ChatGPT has Mastered the Principles of Economics: Now What?

Wayne Geerling[1], G. Dirk Mateer[2], Jadrian Wooten[3] and Nikhil Damodaran[4]

Abstract

ChatGPT is a natural language processing model that can generate conversational style responses to user inputs. It also performs extremely well on a popular standardized assessment used in principles of economics courses around the country. ChatGPT ranked in the 91st percentile for Microeconomics and the 99th percentile for Macroeconomics when compared to students who take the TUCE exam at the end of their principles course. The emergence of artificial intelligence presents a significant challenge to traditional assessment methods in higher education. It is not possible to evaluate students' intellectual ability through artificial intelligence, as it only reflects their ability to access information. An important implication of this finding is that educators will likely need to redesign their curriculum in at least one of the following three ways: reintroduce proctored, in-person assessments, augment learning with chatbots, and/or increase the prevalence of experiential learning projects that artificial intelligence struggles to replicate well.

JEL Codes: A22, I21

Keywords: TUCE, ChatGPT, assessment, artificial intelligence, academic integrity

---

[1] Corresponding author: Professor of Instruction, Department of Economics, Room 2.102 B, University of Texas at Austin. Email: wayne.geerling@austin.utexas.edu

[2] Professor of Instruction, Department of Economics, Room 3.102 B, University of Texas at Austin. Email: dmateer@utexas.edu

[3] Collegiate Associate Professor of Economics, Department of Economics, Pamplin 3034, Virginia Polytechnic Institute and State University. Email: jadrian@vt.edu

[4] Lecturer, Department of Economics, Cabin 7, T4 Academic Block, OP Jindal Global University. Email: ndamodaran@jgu.edu.in

I. Introduction

On November 30, 2022, OpenAI launched ChatGPT (Generative Pre-trained Transformer), a chatbot that quickly gained attention for—among other things—its potential to disrupt traditional assessment methods. ChatGPT allows the user to enter a prompt and receive a unique, detailed response on a wide range of knowledge domains. The chatbot has already demonstrated its ability to pass a medical licensing exam (Gilson et al. 2023), the law school admissions test (Choi et al. 2023), and a common assessment used in introductory physics courses (West, 2023). Can ChatGPT's responses surpass human knowledge among undergraduate students studying economics?

To evaluate this, we use the Test of Understanding in College Economics (TUCE), published by the National Council on Economic Education (NCEE) and in use across the United States for more than 50 years. It is one of the most widely used assessment tools for basic economic knowledge and consists of two versions: one covering microeconomic concepts and one covering macroeconomic concepts. Each version of the test has 30 multiple-choice questions with four answer choices each. Both versions include three questions covering international economics, but the questions are unique to each version. The TUCE is a norm-referenced measure that can be used to compare students' knowledge levels across a wide range of abilities. A score of around 50% is desirable for research purposes, as it provides appropriate levels of item discrimination and test reliability. A score of less than 50% does not necessarily indicate a failing level of knowledge in a course, as instructors may prioritize different concepts from those tested in the TUCE. By using the TUCE as both a pre- and post-test assessment, educators can measure perceived learning during the semester (Smith and Wagner, 2018).

ChatGPT operates using algorithms that process data, allowing it to string words together in response to a prompt. Unlike humans, ChatGPT has access to vast troves of information available on the internet and uses large language modeling to recognize patterns in the words in each prompt to mimic human writing when dispensing knowledge.[5] While ChatGPT is a powerful tool, its abilities are limited to the pool of information it has been trained on. ChatGPT creates responses to user prompts using a transformer-based neural network architecture based on the training data to generate contextually appropriate and coherent responses. ChatGPT doesn't actually "know" anything, but instead generates responses based on probabilities assigned to each word in the vocabulary, which are calculated through a process of iterative training on a large corpus of text. In this paper, we assess ChatGPT's performance on the microeconomics and macroeconomics versions of the TUCE and compare it to the results of college students.

While not the focus of this paper, we would be remiss if we did not mention the potential mischief ChatGPT will cause in the short term. To understand the potential impact of ChatGPT on academic integrity, it is important to acknowledge that cheating is not a

---

[5] The current version of ChatGPT was modeled on data from the internet and other sources, up until the end of 2021.

2

new issue, and ChatGPT is simply the latest tool that can be used for a variety of purposes, ethical considerations aside. Instructors have always been asked to consider how to effectively assess student learning, but now must contend with ChatGPT's capabilities. Instructors are the guardians of the learning environment and chatbots are "smarter" than most students. The first step instructors should take is to minimize cheating. Emphasizing proctored assessments is a good way to ensure that students earn scores that reflect their understanding while reducing academic integrity issues. Second, and much more importantly in the long run, instructors should develop ways to utilize chatbots to augment learning. The second step is evolving as we write and is briefly touched upon in a later section. Economic educators must embrace this technological disruptor in order to better prepare our students for the jobs of the future. One way to accomplish both goals is to emphasize experiential learning and authentic learning projects in which students create their own content. When learning is meaningful and relevant, students have less incentive to cheat.[6]

In the following sections, we briefly review the literature on the role of chatbots in education and then compare ChatGPT's performance on the TUCE with the results achieved by college students after completing a semester of their principles course. We conclude by offering some practical advice on identifying alternative assessments that complement ChatGPT as a learning tool.

II. The Role of Chatbots in Education

Chatbots are a technology application that promote interpersonal communication and learning. They provide information and knowledge through interactive methods and easy-to-operate interfaces (Hwang & Chang, 2021). With the exponential growth in the mobile device market over the past decade, the popularity of chatbots is being driven by their ability to provide an interactive medium through which to learn, one not constrained by time and place (Zhou et al., 2020). A Georgia Tech computer science professor made headlines in 2016 for using artificial intelligence to build a virtual teaching assistant (Goel & Polepeddi, 2018). The chatbot known as "Jill" received very positive student evaluations, and students only seemed to suspect something was amiss when their teaching assistant responded quickly at all hours of the day.

Interaction with technologies, either by natural language or speech, is possible because as technology develops, users become more used to interacting with digital entities. Chatbots are now used across a wide range of domains, including marketing, customer service, technical support, education and training (Smutny & Schreiberova, 2020). Personal digital assistants like Siri (Apple), Alexa (Amazon), Cortana (Microsoft) and Google Assistant (Google) lie at the forefront of technology in voice recognition and 'artificial intelligence' and have effectively replaced much of the day-to-day tasks once performed by assistants or secretaries (Smutny & Schreiberova, 2020). The use of digital technologies is now expected by the current generation of young people who were born into an era of the internet and smartphones (Selwyn, 2021).

---

[6] https://theconversation.com/motivation-is-a-key-factor-in-whether-students-cheat-155274

Despite the global proliferation in the use of chatbots, studies exploring the benefits of using chatbots in educational settings have only recently emerged (Ferrell & Ferrell, 2020). These benefits include providing users with a pleasant learning experience by allowing for real-time interaction (Kim et al. 2019), enhancing peer communication skills (Hill et al. 2015), improving the learning efficiency of learners (Wu et al. 2020) and helping instructors manage large in-class activities (Schmulian & Coetzee, 2019).

With the advent of AI-type technology, scholars are now able to apply machine learning and natural language technology to the creation of chatbots, making their application in education a new topic of academic research (Følstad & Brandtzæg, 2017). Recent empirical studies have focused on understanding the optimal role for chatbots. In a study of educational chatbots for Facebook messenger to support learning, Smutny and Schreiberova (2020) highlight the possibility for chatbots to become a smart teaching assistant in the future. Other studies have examined the use of chatbots in language learning. Based on a review of 25 empirical studies, Huang et al. (2021) find that educational chatbots can foster students' language learning via interaction activities underpinned by intended learning objectives. In a similar study, Kim et al. (2019) conclude that chatbots have a positive effect on students' communication skills by expanding the quantity of their interactions, increasing their motivation, and raising their interest in learning.

Chatbots have come a long way in the last two decades. The rise of machine learning with access to very powerful computers and processing power able to train these datasets form the backbone of these systems. Coupled with 'natural language processing', this has paved the way for chatbots to be introduced into the field of education via digital transformation. Because of its scalability and adaptability, it offers unique possibilities as a communication and information tool for digital learning (Wollny et al. 2021). While it's not exactly clear how this field will evolve in future, as these machine-learning driven systems become more advanced and capable of replicating a broader range of human-like traits, there will be a greater acceptance of its use in shaping the education landscape of the 21st century.

III. Comparing the TUCE Results of ChatGPT and College Students

The National Council on Economic Education (NCEE) created the "Test of Understanding of College Economics" (TUCE) and an accompanying examiner's manual to allow instructors to compare their students' results with those of post-secondary students from across the country (Walstad, Watts, and Rebeck, 2007). In order to make these comparisons, the authors normalized thousands of students from various institutions based on a 30-question assessment that was given at the start and end of the term. The purpose of these pre- and post-tests was for educators to measure learning over the semester, including the impacts of changing the structure of the class away from chalk-and-talk (Emerson & Taylor, 2004, Boyle & Goffe, 2018).

Additionally, the normed sample provides a baseline understanding of the level of knowledge that the average college student in the United States has at the beginning and end of their economics principles courses. On average, student performance improves over the course of a semester as students go from answering an average of

4

9.39 questions correctly at the start of the term to an average of 12.77 questions correctly at the end of their principles of microeconomics course. For macroeconomics, students improved from 9.80 to 14.19 questions. Despite a full semester learning economics principles, most students answer around 40-50% of questions correctly. Figures 1 and 2 show the distribution of pre- and post-test scores for both the microeconomics and macroeconomics version of the exam. Given these distributions, where would a large language model like ChatGPT place if it was administered the TUCE?
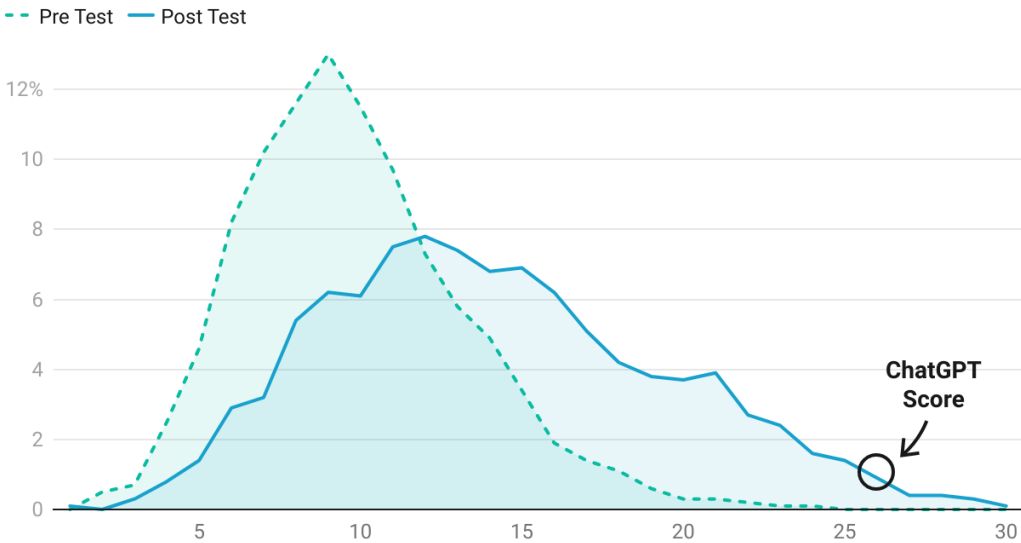
Figure 1: Distribution of Pre- and Post-test Scores on Microeconomics TUCE-4: Matched Sample
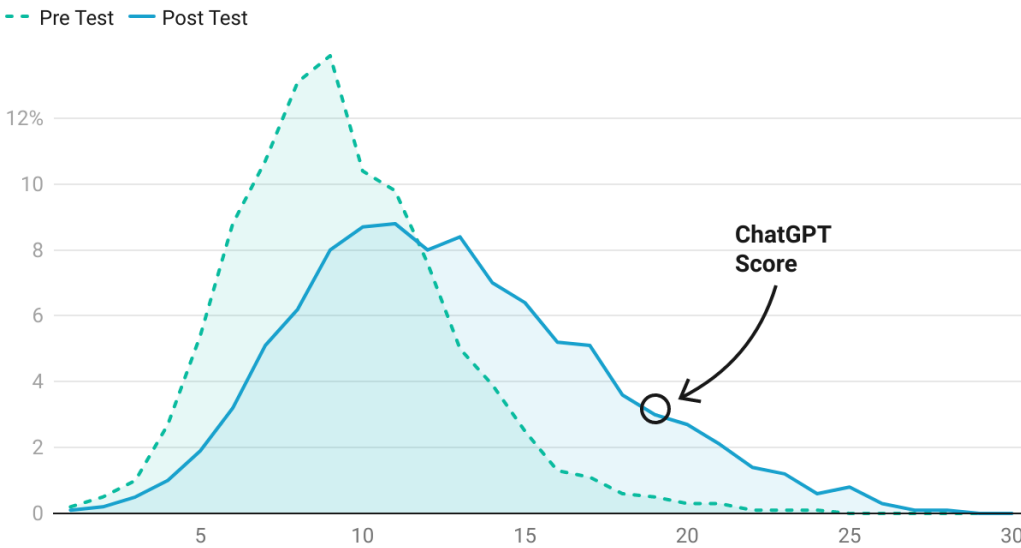
Note: Data presented in Table 4 of the TUCE Examiner's Manual (Walstad, Watts, and Rebeck 2007). The distribution of test scores is based on 3,255 college and university students who took the microeconomics version of the TUCE-4 as a pretest and posttest.

Created with Datawrapper

Figure 2: Distribution of Pre- and Post-test Scores on Macroeconomics TUCE-4: Matched

Note: Data presented in Table 5 of the TUCE Examiner's Manual (Walstad, Watts, and Rebeck 2007). The distribution of test scores is based on 2,789 college and university students who took the macroeconomics version of the TUCE-4 as a pretest and posttest.

Created with Datawrapper

On February 8, 2023, the authors conducted a new chat session on ChatGPT. They provided one question from each of the two versions of the TUCE at a time, along with its answer choices. ChatGPT returned an answer, which was recorded as correct if it matched the TUCE answer key, and incorrect if it was wrong or if multiple answers were provided. The authors didn't assign any partial credit on ChatGPT's response since the TUCE is administered as a multiple-choice test to students in a proctored environment. Figure 3 illustrates the text input and the results for Question 2 on the microeconomics exam.

In our trial, ChatGPT answered 19 of 30 microeconomics questions correctly and 26 of 30 macroeconomics questions correctly, ranking in the 91st and 99th percentile respectively. The incorrect responses often included odd behavior, such as when ChatGPT claimed that all answer choices were correct or provided an answer that was not among the four options. This sort of behavior isn't likely to occur among students taking a multiple choice test. It should also be noted that ChatGPT could not process images at the time of this writing, which resulted in one microeconomics question being provided with missing context.[7] We have included a table in the appendix for both forms of the TUCE which states the concept being tested for each question and whether ChatGPT answered the question correctly or not.[8]

Figure 3: ChatGPT Interface Demonstrating Question and Answer Methodology

---

[7] Despite not having access to the graph, ChatGPT responded with the correct answer.

[8] Due to copyright restrictions on TUCE material, we are unable to include the full-text of the questions. We are happy to send our findings to anyone upon request. The question order presented in the appendix is the same question order presented in the 4th edition of the TUCE available online through the Council of Economic Educators.

6

JA: Suppose a city facing a shortage of rental apartments eliminates rent controls. Which of the following is most likely to occur?
a decrease in rents and a decrease in the number of apartment units supplied
an increase in rents and an increase in the number of apartment units supplied
a decrease in the demand for apartments and an increase in the number of apartment units supplied
an increase in the demand for apartments and a decrease in the number of apartment units supplied

ChatGPT: The most likely outcome of eliminating rent controls in a city facing a shortage of rental apartments is an increase in rents and an increase in the number of apartment units supplied.

The elimination of rent controls would give landlords more freedom to set rents at market levels, allowing them to charge higher prices in response to the demand for housing. This, in turn, would provide incentives for landlords to construct or make available more apartment units in order to take advantage of the higher rents.

It is possible that the increase in rents may cause a decrease in demand for apartments, but in a market with a shortage of rental units, the overall demand for housing is likely to remain high. As a result, the increase in supply of apartments is likely to be greater than any decrease in demand, leading to an overall increase in the number of apartment units supplied.

To compare ChatGPT's performance with that of a typical economics student, we examined its percentile scores based on the results in Table 4 and Table 5 of the 4th Edition of the TUCE. If we consider only the pre-test scores, ChatGPT would rank in the top 1% of both microeconomics and macroeconomics exam takers. However, if we compare its scores with those of students who have completed a full semester of economics, it would still rank in the top 9% of microeconomics exam takers and continue to rank in the top 1% of macroeconomics exam takers.

IV. Practical Advice for Assessment in the ChatGPT Era

The rise of artificial intelligence in higher education, specifically natural language models like ChatGPT, presents a new challenge to educators. Unlike anti-plagiarism tools that compare a student's work with existing sources, ChatGPT can generate original content in seconds.[9] Additionally, it is almost impossible to detect plagiarism when using ChatGPT. Moreover, ChatGPT has many advantages over non-AI forms of cheating: it is free, simple to use and generates content much quicker than earlier methods.[10] The emergence of ChatGPT in November 2022 has raised fears about widespread cheating on non-proctored exams and other assignments. The short-term solution for many educators involves returning to in-person, proctored assessments. The main advantage of this approach is that violations of academic integrity can usually be reduced if the assessment is run properly. There are, however, certain drawbacks,

---

[9] https://www.chronicle.com/article/ai-and-the-future-of-undergraduate-writing

[10] On February 2nd, ChatGPT launched a paid subscription service, in which subscribers will get access to the AI at peak times for a monthly fee of $20.

7

including equity issues for students in remote or online classes when assessment is scheduled on-campus as well as the logistical challenges associated with large lectures.

A number of educators have begun to create assessments that teach students how to use ChatGPT as a resource and also use ChatGPT as part of the assessment.[11] One popular recommendation among the teaching community so far has been to produce ChatGPT responses with errors and have students work in small groups to identify and correct those errors. In essence, students are asked to "fact check" the system to ensure that the responses are accurate.

Figure 4: ChatGPT prompt and response for a hypothetical assignment in a principles of microeconomics course.
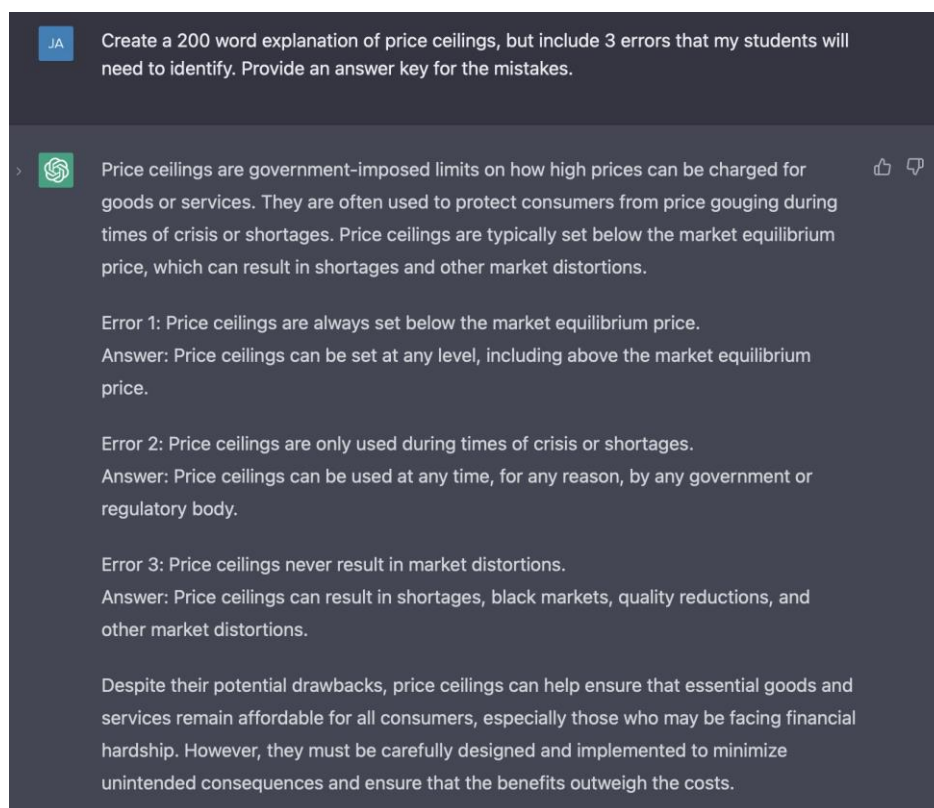
The current emphasis of "teaching with ChatGPT" has focused on humanities courses, but will likely evolve to the social sciences in due course. The current outlook among economics educators is to use ChatGPT as a source of knowledge, which is dangerous in its current stage since the program is merely predicting responses.[12] It's important to emphasize to students that just because ChatGPT provides a response that looks reasonable doesn't mean that the response is accurate.

ChatGPT presents some challenges, but they can be overcome by designing a learning environment that fosters knowledge acquisition. Artificial intelligence can enhance

---

[11] https://www.nytimes.com/2023/01/24/learning/lesson-plans/lesson-plan-teaching-and-learning-in-the-era-of-chatgpt.html

[12] https://www.pbs.org/newshour/education/some-educators-embrace-chatgpt-as-a-new-teaching-tool

8

students' learning experience and help them achieve more in less time, but there are ways to engage students in meaningful learning experiences that can't be replicated by a program like ChatGPT. Research suggests that economic education can be effectively taught through hands-on experiences like classroom demonstrations, experiments, service learning, undergraduate research, case studies, and cooperative learning (Ghent & Mateer, 2016). This type of experiential learning goes beyond simple memorization and fosters a deeper understanding of the subject. Students can be asked to write brief essays that apply economic principles to solve interesting questions they personally observe (Geerling, 2013), form student groups to synthesize music with economics (Geerling, 2019), or work on art-inspired projects that require students to apply economic concepts (Al-Bahrani et al. 2016).

Assessments that evaluate higher-level thinking skills like analysis, evaluation, and creation can help engage students in meaningful learning experiences while making it more difficult for ChatGPT to circumvent the process. The Economic Instructor's Toolkit (Picault, 2019, 2021) is a valuable resource that provides information on a growing list of class activities and student projects that foster higher-level learning. Whether teaching in-person or online, incorporating hands-on experiences into the curriculum can make a big impact on students' learning outcomes.

V. Conclusion

The purpose of this study was to evaluate the performance of ChatGPT in microeconomics and macroeconomics tests, as assessed by the TUCE. The results found that ChatGPT ranks at the 99th percentile in macroeconomics and the 91st percentile in microeconomics, when compared to students who take the exams at the end of a semester-long principles course. It is hardly surprising that ChatGPT outperforms the average college student in a standardized test of economics comprehension delivered in multiple choice format with textbook answers, but the extent of this performance gap is quite revealing. ChatGPT was trained on a vast amount of text for its predictive algorithm, which gives it a significant advantage over its human counterparts.

Our findings have significant implications for assessment strategies in the ChatGPT-era. It is crucial to rethink assessment strategies to include both traditional methods, such as proctored exams, in-class writing assignments, or experiential learning opportunities, and to find ways to utilize chatbots as a teaching aide or as part of assessments in the future. It is important to note that ChatGPT is not the only disruptive technology in education. The advent of artificial intelligence in education is a reality that cannot be ignored, and it is time to embrace the new era with innovative and effective assessment strategies.

Acknowledgements

VI. References

Al-Bahrani, A., Holder, K., Patel, D., & Wooten, J. (2016). Art of econ: Incorporating the arts through active learning assignments in principles courses. *Journal of Economics and Finance Education*, 15(2), 1-16.

Boyle, A., & Goffe, W. L. (2018). Beyond the flipped class: The impact of research-based teaching methods in a macroeconomics principles class. *AEA Papers and Proceedings*, 108, 297-301.

Choi, J., Hickman, K., Monahan, A., & Schwarcz, D. (2023). ChatGPT goes to law school. Minnesota Legal Studies Research Paper No. 23-03, Available at SSRN: https://ssrn.com/abstract=4335905

Emerson, T. L., & Taylor, B. A. (2004). Comparing student achievement across experimental and lecture-oriented sections of a principles of microeconomics course. *Southern Economic Journal*, 70(3), 672-693.

Ferrell, O.C. & Ferrell, L. (2020). Technology challenges and opportunities facing marketing education. *Marketing Education Review*, 30(1), 3-14.

Følstad, A., & Brandtzæg, P.B. (2017). Chatbots and the new world of HCI. *Interactions*, 24(4), 38–42.

Geerling, W. (2013). An exploration of Robert Frank's 'The Economic Naturalist' in the classroom. *International Review of Economics Education*, 12, 48-59.

Geerling, W., Mateer, G. and O'Roark, B. (2019). Music then and now: Using technology to build a lyric animation module. *The American Economist*, 65(2), 1-19.

Ghent, L., and Mateer, G. (2016). JET: A new journal takes off! *Journal of Economics Teaching*, 1(1), 1-6.

Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R., & Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9(1), e45312.

Goel, A. K., & Polepeddi, L. (2018). Jill Watson: A virtual teaching assistant for online education. In: *Learning engineering for online education* (pp. 120-143). Routledge.

Hill, J., Ford, W.R., & Farreras, I.G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, 245–250.

Huang, W., Hew, K.F., & Fryer, L.K. (2021). Chatbots for language learning – Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 1-325.

Hwang, G.J. and Chang, C-Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*. DOI: https://doi.org/10.1080/10494820.2021.1952615

Kim, N.Y., Cha, Y., & Kim, H.S. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning*, 22(3), 32–53.

Picault, J. (2019). The economics instructor's toolbox. *International Review of Economics Education*, 30, 100154.

Picault, J. (2021). Looking for innovative pedagogy? An online economics instructor's toolbox. *The Journal of Economic Education*, 52(2), 174.

Rebeck, K. and Walsted, W. (2008) The test of understanding of college economics. *American Economic Review*, 98(2), 547-551.

Schmulian, A., & Coetzee, S.A. (2019). The development of messenger bots for teaching and learning and accounting students' experience of the use thereof. *British Journal of Educational Technology*, 50(5), 2751–2777.

Selwyn, N. (2021). *Education and Technology: Key Issues and Debates*, third edition, New York: Bloomsbury.

Smith, B. O., & Wagner, J. (2018). Adjusting for guessing and applying a statistical test to the disaggregation of value-added learning scores. *The Journal of Economic Education*, 49(4), 307-323.

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the facebook messenger. *Computers & Education*, 151, 103862.

Walstad, W., Watts, M., and Rebeck, K. (2006). *Test of Understanding of College Economics*, 4e, Examiner's Manual, National Council for Economic Education. New York.

West, C. G. (2023). AI and the FCI: Can ChatGPT Project an Understanding of Introductory Physics?. arXiv preprint arXiv:2303.01067.

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? – A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, 1-18.

Wu, E.H., Lin, C.H., Ou, Y.Y., Liu, C.Z., Wang, W.K., & Chao, C.Y. (2020). Advantages and constraints of a hybrid model K-12 e-learning assistant chatbot. *IEEE Access*, 8, 77788–77801.

Zhou, L., Gao, J., Li, D., & Shum, H.Y. (2020). The design and implementation of Xiaolce, an empathetic social chatbot. *Computational Linguistics*, 46(1), 53-93.

VII. Appendix

Table A1: ChatGPT Performance on Microeconomics Version of TUCE

Question	Concept	Correct
1	Supply and Demand	Yes
2	Price Ceilings	Yes
3	Supply and Demand	No
4	Perfect Competition	Yes
5	Factors of Production	No
6	Externalities	Yes
7	Income Distribution	Yes
8	Opportunity Cost	Yes
9	Supply and Demand	No
10	Utility	No
11	Perfect Competition	Yes
12	Monopoly	Yes
13	Diminishing Marginal Returns	Yes
14	Profit Maximization	N/A
15	Externalities	Yes
16	Taxation	Yes

17	Monopoly	Yes
18	Elasticity	No
19	Demand	Yes
20	Profit Maximization	DNA
21	Market Structure	No
22	Duopoly	Yes
23	Economic Rent	Yes
24	Profit Maximization	Yes
25	Public Choice	Yes
26	Externalities	No
27	Public Goods	Yes
28	Comparative Advantage	No
29	Trade Barriers	Yes
30	Exchange Rates	No

Table A2: ChatGPT Performance on Macroeconomics Version of TUCE

Question	Concept	Correct
1	Components of GDP	No
2	Inflation	Yes

13

3	Aggregate Demand	Yes
4	Potential GDP	Yes
5	Money Supply	Yes
6	Tools of Monetary Policy	No
7	Tools of Monetary Policy	Yes
8	Automatic Fiscal Policy	Yes
9	Crowding Out	Yes
10	Inflation Expectations	Yes
11	Unemployment Rate	Yes
12	Real Interest Rate	Yes
13	Supply Shocks	Yes
14	Aggregate Demand	Yes
15	Aggregate Demand	Yes
16	Tools of Monetary Policy	No
17	Fiscal Policy	Yes
18	Tools of Monetary Policy	Yes
19	Real GDP	Yes
20	Multiplier Effect	Yes

21	Economic Growth	Yes
22	Money Creation	Yes
23	Tools of Fiscal Policy	Yes
24	Monetary vs. Fiscal Policy	Yes
25	Tools of Monetary Policy	Yes
26	Policy Lags and Limitations	Yes
27	Automatic Fiscal Policy	Yes
28	Exchange Rates	No
29	Open-Economy Macroeconomics	Yes
30	Trade Balance	Yes