

SOME AI TOOLS FUNCTIONALITY ABOUT GENERATED FOREIGN LANGUAGE TEACHING MATERIALS

Ivelin Georgiev Ivanov

College-Dobrich, Konstantin Preslavsky University of Shumen, Bulgaria, ivelin.ivanov@shu.bg

Petranka Todorova Ivanova

College-Dobrich, Konstantin Preslavsky University of Shumen, Bulgaria, petranka.ivanova@shu.bg

Zhivka Stoyanova Ilieva

College-Dobrich, Konstantin Preslavsky University of Shumen, Bulgaria, zh.ilieva@shu.bg

Abstract: The paper is a further work closely related to previous articles where detailed information was given about the way children's texts for teaching foreign language were generated by means of AI. One of these articles discusses their readability degree with suggestions matching the targeted age group supported by evidences from some online tools for calculating readability indices. The purpose of the current study is to depict another aspect concerning the functionality of the same AI tools, i.e. ChatGPT, DeepL and Grammarly based on previously discussed texts for young and very young learners. Their effectiveness is viewed in terms of reliability and authenticity of the obtained materials. Whereas a plenty of studies have been conducted to discern the authenticity of texts created by human activity among others generated by artificial intelligence, we haven't come across many researches examining the AI judgment about its own production. Although it would be strange to classify it as an awareness, a closely related to the idea term might be used. To register the fact accordingly would be enough.

Our search provides data with the help of AI instruments in terms of plagiarism – a topic discussed while counting on QuillBot and ZeroGPT. The pivotal point is related to the AI capability, or rather degree of capability, to detect authenticity of its own generated materials. The applications mentioned above were used to check for plagiarism of the generated by ChatGPT text and the edited by DeepL and Grammarly versions. For these tests only free versions of the indicated detecting tools were used. The methodology followed included two stages, one for each application with one and the same steps: the story generated by ChatGPT 4o free was tested via ZeroGPT; the originally generated by ChatGPT 4o free story and then edited by DeepL was tested via ZeroGPT; the originally generated by ChatGPT 4o free story, edited by Grammarly, was tested via ZeroGPT; the story generated by ChatGPT 4o free was tested via QuillBot; the originally generated by ChatGPT 4o free story and then edited by DeepL was tested via QuillBot; the originally generated by ChatGPT 4o free story, edited by Grammarly, was tested via QuillBot.

Some of the results gave evidence for a certain need to be on the alert in case of trusting AI while at the same time others provided rationale for surprisingly clear and explicit categorization of the materials. Comparing the results by ZeroGPT and QuillBot we receive more realistic analysis by QuillBot, although in one out of three tests it is incorrect.

The article touches the issue about the way AI detector products work in search of explanations about the results received. The overall conclusion on the basis of our small scale study is that they are still not completely reliable but leverages feel to be closely at hand.

Keywords: AI tools, generated stories, young learners, plagiarism.

1. INTRODUCTION

The development of AI makes it a partner to learners and teachers. With good prompts teachers create lesson plans, quizzes, interesting tasks, useful texts, stories that can combine familiar characters in new situations or new characters in a familiar context. Following this current trend in language education, we generated a story and its variants using ChatGPT, DeepL and Grammarly and tested it for AI generated text through QuillBot and ZeroGPT.

Some authors (Rocha et al., 2015) mention the appearance of eInk as a turning point for utilizing a large number of sources through a single small device with the ability to provide the meaning of any unfamiliar word a student comes across. Ten years later other inventions reached the remarkable high place of admiration waiting for the pros and cons to finally find their recognized place. As reported by Kopczynski and Silvia (2024), a poll in higher education revealed great reliance on AI for purposes such as plagiarism detection, information for students, etc.

The danger to create false information raises questions about ChatGPT reliability (Stockwell, 2024, p. 3, Kopczynski & Silvia, 2024). In addition to ChatGPT hallucination, the conscious cheating on the part of students presenting generated materials to be assessed as their own, outline the negative side of the coin. How far students are aware of the real origin of the texts they read has been examined in a paper by Sofronieva et al. (2025). The results from their study are congruent with other authors', as they pointed out, giving detailed information about the difficulty that people meet in order to guess it right which one involves AI and which one human mind. Gotoman et

al (2025) after analyzing more than 30 resources also reach the conclusion that “despite most detectors attaining accuracy above 50%,” they are inaccurate and unreliable. They (Gotoman et al., 2025) advise “using them carefully and relying on human judgment”. Weber-Wulff et al. (2023) also find out that AI detection tools are “neither accurate nor reliable”. Chaka’s (2024) findings are similar: he confirms “the inconsistency of the detection efficacy of all the tested AI detectors and all the tested anti-plagiarism detection tools.” Unreliability of AI detectors is stressed by authors as Edwards (2023) and Fowler (2023). Others like Glushka (2023) and Shah et al. (2023) find some reliable instruments.

Xia (2025) underlines the importance to teach students to be able to evaluate the AI generated content. Such significance becomes even much more considerable in case AI doesn’t give reliable judgement about its own generated materials. Debates about AI integration in education continue but the sense of inevitability seems to become more and more realistic. A fresh example is Harakchiyska and Vassilev’s (2024) research on pre-service foreign language teachers’ willingness to embrace new technology for their work at school. Although considering ethical issues and essential skills acquisition to use it appropriately, the overall readiness to implement AI in foreign language classes was declared. Therefore, the question seems to be no longer whether but rather how.

Briggs’ (2025) comparison of leading AI detector products such as Copyleaks, Unicheck, Quetext, and Strike plagiarism gives the opportunity to select appropriate one for the purpose needed. But more important are the pointed out advantages of AI detectors over the traditional plagiarism checkers. The way a typical AI detector works is described as going through several stages, i.e. the text being broken into small units (phrases) and further on matched with strings of words in the database. The matches, if any, are highlighted and their percentage is calculated. This process, as it was indicated, has its weaknesses while the leveraged AI detectors are of a higher level being capable of semantic analysis, natural language processing (related to writing styles and document changes), citation analysis, and pattern recognition. Moreover, some can recognize images, make code analyses and identify authors of text pieces (Briggs, 2025).

2. MATERIALS AND METHODS

As it was already pointed out the subject of the present paper leans on materials generated by AI as the procedure was depicted in our previous works (Ivanova et al, 2025, Ilieva et al. 2025 in print). The resulted stories were tested for plagiarism using only free versions of the indicated tools. The methodology followed included two stages, each containing three steps. The latter were as depicted below:

1. The story generated by ChatGPT 4o free was tested via ZeroGPT;
2. The originally generated by ChatGPT 4o free story and then edited by DeepL was tested via ZeroGPT;
3. The originally generated by ChatGPT 4o free story, edited by Grammarly, was tested via ZeroGPT;

The next stage included checking the same texts but this time via another application:

1. The story generated by ChatGPT 4o free was tested via QuillBot;
2. The originally generated by ChatGPT 4o free story and then edited by DeepL was tested via QuillBot;
3. The originally generated by ChatGPT 4o free story, edited by Grammarly, was tested via QuillBot.

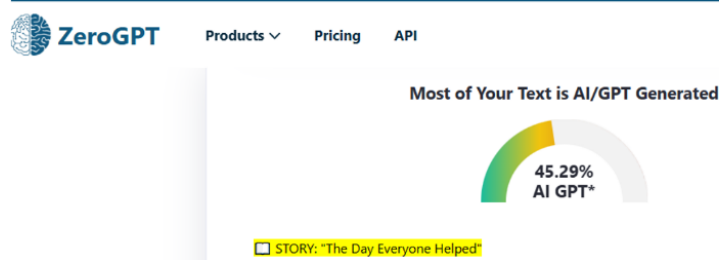
Each step of the testing was confirmed by figures showing the results given by the detective tool.

3. RESULTS AND DISCUSSION

Being acquainted with all the stages of the research, the outcome seemed somehow weird.

Results obtained with the help of ZeroGPT:

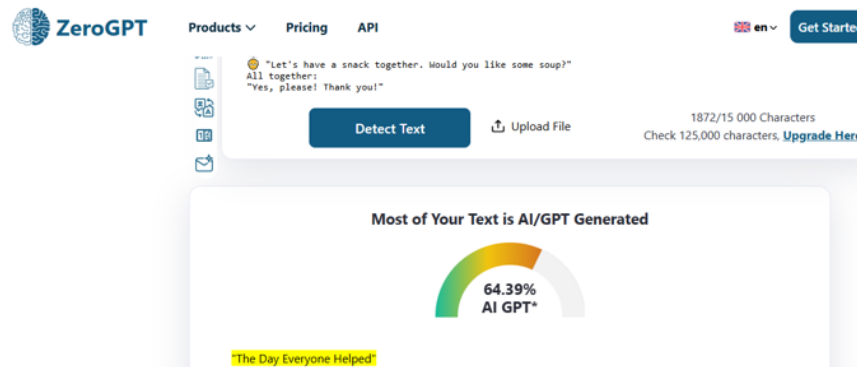
Figure 1. ZeroGPT about the story generated by ChatGPT 4o free



As we know the story was generated completely by ChatGPT 4o free, the percentage shown in figure 1. puzzled a lot. It was not clear whether the given by us initial directions were taken into consideration while calculations occurred. They might seem to be the single excuse for such inaccuracy.

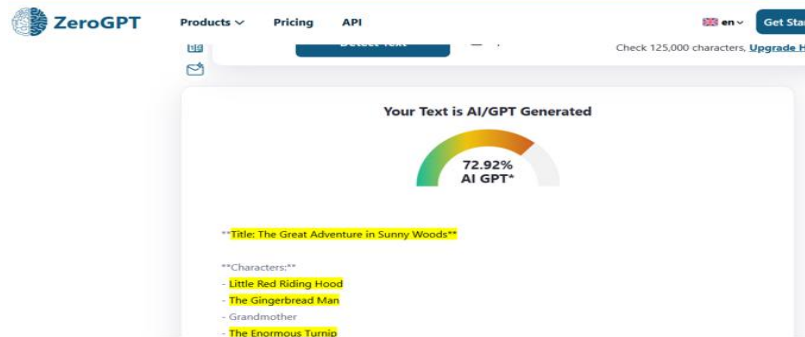
Figure 5 presented the calculations of DeepL's version.

Figure 2. ZeroGPT about the version of DeepL of the original story generated by ChatGPT 4o free



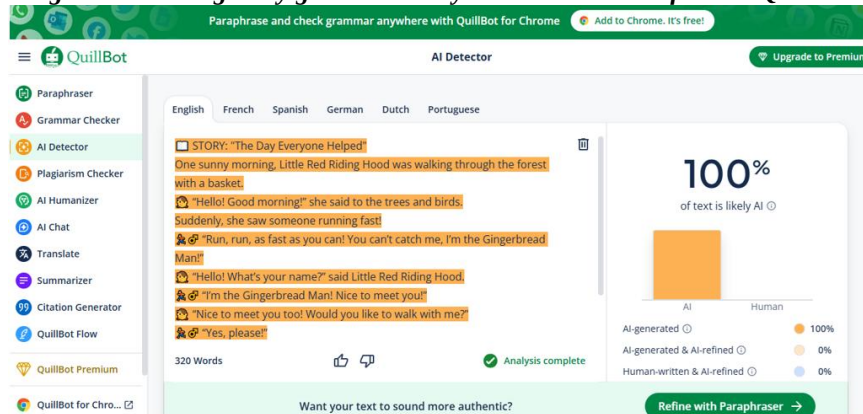
As the text had been transformed once again by another AI tool, the percentage recognized as generated explicitly increased.

Figure 3. ZeroGPT about the transformed story originally generated by ChatGPT 4o free by Grammarly



Grammarly version seemed to be recognized the easiest of all as generated by AI and the closest to the truth. The percentage given as 72,92% is inexplicable again, though.

Figure 4. The originally generated story with ChatGPT 4o free via QuillBot



At first glance the text in figure 4. was identical with the one in figure 5. In fact, the change of the text excluded the first part and edited only the further part of the story. The difference in the percentage, together with the different number of words counted in the two versions both confirm the stated above. DeepL version seemed to transform the

story into a more human-like composition as the number of per cents went down in comparison to the originally generated story by ChatGPT.

Figure 5. DeepL version of the generated story with ChatGPT 4o free via QuillBot

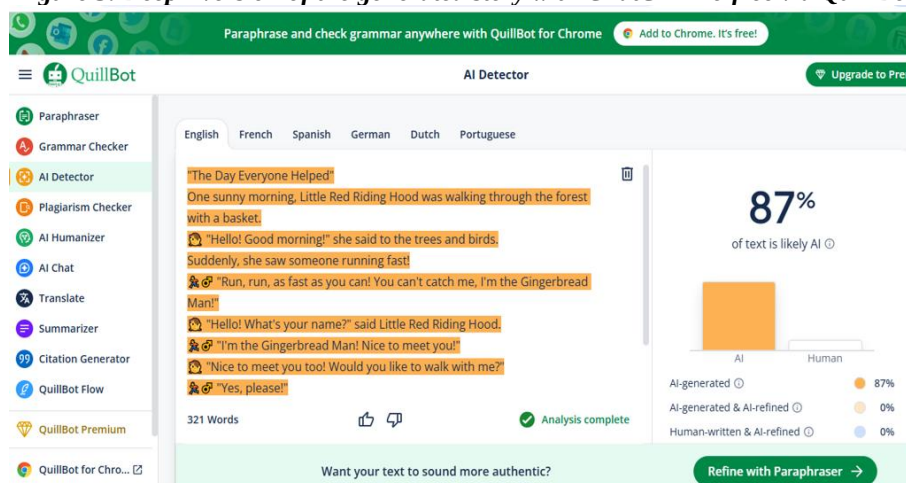
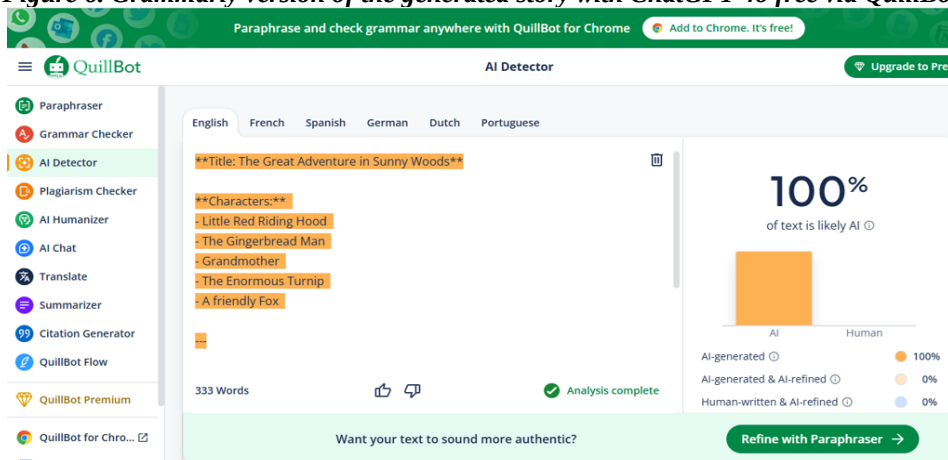


Figure 6. Grammarly version of the generated story with ChatGPT 4o free via QuillBot



This was the most realistic result because in the Grammarly version of ChatGPT story everything was created, or more precisely generated, by AI.

5. CONCLUSIONS

The results confirm what any AI assistant supplies when asked if AI detectors are completely reliable. Currently, they cannot guarantee the truthfulness and correct information they provide. It is rather the likelihood they account. In our small scale research comparing results by ZeroGPT and QuillBot we receive comparatively more realistic results by QuillBot, although in one out of three tests it is incorrect. As technology evolves more than rapidly, we express no doubt the things would change shortly. Hopefully, it will be in a positive direction. This opinion is supported by the works of researchers as Glushka (2023) and Shah et al. (2023). Although such findings are optimistic, we claim our study remains meaningful as it gives information on reliability of some of the applications at least so far. Since the topic in literature has not reached its extensive point on all the existing tools, we hope to have contributed for the role of two of them.

ACKNOWLEDGEMENTS

This work is part of the Gamification in the educational process Project, supported by Shumen University under Grant RD -08-122/ 06.02.2025.

REFERENCES

- Briggs, M. (2025). AI Detectors vs Traditional Plagiarism Checkers: What's the Difference?, Published January 23, 2025 Updated January 24, 2025. <https://tweakyourbiz.com/posts/ai-detectors-vs-traditional-plagiarism-checkers-whats-the-difference>
- Chaka, Ch. (2024). Reviewing the performance of AI detection tools in differentiating between AI-generated and human-written texts: A literature and integrative hybrid review. *Journal of Applied Learning & Teaching* 7(1):1-12.
- Edwards, B. (2023). Why AI detectors think the US Constitution was written by AI. *Ars Technica*. <https://arstechnica.com/information-technology/2023/07/why-ai-detectors-think-the-us-constitution-was-written-by-ai>
- Fowler, G.A. (2023). We tested a new ChatGPT-detector for teachers. It flagged an innocent student. The Washington Post. <https://www.washingtonpost.com/technology/2023/04/01/chatgpt-cheating-detection-turnitin>.
- Glushka, J. (2025). How To Check If Something Was Written with AI (ChatGPT), Gold Penguin. <https://goldpenguin.org/blog/check-for-ai-content/>
- Gotoman, J.E.J., Luna, H.L.T., Sangria, J.C.S., Santiago Jr, C.S., Barbuco, D.D. (2025). Accuracy and Reliability of AI-Generated Text Detection Tools: A Literature Review. *American Journal of IR 40 and Beyond* 4(1):1-9.
- Harakchiyska, Ts., Ts. Vassilev. (2024). Pre-Service Teachers' Perceptions of AI and Its Implementation in the Foreign (English) Language Classroom. Az-buki National Publishing House for Education and Science.04-10-2024. <https://doi.org/10.53656/str2024-5s-22-pre>
- Kopczynski, K. & Silvia, C. (2024). Use of AI in a World Language Classroom. *Creative Education*, Vol.15 No.12, December 2024. DOI: 10.4236/ce.2024.1512163 <https://www.scirp.org/journal/paperinformation?paperid=138662>
- Rocha, J., Soares, A., Honorato, M., Lima, L., Costa, N., Moreira, E. and Costa, E. (2015) Computers and Language Learning. *Creative Education*, 6, 1456-1465. doi: 10.4236/ce.2015.613146.
- Shah, A., Ranka, P, Dedhia, U., Prasad, S., Muni, S., Bhowmick, K. (2023). Detecting and Unmasking AI-Generated Texts through Explainable Artificial Intelligence using Stylistic Features. *(IJACSA) International Journal of Advanced Computer Science and Applications*, (14)10:1043-1053.
- Sofronieva, E., C. Beleva, G. Georgieva. (2025). Use of Artificial Intelligence in Foreign Language Teaching. Az-buki National Publishing House for Education and Science. 01.04. 2025. <https://doi.org/10.53656/ped2025-1s.03>
- Stockwell, G. (2024). ChatGPT in Language Teaching and Learning: Exploring the Road We're Travelling. *Technology in Language Teaching & Learning*, 6(1), 2273 <https://doi.org/10.29140/tltl.v6n1.2273>
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, Th., Guerrero-Dib, J., Popoola, O., Šigut, P., Waddington, L. (2023). Testing of detection tools for AI-generated Text. *International Journal for Educational Integrity*, 19:26.
- Xia, X. (2025). What to teach and how to teach: Artificial intelligence education in primary and secondary schools. *Science Insights Education Frontiers*, 28(1):4613-4626.