Email:

conference@houghtonstreetreview.com

Website:

www.houghtonstreetreview.com/index.php/conference

© 2026 International Online Conference on Multidisciplinary Academic Research. All Rights Reserved.

Editor-in-Chief:

Dr. Ewelum Johnson Nnadi

Managing Editors:

1.Harold Jan R. Terano

2.Dr. Mandakini Singla

3.Massoud Kaykhaii







GENERATIVE ARTIFICIAL INTELLIGENCE AS AN INTERLOCUTOR IN PHILOSOPHY EDUCATION: OPPORTUNITIES AND RISKS FOR CRITICAL THINKING

Biybalaev Islombek Abdunazarovich
Independent researcher, Bukhara International University
E-mail: islombekbiybalayev@gmail.com
ORCID: https://orcid.org/0009-0008-6594-3567
https://doi.org/10.5281/zenodo.23569683

Abstract: The thesis considers generative artificial intelligence (ChatGPT and other large language models) as a possible interlocutor in philosophy education and analyses its opportunities and risks for students' critical thinking. Drawing on I. Kant's call "Sapere aude" and current international guidance, the thesis shows that the main danger lies not only in academic dishonesty but in cognitive offloading – handing over to the machine the very act of judgement. The author proposes a "verify – challenge – reconstruct" protocol in which the model's answer becomes an object of critical analysis rather than a ready-made result.

Keywords: generative artificial intelligence, large language models, ChatGPT, critical thinking, philosophy education, cognitive offloading, Sapere aude

Annotatsiya: Tezisda generativ sun'iy intellekt (ChatGPT va boshqa katta til modellari) falsafa ta'limida suhbatdosh sifatida ko'rib chiqilib, uning talabalar tanqidiy fikrlashi uchun imkoniyat va xavflari tahlil qilinadi. I. Kantning "Sapere aude" chaqirig'i va zamonaviy xalqaro tavsiyalarga tayanib, asosiy xavf faqat akademik halollik buzilishida emas, balki kognitiv yukni mashinaga o'tkazish, ya'ni mustaqil hukm chiqarishning o'zini unga topshirib qo'yishda ekanligi ko'rsatiladi. Muallif tomonidan model javobini tayyor natija emas, balki tanqidiy tahlil obyektiga aylantiradigan "tekshirish – e'tiroz bildirish – qayta qurish" protokoli taklif etiladi.

Kalit so'zlar: generativ sun'iy intellekt, katta til modellari, ChatGPT, tanqidiy fikrlash, falsafa ta'limi, kognitiv yukni o'tkazish, Sapere aude

Аннотация: В тезисах генеративный искусственный интеллект (ChatGPT и другие большие языковые модели) рассматривается как возможный собеседник в философском образовании, анализируются его возможности и риски для критического мышления студентов. Опираясь на призыв И. Канта «Sapere aude» и современные международные рекомендации, показано, что главная опасность заключается не только в нарушении академической честности, но и в когнитивной разгрузке – передаче машине самого акта суждения. Автором предложен протокол «проверка – возражение – реконструкция», в котором ответ модели становится объектом критического анализа, а не готовым результатом.

Email:

conference@houghtonstreetreview.com

Website:

www.houghtonstreetreview.com/index.php/conference

© 2026 International Online Conference on Multidisciplinary Academic Research. All Rights Reserved.

Editor-in-Chief:

Dr. Ewelum Johnson Nnadi

Managing Editors:

1.Harold Jan R. Terano

2.Dr. Mandakini Singla

3.Massoud Kaykhaii







Ключевые слова: генеративный искусственный интеллект, большие языковые модели, ChatGPT, критическое мышление, философское образование, когнитивная разгрузка, Sapere aude

Introduction

In 1784 Immanuel Kant defined enlightenment as humanity's emergence from self-incurred immaturity, the inability to use one's own understanding without the guidance of another, and gave it the motto "Sapere aude!" – "Have courage to use your own understanding!" [1]. Kant also noted how comfortable immaturity is: if I have a book that understands for me, I need not think myself. Today a student of philosophy has something more convenient than a book – a chatbot that can write an essay on the problem of free will in a few seconds. Does generative artificial intelligence become a new "guardian" who thinks instead of the student, or can it become a partner that provokes thinking? For philosophy, which in Uzbek higher education is taught to students of all fields as a general subject, this question is directly connected with its main educational purpose.

Main part

E. Kasneci and colleagues, reviewing the opportunities and challenges of large language models for education, point out that such models can support personalised learning, help with writing and explanation, and generate practice materials, but they also produce plausible yet incorrect information, may reproduce biases from training data, and can lead learners to rely on them excessively [2]. The authors emphasise that the effective use of these models requires competences from both teachers and students, including an understanding of their limitations. UNESCO's Guidance for Generative AI in Education and Research calls for a human-centred approach, recommends an age limit for independent use of such tools in the classroom, and stresses that generative AI should be used in ways that strengthen, not replace, human agency and critical thinking [3].

For philosophy education, the most serious risk can be described as cognitive offloading: the student transfers to the model not only routine work, such as searching for definitions, but the act of judgement itself – deciding which argument is stronger, which objection is decisive, what position to take. A fluent and confident text produced by a model is particularly tempting in philosophy, where the answer often looks like a balanced review of "different points of view". In D.F. Halpern's terms, critical thinking is purposeful, reasoned and goal-directed thinking that relies on skills and a disposition to use them [4]; if the disposition to make one's own effort is lost, skills stop developing. P. Freire's critique of "banking" education, in which knowledge is deposited into passive students [5], is also relevant: a chatbot used as an answer machine reproduces the banking model in a new technological form.

At the same time, a language model has real didactic potential as an interlocutor. It can play the role of an opponent who defends a thesis the student rejects; it can produce several versions of an argument for comparison; it can be asked to formulate the strongest objection to the student's own position. Precisely because the model can be wrong, its answers can become material for checking. S. Wineburg and S. McGrew showed that professional fact-checkers evaluate digital information by lateral reading – leaving the page and checking what other sources say about it – rather than by careful reading of the page alone [6]. This strategy is transferable to work with AI: the

Email:

conference@houghtonstreetreview.com

Website:

www.houghtonstreetreview.com/index.php/conference

© 2026 International Online Conference on Multidisciplinary Academic Research. All Rights Reserved.

Editor-in-Chief:

Dr. Ewelum Johnson Nnadi

Managing Editors:

1.Harold Jan R. Terano

2.Dr. Mandakini Singla

3.Massoud Kaykhaii







model's statement about, for example, Ibn Sina's view of the soul or Beruniy's method of inquiry should be checked against textbooks and primary sources rather than accepted on trust.

On this basis, the author proposes the "verify – challenge – reconstruct" protocol for seminar work with a language model. At the verify stage, students receive a model-generated answer to a philosophical question (for example, "Is human consciousness reducible to the brain?") and must check its factual and conceptual claims: are the philosophers and positions named correctly, are terms used accurately, are any sources invented? Errors and doubtful statements are marked in a shared document using Google Docs comments or a Hypothesis annotation layer. At the challenge stage, students identify the hidden assumptions of the answer, following S. Brookfield's emphasis on uncovering and checking assumptions [7], and then question the model itself, demanding justification, counterexamples and the strongest opposing view; the goal is to find where the model's reasoning becomes superficial or circular. At the reconstruct stage, the model is set aside, and each student writes his or her own short argument on the question, stating explicitly what was accepted from the model's answer, what was rejected and why. Only this final text is assessed as the student's work, while the documented checking and questioning serve as evidence of the thinking process.

The protocol shifts the focus of assessment from the product, which a model can easily imitate, to the process of judgement, which it cannot perform on the student's behalf. It also supports the development of digital competence among teachers and students as described in the DigCompEdu framework [8]. Its limitations should be taken into account: verification requires access to reliable sources and time; models change quickly, so specific examples of errors become outdated; students may use AI at the reconstruct stage as well, which makes oral defence or in-class writing advisable. Institutional rules on the acceptable use of AI also need to be clearly communicated. The effectiveness of the protocol will be examined in experimental work [the author's experimental data will be added].

Conclusion

1. Generative AI in philosophy education is ambivalent: it can serve as a provocative interlocutor or as a "guardian" that thinks for the student. 2. The main risk for critical thinking is cognitive offloading of judgement, not only plagiarism. 3. Kant's "Sapere aude" can serve as a guiding principle: AI should be used to strengthen, not replace, the courage to use one's own understanding. 4. The author's "verify – challenge – reconstruct" protocol turns the model's answer into an object of critical analysis; it is recommended to include it in seminars on epistemology, philosophy of mind and ethics, and to develop clear course rules on AI use together with assessment of the reasoning process.

References

1. Kant I. An Answer to the Question: What is Enlightenment? (1784) // Practical Philosophy / trans. and ed. M.J. Gregor. – Cambridge: Cambridge University Press, 1996.
2. Kasneci E., Seßler K., Küchemann S. et al. ChatGPT for good? On opportunities and challenges of large language models for education // Learning and Individual Differences. – 2023. – Vol. 103.

Email:

conference@houghtonstreetreview.com

Website:

www.houghtonstreetreview.com/index.php/conference

© 2026 International Online Conference on Multidisciplinary Academic Research. All Rights Reserved.

Editor-in-Chief:

Dr. Ewelum Johnson Nnadi

Managing Editors:

1.Harold Jan R. Terano

2.Dr. Mandakini Singla

3.Massoud Kaykhaii







3. UNESCO. Guidance for Generative AI in Education and Research. – Paris: UNESCO, 2023.
4. Halpern D.F. Thought and Knowledge: An Introduction to Critical Thinking. – 5th ed. – New York: Psychology Press, 2014.
5. Freire P. Pedagogy of the Oppressed. – New York: Herder and Herder, 1970.
6. Wineburg S., McGrew S. Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information // Teachers College Record. – 2019. – Vol. 121, No. 11.
7. Brookfield S.D. Teaching for Critical Thinking: Tools and Techniques to Help Students Question Their Assumptions. – San Francisco: Jossey-Bass, 2012.
8. Redecker C. European Framework for the Digital Competence of Educators: DigCompEdu / ed. Y. Punie. – Luxembourg: Publications Office of the European Union, 2017.