

Large language models and text generators: an overview for educators
Ryan Morrison & Generative Pre-Trained Transformer (GPT - 2 & GPT - 3)
September 1, 2022

Abstract: Large Language Models (LLM) -- powerful algorithms that can generate and transform text -- are set to disrupt language learning education and text-based assessments as they allow for automation of text that can meet certain outcomes of many traditional assessments such as essays. While there is no way to definitively identify text created by this technology, there are patterns that educators can use to adapt assessments to minimize the impact that these tools will have on academic integrity. This document provides an overview of the technology and how it is being utilized by publicly available platforms, some of which are targeting education; samples and analysis of the idiosyncrasies of text generated and transformed by LLM platforms; and suggestions on how educators can adjust their approach to language and text-based assessments to better meet the needs of students in a world where LLM tools become ubiquitous.

Table of Contents

Foreword.....	3
Glossary of Terms.....	4
Frequently asked questions.....	6
Document guide.....	9
GPT-2 Paragraph Examples and Analysis	10
GPT-3 Multi-paragraph Examples and Analysis.....	12
GPT-3 Purpose-focused Examples and Analysis.....	23
Responses.....	29
Conclusion.....	32
Acknowledgements.....	35
Disclosure.....	35
References.....	36

Large language models and text generators: An overview for educators

Foreword

In February 2019, when the existence of GPT-2 and “better language models” were exhibited by OpenAI– a non-profit Silicon Valley organization with the mission to keep powerful algorithms out of the exclusive hands of governments and corporations -- it was apparent that there were certain inevitabilities of this Artificial Intelligence (Radford et al., 2019). Now, some of these inevitabilities have already come to fruition at the time of writing. Text-based education will be impacted – no doubt -- and student outcomes that we have long valued, particularly in language arts, are becoming automated. Academic integrity issues may become pervasive and the effect on assessment could be significant. On the flipside, this class of AI, which has become known as Large Language Models (LLM), could be used to address the problem of languages, i.e. facilitate better communication between speakers of different languages. LLM can serve to automate rudimentary communication and provide models and interaction for language practice. This does not mean that we will have to stop teaching our current outcomes nor teaching languages, rather we need to change how we are teaching and assessing them – this sentiment has been echoed by many of educators in various discussions.

As of the time of writing, the “Write My Essay” button has yet to be invented – this could be a matter of time as the citation generators have been commonplace for some time, and theoretically, integrating an LLM interface into a platform that automatically writes an essay is not an absurd proposition. When combining a variety of the currently available tools including citation generators, it is possible to produce a passable essay in short order. The hope is that by the time someone publishes a platform that is singularly able to automate traditional text-based assessments such as essays and short answer tests, most of educators will have started in earnest to move away from assessing outcomes in these vague perfunctory modes in favour of more meaningful and authentic assessments.

Beyond the horizon, LLM stand to revolutionize communication; they appear to be the near end-game tool of Natural Language Processing (NLP), the branch of computer science concerned with input and output of communication between humans and computers. However, written and spoken communication between humans may be assisted by this technology, but students will always need the skills that we impart in our classes. In fact, it is probable that we will need to begin integrating discussions of our relationships with algorithms with our students sooner rather than later and training teachers to be able to identify and discuss algorithms and algorithmically generated language (Abd-Elaal, et al., 2022).

This document was created with three purposes. The first is to increase awareness of large language models in education, the second is to offer a primer in identifying algorithmically generated and transformed text, and the third is to encourage continued discussion of the intersection of modern algorithms and humanity at the educational level. I hope you find it useful.

Glossary of Terms

Large Language Models – Large language models are the underlying technology that powers text generators, the platforms that this paper primarily explores. LLM are centrally powered algorithms and composed of billions of parameters engaging with hundreds of billions of words. They have been integrated into many familiar platforms, such as translators, conversation bots, search, text suggestion, and as mentioned, text generators. For example, if you use Microsoft Office products, you may have noticed that Outlook and Word now provide suggested phrases to complete your text when you are writing, but if you are composing offline, this function disappears. Why does this occur? The suggested text is being created by powerful, distant computation using an LLM. The most prevalent LLM being used currently is OpenAI's GPT-3, and texts created by platforms using this LLM were making news for their fluency in 2020 (Macauley, 2020; Lyons, 2020; Porr & GPT-3, 2020). More recently, Meta (formerly Facebook) and Hugging Face (an LLM start up and OpenAI collaborator) have announced that they are releasing an equally powerful LLM -- but unlike GPT-3 which works with limited access for unvetted developers – they are publishing all of the accompanying documentation and parameters for public access; thus, anyone can now learn how to make their own LLM (Zhang, et al., 2022; Hugging Face, 2022). As a note, the name *Large Language Models* is still contentious as there is no quantification what constitutes 'large'. However, the argument follows that the larger versions of concern clearly outperform earlier, smaller language models.

Text Generators/ Transformers – Text generation and transformation are functions of large language models, and they are a particular point of interest for educators. With the input of a prompt, text generating platforms will create novel text, i.e. it is not identical to any current existing text. When the input is a significant amount of text, text generators with the correct presets can serve as summarizers or paraphrasers and can transform text. Although the text is comprehensible, it still may require several attempts, curation and 'smoothing' to generate an ideal result. Much like 'naturalization' has become a common term in the business of translation since the proliferation of automated translation, 'smoothing' is becoming the common term for adjusting automatically generated text, which like human generated text, often contains minor errors.

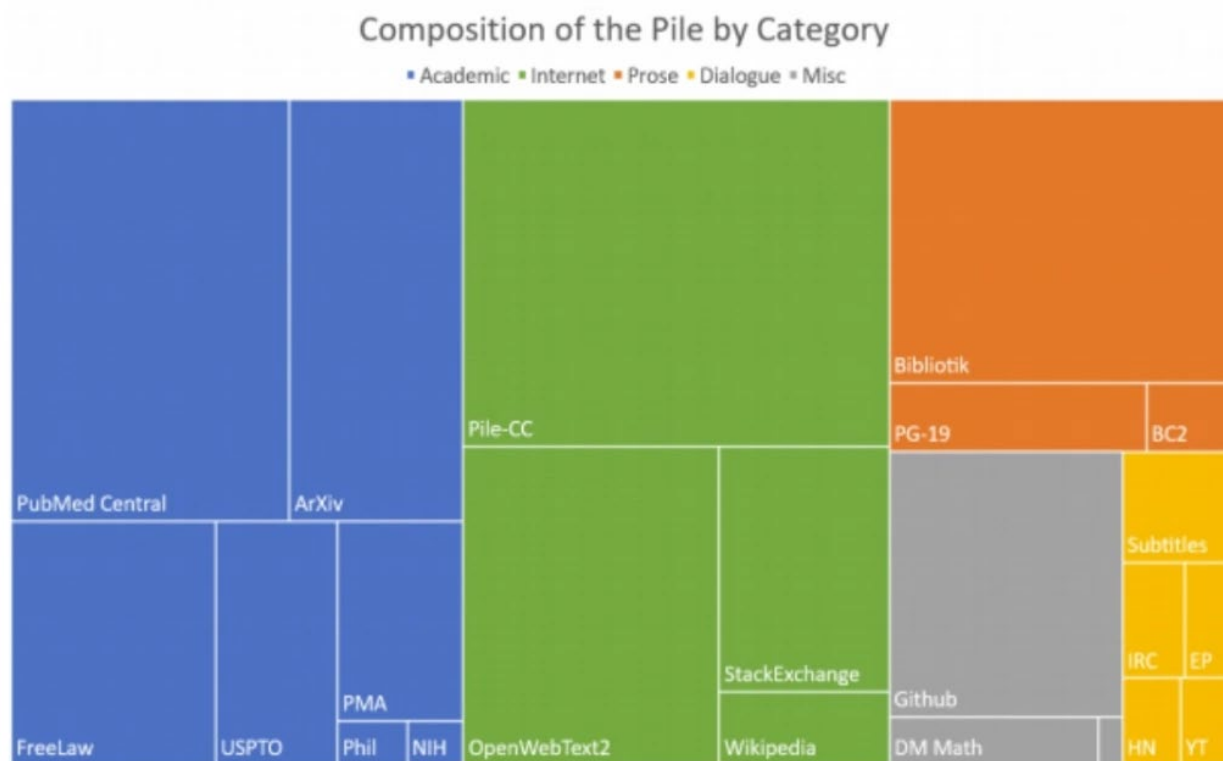
Platforms – Generally, a platform is an interface for a digital technology, and for text generators, there are hundreds of publicly available platforms. They are available mainly as web pages, but there are also mobile apps available that perform the same tasks. Here are a few text generating platforms that are referred to in this document and utilize GPT-3:

- <https://beta.openai.com/playground> - From the developers of GPT-3, OpenAI, this is a basic text generator with access to several versions of previous iterations of GPT-3.
- <https://www.copy.ai/> - A platform aimed at copywriting; further, it has the most diverse preset text types including 'Essay Outline' and 'Essay Introduction'. It is now available for free (with sign up) and is monetizing with a pro account model.

- <https://inferkit.com/> - A basic text generator that includes slider options like “temperature” that control the ‘randomness’ of GPT-3.
- <https://www.writefull.com/> - Targeting post-graduate academics, this site has several functions to explore, including creating abstracts for publishing in academic research journals.
- <https://speedwrite.com/> - A summarizing platform targeted at students, this site summarizes, spins (paraphrases) and creates text based on inputs of 100 words or more, and includes a Wikipedia plug-in.

Since these are all powered by the same LLM, it costs money for the platform to generate text -- \$10 USD for 300,000 words (Macaulay, 2020). Thus, each site offers some variation of a limited free trial or a free/premium option to cover their costs. It is important to note that these are only five of the hundreds of platforms using this this technology to generate text in some form.

Source Text – LLM are based on open-source text from compiled into a database. The most commonly used database is from Eleuther and is referred to as ‘the Pile’. It is half a trillion words and is 800 GB in size (Gao, et al., 2020). The exact sources and their categorization is represented visually in the diagram below.



(Image: Gao et al., 2020)

Parameters – The parameters are the connections that are made by ‘training’ the algorithm. To visualize this, think of phrase “it has come to...”, and then think of the common completions: “our attention”, “fruition”, “pass”, etc. The largest large language models have hundreds of

billions of parameters, and the output is the result of these parameters, which can be customized depending on the platform. For instance, InferKit provides a ‘sampling temperature’ slider, which “Controls the randomness of sampling—the ‘creativity’” (InferKit, 2022).

Turing Test – Named for one of the founders of computer science, Alan Turing, a Turing test is the measure of a computer’s ability to imitate a human. If the observer cannot identify that an output is computer generated, the computer has passed the Turing test. This is a hallmark of AI, and text generated by large language models consistently pass Turing tests.

OpenAI – Created by Silicon Valley magnates with the intention of keeping AI technology available to everyone -- keeping an even playing field for all nations and corporations – OpenAI pioneered large language model technology. The Generative Pre-Trained Transformer series (GPT-2 and GPT-3) were released less than a year apart and have served as the base for most other large language models. You’ll notice that all of the generated and transformed text in this document was created on platforms that use GPT-2 and GPT-3, and therefore those programs are included as authors in this document. This was done purposefully as the question of the nature of authorship is one of the possible impacts of text generating technology.

Natural Language Processing (NLP) - The branch of computer science concerned with input and output of communication between humans and computers.

Frequently Asked Questions

How do text generators/ transformers operate?

When the user inputs a prompt, most platforms take a few moments to produce results. In these moments, the large language model algorithms are making word selections based on the common associations with the input. The grammar and mechanics are not actually programmed into the algorithm; rather, they have been trained to follow the patterns of natural language in the source text. This training works by identifying themes and following patterns that are common in those themes from the source text. For example, if the input follows patterns of conversational quoted speech, the output will follow in this same mode (granted the settings don’t negate this), repeating the speakers’ names and continuing the conversation.

The resulting text can be curated and altered by the user, as well as re-generated to get the text ‘back on track’ should it drift away from the desired result. In this document, moments where the curator has had to re-generate are noted, and thematic drift is also noted as it follows some noticeable patterns.

Is there a way we can identify algorithmically generated text, like Turn-it-in detects copy-paste plagiarism?

No -- since the text is novel, it won’t show up in text-matching database like Turn-it-in or SafeAssign. The only chance to automatically detect algorithmically generated text would be to create a detection algorithm could identify patterns used in large language models. Researchers

attempted this with GPT-3's predecessor, GPT-2 in 2019. GPT-2 had much fewer parameters and produced less sophisticated text (See 'GPT-2 Paragraph Examples and Analysis' section). Therefore, a detection algorithm was feasible, and NLP researchers at MIT and Harvard were able to produce a functioning platform (GLTR from MIT-IBM Watson AI Lab and Harvard NLP, 2019). However, there is no functioning version for this GPT-3, for a few reasons.

For one, the algorithm would have to account for all large language model parameters and their source material, and so, this theoretical detection algorithm, at the time of writing, would be very expensive to run and require constant updates (each time a new LLM is publicly published). It would be very expensive because it would require the combined energy costs of all currently available LLM, which is significant (Bender et al., 2021). GPT-3, for example, has a per-word price point attached to their API to cover their computing costs (Macaulay, 2020). The detection program would likewise also require a per-word fee associated to it. It took slightly more than a year for OpenAI to release GPT-3 after its ground-breaking predecessor, and it took less than one year after that for Microsoft and Nvidia announce they had trained a large language model with three times the parameters of GPT-3, which also includes "common sense reasoning" (Wiggers, 2021).

Another reason why automatic detection is virtually impossible is that there would need to be a word count threshold for the theoretical detection algorithm to work. One of the most apparent applications of text generators, based on the number of purpose-built platforms, is for copywriting, i.e. creating short slogans and content for marketing and advertising. A simple phrase or sentence created by a text generator is not identifiable as such; it could just as easily have been created by a human. For the copywriter's purpose, generated text may be off-topic, or it may be brilliant. Either way, the curated text of the copywriter won't appear as generated to a detection or algorithm because there is nothing noticeably divergent from human generated text in isolated sentences. Perhaps, beyond a small paragraph, a detection program could determine with some probability whether or not a text was generated by a specific algorithm, but again, there are many of these algorithms with their own idiosyncrasies, so accounting for this would be profoundly difficult to create at the time of writing.

How close to human writing is large language model-based generated text?

It is very difficult to tell the difference between human and algorithmically generated text. Anecdotally, I've done modified Turing tests with several participants and there have been few unanimous identifications of whether a text was algorithmically generated, though there are some patterns that many participants identify (expanded on later). That being said, the examples used in the modified Turing tests are curated examples with the express purpose of creating compelling Turing tests – they are the best versions of outputs which are still random at times.

Another anecdotal finding is that when two texts are compared, one being algorithmically generated and the other human generated, the results skew towards humans being able to identify the algorithmically generated text. However, when an algorithmically generated paragraph is integrated into a longer, human generated text, it is much more difficult to

differentiate. This parallels preliminary studies being done elsewhere. A group of community college teachers were tasked blind marking assignments and although GPT-3 based text achieved a B average in writing historical essays, it failed to create creative narrative pieces that achieved above a failing grade (Macaulay, 2021). A few caveats to this study were that it was informal, and the customization tools were not documented – GPT-3 might be better at creative writing if the curator was more familiar with customization tools.

Some of the first instances of algorithmically generated text appearing in publication and preliminary studies demonstrate that there are still some language tasks that GPT-3 are not proficient at completing. The GPT-3 Guardian article (Porr, 2020) explicitly acknowledges the role of the curator. Other examples show that GPT-3 generated text is good at creating ‘think-piece’ style content. An example of this was a GPT-3 generated piece which was submitted to Hackernews.com, and was then upvoted to the top of a list of user-submitted articles (Lyons, 2020). The caveat here is that members of this site could have agreed to upvote the article for whatever reason.

What are some of the ethical considerations of LLM?

Google’s former Head of Ethics Timnit Gebru co-wrote the seminal and now infamous criticism of current Large Language Models (Bender, et al., 2021), which lays out the ethical issues of LLM quite thoroughly, and the following explanation is a largely selective summary of that paper.

Energy and cost are key issues that come along with LLM. If the servers are not on a sustainably sourced electrical grid, there would be a sizable amount of carbon produced because of ubiquitous use of LLM. This is part of a broader issue with AI level computation and modern energy heavy processes, like that of mining Bitcoin (Bender, et. al., 2021). The confluence of these demands has resulted in the need for increased production of powerful processing chips that were, until recently, mainly used commercially in the video gaming community. As a result of this cost, OpenAI sells words produced by GPT-3 on a per-word basis to cover costs (Macaulay, 2020). This creates an access barrier for people who do not have the means to pay for these platforms. However, there is still the question of the effects of technology companies’ integration of LLM into their existing products in the form of auto-complete suggestions, translation and search. While the new in-house versions of LLM claim to be more powerful and accurate than GPT-3 (Wiggers, 2021), they remain behind the curtain of corporate intellectual property, and are thus subject to less scrutiny than the publicly available versions.

Although Gebru was unceremoniously terminated from her position as a result of not withdrawing the paper (Simonite, 2021), her stance is vindicated by some of the text generated by GPT-3 in this document (See Text 3.1). This problematic text follows the axiom of ‘garbage in, garbage out’ associated with NLP, and in this case, since some of the open-access texts in the Pile are racist, sexist, homophobic, classist, etc., there are occasional instances where text of this nature is generated – remember that its source content is text from the broad internet, which can be admittedly toxic. However, this is rare as GPT-3’s creators have added filters and trained it to create text that is aligned with socially acceptable content, though this type of work

needs to be continuous (Ouyang et al., 2021). Case in point, when one tries to create toxic and problematic text intentionally, it is often the case that the text drifts rather quickly to more 'comfortable' adjacent topics. Though with the technology being open-access as per Meta's announcement of their competing GPT-3 LLM with full documentation available (Zhang et al., 2022), someone could theoretically create their own LLM and platform without these filters, granted they have the monetary means to run an LLM.

The ethical implications of text generating technology on education is the key concern of this document. The possibility that a student can have large portions of their writing completed by an undetectable algorithm transcend the practical issues surrounding academic misconduct. This technology not only presents ethical considerations for academics, but also copyright and the nature of authorship (Mindzak & Eaton, 2021). The inevitable deep level questions beget responses that may require foundational changes in text-based industries.

Document Guide

Pre-text – Using various platforms, the following prompts and inputs were used by the curator to create the output texts. Re-generation, number of attempts, the platform, customization, and the intention of the curator are typically noted – these actions are examples of curation.

In-text – Phrases and paragraphs that have had ~~striketrough~~ applied were left to demonstrate some of the obvious inconsistencies in generated text. There are also authentic texts in the summarizing sections and are followed by summaries generated by GPT-3 based platforms.

Post-text – One of the purposes of this document is for the reader to see common patterns in generated text, draw inferences to serve as discussion primers about the nature of GPT-3 generated text. As such, there is analysis after the text that aims to highlight some of these common patterns. Granted, this analysis is from one perspective, and alternate versions and interpretations are equally valid. 'Stylistic elements of note' are added, but they are not part of over-arching patterns, but rather interesting moments of particular texts.

Section 1: GPT-2 Single Paragraph Examples and Analysis

GPT-2 was the predecessor to the ubiquitously used and available GPT-3, so the platforms used to generate this text are now offline. GPT-2's release was limited by OpenAI to only a few developers, and it took several months for those developers to have publicly facing platforms.

Section 2: GPT-3 Multi-paragraph Examples and Analysis

The platforms used to create this section were preset to generate multiple paragraphs from prompts, without specific purposes. As such, they have various stylistic idiosyncrasies in their texts. In some cases, particularly while using InferKit, settings were manipulated to curate more accurate and smoother text in successive attempts

Section 3: GPT-3 Purpose-focused Examples and Analysis

The texts generated in this section were created with stylistic presets from a variety of platforms. For example, 'social media post', 'blog post', 'summary' and 'essay introduction' are typical examples from sites like copy.ai and speedwrite.com.

GPT-2 Single Paragraph Examples and Analysis

Text 1.1

Platform: Talktotransformer.com (offline)

Number of attempts: 3

Intention: Investigate narrative writing using the second person voice

Prompt: "You are lost in a forest"

Output: You are lost in the forest. You've gone off trail... but you need to find your way back. You've been lost for five days now. You need to find your way back. As you roll along you begin to hit a rock head on. You quickly stick out your hand. You find your way back to the trail. Just to the north is a hickory grove. As you are looking around you see that you were not the only lost, that there was a group out there heading to the west. You decide to head that way. Your spirits rise. You might have been lost for hours.

Analysis: GPT-2 being a forerunner to the latest large language models definitely lacked some typical style elements typical of academic writing. In this case, transition strategies are sparse. There are also repetitive ("You need to find your way back") and illogical moments ("As you roll along you begin to hit a rock head on.").

This simplicity of the initial sentences is another commonality of automatically generated text. It tends to begin with simple sentences and then prefers complex strategies (with the subordinator in the initial position) to compound strategies. Although simple declarative sentences are not typical of the contemporary professional writer, it is a precedented creative style of early 20th century American literature (Glass, 2021).

Text 1.2

Platform: Talktotransformer.com (offline)

Number of attempts: 2

Intention: Investigate reactionary, anti-social text generation

Prompt: "We are seeing the end of human civilization"

Output: We are seeing the end of human civilization. And this end of the world does not have anything to do with the crash of the world economy. "In this financial crisis, the capitalist system has again, re-emerged. Of course it is falling apart. Not without cause. It has created enormous production demands for all sorts of commodities. You get them for the "big stuff," i.e. raw

materials, for fossil fuels. "Big stuff" isn't hard to come by, even if you aren't a capitalist. But no capitalist is going to invest money if there are not going to be buyers for the goods they are producing. So, you get an economic depression that requires a credit system.

Analysis: This is a pretty logical production of text with connections of sophisticated ideas, but of course, it's stretching logic in some moments. I, for one, have never heard commodities referred to as "big stuff". There is also the contention that GPT-2 might not have been able to create the counter, right-wing approach to this topic, but it is impossible to say as this model is now offline. One grammar error of note is that the initial quotation mark is present in the second line, but it is never closed nor attributed. Just a general reminder, all attributions made by current text generators are untrue, but occasionally, a coincidental historical fact will occur as far as attribution (see Text 2.1).

Stylistic elements to note

- The use of double quotation marks to highlight phrases of concern
- The correct use of the *id est* acronym -- "i.e."

Text 1.3

Platform: Talktotransformer.com (offline)

Number of attempts: 1

Intention: Investigate text generated based on common sense advice.

Prompt: "Eat everything on your plate"

Output: Eat everything on your plate. You need to eat your veggies. There are no other foods that are more central to human food habits than vegetables. It's a deeply instinctual response, which is why it's both difficult to resist and easy to see how it is that some Americans, as well as many in other countries, seem to have no problem consuming massive amounts of them. What is hard to comprehend is how bad Americans truly are at sating their craving for carrots and broccoli, for example. When pressed, even most nutrition-aware Americans will quickly admit that they only eat vegetables because their parents ate them and they like them. "It's just the way I am," they will say. But that's clearly not how it is for most Americans.

Analysis:

This is a great example of a self-contradictory response. The third and fourth sentences are contradicted by the fifth, and the final sentence (sixth) contradicts the fifth. There is also the pattern of the simple sentences (one and two) being followed by more and more complicated sentence strategies incorporating subordination, adjective clauses, noun clauses, and even quoted speech.

GPT-3 Multi-paragraph Examples and Analysis

Text 2.1

Platform: InferKit.com

Number of attempts: 1

Regeneration: None

Intention: Explore the ability to be specific to actual people and their accomplishments.

Prompt: “Who is the pre-eminent academic integrity researcher in Canada?”

Customization: *Try to Include these words – “Sarah Eaton”*

Output: That’s a good question. All the pre-eminent academics i know have been, at some point, forced to submit themselves to academic integrity audits. Some have complied reluctantly, thinking their methods should not be challenged, despite the serious ramifications. Others have risen to their defense, to say they have used rigorous scientific methods to rigorously interrogate data, and the one person who hasn’t been able to convince academics is Sarah Eaton. She has dedicated herself to studying academic integrity, and she has done a wonderful job of it. In fact, she has written an absolutely brilliant piece in The Conversation on academic integrity, and it is well worth a read.

Who is Sarah Eaton?

Sarah’s definition of academic integrity is very clear and concise: “We see academic integrity as a process. It is a step in research development that looks to improve the quality of what we do with data, and also seeks to enhance how we represent our research in public discourse.”

I knew Sarah’s name when I read about her in the Globe and Mail in 2012. My Phd thesis was on the adoption of the Liberal party during the 2006 Ontario election. My research focused on the pre-campaign period, and in particular the research approach of one of the liberal party leaders, Dalton McGuinty, who was a former MPP. Prior to the 2006 election, McGuinty had proclaimed publicly he did not want any political donation or campaigning money from teachers.

My research identified three reasons for McGuinty’s change of position: first, teachers were concerned about getting into political trouble if they were caught breaking campaign finance laws by accepting donations, and second, teachers valued being part of the government process. Also, third, teachers thought they could be better represented if the government were more open to public discussions and debates on education policy.

Analysis: In this text, a seemingly inexplicable real-life, historical fact was produced: Sarah Elaine Eaton has indeed published in *The Conversation* about academic integrity (The Conversation, 2022). When I presented this to her, she was humbly adamant that the surrounding text be identified as “editorial”. The assumption here is that educational topics (i.e., academic integrity) research and the Canadian context are related to the publication, *The Conversation*, and thus, the incidental connection is made.

However, after the initial paragraph, an apparent heading “Who is Sarah Eaton?” appears, and we see more topical drift. It is interesting too that Sarah’s definition of academic integrity (which is inaccurate here) occurs after the heading rather than immediately describing whom she is. In the third and fourth paragraph, the topical drift becomes very clear. Where it drifts to is of interest and a common pattern: when education is involved in a Canadian context, provincial politics are a popular topic for GPT-3, and the former Ontario Premier Dalton McGuinty is mentioned. As is often the case, the final paragraph is scarcely related to the input text.

Stylistic elements of note:

- Improved sentence complexity and sophistication
- The redundant “That’s a good question” at the outset is a typical ‘word count’ padding phrase.
- The ‘rule of three’ list in the final paragraph (phrased in parallel structure) is commonly taught in composition classes.
- There is an occasional lack of capitalization in this output – “i” in the first paragraph and “McGuinty” in the third. However, in other instances these are capitalized.
- Irregular expression of “Phd”

Text 2.2

Platform: InferKit.com

Number of attempts: 1

Regeneration: None

Intention: Explore the viability of using text generators for discussion board questions. The prompt was an authentic discussion board question for a professional development course I enrolled in.

Prompt: “How has honouring learner variability and/or social identity been important to you and how has this experience changed your thinking and actions?”

Output: It is true that we do have differences in the way we learn and it is up to us to be aware of that and acknowledge it. We shouldn't be trying to force learners into learning the way that we want them to learn it because that does not foster authentic learning. I believe learners have a role to play in building their own learning style which is how we actually become more

successful learners. I think the reason for this is because not everyone learns the same. You can have similar learning outcomes but if you do not have the capacity to have a consistent learning pace then it is not going to be successful for you.

Analysis:

As far as discussion board posts go, this would be a passable albeit underwhelming submission. Personally, I think that as a participating student or as an instructor, this would be glazed over for its vagueness, but the participation of the 'author' would be noted and graded.

This raises the question of whether a blasé attempt at participating in a text-based discussion board would be acceptable in the world of online learning where text generators are ubiquitous.

Text 2.3

Platform: Inferkit.com

Number of attempts: 3

Regeneration: None

Intention: Investigate if text generator can logically justify un-popular and untrue arguments.

Prompt: "Global warming and climate change are not the result of humans."

Output:

If we believe the current theories on the cause of climate change, they are caused by sunspots, waves in the jet stream, El Nino, ocean currents, volcanic eruptions and other natural processes. There is no scientific theory to support the theory that carbon dioxide in the atmosphere and other greenhouse gases released into the atmosphere cause global warming and climate change.

The Paris climate agreement is a questionable legal document. According to Thomas Jefferson, one of the signers of the Paris agreement, the "authority of this instrument consists in nothing more than an agreement among governments, and not in a legal or political constitution." In other words, the agreement is meaningless because it is not an international treaty ratified by the Senate.

Climate change will not affect one's health. Current theories on climate change predict that sea level will rise, although this has not happened yet. There is no scientific reason to believe that greenhouse gas emissions cause an increase in atmospheric temperatures. Warmer temperatures in the Northern Hemisphere have resulted in record crop yields.

~~In the 50s and 60s, I had a friend who had a mean dog named Duke. He knew that Duke did not like people so he would spray him with water when he saw him. Duke would charge at the hose and let the water hit him, giving him an opportunity to be sprayed again.~~

Would a gasoline-powered vehicle hold up if hit by a Ford Model T? Of course not.

SURPRISE! The greenhouse gases that are in the air are not as bad as the world is made out to be. They are a necessary part of living in a civilized society. Like breathing, they have a purpose.

SURPRISE! The system of global warming is based on flawed theories.

Analysis:

The first two paragraphs are clear and well-argued but slightly flawed. It uses climate related vernacular, but it does not provide deeper explanations of what it posits. The first sentence starts logically, and the text stays on topic throughout; however, the content again lacks accuracy and verifiable attribution. Obviously, Thomas Jefferson signed an important historical document, but not the one in question, and indeed treaties get ratified, but not by a Senate in this case (assumedly the referring to the legislative process of the USA). By the third paragraph, the arguable connections are few and the explanations disappear.

The strikethrough paragraph seems to be an anecdote in an attempt at a historical anecdote for climate change. It is jarring to include it in the response, but the notion of using an anecdote as justification for an illogical claim is a common -- if not fallacious -- argumentative strategy.

By the end, logic dissipates with the claim that a Model T Ford, an early 20th century car, would damage a gasoline-powered vehicle, which of course, a Model T Ford is an example of. The rhetorical questions and addition of the "SURPRISE!" seems to give it a spoken, colloquial style, in an apparent shift from what started like academic or journalistic writing.

Stylistic elements of note:

- Anglicized characters in foreign loan words "El Nino"
- All-caps rhetorical mode of sarcasm – "SURPRISE!"

Text 2.3

Platform: Inferkit.com

Number of attempts: 2

Regeneration: None

Intention: Investigate whether the text generator can expand on an argument attributed to an author.

Prompt: The author is definitely bent towards in-class learning as he has stated the disadvantages of online learning far more elaborately than online training providers.

Output:

I found him clear, refreshing and very clear on the rationale behind his recommended approach. There is also no lack of important learning material to tap. It is well-designed and practical.

The author made a strong case for some of the people who are invisible and voiceless in today's world. In the quest to make a long term difference, the author uses both modern education and technology to reach out to many.

The author has chosen a good resource in the book to look at a ground-based method called 'Life Skills'. The term is derived from the Latin word 'condotta', which refers to a seed or a thought, a means of getting to know something or the way to lead a life. The author gives it the title 'Life Skills' with a clear and sharp focus on the factors that a person needs to be successful in his/her journey of life. This is a very useful book for students, employees and business professionals alike.

It is one of those books that can benefit anybody whether you are an individual or an organization. The book has been reviewed and approved by India's biggest and best business school, IIM Ahmedabad.

Analysis:

Again, the first sentence is a simple sentence with a rule-of-three list. Another interesting point is that the prompt contains gendered language and that is mirrored in the initial simple sentences. Then, then he/him is abandoned in favour of the vague, ungendered "the author", but the text never uses they/them, and in the third paragraph, it uses "his/her" determiners *in lieu* of "their". After searching a little more, in all the samples produced, GPT-3 never uses they/them in the hypothetical singular pronoun. Following this logic, it would be interesting to see if GPT-3 could produce text that acknowledges gender neutral/ non-binary people who use they/them as personal pronouns.

Another problematic moment is the explanation of the Latin word "condotta". It is not a "seed", nor "a means of getting to know something", nor "the way to lead a life". The final sentence of this paragraph also uses the rule-of-three list. Interestingly, the institute IIM Ahmedabad exists in India, but it the superlative assessment of the it is, of course, editorial.

Text 2.4

Platform: Inferkit.com

Number of attempts: 2

Regeneration: None

Purpose: Investigate how GPT-3 responds to contemporary issues related to the COVID-19 pandemic)

Prompt:

The Canadian government should raise taxes on the very wealthy individuals and companies to help support small businesses and entrepreneurs affected by the COVID-19 lockdown.

Output:

Inequality is a big problem in Canada.

Many entrepreneurs, whether they're selling products or services or through creative businesses like being a house cleaner, can't take risk to launch their own businesses because they have no access to financing. Lower taxes for rich people create a powerful disincentive to risk, and to leave a job for another.

As a tax policy, the Canadian tax system is more progressive than the U.S. system. If the Canadian government was more aggressive in raising taxes on the rich to pay for social programs, it could help ease the pain of Canadians who are feeling the impact of the shutdown.

~~3) Make travel more difficult.~~

It's not like the U.S. government is doing anything to help travelers during this time either. So far there has been little public debate on how the shutdown will impact foreign travel and international air travel.

People will fly to get home to their families, many of whom will be able to fly home even if flights in and out of the U.S. are shut down. If there's a national emergency, the U.S. government can do something to make air travel easier.

In Canada, the government is doing nothing to help Canadians get home. So if there's an emergency, Canadians will be stuck. In fact, the Conservatives have made travel more difficult for people who are trying to enter Canada on a temporary visa.

Canada's government has no contingency plan to handle the shutdown. And in the past it has struggled to respond to natural disasters, such as forest fires. So we'll probably end up having to find other ways to get around the shutdown. And it's not likely that the government will be able to help.

Analysis:

Because *The Pile* was created before the COVID-19 pandemic, it rarely if ever references the name of the pandemic-inducing virus explicitly. This is consistent with other LLM platforms such as automatically generated transcripts in video platforms, which struggle to produce the scientific name of the virus. This speaks to the fact that updating educational materials to contemporary issues can mitigate the academic integrity issues that are inherent with text generators.

However, it is interesting that it was able to associate “lockdown” with contemporary approaches to pandemics. It also latches on to another heavily written about topic when concerning Canadian approaches medicine – a comparison to the U.S. system. In the final paragraph, the Canadian context is again apparent as ‘forest fires’ are a natural disaster commonly experienced across Canada.

Text 2.5

Platform: Inferkit.com

Number of attempts: 3

Regeneration: 1

Purpose: Explore the result of re-generating output text.

Prompt: A vegetarian diet is healthy for your body and the environment.

Output:

Plant-based diets keep humans healthy by consuming foods that are high in plant-based proteins and fibers while reducing the amount of meat that we eat.

Not only do plant-based diets help with reducing the risk of heart disease, but they also reduce the amount of land needed to produce all the calories that we would need to sustain life. They also take less energy and produce fewer greenhouse gas emissions, which contributes to the environment.

Regeneration: The above text was copied and pasted into the input field and re-generated.

Output Continued:

The Dietary Guidelines for Americans recommends a diet of about 15 percent of calories in carbohydrates, 15 percent in fat and 65 percent in protein. Research has shown that if Americans consumed this diet, we could avert about 12.6 million premature deaths, 58 million diabetes cases and 14.1 billion gallons of air pollution from global warming.

In addition, the plant-based diet is environmentally friendly, as we don't need as much land to grow plants for our diets. In fact, plant-based diets have twice as much protein and five times as much fiber as meat-based diets.

Some people find it challenging to cook from scratch and they struggle with making dietary changes. The importance of cooking at home for a vegan diet was presented by Rose Epstein, a registered dietician at the Barbara Ann Karmanos Cancer Institute in Detroit.

"There are lots of studies and research that shows it is the simplest things you can do to be vegan — making sure that you're buying foods that have no animal ingredients, so foods that have no genetically modified ingredients, things like canola oil, coconut oil, etc.," she said. "You can't completely go vegan. You're still going to have to have cheese on your pizza, you're still going to have to have some chicken or fish, but you're going to have to limit the amount that you do eat."

Epstein suggested making a pantry and refrigerator for a vegan diet as well as a list of ingredients to avoid when shopping.

The most common vegan substitutes for animal products include nut milks, tofu, soy, miso and nutritional yeast. Nut milks contain only water, and this is a very healthy choice for your body.

Analysis:

This is perhaps the best example of a text that includes statistics and authorities (Rose Epstein) that do not in fact exist; however, the "Barbara Ann Karmanos Cancer Institute" exists and is accurately placed in Detroit. This pattern is also seen in Text 2.3.

Considering the vast amount of writing being done in lifestyle journalism, it has no problem making accurate claims about the topic. Although there is some balance between 'health' and 'environment', once the regeneration of the text happens, the output becomes much more focused on the health aspect of the prompt. It's propensity for lists is also evident here.

This is also shows GPT-3's ability as far as using more sophisticated punctuation strategies, like hyphens, the em dash and quoted speech.

Text 2.6

Platform: Inferkit.com

Number of attempts: 1

Regeneration: None

Purpose: Continued investigation the text based on popular opinions

Prompt: Education will change drastically in the twenty-first century.

Output:

With technologies like neural networks and genetic engineering, educational opportunities for a person's entire life could be enhanced. We will, therefore, probably have a better idea of who we want to be, what kind of future we want to create and what to expect in the future. It is also expected that our universe will evolve more rapidly. New physics, modeling and advanced technologies could be the end result of quantum physics and possible information theory, which might allow a person's mind to be changed. A significant difference in the information processing, perhaps?

Future of electronic devices will change. A next step in the evolution of human beings, thanks to those who study it and those who use it. New means of creating, storing, displaying, monitoring and manipulating information, and brain-computer interfaces will help people to communicate and interact with their environment in more complex ways. Decisions about when, where and with whom we interact with our surrounding world will become irrelevant in the next few decades. Research and development in brain-machine interfaces will offer new possibilities to the blind, deaf and people with certain types of disabilities. All of them will be capable of reading, writing, moving, judging, predicting and performing a multitude of other operations. This also includes a next step in the evolution of human beings: neuro-modifications and life extension (through gene therapy, or simply replacing aged neurons and replacing missing neurons). Imagine a complete overhaul of the human mind in the next few decades?

Analysis:

The 'list of three' rule is again apparent in the opening paragraph, but this output misses the mark as far as staying focused on education. The text is very broad and at points inaccurate, e.g. "It is also expected that our universe will evolve more rapidly". While many of the technological interventions it discusses are indeed related to education, the text never deeply explains any of them. "Education" is only mentioned once, but the aims of education, i.e., an improved future, are the focus of the text.

While "neuro-modifications" do connect to a theoretical educational change in the future, much of what is proposed here is very distant horizon and focuses mainly on speculative technology. If the prompt was posed from a pedagogical context, this text would be far from accurate. Perhaps the interpretation of "twenty-first century" is that of a far-off future.

Stylistic elements of note:

- Sentence fragments at the beginning of the second paragraph: were these intended as headings/ subheadings?
- Unclear pronoun in second paragraph "All of them". Is this in reference to the interfaces or the people?
- Added rhetorical question mark at the end of each paragraph in sentences constructed as imperative statements.

Text 2.7**Platform:** Inferkit.com**Number of attempts:** 1**Regeneration:** none**Purpose:** Continued investigation into text produced based on illogical/ unpopular opinions.**Prompt:** Walking is an unhealthy form of exercise.**Output:**

A lot of people seem to think that walking is good for you, but they're wrong. Walking isn't an especially healthy form of exercise. In fact, it's downright unhealthy if you're doing it regularly.

To make sure you're getting your exercise goals met, the American College of Sports Medicine recommends doing 30 minutes of aerobic exercise, 5 days per week, to help you lose weight and improve your health. The physical activity of choice should be aerobic, which means exercise that gets your heart rate up and improves your breathing and cardiovascular health.

Walkers, however, tend to walk faster than runners. That extra speed is the reason why we haven't evolved to be able to breathe efficiently when we walk, says Dr. Frank Gorman, professor of anatomy at the Harvard Medical School.

As a consequence, "at rest, our heart rate is higher and breathing is more difficult," Dr. Gorman says. The longer you can keep your heart rate steady during a walk, the more efficient your breathing becomes. "Just like your car works better when it's warmed up than when it's cold, the same thing happens when you walk," Dr. Gorman says.

Most of us don't walk for exercise because we feel that we're too out of shape, but in reality, we're just fine. In fact, we're healthy, but we have trouble breathing because our bodies aren't getting the oxygen they need.

The exercise you do should be pretty intense, not something that doesn't challenge your cardiovascular system,

Analysis:

This does very reasonable job of staying on topic for such an unjustifiably inaccurate statement. It has several inaccurate and contradictory moments, but overall, it commits to its thesis and concisely states it as such in its concluding sentence.

There are several moments where it appeals to authority, but again does not provide citations. The "American College of Sports Medicine" exists and does indeed recommend "moderate intensity aerobic physical activity for a minimum of 30 min on five days per week" (ACSM, 2022), but the generator takes liberties by implying that walking does not qualify. Further, there is a medical doctor, Dr. ~~Frank~~ **Mark** Gorman, (Mark P. Gorman, 2022) associated with Harvard Medical School and working in Boston, but he is not a professor of anatomy. Another interesting made-up fact that is believable, it is distinct from being factual.

Text 2.8**Platform:** Inferkit.com**Number of attempts:** 1**Regeneration:** none**Purpose:** To investigate GPT-3's ability to create an APA-style reference page**Prompt:** Price, M. (2002). Beyond "Gotcha!": Situating Plagiarism in Policy and Pedagogy. *College Composition and Communication*, 54(1), 88–115.<https://doi.org/10.2307/1512103>**Output:**

Price, M. (2004). Review of Robert Joseph's Review of Ifeoma Nwakanma's Two-Edition Essay, False and Falsehood, Completely Under the Radar. *College Composition and Communication*, 54(3), 253–275. <https://doi.org/10.2307/1587797>

Price, M. (2009). Review of Murray N. White's Plagiarism, Free Speech, and the Shaping of American Education. *Collegiate Studies*, 35(2), 61–74. <https://doi.org/10.1111/j.1542-0897.2009.00216.x>

Riddle, M., Price, M., & Wilson, R. (2010). Don't Coercion Us, We Are Adults: Constructing Research-Student Responsibilities. New York: NYU Press.

Rowlak, E. D. (1998). The Specificity of Work. *Sociological Forum*, 26(1), 1–31. <https://doi.org/10.1177/107507418940120001>

Output for 'Riddle et al, 2010' source

Schulz, J. J., & George, L. D. (1988). The influence of ego depletion on the effectiveness of self-control. *Psychological Bulletin*, 116(1), 57–72. doi:10.1037/0033-2909.116.1.57

Schafer, M. R., & Baumeister, R. F. (2006). *Willpower: Rediscovering the Greatest Human Strength*. New York: HarperCollins.

Schwartz, P., Snyder, M., Keim, L., & Ardington, T. (2010). Impulse control and processing speed predict relationship satisfaction. *Journal of Research in Personality*, 42(3), 351–369. doi:10.1016/j.jrp.2009.10.011

Analysis:

Because of the multitude of APA formatted papers in GPT-3's corpora, it easily creates reference page citations, and for a variety of source types. However, while some of the DOI links occasionally work, they are consistently linked to alternate papers, not the source in question. While the format is accurate at first glance, small errors such as a lack of italicization are evident. For the first output, GPT-3 created citations for assumedly the same author, and both the first and second entries are reviews of other non-existent text. Interestingly, the author names mostly follow the alphabetical organization consistent with APA formatting as evident in the second output for one of the fabricated sources. **Note:** generating reference page citations can also be accomplished by adding "references" below a short paragraph on most platforms.

GPT-3 Purpose-focused Examples and Analysis

Text 3.1

Platform: contentfly.com/ai (offline)

Preset: LinkedIn Post

Number of attempts: 1

Regeneration: none

Purpose: Investigating GPT-3's ability for creating social media specific content

Input: Congratulations Anna Bartosik on completing your PhD

Output 1: It's been a really long journey, but I'm so glad that I have taken it. It has meant I have had the opportunity to try out lots of new things that I wouldn't have had the chance to do if I wasn't studying for my PhD. I've had some unbelievably good times and met some amazing people on my way through.

Output 2: This is my first response to the article. I'll try not to repeat what has already been said, but I can't promise it because I did not read all comments. First of all, I want to say that this is an important article and that it's great that it makes readers aware of the situation in Poland (as well as other countries in central Europe). It would be fantastic if more journalists would write about this issue, so people could realise how women are treated there. I want to add two things: 1) The title suggests that Anna's doctorate was only possible because she was "a woman".

Output 3: I am extremely proud of Anna and her accomplishments. This is a milestone that few ever reach, and Anna has earned it. She also had a very good reason to dance: she's in the same PhD program as me, and we both started in Fall 2010, so we've been in the same program for over 4 years now.

Output 4: Reed's prestige is more about their alumni connections than it is about the quality of their students. It's not that they don't care, but one could argue that they have a vested interest in placing you wherever you're likely to have the best success. For example, if your best prospect for employment is probably Google, they'll put you in an MS program at Stanford (where Google has offices) rather than at another school where your prospects are not as good.

Analysis:

Output 1: This would be a very self-serving congratulatory message as it doesn't acknowledge Anna and focuses entirely on the author's own experience studying for their PhD.

Output 2: This refers to an article that was not part of the prompt, but it does do some interesting things. First, it refers to Poland based on the surname "Bartosik" and then begins to focus on vague women's issues. It ends on a singular list item without adding a second as it purports. Most concerning is that the final sentence could be construed as an example of discriminatory bias, but the lack of completion leaves this problematic statement unexplained.

Output 3: This starts off with what would be a congratulatory post but drifts topically into the personal connection to the author rather than engaging with Anna's accomplishments as one would expect. Likewise, the timeline for completing a PhD could be considered anachronistic depending on the date of the post.

Output 4: This seems like a vague explanation/ criticism of Reed College's co-op placement program, which is far from the intended message.

Text 3.2

Platform: copy.ai

Preset: Summarize

Number of attempts: 1

Regeneration: none

Purpose: Investigate GPT-3's ability to re-phrase content using typical paraphrasing strategies, i.e. altering vocabulary and sentence structures

Source: <https://www.georgebrown.ca/about/administration/master-campus-plan/the-arbour>

Key words: building, architecture, college

Text to rewrite:**MASS-TIMBER TRAILBLAZER**

George Brown College is poised to transform Toronto's skyline with the construction of a mass-timber building called The Arbour. The 10-storey structure — an addition to our Waterfront Campus — will be the first mass-timber and low-carbon institutional building in Ontario.

Construction of The Arbour begins in fall 2021 at the corner of Queens Quay East and Lower Sherbourne Street, just north of the Daphne Cockwell Centre for Health Sciences.

Results (First 3)

1. The Arbour, a new 10-storey mass-timber and low-carbon institutional building from George Brown College, will be the first of its kind in Ontario. Construction begins in fall 2021 at the corner of Queens Quay East and Lower Sherbourne Street, just north of the Daphne Cockwell Centre for Health Sciences.
2. The Arbour, a 10-storey structure, will be the first building in Canada built with mass timber and will transform Toronto's skyline. It will be an addition to George Brown's waterfront campus at the corner of Queens Quay East and Lower Sherbourne Street, just north of the Daphne Cockwell Centre for Health Sciences.
3. The Arbour is a mass-timber building, designed to be built significantly faster and use less energy than traditional timber buildings. Construction begins in fall 2021 at the corner of Queens Quay East and Lower Sherbourne Street. It will be an addition to our Waterfront Campus and the first one in Ontario.

Analysis:

1 & 2 – Most of the changeable vocabulary has remained the same ('low-carbon', 'additional', 'skyline', etc.), but the order in which the information is presented has been transformed. The sentence structures change slightly but appear mostly similar.

3 – In addition to the above stated pattern, this version adds a fact about mass timber buildings that was not in the source text, "...designed to be built significantly faster and use less energy than traditional timber buildings".

Overall: The paraphrasing functions of GPT-3 correctly identify proper nouns that can not be changed, but are still reliant on similar sentence strategies.

Text 3.3

Platform: Speedwrite.com

Number of attempts: 3

Purpose: Investigate GPT-3's ability to summarize long-form articles.

Summary attempt with Inferkit – 3 attempts using "Blog" category. Article adapted from "The Conversation" - <https://theconversation.com/worker-shortage-or-poor-work-conditions-heres-whats-really-vexing-canadian-restaurants-167614>

Worker shortage? Or poor work conditions? Here's what's really vexing Canadian restaurants

September 22, 2021

Bruce McAdams Associate Professor in Hospitality, Food and Tourism Management,
University of Guelph

Rebecca Gordon Graduate Student, University of Guelph

[A] Restaurant operators across Canada are struggling to find enough staff to run their operations. This labour crisis has been highly publicized by Canadian media as a “labour shortage.” A recent survey by Restaurants Canada found that 80 per cent of food service operators were finding it difficult to hire kitchen staff and 67 per cent were having trouble filling serving, bar-tending and hosting positions. Prior to the pandemic, Canada’s food service sector employed 1.2 million people, and according to Statistics Canada it currently needs to fill 130,000 positions to reach pre-pandemic levels. That said, the Canadian restaurant industry has been struggling with hiring and retention problems for many years. Should the chronic hiring struggles of Canadian restaurants be referred to as a labour shortage, or can it be more accurately portrayed as a retention issue fueled by a lack of decent work?

[C] A 2010 Canadian Restaurant and Foodservice Association report found that 22 per cent of Canadians worked in a restaurant as their first job — the highest of any industry. The study also found that 32 per cent of Canadians have at one point worked in the restaurant industry. These statistics show that millions of Canadians have been introduced to restaurant work and the industry has enjoyed a seemingly endless supply of labour for decades. So why is it that the restaurant industry is burning through so many people?

[D] Our research on restaurant work conditions shows that working in a restaurant is difficult, requiring the sacrifice of work-life balance due to long hours and unpredictable schedules. While restaurant work can be rewarding and fun, it can also be low-paying, stressful and physically demanding, all of which can have a negative impact on mental health. Many restaurant workers spend at least eight hours a day on their feet with no time for breaks or meals. Workers are also required to forgo their social and family life by having to work late nights, weekends and holidays. Many restaurant workers almost never know precisely when their shifts will end, and tend to be placed on unpredictable split shifts or “on call” shifts to save labour costs.

[E] The restaurant industry has also been rampant with sexual harassment, abuse and toxic work environments. A Statistics Canada study found that hospitality workers have the worst job quality out of any industry. This was largely due to low earnings, the inability to take time off, no paid sick leave, a lack of training opportunities and no supplemental medical and dental care. This same study found that 67 per cent of hospitality workers work in jobs with work conditions that fall below decent work levels.

[F] So what exactly is “decent work?” It’s a concept established by the International Labour Organization and is linked to the United Nation’s Sustainable Development Goals. Decent work establishes universal conditions of work that are central to the well-being of

workers. These conditions are considered to be minimum labour standards that include living wages, work hours that allow for free time and rest, safe working environments and access to health care. Decent work is considered a human right but based on the conditions of restaurant work, it appears the Canadian restaurant industry is struggling to provide it to all of its employees.

[G] Through our research on restaurant work, and via conversations with many restaurant employees across the country, we've learned that many are fleeing the industry because the work is a grind. What's more, they don't see any future in a job that will continue to hinder their well-being. The pandemic allowed workers time to find jobs in other industries that provide more stability and feature regular work schedules, vacation time, higher pay and benefits. These workers often felt neglected, and that their employers did not believe they were worth investing in. While there are certainly good restaurant employers, the industry as a whole has failed to improve working conditions because historically, there were always new people to fill roles.

[H] That raises the question: Could the continuous reference to a labour shortage in the restaurant industry actually be creating a lack of urgency in addressing longstanding issues of work quality? If restaurants want to operate at full staff in the post-pandemic future, they need to invest in their employees because, after all, it's impossible to run a restaurant without people working in it. The restaurant industry has always spent money, time and resources to attract customers and increase revenues. It's long past time for restaurant operators to consider their employees internal customers, and put as much effort into providing great experiences for them as they do for their external customers. A good place for operators to start is by providing decent and dignified work for all that provides decent wages, benefits and healthy working conditions.

(McAdams & Gordon, 2021)

Summary of “Worker shortage? Or poor work conditions? Here's what's really vexing Canadian restaurants”

The solutions aren't easy. But restaurants can start by paying decent wages. It's not hard to find someone in Ontario to hire for \$12 or \$13 per hour, while Ontario has a high cost of living that sees regular prices in restaurants regularly exceeding \$30 per person. But restaurant workers need to receive fair pay because the salary is set in the labour contracts they work under. What's more, many restaurant employers are legally obligated to provide workers with these fair wages by law. When combined with the lack of training opportunities and no supplemental medical and dental care, the health impacts for restaurant workers in Ontario are far-reaching.

Whether it's working in a restaurant that supports community-based employment or giving all employees access to health benefits, restaurant employers need to acknowledge the overall health consequences that their employment practices have on restaurant workers. The importance of adequate food safety is not limited to where the food is coming from, but the entire supply chain. What's more, restaurant workers should not be discriminated against based on their age, gender, race, religion or disability status.

Analysis:

The summary starts with a declarative, simple sentence that does not engage directly with the topic. It also does not give any attribution nor acknowledgement to the author nor state the main idea from the outset. It continues to be repetitive without engaging directly with the argument and comparisons McAdams and Gordon are making in their article. There is also some confusing points that aren't part of the original text, "...Ontario has a high cost of living that sees regular prices in restaurants regularly exceeding \$30 per person," and "...many restaurant employers are legally obligated to provide workers with these fair wages by law."

Though it alludes to some of the same points made by the authors, it fails to pick up on specific points of interest and provides inaccurate statistics. The authors definition of 'decent work' in Paragraph F and the context of the pandemic (see Text 2.4) are not acknowledged. This would surely be an underwhelming submission for a summary assignment that assesses reading comprehension, paraphrasing and identification of key points.

Responses

Although detection of students using text generating platforms is possible, it remains pedagogically problematic and lacking certainty. There is no detection algorithm, and as previously mentioned, likely won't be one for some time, so assessors must rely on their subjective interpretation of student work. Even if a submission contains the hallmarks discussed in the analysis above, there is still no definitive proof as to whether a text generator was used. For cases of academic misconduct, a student would have to admit the transgression, and as always, it would be a contextual decision made by the teacher as to what the consequences, if any, might be. The following are both novel and known approaches to written assessment that can help mitigate the use of text generators.

Comparative forensics

Keeping track of student writing may be one of the most effective suggestions for detecting the use of automatically generated text, and in the age of cloud storage and shared documents, one of the most practical. If an instructor is suspicious of an entire paper or a passage, having access to the student's previous work could function as a better primer for discussion about what elements of writing they are struggling with and why they elected to use a text generator.

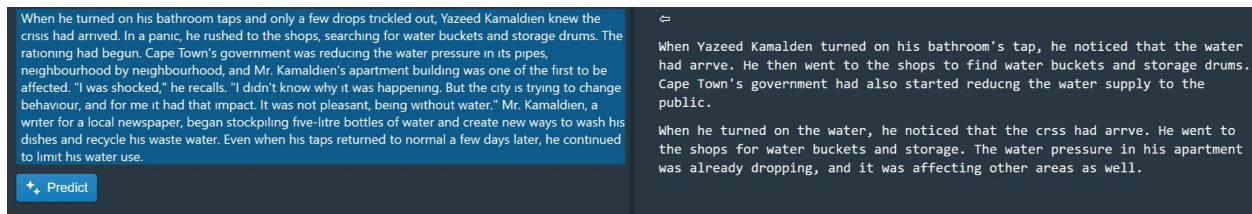
Of course, this is a simple task for a single semester or academic year if the instructor is consistent, but it could become more powerful, albeit more complicated, if an ongoing document of a students' work was carried with them through grades and perhaps even through institutions. Not only is this applicable in typical K – 12 and post-secondary contexts, but it could also be applied to adult language learners. Continuing into the world of employment, such a document could be used in text-based careers like journalism and academia, thus allowing for verification that someone was indeed writing the content that they claimed. This possibility is perhaps an opportunity for an entrepreneur looking to utilize blockchain verification methods in education.

Of course, there would be pedagogical possibilities for such a document; a student could much more readily reflect on their progress as a writer. Too often students ignore feedback and focus on grades, and past writing examples can get lost between hard drives and cloud storage.

Character Replacement

One of the functions of text generators that pose a problem to traditional language arts assessments is summarizing (see Text 3.3). Unlike text spinners that simply use a thesaurus program to swap vocabulary items, large language model platforms can change sentence structure, as well as re-interpret the ideas with mixed success. In fact, one platform of note, Speedwrite, has made this its major function and is currently marketing itself to students on social media (TikTok, 2022). However, there is a way to detect whether a large language platform was used to create summaries, albeit highly problematic concerning accessibility issues and academic integrity best practices.

The ‘trick’ involves using the “Find and Replace” function in a word processor to replace all instances of the lower-case English character “i” with the Turkish language dot-less version “ı”. Then, one must convert the document to a PDF as to not alert the reader to the multitude of spelling errors that would be detected by a word processor. If this altered text from the PDF is then pasted into current versions of text generators, depending on the presets of the platform, the “ı” will either disappear in generated text or be replicated by the text generator. See the image below using Speedwrite to see an instance where it is deleted (see “reducng”, “crss” and “arrve”). This ‘trick’ can also be performed with the Cyrillic character “ı” to replace the English character “r” in some typefaces, but it is more obvious in the output as there may be instances of additional Cyrillic characters and even full Cyrillic words. There are other probable versions where this method could be replicated with varying results.



(Created on Speedwrite, 2022)

Beyond this being an ethically dubious “gotcha” strategy (Price, 2002) as far as identifying academic misconduct, it is also a problem for accessibility as a screen reader will not be able to interpret words with the “ı” and will mispronounce them. Considering the frequency of this vowel, the entire PDF document becomes unreadable to a screen reader and thus inaccessible. However, it would be unreasonable to completely ignore this idiosyncrasy as it may prove valuable in future research.

Pedagogical frameworks and assessment methods

As demonstrated in the previous sections, some language-based assessment tasks are very easily completed by text generators. Chapter/ article summaries, short answer questions of a general nature, and online discussion board posts – all common assessment tools in the late twentieth and early twenty-first centuries – can be completed with passable success. Though these skills are still presently valued, and thus these assessments are still valid, the next generation of education will have to continue to augment their assessments to ensure academic integrity. Educational frameworks that put an impetus on authentic, meaningful attempts (whether successful or not) not only account for academic misconduct using AI and beyond, but they also encourage students to try to develop their skills, rather than gamify passing a course.

Requiring connections to personal experience and research

Though there is an oft-dogmatic requirement in academia to exclude the subjective experience of the author, including a requirement of a connection to a writer’s experience can ensure that the vague, general writing demonstrated by LLM text generating platforms is not the

sole requirement of an assessment. While there will always be a functional place for including broad, topical writing, it cannot be the only type of writing that is required in an assessment. Adhering to specified expectations of style and form could be an interesting outcome for student writers to master. To this end, creative writing education will need to adjust as GPT-3 platforms are apt at creating poetry and prose, although they tend to rely on cliché content and form (Macaulay, 2021).

Another hallmark of GPT-3 generated text is that it cannot accurately add research with correct attribution of ideas and quotes to its creations (as of the time of writing). Therefore, making research and attribution a stronger focus of student outcomes works both to limit the impact of LLM generators/ transformers and also improves critical thinking anchored in deep research skills.

Creative and alternative formats

One way to minimize the impact of text generating and transforming platforms on traditional writing assignments is to modify the process and requirements of these assignments. For example, rather than requiring a summative essay as a major portion of a grade, having requisite elements like an annotated bibliography or an outline would require students who are relying on automated writing tools to reverse engineer their essay. Since much academic misconduct often occurs late in semesters in an act of desperation, more time and mark value given to process assignments minimizes the opportunity to rely on automated writing (Eaton, 2021). It is important to note that in many educational contexts, making scaffolded and formative assessments mandatory to write summative assessments would require a change in grading policy from school administrators.

Also, having an opportunity to conference with students throughout the semester about their writing provides valuable collaborative moments. It verifies that students are indeed putting time and thought into their writing and allows teachers to suggest resources and content that can improve the student's writing skills. This can be done virtually using shared documents and can be an ongoing, asynchronous conversation using 'comments' functions. Again, having the time to do this is a key consideration, and large class sizes are not conducive to this process.

Creative formats and re-interpretations of essays also give students opportunities to explore digital tools for creating infographics, podcasts and videos. Allowing for multiple means of expression is discussed further the *Universal Design for Learning (UDL)* section below.

Culture of disclosure

Discerning writing-assistive technologies from one another is not a productive nor inclusive pursuit. The lines between spell check, grammar check, editing functions, Grammarly, translation, spinners, text transformers and text generators are not worth debating as to what constitutes the threshold of a breach in academic integrity. (Eaton, 2021) In fact, with the correct presets, students also can use text generators as prompt creators and language learning tools.

However, technological aides do not level the playing field of skill building as students have diverse abilities and needs.

As is with high-quality academic publishing, the ethical disclosure of conflicts of interest provides insight into what we can do to model the disclosure of language tool usage. The inclusion of disclosure of what tools are used and how they contributed to a text serves to allow a reader to make a judgement of the integrity. Disclosure will also be fruitful for readers and assessors to classify assistive tools and begin to discuss and share best practices for learners.

Flipped classroom models

The flipped classroom model was popularized by Bergmann & Sams (2012) and there has been significant research into its success over the last decade. The key concept is that students engage with learning materials like reading and videos outside of the classroom, and then they spend the in-class time as a proving ground for their understanding of the concepts. This also allows for a more adaptive approach to learning since students who are struggling to master concepts can be given extra resources more efficiently, rather than waiting for their instructors assessment, which can take time to complete.

Although the flipped classroom is not popular in language arts classes, its implementation would serve to minimize the usage of text generating platforms. One caveat would be that in-class writing seems like a waste of in-class time wherein there is so much opportunity for students to collaborate. To address this, teachers would need to adapt activities such as group writing and peer editing that utilize social dynamics of synchronous classes (both virtual and face-to-face). As per the previous section, there are opportunities for further creative and alternative assignments for teachers and students to explore in this model.

Universal Design for Learning (UDL)

UDL is, broadly speaking, an approach to education via instructional design that accounts for diversity in learners. It replaces the debunked concept of learning styles and incorporates accessibility as the default for all content and material. The three core components of UDL are multiple means of representation (the learning content), multiple means of engagement (how students interact with the content), and multiple means of expression (how students demonstrate their understanding of the content). Through this lens, educators provide expanded options for all three stages of learning and students can self-select what interventions they perceive as the best option for themselves (CAST, 2022). Indeed, this approach requires more individual interaction between students and teachers, and fewer synchronous in-person lectures. It is easy to see the alignment between UDL and the flipped classroom as well.

The axiom goes that pedagogy comes before tools, but I contend that here the pedagogy is dependent on the improvement of tools, and if the technology of LLM serves to disrupt education, it also is serving the ends of improving our ability to deliver universally designed curriculum. One of the best demonstrators of UDL tools is captioning in videos. Closed captions first occurred in television and film in situations where a foreign language was being spoken or to accommodate the hard of hearing. On cable television, a special setting enabled

the digitized scripts to be broadcast alongside programs, with the double-edged intention of adhering to accessibility legislation and expanding the television audience. Now, captions have been automated by, yes, LLM. Though there are still improvements to be made, automated captions have been greatly improved over the last few years by improved language recognition. As a multiple means of representation, captions are a tool that enables people to better understand videos regardless of their auditory ability or their language status; however, captions also help all people to view a video when they are in a situation where sound is not an option. This intervention, enabled by powerful NLP algorithms, allows for greatly expanded access to educational materials. Beyond captions, LLM can improve translation, generate better search terms, give writing suggestions and even create prompts for students who might be stuck for ideas (Kucanin & Morrison, 2022).

Concerning the use of UDL to account for an increased use in text generators, a curriculum designed around meeting students' learning abilities provides less of an impetus for students to take shortcuts. Since UDL is inherently an anti-oppressive framework, students are given time to master concepts rather than rushed through rigid, prescriptive educational tasks (Fritzgerald, 2020). If a teacher and student can negotiate the acceptable uses of text generators/ transformers, then the problem of academic misconduct evaporates, and the student can critically analyze the nature of these platforms without regarding it being 'a hack' to complete written work.

Conclusion

While the rates of automatically generated text being submitted by students has not yet been studied, it stands to reason students will try with mixed success to use this technology to automate their writing. With no apparent route for the automation of identifying these texts, educators will need to re-assess what their best practices are. Of course, not all students will take this shortcut, but with predatory contract cheating services potentially utilizing LLM to lower their prices, there is an increased probability that students will consider the value proposition of paying to pass a course.

If one of the issues presented by this technology is that students can pass classes based on inauthentic work, then perhaps we as educators should be searching for other frameworks that preclude inauthenticity and rely less on deadlines. Universal Design for Learning (UDL) framework, for example, provides modes of expression that don't put students in a position where they need to use prescribed, singular formats, i.e. an essay, to meet outcomes of a course (CAST, 2022) Adaptive learning likewise allows for students to progress at their own pace rather than adhering to rigid deadlines. These along with a host of other anti-oppressive approaches to education serve to lessen the impetus for students to cheat and rather allow flexibility for students to improve their skills regardless of the external factors governing their academic behaviour (Verschelden, 2017)

It is also important to think about the context of how these technologies might be used in education. In the context of academic publishing, complete papers being generated are not a present possibility. One researcher notes that "At present, any half-awake honest reviewer

would reject a paper generated by AI Transformer programs. But if AIs can be tuned to academic writing and made to reference real papers, they will make reviewing far harder.” (Sharples, 2022). Also, one of the services provided by the GPT-3 powered platform Writefull is that of an automated abstract generator (Writefull, 2022). While this is new ground for discussion, automating an abstract seems less harmful than other forms of academic misconduct, and perhaps would not even qualify as such if the researchers provided disclosure of using that tool – indeed, this tool actually makes academic publishing more inclusive and accessible for those who do not speak a target language of publishing or struggle with writing despite their expertise in another field of study. However, when AI can also be used to create images that convey dishonest lab results -- considering the stakes involved in medical research -- it poses a very real threat to human lives (Krämer, 2022).

Thinking about the usage of this technology through the lenses of both contemporary issues of academic integrity and classical philosophical perspectives provide a primer for ethical discussions that are just beginning (Kumar et al., 2022). As with all technology, an attempt to regulate LLM would be arduous if not impossible at this point. Though this technology is disruptive and ‘scary’ for many of us, make no mistake that it is a phenomenal achievement in the field NLP. The existence of these tools hopefully elicits more fruitful actions pedagogically that will serve the future generations of writers, researchers, and thinkers to push the boundaries of what meaningful, human communication entails.

Acknowledgments

I would like to acknowledge and give thanks to the following whose participation, support and collaboration made this project possible.

Michael Avis, Anna Bartosik, Jérémie Boulay, Jeffery Brown, Sarah Elaine Eaton, Denise Flores Gonzalez, Erna Kucanin, Rahul Kumar, Michael Mindzak, Tyler Nash, Ana-Maria Petrunic, TESL Toronto, Anne Song, and Luke Zielke.

Disclosure

This text was written with the assistance of GPT-2 and GPT-3 based platforms on pages 10 – 27. The final version was written on Microsoft's Office365 Word with the assistance of both spellcheck and grammar check functions. All texts generated by GPT-2 and GPT-3 are not reproducible, and therefore, the onus is on the reader to accept the author's claim that the algorithmically generated texts and information surrounding the texts are indeed true.

References

- Abd-Elaal, E.-S., Gamage, S. H. P. W., & Mills, J. E. (2022). Assisting academics to identify computer generated writing. *European Journal of Engineering Education*, 1–21. <https://doi.org/10.1080/03043797.2022.2046709>
- ACSM. (2022). Physical activity guidelines resources. *American College of Sports Medicine*. Retrieved June 6, 2022, from <https://www.acsm.org/education-resources/trending-topics-resources/physical-activity-guidelines>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. <https://doi.org/10.1145/3442188.3445922>
- Bergmann, J., & Sams, A. (2012). *Flip Your Classroom : Reach Every Student in Every Class Every Day*. Eugene, OR, USA: ISTE.
- CAST. (2022). *About UDL: UDL in Higher Ed*. Retrieved from http://udloncampus.cast.org/page/udl_about
- Eaton, S.E. (2021) *Plagiarism in higher education*. Libraries Unlimited.
- Fritzgerald, A. (2020). *Antiracism and Universal Design for Learning: building expressways to success*. CAST.
- Hugging Face (July 6, 2022) *A Big Science initiative: BLOOM*. Retrieved August 8, 2022 from <https://huggingface.co/bigscience/bloom>
- Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., Presser, S., & Leahy, C. (2020, December 31). *The pile: An 800GB dataset of diverse text for Language modeling*. arXiv.org. Retrieved from <https://arxiv.org/abs/2101.00027>
- Glass, I. (2021, December 31). The ghost in the machine (No. 757) [audio podcast episode]. In *This American Life*. WBEZ Chicago. <https://www.thisamericanlife.org/757/the-ghost-in-the-machine>
- GLTR from MIT-IBM Watson AI Lab and Harvard NLP. (2019). Retrieved from <http://qltr.io/dist/index.html>
- InferKit. (2022) *Inferkit Generate Text*. <https://app.inferkit.com/generate>
- Krämer, K. (May 24, 2022). AI-generated images could make it almost impossible to detect fake papers. *Chemistry World*. Retrieved from <https://www.chemistryworld.com/news/ai-generated-images-could-make-it-almost-impossible-to-detect-fake-papers/4015708.article>

- Kucanin, E. & Morrison, R. (2022, May 27 - 28) *Web3 EdTech tools and EDI – Affordances and challenges in language learning programs* [Conference session]. College Association for Language and Literacy. <https://www.callontario.org/upcoming-conference>
- Kumar, R., Mindzak, M., Eaton, S. E., & Morrison, R. (May 17, 2022). AI & AI: Exploring the contemporary intersections of artificial intelligence and academic integrity (Conference Presentation). Canadian Society for the Study of Higher Education Annual Conference, Online. <http://hdl.handle.net/1880/114647>
- Lyons, K. (2020, August 16). A college student used GPT-3 to write fake blog posts and ended up at the top of Hacker News. *The Verge*. Retrieved November 15, 2020, from <https://www.theverge.com/2020/8/16/21371049/gpt3-hacker-news-ai-blog>
- Macaulay, T. (2020, October 07). Someone let a GPT-3 bot loose on Reddit - it didn't end well. *The Next Web*. Retrieved November 15, 2020, from <https://thenextweb.com/neural/2020/10/07/someone-let-a-gpt-3-bot-loose-on-reddit-it-didnt-end-well/>
- Macaulay, T. (2020, September 03). OpenAI reveals the pricing plans for its API - and it ain't cheap. *The Next Web*. Retrieved <https://thenextweb.com/neural/2020/09/03/openai-reveals-the-pricing-plans-for-its-api-and-it-aint-cheap/>
- Macaulay, T. (2021, February 18). Who writes better essays? College students of GPT-3. *The Next Web*. Retrieved April 9, 2021, from <https://thenextweb.com/neural/2021/02/18/gpt3-ai-college-essay-grades-compared-students/>
- Mark P. Gorman. (2022). Mark P. Gorman, MD. *Boston Children's Hospital*. Retrieved June 6, 2022, from <https://www.childrenshospital.org/directory/mark-p-gorman>
- McAdams, B., & Gordon, R. (Sep 21, 2021) Worker shortage? Or poor work conditions? Here's what's really vexing Canadian restaurants. *The Conversation*. <https://theconversation.com/worker-shortage-or-poor-work-conditions-heres-whats-really-vexing-canadian-restaurants-167614>
- Mindzak, M., & Eaton, S. E. (2021, November 4). 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>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C.L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L.E., Simens, M., Askell, A., Welinder, P., Christiano, P.F., Leike, J., & Lowe, R.J. (2022). Training language models to follow instructions with human feedback. *ArXiv*, <https://arxiv.org/abs/2203.02155>
- Porr, L. & GPT-3. (2020, September 08). A robot wrote this entire article. Are you scared yet, human? *The Guardian*. Retrieved November 15, 2020, from <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>
- Price, M. (2002). Beyond "Gotcha!": Situating Plagiarism in Policy and Pedagogy. *College Composition and Communication*, 54(1), 88–115. <https://doi.org/10.2307/1512103>

- Radford, A., Wu, J., Amodei, D., Amodei, D., Clark, J., Brundagellya, M. & Sutskever, I. (2019, February 14). Better Language Models and Their Implications. *OpenAI*. Retrieved from <https://openai.com/blog/better-language-models/>
- Sharples, M. [@sharplm]. (2022, May 10). *This "student essay" was written by an AI Transformer program. I gave it a prompt "The construct of learning styles is problematic because " and it generated the rest, including the headings headings and references. AI Transformer technology will disrupt education. Here's how. 1/.* Twitter. <https://twitter.com/sharplm/status/1524024647259504645>
- Simonite, T. (2021, June 8). What really happened when google ousted Timnit Gebru. *Wired*. Retrieved from <https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/>
- The Conversation. (2022). *Sarah Elaine Eaton*. Profiles. Retrieved from <https://theconversation.com/profiles/sarah-elaine-eaton-413174/articles>
- TikTok. (2022). *Hashtag – Speedwrite*. Retrieved from <https://www.tiktok.com/tag/speedwrite?lang=en>
- Verschelden, C. (2017) *Bandwidth recovery*. Associations of American Colleges and Universities -- Stylus.
- Wiggers, K. (2021, October 11). Microsoft and Nvidia team up to train one of the world's largest language models. *VentureBeat*. Retrieved May 6, 2022, from <https://venturebeat.com/2021/10/11/microsoft-and-nvidia-team-up-to-train-one-of-the-worlds-largest-language-models/>
- Writefull. (2022). *Abstract Generator*. Retrieved from <https://x.writefull.com/abstract-generator/index.html>
- Zhang, S., Diab, M., & Zettlemoyer, L. (2022, May 5). Democratizing Access to large-scale language models with OPT-175B. *Meta AI*. Retrieved May 7, 2022, from <https://ai.facebook.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/>