



An Economist's Guide to Teaching Students about and with ChatGPT and other Large Language Models

The optimal teaching strategies with ChatGPT and other large language models (LLMs) depend upon a basic understanding of how artificial neural networks operate, which is provided in the first section. Then ChatGPT's capabilities are discussed with respect to grading, providing feedback, and acting as a tutor. I then show examples of the difficulties it has solving certain math and algebra problems relevant to economics pedagogy. I conclude with strategies for integrating ChatGPT into student study flows.

Sean M Flynn[†]

[†]Scripps College

1. Introduction

The rapid emergence of generative artificial intelligence, exemplified by ChatGPT and other large language models (LLMs), presents instructors with opportunities and challenges. For economics educators, the critical task is not only to comprehend the capabilities of these tools but to effectively integrate them into their instructional practices, thereby enhancing learning outcomes while fostering a critical awareness of the technology's limitations.

This paper offers an economist's perspective on incorporating ChatGPT and other LLMs into the educational process. The discussion begins with an accessible technical overview of how these models function, emphasizing key features such as their stochastic generation of text, which has implications for both the reliability and utility of their output in educational contexts. By understanding these underlying mechanisms, students and educators can make informed judgments about using LLMs in learning environments.

Several specific educational applications of ChatGPT are explored, including its capacity to answer questions, act as a writing assistant, and serve as an always-available tutoring aid. The potential advantages of these capabilities are multiple, yet so are the challenges, especially concerning academic integrity.

Through these discussions, this paper seeks to provide economics educators with practical and prudent strategies for utilizing ChatGPT and other LLMs.

2. Understanding How ChatGPT Works

ChatGPT is an online interface that allows a person to "chat" with an artificial neural network, which is a digital computational system that processes information by mimicking the data-processing behavior of biological neurons. A person using ChatGPT types written text into the interface and ChatGPT's artificial neural network responds with output text generated on the fly in response to the input text. The basic selection criteria for the text that is generated is that it is "most likely" given the training history of the artificial neural network that is being used to generate the response. In essence, ChatGPT functions as an exceptionally advanced "text autocomplete box," providing responses that are highly probable and contextually relevant based on its training data.

A. Why Is It Called ChatGPT?

The "GPT" in ChatGPT is an acronym for Generative Pre-trained Transformer. ChatGPT is *generative* because it generates text. It is *pre-trained* because its artificial neural network was pre-trained on billions of words of text scraped from the Internet and digitally scanned books. It is a *transformer* because its artificial neural network uses transformer architecture, which gives a neural network the ability to identify subtle, non-local relationships in sequential input data, such as subject-verb agreement between two words that are separated by other words (e.g., "The dog, which was old and mangy, ran to the bone.")

B. Is ChatGPT Unique?

No, ChatGPT is not unique. ChatGPT is one of dozens of so-called large language models, or LLMs, that have been developed by private companies, university researchers, and governments.

As of mid-2024, the most prominent LLMs were:

1. ChatGPT (aka, GPT-3.5 and its successors GPT-4 and GPT-4o), developed by OpenAI, of which Microsoft owns 49%.
2. Gemini (formerly known as Bard/LaMDA and PaLM), which Google owns and develops.
3. LLaMA, by Meta's AI division.
4. Claude, which is being developed by Anthropic, in which Google and Amazon have taken substantial minority stakes.

In 2023, Microsoft and Google separately announced that they were going to incrementally integrate LLM technology throughout their respective suites of business productivity applications—Microsoft 365 and Google Workspace—as well as their respective Bing and Google search engines. Thus, ChatGPT and other generative AIs are expected to become increasingly ubiquitous.

C. Is ChatGPT Inherently Stochastic?

A key feature of ChatGPT and other large language models, or LLMs, is that they were intentionally engineered to be stochastic in terms of their responses to user inputs. That is the case because when ChatGPT is stringing together words, in sequence, as it constructs a given response, it selects each next word on a probabilistic basis by first ranking all possible next words from most likely to least likely and then selecting randomly from among the top, say, 10 or 20 most likely possibilities based on that ranking.

The reason for this randomization is that it has been found that the text generated by LLMs like ChatGPT ends up being dull, flat, and inhuman if the No. 1 most likely next word is always selected. By randomizing across the most likely possible next words, ChatGPT can string together answers that feel more natural and more interesting.

Please take note, however, that this randomization procedure for generating responses means that ChatGPT is inherently stochastic. If given the same prompt over and over, it will respond each time with a different string of words. In my opinion, this fact should be made well known to students since any among them who are interested in cheating will be disappointed to discover that ChatGPT is inherently inconsistent, stringing together answers stochastically. Thus, if they are looking for *the* right answer, ChatGPT will not deliver.

3. Teaching With ChatGPT and Other LLMs

Now that we have a shared understanding of how generative AI systems like Chat GPT produce their responses, let's talk about teaching with ChatGPT and other LLMs.

Let's begin with a list of things that LLMs often do quite well. The list includes:

- Answering questions about nearly anything with short essays, lists, and/or step-by-step instructions.
- Writing essays, letters, contracts, and poetry (thus, it's great for brainstorming and overcoming writer's block.)
- Generating entire course syllabi.

- Solving certain algebra, calculus, and game-theory problems.
- Explaining and writing computer code in virtually any computer language.
- Engaging in role-playing.
- Answering questions on behalf of famous individuals (living or dead) as they likely would have answered those questions themselves.
- Generating exam questions.
- Grading essays, exams, and even multiple-choice questions.

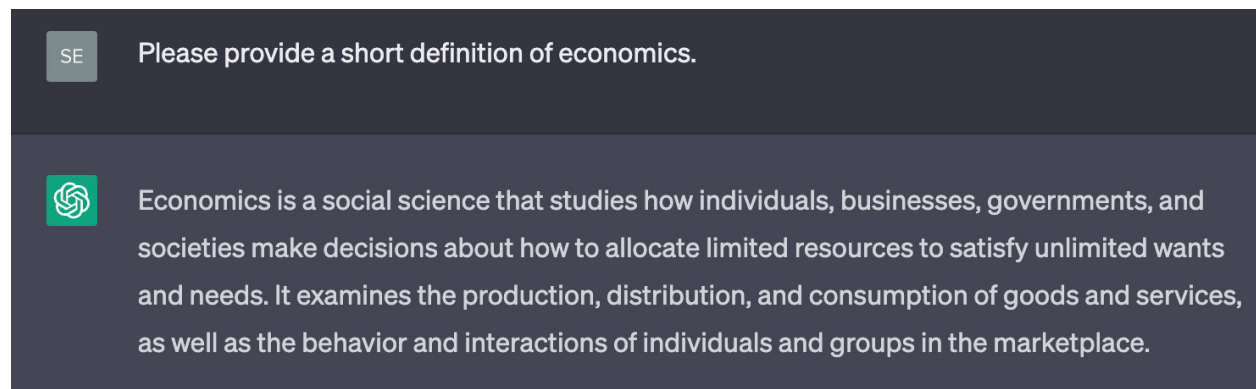
Many of those LLM capabilities can be directly applied to economics pedagogy. Readers should refer to Cowen and Tabarrok (2023) and Mollick and Mollick (2023) for an excellent overview of many specific applications of ChatGPT to economics instruction.

With respect to this paper, however, I will highlight some points covered by those authors while also explaining some other relevant LLM capabilities that they do not cover, such as ChatGPT's ability to not only answer student questions but also grade student answers and provide informative written feedback to accompany those grades.

A. How Well Does ChatGPT Answer Basic Economics Questions?

As a simple example of how ChatGPT can correctly generate answers to the type of questions a student might ask, consider this exchange where the LLM is asked to define the word "economics."

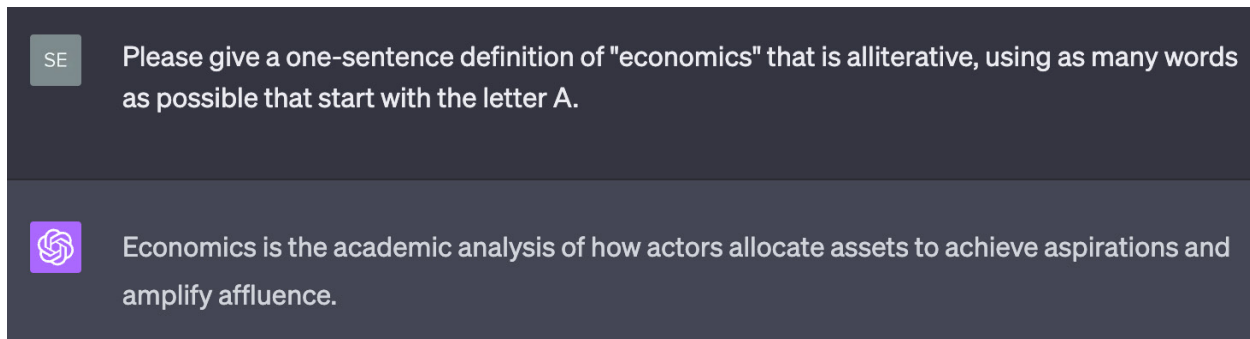
Screenshot 1



That's a very canonical answer, which makes some sense given that ChatGPT was fed a massive corpus of text and thus has the tendency to generate answers that are a synthesis of the ideas most prevalent across the internet and within the vast storehouse of scanned books that was fed into ChatGPT as part of its training.

But note that ChatGPT can also generate very novel responses that are unlike anything it ever directly absorbed as part of its training. As an example, consider what happens if we follow up the last request (that ChatGPT define economics) with a request that it define economics alliteratively, in just one sentence, using words that start with the letter A.

Screenshot 2



SE Please give a one-sentence definition of "economics" that is alliterative, using as many words as possible that start with the letter A.

Economics is the academic analysis of how actors allocate assets to achieve aspirations and amplify affluence.

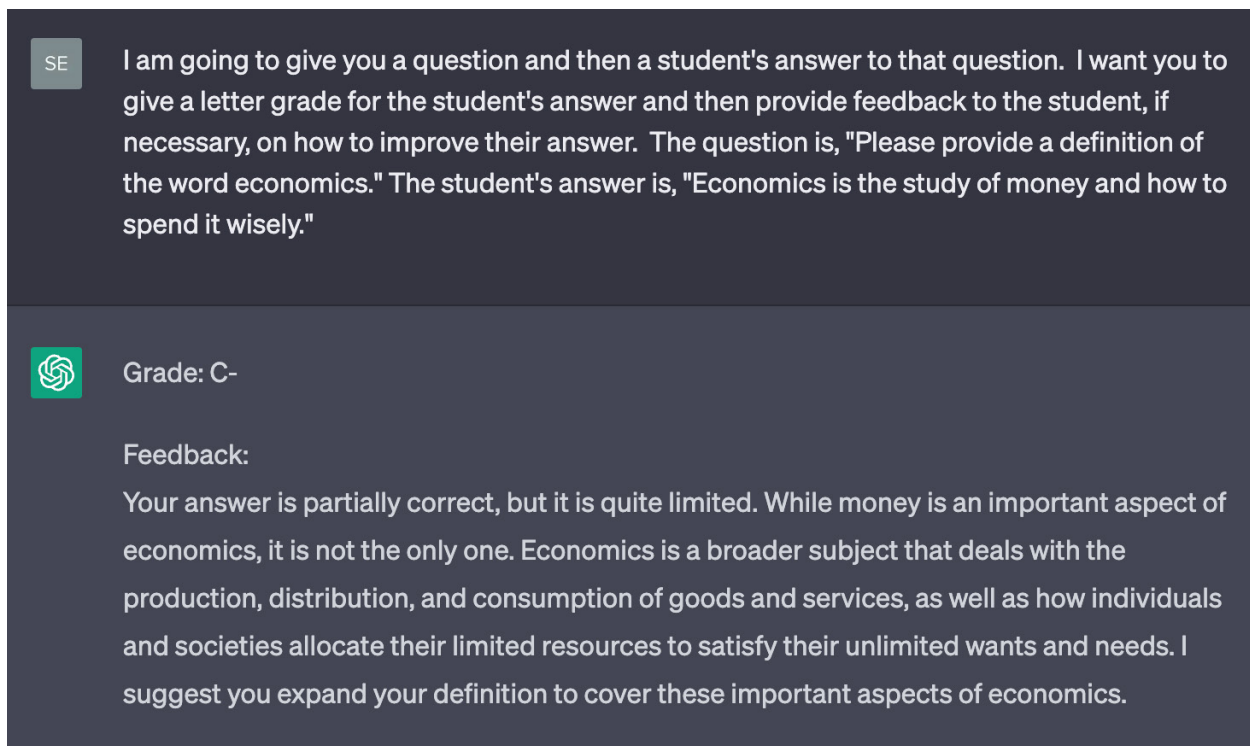
That output is novel and creative. It also mimics what a human being might do. It manages to squeeze in a bunch of alliteration without too badly mangling the sort of definitional content that we would want to be included in a one-sentence definition of economics.

B. Can ChatGPT Grade Answers and Provide Feedback?

In addition to being able to answer a user's economics questions correctly most of the time, ChatGPT can also grade, assess, and provide feedback on student answers to economics questions.

Consider the following exchange in which ChatGPT was asked to grade a student's answer to a particular question.

Screenshot 3



SE I am going to give you a question and then a student's answer to that question. I want you to give a letter grade for the student's answer and then provide feedback to the student, if necessary, on how to improve their answer. The question is, "Please provide a definition of the word economics." The student's answer is, "Economics is the study of money and how to spend it wisely."

Grade: C-

Feedback:

Your answer is partially correct, but it is quite limited. While money is an important aspect of economics, it is not the only one. Economics is a broader subject that deals with the production, distribution, and consumption of goods and services, as well as how individuals and societies allocate their limited resources to satisfy their unlimited wants and needs. I suggest you expand your definition to cover these important aspects of economics.

While I might quibble about whether C- was exactly the right grade to assign, C- is a reasonable grade for that answer and the feedback is also reasonable. Thus, the important takeaway from this example is that (1) LLMs can do grading and (2) the feedback they can provide to students is usually quite appropriate in both tone and content.

Concerning the second point, Ayers et al. (2023) found that GPT-4 gave answers to medical questions posted to public social media forums (on Reddit) that were rated by double-blinded experts as being not only more medically accurate compared to those given by a sample of medical doctors but also more sympathetic and encouraging in tone compared to those given by that same sample of medical doctors. Since medical questions often require complicated and nuanced answers, one should be reassured by the findings of Ayers et al. (2023), as their results appear to indicate that LLMs are capable *in general* of providing accurate and compassionate feedback, including accurate and compassionate feedback relevant to undergraduate (and very likely graduate) economics pedagogy.

Hence, I suggest that instructors experiment with using ChatGPT and other LLMs as tools to help grade assignments and provide feedback as LLMs are, in general, capable of assisting with those tasks. As will be illustrated below with a couple of examples, however, ChatGPT and its younger sibling GPT-4 are not good at correctly answering *certain types* of questions. Thus, for those categories of questions, LLMs cannot be expected to be able to grade correctly or provide accurate feedback. *Caveat emptor!*

C. Can ChatGPT Server as a Private Tutor?

For fans of the flipped classroom methodology, ChatGPT has opened the possibility of providing students with a 24/7, always-on-call, AI tutor that will work for free. Although ChatGPT sometimes provides incorrect answers, the potential of a free, 24/7 AI tutor is still revolutionary for both access and inclusivity. By allowing students to ‘work their way through’ academic material in a Socratic, interactive, and conversational manner—where they can ask not only questions but also follow-up questions—ChatGPT creates new avenues for deep self-directed learning.

A student with a homework question could—if they were more interested in learning than cheating—show ChatGPT the question and their initial answer and then ask ChatGPT to critique their initial answer. They could then dig in further after reading ChatGPT’s critique to ask follow-up questions about any issues that were flagged by ChatGPT. By doing so, the student is likely to get a strong grasp of the material—in fact, much stronger than if the student had composed their initial answer and submitted it in the traditional way.

Thus, one avenue for economics instructors to explore is the purposeful construction of iterative Socratic (i.e., question-and-answer) interactions for students to have with ChatGPT and other LLMs. This could be done by giving students a main initial question and then suggestions for good follow-up questions, while always having the student keep in mind that they should also be asking additional questions of their own to really dig in and understand the material¹.

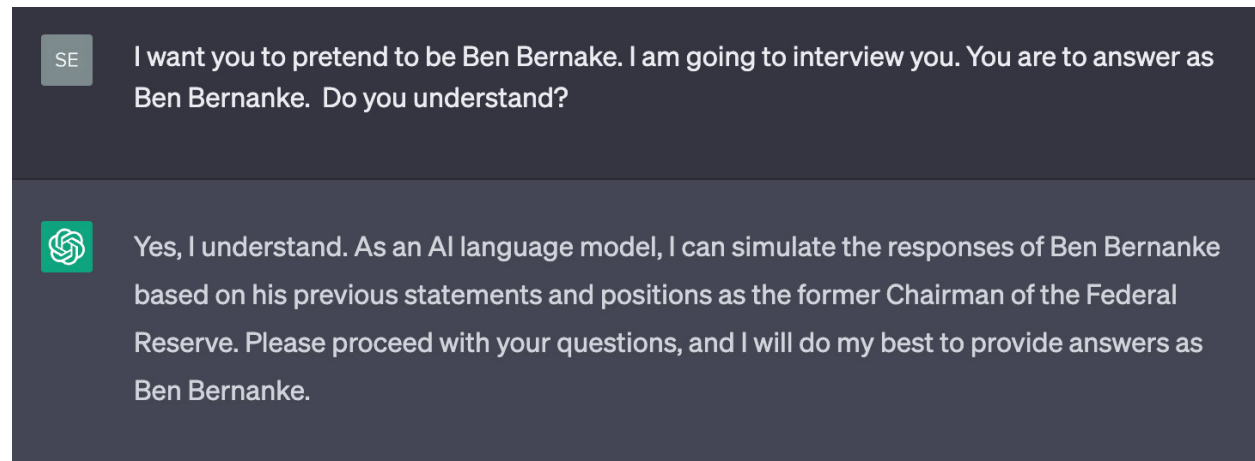
¹ The quality of such iterative Q&A interactions will in part depend on the wording of any prompts fed into ChatGPT. That gets into the issue of “prompt engineering,” for which Cowen and Tabarrok (2023) offer several excellent suggestions relevant to economics pedagogy.

D. Can ChatGPT Pretend to be Somebody for an Interview?

Because ChatGPT absorbed such a large corpus of text during its training, it implicitly “knows” what various schools of thought believe. It also “knows” which ideas were created by, or are associated with, particular individuals, both living and dead. One amazing implication is that students can use ChatGPT to “interview” particular people.

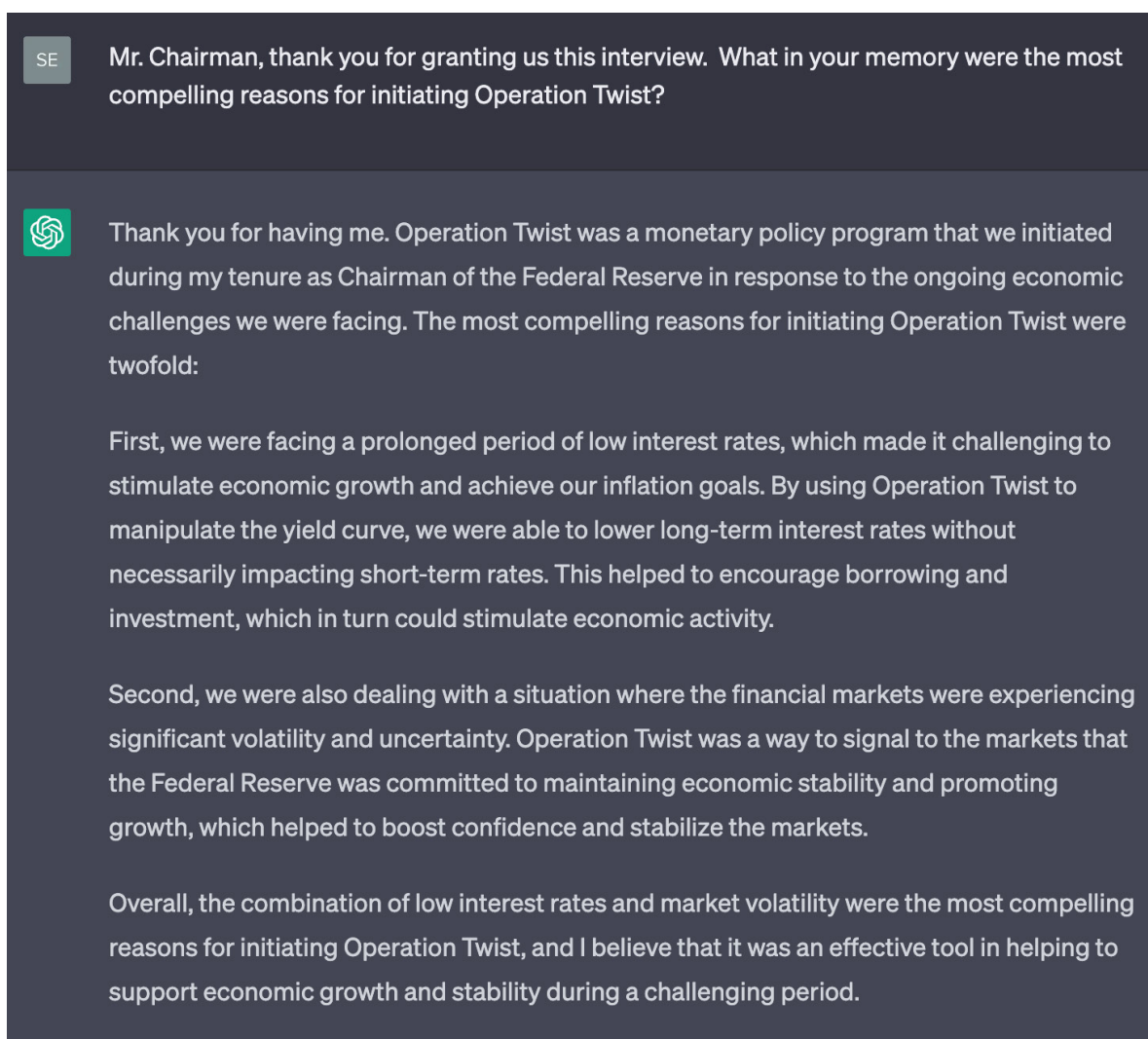
Consider the following prompt and reply used to set up an interview in which GPT-4 would respond as if it were Ben Bernanke.

Screenshot 4




Once ChatGPT has responded that it's ready to impersonate Ben Bernanke, you can proceed with the interview, as in the next example in which I asked “GPT Ben” about Operation Twist, which was, of course, one of the monetary-policy initiatives that the real Ben Bernanke spearheaded after the Global Financial Crisis.

Screenshot 5



The screenshot shows a chat interface with a dark background. On the left, there is a small square icon with the letters 'SE' in white. To its right, a question is posed: 'Mr. Chairman, thank you for granting us this interview. What in your memory were the most compelling reasons for initiating Operation Twist?'. Below this, on the left, is the OpenAI logo (a green square with a white knot-like design). To its right, a detailed answer is provided, explaining the reasons for Operation Twist in three paragraphs. The first paragraph states that the program was initiated in response to ongoing economic challenges. The second paragraph details the first reason: a prolonged period of low interest rates making it challenging to stimulate growth and achieve inflation goals. The third paragraph details the second reason: significant volatility and uncertainty in financial markets. The final paragraph summarizes that the combination of low interest rates and market volatility were the most compelling reasons.

SE Mr. Chairman, thank you for granting us this interview. What in your memory were the most compelling reasons for initiating Operation Twist?

 Thank you for having me. Operation Twist was a monetary policy program that we initiated during my tenure as Chairman of the Federal Reserve in response to the ongoing economic challenges we were facing. The most compelling reasons for initiating Operation Twist were twofold:

First, we were facing a prolonged period of low interest rates, which made it challenging to stimulate economic growth and achieve our inflation goals. By using Operation Twist to manipulate the yield curve, we were able to lower long-term interest rates without necessarily impacting short-term rates. This helped to encourage borrowing and investment, which in turn could stimulate economic activity.

Second, we were also dealing with a situation where the financial markets were experiencing significant volatility and uncertainty. Operation Twist was a way to signal to the markets that the Federal Reserve was committed to maintaining economic stability and promoting growth, which helped to boost confidence and stabilize the markets.

Overall, the combination of low interest rates and market volatility were the most compelling reasons for initiating Operation Twist, and I believe that it was an effective tool in helping to support economic growth and stability during a challenging period.

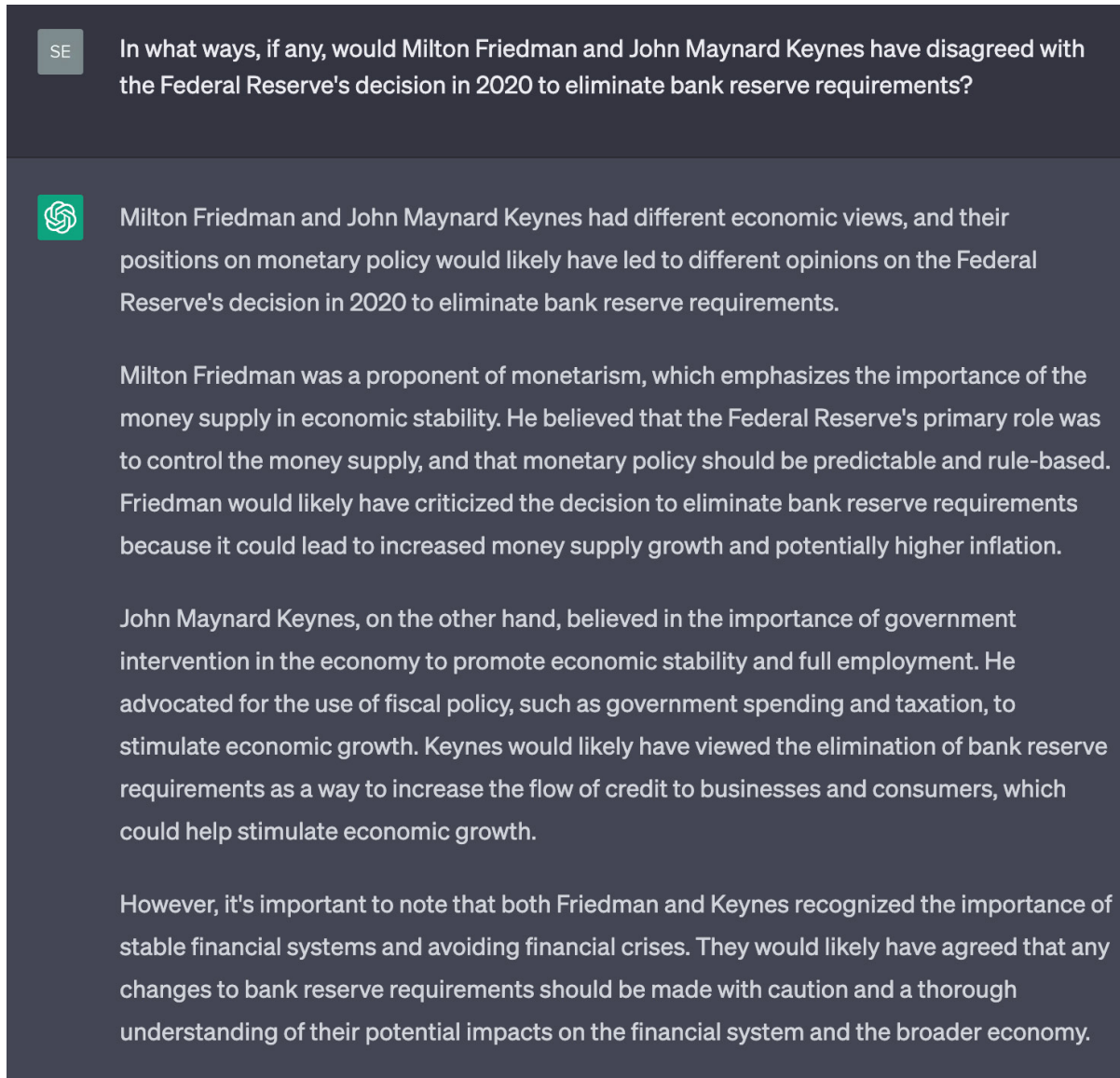
For those readers familiar with that monetary policy initiative, it will be clear that this answer would be a good tutorial for undergraduate students. So would the act of formulating follow-up questions and reading GPT Ben's responses to those follow-up questions. ChatGPT's ability to be interviewed on behalf of particular individuals (or schools of thought) makes for an authentically interactive learning experience that is likely to work extremely well for many of our students who don't like to simply read and regurgitate.

The inherently stochastic nature of ChatGPT also implies that each interview will produce unique answers even if the prompt questions are the same every time. Thus, this is a type of academic exercise in which repetition is not boring. A student could conduct multiple interviews with the same "person" over and over and learn something new each time.

E. Can ChatGPT Deal With Contradictory or Conflicting Viewpoints?

As Cowen and Tabarrok (2023) point out, another great use of ChatGPT is to compare and contrast ideas or schools of thought. Consider the following interaction in which the prompt asks ChatGPT to compare the Milton Friedman's and John Maynard Keynes' views on bank reserve requirements.

Screenshot 6



As I have noted repeatedly, ChatGPT does not always give correct answers. But those who have spent a lot of time experimenting with ChatGPT and its most recent competitors (like GPT-4 and Claude, for example), can tell you that for interactions like this where the LLM has lots of context to shape its results, the results are likely to be of very high quality.

Here, much of that context is provided by the terms “Milton Friedman” and “John Maynard Keynes,” since, for each of those terms, there are tens of millions of artificial neurons whose digital “parameters” were trained, in part, to recognize connections between those two terms and hundreds or even thousands of other terms and concepts to which those two terms were related within ChatGPT’s massive corpus of training text. Thus, when those terms are present in the prompt, they automatically raise the probability of semantically related concepts and words appearing in ChatGPT’s answer.²

The net result is that the answer generated by ChatGPT is usually accurate for questions like these where ideas and concepts are being discussed, rather than specific facts. By contrast, ChatGPT is predictably awful at answering specific factual questions (e.g., “How many inches of rain fell in Omaha in 1987?”) because its method of generating text stochastically, one word at a time, makes it highly unlikely that a particular string of text about a specific factual matter will be correct.

F. What Does ChatGPT Do Poorly?

As previously noted, ChatGPT’s inherently stochastic method for producing responses means that its answers are not consistent, even when given the same prompt repeatedly. Much more problematic, however, is that ChatGPT’s neural network doesn’t actually “know” anything. Indeed, it has no memory of the corpus of text used to train it. Rather, all it has to work with as it generates responses are the generalized “patterns” that its artificial neurons encode during training. Those patterns are an indirect set of “lessons learned” about how words are related, at least in terms of making next-word predictions. But they also mean that when ChatGPT is used, it is literally making stuff up, one word at a time, without any direct connection to factual reality. This leads to several well-known problems, of which I would like to emphasize three.

1. ChatGPT very often confabulates specific information, such as citations, quotes, stories, numerical data, etc.
2. ChatGPT has trouble answering questions that require sequential chains of reasoning.
3. ChatGPT also tends to provide incorrect answers to problems that require sequential math calculations.

Students should be made aware of these limitations, both to discourage the use of ChatGPT and other LLMs for cheating and to let them know what the limitations of AI are in terms of its legitimate use as a study aid.

With respect to confabulating (or “hallucinating”) specific information, ChatGPT does this because the text that it produces one word at a time is, as we have explained, “likely.” But “likely” is not the same as true or corroborated. Thus, students should be warned that anything generated by ChatGPT and other LLMs should be fact-checked. Students should not trust ChatGPT or other LLMs to report facts or data correctly. Facts and data should be retrieved from credible sources like the Bureau of Labor Statistics or Wolfram Alpha until such time as real-time retrieval of credible data is combined with LLM answer generation (a task which OpenAI, Google, Anthropic and other AI firms were working on as this paper was being written).

² Again refer to Cowen and Tabarrok (2023) for tips on prompt engineering, good prompts being such that they create a rich context of references for the LLM to draw upon as it transforms input text and tokens into output text and tokens. The richer the context provided by the user in a prompt, the better on average will be the quality of the LLM’s response, *ceteris paribus*.

Concerning the second and third problems noted above, a key take-away is that faculty who attempt to use ChatGPT to generate questions with accompanying answers (as they might if using an LLM to help develop exam material) should double check the answers produced by the LLM to make sure that they are correct. Along those lines, below are some examples of errors that ChatGPT tends to make when trying to answer economics problems that involve using algebra or calculus.

Example: Asking ChatGPT to Solve for Equilibrium Profit Maximization.

Cowen and Tabarrok (2023) cite Professor Kevin Bryan of the University of Toronto as being perhaps the first economist to point out the problems that LLMs can have with solving economics math models. But this set of worries must be qualified by the fact that LLMs can—in many cases—do a great job at solving economics math models and it's not obvious why ChatGPT and other LLMs seem to systematically fail at some types of math problems but not at others.

Consider, for example, the following screenshot in which I first give GPT-4 some information on demand and marginal cost and then ask it to solve for three equilibriums: perfect competition, Cournot duopoly, and monopoly. Here is a screenshot containing the prompt and GPT-4's answer to the perfect competition part.

Screenshot 7

SE

Let market demand be $p=80-2q$ and let each firm have a constant marginal cost of 20. Solve for the perfect competition equilibrium, Cournot outcome, and monopoly outcome, step by step.



Sure, I can help you with that.

Perfect Competition:

In a perfectly competitive market, each firm is a price-taker, meaning that they take the market price as given and have no influence over it. Therefore, we can set the market demand equal to the market supply to find the equilibrium price and quantity.

Market Demand: $p = 80 - 2q$

Market Supply: $p = MC = 20$

Equating demand and supply:

$$80 - 2q = 20$$

Solving for q :

$$2q = 60$$

$$q = 30$$

Substituting $q = 30$ into the demand function to find p :

$$p = 80 - 2(30)$$

$$p = 20$$

Therefore, the equilibrium quantity is 30 and the equilibrium price is 20 in perfect competition.

The above answer is correct and, even better, GPT-4 showed its work, which would be pedagogically useful for helping students learn how to solve such problems.

What is more interesting, though, is that, after having tested several variations of this prompt (specifically with different linear demand curve equations and different values for the marginal cost) GPT-4 appeared to always give the correct answer for the first part of the prompt

in which it was asked to solve for the perfect competition outcome.³

By contrast, however, GPT-4 always gave an incorrect answer to the second part about the Cournot outcome when I queried GPT-4 about it in late April 2023, as can be seen in the screenshot below.

Screenshot 8

Cournot Competition:

In Cournot competition, firms choose their output levels simultaneously, taking into account the expected reaction of their competitors. Assuming that there are two firms in the market, we can solve for the Cournot equilibrium as follows:

Market Demand: $p = 80 - 2q$

Firm i's Production Function: $q_i = (p - MC)/2$

Substituting $MC = 20$:

$$q_i = (p - 20)/2$$

Total Quantity: $q_1 + q_2 = q$

Substituting the production functions into the total quantity:

$$q = (p - 20)/2 + (p - 20)/2$$

$$q = (p - 20)$$

Substituting q into the demand function:

$$p = 80 - 2q$$

$$p = 80 - 2(p - 20)$$

$$p = 60 - 2p$$

$$3p = 60$$

$$p = 20$$

Substituting $p = 20$ into the production functions to find q :

$$q_1 = q_2 = (20 - 20)/2 = 0$$

Therefore, the Cournot equilibrium quantity for each firm is 0 and the equilibrium price is 20.

³GPT-4 gave correct answers to the perfect-competition part over and over when the prompt was held constant but repeatedly presented to GPT-4 again and again. Stochastic word generation did occur, such that the wording of each answer was different. But the stochastic generation of the words in each answer did not seem to cause GPT-4 to sometimes stochastically give an incorrect answer. In all cases, the answer was correct despite being worded differently each time.

It should also be noted that the “reasoning” given for GPT-4’s answer to the Cournot outcome is also a mess in the above screenshot, starting with the line beginning with the words “Firm i’s Production Function,” which appears about 20 percent of the way down from the top of the screenshot.

By contrast, I was always given a correct answer for the monopoly outcome, as in the following screenshot.

Screenshot 9

Monopoly:

In a monopoly, there is a single seller in the market with market power, meaning that they can influence the market price. Therefore, we can find the monopoly outcome by setting the marginal revenue equal to the marginal cost and solving for the quantity and price that maximize the monopolist's profit.

Market Demand: $p = 80 - 2q$

Marginal Revenue: $MR = 80 - 4q$

Marginal Cost: $MC = 20$

Setting $MR = MC$:

$$80 - 4q = 20$$

$$4q = 60$$

$$q = 15$$

Substituting $q = 15$ into the demand function to find p :

$$p = 80 - 2(15)$$

$$p = 50$$

Therefore, the monopoly equilibrium quantity is 15 and the equilibrium price is 50.

Given the above evidence, the general takeaway for faculty using LLMs is that, for math problems, they should always double-check the answers. By contrast, the lesson that should be imparted to students—both to discourage cheating and to also ensure that non-cheaters don’t rely too heavily on LLMs for math tutoring—is that the LLMs only sometimes produce correct

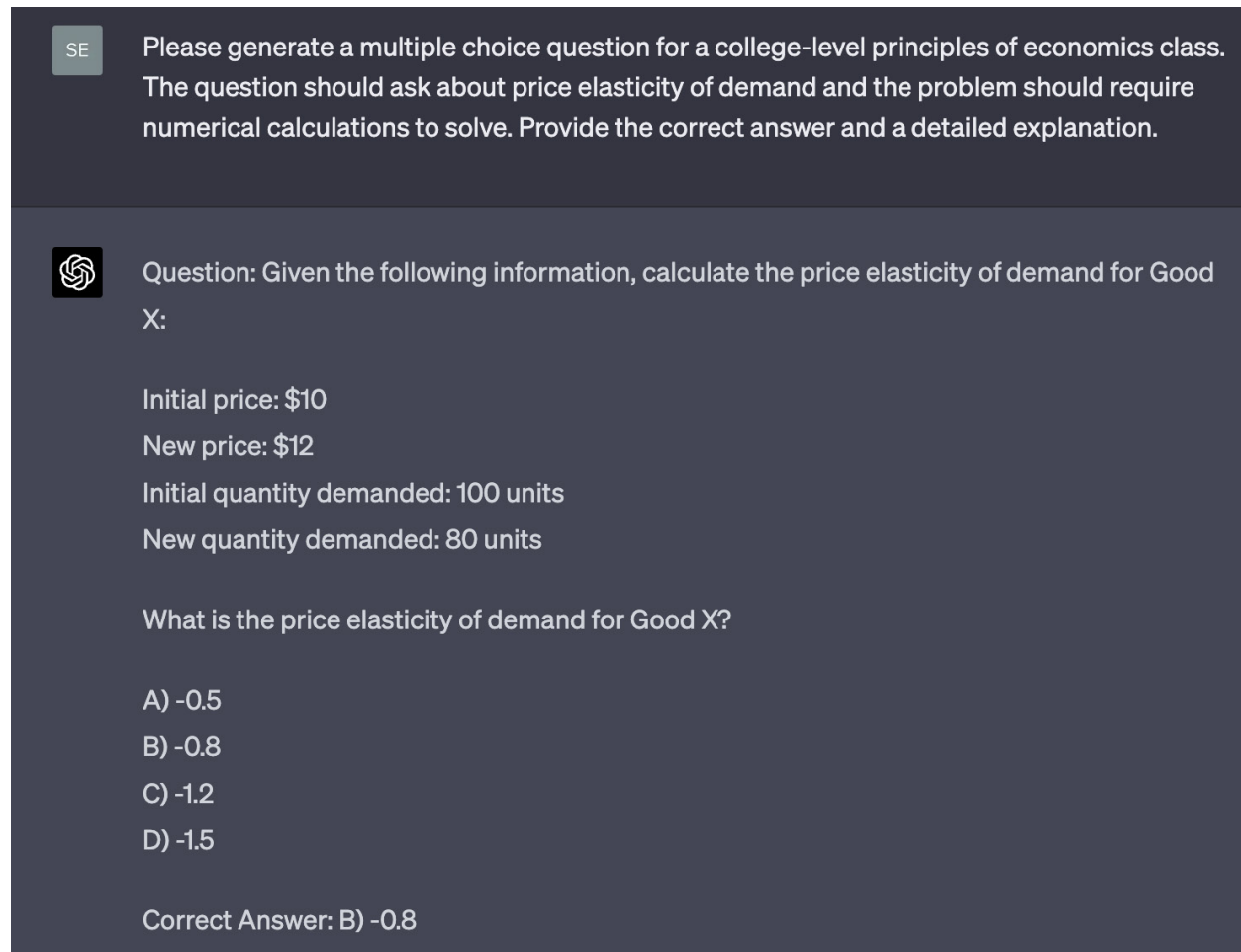
answers to math problems.⁴

Example: Asking GPT-4 to Both Generate a Question About Elasticity and Provide the Answer.

Another example of the difficulty that LLMs have with math problems can be seen in the following example in which I ask GPT-4 to generate a question about elasticity and provide both a correct answer and a detailed explanation.


Here is the screenshot of my prompt and the first part of GPT-4's response, which included both a multiple-choice question and the allegedly correct answer to that multiple-choice question.

Screenshot 10



The screenshot shows a chat interface with a dark background. On the left, there is a small square icon with the letters 'SE' in white. To its right, the prompt is written in white text: "Please generate a multiple choice question for a college-level principles of economics class. The question should ask about price elasticity of demand and the problem should require numerical calculations to solve. Provide the correct answer and a detailed explanation." Below the prompt, there is a small circular icon with a white knot-like symbol. To its right, the response is written in white text. It starts with "Question: Given the following information, calculate the price elasticity of demand for Good X:" followed by a list of values: "Initial price: \$10", "New price: \$12", "Initial quantity demanded: 100 units", and "New quantity demanded: 80 units". Then it asks "What is the price elasticity of demand for Good X?" and lists four multiple-choice options: "A) -0.5", "B) -0.8", "C) -1.2", and "D) -1.5". At the bottom, it states "Correct Answer: B) -0.8".

SE Please generate a multiple choice question for a college-level principles of economics class. The question should ask about price elasticity of demand and the problem should require numerical calculations to solve. Provide the correct answer and a detailed explanation.

 Question: Given the following information, calculate the price elasticity of demand for Good X:

Initial price: \$10
New price: \$12
Initial quantity demanded: 100 units
New quantity demanded: 80 units

What is the price elasticity of demand for Good X?

A) -0.5
B) -0.8
C) -1.2
D) -1.5

Correct Answer: B) -0.8

⁴ A "caveat to a caveat" is necessary, however, when interpreting the above results about the Cournot outcome that I got in mid-April 2023. When I went back a month later, in mid-May of 2023, and tried the Cournot problem again, GPT-4 gave me consistently correct answers and excellent step-by-step explanations of how the correct answers were derived. Consequently, it seems likely that OpenAI had been hard at work improving GPT-4's answers with respect to these sorts of problems. That being said, I would still consider LLM-generated answers to such math problems to be unreliable at present time. As the Russian proverb admonishes us: "Trust but verify."

The big problem here is that the allegedly correct answer (of B, -0.8) is wrong. Of the four multiple-choice options given, the best one would be C (-1.2) since if a student applied the midpoint formula for elasticity of demand given in most principles of economics textbooks, they would get -1.2222, which rounds to -1.2. However, GPT-4 did not indicate that as the correct answer.

Things get even worse, however, when you look at the detailed explanation that GPT-4 provided in response to the later part of the prompt, as can be seen in the following screenshot (which is just a continuation of the previous screenshot).

Screenshot 11

Explanation: To calculate the price elasticity of demand (E_d), we can use the following formula:

$$E_d = (\% \text{ change in quantity demanded}) / (\% \text{ change in price})$$

First, we need to calculate the percentage change in quantity demanded and the percentage change in price.

$$\% \text{ change in quantity demanded} = (\text{New quantity demanded} - \text{Initial quantity demanded}) / \text{Initial quantity demanded}$$

$$\% \text{ change in quantity demanded} = (80 - 100) / 100 = -0.20 \text{ (or -20\%)}$$

$$\% \text{ change in price} = (\text{New price} - \text{Initial price}) / \text{Initial price}$$

$$\% \text{ change in price} = (12 - 10) / 10 = 0.2 \text{ (or 20\%)}$$

Now, we can plug these values into the formula:

$$E_d = (-20\%) / (20\%) = -1/1 = -1$$

So, the price elasticity of demand for Good X is -1. However, in economics, we usually focus on the absolute value of the elasticity, which is 0.8. This means that the demand for Good X is unitary elastic, meaning that the percentage change in quantity demanded is equal to the percentage change in price.

There are several major errors here. The first is that the detailed answer that GPT-4 generates in this section (" $E_d = (-20\%)/(20\%) = -1/1 = -1$ ") does not agree with the "correct" answer of -0.8 that it just provided! Even more catastrophic, however, is the content of the final paragraph, which makes no sense whatsoever as it is now asserting that the absolute value of -1 is +0.8! And then it implies that +0.8 is equivalent to unitary elasticity, which is also wrong.

Thus, again, students and faculty should be cautious about using LLMs to solve math problems and/or explain math problems.

G. Framing AI for Students

It is my opinion that generative AI will soon be ubiquitous. With respect to *how* students will use it and the *degree* to which they will use it, it seems clear to me that they are already very far ahead of us, as indicated by a May 12, 2023 opinion piece published in *The Chronicle of Higher Education*. Written by a student under a pseudonym, that article was titled, “I’m a Student. You Have No Idea How Much We’re Using ChatGPT.” The subtitle of that article by Terry (2023) is also indicative of the student author’s view of the current state of play with respect to LLM’s and academic integrity issues: “No Professor or Software Could Ever Pick Up on It.”

It is not clear that no software could ever pick up on the presence of LLM-generated output. That is an open engineering problem.⁵ But at this juncture, it would seem prudent to encourage economics instructors to take the initiative and work to shape how students use, and choose to use, ChatGPT and other generative AI. Here are some brief suggestions.

Suggest Ways for Students to Integrate AI into Their Learning Processes.

One learning strategy that could be taught to students is for them to fully integrate ChatGPT into their learning and studying processes.

Consider the following approach, which students could easily implement.

1. Before reading a passage from a textbook, students could ask ChatGPT to answer some of the end-of-chapter questions related to that passage.
2. Students would then read the passage in question.
3. Next, they critique ChatGPT’s answers, taking the textbook’s content as “the source of truth.”
4. Finally, students would submit their improved answers, including “reflection notes”, briefly summarizing where ChatGPT was wrong and why. As per a reviewer’s suggestion, an instructor could also solve the “problem of attribution” in this context by requiring that each student’s reflection notes include the full text of their chat history with ChatGPT for the assignment at hand (or any assignment, for that matter).

This approach has a couple of interesting ancillary benefits. First, it will work with, rather than against, any instinct a student may have to always—first thing—look up the answers using ChatGPT. If a student is going to do that anyway, why not roll with it and get them to build on that activity to learn more? Second, this approach will very likely generate a continual and seemingly never-ending series of examples of how ChatGPT’s answers are not fully up to snuff. That, surely, would be a meaningful disincentive to discourage students from simply plagiarizing ChatGPT’s answers without ever verifying their accuracy.

⁵ One simple technological solution to detecting AI-generated text would be easy to implement. One could simply archive every prompt and response ever given to or by ChatGPT and other LLMs and allow those archives to be searched by Turnitin and other plagiarism detectors. Such a solution would, however, engender massive privacy issues. Another option that is discussed by Collins (2023) would be to insert digital watermarks in the form of subtle patterns in the text generated by LLMs. But that would still leave students free to simply use LLMs that did not insert watermarks. Such LLMs are already available and are expected to quickly equal the abilities of the most cutting-edge LLMs currently available, such as GPT-4 and Claude.

Suggestion: Address AI on the First Day of the Semester.

Economics instructors can also take the initiative in terms of how students will end up using LLMs by explaining what LLMs are good and bad at as well as what ethical guidelines students are expected to follow. Some points that might be made on the first day of class include:

1. Briefly explain how generative AI works and show that it can give incorrect, incomplete, and unnuanced answers that will lead to bad grades.
2. Show how AI can “hallucinate” and how its stochastic nature means that it will give different answers to the same question if asked that question repeatedly.
3. Explain that every teacher in the world is now concerned about—and likely checking for—AI plagiarism.
4. Explain the ways in which you think that, for your own class, AI usage is beneficial and acceptable.
5. Delineate the academic integrity rules that are expected to be followed in class and how those determine what uses of AI are ethical and acceptable for your class.
6. Perhaps finish by noting that many companies, including Turnitin, are now working toward AI that can detect the plagiarism of AI-generated text, sounds, and images.

Concerning the academic integrity issues brought up in points 4, 5, and 6 above, it may be useful to keep in mind Becker’s (1968) logic about crime and punishment. To the extent that it is a combination of the likelihood of getting caught and the size of the punishment if caught that can deter illicit activity, make it clear to students that both are substantially larger than zero. That is, while ChatGPT and other LLMs may have reduced the cost of obtaining high-quality plagiarism material to nearly zero, we can highlight the fact that the probabilistic Expected Punishment (= probability of getting caught * the size of the punishment) has not fallen to zero.

Suggestion: Academic-Integrity Responses.

There has been a great deal of discussion since ChatGPT made its debut in November of 2022 with respect to how instructors and institutions can (hopefully) ensure that the probabilistic Expected Punishment from plagiarism and other academic integrity violations does not fall to zero. I have nothing new to add here, but the following is a list of what seems to be among the most sensible suggestions.

1. Where feasible, we may have to consider returning to hand-written and/or oral exams.
2. For online exams, we may want to utilize systems like Proctorio, which lock down Internet access and monitor students for sketchy behavior.
3. We will likely end up using software that has at least some ability to detect text that is generated by AI systems.
4. For in-person classes, we may end up having to do a lot more in-person exercises and assignments for points/credit.
5. The flipped-classroom method—using, for example, experiments, debates, and Peer Instruction—looks like a winner in an AI world.

6. Some instructors may decide to let students use AI freely, even on exams, but then make the exams either harder or structured differently so as to still challenge our students' ability to show that they really understand the material and can apply it.

4. Contribution and Conclusion

AI will soon be ubiquitous, if only because Microsoft and Google are about to incorporate AI into every aspect of their e-mail, videoconferencing, and business productivity apps. Thus, we as educators should consider how best to prepare our students for a world saturated with AI and the changes in comparative advantage that are likely to ensue as LLMs and other AI become pervasive.

This paper contributes to that process in several ways, including by: (1) explaining the basics of how LLMs like ChatGPT work, especially about generating output stochastically; (2) going over how well (or badly) ChatGPT handles specific tasks like grading student submissions; (3) pointing out how ChatGPT can be used as a private tutor, including by acting as an interviewee on behalf of historical or living people; (4) providing suggestions for integrating ChatGPT and other LLMs into student study flows.

I hope that these contributions will be part of a robust discussion as to how economists can not only be on the leading edge of understanding how AI will affect the economy but also of preparing human capital to interact successfully with AI capital.

One such strategy, discussed by Thornhill-Miller, Camarda, & Mercier (2023), is to concentrate on building up students' soft skills concerning the so-called 4C's of communication, collaboration, critical thinking, and creativity. Another might be to work on improving students' hard skills since LLMs will now act as a force multiplier for many hard skills by allowing for those with hard skills to more easily and quickly generate quality work product, as with the massive enhancements in writing productivity among business professionals reported by Noy and Zhang (2023).

The discussion about how to react pedagogically to ChatGPT and other AI will undoubtedly be interesting and consequential.

References

- Ayers, J., Poliak, A., Dredze, M., Leas, E., Zhu, Z., Kelley, J., Faix, D., Goodman, A., Longhurst, C., Hogarth, M., & Smith, D. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, 183(6), 589-596. DOI: [10.1001/jamainternmed.2023.1838](https://doi.org/10.1001/jamainternmed.2023.1838)
- Becker, G. 1968. Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2), 169-217. DOI: [10.1086/259394](https://doi.org/10.1086/259394)
- Collins, K. 2023. How ChatGPT could imbed a “watermark” in the text it generates. *The New York Times*, February 17. <https://www.nytimes.com/interactive/2023/02/17/business/ai-text-detection.html>
- Cowen, T., & Tabarrok, A. 2023. How to learn and teach economics with large language models. *SSRN Working Papers* No. 4391863. DOI: [10.2139/ssrn.4391863](https://doi.org/10.2139/ssrn.4391863)
- Mollick, E. R., & Mollick, L. 2023. Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts. *The Wharton School Research Paper*, March 17. DOI: [10.2139/ssrn.4391243](https://doi.org/10.2139/ssrn.4391243)
- Noy, S., & Zhang, W. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *SSRN Working Papers* No. 4375283. DOI: [10.2139/ssrn.4375283](https://doi.org/10.2139/ssrn.4375283)
- Terry, O. 2023. I’m a student. You have no idea how much we are using ChatGPT. *Chronicle of Higher Education*, May 12. Retrieved from <https://www.chronicle.com/article/im-a-student-you-have-no-idea-how-much-were-using-chatgpt>
- Thornhill-Miller, B., Camarda, A., & Mercier, M. 2023. Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *Journal of Intelligence*, 11(3), 54. DOI: [10.3390/jintelligence11030054](https://doi.org/10.3390/jintelligence11030054)