

ChatGPT (and GPT-4) in Accounting Education?

By Irene Wiecek, FCPA, FCA



Irene Wiecek is a Professor, Teaching Stream, Accounting Area, Director of the Master of Management & Professional Accounting (MMPA) Program and Director of the BIGDataAIHUB at the Institute for Management & Innovation at the University of Toronto Mississauga, with a cross-appointment to the Rotman School of Management.

In my humble opinion, as an accounting academic, AI-powered, large language models such as ChatGPT (and GPT-4) are here to stay – and they will keep getting more powerful.¹ That, to me, does not represent uncertainty. It represents certainty. As educators and researchers, we have a unique opportunity to think about how these technologies can be used to make a better, while still human-centred, life. And our students will and should lead the way. How do we navigate these waters? Given the proliferation of technologies such as ChatGPT (and GPT-4), how do we best prepare our graduates to enter the changing work force and add value to society?

There are many conversations going on at present about what we should do about the rise of such technologies in the classroom (with some individuals expressing positive and supportive views of the technologies and some voicing serious concerns). But the real question is not whether we should ban these from our classrooms and broader learning environments. Rather, the real question is how we help our students use ChatGPT and other similar technologies to augment human intelligence, in a responsible and ethical way, with the emphasis on the latter.

Will technologies such as ChatGPT take away jobs? Of course. But this is not a bad thing. The technologies were designed to simplify tasks and streamline workflows. It should not be a surprise then, that jobs will certainly be changing. New technologies always cause shifts in the workforce and change how many people learn and do their jobs. The flip side of this coin is that they also create new opportunities. We need to be sorting out just what these are. What human skills and competencies will be prized as we move deeper into the technology dominated space that our world is fast becoming?

In an academic environment, how we create optimal learning environments and do research is already changing, with many of us embracing more and more technologies to become more productive and glean deeper insights. Are we embracing these technologies to automate many

¹ ChatGPT is powered by GPT-3.5 models and was released in November 2022 by OpenAI. The GPT-4 model was released subsequently, in March 2023, and is currently available for a fee, through “ChatGPT Plus”. GPT-4 also powers Bing (Microsoft’s search engine).

of our own tasks and streamline our own workflows? Or are we doing so to make a better learning environment for our students? Or is it both? I believe we need to be doing both. To truly and authentically engage in a conversation about using this technology in our classrooms, it is helpful to have a basic and common understanding of the technology itself. We will first look at what ChatGPT (GPT-3.5) is, and what it is not.

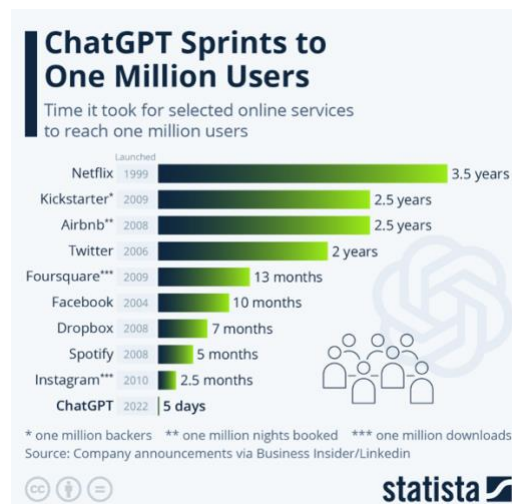
We will then have a look at (what I believe is) the inevitability of allowing it in our classrooms and, finally, how we might use Chat GPT (and GPT-4) to provide learning opportunities for our students.

What ChatGPT (GPT-3.5) Is

It is a chatbot. More specifically, ChatGPT is a large language model chatbot. It is meant to simulate human-like conversations and text. GPT stands for “generative, pre-trained transformer.” “Generative,” because it generates human-like interactions; “pre-trained” because its outputs are based on its prior training (trained using machine learning on very large amounts of internet data and data licensed from third-party providers, only up to September 2021); and finally, “transformer” because it uses its underlying architecture to transform parts of words (referred to as “tokens”) into what we would call “natural” language” (i.e., language that would naturally be spoken by humans in response to a user query or prompt).

After initial training, the model (or algorithm) is fine-tuned by using human feedback, in a process called reinforcement learning from human feedback RLHF. The model is constantly being updated and upgraded to make the output more human-like.

Its uptake has been shockingly rapid. After its release in November 2022, it gained one million users within 5 days. The chart below² shows a comparison with other technologies. Enough said.



² <https://www.statista.com/chart/29174/time-to-one-million-users/>.

What ChatGPT Is Not

By itself, ChatGPT (GPT-3.5) is not a search engine. It is not connected to the internet unless plugins are used to augment its capabilities (more on that later). Search engines are programmed to crawl the internet looking for things, based on a search query that is input by a user. Historically, search engines have returned a ranked list of links to sites that pertain to the query. The user must then sift through the list, perhaps visiting the linked websites, to continue to try to find an answer to the search query. It can be a time-consuming process with many dead ends and information provided that is not actually relevant to the user query.

By comparison, ChatGPT relies only on its training to respond to the user query with human-like conversations. During training, it was fed hundreds of gigabytes of data, largely from the internet and including books, the Wikipedia body of knowledge and research articles. This is somewhat similar to the way many of us learn, i.e., by reading a book for instance. When prompted, ChatGPT will rely on only what it has digested during its training, much like students might respond to questions after reading a textbook, basing their responses on what they have read in the book.

On its own, it cannot access up-to-date information. Search engines can access all data that is publicly available on the internet and can do so in real time. If you wanted to know a current stock price or the temperature outside at the current moment, a search engine can quickly find that for you. Since Chat GPT has been trained only on certain information and only up to September 2021, it does not “know” about events that happened after that date. These events were never included in its training material so it cannot have a conversation about them. This would be like asking your students to complete an in-class test on material not yet covered in a course. ChatGPT would not be able to tell you the temperature at the current moment although it would be able to tell you what the historic average temperature might be.

It is not always right, nor are its answers always acceptable. ChatGPT suffers from what are known as “hallucinations,” i.e., it returns false information and is wrong in quite a few cases. Just why is this? It has to do with the way that the chatbot compiles its output (responds to a user query). Recall from above that, when prompted with a query, the chatbot is not “going out to search the internet.” Instead, it is relying on its training data to compile an answer.³

Suppose the user asks why AI should be embedded in accounting education. ChatGPT may begin to answer by starting with the same words as the user prompt (query). Logically, the answer might start with “AI should be embedded in accounting education because it will enhance....” Now, it must decide what goes next. In general, it goes back into its training material and, given the context of the query, it looks at the probability that any given word (or part of a word) might follow the words in the sentence it has already started.

³ This is based on a very interesting article which really helped me think about the issue of “hallucinations.” A more technical discussion is contained in the article in the following link:

<https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>.

While the process is a bit more complex than this, in essence, it will then identify and rank possible words (or parts of words) that are likely to follow the words it already has in its sentence (considering the probabilities of each word occurring in its training material). Given its training data, if the next most likely word is “cognitive,” it would add that word and start looking for the next word that would follow “cognitive.” That next word might be “development” for instance. It would then continue, looking for each successive word, thus creating a sentence or two.

The algorithm is built so that it does not always use the most likely next word. This is so that the output is not always predictable. Instead, it might use a word that is likely, but not the most likely. This introduces some randomness in the responses. After all, if asked the same question, your students would not always answer in exactly the same manner. This results in interactions with the chatbot that are more interesting and human-like.

Therefore, although the algorithm has been trained (on lots of data but only up until September 2021) to mimic human language and consider context, it will not always give the same answer and the answer might not always make sense when taken in totality. Recall that, as it compiles the answer, it is adding one word (or part of a word) at a time, based on its probability of occurrence from the training data. It is not “thinking” nor “sense-making” as we humans would do. Because its training data was defined (by someone), bias in the data also exists. Ambiguity in the user prompt (query) might also create problems. Thus, the answers must be taken with a grain of salt.⁴

Students generally embrace emerging technologies – and much faster than many faculty members – especially when it helps them get their homework done or streamlines their studies.

ChatGPT is not GPT-4. So far, we have been talking about ChatGPT (which is based on GPT-3.5 models). In March of 2023, OpenAI released GPT-4 in beta mode. While GPT-4 is based on GPT-3.5, it is more powerful than its predecessors. With GPT-4, many of the limitations identified above with ChatGPT have been sorted out, mitigated or may even disappear.⁵

It can also operate with plugins (software add-ons that enhance productivity). These plugins can connect the chatbot in real time to other applications and the internet. Depending on the plugin, this could, for instance, allow GPT-4 to retrieve real time information and/or do things like book flights and order food. GPT-4 acts as a chatbot interface with these other applications

⁴ The OpenAI website lists a more fulsome discussion of the limitations of ChatGPT, [Introducing ChatGPT \(openai.com\)](https://openai.com).

⁵ <https://arxiv.org/abs/2303.08774> .

and services. This is a very useful augmentation of its capabilities and, by allowing the chatbot to connect to the internet and take actionable steps, its usefulness increases significantly.

OpenAI has also made the chatbot technology available to developers as an application programming interface (API), which allows developers to build applications and services using both the GPT-3.5 and GPT-4 models. For instance, OpenAI is collaborating with Khan Academy in a pilot project to use GPT-4 to power Khanmigo, to provide personalized tutoring or support to students.⁶ Thus, the technology will continue to become even more pervasive.

GPT-4 also differs from ChatGPT (which is text-based only), as it is multimodal, accepting both text and image inputs.⁷ GPT-4 is currently available through “ChatGPT Plus” for a fee. As such, it may not be widely used (yet) by students. GPT-4 also powers Microsoft’s search engine Bing (which is available to all), allowing for a more conversation-like interface between the search engine and the user.

With the release of GPT-4, the lines between large language models (including ChatGPT and GPT-4) and search engines are starting to blur and the chatbot technology is now being integrated with other technologies (visibly or invisibly).

Is The Use of Such Technologies in Our Classrooms And Learning Environments Inevitable?

Now that we have a bit of a better understanding of the technology itself (including how it is rapidly changing and morphing and how pervasive it is becoming), we can start to think about its use in the classroom. Below are some thoughts.

Keeping up with our students in terms of adopting new technologies. If there is one thing that I have learned, in my many years at the university, it’s that my students generally embrace emerging technologies (and much faster than many faculty members), especially when it helps them get their homework done or streamlines their studies. Use of ChatGPT is no exception.⁸ What choice do we have, as educators, other than to similarly embrace the technologies? If not only to maintain academic integrity, should we not be using ChatGPT and GPT-4 to help us to create a better student learning experience? We could certainly use it to streamline our own workflows so that we free up our time for more individualized student interactions. I have started down this path and it can be a real time-saver for faculty members.

The changing workplace. The workplace is increasingly becoming more and more automated, with rote and predictable tasks being supplanted by technologies such as these. In fact, many accounting-related tasks currently being performed by humans are at high risk of being replaced with large language models such as ChatGPT. In a recent study looking at the impact of large language models on the labour market, the authors noted that most occupations will be

⁶ <https://openai.com/customer-stories/khan-academy>.

⁷ As an example, a picture of ingredients such as milk, eggs and flour was input to GPT-4 asking what could be made with the ingredients pictured. GPT-4 responded, noting that things such as pancakes, waffles and crepes could be made.

⁸ <https://stanforddaily.com/2023/01/22/scores-of-stanford-students-used-chatgpt-on-final-exams-survey-suggests/>.

affected to some degree, and some professions are significantly exposed, including those such as the accounting profession (and tasks such as auditing and tax preparation).⁹

If we are hoping to prepare our graduates for entry into the workforce, we need to be cognizant of the fact that tasks in certain professions (including accounting) will change significantly. New technologies can and should be harnessed, where possible, to augment human intelligence, optimize performance and streamline processes. Machines can certainly do many (but not all) things faster and more consistently than humans and, therefore, curricula need to focus on the value that humans bring to any given role, understanding that this will continue to change as technologies such as ChatGPT become more powerful and pervasive. Our students will have to understand how to use and interact with these technologies.

We have a great opportunity to help our students understand what it means to use technologies such as ChatGPT (and GPT-4) in an ethical and responsible way.

Encouraging our students to be technologically savvy. The new CPA Competency Map (CM2.0)¹⁰ in Canada notes that being “tech savvy” is a required skill and competency for newly qualified CPAs. If your students are CPA bound, they therefore have no choice but to walk this path but, as noted earlier, more and more jobs are requiring this skill anyway, as things continue to become more automated. What does being tech savvy mean though? That is a good question. Does it mean knowing how to use data analytics platforms such as Tableau or (advanced) Excel? That is certainly part of it. It seems like we have been on a journey the last few years to revamp curriculum to incorporate data analytics. Most schools and employers increasingly prize data analytics skills and happily incorporate them into curriculum and training. I would be surprised to see university and college level programs that do not use data analytics platforms and Excel.

How is ChatGPT different though? It is yet another tool that we can use to make better decisions and better utilize our time. Why would we argue that data analytics skills supported by technologies (such as Tableau or Excel) are good (and part of being tech savvy), but language skills supported by technologies (such as ChatGPT) are not good (not part of being tech savvy)?

Many language skills related tools are already embedded in platforms such as Word (think spellcheck).¹¹ Microsoft has recently announced that it will bring AI into Outlook, Word, Excel and PowerPoint.¹² According to Microsoft, the new features will be built on the same technology that underpins ChatGPT. This means that ChatGPT technologies will soon move from being “cutting edge” to being fully-integrated mainstream, everyday tools. If we allow the use of Word and Excel in our classrooms, we will, by default, be allowing the use of ChatGPT technologies once this is put in place. Some might not even understand that these well-used

⁹ [GPTs are GPTs: An early look at the labor market impact potential of large language models \(openai.com\)](https://openai.com).

¹⁰ <https://cpaleadstheaway.ca/>.

¹¹ I note, while writing this, that Word also now checks for clarity, conciseness, inclusiveness and sensitive geopolitical references. I wonder if, pretty soon, it will check for factual correctness and suggest additions!

¹² <https://www.cnn.com/2023/03/16/tech/openai-gpt-microsoft-365/index.html>.

mainstream applications are supported by the ChatGPT technology as it might be behind the scenes (much like Microsoft's Bing search engine is currently being powered by GPT-4).

The challenge for academics is to decide just how deep an understanding we need when it comes to technologies such as this (whether they are embedded in other applications or not). Certainly, some basic understanding of the risks and opportunities of a specific technology such as ChatGPT is required. If academics and students do not understand the technology sufficiently, then they might not understand the risks associated with using it.

Some of the risks associated with ChatGPT include the following:

- Potential privacy breaches, not only with user data (including payment information) but also user chats.¹³
- The risk of “hallucinations” resulting in incorrect and/or fictitious responses as discussed earlier.
- Outdated information (given that the training data only goes up to September 2021). Students should understand which version of the underlying model they are using since newer versions are more reliable and have greater flexibility.
- Unintended bias in the training information, resulting in biased responses.¹⁴

Managing the Associated Risks When Using ChatGPT

Would your students know when and why to put the guardrails up and how they might manage the associated risks when using ChatGPT?

Rethinking academic integrity in a technology-driven world. During covid-19, many post-secondary institutions observed increases in the incidence of academic misconduct, often involving technologies. Existing control systems to monitor use of unauthorized aids and plagiarism (two prominent forms of academic misconduct) appear increasingly to be inadequate in the face of many newer technologies. Academics understand the need to continually update our control systems to combat academic misconduct, but how do we get ahead of emerging and more powerful technologies that are increasingly available to our students? It feels like a battle that we are destined not to win. Being vigilant and following through on cases of academic misconduct are extremely important (and we will continue to do so) but are there other tools that we can also use?

We could certainly update the definition of academic misconduct as needed. Some schools are claiming that use of chatbots such as ChatGPT should be cited as an academic offense (specifically linking its use with plagiarism). Others are going further, publicly banning its use¹⁵.

¹³ Indeed, on March 24, 2023, OpenAI noted that they had taken the chatbot offline to fix a bug in the open-source library which allowed both of these to occur <https://openai.com/blog/march-20-chatgpt-outage>.

¹⁴ Sam Altman, CEO of OpenAI posted this on Twitter. In addition, potential social biases are noted as limitations on the OpenAI website.

¹⁵ <https://www.reuters.com/technology/top-french-university-bans-use-chatgpt-prevent-plagiarism-2023-01-27/>.

Is banning the technology the answer? Where do we draw the line and which technologies are on the “wrong side” of that line?¹⁶

Other schools are using curriculum design as a tool to combat academic integrity issues, shifting for instance, the nature of the course deliverables, putting emphasis on handwritten essays and/or oral examinations where academic misconduct is less likely to occur.¹⁷

One thing that is becoming abundantly clear is that historic ways of administering, designing and delivering programs and courses are non-sustainable in a tech-enabled environment. Maintaining academic integrity using these older approaches and practices has reached a tipping point. We need to step back and look at our learning environments through a technology lens.

Careful design of curriculum is critical – more so now than ever. Which human skills and competencies would we like to encourage in a world where machines and new technologies are increasingly being employed? How can we best provide learning and assessment opportunities to allow and encourage our students to develop these, while embracing (perhaps) a fresh take on academic integrity? It is quite a challenge but certainly an interesting and worthy one.

Using ChatGPT to Provide Learning Opportunities That We Would Like to Encourage

As introduced earlier, one way to think about incorporating technologies such as ChatGPT into the classroom is to consider the skills and competencies we want to encourage in our students. These might be skills and competencies that we have always sought to embed (traditional competencies associated with accounting), or they might be new ones. Either way, we have to think about them in the context of technology, and we have to think about the human piece, i.e., how we interact with the technologies, applying unique human skills and competencies and how the technologies can augment human intelligence to achieve efficiencies and better insights.

As an example of a more traditional competence, we might consider the trust that accountants bring to decision-making, for instance around financial statements and sustainability disclosures. Accountants are trusted because they are ethical. They ensure that the data used for decision making is appropriate and that the underlying systems to produce information may be relied upon (i.e., to produce decision-relevant information).

Let’s think about this in a technology-forward environment, where technologies such as ChatGPT and GPT-4 are used. How can people trust accountants if they are using technologies (or systems that are powered by technologies) they do not understand? How is this ethical?

¹⁶ I recall when many schools banned the use of computers and phones in the classrooms and in examinations. Many of us now encourage and support the use of computers and phones in the classroom. I would hazard a guess that we will similarly and eventually have to figure out how to effectively incorporate ChatGPT into our curricula.

¹⁷ Asking students to write their answers using only pen and paper seems extreme to me. In my opinion, it represents a step backwards and an inelegant response to the perceived risks apparently posed by ChatGPT. The cognitive process required to formulate a response to an examination question using only a pen and paper is much different than that required when composing an answer with a computer. Using a pen and paper requires some pre-thought about the structure of the answer, whereas using an application such as Word allows for capturing ideas and subsequent re-arranging and editing to occur. The latter allows for more freedom of thought and creativity (both human centred skills that are ironically supported by using more technology). There is also the speed factor. Using pen and paper is certainly slower. This would make a difference in a time-constrained situation.

How is this responsible? If they do not have the expertise, they at least have to have sufficient understanding to be able to collaborate with someone who does understand the technology. This speaks to the earlier point regarding the need to identify a sufficient level of understanding of the technology – either to use it ethically and responsibly or to be able to collaborate with someone more knowledgeable. This would make an excellent and provocative class discussion.

Another example of a traditional concept that many accounting instructors seek to embed in accounting curriculum is skepticism. Accountants pride themselves on being skeptical – a uniquely human mindset, by the way. Accounting students can apply skepticism when evaluating risks and limitations associated with emerging technologies such as ChatGPT (and GPT-4).

We have a great opportunity to help our students understand what it means to use technologies such as ChatGPT (and GPT-4) in an ethical and responsible way. We (professors, students and professionals alike) are all trying to figure out just what this means. There is no right answer since new technologies keep proliferating and societal views of what is acceptable also continue to evolve. That’s what makes this so interesting.

The more technologies proliferate the learning space (and our communities), the more we should focus on supporting the development of human skills and values (such as curiosity, creativity, agility, adaptability and the ability to collaborate)¹⁸ in our classrooms, providing opportunities to nourish and cultivate them in the safe spaces that our classrooms should be. We have a long road ahead of us – but definitely an interesting one.



¹⁸ <https://cpaleadstheaway.ca/>. See CPA Ethical Mindset, page 13.