Schopnosť softvéru detekovať text generovaný umelou inteligenciou v študentských úlohách

The software's ability to detect AI-generated text in student assignments

Eva Rakovská[1]

Abstrakt
Ako zlepšiť učenie a výučbu pomocou AI je téma, ktorá učiteľov zamestnáva každý deň. Nástroje AI sa vyvíjajú veľmi rýchlo a ich používanie je bezplatné. Mnohé nástroje AI dokážu generovať študentské eseje, vedecké texty, štruktúrované životopisy, motivačné listy a iné typy textov. Môžeme nájsť nástroje AI na prevod textu vytvoreného AI na ľudský text alebo na parafrázovanie pôvodného textu. Preto je veľmi zložité povedať, ktorý nástroj AI je užitočný na získavanie vedomostí alebo je len nástrojom na podvádzanie. Zistiť, či je text vygenerovaný pomocou AI alebo je ľudským originálom, nie je jednoduché hoci na webových stránkach je veľa detektorov na zistenie AI obsahu. Niektoré štúdie spomínajú aj rozdiely v angličtine medzi cudzincami a prejavom rodených hovoriacich. Výsledkom je, že väčšina textov ľudí, ktorí sa naučili angličtinu neskôr je detekovaná ako vygenerovaná AI. Článok predstavuje vybrané detektory AI a používa ich na skúmanie schopnosti detekovať obsah AI v študentských úlohách a ich spoľahlivosť. Témy sme vybrali z dvoch rozličných predmetov a rozdielnych študijných programov. Výsledok nášho výskumu môže ukázať, či je užitočné používať detektor AI v každodennej rutine univerzitných učiteľov. Nakoniec sa snažíme nájsť rovnováhu medzi používaním detektorov AI a detekciou textov študentov človekom počas hodín angličtiny (rôzne zadania).

Kľúčové slová
Generatívna umelá inteligencia, AI detektor, AI vo vzdelávaní, akademická etika, spracovanie prirodzeného jazyka

Abstract
How to improve learning and teaching with AI is a topic that keeps teachers busy every day. AI tools are developed very quickly and are free to use. Many AI tools can generate students' essays, scientific texts, structured CVs, motivation letters, and other types of texts. We can find AI tools for converting AI text to human text or paraphrasing the original text. Therefore, it is very tricky to say which AI tool is useful for gaining knowledge or is only a tool for cheating. How to find out whether the text is AI-generated or human original seems to be not easy; however, there are many AI content detectors on websites. Some studies mention differences between native and non-native English speakers (Liang et al., 2023). The result is that most non-native English speakers' text is detected as AI-generated. The article introduces selected AI detectors and uses them by exploring their ability, and reliability to detect AI-generated content from students' assignments. We chose the topics from two distinguished subjects and different study programs. The result of our research can show whether it is useful to use an AI detector in the daily routine of university teachers. Finally, we try to find the balance between using AI detectors and human-detecting students' texts during English lessons (various assignments).

[1] Bratislava University of Economics and Business, Faculty of Economic Informatics, Department of Applied Informatics, Dolnozemská cesta 1, 852 35 Bratislava, eva.rakovska@euba.sk.

Key words
Generative Artificial Intelligence, AI detector, AI in education, Academic ethics, Natural Language processing

JEL classification
I23, O39

1 Introduction

The rapid development of Artificial Intelligence (AI) brings many challenges for people dealing with AI. In many fields, the question of when and where to use AI is often addressed, so that it is not only useful, but also safe.

The concept of AI started with Alan Turing's idea of a "thinking machine" or "intelligent machine". His imitation game called the Turing test (Muggleton, 2014) started the activity of mathematicians towards the initiative directed to the basic concepts of AI. John McCarthy invited more mathematicians, electrical engineers and psychologists to the Dartmouth Summer Research Project on Artificial Intelligence in 1956 where AI's first concepts were discussed (e.g. symbolic methods, early expert systems, and deductive systems versus inductive systems based on first-order predictive logic). The workshop lasted six weeks and laid the foundations of artificial intelligence. The following years were full of enthusiasm for AI scientists and the idea to solve all problems using AI methods was the central motive of the research. After 20 years Dartmouth Workshop started the gold era of AI in two main directions. The first direction was oriented toward knowledge representation and the development of Expert Systems. Marvin Minsky found the knowledge frame representation in 1975, so object-oriented programming started. Object-oriented modelling and programming are still essential in the development of AI. The second direction started the era of artificial neural networks (NNs) with the idea of an artificial neuron called Perceptron (Rosenblatt, Minsky, Pappert) in 1969. It led to massive research in artificial neural networks. AI development is not only directly increasing, the time known as an "AI winter" was 40 years ago. (Hurwitz, Kaufmann, Bowles, 2015).

Neural networks are perceived as the main concept of AI in current times. Many people perceive NNs as self-learning tools, which can learn the same way as people. Therefore they are perceived as tools with a "magic power". However the NNs are very important in the AI field, they represent only a subset of AI methods. An increased interest in NNs might have started after the year 2000 with the rapid technological progress. The importance of NNs is visible mainly by recognizing the patterns and hidden relations within huge amounts of data.

Although the statistical methods were successfully used for text and language recognition for a long time, the NNs surpassed many previous approaches and techniques in this field. Therefore the NNs are one of the main concepts behind the Natural Language Processing (NLP). As Hurwitz writes "NLP is the capability of a computer system to process text written or recorded in a language used for human communication" (Hurwitz et al., 2015) and deals with unstructured data hidden in the text documents and multimedia. The NNs help find the similarities within the unstructured data and then process the syntax and semantics of language by using grammar systems. The process of recognizing the context of language is not easy and the developers of NLP systems use various combinations of techniques (Aßenmacher, 2020). The NLP boom started after the launch of chatbots searching for answers in web data where the ChatGPT, Bard (now Gemini), and Bing played crucial roles in further research. Generative AI used in the mentioned chatbots, is just at the peak of its development and young people have quickly adopted it. Although it can make life easier in some activities, it brings new problems in other activities.

The problems associated with the ethics of student thesis, essay and project writing are becoming more complicated. The question of when and how to use generative AI to create text is becoming increasingly relevant. Students can use AI tools, but sometimes they lose the ability to communicate, the capacity to make rational or logical decisions, the ability to learn in context, and the skill to learn foreign languages. They rely on mobile and web applications and the sentence: "Google knows everything" has changed to "ChatGPT tells everything". The main objective of the article comes from the KEGA project "ePortfolio as Pedagogy Facilitating Integrative Learning" and the goal of the project is to find out to what extent an e-portfolio will affect students' writing. The students create an e-portfolio and develop digital skills. Therefore, the idea of research is to explore the impact of generative AI on the creation of students' portfolios and whether it is possible to improve their language skills using chatbots.

We experimented with the students studying the English language on various levels. We obtained a set of students` assignments from two universities. The set contains 53 motivation letters (cover letters). Creating motivation letters is very easy for ChatGPT and other chatbots based on Large Language Models (Kasneci, 2023). Therefore, we try to find out whether the students were cheating or whether they used AI as a helpful tool only. Finally, we consider, whether it was possible to say they have completed the assignment successfully. The research started by selecting three AI detectors which are free and then we compare their results. The accuracy of AI detectors is presented in many web articles starting with Forbes ranking (Clark, 2023). Although these evaluations start with commercial software that has a free version, we decided to use none of them. The reason is given toward the end of the article. The main objective of the research is to compare the results from AI detectors, suggest some steps for detecting cheating, and encourage students to use the generative AI correctly and meaningfully.

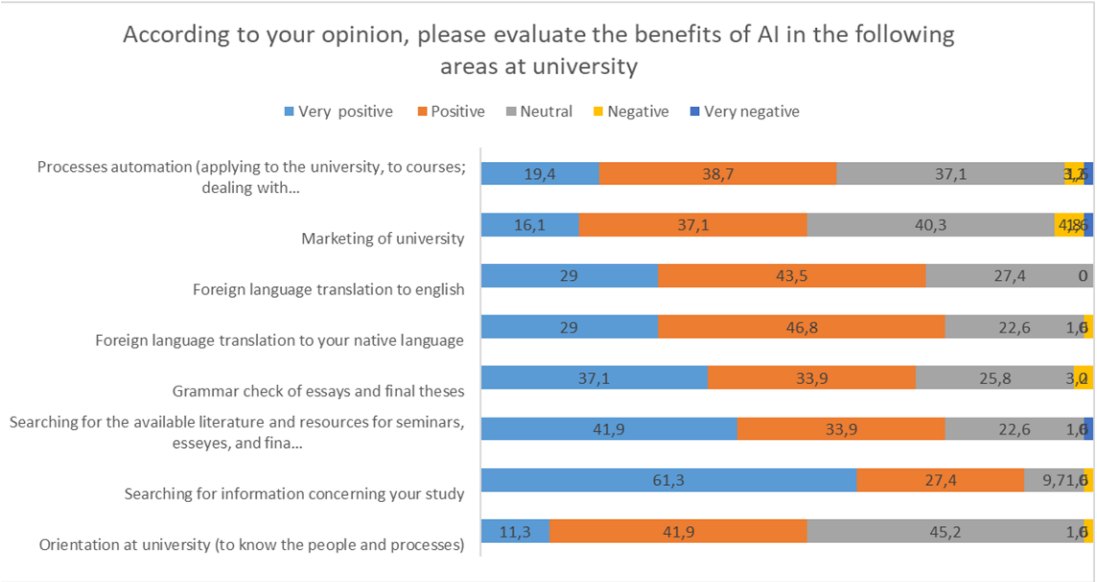
2 Perceiving AI in an academic environment

The Academic environment is excellent for testing new technologies and exploring whether young people understand the role of new technologies, not only in gaming, watching movies, social networks and leisure activities. Young people use AI technologies naturally and intensively, not considering what is behind them. Many times, students in the third year of their study at university do not have a clear idea about cybersecurity or ethical issues concerning AI. Therefore, we started the survey concerning using AI in the university environment in the academic year 2023/24. We continued with a more precise survey in the 2025. A new survey asked more questions and we asked students with higher granularity of possible answers. We distributed the questionnaire to more than 250 students at Bratislava University of Economics and Business (EUBA) and got back the answers from 43 respondents. The survey contains 32 questions but for our actual research, we used a selected part of the survey. These selected questions are focused on student learning and the benefits of AI.

The first section of the survey investigates whether AI has some benefits for students in practical issues (orientation at university, processes, etc.) and whether AI is helpful for learning (mainly writing essays, thesis, seminar works, etc.). The summary of the first section is in Fig. 1.

Fig. 1 shows that using generative AI such as ChatGPT or translators plays a key role in searching information (87,5%), searching resources (75,8%), checking grammar in written assignments (71%), and translation (more than 70%). Further, we asked students whether they knew some AI technologies. Surprising was that the Large Language Model or Natural Language processing is not very well-known despite the students use it often.

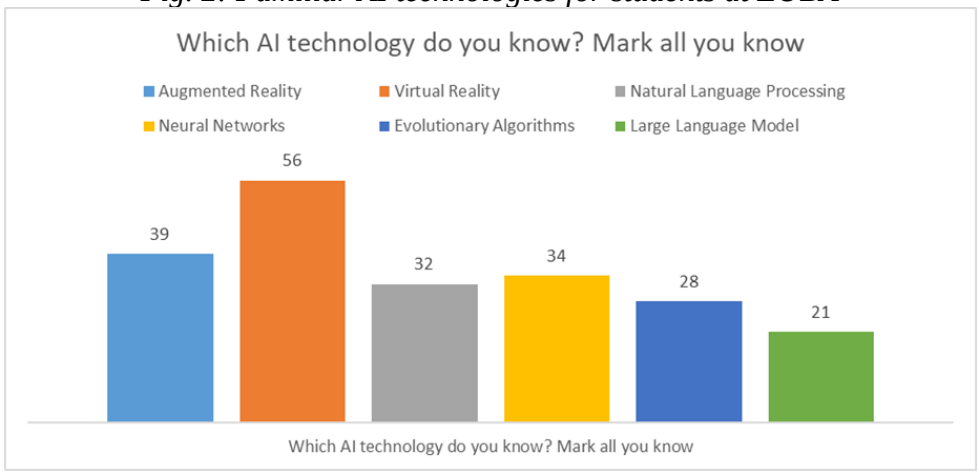
Fig. 1: Summary of the first section of the survey at EUBA - benefits of AI in the university



Source: Authors' own processing

Many students studied Information management, so we assumed they knew more about these technologies. Fig. 2 shows that virtual reality and augmented reality are popular. This result is thanks to playing computer games and using the cameras on mobile phones.

Fig. 2: Familiar AI technologies for students at EUBA

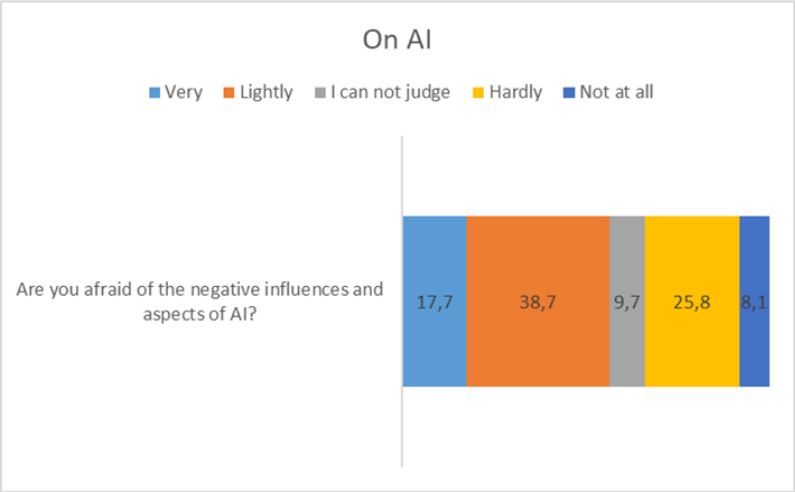


Source: Authors' own processing

The survey contained a special question regarding AI concerns or whether students fear the negative influences and aspects (Fig. 3). More than half of the students worry about AI's negative aspects. So, the discrepancy between perceptions of the benefits of AI and concerns about AI was the first stimulus to investigate whether students also practically respect AI ethics.

Considerable discrepancy in the perception of AI among students lead us to explore how intensive they use AI for writing assignments. We focus on the English lessons and some assignment in which is easy to use ChatGPT. Even though it is possible to use AI detectors for the Slovak language, detectors have a problem distinguishing Slovak grammar correctly.

Fig. 3: Survey question – students` concerns about AI at EUBA



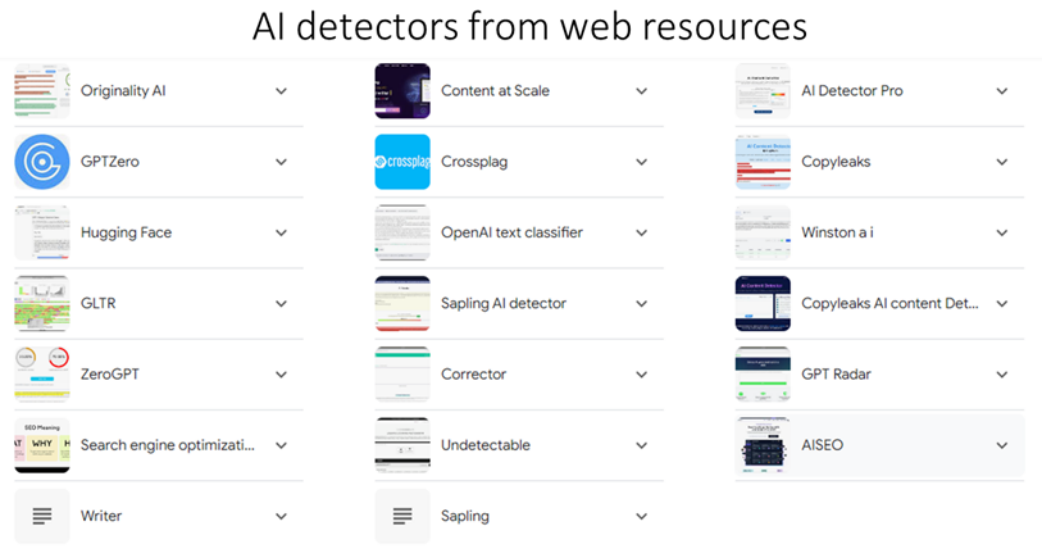
Source: Authors' own processing

3 AI detectors as useful tools in Learning

The history of generative AI started only around the 2020s. As (Toosi et al, 2022) writes it is "Thanks to the convergence of parallel processing, higher memory capacity, and more massive data collection (e.g., big data)" and "Computers are becoming faster. Computing power has continued to double nearly every two years (Moore's law)." Therefore, AI became very interesting for technology investors and the area of testing various technologies began to expand. LLMs such as ChatGPT 3 and then ChatGPT 4 or Bard, Gemini were trained on larger sets of human texts, therefore they began to have a better ability to predict the syntax of different sentences in the natural languages. AI detectors have evolved hand in hand with generative pre-trained AI and their self-learning models are getting more perfect. They use a large knowledge base consisting of human and AI-generated texts. It is difficult to say which AI detector is the best one. Fig. 4 shows the ranking of AI detectors from various web resources in 2024. The Google engine offered this summary to users after searching "AI detectors ranking" (Fig. 4).

The main feature of the AI detector is the accuracy. The principles of the AI detectors are based on machine-learning pre-trained algorithms, and statistical methods, and work with probability. The probability is the factor that makes it impossible to say that any AI detector evaluates text with 100% accuracy. Two main principles applied in AI detectors are important: perplexity and burstiness. Characteristics of these principles are well-described in (Leechuy, 2023): "Perplexity: The unpredictability of the content. AI-generated text tends to have low perplexity, while human writing has higher perplexity. Burstiness: Variation in the length and structure of the sentences. AI content tends to be more steady and have lower burstiness than human writing." Author Leechuy lists multiple features typical of AI-generated text (repetition of the words, repetitive syntax of sentence, unnatural words, generic or impersonal tone, inconsistent verb tense, etc.). Although AI detectors are strong tools, they might fail still in some cases. (Leechuy, 2023).

Fig. 4: AI detectors from web resources



Source: Google 2024

We started our research with a list of AI detectors and their ranking evaluation on various websites. It seemed that AI detectors, which are free of charge, met the requirements best for our research. No regular payment and accuracy over 60% were sufficient, therefore we used the ranking from (Driessen, 2024) "Best AI Detector | Free & Premium Tools Compared". The two best-ranking and free versions were with the 78% accuracy Quillbot (Quillbot, n.d.) and Scribbr (Scribbr, n.d.) and ZeroGPT (ZeroGPT, 2025) with 64% accuracy (Table 1). We tested a few free versions of paid software, but although their paid versions are high-ranking, none recognized AI-generated text correctly (AI-generated covering letter). Or the user interface was not as intuitive as for our chosen detectors.

As previously mentioned the student assignments are included in the project "ePortfolio as Pedagogy Facilitating Integrative Learning". We tried to choose an assignment that could be completed at different English language levels and AI could easily generate that.

Tab. 1: Comparing the chosen AI detectors

	Accuracy	False positive	free	Star ranking
Quillbot	78%	0	yes	3,9 *
Scribbr (free)	78%	0	yes	3,9 *
ZeroGPT	64%	1	yes	3,1 *

Source: Authors' own processing

The project referred to above is implemented in two universities: Trnava University in Trnava and the Bratislava University of Economics and Business (EUBA). We selected students from both universities. Thirty-three students of Trnava University in Trnava study the English language in the 3rd year of their first degree (Bachelor`s degree). Their study programs are:

- Teacher Training of the English Language and Literature in the study field: Teacher Training and Educational Sciences
- English Language and Anglophone Cultures, study field: Philology

Students from the EUBA study at the Faculty of Economic Informatics, the 1st degree of study. The assumption was that English language students would not cheat and would use AI-generated text as little as possible. On the other hand, we supposed that students of EUBA

would use more because they know better the various software. The assignment for both groups asked students to write a cover letter for a job position.

Tab. 2: Results of AI detectors for Trnava University in Trnava

Student ID		***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***
QuilltBot	%AI generated	43	79	19	65	0	100	100	76	0	29	22
Scribbr	%AI generated	100	48	23	24	6	100	100	1	13	25	1
ZeroGPT	%AI generated	13	72	14	23	0	100	87	8	32	37	32
	Trnava University in Trnava group											
Student ID		***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***	***21***	***22***
QuilltBot	%AI generated	80	15	26	25	25	24	0	78	22	28	36
Scribbr	%AI generated	25	2	19	25	0	0	5	76	21	16	16
ZeroGPT	%AI generated	20	21	22	49	59	0	0	54	16	48	35
	Trnava University in Trnava group											
Student ID		***23***	***24***	***25***	***26***	***27***	***28***	***29***	***30***	***31***	***32***	***33***
QuilltBot	%AI generated	68	0	0	22	100	44	69	100	23	0	100
Scribbr	%AI generated	100	7	0	25	100	24	100	88	15	0	100
ZeroGPT	%AI generated	98	9	3	31	100	88	33	45	0	38	100

Source: Authors' own processing

People often use an AI generator to write cover letters. It is more comfortable and the AI generator helps people formulate better ideas without grammar mistakes. Table 2 shows the results of 33 students` cover letters studying the English language at Trnava University. Table 3 shows the results of 21 students` cover letters studying Economic Informatics at the EUBA.

Tab. 3: Results of AI detectors for EUBA

Student ID		***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***
QuilltBot	%AI generated	30	52	100	66	0	100	100	100	100	52	31
Scribbr	%AI generated	34	60	100	100	0	100	100	100	100	60	37
ZeroGPT	%AI generated	13	77	95	84	0	99	100	100	100	9	0
	Bratislava University of Economics and Business											
Student ID		***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***	***21***	
QuilltBot	%AI generated	100	0	0	0	62	30	68	56	78	0	
Scribbr	%AI generated	100	26	6	34	1	32	100	22	81	18	
ZeroGPT	%AI generated	29	25	6	16	0	16	100	0	65	0	

Source: Authors' own processing

3.1 Summary of the experiment

We had two statements before the experiment. First, the students studying the English language and Literature as their main subject, are not cheating. They use AI-generated text on a minimal level.

The results showed that four students (cca 12%) submitted fully AI-generated text; eight (cca 24%) students had a very high percentage of AI text at least from one AI detector (over

75%); it seems that other students used AI text generators less or more in a rational way (e.g., for finding some phrases, better words, grammar syntax, etc.)

The second statement was that the students of Economic Informatics use the AI-generated text often. This statement has been confirmed. Seven students` texts are detected as AI-generated (one-third of the number). One letter of cover was completely human-written (all three detectors evaluate text as 0% AI-generated).

The objective of the experiment was to explore whether it is possible to say that using an AI detector is reliable and useful for teachers. An important observation in research was also the quality of the student's text. How many and which type of grammar mistakes are within the text? It is a very tricky issue; therefore, we used the software Grammarly and did not count the number of mistakes exactly. The advantage of Grammarly is that mistakes in the text are sorted by colour. We used heuristics and visually compared the length of text (the number of words or characters visible in the AI detectors) with the number of mistakes. So, we had categories: no mistakes; number of mistakes from 1 to 5; a few mistakes; many; too many mistakes; and mistakes only in the sentence structure.

The experiment verifies that the AI-generated text usually contains a few syntax errors (incorrect sentence structure). It has never been with zero mistakes in sentence structure. The text marked as human-written with a low percentage usually contained many mistakes (students of EUBA) or the one (totally 0% evaluated) contained only 5 mistakes.

Comparing the three chosen AI detectors, we can say that they gave us more or less similar results in most cases. In certain instances, the results were unexpected and confusing, e.g., 1st case: Quillbot detection was 76%, Scribbr 1%, ZeroGPT 8%; 2nd case: Quillbot 62%, Scribbr 1%, ZeroGPT 0%. These discrepancies depend on the algorithms used in the detector and the user's language level. The mix of sentence structure (e.g., Slovak vs. English sentence structure) and incorrect grammar syntax influenced the reliability of AI detectors. It is not easy to say which of the AI detectors is better or the best, but for experienced teachers, it is valuable to check the students' assignments with AI detectors. Especially, when there are similarities in the set of assignments (two cover letters started with the same sentences); the letter has no mistakes; the letter has only a few syntactic errors, etc. When the teachers have large student groups it is impossible to work individually with the students, to compare speaking abilities with written text (EUBA). The rules of many universities require using the anti-plagiarism system for all assignments. Practically, it is impossible because the topics of assignments are repetitive every year (maybe every second year). Therefore, using AI detectors as supporting tools for student assessment is better and can have several benefits:

- a more fairly assessment of students,
- greater student motivation to handle the assignments independently,
- saving the teacher`s time,
- a change in the routinely done assessment by scoring only,

and maybe more. Although, it is not possible to judge completely the quality of AI detectors, the result of our investigation is that students need to be taught how to use generative AI properly. It seems that 25 to 40% of text in assignments is useful and can help the students improve their language skills (e.g., the student the student uses prepositions incorrectly).

4 Conclusion

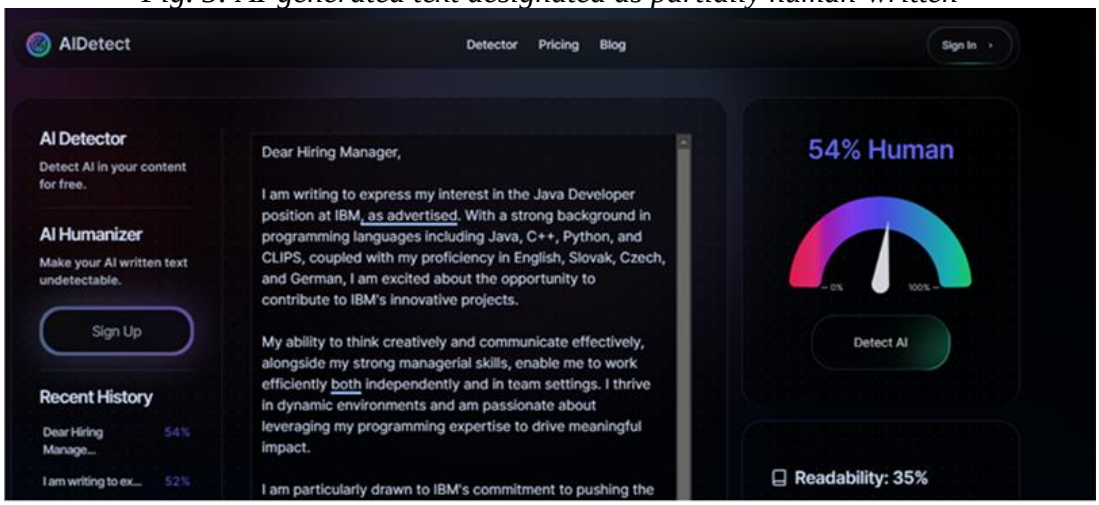
The software tools of generative AI are very powerful, but we cannot rely on them. Their Large Linguistic Models strictly work by using syntax and grammar rules and do not follow the emotional aspects of language. A good example of LLMs is the software Grammarly, controlling grammar errors. Grammarly works with a large amount of data from a text corpus and has a self-learning cycle that follows the users' behaviour; therefore, the model changes in

every cycle and accordingly changes the result. (Grammarly, 2019) That is the reason why the software is changing from the point of view of the user's observation. LLMs include the key principle of Machine learning. The model develops itself. (Aßenmacher, 2020)

AI detectors use similar principles but focus more on statistics and probability. They follow the complexity of sentences, sentence structure, repetitive words, matter-of-fact writing, formal writing, etc. Based on these attributes, the AI detector guesses whether the text is AI-generated or not (Leechuy, 2023). AI detector's reliability is expressed by false positivity (human-written text is detected as AI-generated), and the ranking usually also shows false positivity.

On the other hand, an AI detector can evaluate the AI-generated text that is partially or totally human-written (Figure 5). We also tested false positivity in the process of choosing suitable AI detectors. Our experiment was only the first step in investigating. We plan to explore more student assignments with other content (essays, students' seminar theses from cyber-security, IT, programming, etc.) or in the Slovak language. However, using the AI generative software seems to be easy, we can see in real life many mistakes (maybe hoaxes) produced by using them. Many various software robots operate in the web space and search for something interesting, new, short and bombastic for readers. These bots are products of generative AI, and the common user is not able to notice that the author of the text is not a human.

Fig. 5: AI-generated text designated as partially human-written



Source: (AIdetect, 2024)

The students use webpages for writing seminars, research papers, essays, theses, and so on. They often use the web pages created by robots such as resources for their work without verifying the facts. As our research shows, students of Informatics rely more on AI tools without considering the consequences. This situation is contradictory to the result of the survey question "Are you afraid of using AI?" Many students "are afraid," but they do not understand that AI depends on humans. A human is the core of AI wisdom. AI is based on well and correctly done algorithms and proper usage, and can learn from people. Whatever we put into the AI system starts inference and copies our ideas. That is why it is necessary to start talking and teaching more students what artificial intelligence actually is and how to deal with it.

Resources

1. AIdetect, (2024). Detect AI, Anywhere. https://www.aidetect.com [Accessed 05 May 2025]
2. Assenmacher, M. et al., (2020, September 8). Modern approaches in Natural Language Processing. Modern Approaches in Natural Language Processing. https://slds-lmu.github.io/seminar_nlp_ss20/ [Accessed 05 March 2025]
3. Clark, E., (2023) The 10 Greatest AI Content Detector Tools. https://www.forbes.com/sites/technology/article/best-ai-content-detector-tools/?sh=17a1ff625d5e [Accessed 25 May 2024]
4. Driessen, K., (2024, November 18). Best AI Detector | Free & Premium Tools Compared. https://www.scribbr.com/ai-tools/best-ai-detector/ [Accessed 25 Januar 2025]
5. Google. 2024. Searching: AI detectors ranking. [online] Available at <https://www.google.com/ > [Accessed 25 May 2025]
6. Grammarly. (2019, May 17). How We Use AI to Enhance Your Writing | Grammarly Spotlight. https://www.grammarly.com/blog/how-grammarly-uses-ai/ [Accessed 25 May 2025]
7. Hurwitz, J., Kaufman, M., & Bowles, A. (2015). Cognitive computing and big data analytics. John Wiley & Sons.
8. Leechuy, J., (2024, March 14). How Reliable Are AI Detectors? Claims vs. Reality. https://www.theblogsmith.com/blog/how-reliable-are-ai-detectors/ [Accessed 25 May 2025]
9. Kasneci, E., et al., (2023). Chatgpt for good? On opportunities and challenges of large language models for Education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274 [Accessed 2 May 2025]
10. Muggleton, S. (2014, February 1). Alan Turing and the development of Artificial Intelligence. AI Communications, 27(1), 3–10. https://doi.org/10.3233/aic-130579 [Accessed 2 November 2025]
11. QuillBot, a Learneo, Inc. business (n.d.). Unlock your potential with QuillBot's creative suite. https://quillbot.com/ [Accessed 25 November 2025]
12. Scribbr. (n.d.). Your path to academic success, https://www.scribbr.com [Accessed 25 November 2025]
13. Toosi, A., Bottino, A. G., Saboury, B., Siegel, E., & Rahmim, A. (2021). A brief history of AI: How to prevent another winter (a critical review). PET Clinics, 16(4), 449–469. https://doi.org/10.1016/j.cpet.2021.07.001 [Accessed 25 May 2025]
14. ZeroGPT. (2025). Trusted GPT-5, ChatGPT and AI Detector tool by ZeroGPT, https://www.zerogpt.com/ [Accessed 25 November 2025]