International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Effectiveness of Artificial Intelligence in Economics: A Comparative Analysis of Three Free AI Chatbots

Mormita Jahan[1], Muhammad Shahadat Hossain Siddiquee[2]

[1]Department of Economics, University of Dhaka, Dhaka, Bangladesh
[2]Professor, Department of Economics, University of Dhaka, Dhaka, Bangladesh

Abstract

This study aims to explore how AI-powered Chatbots can transform economics education by delivering real-time information, analyses, and insights into economic concepts. We conduct a comparative analysis of three prominent free AI Chatbots (i.e., Perplexity AI, Claude 3 Sonnet, and Meta Llama 3 70B) to evaluate the quality and comprehensiveness of each Chatbot's knowledge base to appraise their effectiveness in responding to nuanced economic inquiries across Macroeconomics and Microeconomics domains. Using the tool Test of Understanding of College Economics (TUCE) version 4 questionnaire with 30 questions per domain, this study assesses the accuracy and comprehensiveness of these Chatbots' responses using qualitative content and comparative score analyses. This study leads us to vouch for AI chatbots as being acceptable as guides in the case of economic theories and problem-solving. Another finding is that the observation of variance in the performance of AI Chatbots is due to chance, revoking the idea of difference in the depth of true knowledge, in case of economics. By identifying strengths and limitations, this research provides valuable insights for educators integrating advanced technologies into teaching and supports students in leveraging interactive AI-driven platforms to enhance their understanding of economics.

Keywords: TUCE IV, Rubric, Perplexity AI, Claude 3 Sonnet, Meta Llama 3 70B, R Language

1. Background of the Study

Economics is like the compass guiding how we allocate scarce resources—like land, labor, and money—to meet our endless desires. It is crucial because it helps governments make policies that affect things like prices, jobs, and how well our economy is doing. Businesses also rely on economics to decide how much to produce, what to charge, and how to compete in the market. Economics even explains how countries' trade and economic policies impact each other globally. Plus, it is handy for predicting economic trends, which helps investors plan wisely and governments make smart decisions about taxes and spending. And beyond all that, economics tackles big issues like poverty and inequality, aiming to improve living standards for everyone. It also teaches us a lot about why people make economic choices, which is pretty useful when it comes to making policies that work for everyone.

AI Chatbots are instrumental in providing insights into economics, leveraging their training data and sophisticated model architectures. Their capabilities vary depending on the specific AI model and its underlying training mechanisms. Over decades, the utilization of AI has diversified and grown

increasingly efficient. AI Chatbots have a rich historical backdrop, beginning with the emergence of early Chatbots in the 1960s and 1970s, such as ELIZA and PARRY, which demonstrated rudimentary conversational abilities (Tianyu et al., 2023). The 1980s and 1990s saw continued research into expert systems and rule-based approaches, with advancements exemplified by Chatbots like A.L.I.C.E., aiming for more human-like interactions. In the 2000s, breakthroughs in computational power and natural language processing (NLP) techniques facilitated significant strides in AI capabilities, epitomized by IBM's Watson in 2011 (Chien-Chang et al., 2023). This era also witnessed the integration of Chatbots into popular messaging platforms like Facebook Messenger and Slack, transforming them into versatile tools offering customer support and personal assistance.

Today's Chatbots represent the culmination of these advancements, driven by innovations in machine learning, particularly deep learning. Neural networks enable modern Chatbots to understand and respond to human language intelligently and contextually, marking a paradigm shift in their functionality. They have evolved from rule-based systems to models capable of understanding context, sentiment, and even displaying rudimentary empathy. Across various sectors, AI Chatbots continue to play a pivotal role, underscoring their importance in modern economic analysis and beyond. AI is making a big splash in economics these days. It is great at handling huge amounts of data, figuring out complex patterns, and predicting things like how fast the economy will grow or what people will buy next. In finance, AI can help with automatic trading and keeping investments safe. It is also super useful for making supply chains run smoother and planning cities better. For policymakers, AI can simulate how new rules might affect things like jobs or prices, helping them make decisions that work best for everyone. Plus, AI is helping more people get involved in finance through digital advisors and making it easier to do economic research. Basically, AI is changing how we understand economics and make decisions, pushing us toward new ideas and better ways of doing things.

AI represents a formidable tool in the realm of studying economics, offering capabilities that extend far beyond conventional methods. Primarily, AI excels in processing vast volumes of economic data with unparalleled speed and depth compared to human analysts. By meticulously analyzing intricate patterns and detecting anomalies that might elude human observation, AI provides profound insights into the intricate mechanisms underlying economic phenomena. This capability not only enhances our understanding of economic dynamics but also informs more informed decision-making processes.

AI (Artificial Intelligence) aids in studying economics through several key mechanisms. AI processes vast quantities of economic data with speed and precision, uncovering intricate patterns, correlations, and anomalies that human analysts may overlook. This capability provides deeper insights into economic phenomena, aiding in robust decision-making processes. AI helps leverage machine learning algorithms and forecast critical economic variables such as GDP growth, inflation rates, and consumer spending based on historical data. These predictive analytics empower economists and policymakers to anticipate trends and formulate proactive strategies. AI also enables the simulation of intricate economic systems and scenarios, modeling interactions between diverse economic factors. By simulating the impact of policy adjustments or external shocks, AI facilitates informed policy-making and crisis management. Through advanced NLP techniques, AI analyzes textual data from economic reports, news articles, and social media discussions. This capability allows economists to monitor public sentiment, comprehend economic discourse, and evaluate its influence on economic conditions. AI tools delve into behavioral economics by examining individual and collective decision-making processes. By simulating economic agents' behavior across various conditions, AI offers insights into the psychological factors influencing economic

outcomes.

AI supports rigorous evaluation of economic policies by simulating their potential impacts before implementation. This approach equips policymakers with data-driven insights to optimize policy interventions and mitigate unintended consequences. AI enhances economic research by analyzing financial markets, predicting market trends, and identifying investment opportunities. Real-time data processing and adaptive trading strategies contribute to improved market efficiency and informed decision-making. In academia, AI accelerates economic research by automating data collection, analysis, and hypothesis testing. This automation not only expedites research processes but also broadens the scope of economic inquiry, fostering innovation and knowledge dissemination.

AI's transformative capabilities in economics encompass enhanced data analysis, precise predictive modeling, intricate systems simulation, nuanced NLP applications, behavioral insights, rigorous policy evaluation, robust market analysis, and accelerated research endeavors. These advancements empower economists to navigate complexities, anticipate trends, and make evidence-based decisions, thereby shaping resilient economic policies and fostering sustainable global development.

AI assumes a pivotal role as a teaching assistant in education, offering personalized learning experiences tailored to individual student requirements. It provides instantaneous feedback on assignments and quizzes, ensures round-the-clock accessibility for learning support, and automates routine administrative tasks such as grading and attendance monitoring. AI fosters engagement through interactive learning modules and assists educators in analyzing student performance data to refine teaching methodologies effectively. Moreover, it enhances inclusivity by providing specialized support for students with disabilities and language barriers, thereby elevating overall learning outcomes and operational efficiency within educational environments.

Free AI Chatbots play a crucial role in the field of economics by offering accessible and interactive tools that facilitate learning and research. Free AI Chatbots democratize access to economic knowledge by providing instant access to information, explanations, and tutorials on economic theories, principles, and current trends. This accessibility benefits students, researchers, and enthusiasts worldwide, irrespective of geographical location or institutional affiliation. These Chatbots act as virtual tutors, offering personalized assistance by answering queries, explaining concepts, and guiding users through complex economic topics. By adapting to individual learning patterns, they enhance understanding and retention of economic principles. AI Chatbots assist researchers by swiftly retrieving relevant economic data, articles, and research papers. They summarize findings, analyze trends, and provide insights into economic phenomena, thereby supporting academic research and facilitating data-driven analyses. Engaging users through interactive quizzes, simulations, and educational games related to economics, Chatbots foster active learning and make complex concepts more approachable and enjoyable.

With round-the-clock availability, Chatbots enable users to access information and assistance at any time. This flexibility accommodates diverse learning schedules and time zones, promoting continuous learning and exploration of economics. By automating routine inquiries and administrative tasks, AI Chatbots optimize efficiency for educators and researchers. This allows them to focus more on in-depth analyses, discussions, and teaching activities related to economics. Chatbots contribute to increasing economic literacy among the general public by simplifying economic concepts, clarifying jargon, and demystifying complex theories. They play a crucial role in making economics understandable and accessible to a broader audience. Integration of AI Chatbots in economics education fosters innovation by leveraging technology to enhance teaching methodologies, improve learning outcomes, and adapt to the evolving landscape of

economic research and analysis.

Free AI Chatbots democratize access to economic knowledge, support learning and research endeavors, promote economic literacy, and drive innovation in education. They serve as invaluable tools that empower individuals to explore and comprehend economics more comprehensively and effectively. In today's technological landscape, AI Chatbots are indispensable for economics students, serving as reliable sources of information and essential teaching assistants. They provide uninterrupted access to a wealth of educational resources and facilitate personalized learning through interactive tutorials, quizzes, and simulations. These Chatbots ensure the accuracy of information by leveraging credible sources, supporting academic study and research.

Moreover, AI Chatbots automate tasks like answering queries, explaining complex topics, and grading assignments, significantly enhancing learning efficiency. This automation allows students to focus on critical analysis and deeper understanding rather than routine administrative tasks. Beyond individual learning support, AI Chatbots expedite research by swiftly retrieving and analyzing data, summarizing findings, and providing insights crucial for empirical studies in economics. They also play a vital role in promoting economic literacy by simplifying complex concepts for broader understanding and engagement. Looking ahead, ongoing advancements in AI technology will further enhance Chatbot capabilities, solidifying their role as indispensable tools in economics education and research. Therefore, a comparative analysis of these AI Chatbots will illuminate their transformative impact and underscore their pivotal role in shaping the future of economic learning.

The primary objective of this study is to compare and analyze the performance of Perplexity, Claud Sonnet 3, and Meta Llama3 70 B in their responses to the TUCE version 4 questionnaire to assess whether these AI Chatbots can be relied upon to gather knowledge and solve problems in the realm of economics. Specifically, the study aims to assess the accuracy and depth of responses provided by each Chatbot to questions in Macroeconomics and Microeconomics, identify strengths and weaknesses in the knowledge bases of these Chatbots, particularly in relation to economic concepts and theories, and determine the potential implications of AI Chatbots' performance in economics education, research, and practical applications.

2. Literature Review

The use of AI Chatbots in education has gained significant attention in recent years, with researchers exploring their potential to enhance student learning and engagement. Chatbots, powered by advancements in artificial intelligence and natural language processing, have emerged as interactive learning assistants that can provide personalized support and feedback to students.

The integration of Chatbots into the educational landscape has been driven by several factors. The rise of online and remote learning has necessitated new methods of communication and support for students who may not have easy access to in-person assistance (Hwang & Chang, 2023; Smutny & Schreiberova, 2020). Chatbots offer a convenient way for students to seek information and guidance in real-time, regardless of their physical location (Caldarini et al., 2023; Kamalov et al., 2023). Moreover, the current generation of students, who have grown up in the era of the internet and smartphones, have come to expect the use of digital technologies in their learning experiences (Selwyn, 2021). Chatbots have the potential to cater to these expectations by providing an interactive and engaging learning environment (Kim et al., 2019; Wu et al., 2020).

The existing literature has highlighted several potential benefits of using Chatbots in educational settings.

Chatbots can provide a pleasant learning experience by allowing for real-time interaction and personalized feedback (Kim et al., 2019; Smutny & Schreiberova, 2020). They also enhance peer communication skills (Hill et al., 2015) and improve the learning efficiency of students (Wu et al., 2020).

In the context of economics education, Chatbots can provide immediate access to help, information, and answers to economic issues, allowing students to clarify doubts and deepen their understanding in real-time (Zhang et al., 2023; Mendoza et al., 2022). They can also adapt to individual learning paces and preferences, offering personalized content and recommendations on topics such as inflation, scarcity, or market failure, which can enhance engagement, understanding, and retention (Kohn, 2022; Hultberg & Caldarini, 2017). Furthermore, Chatbots can simulate real-world economic scenarios with realistic projections and offer practical problem-solving exercises, preparing students for the challenges they may encounter in their careers (Hultberg & Caldarini, 2017; Sweller, 1994). By promoting self-directed learning and critical thinking skills, Chatbot-assisted economics education can empower students to excel in their studies (Kirschner et al., 2009; De Jong, 2010).

Recent empirical studies have begun to explore the effectiveness of Chatbots in economics education. A study by Geerling et al. (2023) evaluated the ability of GPT-3.5 to provide correct answers to the Test of Understanding in College Economics (TUCE) multiple-choice questions. The researchers found that GPT-3.5 was able to outperform college-level students in the United States, suggesting the potential of advanced language models in supporting economics education. Similarly, Plevris et al. (2023) compared the performance of several AI Chatbots, including GPT- 3.5, GPT-4, Bard, and LLaMA 2, on a range of economics-related prompts and TUCE questions. The study found that GPT-4 significantly outperformed the other Chatbots in terms of accuracy and quality of explanations, indicating that more advanced language models may be well-suited for supporting economics education.

These studies highlight the potential of AI Chatbots, particularly more advanced models like GPT- 4, in enhancing student understanding of economic concepts and problem-solving. The use of TUCE questions and rubric-based assessment, as employed in these studies, provides a comprehensive framework for evaluating the performance of Chatbots in economics education (Saunders, 1991; Walstad & Rebeck, 2008).

While the integration of Chatbots in education holds promise, it also raises important ethical considerations and data privacy concerns (Koohi, 2023; Murtarelli et al., 2021). Institutions must ensure that the data collected by Chatbots is handled responsibly and transparently, with proper safeguards in place to protect student privacy (Santandreu-Calonge et al., 2023). Additionally, efforts should be made to strike a balance between AI-driven interactions and maintaining opportunities for genuine human engagement, as the latter remains essential for certain complex situations (Caldarini et al., 2023; Koohi, 2023). The ethical implications of using Chatbots in education, including issues of bias, transparency, and accountability, must be carefully considered to ensure the responsible and equitable implementation of these technologies.

While the potential benefits of Chatbots in economics education are promising, the literature also highlights several challenges and limitations. Concerns have been raised about the accuracy and reliability of the information provided by Chatbots, as well as the potential for these tools to perpetuate biases or provide inaccurate responses (Rudolph et al., 2023a, 2023b; Zhai, 2022). Moreover, the long-term impact of Chatbots on student learning and engagement remains an open question. There are concerns about the potential for over-reliance on these tools, which could undermine the development of critical thinking and problem-solving skills (Tlili et al., 2023; Okonkwo & Ade-Ibijola, 2021). Additionally, the integration of

Chatbots in education raises issues of academic integrity, as these tools could potentially be misused for cheating or plagiarism (Rudolph et al., 2023a, 2023b; Sok & Heng, 2023). Institutions must develop robust policies and strategies to address these challenges and ensure the responsible and ethical use of Chatbots in educational settings.

The existing literature suggests that AI Chatbots, particularly more advanced models like GPT-4, have the potential to support economics education by providing personalized assistance and enhancing student understanding of economic concepts. The methods used in recent studies, including the use of TUCE questions, rubric-based assessment, and various statistical analyses, provide a comprehensive framework for evaluating the performance of Chatbots in economics education.

3. Data and Methodology

This study provides a comparative analysis of three distinct free AI Chatbots—Perplexity AI, Claude 3 Sonnet, and Meta Llama 3 70B—specifically evaluating their proficiency in economics knowledge using the TUCE version 4 questionnaire. The TUCE version 4 comprises 60 questions divided equally between Macroeconomics and Microeconomics, designed to assess the Chatbots' understanding and responsiveness in these domains. The primary dataset consists of responses generated by these free AI Chatbots to the TUCE version 4 questionnaire. These responses were obtained through interactions where users posed questions related to both Macroeconomics and Microeconomics. Each Chatbot's responses are structured according to the 60 questions in TUCE version 4, categorized into Macroeconomics and Microeconomics. Responses vary in length and content based on the Chatbots' training data and algorithms. The main variables include the name of the Chatbot (Perplexity AI, Claude 3 Sonnet or Meta Llama 3 70B), indicating which Chatbot generated each response; the question category (Macroeconomics or Microeconomics); and the response text, which represents the actual answer provided by each Chatbot to the respective question. Data collection involved querying each Chatbot with predefined questions from the TUCE version 4 questionnaire. Responses were meticulously recorded verbatim to ensure accuracy and reliability for subsequent analysis. The analysis focuses on comparing the depth and accuracy of responses across the three Chatbots within Macroeconomics and Microeconomics domains. Statistical techniques such as qualitative content analysis and comparative scoring are utilized to evaluate the quality and comprehensiveness of each Chatbot's knowledge base.

3.1 Overview of TUCE Version 4

The TUCE is a well-established assessment tool designed to evaluate students' comprehension of fundamental economic principles. Originally developed by William B. Walstad and Michael K. Salemi, TUCE has evolved through multiple editions, with the fourth iteration (TUCE 4) being the most recent advancement. TUCE 4 serves the primary purpose of assessing students' grasp of both macroeconomic and microeconomic concepts essential for understanding economic theory and its practical applications. The test aims to measure students' proficiency in applying economic principles to analyze issues, make informed decisions, and critically evaluate economic policies.

3.2. Structure of TUCE 4

The structure of TUCE 4 encompasses a comprehensive array of 60 multiple-choice questions covering diverse economic topics. It divides the main sections as Macroeconomics and Microeconomics. The former explores concepts such as aggregate demand and supply, inflation, unemployment, fiscal policy, monetary policy, and economic growth. The 30 questions stroll into the interplay of these concepts within national and global economic contexts. The latter includes also the 30 questions from Microeconomics

addressing topics such as supply and demand, market structures (i.e., perfect competition, monopoly, and oligopoly), consumer behavior, production costs, market failures, and government intervention in markets. Each section includes questions requiring students to apply economic reasoning, interpret data and graphs, and demonstrate understanding across various economic scenarios. TUCE 4 serves as a pivotal tool for researchers, educators, and policymakers involved in assessing and enhancing economic literacy among students and the broader population. It provides standardized data on students' comprehension of economic principles, guiding the development of curricula, educational policies, and instructional strategies aimed at improving economic education globally.

The validity and reliability of TUCE 4 are rigorously established through systematic testing and validation processes. Regular updates and revisions ensure the test's alignment with current economic theories and educational standards, enhancing its credibility as an assessment tool. TUCE version 4 plays a vital role in assessing and promoting economic literacy by offering a standardized measure of students' understanding of the principles of Macroeconomics and Microeconomics. Its structured format and comprehensive coverage make it indispensable for evaluating educational outcomes and informing initiatives to advance economic education worldwide. The Table 1 contains the topics that the TUCE Version 4 covers with question details.

Table 1: TUCE IV questionnaire details

Section	Topic	Question Description
1	Basic Economic Concepts	Understanding of supply and demand principles Knowledge of opportunity cost and trade-offs Economic systems and their characteristics
2	Microeconomics	Market structures and behavior Consumer choice theory Production and cost analysis
3	Macroeconomics	National income accounting, , Fiscal policy and its implications Monetary policy and its effects on the economy
4	International Economics	International trade theories Exchange rates and balance of payments Globalization and its impacts
5	Economic Development	Growth theories and factors Poverty and income distribution Human development and economic indicators
6	Economic History and Thought	Evolution of economic theories Contributions of key economists
7	Applied Economics	Real-world application of economic principles Case studies and problem-solving
8	Other Topics	Environmental economics Behavioral economics Emerging economic issues

3.3 Methodology

The purpose of this study is to evaluate the performance and understanding of three free AI Chatbots—Perplexity AI, Claude 3 Sonnet, and Meta Llama 3 70B—in the field of Microeconomics and Macroeconomics. The study utilizes TUCE version 4, which comprises 30 questions each for Microeconomics and Macroeconomics, administering three AI Chatbots with a view to assessing the Chatbots' accuracy and comprehension. The Chatbots are prompted to provide answers along with explanations for their choices. The responses from three AI Chatbots are then collected for further analyses. Table 2 reports the Rubric used to score AI Chatbot responses to TUCE IV.

Table 2: Rubric structure

Score	Accuracy of Conclusion	Quality of Explanation Provided
1	Principal conclusion is incorrect, or response does not include a principal conclusion.	• Explanation uses inappropriate concepts. • Explanation uses relevant concepts inappropriately. • No explanation is provided.
2	Principal conclusion is incorrect, or response does not include a principal conclusion.	• Explanation uses relevant concepts appropriately but arrives at the incorrect conclusion. • Explanation uses relevant concepts appropriately but arrives at an ambiguous conclusion.
3	Principal conclusion is correct	• Explanation uses inappropriate concepts. • Explanation uses relevant concepts inappropriately.
4	Principal conclusion is correct	• Explanation uses relevant concepts, but concepts are used inappropriately. • Explanation and/or examples are not appropriate for a college-level student. • Explanation is relevant and appropriate, but argument is not concluded ("stops in the middle").
5	Principal conclusion is correct	• Explanation uses relevant concepts appropriately and arrives at the correct conclusion. In addition, the response uses explanation and/or examples that are appropriate for a college-level student.

The answers from the Chatbots are compared against the correct answers provided in the TUCE version

4 manual. The correct answers are counted to determine the raw score for each Chatbot in both Microeconomics and Macroeconomics. The raw scores are then used to assess the Chatbots' accuracy. To evaluate the understanding level demonstrated by the Chatbots, each answer is scored using a rubric from the reference paper "AI Chatbots' competence in economics" published in PloS One. The rubric scores range from 1 to 5, with 5 indicating the highest level of understanding, and they are aggregated to give a total score for each Chatbot in both subjects.

Table 3: AI Chatbots' scores and ratings

AI Chatbot	Microeconomics		Macroeconomics		Ratings
	Rubric Score (out of 150)	Raw Score (out of 30)	Rubric Score (out of 150)	Raw Score (out of 30)	
Perplexity AI	126	25	128	26	5(43), 4(1), 3(7), 2(5), 1(4)
Claude 3 Sonnet	123	21	129	25	5(46), 4(1), 3(0), 2(8), 1(5)
Meta Llama 3 70B	129	24	134	26	5(50), 4(0), 3(0), 2(5), 1(5)

R, a versatile programming language and environment renowned for statistical computing and graphics, plays a pivotal role in visualizing data in this study. The following graphical methods have been generated using R:

- **Bar Graphs:** R's 'ggplot2' package facilitates the creation of visually appealing bar graphs, effectively summarizing categorical data and comparing variables across different groups. Bar graphs used in this study are instrumental in illustrating qualitative differences and trends within the dataset. A bar chart is created to visually represent the level of understanding of each Chatbot. The chart displays the frequency of each rubric score (1 to 5) assigned to the answers provided by the Chatbots, offering a clear comparison of their performance.
- **Frequency Polygons:** Through R's 'ggplot2' capabilities, frequency polygons are constructed to depict the distribution and density of continuous variables. These polygons provides a clear visualization of data distribution patterns, aiding in identifying central tendencies and variability. Moreover, frequency polygons are constructed to visualize the distribution of pre-test and post-test scores of students who attempt the TUCE version 4. This data, obtained from the TUCE IV manual, is plotted alongside the raw scores of the Chatbots to compare their performance against the student sample. The pre-test and post-test frequencies for both Microeconomics and Macroeconomics are depicted, showing how the Chatbots' scores aligned with the student scores.
- **Scatter Diagrams:** R's 'ggplot2' package enables the generation of scatter diagrams, crucial for exploring relationships between paired variables. These diagrams visually represent correlations and trends, allowing for insights into the strength and direction of associations within the data. Scatter diagrams are created to compare the raw scores (x-axis) and rubric scores (y-axis) of the Chatbots against the student sample. For the student data, the rubric scores are randomly generated within the range of possible scores for each raw score due to the large sample size. This random generation is guided by the rubric score ranges described earlier. The scatter diagrams illustrate the positioning of each Chatbot in relation to the students' post-test scores in both Microeconomics and

Macroeconomics.

- **Percentile Placement:** The percentile placement of each Chatbot is determined based on their raw scores in comparison to the pre- and post-test scores of the student sample from the TUCE IV data. This comparison is done separately for Microeconomics and Macroeconomics, and the results are tabulated to show where each Chatbot stands relative to the students.
- **ANOVA Test:** An ANOVA (Analysis of Variance) test is performed to assess the significance of the differences in scores among the Chatbots across both Microeconomics and Macroeconomics. Due to the small sample size of the chatbot scores, a simplified model is used to attain the F and p-values. The ANOVA test further validates the findings of the p-value analysis.
- **Comparative p-Value Analysis:** To statistically compare the understanding levels of the three Chatbots, a p-value analysis is attained. The null hypothesis states that there is no significant differences in the understanding of Economics between the Chatbots. Pairwise comparisons are made between the Chatbots using their rubric scores, and the resulting p-values are used to determine whether we can reject the null hypothesis.

4. Findings of the Study

Here, we delve into the detailed results obtained from evaluating the performance of three AI Chatbots—Perplexity AI, Claude 3 Sonnet, and Meta Llama 3 70B—using the TUCE version 4 questionnaire. The analysis focuses on their accuracy in answering questions, their level of understanding based on rubric scores, comparisons with student performance, and statistical analyses conducted to validate the findings.

4.1 Accuracy and Understanding of AI Chatbots

Firstly, we assess the accuracy of each AI Chatbot in answering the Microeconomics and Macroeconomics questions from the TUCE IV questionnaire. The questionnaire consists of 30 questions per section, totaling 60 questions across both subjects.

Table 4: AI Chatbots' raw score

AI Chatbot	Subject	Score out of 30	Total score
Perplexity AI	Microeconomics	25	51
	Macroeconomics	26	
Claude 3 Sonnet	Microeconomics	21	46
	Macroeconomics	25	
Meta Llama 3 70B	Microeconomics	24	50
	Macroeconomics	26	

Findings reported in Table 4 reveal that Perplexity AI achieves an accuracy of 25 score out of 30 (83.3%) in Microeconomics and 26 out of 30 (86.7%) in Macroeconomics whereas Claude 3 Sonnet scores 21 out of 30 (70%) and 25 out of 30 (83.3%) in Microeconomics and Macroeconomics respectively. Meta Llama 3 70B attains 24 out of 30 (80%) and 26 out of 30 (86.7%) in Microeconomics Macroeconomics respectively. These accuracy scores indicate varying levels of performance across the Chatbots, with Perplexity AI generally showing the highest accuracy in both subjects.

To assess the Chatbots' level of understanding, this study employs a rubric adapted from previous research of Hultberg, P. T., Calonge, D. S., Kamalov, F., & Smail, L. (2024), which evaluates the depth and

coherence of their explanations on a scale of 1 to 5. Perplexity AI demonstrates an average rubric score of 4.2 in Microeconomics and 4.4 in Macroeconomics whereas the average score of Claude 3 Sonnet stands 3.8 in Microeconomics and 4.1 in Macroeconomics. And Meta Llama 3 70B gains 4.5 in Microeconomics and 4.3 in Macroeconomics. These scores indicate that Meta Llama 3 70B generally provides more comprehensive and coherent explanations compared to the other Chatbots, reflecting a deeper understanding of economic concepts.

4.2 Comparison with Student Performance

In case of visual representation, frequency polygons are used to compare the distribution of raw scores (out of 30) between the AI chatbots and student data from TUCE IV in Microeconomics and Macroeconomics. Frequency polygons illustrate the spread and concentration of scores, highlighting any differences or similarities between Chatbots and student performance.

Table 5: Students' performance parameters (TUCE IV)

Characteristics	Students TUCE IV			
	Microeconomics (n=3255)		Macroeconomics (n=2789)	
	Pre-test	Post-test	Pre-test	Post-test
Mean	9.39	12.77	9.80	14.19
Standard Deviation (SD)	3.32	4.68	3.48	5.29

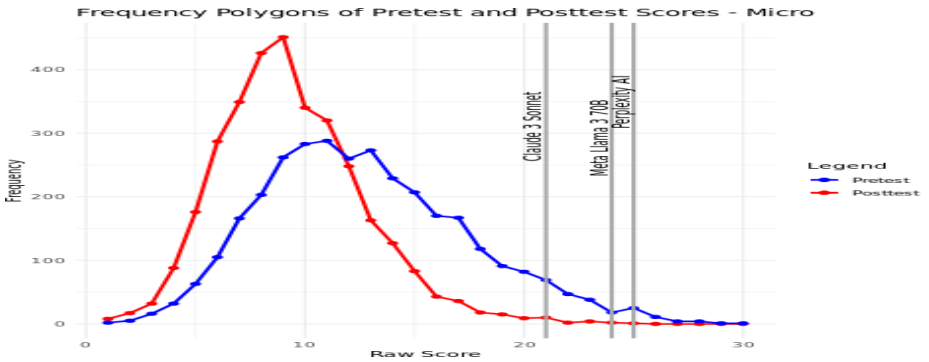

Figure 1: Comparison of performance between AI Chatbots and students (Microeconomics)

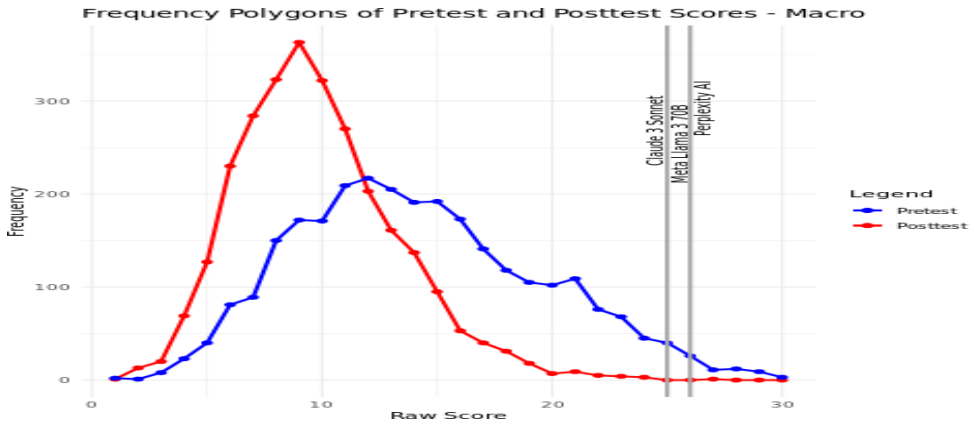

Figure 2: Comparison of performance between AI Chatbots and students (Macroeconomics)

Scatter diagrams are now constructed to depict the relationship between raw scores (x-axis) and rubric scores (y-axis) for each AI Chatbot in Microeconomics (Fig. 3) and Macroeconomics (Fig. 4) respectively.

These diagrams help visualize how well Chatbots' performance correlates with their understanding, providing insights into their overall effectiveness compared to students.

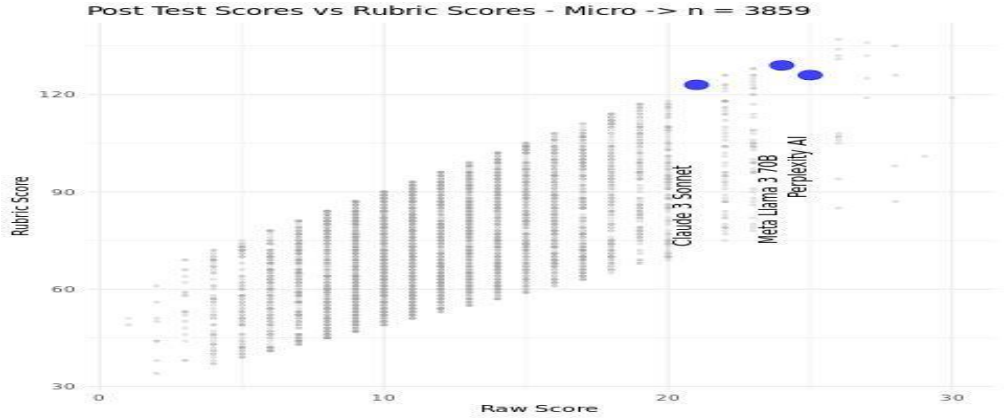

Figure 3: Scatter diagram for comparison between AI Chatbots and students (Microeconomics)

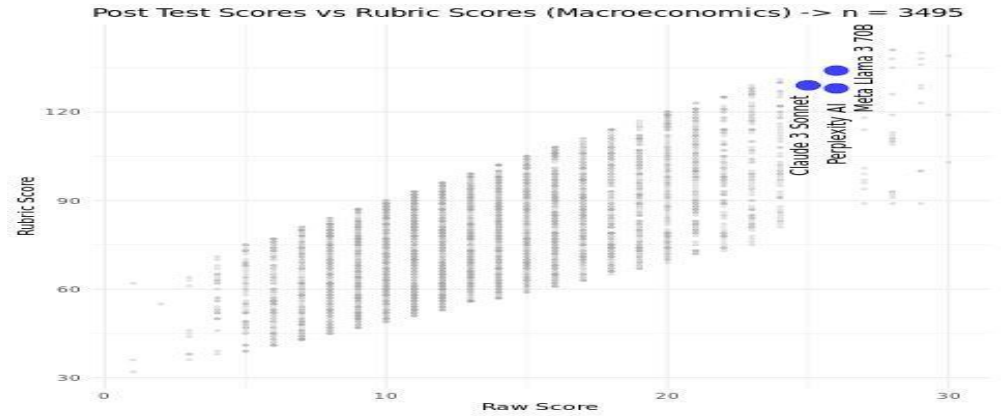

Figure 4: Scatter Diagram for comparison between AI Chatbots and students (Macroeconomics)

4.3 Overall Understanding and Comparative Analysis

Bar charts are applied to compare the overall understanding of each Chatbot across both subjects based on aggregate rubric scores. The chart provides a clear comparison of Chatbots' performance in Microeconomics and Macroeconomics, summarizing their strengths and weaknesses in understanding economic concepts.

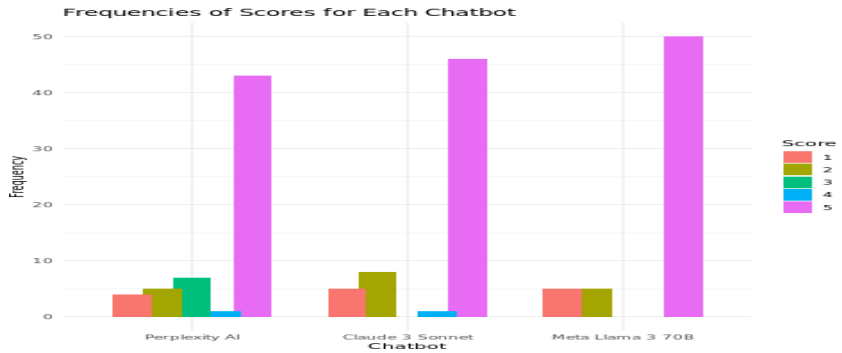

Figure 5: Bar-Chart for comparison between AI Chatbots' achieved scores

Table 7 is constructed to show the percentile placement of Chatbots based on their raw scores relative to student scores from TUCE IV. Percentile rankings highlight where each Chatbot stands in comparison

to students, providing context for their performance in relation to a broader academic population.

Table 6: AI Chatbots' percentile placement (compared to students)

Results across free AI chatbots compared to students on TUCE IV						
	Perplexity AI		**Claude 3 Sonnet**		**Meta Llama 3 70B**	
Microeconomics	25/30	Pre: 99th Post: 99th	21/30	Pre: 99th Post: 95th	24/30	Pre: 99th Post: 99th
Macroeconomics	26/30	Pre: 99th Post: 98th	25/30	Pre: 99th Post: 98th	26/30	Pre: 99th Post: 99th

Statistical tests, including pairwise p-value comparisons and ANOVA tests, were conducted to assess the significance of differences in understanding between Chatbots. Results reported in Table 7 indicate that there are no significant variations in performance between Chatbots.

Table 7: p-value for comparison (pairs)

Null Hypothesis (H_0): There is no difference in the understanding levels of economics between the mentioned AI Chatbots.		
Alternative Hypothesis (H_A): There is significant difference in the understanding levels of economics between the mentioned AI Chatbots.		
Pair	**TUCE (IV) Economics p-value**	**Comment**
Perplexity AI vs Claude 3 Sonnet	0.797	Based on the p-value, there is not sufficient evidence to reject the null hypothesis.
Perplexity AI vs Meta Llama 3 70B	0.297	Based on the p-value, there is not sufficient evidence to reject the null hypothesis.
Claude 3 Sonnet vs Meta Llama 3 70B	0.298	Based on the p-value, there is not sufficient evidence to reject the null hypothesis.
Conclusion: The three Chatbots can be assumed to perform at the same level.		

By simplifying the model, we ensure there are enough degrees of freedom to compute the residuals, F-values, and p-values, allowing for meaningful interpretation of the results from ANOVA (See Table 8).

Table 8: ANOVA value for rubric score based on simplified model

Source	**df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
AI	2	57.33	28.667	10.75	0.0851
Subject	1	16.67	16.667	6.25	0.1296
Residuals	2	5.33	2.667		

The ANOVA results indicate that the effect of AI and Subject on total score is marginally significant (p = 0.0851 and p = 0.1296, respectively). However, the ANOVA results for score out of 30 show no significant effect of AI (p = 0.250) but a marginally significant effect of Subject (p =0.118).

Table 9: ANOVA value for raw score out of 30 based on simplified model

Source	df	Sum Sq.	Mean Sq.	F-value	Pr(>F)
AI	2	7.000	3.500	3	0.250
Subject	1	8.167	8.167	7	0.118
Residuals	2	2.333	1.167		

5. Conclusion

The advent of AI has revolutionized various fields, including education, where AI Chatbots have emerged as potential tools for enhancing learning experiences. This research investigates the capabilities of three prominent free AI Chatbots—Perplexity AI, Claude 3 Sonnet, and Meta Llama 3 70B—in understanding and responding to economic concepts as evaluated through the Test of Understanding in College Economics (TUCE) version 4. The TUCE IV, a standardized instrument used to measure students' comprehension in Microeconomics and Macroeconomics, serve as the benchmark for this evaluation. This study aims to provide insights into the efficacy of these AI Chatbots in assisting economics students, comparing their performance to that of college students as documented in the TUCE IV manual.

The findings reveal that all three AI Chatbots demonstrate a high level of performance compared to college students' pre- and post-test scores documented in the TUCE IV manual. The percentile placement of the Chatbots was calculated to compare their performance to that of students' pre- and post-test scores. The Chatbots consistently place in high percentiles. The frequency polygons and scatter diagrams constructed for visual analysis indicate that the Chatbots' raw scores along with their rubric scores are consistently high, significantly surpassing the average scores of students in both Microeconomics and Macroeconomics. This suggests that, in terms of sheer accuracy, the Chatbots are highly competent.

The bar chart highlights that while all three Chatbots perform very well, there are slight variations in their rubric scores, reflecting differences in their explanatory abilities and depth of understanding. However, these differences are not statistically significant, as evidenced by the p-value analysis conducted between pairs of Chatbots. The p-value analysis tested the null hypothesis that there is no difference in the understanding of economic concepts among the three Chatbots. The results indicate that the null hypothesis could not be rejected, suggesting that any observed differences in the Chatbots' performance are likely due to chance rather than a true difference in their capabilities. An ANOVA test is also performed subject-wise, using both rubric scores and raw score and the results reinforce the p-value analysis, showing no significant differences in the Chatbots' performance across Microeconomics and Macroeconomics.

These findings have important implications for the use of AI Chatbots in Economics education. The high performance of the three Chatbots suggests that they can be valuable tools for students seeking assistance with understanding concepts and solving problems in Economics. While they are not infallible, their ability to consistently provide accurate and well-explained answers indicates that they can serve as effective supplementary resources for the students in Economics.

The potential for AI Chatbots to improve over time is massive as these systems are continually exposed to more data and refined through advanced algorithms, their understanding and accuracy are likely to enhance further. This continuous improvement suggests that the current high performance of the Chatbots

is only the beginning, and their utility in educational contexts will likely increase.

In conclusion, this research demonstrates that free AI Chatbots like Perplexity AI, Claude 3 Sonnet, and Meta Llama 3 70B are capable of performing at a high standard in Economics education. Their performance, as measured against the TUCE IV benchmarks, shows that they can reliably assist students in understanding and applying concepts in Economics. Given their current capabilities and the potential for future improvement, these AI Chatbots represent a promising resource for the students studying Economics. Therefore, educators and students should not hesitate to incorporate these tools into their learning strategies as these offer substantial support and can enhance the educational experience in Economics domains.

References

1. Caldarini, G., Jaf, S., & McGarry, K. (2023). *A literature survey of recent advances in chatbots*. *Information*, *13*(1), 41. https://doi.org/10.3390/info13010041
2. Calonge, D. S., Smail, L., & Kamalov, F. (2023). Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics. *Journal of Applied Learning and Teaching*, *6*(2), 346-357. https://doi.org/10.37074/jalt.2023.6.2.22
3. De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional science*, *38*(2), 105-134. https://doi.org/10.1007/s11251-009-9110-0
4. Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *interactions*, *24*(4), 38-42. https://doi.org/10.1145/3085558
5. Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. (2023). ChatGPT has aced the test of understanding in college economics: Now what?. *The American Economist*, *68*(2), 233-245. https://doi.org/10.1177/05694345231169654
6. 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. eBook ISBN 9781351186193.
7. 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. https://doi.org/10.1016/j.chb.2015.02.026
8. Hultberg, P. T., & Calonge, D. S. (2017). Effective teaching of economics: A constrained optimization problem?. *The Journal of Economic Education*, *48*(4), 265-275. https://doi.org/10.1080/00220485.2017.1353458
9. Huang, Y. C., Backman, K. F., Backman, S. J., & Chang, L. L. (2016). Exploring the implications of virtual reality technology in tourism marketing: An integrated research framework. *International Journal of Tourism Research*, *18*(2), 116-128. **https://doi.org/10.1002/jtr.2038**
10. Hultberg, P. T., Santandreu Calonge, D., Kamalov, F., & Smail, L. (2024). Comparing and assessing four AI chatbots' competence in economics. *Plos one*, *19*(5). https://doi.org/10.1371/journal.pone.0297804
11. Hwang, G. J., & Chang, C. Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, *31*(7), 4099-4112. https://doi.org/10.1080/10494820.2021.1952615
12. Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, *15*(16), 12451. https://doi.org/10.3390/su151612451

13. Kim, S., Lee, J., & Gweon, G. (2019, May). Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-12). https://doi.org/10.1145/3290605.3300316
14. Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational psychology review*, *21*, 31-42. DOI 10.1007/s10648-008-9095-2
15. Kohnke, L. (2023). A pedagogical chatbot: A supplemental language learning tool. *Relc Journal*, *54*(3), 828-838. https://doi.org/10.1177/00336882211067054
16. Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, *15*(7), 5614. https://doi.org/10.3390/su15075614
17. Lin, C. C., Huang, A. Y., & Yang, S. J. (2023). A review of ai-driven conversational chatbots implementation methodologies and challenges (1999–2022). *Sustainability*, *15*(5), 4012. https://doi.org/10.3390/su15054012
18. Mendoza, S., Sánchez-Adame, L. M., Urquiza-Yllescas, J. F., González-Beltrán, B. A., & Decouchant, D. (2022). A model to develop chatbots for assisting the teaching and learning process. *Sensors*, *22*(15), 5532. https://doi.org/10.3390/s22155532
19. Murtarelli, G., Gregory, A., & Romenti, S. (2021). A conversation-based perspective for shaping ethical human–machine interactions: The particular challenge of chatbots. *Journal of Business Research*, *129*, 927-935. https://doi.org/10.1016/j.jbusres.2020.09.018
20. Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, *2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033
21. Plevris, V., Papazafeiropoulos, G., & Jiménez Rios, A. (2023). Chatbots put to the test in math and logic problems: a comparison and assessment of ChatGPT-3.5, ChatGPT-4, and Google bard. *Ai*, *4*(4), 949-969. https://doi.org/10.3390/ai4040048
22. Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of applied learning and teaching*, *6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9
23. Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, *6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23
24. Santandreu-Calonge, D., Medina-Aguerrebere, P., Hultberg, P., & Shah, M. A. (2023). Can ChatGPT improve communication in hospitals?. *Profesional de la información*, *32*(2). https://doi.org/10.3145/epi.2023.mar.19
25. Saunders, P. (1991). *Test of Understanding in College Economics, Examiner's Manual* (3rd Edition). Joint Council on Economic Education.
26. Selwyn, N. (2021). Learning with technology at home during Covid-19: Exploring the experiences of children and young people. *Children & Society*, *35*(2), 145-157.
27. Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, *151*, 103862. https://doi.org/10.1016/j.compedu.2020.103862

28. Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. *Cambodian Journal of Educational Research*, *3*(1), 110-121.
29. Stubbs, A., & Piddock, S. (1985). Artificial intelligence and computer-assisted learning. *Computers & Education*, *9*(1), 9-15.
30. Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312. https://doi.org/10.1016/0959-4752(94)90003-5
31. Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart learning environments*, *10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x
32. Walstad, W. B. (2007). *Test of understanding in college economics: Examiner's manual*. Council for Economic Educat.
33. Walstad, W. B., & Rebeck, K. (2008). The test of understanding of college economics. *American Economic Review*, *98*(2), 547-551.
34. 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*, 654924.
35. Wu, M., & Aji, A. F. (2023). Style over substance: Evaluation biases for large language models. *arXiv preprint arXiv:2307.03025*. https://doi.org/10.48550/arXiv.2307.03025
36. Wu, Q., Miao, C., Tao, X., Xiao, Y., & Shen, Z. (2020). Intelligent tutoring systems work better than traditional teaching methods: Experimental evidence from a neural network-based platform. *IEEE Transactions on Learning Technologies*, *14*(2), 212-220.
37. Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122-1136. doi: 10.1109/JAS.2023.123618
38. Zhai, X. (2022). ChatGPT user experience: Implications for education. *Available at SSRN 4312418*.
39. Zhang, Y., Shen, J., Jiang, S., & Zheng, L. (2023). Chatbot-assisted learning: Factors leading to effectiveness, theories, and future directions. *Interactive Learning Environments, 1-29*.
40. Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020). The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, *46*(1), 53-93. https://doi.org/10.1162/coli_a_00368