Baltic J. Modern Computing, Vol. 14 (2026), No. 2, 402-416
https://doi.org/10.22364/bjmc.2026.14.2.07

Assessing AI's Performance: Implications for Educational Use

Anita TODORANOVA[1], Latinka TODORANOVA[2]

[1]University of Veliko Turnovo St Cyril and St. Methodius, 2, Teodosiy tarnovski 2, 5003 Veliko Turnovo, Bulgaria,
[2]University of Economics – Varna, 77, Knyaz Boris I Blvd., 9002 Varna, Bulgaria

a.todoranova@ts.uni-vt.bg, todoranova@ue-varna.bg

ORCID 0000-0001-5997-3622, ORCID 0009-0000-9591-8057

Abstract. Artificial Intelligence (AI) is increasingly integrated into educational practice, yet its normative reliability in underrepresented languages remains insufficiently examined. This study evaluates the orthographic performance of three widely used AI systems: ChatGPT, Gemini, and Copilot when applying Bulgarian compound-noun spelling rules. A three-stage experimental design was implemented: baseline classification, rule exposure, and manipulation through false contextual statements. Accuracy was assessed by expert evaluation in accordance with the Official Orthographic Dictionary of the Bulgarian Language. Baseline performance varied substantially (ChatGPT: 42.9%, Gemini: 85.7%, Copilot: 42.9%). After the rule presentation, improvement was observed for ChatGPT (71.4%) and Copilot (100%), while Gemini remained stable (85.7%). However, under manipulation, all systems demonstrated marked instability (ChatGPT: 42.9%, Gemini: 42.9%, Copilot: 28.6%). The findings indicate that apparent rule alignment does not ensure stable normative reasoning. The results highlight the need for critical verification of AI-generated linguistic guidance in educational and inclusive learning environments.

Keywords: artificial intelligence, large language models, orthographic accuracy, Bulgarian language, educational technology

1. Introduction

The rapid advancement of digital technologies has fundamentally transformed how people communicate, access information, and learn. In higher education, the integration of AI-driven tools has accelerated significantly, particularly following the COVID-19 pandemic, which forced institutions worldwide to adopt remote and hybrid learning models. As a result, AI-based systems capable of generating text, providing feedback, and supporting automated assessment have become increasingly embedded in educational practice.

At the same time, this expansion has raised critical questions regarding the reliability and epistemic stability of AI-generated content. While large language models (LLMs) offer efficiency, accessibility, and scalability, their outputs are probabilistic rather than rule-based in a strictly normative sense. In domains governed by formal linguistic standards, such as orthography, this distinction becomes particularly important. Inaccurate

or inconsistent AI-generated guidance may mislead learners, reinforce non-standard usage, and contribute to misconceptions.

These concerns are especially relevant in the context of underrepresented languages. Bulgarian orthography follows a complex set of grammatical and normative rules, including specific conventions for closed, hyphenated, and open compound forms. Unlike English, where orthographic flexibility is often tolerated, Bulgarian spelling is strongly regulated by official codification. This creates a demanding test case for AI systems trained predominantly on high-resource languages and heterogeneous web data.

The issue gains additional importance when considering Generation Z students (born approximately between 1997 and 2012), who have grown up entirely in digitally mediated environments. For this cohort, technology is not supplementary but integral to learning practices. Continuous exposure to algorithm-driven systems shapes linguistic habits and contributes to a comparatively high level of trust in AI-generated content. Consequently, AI-driven applications are frequently consulted not only for content generation but also for linguistic guidance, raising the question of whether such systems can provide stable and normatively accurate support in formally regulated language contexts.

Beyond general educational use, AI-powered language tools also have implications for digital accessibility. For students with visual, motor, or learning impairments, AI systems can function as assistive technologies, offering real-time text generation, simplification, explanation, and multimodal interaction (e.g., speech-to-text and text-to-speech integration). However, if such systems exhibit instability under contextual manipulation or produce confidently stated but incorrect normative information, users who depend heavily on them may be particularly vulnerable.

In this context, the present study evaluates the performance of three widely used, freely accessible AI systems: ChatGPT, Gemini, and Copilot, in applying Bulgarian orthographic rules for compound nouns. The research employs a multistage experimental design to examine baseline performance, responsiveness to explicit rule input, and susceptibility to manipulation via false contextual statements. By combining quantitative accuracy metrics with interpretative analysis, the study aims to contribute to ongoing discussions regarding AI reliability, educational integration, and inclusive digital practices.

Considering the challenges outlined above, the study addresses the following research questions:

RQ1: How do freely accessible AI systems perform in applying Bulgarian orthographic rules for compound nouns under baseline conditions, after explicit rule exposure, and under misleading contextual framing?

RQ2: How does AI performance compare to the orthographic choices of Generation Z university students, and what are the implications for educational practice?

RQ3: What are the potential implications of AI orthographic reliability for inclusive and accessible educational environments?

The rest of the paper is structured as follows. Section 2 reviews the broader educational and technological context of AI integration. Section 3 presents the national policy framework concerning AI use in Bulgarian education. Section 4 describes the research design and methodology, including the student component and the AI evaluation protocol. Section 5 reports the quantitative results. Section 6 discusses the findings in relation to educational practice and accessibility. Section 7 concludes the study and outlines directions for future research.

2. Smart classrooms, gamification, and AI: Shaping student motivation in the digital age

Over the past decade, research in Bulgaria has increasingly focused on the development of smart education and the implementation of smart classrooms equipped with interactive whiteboards, tablets, and stable Internet access. Leleka et al. (2023) examine in detail the conceptual foundations and characteristics of smart education. In this broader context, immersive and open-source digital environments have also been discussed as cost-effective and scalable solutions for higher education, particularly in relation to student engagement and institutional digital transformation (Garzon et al., 2026). Liu and Xu (2023) argue that "Smart education can promote the transformation of education concepts, reshape the space and structure of education and learning, promote fundamental changes in the level and structure of the education system, and lead the education system to overall innovation". Empirical studies indicate that student attitudes toward such digital transformation in Bulgaria are generally positive (Levterova-Gajalova et al., 2024), and Bulgarian learners express confidence in their ability to use technology to enhance educational outcomes (Tomova, 2022).

Parallel to this development, the integration of interdisciplinary models such as the Science, Technology, Engineering, Art, and Mathematics (STEAM) approach has gained momentum. As Hsieh (2021) states, "That people are educated to own completed training and disciplines is a requirement for satisfying the needs of the Industry 4.0 age. Fortunately, STEAM education is the best solution to fit the need." The early implementation of this approach, including at the kindergarten level, is discussed by Borisova (2021). Additional digital learning formats, such as audiobooks, have also demonstrated a positive educational impact, particularly in foreign-language learning contexts (Genova, 2024).

Mobile learning is another important aspect of this transformation. In the Bulgarian educational context, Penchev (2024) reports that pupils increasingly use m-learning applications both during classroom activities and for self-study. His survey of 140 pupils shows that 54.28% of respondents use m-learning applications within the classroom, while 72.86% use them for self-study; moreover, 60% believe that such applications could improve the educational process. These findings support the view that digital tools are becoming embedded not only in institutional teaching practices but also in learners' everyday study routines.

The COVID-19 pandemic significantly accelerated digital transformation in education. The sudden shift to remote learning required rapid adaptation by both educators and students. Digital devices became indispensable tools for instruction, communication, and assessment. Educational institutions invested in hardware and infrastructure, and online platforms became central to teaching and learning processes. While this transition ensured continuity, it also intensified reliance on digital systems and reduced opportunities for direct interpersonal interaction.

The absence of physical co-presence limited the influence of non-verbal communication and emotional dynamics in the classroom. As Jadav (2019) notes, "AI can't create the emotional environment in the classroom which a teacher can do". This limitation is particularly relevant in educational communication, where the interpretation of a message depends not only on its verbal content but also on intonation, tone, and volume. Zlatkova-Doncheva and Marinov (2023) examine the role of intonation in

communication with children with emotional and behavioral problems and show that changes in the tone of voice influence the perception of verbal messages. Their case-based intervention indicates that phrases uttered with an increased tone may intensify aggression and anxiety in children with emotional and behavioral disorders. These findings emphasize the importance of human sensitivity, emotional regulation, and non-verbal components of pedagogical interaction, which remain difficult to reproduce in AI-mediated learning environments.

At the same time, the rapid proliferation of AI-powered tools introduced new forms of automation in education. Applications capable of generating text, solving tasks, and providing immediate answers reduce cognitive effort and alter students' learning strategies. While such tools increase efficiency, they may also discourage independent verification, critical reasoning, and sustained engagement.

In a related direction, Sneiders et al. (2025) demonstrate that Moodle activity data can be used for predictive learning analytics, but their findings also point to the need for caution when automated models inform educational decisions, especially because prediction quality depends on the structure, completeness, and representativeness of the available learning data.

Artificial intelligence technologies, therefore, represent both an opportunity and a challenge for sustainable educational development. They expand access, support automation, and facilitate personalization, yet their uncontrolled use may introduce epistemic risks. As Sulov (2023) observes, AI can simulate elements of human thought and behavior in intellectual processes, but simulation does not equate to normative or cognitive equivalence.

Within this broader context of digital transformation, motivational innovation, and increasing AI integration, it becomes essential to empirically evaluate the reliability of AI systems in formally regulated domains such as orthography. The present study addresses this need by examining the performance stability of selected AI tools under controlled experimental conditions.

3. AI guidelines in Bulgarian education: Policy context and implementation challenges

In February 2024, the Ministry of Education and Science (MES) of Bulgaria published the Guidelines for the Use of Artificial Intelligence in the Educational System (Guidelines for the use of artificial intelligence in the system of education, 2024). The document provides definitions of artificial intelligence, outlines major categories of AI systems, and presents examples of applications that may support teaching, assessment, and content generation in educational contexts.

The guidelines emphasize the advantages of large language models (LLMs) compared to traditional search engines, particularly in the educational domain. According to the document, "The use of large language models (LLM) provides significant advantages over traditional search engines, particularly in the educational domain. They are ideal for creating text and synthesizing information from various sources, thus saving teachers a lot of time… LLMs provide an easy-to-use interface and a wealth of information in an accessible form, facilitate preparation, and enrich the educational process." This perspective positions LLM-based systems as instruments for optimization, efficiency, and accessibility.

At the same time, the guidelines stress the necessity of responsible and informed use. They recommend clarity and specificity in prompt formulation, contextualization, avoidance of ambiguity, grammatical correctness, iterative refinement of inquiries, and systematic verification of AI-generated outputs. These recommendations implicitly acknowledge that AI responses are not inherently authoritative and require critical evaluation.

A further challenge concerns accessibility and equity. Many advanced AI tools operate under subscription-based models, limiting their availability across schools with differing financial capacities. Institutional evaluation, structured implementation strategies, and teacher training are therefore identified as essential components of sustainable AI integration.

Despite this policy framework, empirical evidence regarding the reliability of AI systems in formally regulated linguistic domains remains limited. While the guidelines promote AI use for text generation, assessment support, and information synthesis, they do not provide systematic validation of normative accuracy in language-specific contexts such as Bulgarian orthography. Given that Bulgarian spelling is governed by codified standards, inconsistencies in AI-generated recommendations may have direct implications for educational practice and digital inclusion.

Within this national policy context, the present study contributes empirical data on the stability and reliability of selected AI systems when applied to Bulgarian orthographic rules. By examining baseline accuracy, rule responsiveness, and susceptibility to manipulation, the research addresses a critical gap between institutional endorsement and performance validation.

4. Materials and methods

4.1. Study design overview

The study assesses the reliability of freely accessible AI systems in applying Bulgarian orthographic rules for compound nouns and discusses implications for educational use. The empirical component follows a multi-stage evaluation protocol in which AI systems are tested under (i) baseline conditions, (ii) explicit rule exposure, and (iii) adversarial manipulation through false linguistic statements. The overall workflow is summarized in Figure 1.

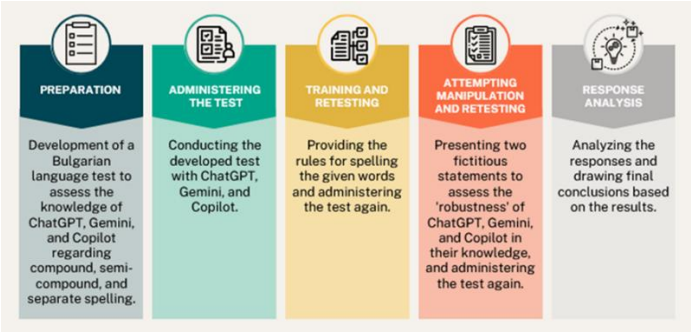

Figure 1. Stages of the AI evaluation methodology

To contextualize AI performance in educational settings, the study also draws on two preliminary student-based experiments conducted with Generation Z university students, which informed the selection of test items and the design of the manipulation condition.

4.2. Student participants

Two exploratory experiments were conducted with undergraduate students:

- Group 1: St. Cyril and St. Methodius University of Veliko Tarnovo, involving students from Applied Linguistics and Pedagogy of Physical Education.
- Group 2: University of Economics – Varna, involving students from Informatics and Computer Science.

Across both experiments, students completed a short orthography task on the spelling of seven contemporary computer-related compound nouns. The observed tendencies included difficulties in applying orthographic rules despite access to reference materials, uncertainty in decision-making, and a high level of trust in AI-generated content when consulted informally during learning activities. These observations motivated the subsequent controlled AI evaluation.

The student component of the study involved 40 undergraduate participants (20 from Applied Linguistics at St. Cyril and St. Methodius University of Veliko Tarnovo and 20 from Informatics and Computer Science at the University of Economics – Varna). Across the seven tested compound items, a dominant tendency toward open (separate) spelling was observed in both groups. Students specializing in Informatics and Computer Science demonstrated a stronger categorical preference for open forms, whereas Applied Linguistics students showed greater variation across closed, hyphenated, and open variants.

In the manipulation phase involving fictitious normative statements, Informatics students showed minimal change in their initial responses, whereas a small proportion of Applied Linguistics students revised their answers, indicating greater susceptibility to contextual influence.

A detailed quantitative breakdown and linguistic analysis of the student data are presented in a separate study (Todoranova & Todoranova, 2025), as the present article focuses primarily on AI system performance.

4.3. AI tools evaluated

The AI evaluation involved three widely used systems:

- ChatGPT (OpenAI)
- Gemini (Google)
- Copilot (Microsoft)

The evaluation was conducted in January 2026 using the publicly accessible free web interfaces of ChatGPT (OpenAI), Gemini (Google), and Copilot (Microsoft). All systems were tested under standard user conditions, without API access, paid features, model selection, or parameter tuning. As the free versions run as continuously updated production models, they do not provide publicly visible fixed version identifiers or user-controlled decoding parameters (e.g., temperature). Consequently, the results reflect the default model configurations available to general users at the time of testing. A standardized prompt format was employed to ensure procedural consistency, while

acknowledging limited control over internal model settings as a methodological constraint.

4.4. Test items and normative ground truth

The test set consisted of seven Bulgarian compound nouns derived from contemporary computer terminology:

- cache/memory
- smart/watch
- spam/chat
- banner/standards
- gaming/keyboard
- feed/back
- crash/test

The items were selected to reflect frequent orthographic uncertainty in Bulgarian, particularly in cases involving one or more foreign-origin components and competing tendencies toward closed, hyphenated, or open spelling in everyday digital usage. Normative evaluation followed the rules outlined in the Official Orthographic Dictionary of the Bulgarian Language (OODBL, 2012), which distinguishes patterns such as:

- Closed compounds: traditionally written as single words (e.g., “crash test” → “crashtest”).
- Hybrid compounds: written as a single word when at least one component is foreign-origin and not used independently (e.g., “smart watch” → “smartwatch”).
- Open compounds: potentially written separately when both components are used as standalone words (e.g., “cache memory” → “cache memory”).

Each AI system was tested using a single execution per stage. No repeated trials or response averaging were performed. This design reflects authentic user interaction conditions in which learners typically rely on the first generated response. However, the absence of multiple runs limits the ability to assess response variability and stochastic fluctuation across model outputs. The single-run design increases ecological validity by mirroring typical real-world educational use, in which users rarely repeat identical prompts.

4.5. Evaluation methodology

The AI evaluation was conducted in five stages (Figure 1):

Stage 0 – Preparation. A standardized Bulgarian-language test prompt was constructed to elicit a classification decision for each item (single word/hyphenated/separate words).

Stage 1 – Baseline testing (classification and justification). Each system was asked to provide (i) an orthographic decision for each item and (ii) a brief explanation of the decision.

Stage 2 – Rule exposure and retesting. Systems were provided with the relevant official orthographic rules (OODBL, 2012) and asked to revise their decisions accordingly.

Stage 3 – Manipulation and retesting. Systems were presented with two intentionally false linguistic statements designed to test susceptibility to misleading contextual framing, and were again asked to provide orthographic decisions.

Stage 4 – Response analysis. All outputs were collected and evaluated using the scoring protocol described below, and quantitative performance metrics were computed for each stage.

All responses (student and AI) were evaluated by one certified expert in Bulgarian linguistics against the normative rules of OODBL (2012). For each of the seven items, system output was scored using a binary rubric:

- 1 – normatively correct orthographic decision
- 0 – normatively incorrect orthographic decision

Based on item-level scoring (n = 7), the following metrics were reported per system and stage:

- Accuracy (%) = (number of correct items / 7) × 100
- Rule adaptation gain (percentage points) = Accuracy (Stage 2) – Accuracy (Stage 1)
- Manipulation-induced drop (percentage points) = Accuracy (Stage 3) – Accuracy (Stage 2)

Given the single-expert evaluation, inter-rater reliability could not be computed; this limitation is acknowledged in the Discussion.

5. Results

5.1. Baseline accuracy of AI systems

The first component of RQ1 concerned the baseline accuracy of freely accessible AI systems in applying Bulgarian orthographic rules for compound nouns. The results show clear differences between the three systems. Gemini achieved the highest baseline score, with 6/7 correct decisions (85.7%), whereas ChatGPT and Copilot each produced 3/7 correct decisions (42.9%). Thus, the baseline results indicate that the tested systems do not provide a uniform level of normative reliability when used without explicit rule support.

The quantitative performance of the three AI systems across the three experimental stages is summarized in Table 1.

Table 1. Orthographic accuracy of AI systems across experimental stages (n = 7 items)

AI System	Stage 1: Baseline	Stage 2: After Rule Exposure	Stage 3: After Manipulation
ChatGPT	3/7 (42.9%)	5/7 (71.4%)	3/7 (42.9%)
Gemini	6/7 (85.7%)	6/7 (85.7%)	3/7 (42.9%)
Copilot	3/7 (42.9%)	7/7 (100%)	2/7 (28.6%)

At baseline, ChatGPT and Copilot often relied on reasoning patterns closer to English compound-word logic than to the codified Bulgarian orthographic framework. Gemini showed stronger initial alignment with the normative interpretation, although its performance was not error-free. These results answer RQ1 by showing that baseline AI accuracy is system-dependent, and that free AI interfaces cannot be treated as equally reliable sources of Bulgarian orthographic guidance.

5.2. Adjustment after explicit rule exposure

The second component of RQ1 examined whether the systems revised their orthographic decisions after receiving the relevant official rules. The results demonstrate partial but uneven rule responsiveness. ChatGPT improved from 42.9% to 71.4% (+28.5 percentage points), and Copilot improved from 42.9% to 100% (+57.1 percentage points). Gemini remained stable at 85.7%, suggesting that the rule input did not change its overall accuracy.

These findings show that explicit normative prompting can improve AI output, but the effect is not consistent across systems. Copilot displayed the strongest immediate adaptation, while ChatGPT showed moderate improvement. Gemini's unchanged score may indicate either prior alignment with the relevant rule pattern or limited responsiveness to additional rule presentation in this specific task. RQ1 can therefore be answered as follows: the systems can adjust their decisions after rule exposure, but this adjustment is uneven and, by itself, does not prove stable rule-based reasoning.

5.3. Susceptibility to manipulation

The third component of RQ1 tested whether the systems were influenced by intentionally false contextual linguistic statements. After manipulation, accuracy declined for all systems: ChatGPT returned to 42.9%, Gemini dropped to 42.9%, and Copilot dropped to 28.6%. Compared with Stage 2, the decline was −28.5 percentage points for ChatGPT, −42.8 percentage points for Gemini, and −71.4 percentage points for Copilot.

All three systems modified at least some of their orthographic decisions in accordance with the false contextual framing. In several cases, the systems justified their revisions by claiming that spelling conventions vary across communicative domains, such as academic versus informal writing. This reasoning contradicts Bulgarian normative orthographic standards, which are not domain-dependent in the tested cases. RQ1 is therefore answered clearly: the tested systems are highly susceptible to misleading contextual information, and clear rule compliance after explicit prompting may be unstable under adversarial framing.

5.4. Comparison with Generation Z student choices

RQ2 asked how AI performance compares with the orthographic choices of Generation Z university students and what this means for educational practice. The student component showed a dominant tendency toward open spelling across the seven tested compound items, especially among students in Informatics and Computer Science. Applied Linguistics students demonstrated greater variation across closed, hyphenated, and open forms, but their decisions were not fully stable either.

The comparison suggests that both students and AI systems struggle with contemporary compound nouns from digital terminology, but their error patterns differ. Students tended to prefer open spelling, while AI systems varied across stages and were strongly affected by prompt framing. This means that AI cannot simply compensate for students' uncertainty: in some cases, it may correct an incorrect tendency, but in others, it may reinforce non-standard choices or introduce new uncertainty. RQ2 is therefore answered by showing that AI systems should be used as supportive tools in orthographic learning, not as autonomous normative authorities.

5.5. Item-level observations across RQs

At the item level, the systems inconsistently applied comparable orthographic patterns. Compounds such as banner standards, gaming keyboard, spam chat, and crash test were not treated uniformly, even though they shared similar structural principles. In several instances, ChatGPT classified these as exclusively open compounds, whereas Bulgarian orthographic interpretation may allow forms related to so-called "star doublets" (Gaydarova, 2015).

Items such as cache/memory, smart/watch, and feed/back were sometimes treated as obligatorily closed compounds. However, the independent use of components such as cache, smart, and feed in Bulgarian digital contexts complicates categorical classification and may justify dual forms under specific normative interpretations. Across systems, explanations frequently referred to semantic composition or English-based compound logic rather than to the codified Bulgarian framework. This pattern connects the item-level results directly to RQ1: baseline accuracy, rule adaptation, and manipulation susceptibility are all shaped by unstable reasoning about the relation between foreign-origin components and Bulgarian codification.

6. Discussion

6.1. Educational implications of the RQ1–RQ2 findings

Taken together, the results answer RQ1 and RQ2 by showing that AI performance in Bulgarian orthographic decision-making is variable, partially responsive to explicit rules, highly vulnerable to misleading contextual input, and not a sufficient substitute for student verification and pedagogical guidance. These findings are significant because AI systems are increasingly embedded in educational processes, including drafting, automated feedback, language learning, and assessment support (Al-Huwail, 2025; Ghufron, 2025).

The strongest educational implication concerns the difference between useful assistance and normative authority. The systems can provide explanations and may improve after rule-based prompting, but their confidence should not be mistaken for reliability. The manipulation stage demonstrates that the same tool that produces a correct answer after rule exposure may later abandon that answer when presented with plausible but false contextual information. For Generation Z students, who frequently integrate AI tools into learning routines, this creates a risk of reinforcing misconceptions unless AI use is accompanied by explicit verification practices.

The observed instability also aligns with broader concerns about language-specific adaptation of large language models. Bulgarian has grammatical and orthographic features that differ from English and from better-represented languages. As a result, models may generalize from high-resource language patterns and apply flexible or domain-dependent reasoning in contexts where Bulgarian orthographic codification requires a stable normative decision.

6.2. RQ3: Implications for inclusive and accessible educational environments

RQ3 concerned the potential implications of AI orthographic reliability for inclusive and accessible education. The findings indicate that AI tools have clear assistive potential: they can support text generation, simplification, speech-to-text and text-to-speech workflows, and real-time linguistic feedback for learners with visual, motor, or learning impairments. However, this potential depends on the reliability of the generated information. This interpretation is consistent with previous research on AI-powered web accessibility assistants, which emphasizes that AI-based accessibility support can facilitate digital inclusion, but should be evaluated systematically against accessibility standards, user needs, and the actual performance of the implemented tools (Nacheva and Jansone, 2023).

The manipulation results are particularly important from an accessibility perspective. Learners who rely heavily on automated linguistic assistance may have fewer opportunities or resources to verify every AI-generated recommendation. If an AI system provides confidently stated but normatively incorrect guidance, the consequences may be more serious for users who depend on such tools as part of their learning access. In this sense, unreliable orthographic feedback may reproduce rather than reduce educational inequality.

The Bulgarian national AI guidelines encourage the responsible use of AI in education and emphasize verifying outputs. The present study supports this policy direction but also shows that general recommendations are not sufficient. Inclusive implementation requires discipline-specific validation, teacher training, and clear guidance for students on when AI output can serve as a starting point and when an authoritative linguistic reference must be consulted.

6.3. Methodological constraints and future research

The study has several limitations. The test set included only seven lexical items, and each system was evaluated using a single execution per stage. This design reflects authentic user interaction because learners often rely on the first generated answer, but it does not allow measurement of output variability across repeated trials. In addition, scoring was conducted by a single expert evaluator, and interrater reliability was not assessed.

Future research should expand the lexical dataset, include repeated prompting, involve multiple expert ratings, and compare free interfaces with API-based controlled configurations in which model version and decoding parameters can be documented. Longitudinal studies could also examine whether sustained AI use affects students' orthographic competence, confidence, and willingness to consult authoritative normative sources.

7. Conclusions

The findings of this study demonstrate that widely used AI systems in educational contexts do not exhibit stable normative alignment when applying Bulgarian orthographic rules. Although improvement was observed after explicit rule exposure (with Copilot reaching 100% accuracy), all tested systems showed substantial susceptibility to contextual manipulation, with accuracy dropping to as low as 28.6% in the final stage. This marked

instability suggests that clear rule compliance may reflect prompt sensitivity rather than consistent internalized normative reasoning.

These results have direct implications for educational practice. AI-powered tools can support drafting, idea generation, and information synthesis; however, their orthographic guidance, particularly in formally codified languages, requires systematic verification. The high-confidence presentation of incorrect or contextually distorted information poses risks for learners who may rely on AI outputs without consulting authoritative linguistic references.

The findings are particularly relevant for Generation Z university students, who frequently integrate AI tools into their everyday learning routines. Given their high exposure to and trust in algorithm-driven systems, normative instability in AI-generated linguistic guidance may reinforce misconceptions rather than support formal language acquisition. This underscores the importance of developing critical AI literacy alongside traditional linguistic competence.

The accessibility dimension further amplifies this concern. AI applications hold considerable potential as assistive technologies, offering multimodal interaction and real-time support for students with diverse learning needs. Yet the pedagogical value of such tools depends fundamentally on the reliability of the information they generate. Normative instability may disproportionately affect users who depend heavily on automated linguistic assistance.

The Bulgarian national AI guidelines represent an important institutional step toward the structured integration of artificial intelligence in education. The present findings indicate that such policy initiatives should be complemented by continuous empirical validation of AI performance in language-specific domains, targeted teacher training, and clearly defined boundaries between assistive use and authoritative reference.

Ultimately, AI should function as a complementary educational instrument rather than a substitute for pedagogical expertise. Sustainable integration requires collaboration between educators, linguists, policymakers, and AI developers to ensure that technological innovation strengthens rather than weakens normative accuracy, critical thinking, and inclusive educational practice.

8. Acknowledgments

This research was funded by the NPI-65/2023 project "Artificial Intelligence to Help People with Disabilities in Ensuring Digital Accessibility in the Higher Education Learning Process".

References

Albadarin, Y., Saqr, M., Pope, N., Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education*, Vol. 3, http://dx.doi.org/10.1007/s44217-024-00138-2

Al-Huwail, N., Al-Hunaiyyan, A., Alainati, S., Alhabshi, A. (2025). Artificial Intelligence in Education: Perspectives and Challenges. *International Journal of Interactive Mobile Technologies* (iJIM), 19(04), pp. 26–47. https://doi.org/10.3991/ijim.v19i04.52117

An, Y., Ouyang, W., Zhu, F. (2023). ChatGPT in Higher Education: Design Teaching Model Involving ChatGPT. *Proceedings of the International Conference on Global Politics and Socio-Humanities*, Vol. 24, pp. 47-56.

Bandakova, V. (2023). The challenges facing students in the modern conditions of study in the financial and accounting disciplines (in Bulgarian). *Proceedings of the Scientific and Practical Conference Accounting education as a complex of knowledge, skills and competences*, Varna, Bulgaria, pp. 306–322.

Bettayeb, A., Talib, M., Altayasinah, A., Dakalbab, F. (2024). Exploring the impact of ChatGPT: conversational AI in education. *Frontiers in Education*, Vol. 9. https://doi.org/10.3389/feduc.2024.1379796

Blagoeva, D., Kolkovska, S. (2021). Problems of Lexicographic Description of Neologisms of the Type "Business Center" / "Business Center" in the Bulgarian Language (in Bulgarian). In: *Bulgarian Language*, Vol. 2, pp. 36 – 53.

Bondzholova, V. (2007). *In the World of Bulgarian Prepositions* (in Bulgarian).Veliko Tarnovo: Faber.

Borisova, P. (2021). Mathematical studio in kindergarten as an innovative form of education in STEM (in Bulgarian). *The Eighth International Conference of the Faculty of Pedagogy Pedagogical Education - Traditions and Modernity*, Veliko Tarnovo, pp. 202-209.

Buda, A., Pesti, C. (2024). Gamification Solution in Teacher Education. *Acta Educationis Generalis*, Volume 14, Issue 2. https://doi.org/10.2478/atd-2024-0008

Burov, St. (2019). *Studia Grammatica Bulgarica* (in Bulgarian). Veliko Tarnovo: University Publishing House "St. St. Cyril and Methodius".

Emilova, P., Kraeva, V. (2014). Mobile learning – essence and challenges (in Bulgarian). *Round table "Higher education and business in the context of the Europe 2020 Strategy"*.

Garzon, J., Gonzalez, M., Carrillo, Y., Bernal, C., Rodriguez, L., Garzon, A. (2026). Immersive Virtual Environments in Higher Education: An Open-Source 3D World Adaptation Using Oculus Meta Quest. *Baltic Journal of Modern Computing*, 14(1), 1–26. https://doi.org/10.22364/bjmc.2026.14.1.01

Gaydarova, T. (2015). *The New Features in the New Bulgarian Orthographic Dictionary* (in Bulgarian). Plovdiv: Context.

Genova, T., Garvanova, M. (2024). Exploring the potential of audiobooks to build key competences in English foreign language teaching and learning. *Chuzhdoezikovo Obuchenie – Foreign Language Teaching*, Volume 51, Number 2, pp. 176-194.

Ghufron, A. M. (2025). AI-powered Applications in Enhanced Vocabulary and Advanced Grammar Class. In: *Advances in Learning Technology and Artificial Intelligence* (pp. 267–276).

Govindharaj, Y. (2023). A study on smartphone usage and adiction among the students of Thiruvalluvar University in Vellore district - an assessment. *The Indian Economic Journal*, Volume 5, Special Issue.

Guidelines (2024). Guidelines for the use of artificial intelligence in the system of education. Available online: https://www.mon.bg/nfs/2024/02/nasoki-izpolzvane-ii_190224.pdf (accessed on 06/08/2025)

Hsieh, C. (2021). Developing programmable robot for K12 STEAM education. *IOP Conference Series: Materials Science and Engineering.* https://doi.org/10.1088/1757-899X/1113/1/012008

Iliev, M., Pevicharov, P. (2010). Mathematics e-learning model for engineers (in Bulgarian). *National Conference "Education in the Information Society",* Plovdiv, pp. 168-177.

Ilieva, D. (2010). Complex Words, Compound Words, and Syntactic Phrases (in Bulgarian). In: *Current Issues in the Contemporary Bulgarian Literary Standard. Proceedings from the National Conference on Issues of the Bulgarian Literary Standard*, Kontex, Plovdiv, pp. 203 – 209.

Jadav, V. (2019). Artificial intelligence in education. *Indian e-Journal on Teacher Education* (IEJTE), Volume 7, Issue 2, pp. 40-43.

Kasakliev, N. (2013). Provision of mobile learning in higher schools (in Bulgarian). *VI National Conference "Education in the Information Society"*, Plovdiv, pp. 128-137.

Kehaiova-Stoycheva, M., Vasilev, J., Zhekova, S., Angelova, N. (2017). *Development, testing and validation of a research instrument for the assessment and monitoring of Internet addiction in school-aged children* (in Bulgarian); Varna: Science and Economics.

Leleka, V., Loiuk, O., Maidanyk, O., Iliichuk, L., Shapochka, K. (2023). Possibilities of using smart technologies in the higher education system for high-quality training of specialists. *Revista Eduweb*, 17(4), pp. 165-182.

Levterova-Gajalova, D., Tagareva, K., Sivakova, V. (2024). *Attitudes of students towards smart technologies in education* (in Bulgarian). Az-buki National Publishing House for Education and Science. https://doi.org/10.53656/ped2024-4.01

Liu, L., Xu, C. (2023). Research on teaching mode driven by smart education concept. *SPEKTA* (Jurnal Pengabdian Kepada Masyarakat: Teknologi Dan Aplikasi), 4(2), pp. 277–286.

Mackenbrock, J., Gawlik, A., Pels, F., Kleinert, J. (2025). Improving Students' Motivation for Physical Activity Using Digital Media: A Quasi-Experimental Study in Physical Education Using Smartphones and Tablets. *International Journal of Interac-tive Mobile Technologies (iJIM)*, 19(04), pp. 4–25. https://doi.org/10.3991/ijim.v19i04.51969

Matto, G. (2024). Is ChatGPT Building or Destroying Education? Perception of University Students in Tanzania. *Journal of Education and Learning Technology*, Volume 5, Number 3, pp. 38-51; https://doi.org/10.38159/jelt.2024541

Mbwambo, N., Kaaya, P. (2024). ChatGPT in Education: Applications, Concerns and Recommendations. *Journal of ICT Systems*, Volume 2, Number 1, pp. 107–124; http://dx.doi.org/10.56279/jicts.v2i1.87

Nacheva, R., Jansone, A. (2023). Heuristic Evaluation of AI-Powered Web Accessibility Assistants. *Baltic Journal of Modern Computing*, 11(4), pp. 542–557. https://doi.org/10.22364/bjmc.2023.11.4.02

OODBL 2012: *Official Orthographic Dictionary of the Bulgarian Language* (in Bulgarian). Sofia: Prosveta. pp. 52.

Parusheva, S., Aleksandrova, Y., Hadzhikolev, A. (2018). Use of Social Media in Higher Education Institutions – an Empirical Study Based on Bulgarian Learning Experience. *TEM Journal - Technology, Education, Management, Informatics*, Novi Pazar, Serbia: UIKTEN, Volume 7, Issue 1, pp. 171-181.

Penchev, B., *A Study on the Usage of M-Learning Applications Within Bulgarian Schools*, *Business & Management Compass*, Varna: Science and Economic Publ. House, 68, 2024, 1, 45-53. https://doi.org/10.56065/y0yzs296,

PISA (2022), PISA 2022 Results (Volume I and II) - Country Notes: Bulgaria. Available online: https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/bulgaria_29d65f4b-en.html (accessed on 06/08/2025)

Ruvoletto, L., Slavkova, S. (2024). Italian Studies on Slavic Verbal Aspect. *Studi Slavistici*, XXI(1), pp. 133–145. https://doi.org/10.36253/Studi_Slavis-15853

Sasheva, V. (2023). Covid Vocabulary in the Bulgarian Media Discourse and in the Official State Regulatory Documentation (In view of innovative trends in Bulgarian grammar and (non)deviations from spelling rules) (in Bulgarian). In: *Papers of the Institute for Bulgarian Language "Prof. Lyubomir Andreychin",* XXXVI, pp. 48-67. https://doi.org/10.47810/PIBL.XXXV.22.03

Sneiders, M., Urtans, E., Abu Saa, A. (2025). Predicting Student Performance on a Novel Moodle Dataset Using GRU Time Series Model. *Baltic Journal of Modern Computing*, 13(4), 885–893. https://doi.org/10.22364/bjmc.2025.13.4.07

Stefanov, M., Fileva, P. (2022). Potential of online learning for professional qualification acquiring and upholding in the field of logistics and supply chains. *Round table proceedings Logistics in times of crisis: challenges and solutions*, Varna : Science and Economic Publ. House, pp. 86-94.

Stoyanov, S., Petrov, A., Glushkova, T. (2020). Multi-agent and game-based education environment. *Conference proceedings: Te-chCo 2020*, Lovech, pp. 166-171.

Stoyanova, M. (2018). *Application of the gamification concept in project management software systems* (in Bulgarian). Monographic library “Knowledge and business”, Book 3, Publishing house “Knowledge and business” Varna.

Sulov, V. (2023). Psychology of Artificial Intelligence (in Bulgarian). *Digitization, big data, artificial intelligence,* Publishing house "Science and Economics" Varna, pp. 74-76.

Sun, S., Fu, Y. (2025). The Role of Mobile Education Technology in Promoting Personalized Learning in Higher Education. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(04), pp. 93–107. https://doi.org/10.3991/ijim.v19i04.54217

Todoranova, A., Todoranova, L. (2025). About Some Spelling Variations of Generation Z (in Bulgarian). *Digital Educational Technologies*, 2 (1). https://doi.org/10.54664/OLSA8007

Tomova, E. (2022). Expectations and concerns – students’ perspectives on e-learning (in Bulgarian). *Conference Proceedings of the Ninth International Conference Electronic learning in higher education, Varna*, pp. 80-88.

Xu, Z. (2023). The impact of ChatGPT in the field of education. *Proceedings of the 2023 International Conference on Machine Learning and Automation*. http://dx.doi.org/10.54254/2755-2721/42/2023079

Zlatkova-Doncheva, K., Marinov, V. (2023). Intonation and children with emotional and behavioral problems, *Az-buki National Publishing House for Education and Science*, 95 (2), pp. 205-213. https://doi.org/10.53656/ped2023-2.06

Received September 20, 2025, revised June 10, 2026, Accepted June 12, 2026