The New York Times. <https://www.nytimes.com/2022/04/15/magazine/ai-language.html>
- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1(1), 1–26. <https://doi.org/10.17239/jowr-2008.01.01.1>
- Lee, C. (2022, December 22). *How to uphold academic integrity in remote learning*. Turnitin. <https://www.turnitin.com/blog/how-to-uphold-academic-integrity-in-remote-learning>
- Mak, A. (2019, February 22). *When is technology too dangerous to release to the public?* Slate Magazine. <https://slate.com/technology/2019/02/openai-gpt2-text-generating-algorithm-ai-dangerous.html>

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., Ko, R., & Sanghvi, S. (2022, April 5). *Jobs Lost, jobs gained: What the future of work will mean for jobs, skills, and wages*. McKinsey & Company.

<https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages>

McKnight, L. (2022, September 15). *To succeed in an AI world, students must learn the human traits of writing*. The Conversation. <https://theconversation.com/to-succeed-in-an-ai-world-students-must-learn-the-human-traits-of-writing-152321>

McKnight, L. (2022, October 14). *Eight ways to engage with AI writers in Higher Education*. THE Campus Learn, Share, Connect.

<https://www.timeshighereducation.com/campus/eight-ways-engage-ai-writers-higher-education>

Mindzak, M., & Eaton, S. E. (2022, September 15). *Artificial Intelligence is getting better at writing, and universities should worry about plagiarism*. The Conversation.

<https://theconversation.com/artificial-intelligence-is-getting-better-at-writing-and-universities-should-worry-about-plagiarism-160481>

Mondschein, K. (2023, January 3). *Avoiding cheating by AI: Lessons from medieval history*. Medievalists.net. <https://www.medievalists.net/2023/01/chatgpt-medieval-history/>

White, D. J., & Skorburg, J. A. (2023, February 28). *ChatGPT killed the student essay? philosophers call bullshit*. The Conversation. <https://theconversation.com/chatgpt-killed-the-student-essay-philosophers-call-bullshit-200195>