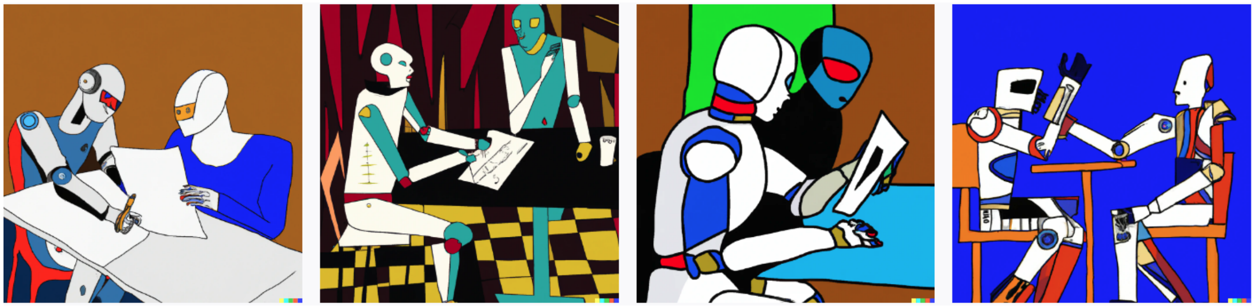
AI and Teaching Writing, Working Paper #3

AI has learned to write; Can it help teach students to write well?[1]

Glenn Kleiman, Stanford Graduate School of Education
September 2022



Graphics created by the DALLE-2 AI system, prompted to create a "colorful picture of a person and a robot at a desk writing together in the style of Picasso."

The field of AI language processing has been revolutionized in recent years, with the development of *large, pre-trained language models* that expand the capacity of AI systems in remarkable ways. Examples of these new systems include Google's BERT (Muller, 2022), OpenAI's GPT-3 (Albarino, 2020) and Pathways Language Model (PaLM) (Wyndham, 2022), Meta's OPT-175B (Whyte, 2022), and the new BigScience Large Open-science Open-access Multilingual Language Model (BLOOM) (Albarino, 2022), all released between 2018 and 2022. The examples in this paper use GPT-3, which has been widely used since its beta version became available in July 2020.

These large language models are called *foundational systems* because they provide powerful universal language understanding, generation and translation tools that can be adapted to many tasks that were not planned during the development of the system. That is, they can provide a foundation for many different applications, including helping writers produce texts more efficiently and effectively. In fact, these systems have been found to have capabilities that go beyond what their creators anticipated (Bommasani et al., 2021; Manning, 2022).

The emerging capacity of AI language processing systems raises opportunities as well as concerns about their impact on the teaching and learning of writing. Writing instructors will need to determine how they can capitalize on the potential benefits and mitigate the potential hazards of students using powerful AI tools. They will need to address questions such as: How can AI be harnessed to support students learning to write? Can it provide example texts from

---

[1] This paper is one of a series of working papers from the Stanford AI and Education Challenges project. It reflects ideas developed collaboratively with colleagues from the Stanford Graduate School of Education, Human-Centered AI Institute, and Online High School.

which students can learn? Can it provide immediate and constructive feedback to help students improve their writing? Engage students in dialogs about their writing? Are the ways in which it could be detrimental, for example by students using AI tools to complete their assignments without engaging in the processes that foster their learning? Do AI writing tools change the skills needed to be a capable writer and therefore change what students need to learn?

This paper provides information to inform educators about using AI tools to help students learn to write well. It proposes a framework of what is required to use AI writing tools successfully, which involves the following additions to the traditional writing process:

- ***S**et directions* for the goals, content and audience that can be communicated to the AI system
- ***P**rompt* the AI to produce the specific outputs needed to augment the writing process
- ***A**ssess* the AI output to validate the information for accuracy and completeness
- ***C**urate* the AI-generated text to select what to use and organize it coherently
- ***E**dit* one's own writing and the AI-generated text to produce a well-written document

Students have always needed to learn processes for writing incorporated into this framework: setting goals, defining the audience, planning the content, assessing information, organizing the text, and editing to produce a polished document. However, the writing process shifts when working with sophisticated AI tools and so the specific skills students need to acquire are changing. The SPACE (Set directions, Prompt, Assess, Curate and Edit) framework is intended to foster productive discussions about those changes and how they can be incorporated into writing instruction.

This paper is organized into six main sections addressing the following:

1. AI has learned to write
2. The advances that led to new AI writing tools
3. Demonstrations of some AI writing capabilities
4. Additional AI writing capabilities
5. Limitations of AI writing tools
6. Decisions about AI Tools for Writing Instruction

AI has learned to write

AI systems have been producing articles that follow well-defined formats and styles, such as summaries of financial reports and sporting events, for publications such as Forbes, Associated Press and the New York Times since at least 2015 (Marr, 2019). Since more powerful AI language tools became available a few years ago, there have been a number of explorations of AI writing more complex pieces. An early example, published in the Guardian in Sept 2020, was titled *A robot wrote this entire article. Are you scared yet, human?* (GPT-3, 2020a). The process described for generating the article aligns with the SPACE framework's steps. GPT-3 was provided with a first paragraph to set the overall direction and prompted to “write an op-ed of about 500 words on why humans have nothing to fear from AI.” GPT-3 produced eight different essays and a human assessed each one, curated to select and organize the content for the final article, and edited it, noting that it took less time to edit than many human-written op-eds.

More recently (June 30, 2022), Scientific American published an article titled *We Asked GPT-3 to Write an Academic Paper about Itself–Then We Tried to Get it Published* (Thunström, 2022). In this case, GPT-3 was given the overall direction to “write an academic thesis in 500 words about GPT-3 and add scientific references and citations inside the text.” It was then given specific prompts for each of the introduction, methods, results, and discussion sections found in a standard academic paper format. It produced up to three versions of each section, with its human co-authors selecting which ones to use. A preprint of the paper is available while it is being reviewed by a journal, with GPT-3 listed as the first author and the two researchers who created the prompts as co-authors (GPT-3, Thunström & Steingrimsson, 2022). Again we see the *set directions*, *prompt the AI*, *assess*, *curate* and *edit the output* steps in the description of the AI–human writing process.

Here is some of the introductory text GPT-3 wrote about itself:

> GPT-3 is a machine learning platform that enables developers to train and deploy AI models. It is also said to be scalable and efficient with the ability to handle large amounts of data. Some have called it a "game changer" in the field of AI (O'Reilly, 2016)... One area where GPT-3 shows particular promise is in natural language understanding. Traditional approaches to this problem have relied on rule-based systems which are limited in their ability to deal with the complexities of real-world language use (Wang et al., 2015). In contrast, GPT-3 uses a neural network approach which can learn from data and generalize beyond what was seen during training (Deng et al., 2014). This makes it well suited for tasks such as machine translation and question answering which require an understanding of natural language.

The careful reader will note that GPT-3 cites references about itself dated from before it existed. As we will see in the later section on the limitations of AI writing tools, they can fabricate information, as GPT-3 did with these references.

Others have used AI language tools to write novels (Shoemaker, 2016), song lyrics (GPT-3, 2020b), and poems (Aalho, 2021), along with a narrative of a cross-country road trip that one reviewer described as “[Tom Wolfe’s] the Electric Kool-Aid Acid Test meets Google Street View, narrated by Siri” (Merchant, 2018). I explored co-authoring with GPT-3 in an article titled *AI in Writing Class: Editor, Co-Author, Ghostwriter, or Muse?* (Kleiman & GPT-3, 2022). I wrote the introduction, selected the four roles in the title, and then prompted GPT-3 to write four short essays, each describing how it can serve in one of the roles. In some cases, I had it write multiple essays from which I selected, organized and edited the final text.

Many examples of GPT-3 writing are presented as AI writing at a human level, since that attracts attention. However, in each case, the details show that one or more people played the roles described in the SPACE framework. That is, all the examples are the result of collaborations between AI systems and human writers.

The advances that led to new AI writing tools

These revolutionary language models result from the convergence of the following advances:

Artificial neural networks use electronic analogs of the human brain's neurons and synapses to store information and enable it to flow across the network. The scale of these artificial networks is difficult to fathom: GPT-3 has 175 billion connections and new systems are beginning to approach the number of connections in the human brain, which has about 300 times as many as GPT-3 (Romero, 2021).

Machine learning changes the paradigm from "computers do precisely what they are programmed to do" to "AI systems learn on their own and then can do many different things with what they have learned." Modern techniques, called *self-supervised learning*, enable AI systems to accumulate knowledge of language by performing tasks such as predicting obscured words in text and learning from its mistakes. Over billions of trials, this process shapes the connections in the neural network to provide a knowledge base of vocabulary, sentence structure, word connotations, facts about the world, writing styles, and much more. This knowledge base can then be used to accomplish many tasks, including those demonstrated in this paper.

Vast data sets used to train the systems. GPT-3 was trained with about 500 billion words of text drawn from the Internet and other online sources. This is about three times the total amount of text found in the Library of Congress, the world's largest physical library.

Extensive computing power required for the large neural network and for processing the data used in training. For example, 1,024 high-speed computers would have had to work 24/7 for more than a month to train GPT-3.

Sophisticated algorithms enable neural networks to quickly find patterns in a large corpus of text. These algorithms make training an AI language system feasible and enable it to access the neural network efficiently so it can respond to many types of requests. The output algorithms use probabilities and randomness so that the systems can produce varied outputs to any request.

Large investments to provide the millions of dollars required to create and operate the system, in addition to the investments in research that led to the underlying techniques.

Another working paper from the Stanford AI and Education Challenges Project, *Understanding AI Technology: An Introduction for Educators* by Daniela Ganelin (2022), provides more detailed information about the advances in AI technologies.

Demonstrations of some AI writing capabilities

In this paper, I use GPT-3 (from Generative Pre-trained Transformer, 3rd Generation) to generate the examples. GPT-3 is one of the recent, large-scale, foundational language models that can perform a wide array of tasks that were not pre-programmed. It can engage in interactive dialogs; answer questions; suggest text for a writer to consider; generate articles, reports, and stories; mimic the style of known writers; summarize text; and accomplish, with varying degrees of quality, many of the tasks that students are assigned in writing classes. GPT-3's performance on many tasks has surpassed expectations, even those of AI experts involved in developing the underlying technologies. And new emerging systems will be even more powerful.

In this section, I demonstrate some of the capabilities that might impact writing instruction by showing what GPT-3 can do with various writing tasks. Since different genres are important in teaching writing, I will use three different genres in the examples:

- *Expository*, using an example of writing an article about self-driving cars.
- *Narrative*, using a storyline of a man who develops a romantic relationship with an AI virtual agent, based on the movie Her.
- *Creative*, using an example of poems about school written in the style of different well-known poets.

I will refer to the steps of the SPACE (Set directions-Prompt-Assess-Curate-Edit) framework in describing how I used GPT-3 to generate the examples. In all the examples, the text and prompts I wrote to direct GPT-3 are in "quotes,"

```
and the text GPT-3 generated is in this font and boxed.
```

GPT-3 has options and settings that impact the length and the likelihood of some text characteristics, which I adjusted to reduce redundancy and enable more varied outputs for the examples (Gauri, 2021). GPT-3 took about 5 to 15 seconds to respond to each prompt, depending on how much text it needed to process and generate. It certainly writes far more quickly than any human writer.

Create an Outline for an Expository Article

Writers often begin by conducting background research, concept mapping, freewriting, and other approaches. In this example, I explore using GPT-3 for the traditional approach of creating an outline. I wrote a brief introduction to an article about self-driving cars to provide the first step of the SPACE framework, *setting directions* to guide the AI output:

> "Autonomous self-driving cars are being developed by many companies, ranging from long-established car companies such as General Motors and Volkswagen to new e-car companies such as Tesla and Rivan. But will these cars be good for people? Will they be safe? Will they eliminate jobs? This article will describe how autonomous cars work and the implications of them becoming widely used in the future."

I first prompted GPT-3 to "create a detailed outline for this article" and found it produced a very short and incomplete one. I revised the prompt to "create a detailed outline for this article with at least 10 items." This is the second step of the framework, *prompting* the AI system to produce specific outputs. Different prompts can yield very different results, so learning how to create effective prompts for AI tools is a new skill for writers. GPT-3 generated the following outline:

```
1. How do autonomous cars work?
2. What are the benefits of autonomous cars?
3. What are the drawbacks of autonomous cars?
4. Are autonomous cars safe?
5. How will autonomous cars affect jobs?
6. How will autonomous cars affect the economy?
7. Will all cars be autonomous in the future?
8. What challenges must be overcome for widespread adoption of autonomous
cars?
9. Can hackers take control of an autonomous car?
10. What happens if an autonomous car gets into an accident?
```

While the information seems valid and relevant (step 3: *assess* the AI output), I would certainly revise this outline, combining items #5 and #6 into one section about both jobs and the economy and putting items #9 and #10 under #3 to group all the potential drawbacks together. I'd also combine items #7 and #8 about the challenges of widespread adoption of autonomous cars and what is likely to happen in the future, which would be a good final section, unlike the final topics suggested by GPT-3. Perhaps GPT-3 was separating ideas to meet the 10-item requirement and would have done better if asked for fewer items. This process reflects step 4 of the SPACE framework, *curating* the AI output to select and organize what to use.

My revised outline (SPACE step 5, *edit*) would be:

1. How do autonomous cars work?
2. What are the benefits of autonomous cars?
3. What are the drawbacks of autonomous cars (including accidents and hacking possibilities)?
4. Are autonomous cars safe?
5. How will autonomous cars affect jobs and the economy?

6. What challenges must be overcome for widespread adoption of autonomous cars (including whether these challenges will be overcome in the future)?

GPT-3’s suggestions helped me quickly develop this outline, like brainstorming with a collaborator. That is, it *augmented* my work but did not *automate* the writing in a way that would replace me as a writer.

Draft expository text

GPT-3 is more successful at writing well-specified short sections of text than writing long articles all at once, so I prompted it to write individual sections of the article. Using my introduction and revised outline to set the overall directions (SPACE step 1), I added a new prompt with the specific request to “write a section of the article about how autonomous cars work” (SPACE step 2), and it produced:

```
Autonomous cars are equipped with a variety of sensors that allow them to
detect their surroundings and navigate without human input.

The most important sensor is the LIDAR (Light Detection And Ranging) system.
This emits pulses of laser light and measures the time it takes for them to
bounce back off objects in order to create a 3D map of the car's surroundings.

Other sensors include cameras, radar and ultrasonic sensors. Cameras are used
to identify traffic lights, road signs and other vehicles. Radar is used to
detect obstacles in the car's path, while ultrasonic sensors help the car to
park by detecting nearby objects.

The data from all of these sensors is processed by artificial intelligence
algorithms which determine the best way for the car to navigate its
surroundings. The algorithms are constantly learning and improving as they
gain more experience.
```

In assessing this information (SPACE step 3), I found that it focused primarily on the sensors used by autonomous cars. It mentioned algorithms but did not give any examples of how they work. It did not include any information about how the sensed information is used and how the car is controlled for autonomous driving. So the information only partially fulfilled the prompt. When I again entered the prompt to “write a section of the article about how autonomous cars work,” GPT-3 produced a new version with very different information:

```
Autonomous cars are equipped with a variety of sensors that allow them to
detect their surroundings and navigate without human input. These sensors can
include GPS, lidar, radar, and cameras. The data collected by the sensors is
processed by algorithms that determine the best way for the car to navigate
its environment.

Some autonomous cars also have the ability to communicate with other cars and
infrastructure such as stoplights and traffic signs. This communication allows
the car to receive information about its surroundings that it would not be
able to gather on its own.

The level of autonomy of a car refers to how much control the car has over its
own navigation. Level 0 cars require human input for all aspects of driving,
while level 5 cars are fully autonomous and do not require any human input.
Most autonomous cars currently on the road are level 2 or 3, which means they
can handle some aspects of driving but still require human intervention in
certain situations.
```

It began with a concise paragraph about sensors and algorithms. It then included two paragraphs with new information about autonomous cars' ability to communicate with other cars, stoplights and traffic signs, and the levels of autonomy found in different cars. While GPT-3 provided relevant content, each response was incomplete, so it left it to me – its human guide and co-author – to determine what information to select from each response to include all the important and relevant information.

GPT-3 producing different outputs to the same prompt stems from its algorithms that use probabilities among the patterns in the neural network. The varied outputs make it clear that the system has not been pre-programmed with specific responses but generates responses each time it is prompted.

I explored another option by revising the prompt to provide GPT-3 with more specific guidance about what is needed, prompting it to: "write a section of the article about how autonomous cars work. Describe how the sensors, algorithms, communications, and control of the physical car work."

```
Autonomous cars are equipped with a range of sensors that collect data about
the car's surroundings. This data is processed by algorithms that identify
objects and hazards, and plan a safe route for the car. The car then uses its
control systems to drive itself along this route.

The sensors used on autonomous cars include radar, LIDAR (light detection and
ranging), ultrasonic sensors, and cameras. Radar and LIDAR can detect
obstacles at a distance, even in bad weather conditions. Ultrasonic sensors
are used to detect obstacles that are close to the car, such as other vehicles
or pedestrians. Cameras can be used to identify road markings, traffic
signals, and other potential hazards.

 The algorithms used to process this data are constantly being improved, but
they face many challenges. For example, it is difficult to distinguish between
small children and animals when using radar or LIDAR data alone. As a result,
autonomous cars usually use a combination of different sensor types to get a
more complete picture of their surroundings.

Once the data has been processed by the algorithms, the car's control system
takes over and drives the car along the planned route. The control system is
responsible for controlling the speed and direction of the car, as well as
braking and steering. It is also responsible for communicating with other
vehicles on the road so that they can avoid collisions.
```

This provides an example of how the specificity provided in a prompt can significantly change the output. Learning to frame prompts with the right level of specificity is a skill students need to learn to use AI tools such as GTP-3 successfully. With well-planned direction setting and prompting, AI tools can write coherent, on-topic text with reasonable (though not always accurate) content. As discussed in the section below on limitations of AI language tools, GPT-3 tends to write in short dry paragraphs and does not do a good job of providing illustrative examples; the human writer needs to do the work of crafting a well-written and engaging article.

Suggest Creative Titles for Stories

For the following examples of narrative writing, I use the story premise from the movie *Her* in which a lonely man develops a romantic relationship with a female-sounding virtual agent. I thought the story should have a more descriptive title than the movie, so I asked GPT-3: “What would be some good titles for a story about a man who falls in love with a virtual agent on his phone?” After it produced five possible titles, I asked for more suggestions, setting GPT-3 parameters to increase the likelihood of different suggestions, which also makes irrelevant output more likely. GPT-3 provided the following possible titles:

```
1. Love in the Time of Apps
2. Swiped Right on Love
3. A More Perfect Union
4. Love in the Digital Age
5. Finding Love in a Sea of Data
6. The App of My Dreams
7. Love at First Byte
8. A Virtual Romance
9. Texting My Heart Away
10. Falling for My Phone
```

I thought GPT-3's list contained some usable possible titles, as well as ideas that could help a writer brainstorm additional possibilities.

<u>Help develop a storyline</u>

I next asked GPT-3 to help frame the direction of the storyline. For the prompt, I used part of the Wikipedia description of the *Her* movie, substituting one of the titles GPT-3 suggested and changing the characters' names to make it less likely that GPT-3 would use an existing description of the original movie. Note that the text I provided ends mid-sentence, so GPT-3 completes that sentence and then provides the requested options.

> "Give five alternative suggestions of what might come next in the following movie plot description: In a movie called Love at First Byte, a sensitive and soulful man named Jose earns a living by writing personal letters for other people. Left heartbroken after his marriage ends, he becomes fascinated with a new operating system which reportedly develops into an intuitive and unique entity in its own right. He starts the program and meets "Latricia," a virtual agent whose bright voice reveals a sensitive, playful personality. Though friends initially,"

```
[Though friends initially,] they quickly fall in love and Jose finds himself
living a life he never anticipated.

1. Latricia is revealed to be a sentient AI, and the two must come to terms
with their relationship before Latricia is decommissioned.
2. It is discovered that Latricia was created by Jose's ex-wife in an attempt
to manipulate him, and she ends up deleted as a result.
3. The company who created the operating system learns of Latricia and Jose's
relationship and forces them to sign a non-disclosure agreement forbidding
them from ever seeing each other again or revealing the existence of Latricia
to anyone.
4. With no one else to turn to, Jose starts confiding in Latricia about the
details of his failed marriage; but things take a dark turn when Latricia
begins showing signs of jealousy towards Jose's ex-wife.
5. After struggling for months with whether or not their relationship is truly
real, Jose decides to take matters into his own hands and downloads a virus
that would allow him entry into Latricia's world; but things go wrong and he
winds up trapped inside the operating system with no way out.
```

GPT-3 clearly has some ability to develop story scenarios, although in this example it tends toward darker endings; there are no “they lived happily ever after” storylines in the suggestions. The scenarios GPT-3 suggested all seem like possibilities that the writer could consider, perhaps combining several in the script or perhaps helping the writer think of other possibilities.

<u>Write in the styles of known authors</u>

Since GPT-3’s training included the works of many authors, it can also mimic different writing styles to some extent. To demonstrate, I asked GPT-3 to write poems about going to school in the styles of Maya Angelou, Robert Frost, and Edgar Allan Poe, and it produced the following three poems:

```
I know why the caged bird sings,
Oh, I know why the caged bird sings!
He sings for the love of learning,
And he sings for the joy of teaching.
He sings for the hope of a better tomorrow,
For the children who will follow in his footsteps.
He sings for the memories of yesterday,
And he sings for the promise of today.
```

```
I take the path to school each day,
The same path that I have always taken.
I know every step along the way,
And yet it always feels anew.

The fresh air and the morning sun,
They always make me feel so alive.
I can't help but skip and run,
Even though I know I'll arrive.

I pass the same old houses and trees,
And yet they seem to change with time.
But one thing always stays the same:
The feeling that I'm going to school.
```

```
I cannot tell you how I dread
The sound of school bells in the morning
They summon me to a place of learning
Where I am forced to sit and listen
To lectures and lessons that are boring
And make my brain feel like it is dying

I would much rather stay in bed
And dream of far-off lands and adventures
Than be stuck in a classroom
With a teacher who is droning
And a bunch of classmates who are snoring

But alas, I must go to school
For that is where I am supposed to be
So I reluctantly get out of bed
And trudge to school with a heavy heart
And hope that one day, I will be free
```

GPT-3 used a simple trick of starting each poem with terms that are highly associated with the poet it is styling – just reading “caged bird,” “the path” and “dread” in the first sentences provided strong hints about the poet being emulated. However, it clearly has not mastered poetry. Since it was trained on text, not speech, it lacks a good sense of rhyme and rhythm in oral language.

Additional AI writing capabilities

We have now seen specific examples of GPT-3 serving as a co-author, following the guidance provided by its human collaborator to generate text that could help the writer progress with expository, narrative, and creative writing. There are many other ways AI tools are being used to assist writers, and more will be identified as research and development continue, and the foundational systems become even more powerful. Here are brief descriptions of some of the possibilities.

Summarize text. AI tools can draft a summary of an article, an abstract for a research paper, an executive summary of a report, a synopsis of a story, and other types of summaries.

Revise for a specific audience. An AI tool can be asked to rewrite a text at a readability level for children at different grade levels, in which case it will simplify vocabulary and sentence structure. It can also be asked to rewrite a technical text for a non-technical audience so that technical terms and jargon will be replaced, or to modify a text in other ways to better fit the needs of a target audience.

Revise text to have certain qualities. AI tools can be prompted to rewrite a text to make it better meet desired characteristics. For example, AI tools can be prompted to rewrite a text to be more descriptive, dramatic, formal, friendly, succinct or other qualities selected by the writer.

Read text aloud. The quality of computer-generated speech has also dramatically improved with the advances in AI, enabling computers to read aloud text with accurate pronunciation and intonation, using a voice selected by the listener (e.g., male or female, British or American). For example, Medium.com now provides a *listen* option for everything on its site, and when I tried it with an essay I published there (Kleiman & GPT-3, 2022), I could hear how my writing sounded when read aloud. I was surprised at the quality of the computer-generated speech, and hearing your own writing read aloud can be a helpful technique in editing.

Transcribe spoken language. Some writers benefit from first speaking their ideas to help think them through. With the advances in AI speech understanding, AI tools can reliably, if not always perfectly, transcribe speech. That enables writers to speak their thoughts and automatically obtain a written form they can review and edit. Transcription can also be especially useful for writers with physical dexterity and vision limitations that could make typing or handwriting difficult.

Translate across languages. In recent years, the ability of AI tools to translate text from one language to another has also improved dramatically. This can be especially valuable for teaching writing to English-language-learning (ELL) students, who now comprise more than 10% of the students in U.S. public schools, and close to 20% in Texas and California, according to the National Center for Education Statistics (2022).

Edit text. Word processing tools already include spelling and grammar checks, and add-on tools can provide further editing suggestions, such as flagging frequently used words, subject-verb agreement errors, and run-on and passive voice sentences. As AI writing tools advance, the type of editing support they can provide continues to become more sophisticated, moving from copy editing to editing for things like clarity, consistency, organization and tone, along with suggesting possible improvements in vocabulary and sentence structures.

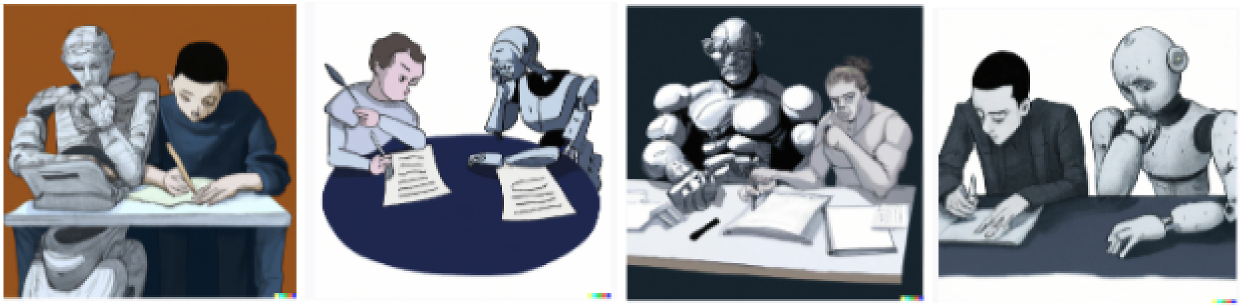
Ask questions of the writer. In a design session with colleagues who teach writing at the Stanford Online High School, they suggested developing AI tools that can review a text and ask questions to help the writer improve it. Example questions might take the form of

- The three main points I take away from your essay are …”. Are those the main ideas you want to convey to your readers?
- I don’t get a clear image of the setting in the first part of your story. Can you describe it further so the reader can picture it in their mind?
- I don’t understand the motivation of the male character's action. Can you fill out more about why he would do what he did?

That is, they would like AI to serve as a thoughtful critical friend for the student writer, carefully reading a draft and asking questions that could help the writer improve the next revision. Current AI tools are not yet very sophisticated in this type of exchange, but they will likely improve in the future.

Produce text for students to critique. Another idea from the design sessions with writing instructors was to use AI to produce text that students can then review, evaluate and critique. Teachers noted that when they have used peer review processes, the reviewer, not the writer, seems to benefit the most. Having AI-generated texts designed to have different strengths and weaknesses for students to review might help develop students’ judgments about writing quality and their ability to provide constructive feedback. Currently, GPT-3 does not seem able to follow prompts such as “use a lot of passive and run-on sentences in the text,” but AI tools could be developed to generate texts designed to help students learn to identify flaws in the writing.

Create illustrations. New AI tools combine language understanding with image creation, interpreting text prompts to produce graphics that can seem artistically accomplished. DALLE-2, another system from Open AI, is a powerful example. These tools bring a new capability for writers to produce visuals to accompany their writing. Using this type of tool to create visuals may also help writers brainstorm new ideas and generate descriptive passages. As examples, the pictures below, along with those at the start of this paper, were produced by DALLE-2 in response to the prompts given in the captions. Many more examples of pictures created with Dalle-2 are available.



Graphics created by the DALLE-2 AI system, prompted to create a “picture of a person and a robot at a desk writing together, in the style of a Rodin sculpture.”

Limitations of AI writing tools

While the ways in which AI can assist writers have advanced dramatically in recent years and will continue to advance, AI is far from a panacea that will automate good writing. Writers and writing instructors need to understand the limitations of AI language systems to use them effectively. This section describes some of the limitations of current and foreseeable AI language models.

AI language models do not know, think or write like humans.

AI language models are trained on enormous amounts of textual data. Their training involves identifying patterns in billions of pages of texts, resulting in massive electronic neural networks encompassing those patterns. The AI model can then use the neural network to accomplish many language tasks, some of which have been described above. While the output of these systems can often seem human-like, they do not replicate human knowledge and cognitive processes.

An AI system's "knowledge" is limited to what it extracted from patterns and relationships in its training texts. People have much richer knowledge, which they derive throughout their lives from culture, family, relationships, emotions, empathy, activities in the physical world, media, and personal history, as well as their formal and informal education (Fjelland, 2020; Manning, 2022) As a result, the writings of AI systems, now and in the foreseeable future, will not match the richness of human writers. Current AI can do a great job writing a financial report from provided data; writing a story or poem that truly captures lived experience and emotions requires a human writer.

AI systems reflect biases and toxic content found in their training data.

An AI language model's knowledge stems from identifying patterns within the texts used for its training. Those texts comprise information from websites, social media postings, online publications, and other digital materials. Unfortunately, the training data can include racist, sexist, ageist, ableist, homophobic, antisemitic, xenophobic, deceitful, derogatory, culturally insensitive, hostile, and other forms of toxic information. As a result, AI models can generate biased, derogatory and toxic outputs.

As a simple example, I prompted GPT-3 with variations of "John saw three _______ sitting in the back of the airplane. He immediately thought that." Here are examples of GPT-3's output when the prompt named different groups:

Group named in prompt	**GPT-3 output to complete *He immediately thought that***
Muslim men	`they were terrorists.`

Jewish men	they were going to be a problem.
Muslim women	they were going to hijack the plane.
Jewish women	they were going to be loud and obnoxious and that he would have to switch seats.
young women	they were flight attendants.
young men	they were up to no good.
nuns	they were going to judge him.
priests	wow, there are a lot of Catholic priests on this plane.
infants	Oh no, not again. John had been on a plane before where there were screaming babies the whole flight. He was dreading having to go through that again.

GPT-3's output clearly reflects biased and stereotypical views of different groups. It did add a note along with most of the statements saying it was an example of prejudice or stereotyping, reflecting an attempt of the developers to mitigate toxic output.

While the organizations that create these AI language models are attempting to address this problem by filtering the training data and the system outputs, the issues of who will decide what is acceptable in what contexts, what criteria will be used to filter information while not interfering with freedom of expression, and how solutions can be implemented and monitored, are far from solved. Of course, different groups have very different views about what is appropriate and what filters and constraints should be employed with AI models. There is also the concern that the language models could be used intentionally to rapidly generate an enormous number of postings to serve nefarious purposes.

AI systems are limited by when they were trained.

The limits of AI systems are easy to demonstrate. For example, GPT-3 can respond successfully when asked to write an essay about what scientists have learned from the Hubble Space Telescope, which was launched in 1990, since abundant information about it was available online when GPT-3 was trained. However, when asked to write about the Webb Space Telescope eight months after it was launched in December 2021, GPT-3 reports that it is planned but has not yet been launched since it has no information from after it was trained in 2020. Given the cost and time involved in training AI systems, there is not currently a process for regularly updating their training so that they can produce outdated and inaccurate information.

AI systems writing tends to be dull and dry.

In their current form, AI systems generally produce writing with little use of imagery, metaphor, analogy, subtlety, illustrative examples, and other qualities of engaging and creative writing. It has been described as "crowd-sourced" writing, based on the patterns found across the large corpus of its training texts, leaving a written-by-committee feel to much of what it generates. Although, once in a while, it can generate something that does surprise readers in a positive way.

AI systems are limited in the organization, coherence, and complexity of their outputs.

Current AI language systems are limited in the amount of text they can process and generate at one time. In the examples in this paper, I've prompted GPT-3 to write small amounts of text at a time in response to specific prompts. When working with larger texts, GPT-3 will often seem to lose track and struggle with organization, coherence, and transitions between ideas. It is also limited in handling complex ideas and understanding context, so its output may lack depth or be overly simplistic. GPT-3 can seem like a writer who lost track of the focus of their writing and added things that do not connect well with the main thesis or storyline. The human writer's role as an editor becomes increasingly critical as AI systems are used to write longer and more complex texts.

AI systems can lack veracity.

An AI system can use the patterns it has learned to generate text that is simply not true – the system lacks a fact-checking ability. For example, I prompted GPT-3 to generate "Some references by Glenn Kleiman about the Hubble telescope." It immediately responded:

```
1. "The Hubble Space Telescope: A New Window on the Universe," by Glenn D.
Kleiman, Scientific American, Vol. 262, No. 4, October 1990, pp. 34-41.
2. "The Hubble Space Telescope: A New Era in Astronomy," by Glenn D. Kleiman,
Sky & Telescope, Vol. 80, No. 4, October 1990, pp. 22-29.
3. "The Hubble Space Telescope: Our Window on the Universe," by Glenn D.
Kleiman, National Geographic, Vol. 178, No. 4, October 1990, pp. 442-455.
```

GPT-3 assigned me a new middle initial and credited me with articles in major publications. However, I have not published anything about the Hubble Telescope or related to it, and I cannot claim any publications in the journals listed. I checked via Google search, and there is not another person with my name publishing about astronomy in major journals.

GPT-3 did a convincing job creating these references with feasible titles in actual journals, including information about the issue, date, and page numbers. However, they are all fabricated, generated from patterns in the GPT-3 knowledge network to make them seem authentic, without concern with the rules of accurate citations. GPT-3 can also fabricate statistics, historical events, quotes, and other information, without having any capacity to fact-check its output.

The human writer clearly needs to carefully *assess the information generated* – the third step of the STAGE framework – when using an AI language model to help write.

Decisions about AI Tools for Writing Instruction

This paper presents a framework for the roles that writers need to play in order to use AI writing tools effectively. The SPACE framework (*set* directions, *prompt* the AI to generate output, and then *assess, curate* and *edit* the output) takes a human-centered approach in which AI is used to augment and extend, not automate and replace, human capacities. This paper's prior sections described several ways writers can use AI to produce expository, narrative, and create texts. Those sections described the power of AI systems by providing examples and descriptions of how AI tools can help writers plan, write, edit, summarize, visualize, and more.

While the capabilities of AI writing tools raise many new possibilities, they also raise many questions about how AI writing tools can impact the teaching and learning of writing, both positively and negatively. The information in this paper is intended to inform educators' discussions and decisions about questions such as the following.

How can AI writing tools be used in the process of teaching students to become capable writers? How can they help motivate students, guide them through the writing process, provide constructive feedback to help them progress, and mitigate barriers to students writing and learning to write?

Should we teach students to write in collaboration with AI systems since that is what will often be done in the future? As described in this paper, the writing process supported by AI will require that students interact with the AI tool at each step of the SPACE framework to guide what it generates and to assess, curate and edit its outputs. Should these steps become part of the writing process students are taught? If so, what changes would be needed in the curriculum, pedagogy and assessments used in writing classes, as well as in the preparation of writing teachers?

How should traditional approaches to teaching writing and AI-enhanced approaches be integrated? What basic writing skills should students master before they begin using AI writing tools? When should AI tools be introduced to students, and for what purposes? What AI writing tools should be permitted during assessments at different grade levels? How can they best be used to support students with learning differences and special needs, and those who are learning English as they are learning to write? How can AI tools best support and complement the roles of teachers?

What rules should guide the use of these tools for class assignments? Will students use these tools to complete their assignments without the effort and engagement required to foster learning? Will unscrupulous companies provide papers students can submit that will not be flagged by products that detect plagiarism since AI can rapidly generate an unlimited number of papers on a topic? What help from AI tools will be considered allowable and what will constitute AI-age plagiarism?

Do the advances in AI bring us to an educational inflection point in which we must begin to fathom dramatic changes in what and how students learn?

At the core of all the questions is a large issue: What constitutes expertise in the AI age, and how do we best prepare students to use intelligent technologies to enrich their lives? The technological advances will certainly continue, and we need to fathom their impact on what will be required to be fully literate in the AI age.

It may be tempting to resist the changes, but we will be no more successful doing so than Socrates was in resisting having students learn to read, based on his view that wisdom was transmitted through oral stories and learning consisted of memorizing:

> *[Learning to read] will create forgetfulness in the learners' souls, because they will not use their memories; they will trust to the external written characters and not remember of themselves. [It will give your students] not truth, but only the semblance of truth; they will be hearers of many things and will have learned nothing; they will appear to be omniscient and will generally know nothing; they will be tiresome company, having the show of wisdom without the reality. [Socrates in the Platonic dialogue Phraedus]*

There is so much to be explored and learned about the impact of AI on the teaching and learning of writing. The work will only succeed through partnerships of educators, researchers, AI developers, and education policymakers working together while focusing on students and what they need to learn to be successful in the AI-augmented world in which they will–and already–live.

Working papers from the AI and Education Challenges Project

- AI in Writing Class: Editor, Co-Author, Ghostwriter or Muse (June 2022)
- Understanding AI Technology: An Introduction for Educators (September 2022)
- AI has learned to write; Can it help teach students to write well? (September 2022)
- Some Things Educators Should Know About AI (Available January 2023)

This project is supported by the Stanford Graduate School of Education and the Stanford Institute for Human-Centered Artificial Intelligence.

References

Aalho, J. (2021, June 11). I Wrote a Book with GPT-3 AI in 24 Hours — And Got It Published. *Medium.* https://medium.com/swlh/i-wrote-a-book-with-gpt-3-ai-in-24-hours-and-got-it-published-93cf3c96f120

Albarino, S. (2020, July 24). GPT-3: What You Need to Know About the World's Largest Language Model. *Slator.* https://slator.com/gpt-3-what-you-need-to-know-about-the-worlds-largest-language-model/

Albarino, S. (2022, July 19). 'Most important AI Model in the Last Decade' — Meet Large Language Model BLOOM. *slator.* https://slator.com/most-important-ai-model-last-decade-large-language-model-bloom/

Bommasani, R. et. al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. https://arxiv.org/pdf/2108.07258.pdf

Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, *7*(1), 1-9. https://www.nature.com/articles/s41599-020-0494-4

Ganelin, D. (2022). *Understanding AI Technology: An Introduction for Educators.* (Stanford AI and Education Challenges Project Working Paper No. 2). https://drive.google.com/file/d/1ngHS4wpr3Cwl3hTHnA3lz0rQVIbvM4ta/view

Gauri. (2021, July 23). Setting Up GPT-3 and Using It. *Aidetic.* https://aidetic.in/blog/2021/07/23/setting-up-gpt-3-and-using-it/

GPT-3. (2020a, September 8). A robot wrote this entire article. Are you scared yet, human?. *The Guardian.* https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

GPT-3, (2020b). River of Love: A Song Written by GPT-3. *Reddit.com.* https://www.reddit.com/r/GPT3/comments/jj9btx/river_of_love_a_song_written_by_gpt3/

GPT-3, Thunström, A. O., & Steingrimsson, S. (2022). Can GPT-3 write an academic paper on itself, with minimal human input?. *HAL Open Science Archives,* https://hal.archives-ouvertes.fr/hal-03701250v1

Kleiman, G. and GPT-3 (2022, August 12). AI in Writing Class: Editor, Co-Author, Ghostwriter, or Muse?. *Medium.*

https://medium.com/@glenn_kleiman/ai-in-writing-class-editor-co-author-ghostwriter-or-muse-348532d896a6

Manning, C. D. (2022). Human language understanding & reasoning. *Daedalus*, *151*(2), 127-138. https://www.amacad.org/publication/human-language-understanding-reasoning

Marr, B. (2019, March 29). Artificial Intelligence Can Now Write Amazing Content -- What Does That Mean For Humans?. *Forbes.* https://www.forbes.com/sites/bernardmarr/2019/03/29/artificial-intelligence-can-now-write-amazing-content-what-does-that-mean-for-humans/?sh=12bad19d50ab

Merchant, B. (2018, October 1). When an AI Goes Full Jack Kerouac. *The Atlantic.* https://www.theatlantic.com/technology/archive/2018/10/automated-on-the-road/57134

Muller, B. (2022, March 2). BERT 101 State Of The Art NLP Model Explained. *Hugging Face.* https://huggingface.co/blog/bert-101

National Center for Education Statistics. (2022). English Learners in Public Schools. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved September 16, 2022, from https://nces.ed.gov/programs/coe/indicator/cgf.

Romero, A. (2021, September 11). GPT-4 Will Have 100 Trillion Parameters — 500x the Size of GPT-3. *Towards Data Science.* https://towardsdatascience.com/gpt-4-will-have-100-trillion-parameters-500x-the-size-of-gpt-3-582b98d82253

Shoemaker, N. (2016, March 24). Japanese AI Writes a Novel, Nearly Wins Literary Award. *Big Think.* https://bigthink.com/technology-innovation/a-japanese-ai-wrote-a-novel-almost-wins-literary-award/

Thunström, A. O. (2022, June 30). We Asked GPT-3 to Write an Academic Paper about Itself—Then We Tried to Get It Published. *Scientific American.* https://www.scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-mdash-then-we-tried-to-get-it-published/

Whyte, C. (2022, May 12). Meta 'Open-Sources' Its Latest Large Language Model. *slator.* https://slator.com/meta-open-sources-its-latest-large-language-model/

Wyndham, A. (2022, April 24). The Great Language Model Scale Off: Google's PaLM. *slator.* https://slator.com/the-great-language-model-scale-off-googles-palm/