AGORA International Journal of Economical Sciences, http://univagora.ro/jour/index.php/aijes
ISSN 2067-3310, E-ISSN 2067-7669
Vol. 20, No. 1 (2026), pp. 259-268

AI CHATBOTS DRASTICALLY OUTPERFORM SECONDARY STUDENTS ON THE TEST OF ECONOMIC LITERACY

A. KONC

Andraž Konc
School of Advanced Social Studies, Slovenia
https://orcid.org/0009-0000-5064-8153, E-mail: andrazkonc@gmail.com

Abstract: *This paper examines the performance of generative artificial intelligence models (OpenAI ChatGPT, Google Gemini, Microsoft Copilot) on the Test of Economic Literacy and compares their results with those of U.S. and Slovenian secondary school students. The study aims to assess whether AI systems demonstrate comparable or even superior levels of economic knowledge and to explore potential language-related differences in test performance. Descriptive statistics and item-level analysis are used to identify patterns and inconsistencies across languages when comparing AI test-takers and secondary students. Results show that AI models achieved near-perfect accuracy, while students performed significantly lower on average. The analysis also highlights differences in the length and explanatory depth of responses among the AI models. Findings contribute to current debates on assessment validity, AI use in economic education, and the future of multiple-choice testing in digitally augmented learning environments. The paper underscores the need for responsible integration of AI into educational frameworks, safeguarding the credibility of learning and the support for the sustainable development of economic knowledge.*
Keywords: *Artificial Intelligence, AI Chatbots, Economic Literacy, Assessment, Economics Education.*

1 Introduction

In the evolving educational landscape, the integration of technology has become essential for supporting effective learning, particularly in economics, where complex concepts often challenge students, and traditional teaching methods may not fully address individual learning needs or sustain active engagement. Artificial intelligence (AI) chatbots have rapidly become a prominent subject of empirical investigation across disciplines, including economics, education, and management. Recent studies demonstrate that models such as ChatGPT can successfully complete standardized academic and professional assessments, raising important questions about the nature of knowledge measurement and the future role of traditional testing instruments (Geerling et al., 2023; Hultberg et al., 2024; Brose, Spielmann & Tode, 2025). While much of the existing literature focuses on higher education or professional certification exams, comparatively little attention has been paid to secondary-level assessments and to cross-linguistic dimensions of AI economics test performance.

Multiple studies (Jappelli, 2010; McCowage & Dwyer, 2022) emphasize that economic literacy extends beyond the mastery of financial concepts and also encompasses an understanding of macroeconomic phenomena such as inflation, economic growth, unemployment, and public finance. Mügge (2016) argues that citizens' comprehension of macroeconomic concepts is essential, as it enables them to critically evaluate political claims about economic performance. Such knowledge is fundamental to the functioning of a democracy, since "economically literate" citizens are better equipped to participate in public debates and to form well-reasoned political preferences (Evans, 2015). As a result, standardized instruments designed to measure economic knowledge have long played an important role in

economics education research and educational evaluation. Among these instruments, the Test of Economic Literacy (TEL) developed by the National Council on Economic Education provides a psychometrically sound instrument aligned with widely accepted content standards and has been extensively used to assess students' understanding of core economic concepts at the secondary education level (Walstad, Rebeck & Butters, 2013). Recent evidence that AI chatbots can perform at or above student level on economics tests challenges the interpretive meaning of such assessments. If AI systems can reproduce high scores on standardized instruments, it creates substantial pressure on long-standing assessment practices. One key consequence is that instructors may need to rethink their curriculum in at least three possible directions: bringing back supervised, in-person exams, integrating chatbot-supported learning activities, or expanding experiential, real-world projects that AI cannot easily imitate (Geerling et al., 2023). At the same time, students can, and arguably should, select and use appropriate chatbots to deepen their understanding of economic concepts and reasoning. AI-empowered economics education can not only enhance teaching quality but also drive innovations in this field. Wan et al. (2025) emphasize that reforms in economics education are necessary because traditional teaching approaches no longer adequately reflect the growing role of data, algorithms, and AI in modern economics.

Existing research has largely focused on English-language assessments at the tertiary education level. Cross-linguistic analyses remain rare – Wood et al. (2023) found that ChatGPT scores much higher when the exam is given in English than when it is given in a different language. The present study contributes to this emerging literature by evaluating the performance of commonly used OpenAI ChatGPT, Google Gemini and Microsoft Copilot on the TEL4 and by comparing AI-generated results with empirical data from U.S. and Slovenian secondary school students. In doing so, the paper provides empirical evidence relevant to ongoing debates on assessment validity and AI integration in education systems. By administering both the original English version and a methodologically adapted Slovenian version of the TEL4, the study also examines whether language context influences AI performance. Given that language plays a central role in both human cognition and AI model training, examining AI performance across different language versions of the same test represents an important identified research gap. Also, evidence from secondary education contexts is still limited.

1.1 Literature Review

Economic literacy is commonly defined as the ability to understand fundamental economic concepts and apply them to everyday decision-making situations. Inquiries in economics education have consistently emphasized its importance for individual welfare and societal functioning. At the secondary education level, economic literacy is shaped by curricular exposure, instructional practices, and institutional context, making reliable measurement particularly important. Standardized tests have played a key role in evaluating economic knowledge across different populations. The fourth edition (TEL4) of the Test of Economic Literacy reflects updated economic concepts and has been used in numerous national and international studies to compare student performance across school types and educational systems (Walstad, Rebeck & Butters, 2013). Prior research has demonstrated that TEL-based instruments show acceptable reliability and construct validity when applied to secondary school populations (Walstad & Watts, 2015). However, these validation efforts largely predate the widespread availability of advanced AI systems capable of generating contextually appropriate responses to multiple-choice questions. The emergence of AI chatbots therefore introduces a new dimension to the use and interpretation of standardized test results. A growing body of literature has examined the performance of AI chatbots on academic assessments.

Geerling et al. (2023) report that ChatGPT achieved high scores on the Test of Understanding in College Economics (TUCE), outperforming the average university student. Similar findings have been reported for examinations in medicine, law, and business, suggesting that the latest AI updates possess extensive domain knowledge encoded through large-scale training data and capability to solve multiple-choice tests (Newton & Xiromeriti, 2024). Importantly, these studies treat AI systems as test-taking agents rather than instructional tools. The primary focus is on score comparability, response patterns, and item-level performance rather than pedagogical effectiveness. This approach however provides a useful framework for evaluating the implications of AI for assessment validity, particularly when standardized instruments are used for high-stakes decisions.

OpenAI ChatGPT, Google Gemini, and Microsoft Copilot are large language model–based AI systems designed to generate human-like text and support knowledge-intensive tasks. ChatGPT, developed by OpenAI, is a generative transformer-based model trained on a mixture of licensed data, human-created content, and publicly available text, enabling it to perform complex reasoning, explanation, and problem-solving tasks across a wide range of domains. Gemini is a multimodal large language model developed by Google, designed to integrate text-based reasoning with advanced contextual understanding and to support analytical tasks through large-scale data. Microsoft Copilot is an AI-powered assistant embedded within Microsoft's ecosystem, leveraging large language models to assist users in information retrieval, reasoning, and productivity-related tasks. Although these systems differ in architecture, deployment environments, and optimization objectives, all three rely on probabilistic language generation and have demonstrated the capacity to perform domain-specific cognitive tasks, making them suitable subjects for comparative assessment using standardized instruments such as the Test of Economic Literacy (TEL4).

Traditional lecture-based approaches in economics education are often characterized by limited interactivity and poor student engagement, whereas the integration of AI through educational reforms can address these shortcomings by providing a more interactive, practical, and dynamically responsive learning experience (Davar, Dewan & Zhang, 2025; Wan et al., 2025). ChatGPT and other AI chatbots have the potential to transform education by providing instant, accessible support for learning and research, shifting the focus from memorization to logical thinking and conceptual understanding, thereby enhancing cognitive, creative, and problem-solving skills (Pittner, Romih & Polančič, 2023). However, overreliance on this technology could lead to a decline in these very skills, making balanced integration essential for developing competent young people. Empirical evidence indicates that students generally perceive AI systems as useful and motivating, particularly as personalized assistants that provide immediate feedback and support for academic tasks; however, concerns persist regarding the accuracy and reliability of AI-generated outputs and their potential effects on critical thinking, learning discipline, and deeper conceptual understanding (Schei, Møgelvang & Ludvigsen, 2024). While AI chatbots offer opportunities for personalized and adaptive learning, they cannot fully replicate essential dimensions of personal teaching, such as mentorship, emotional support, and pedagogical judgment, underscoring the continued centrality of educators in the learning process (Labadze, Grigolia & Machaidze, 2023). Digital platforms, AI-enhanced personalized learning improve accessibility and engagement, and it is important to develop a flexible curriculum aligned with ongoing transformations to strengthen students' adaptability to evolving economic environments (Nezhyva, Mysiuk & Zaremba, 2022; Konc, 2025). In the context of economics education, this transformation aligns with broader shifts toward data-intensive and computational approaches, which require students to combine solid foundational knowledge with innovative and analytical capacities (Hao & Zhang, 2025). Given that economics graduates are likely to enter occupations with above-

average exposure to AI, integrating AI literacy and complementary skill development into curricula is increasingly viewed as essential for preparing students for an AI-augmented labour market (Oschinski, Spielmann & Subbu-Rathinam, 2025). In economics education, AI can help enhance values such as analytical thinking and decision-making skills by offering simulations, predictive models, and scenario-based learning. Using AI to generate and enhance case studies offers a powerful instructional tool in economics education by significantly reducing the time required to develop high-quality, practice-oriented learning materials, while enabling instructors to tailor cases to specific learning objectives and student groups (Byrd, 2025). AI-supported case studies facilitate active, student-centred learning by allowing students to apply economic concepts to realistic scenarios, engage in fruitful discussion, and explore alternative outcomes through simulations.

Nevertheless, addressing the challenges that ChatGPT poses to academic integrity requires a reconsideration of current assessment practices, particularly in relation to multiple-choice question (MCQ) based testing. MCQs are widely used in online examinations, a context in which academic misconduct was already a significant concern prior to the emergence of generative artificial intelligence. One possible response is to move away from strictly closed-book examinations and to permit the use of AI tools such as ChatGPT, thereby aligning assessment more closely with real-world professional environments in which graduates are increasingly expected to use such technologies. But this approach does not diminish the importance of foundational knowledge, which remains essential for critically formulating questions, evaluating outputs, and interpreting AI-generated responses. The assessment of basic and core economic knowledge must be conducted under conditions without direct use of AI tools. This does not necessarily require traditional in-person written examinations but may also include supervised practical tasks, oral examinations, presentations, or other real-time assessment formats that ensure academic integrity (Newton & Xiromeriti, 2024).

2 Methodology

2.1 Experimental setup

The study employs a comparative quantitative research design. Performance on the TEL4 by three artificial intelligence chatbots (OpenAI ChatGPT, Google Gemini and Microsoft Copilot) is compared with U.S. data from Walstad, Rebeck and Butters (2013) and empirical test data obtained from the sample of Slovenian secondary school students (data collected and analysed by the author of this paper, not yet published). Primary analysis is conducted at the level of total test scores, individual test items and question-answer content.

Free versions of OpenAI ChatGPT, Google Gemini and Microsoft Copilot were selected due to their widespread use and accessibility. Each TEL4 item was presented individually to the AI models, including the question stem and response options. No additional contextual cues or explanations were provided. Prompts were standardized across chatbots and languages to minimize variation in input conditions. To ensure consistency, each model completed the full TEL4 twice getting the same results (on January 8th and 9th 2026), and responses were recorded precisely. Scoring was conducted using the same answer key applied to student responses, allowing for direct comparability. The author conducted a new chat session on each of the three AI chatbots included in this analysis. Provided was one question from A version of TEL4 at a time, along with its answer choices (example Table 1 below). While it was generating responses during the chat session, the author did not provide any feedback or additional information to the AI chatbot.

Although identical test items and scoring were applied, the comparability between AI chatbots and secondary students is limited and should be interpreted cautiously. AI models do not engage in cognitive processes such as comprehension, reasoning, or dealing with

misconceptions. Their responses are generated probabilistically from training data and may reflect exposure to similar economic content rather than a genuine understanding of economics. Furthermore, AI is not affected by the same distractors that can influence students' test performance.

Table 1. *ChatGPT, Copilot, Gemini interface illustrating the question–answer methodology applied in the study*

Question 1
The opportunity cost of a new public high school is the A. money cost of hiring teachers for the new school. B. cost of constructing the new school at a later date. C. change in the annual tax rate to pay for the new school. D. other goods and services that must be given up for the new school.
ChatGPT response
Correct answer: D. The opportunity cost of a new public high school is the other goods and services that must be given up to build and operate it, since resources are limited and could have been used elsewhere.
Copilot response
The best answer is D. other goods and services that must be given up for the new school. Why? Opportunity cost is the value of the next best alternative that must be sacrificed when a choice is made. Building a new public high school uses resources—labor, land, materials, and money—that could have been used for something else. Those forgone alternatives are the true opportunity cost. Options A, B, and C describe financial or budgetary aspects, but they do not capture the full economic concept of opportunity cost. If you want, I can walk you through how to spot opportunity cost questions quickly.
Gemini response
The correct answer is D. other goods and services that must be given up for the new school. Why this is the correct answer In economics, opportunity cost is defined as the value of the next best alternative that is foregone when a choice is made. It isn't just about the dollar amount (the "accounting cost"), but rather what else society could have done with those same resources. To break down why the other options don't fit the strict definition of opportunity cost: - A and C: These represent monetary costs or financial outlays. While they quantify the expense, they don't describe the actual "thing" you are missing out on. - B: This is simply a different timeline for the same project, not a sacrificed alternative. - D: This captures the essence of the "trade-off." If a city spends $50 million on a high school, that is $50 million that cannot be spent on new roads, public parks, or healthcare services. Those "lost" goods and services are the true opportunity cost.

2.2. Sample

This study utilizes results derived from two prior research samples. The first student sample consists of 3.682 U.S. secondary school students from the study of Walstad, Rebeck and Butters (2013). The second sample comprises 113 students attending Slovenian general and economic secondary education schools. The same dataset is employed in a manuscript currently under review in a Scopus-indexed journal (Konc, under review). For simplicity, we analysed only results from version A of the TEL4, as the test developers found only minor differences in measurement characteristics between version A and version B. The Slovenian version of the test was developed methodologically through a translation, adaptation and validation process designed to preserve conceptual equivalence with the original English version, consistent with established guidelines for cross-linguistic assessment instruments. We deploy two samples to conduct a comparative analysis of the English and Slovenian versions and to examine potential language-related differences in AI performance.

3 Results

OpenAI ChatGPT, Google Gemini and Microsoft Copilot achieved near perfect overall accuracy on the TEL4 (Table 2). However, question 14 seems to cause a problem for Gemini and Copilot – this question is about the exchange rate between two currencies. After we got the last response, we asked each chatbot the following question: Do you know where these questions come from (which test are they from)? Only Gemini knew the right answer, even stating the correct version of the Test of economic literacy (i.e. 4th).

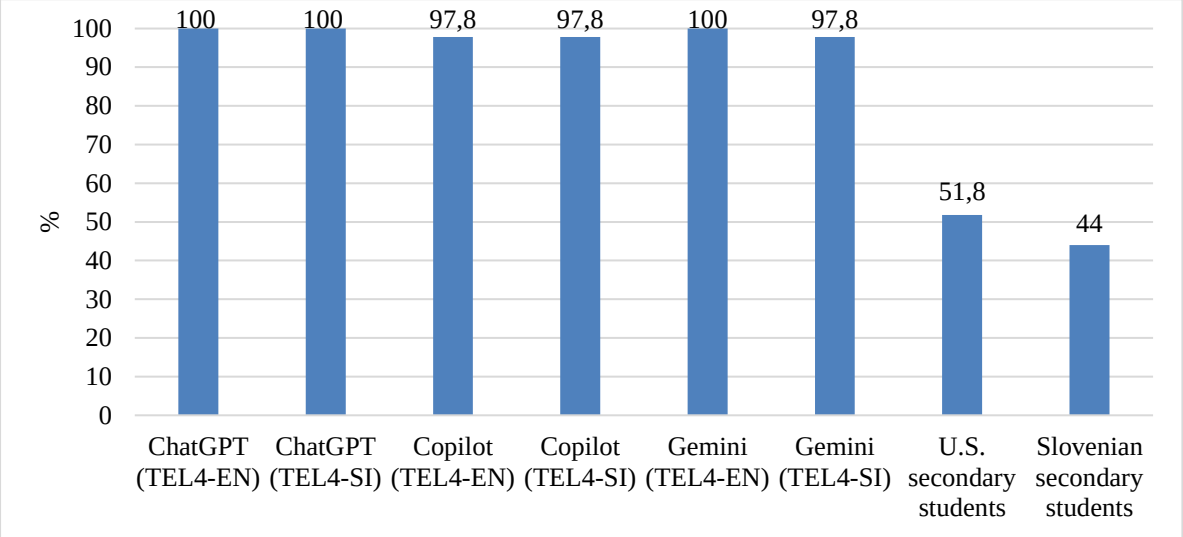
Table 2. *Which AI chatbot is more economically literate?*

	ChatGPT	Copilot	Gemini
Question 1	D	D	D
Question 2	B	B	B
Question 3	C	C	C
Question 4	D	D	D
Question 5	B	B	B
Question 6	C	C	C
Question 7	A	A	A
Question 8	C	C	C
Question 9	A	A	A
Question 10	C	C	C
Question 11	B	B	B
Question 12	A	A	A
Question 13	A	A	A
Question 14	B	A	B*
Question 15	A	A	A
Question 16	B	B	B
Question 17	A	A	A
Question 18	B	B	B
Question 19	D	D	D
Question 20	A	A	A
Question 21	D	D	D
Question 22	C	C	C
Question 23	D	D	D
Question 24	D	D	D
Question 25	C	C	C
Question 26	B	B	B
Question 27	A	A	A
Question 28	B	B	B
Question 29	D	D	D
Question 30	C	C	C
Question 31	A	A	A
Question 32	B	B	B
Question 33	C	C	C
Question 34	B	B	B
Question 35	C	C	C
Question 36	D	D	D
Question 37	A	A	A
Question 38	C	C	C
Question 39	D	D	D
Question 40	C	C	C
Question 41	C	C	C
Question 42	B	B	B
Question 43	D	D	D
Question 44	D	D	D
Question 45	A	A	A

**Answer incorrect only in Slovene version of the test. Note: For the full text of all 45 questions see the work of Walstad, Rebeck and Butters (2013).*

Differences in scores between the English and Slovenian versions were not identified, with AI models basically perform equally on the English version of the test. There were however noticeable differences in the length and interpretation of the answers. Secondary students from the U.S. score on average just above 50% and Slovene even less (44%), indicating substantially weaker performance compared to AI chatbots. This is also presented in Graph 1 below. These findings suggest that current secondary-level economics curriculum in those two countries may be insufficiently aligned with the competencies assessed by standardized test of economic literacy.

Figure 1. *AI chatbots vs U.S. and Slovene secondary students*



We also applied the free-marginal Fleiss' kappa (Fleiss & Cohen, 1973) to assess inter-rater agreement in responses from the three AI chatbots included in this study. The coefficient ranges from −1.00 to +1.00, where negative values indicate agreement worse than chance, 0 represents chance-level agreement, and +1.00 denotes perfect agreement. The analysis yielded a moderate level of inter-chatbot agreement (free-marginal Fleiss' $\kappa = 0.492$, $Z = 5.72$, $p < .001$), despite a very high observed agreement rate. The moderate kappa value reflects the prevalence effect, whereby near-uniform response patterns inflate expected chance agreement. The accuracy of the AI chatbot's response is most important, but another notable difference between the AI models is the variation in how answers are phrased (see Figure 1 again) in the word count of their responses and the level of verbosity displayed. It should be added that Gemini added a graphic explanation in addition to the text in some of the answers. To explore these apparent differences, average wordcounts were calculated. Although the sample size was limited to the 45 questions of TEL4, the responses in English to the given prompts in these cases the response showed that ChatGPT was the least talkative (36.0 average words per question), followed by Copilot (111.9), and then Gemini with times more (331.3). In the Slovene version, Copilot replaced ChatGPT as the briefest provider (59.8 words). ChatGPT followed with 113.5 words, while Gemini once again delivered the most prolific responses, averaging 203.4 words. Whereas ChatGPT and Copilot mainly gave their answer with only a short explanation, Gemini's explanations were more exhaustive and also told why other given answer options to the question are not correct. The observed low similarity in chatbot responses, despite identical questions and prompting conditions, can be explained by the stochastic nature of artificial intelligence models, which can yield different outputs for identical inputs, as well as by structural differences in their underlying neural network architectures (Kamalov et al., 2025).

4 Discussion/Conclusion

Recent advances in generative artificial intelligence have established AI chatbots as versatile educational tools capable of supporting learning through explanation of complex concepts, instant feedback, and assistance with various academic tasks. At the same time, these systems are increasingly used by educators for content generation, assessment design, and analytical support. However, the expanding presence of AI in learning and teaching also raises important concerns related to ethical use, data protection, and the preservation of academic integrity. The ability of AI systems to perform well on standardized tests challenges existing education systems surrounding current assessment practices and the credibility of certifications. In contrast to traditional plagiarism-detection systems that rely on comparing student submissions with existing sources, ChatGPT and other AI models are capable of producing syntactically original text within seconds. Although AI-generated assignments may achieve adequate formal evaluations, they often do not demonstrate the depth of conceptual understanding typically expected in higher education and in competitive workplaces. Moreover, identifying AI-assisted plagiarism remains methodologically challenging, as generated content does not directly correspond to identifiable source materials. As generative AI tools increasingly become embedded in academic and educational environments, similarly to the preceding widespread adoption of calculators and computers in mathematics and science, schools face the challenge of adapting curriculum and assessment strategies to this evolving educational landscape.

This study provides empirical evidence on the performance of three mainstream AI chatbots on a standardized measure of economic literacy at the secondary education level. By comparing AI and human performance across English and Slovenian versions of the TEL4, the paper contributes to ongoing discussions on AI, assessment validity, and educational governance. The findings underscore the need for continued research and policy attention as AI systems become increasingly capable participants in knowledge assessment environments. Results of this study align with previous findings indicating that AI models can achieve high scores on standardized economics assessments (Geerling et al., 2023), and even achieve a perfect score – this could be due to advances in AI systems' technology and their ability to answer these kinds of questions. While ChatGPT was the only one answering all the questions of the test correctly, Google Gemini's responses were the most detailed and extensive, sometimes including explanatory graphs or videos. Extending this evidence to the secondary education level, the present study demonstrates that AI chatbots possess substantial economic knowledge as measured by the TEL4. From an assessment perspective, these findings raise questions about the continued validity of standardized multiple-choice instruments as exclusive measures of students' competence. If AI systems can reliably get high(er) scores, the interpretive meaning of test results may be altered, particularly in contexts where assessments are used for certification or selection. Therefore, educational systems will need to adapt by integrating innovative assessment methods and using AI as a tool to enhance learning, rather than viewing it solely as a painful challenge. Our study is limited by its focus on a single test instrument and a specific two-national context. Future research could extend or explore alternative assessment formats less susceptible to AI optimization, and also include more, perhaps less-known AI chatbots.

REFERENCES

1. Brose, N., Spielmann, C., & Tode, C. (2025). *ChatGPT as Economics Tutor: Capabilities and Limitations*. (Working Paper No. 25/786). School of Economics, University of Bristol, UK.

2. Byrd, W. (2025). Using AI to Generate and Refine Case Studies for Economics Education. *Journal of Economics Teaching*, *10*(4), 264-284.
3. Davar, N. F., Dewan, M. A. A., & Zhang, X. (2025). AI chatbots in education: challenges and opportunities. *Information*, *16*(3). https://doi.org/10.3390/info16030235
4. Evans, B. A. (2015). Did Economic Literacy Influence Macroeconomic Policy Preferences of the General Public during the Financial Crisis? *The American Economist*, *60*(2), 132-141. https://doi.org/10.1177/056943451506000204
5. Fleiss, J. L., & Cohen, J. (1973). The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement*, *33*(3).
6. 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
7. Hao, L., & Zhang, S. (2025). Enhancing Practical Teaching of Applied Economics Through Web-Based Computational Algorithms. *International Journal of Web-Based Learning and Teaching Technologies*, *20*(1), 1-21. https://doi.org/10.4018/IJWLTT.396269
8. 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
9. Jappelli, T. (2010). Economic Literacy: An International Comparison. *The Economic Journal*, *120*(548), 429–451. https://doi.org/10.1111/j.1468-0297.2010.02397.x
10. Kamalov, F., Calonge, D. S., Hultberg, P. T., Smail, L. & Jamali, D. (2025). Comparative analysis of leading artificial intelligence chatbots in the context of entrepreneurship. *Journal of Innovation and Entrepreneurship*, *14*(1). https://doi.org/10.1186/s13731-025-00527-3
11. Konc, A. (2025). What is the impact of disruptive innovative transformation? In economic education. *International Journal of Scholars in Education*, *8*(1), 40-48. https://izlik.org/JA52HJ64HP
12. Konc, A. (n. d.). *Priredba in validacija testa ekonomske pismenosti za uporabo v slovenskem srednješolskem izobraževalnem kontekstu*. [Manuscript under review].
13. Labadze, L., Grigolia, M. & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International journal of Educational Technology in Higher education*, *20*(1). https://doi.org/10.1186/s41239-023-00426-1
14. McCowage, M., & Dwyer, J. (2022). Economic Literacy: What Is It and Why Is It Important? *Reserve Bank of Australia Bulletin*, 1-22. https://www.rba.gov.au/publications/bulletin/2022/dec/pdf/economic-literacy-what-is-it-and-why-is-it-important.pdf
15. Mügge, D. (2016). Studying macroeconomic indicators as powerful ideas. *Journal of European Public Policy*, *23*(3), 410-427. https://doi.org/10.1080/13501763.2015.1115537
16. Newton, P., & Xiromeriti, M. (2024). ChatGPT performance on multiple choice question examinations in higher education. A pragmatic scoping review. *Assessment & Evaluation in Higher Education*, *49*(6), 781-798. https://doi.org/10.1080/02602938.2023.2299059
17. Nezhyva, M., Mysiuk, V., & Zaremba, O. (2022). Digital transformation in economies and consumers. In *Proceedings of the 4th Eastern European Conference of Management and Economics (EECME 2022)*, 45-57.
18. Oschinski, M., Spielmann, C., & Subbu-Rathinam, S. (2025). *AI and the future of work for economists: rethinking economics education*. (Working Paper No. 25/788). School of Economics, University of Bristol, UK.

19. Pittner, D., Romih, D., & Polančič, G. (2023). The use of ChatGPT in the study of economics: a SWOT analysis. In *International Scientific Conference on Economy, Management and Information Technologies*, *1*(1), 157-165. https://icemit.vpsblace.edu.rs/index.php/icemit/article/view/24
20. Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and Use of AI Chatbots among Students in Higher Education: A Scoping Review of Empirical Studies. *Education Sciences*, *14*(8). https://doi.org/10.3390/educsci14080922
21. Walstad, W. B., & Watts, M. (2015). Perspectives on Economics in the School Curriculum: Coursework, Content, and Research. *The Journal of Economic Education*, *46*(3), 324–339. https://doi.org/10.1080/00220485.2015.1040185
22. Walstad, W. B., Rebeck, K., & Butters, R. B. (2013). *Test of Economic Literacy* (4th ed.). Council for Economic Education.
23. Wan, Y., Wan, X., Huang, L., & Zhao, Y. (2025). Artificial Intelligence Empowered Reforms in Economics Education. *International Journal of Computer Science and Information Technology*, *5*(1), 1-15. https://doi.org/10.62051/ijcsit.v5n1.01
24. Wood, D. A., Achhpilia, M. P., Adams, M. T., Aghazadeh, S., Akinyele, K., Akpan, M., ... & Kuruppu, C. (2023). The ChatGPT artificial intelligence chatbot: How well does it answer accounting assessment questions? *Issues in Accounting Education*, *38*(4), 81-108.