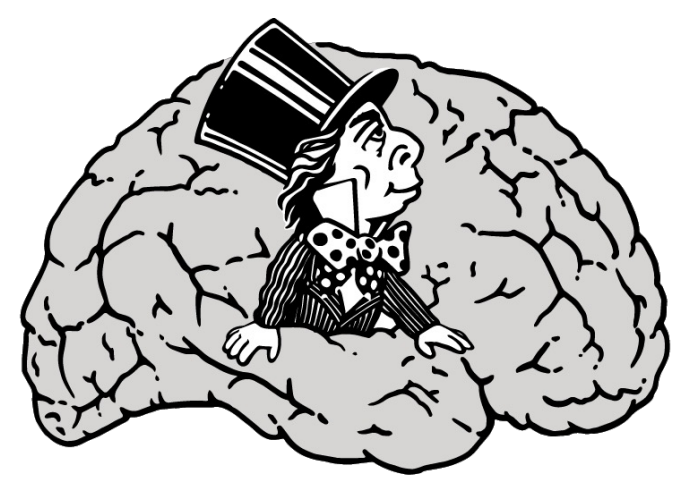
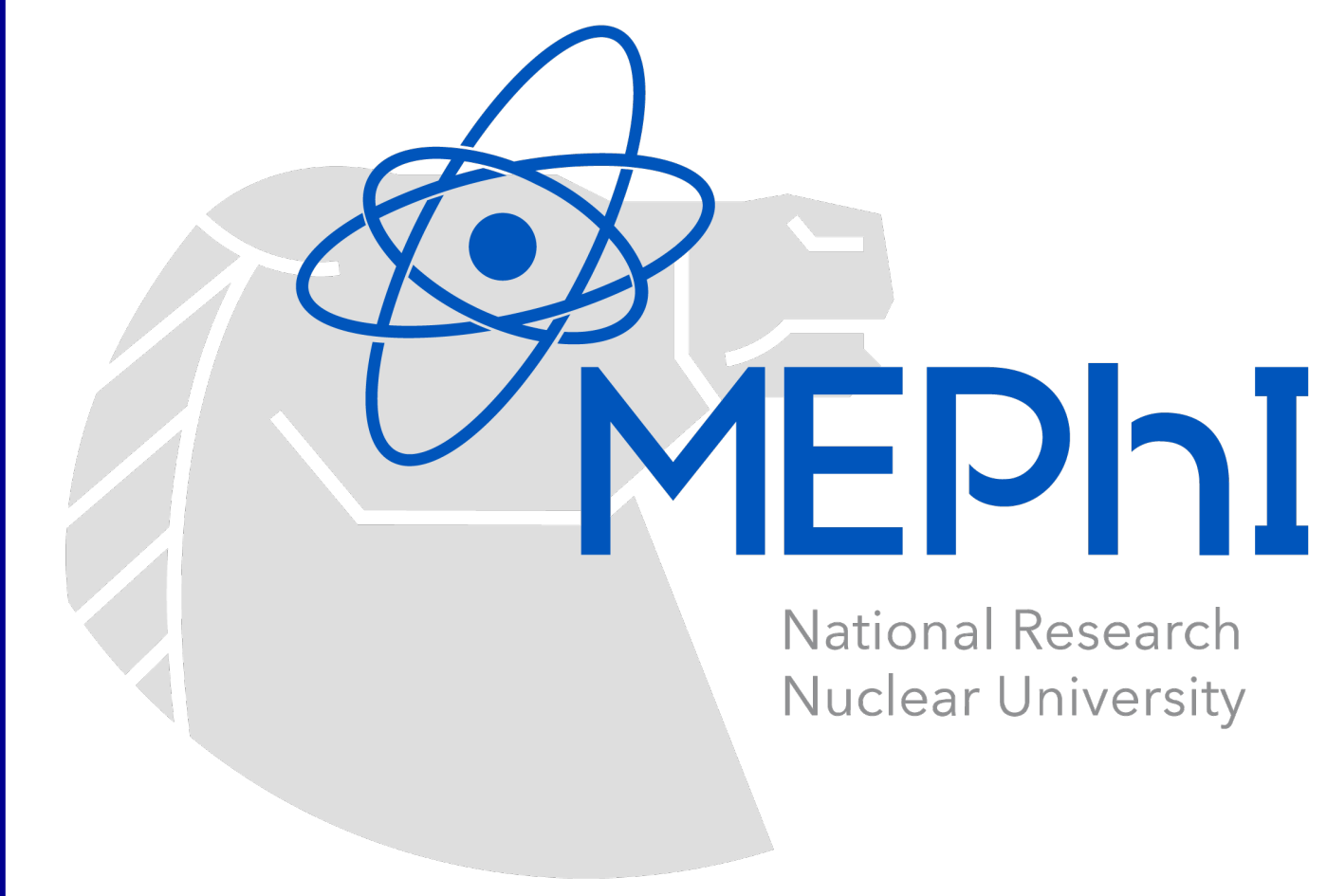




A Cognitive-Architecture-Based Virtual Tutor Supporting Self-Regulated Learning

A.A. Dolgikh, D.L. Khabarov, A.V. Samsonovich: National Research Nuclear University MEPhI



INTRODUCTION

Recently, integrating cognitive architectures with Large Language Models (LLMs) has gained attention in developing intelligent tutoring systems (ITS) (Sarker, 2022; Niu et al., 2024; García-Méndez et al., 2024). Simultaneously, advancements in LLMs have improved conversational agents, enhancing tutor adaptability and interactivity (Dong, 2023). Integration of these technologies resulted in tutors that facilitate deep learning through a meaningful dialogue (Steinert et al., 2024). However, developing cognitively plausible virtual tutors with strong language understanding remains challenging, requiring interdisciplinary approaches from Cognitive Science, AI, and ITS research.

This work introduces a Virtual Tutor designed to enhance students' essay writing skills. The system integrates the eBICA cognitive architecture (Samsonovich, 2013, 2018, 2020) with advanced LLMs, specifically GPT-4o, and is grounded in the principles of Self-Regulated Learning (SRL: Zimmerman & Kitsantas, 2005; Panadero, 2017). By combining these technologies, the system fosters the development of adaptive, context-sensitive learning strategies that promote learner autonomy and critical thinking. Results demonstrate a lasting positive impact of the Virtual Tutor on essay quality.

It is expected that the developed Virtual Tutor will be equally effective in other educational domains, highlighting the substantial potential of combining SRL, cognitive modeling, and advanced LLMs in educational technology.

Keywords: ITS, SRL, HCI, LLM, multimodal interface, intelligent virtual agents, ethical AI

REFERENCES

- Dong, C. How to build an AI tutor that can adapt to any course and provide accurate answers using large language model and retrieval-augmented generation // arXiv, 2023. No.2311.17696.
- García-Méndez, S., de Arriba-Pérez, F., Somoza-López, M. D. C. (2024). A review on the use of large language models as virtual tutors // Science & Education, 2024. Vol. 2024. P. 1-16.
- Panadero, E. A review of self-regulated learning: Six models and four directions for research // Frontiers in Psychology, 2017. Vol. 8, P. 422.
- Niu, Q., Liu, J., Bi, Z., et al. Large language models and cognitive science: A comprehensive re-view of similarities, differences, and challenges. arXiv, 2024. No.2409.02387.
- Samsonovich, A.V. Socially emotional brain-inspired cognitive architecture framework for artificial intelligence // Cognitive Systems Research, 2020. Vol. 60, P. 57-76.
- Samsonovich, A.V. Schema formalism for the common model of cognition // Biologically Inspired Cognitive Architectures, 2018. Vol. 26, P. 1-19.
- Samsonovich, A.V. Emotional biologically inspired cognitive architecture // Biologically Inspired Cognitive Architectures, 2013. Vol. 6, P. 109-125.
- Sarker, I. H. AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. SN Computer Science, 3, 1–20, 2022.
- Steinert, S., Avila, K. E., Ruzika, S., Kuhn, J., Küchemann, S. Harnessing large language models to develop research-based learning assistants for formative feedback // Smart Learning Environments, 2024. Vol. 11. No. 1. P. 62.
- Zimmerman, B. J., Kitsantas, A. Homework practices and academic achievement: The mediating role of self-efficacy and perceived responsibility beliefs // Contemporary Educational Psychology, 2005. Vol. 30/ No. 4. P. 397-417. doi: 10.1016/j.cedpsych.2005.05.003

METHODS & DESIGN

Virtual Tutor consists of the following parts: eBICA component, Visualization component (based on UE5), Interaction Module (RESTful API-based), and Generation module utilizing OpenAI services (Fig. 1).

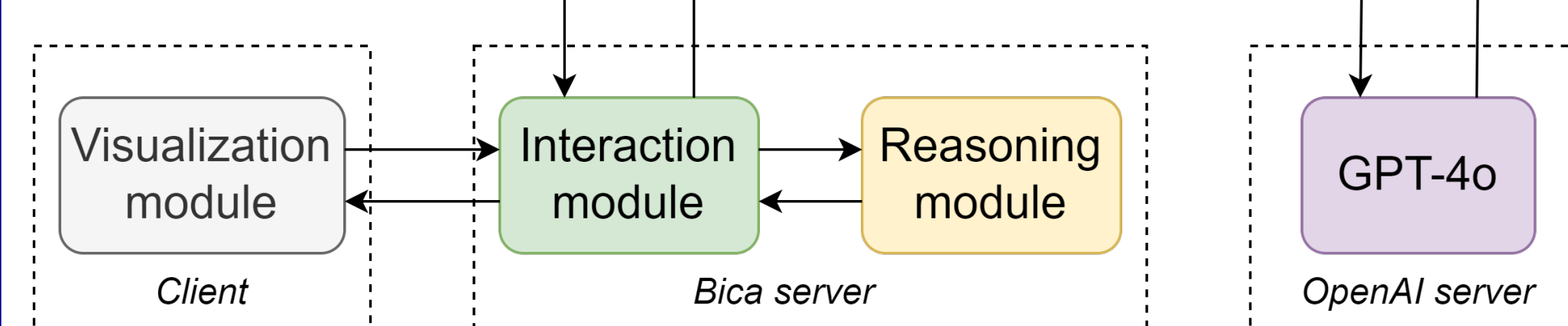


Fig. 1. Overview of the Virtual Tutor architecture.

The eBICA cognitive architecture (Fig. 1: BICA Server) makes decisions and controls other components. The module implements SRL using moral schemas of eBICA (Fig. 2). The module runs on a Debian 12 laptop (3.3 GHz, 2 GB RAM).

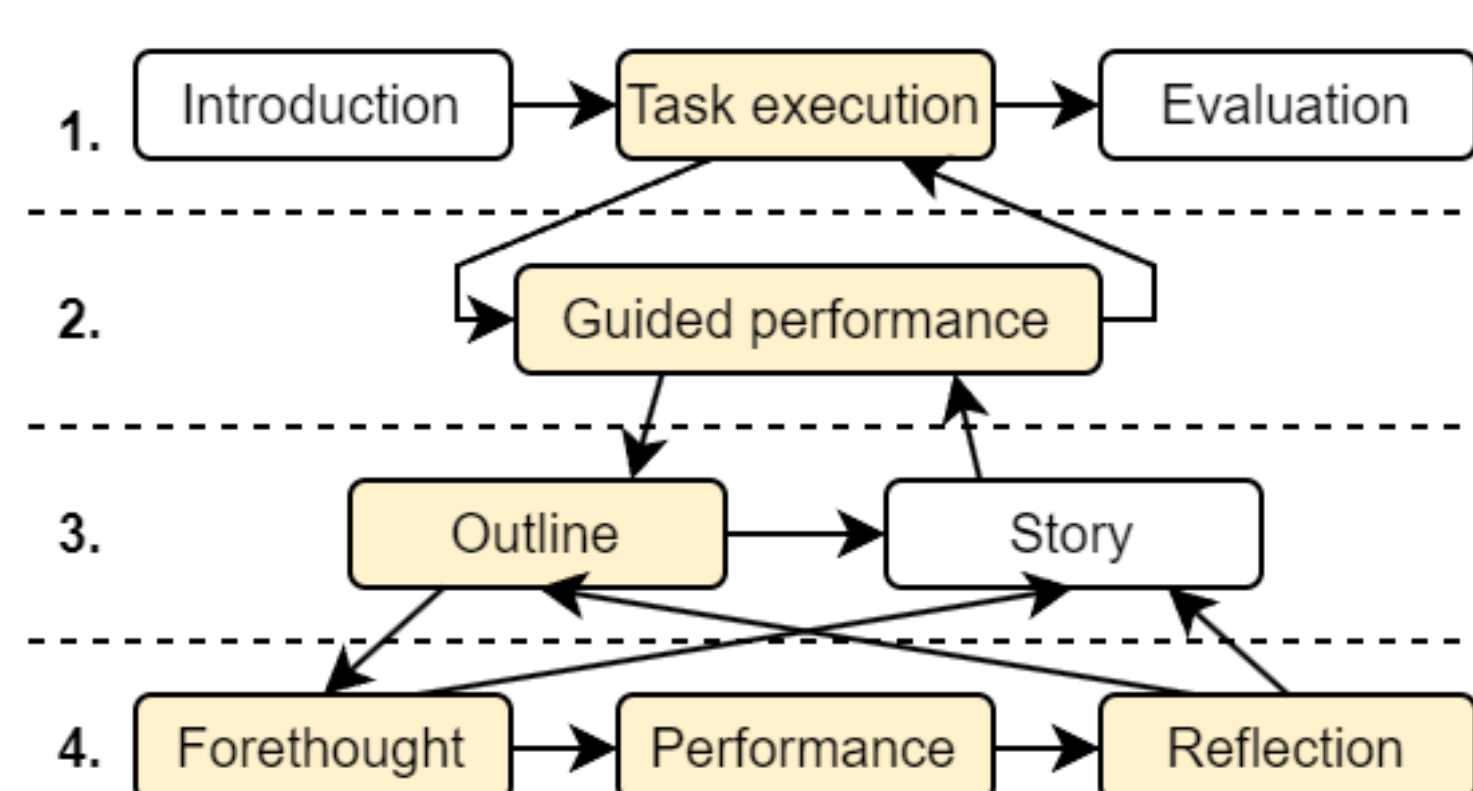


Fig. 2. Moral schema of Virtual Tutor

Moral schemas rely on semantic spaces defined in terms of **intensions** (Fig. 4). Here an **intension** denotes a connotation, or the meaning of a speech act considered as a "sign".

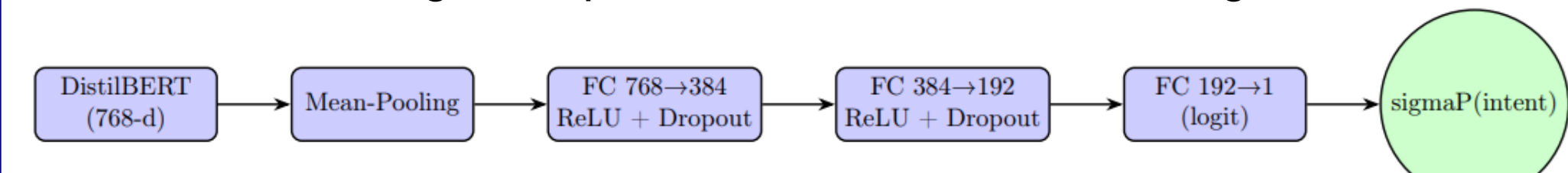


Fig. 3. The neural network architecture responsible for labeling speech acts along intension scales (intension antonym pairs).

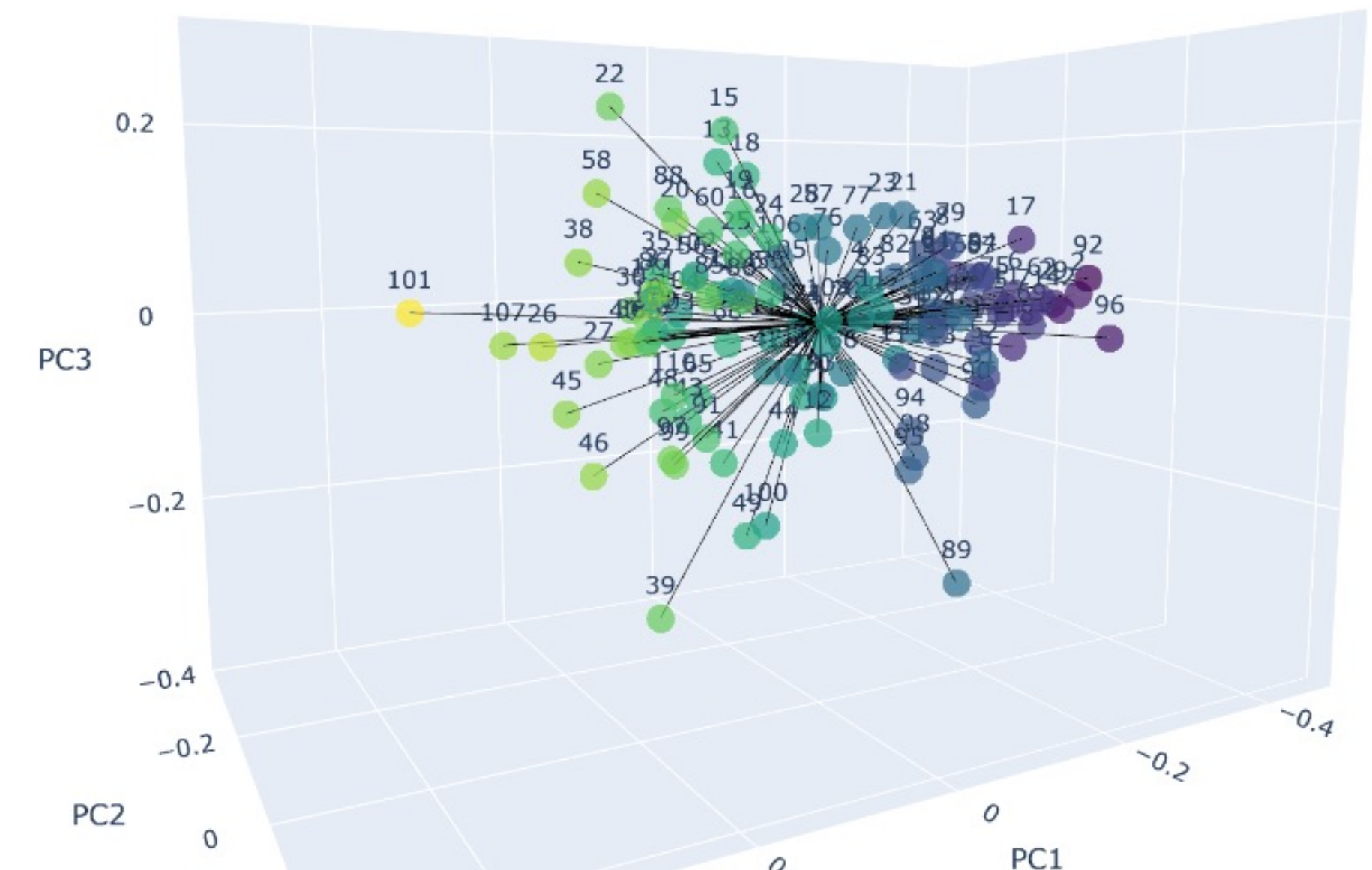


Fig. 4. Projection of intension scales onto the first 3 principal components in the computed embedding (lines originate at 0).

Experimental Design and Procedures

Participants were 48 undergraduate (4th year) NRNU MEPhI students majoring in Software Engineering: 7 women and 41 men, all native Russian speakers, white, 21 to 23.

Essay topics were selected from the participants' field of study. They were: (1) "The role of open-source projects in the development of software ecosystems", (2) "The importance of documenting code and projects".

The main three factors influencing essay writing were: **tutor** (whether Virtual Tutor was used or not), **topic** (whether the topic of the essay was 1 or 2), and **order** (was it the first or the second essay written by a particular student).

Procedure. The 48 participants were divided at random into 4 groups of equal size, assigned to 4 separate classrooms. Each group was assigned a unique task configuration to balance all combinations of the 3 factors. Each essay writing took 30 min.

Interaction with Virtual Tutor. Students used a browser chat window on personal laptops to interact with the Tutor (Fig. 5). Entered essay text was analyzed by the Tutor for a feedback.

Result evaluation. Each essay was assessed on 16 criteria using GPT (OpenAI) via custom prompts, scoring 0-100 on a Likert scale. Automated scores were spot-checked and validated manually. All were deemed acceptable. In addition, the validity of the method was confirmed in a separate study, showing a strong agreement between GPT-generated and human expert scores (Pearson $r = 0.84-0.87$, 95% CI).

Criteria. The 16 criteria (labeled a-p: Table 3) included two groups, one (a-h) globally descriptive, and another (i-p) focused on specific features. The focused group was used by Virtual Tutor in generating a guiding feedback for the student. The average score served as the measure of essay quality.

RESULTS & ANALYSIS

The primary objective of the experiment was to test the hypothesis that Virtual Tutor improves the quality of essays written by students. First, collected data were subject to the ANOVAN variation analysis (Table 1). It was found that out of the three factors (**tutor, topic, order**) only **tutor** had a significant effect on the average essay score: the quality of writing with the help of Virtual Tutor was higher ($p < 0.011$). No significant interaction between the three factors was found.

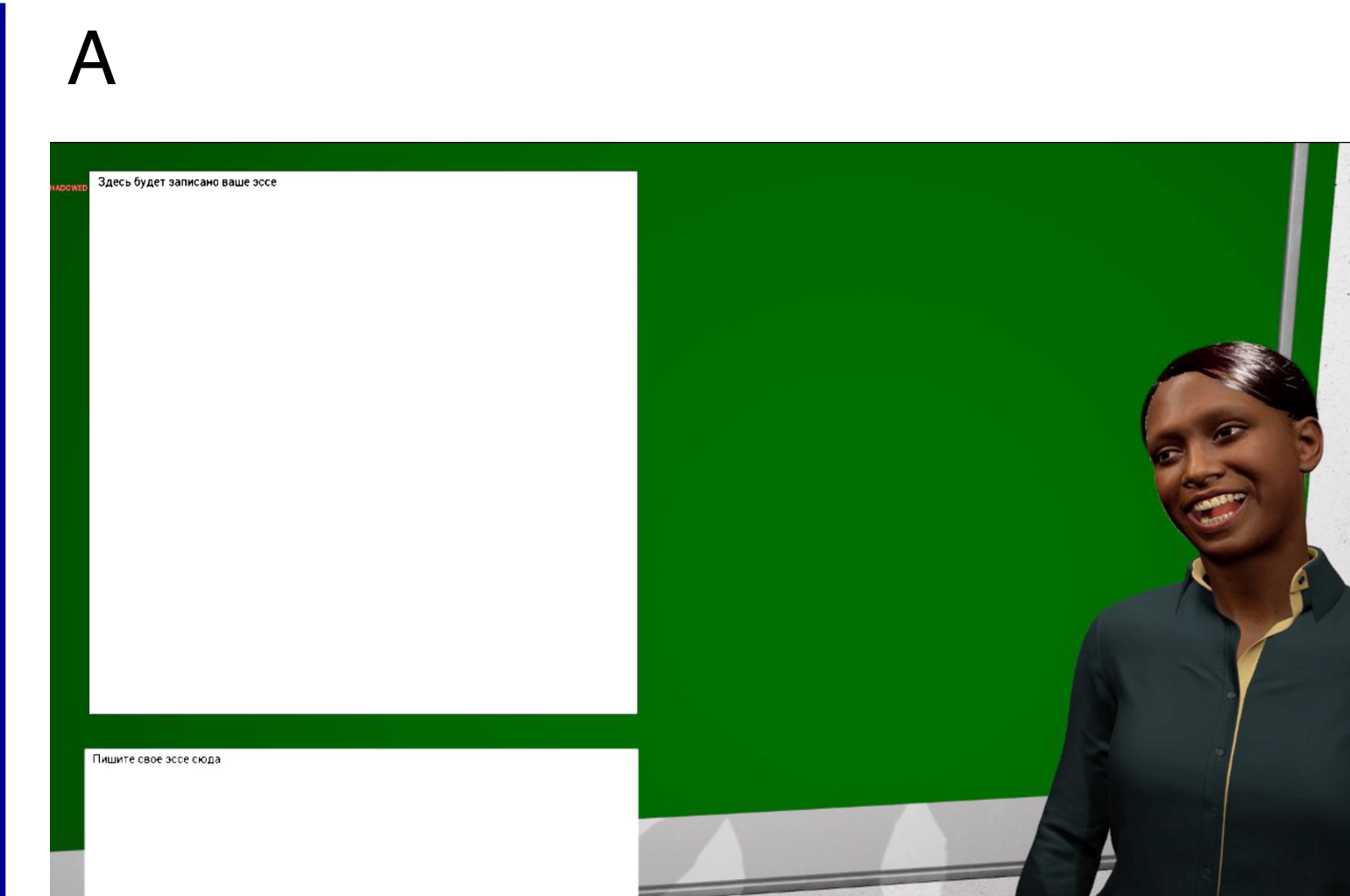


Fig. 5. A: User interface of Virtual Tutor. B: Student working with Virtual Tutor. Video is available: <https://youtu.be/QGga-DikGbl>

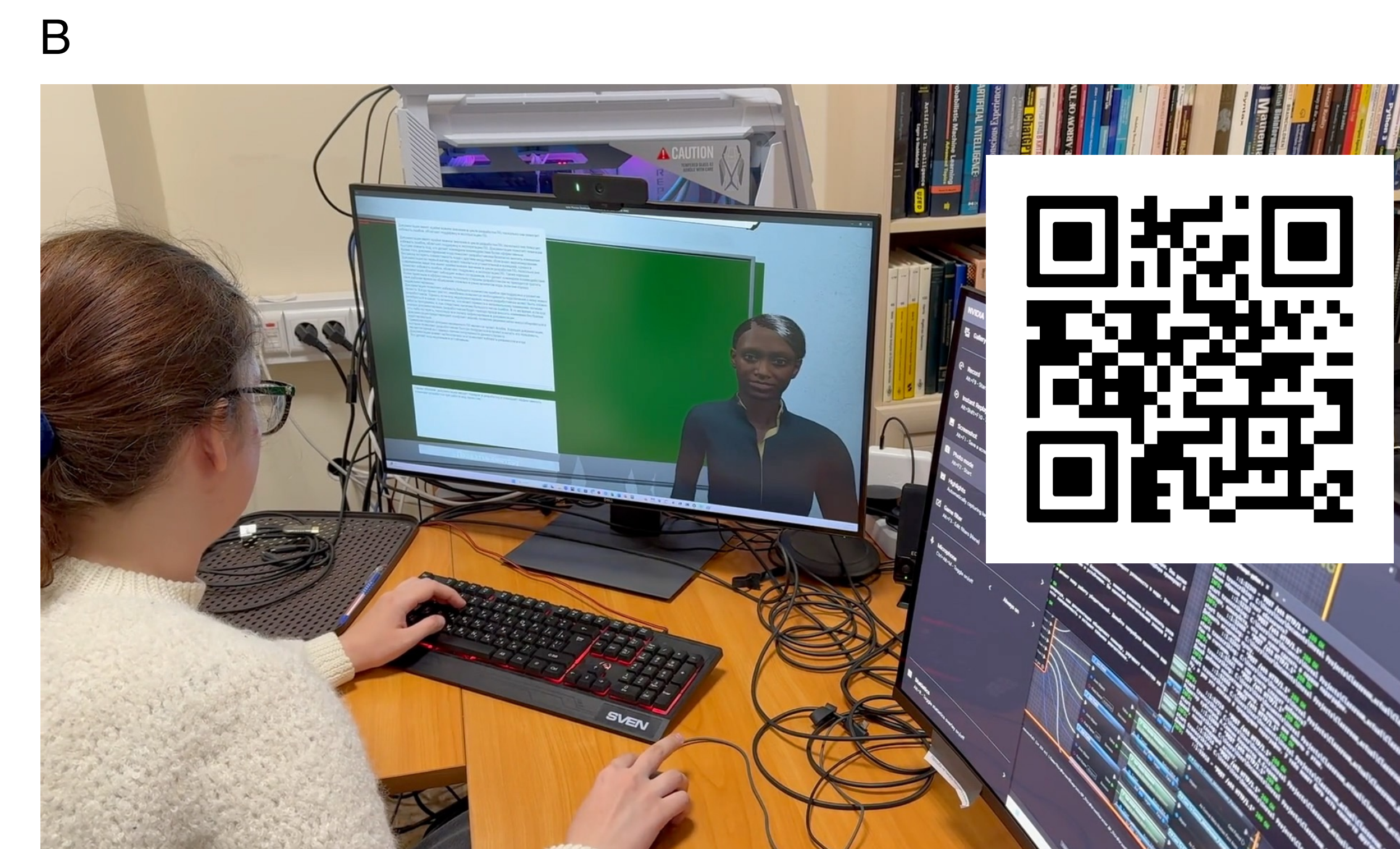


Table 1. ANOVAN table (d.f. stands for "degrees of freedom").

Factor	Sum of Squares	d.f.	Mean Square	F-score	P-value
Tutor (used or not)	273.75	1	273.753	6.76	0.0109*
Topic (1 st or 2 nd topic)	31.98	1	31.978	0.79	0.3766
Order (1 st or 2 nd essay)	48.61	1	48.611	1.20	0.2762
tutor:topic	22.31	1	22.310	0.55	0.4599
tutor:order	0.47	1	0.472	0.01	0.9142
topic:order	18.47	1	18.475	0.46	0.5012

Table 2. Comparison by Initial Tutor Condition.

Group	Mean Score (Task 1)	Mean Score (Task 2)	T-statistic	P-value
Started with Tutor	78.71	76.78	1.0547	0.2971
Started w/o Tutor	75.5	80.20	-2.6443	0.0124*

Table 3. Essay evaluation criteria.

First group of criteria	Second group of criteria
(a) Content of the essay is adequate to its title.	(i) Definition of the subject with intuitive explanations.
(b) The topic is sufficiently covered.	(j) Discussion of the background and related topics.
(c) Main points are well-chosen.	(k) Discussion of questions or problems and existing solutions.
(d) Arguments are sound.	(l) Characterization of the state of the art.
(e) The work is complete.	(m) Description of methodologies or technical details.
(f) Organization of the essay is appropriate.	(n) Discussion of a broad impact on technology/society.
(g) The writing makes good sense, is continuous and logical.	(o) Discussion of future trends and perspectives.
(h) Authors own point of view is articulated.	(p) Conclusions, a bottom-line or take-home message.

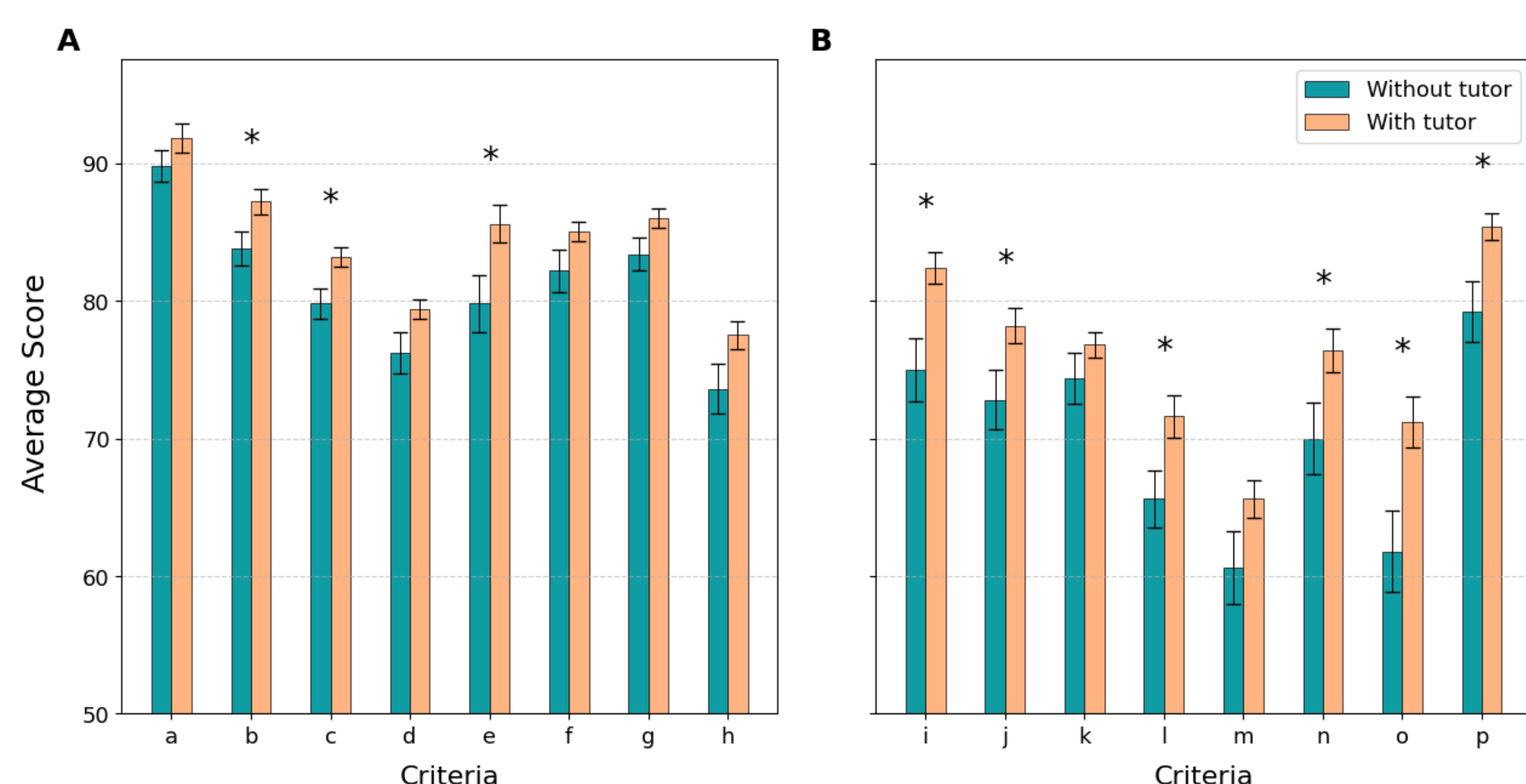


Fig. 6. Comparison of average scores by criterion A - Average Score for criteria (a)–(h), B - Average Score for criteria (i)–(p).

Criterion-level comparison (Group: Started Without Tutor) reveals a systematic improvement across nearly all evaluated aspects of the essays (Fig. 6).

The largest positive effect (uplift of 15.2%) was observed in criterion (o): "Conclusions, a bottom-line or take-home message" ($p < 0.0094$). It shows that Virtual Tutor plays a critical role in prompting students to adopt a more forward-looking and analytical writing approach.

In contrast, **smaller improvements** (around 2–4%) were observed in more fundamental or surface-level aspects such as: (a) "Content of the essay is adequate to its title" (+2.2%, $p < 0.21$), (f) "Organization of the essay is appropriate" (+3.4%, $p < 0.11$), and (g) "The writing makes good sense, is continuous and logically consistent" (+3.1%).

Several other criteria also demonstrated **notable improvements** (7% -10%), including (e) "The work is complete", (i) "Definition of the subject with intuitive explanations", (j) "Discussion of the background and related topics", and (p) "Conclusions, a bottom-line or take-home message" criteria. Cluster analysis using k-means and silhouette in Matlab did not find any significant clustering of the 16 criteria.

CONCLUSIONS

Main findings can be summarized as follows.

- Virtual Tutor significantly improved the quality of written essays ($p < 0.011$).
- Effect persisted following a single experience with Virtual Tutor.
- The most substantial impact (15.2% improvement) of Virtual Tutor was on the criterion (o): "Conclusions, a bottom-line or take-home message". Moderate but consistent improvement (7-10%) was observed for criteria that require higher-order thinking, specifically: (e) completeness, (i) subject definition, (j) background discussion, and (p) clear conclusions.

Take-home message: The BICA+LLM+SRL integration works for educational technologies.

INFO & ACKNOWLEDGMENTS

- Anatoly A. Dolgikh: tolick.dolgih2013@yandex.ru
- Dmitry L. Khabarov: vip.dima020210@mail.ru
- Alexei V. Samsonovich: alexei.samsonovich@gmail.com
- Project URL: https://gitlab.com/bica_lab/virtual_tutor

The authors are grateful to the following NRNU MEPhI students who made their major contributions to the Virtual Tutor project: Anastasia Navrotskaya (metahuman-based visualization in Unreal Engine 5), Fyodor Bulygin (emotion & gaze recognition from Zoom video), Aleksey Mikhnev (initial prototype design).