

University "St. Kliment Ohridski"
Faculty of Education



TEACHER

International Journal of Education

ISSN 1857-8888



vol. 26

November 2023

University “St. Kliment Ohridski“ – Bitola
Faculty of Education – Bitola

TEACHER
International Journal of Education

Vol. 26
Bitola, November, 2023

Publisher

Faculty of Education – Bitola

Executive and Editor-in-chief

Ljupco Kevereski, PhD, University St. Kliment Ohridski, Republic of North Macedonia

Editorial Team

Ljupco Kevereski, PhD, University St. Kliment Ohridski, Republic of North Macedonia, ljupco.kevereski@uklo.edu.mk (Editor in-chief)

Jove Dimitrija Talevski, PhD, University St. Kliment Ohridski, Republic of North Macedonia, jove.talevski@uklo.edu.mk

Dance Sivakova-Neshkovski, PhD, University St. Kliment Ohridski, Republic of North Macedonia, dance.sivakova@uklo.edu.mk

Violeta Janusheva, PhD, University St. Kliment Ohridski, Republic of North Macedonia, violeta.januseva@uklo.edu.mk

Jasminka Kochoska, PhD, University St. Kliment Ohridski, Republic of North Macedonia, jasminka.kochoska@uklo.edu.mk

Josif Petrovski, PhD, University St. Kliment Ohridski, Republic of North Macedonia, josif.petrovski@uklo.edu.mk

Editorial Board

Grozdanka Gojkov, Higher Professional School for Education of Teachers „Mihailo Palov“ Serbia, g_gojkov@mts.rs

Marjan Blažič, PhD, Faculty of Health Sciences, University of Novo Mesto, Slovenia, marjan.blazic@uni-nm.si

Ante Kolak, PhD, University of Zagreb, Croatia, akolak@ffzg.unizg.hr

Svetlana Kurtesh, PhD, Faculty of Arts and Humanities, University of Madeira, Portugal

Danimir Mandić, PhD, University of Belgrade, Serbia, danimir.mandic@uf.bg.ac.rs

Danijela Kostadinović, PhD, University of Nish, Serbia, danijela.kostadinovic@filfak.ni.ac.rs

Jasmina Starc, PhD, University of Novo Mesto, Slovenia, jasmina.starc@uni-nm.si

Mojca Jurishevich, PhD, University of Ljubljana, Slovenia, mojca.jurisevic@pef.uni-lj.si

Anton Ilica, PhD, University „Aurel Vlaicu“, Romania, ilica_anton@yahoo.com

Eva Szórádová, PhD, Pedagogical faculty, Constantine the Philosopher University in Nitra, Slovakia, eszoradova@ukf.sk

Lazar Stoshić, PhD, University UNION Nikola Tesla, Serbia, lazarstosic@yahoo.com

Alla Belousova, PhD, Don State Technical University, Russia, belousovaak@gmail.com

Irina Abakumova, PhD, Don State Technical University, Russia, abakira@mail.ru

Tom Jovanovski, PhD, USA

Zlatko Zhoglev, PhD, University St. Kliment Ohridski, Republic of North Macedonia, zlatko.zhoglev@uklo.edu.mk

Valentina Gulevska, PhD, University St. Kliment Ohridski, Republic of North Macedonia, valentina.gulevska@uklo.edu.mk

Biljana Gramatkovski, PhD, University St. Kliment Ohridski, Republic of North Macedonia, biljana.gramatkovski@uklo.edu.mk

Milena Pejchinovska-Stojković, PhD, University St. Kliment Ohridski, Republic of North Macedonia, milena.pejcinovska@uklo.edu.mk

Nikoleta Momčilović, Faculty of Philosophy, University of Belgrade, Serbia, nikoleta.momcilovic@filfak.ni.ac.rs

Namita Subiotto, Faculty of Arts, University of Ljubljana, Slovenia,
namita.subiotto@ff.uni-lj.si

English Language Editor

Stela Bosilkovska, MA, University St. Kliment Ohridski, Republic of North Macedonia,
stela.bosilkovska@uklo.edu.mk

Technical Editor

Josif Petrovski, PhD, University St. Kliment Ohridski, Republic of North Macedonia,
josif.petrovski@uklo.edu.mk

CIP - Cataloging in Publication,
National and University Library "St. Kliment Ohridski" – Skopje.

TEACHER: Journal of the Faculty of Education – Bitola /
[Editorial Board Acad. Grozdanka Gojkov ...] Year XXI, No. 2 (2023) -
- Bitola: Faculty of Education, 2023 -. - 29 cm., 65 p.

Unspecified

ISSN 1857- 8888 (online)

University “St. Kliment Ohridski“ – Bitola, North Macedonia
Faculty of Education – Bitola, North Macedonia

Address:

Faculty of Education
ul “Vasko Karangelevski“ b. b.
7000 Bitola, North Macedonia
Tel/Fax. ++ 389 47 253 652; 203 385

With the opinion of the Ministry of Culture no. 07-2699/2 from 15.04.1998, for the journal “Teacher” is paid preferential tax rate. In accordance with Article 20, paragraph 8 of the VAT Law (Official Gazette 44/99), for the journal “Teacher” is paid a tax of 5%.

The journal has no commercial nature.

Table of contents

BRITISH ENGLISH VERSUS AMERICAN ENGLISH PREFERENCE BY UNIVERSITY STUDENTS OF EFL	6
Ana Koceva, Dafina Kostadinova, Tanya Tabutova	
INVESTIGATING L2-L1 TRANSFER AMONG BILINGUAL STUDENTS IN THE REPUBLIC OF NORTH MACEDONIA	11
Stojanoska Marija, Shalevska Elena, Talevska Marija	
AI LANGUAGE MODELS, STANDARDIZED TESTS, AND ACADEMIC INTEGRITY: A CHAT (GPT)	17
Elena Shalevska	
EXPLORING EFL TEACHERS' PERCEPTIONS: INTEGRATING CHILDREN'S LITERATURE IN THE LANGUAGE CLASSROOM.....	26
Dushko Talevski, Elena Shalevska	
JUXTAPOSITION OF THE SOUND SYSTEMS OF THE ENGLISH AND MACEDONIAN AND THE POSSIBILITIES FOR INTERFERENCE.....	31
Stela Bosilkovska, Violeta Janusheva	
THE IMPERATIVE FOR OCCUPATIONAL THERAPY IN CHILDREN WITH LEARNING DISABILITIES – A LITERATURE REVIEW	41
Denis Arsovski, Tanja Jovanovska, Viktorija Prodanovska-Stojchevska, Izabela Filov, Domnika Rajchanovska, Angelka Jankulovska, Daniela Petkovska	
THE IMPORTANCE OF THE TEACHER IN ENCOURAGING CREATIVITY AMONG THE GIFTED STUDENTS IN PRIMARY EDUCATION.....	46
Vesna Stojanovska	
PANNONIC LEGENDS – ATTRIBUTION	49
Anita Angelevska	
COMPARISON OF SPEAKERS LIZ TRUSS AND BORIS JOHNSON	53
Sofija Trampevska, Biljana Gramatkovski	
USING STEM EDUCATION IN SCHOOLS	62
Josif Petrovski	

AI LANGUAGE MODELS, STANDARDIZED TESTS, AND ACADEMIC INTEGRITY: A CHAT (GPT)

Elena Shalevska

University "St. Kliment Ohridski" – Bitola, North Macedonia
elena.shalevska@uklo.edu.mk

Abstract

Language models' popularity is on the rise, and with that, concerns about academic integrity in the times of such advanced Artificial intelligence (AI) tools are on the rise, too. Considering such concerns, this small study, which employs both qualitative and quantitative methods, thoroughly examines the role of language models, particularly ChatGPT, in the context of academic integrity. By assessing the accuracy of test answers generated by said language model, on questions from the state-issued high-school graduation English exam in N. Macedonia, and analyzing parts of essays generated using various prompts, the study aims to explore the potential implications of such AI tools on academic integrity in this new tech era.

The study shows that ChatGPT's accuracy in providing test answers is satisfactory, with a minimal number of mistakes and over 80% accuracy on average, on both tests! As for the text/parts of essays generated by the model, the study has shown that the quality of the generated text differed based on the prompts that the user provided and their proficiency in articulating their specific demands. The study also showed that current AI detection remains unreliable at best.

These findings contribute to the ongoing discourse on AI's influence on education and academic integrity, especially in regard to ChatGPT's capabilities to generate content that can pass standardized tests and excel in open-ended writing tasks.

Keywords: Language models, ChatGPT, academic integrity, standardized tests, ESL.

1. Introduction: Language Models and Academic Integrity

With the recent introduction of sophisticated language models like ChatGPT (Conversational Generative Pre-training Transformer), the field of artificial intelligence (AI) has seen somewhat of a revolution. These models, primarily based on the GPT-3.5 and GPT-4 architecture created by OpenAI, have notably changed how people interact with technology. They have given us a glimpse of the future and the language an AI can produce and have shown us how these language models, unlike anything we have seen thus far, can be coherent, contextually relevant, and most of all, similar to human writing. ChatGPT, InstructGPT's sibling model, was made to carry out instructions and offer detailed responses to various prompts. Using deep learning and unsupervised learning techniques, as well as training the model on large text datasets, before continuing with task-specific fine-tuning, OpenAI has created the most sophisticated, free AI chatbot to date. Trained to understand and generate human-like text across multiple languages, ChatGPT does question answering and sentence completion. Its capabilities have taken the world by storm and according to Mollman (2022), the model attracted over a million subscribers in just five days after going live.

2. Broader research context

As technology and AI have evolved, so have the ethical and academic concerns about the use of language models like ChatGPT. Academic integrity, a fundamental principle in education and research, encompasses the values of honesty, trust, fairness, respect, and responsibility. (Fishman, 2014). It emphasizes the importance of proper attribution, accurate citing, objectivity, and originality in education, in general. Academic integrity is something that is and should be highlighted in all levels of education, especially in high schools, as students are old enough to understand these principles of ethical conduct. Citing several authors working in the field, in her literature review, Holden et al. (2021) define academic dishonesty (sometimes referred to as cheating) as the use of unauthorized materials, helping others cheat, misrepresentation of self, and ultimately, all kinds of plagiarism. These definitions clearly show why academic integrity has become a critical concern with

the widespread use of ChatGPT and similar language models. The ease with which the models can generate text that is indistinguishable from authentic student work is worrisome, to say the least.

Some authors, like Susnjak (2022) even go as far as to claim that such language models may forever end the, albeit fragile, online exam integrity. Susnjak study has found that ChatGPT poses a potential risk to the integrity of online exams, particularly in tertiary education where these kinds of exams are becoming more common. The language model, in his view, exhibits critical thinking skills on par with numerous students, and, more worryingly, can produce very realistic, human-like text with little input and prompting. What Susnjak suggests to aid this issue is the implementation of oral exams, vigorous academic integrity measures, as well as the use of AI detectors. He cautions, though, that these AI-text detectors, although useful, as of now, are still not a foolproof way to detect academic dishonesty.

Though exams that are fully online seem to be the ones facing major challenges, other non-online assignments can be subject to ChatGPT-enabled cheating, as well. Take, for example, essay writing. Having the students write an essay on a book or an argumentative essay on a given topic, has been the cornerstone of instruction for decades. Essay writing, or even report writing, has thus far involved meticulous research, hours upon hours of writing and editing, before submitting the final version to one's professor. Yet, with language models such as ChatGPT, essay writing became something that can be done in seconds, within clicks, with nothing but a prompt. As stated, ChatGPT can generate human-like text as an answer to a prompt. The prompt can be as detailed or as simple as the user decides and the chatbot's answers will vary based on the user input. The chief problem with these AI-generated essays is their detection. As mentioned, though valuable efforts have thus far been made by AI-detection software such as GPTZero and Turnitin's AI writing indicator, false positives make these detectors unreliable, which in turn, helps students cheat. It must be noted, that such cheating not only compromises the educational process, as a whole, but also devalues the efforts of the students who don't engage in academic dishonesty and invest time and effort into producing their own, authentic writings. And though according to some papers, ChatGPT can be successfully used in different activities in the language classroom to improve students' communication and writing skills (Shalevska, 2023), ethical considerations remain.

2.1. Standardized Exams in N. Macedonia and Academic Dishonesty

Standardized exams in multiple subjects mark the end of high school education in North Macedonia. These state-issued exit exams are essential for evaluating students' academic progress in the four years of compulsory gymnasium or vocational schools throughout the country. They determine how prepared high schoolers are for higher education and can influence students' chances of being accepted into their higher education institution of choice. One important part of the language exams for both the native Macedonian language, as well as the second, English language, is the essay writing section. These essays test students' communication and language skills as well as their critical thinking and analytical writing skills. Students are expected to write about their views clearly and convincingly, showcasing their ability to craft well-organized arguments with minimal mistakes.

The Macedonian language exam, a compulsory component of the standardized testing regime of N. Macedonia, assesses students' proficiency in the native language. Yet, the essay writing task in this exam evaluates not only language skills but also students' literary interpretation skills. Students are expected to be able to write about a vast number of books and convey their thoughts about given authors, styles, periods, and works through their essays. On the other hand, the English language exam evaluates students' English language skills in general. The inclusion of an essay writing task in this exam does not focus on works of literature but allows students to showcase their language proficiency, as well as their ability to write about diverse topics in non-native English.

Standardized exams everywhere seem to always involve certain fraudulent behaviours, and North Macedonia is no exception. The 2022/2023 English graduation exam caused a lot of controversies as around 7 000 Macedonian high school seniors were discovered to be part of a Viber group in which they shared answers to the test. The students seemed to communicate during the exam itself, which shows worrisome misconduct and alarming disrespect to the no-phones-during-exams policy (Sitel, 2023).

As for the cheating itself, students seemed to cheat on all exam tasks. What makes cheating even easier now, provided that they have access to their phone, are the AI-powered chatbots such as ChatGPT, which can produce high-quality, coherent essays in a matter of seconds.

3. Methodology

This small-scale study uses a mixed-methods approach, combining qualitative and quantitative techniques, to comprehensively investigate the link between academic integrity and language models such as ChatGPT. The study aims to shed light on the potential impact of AI tools on academic integrity by evaluating their ability to provide correct test answers and “help” students with their writing tasks.

3.1. Data Collection

1. *Test Answer Accuracy Assessment*: A subset of questions from the standardized exam in English in N. Macedonia is selected. ChatGPT’s answers are then compared with expert-reviewed answers to evaluate the accuracy of responses produced by the language model. Results are statistically analyzed.

2. *Essay and Text Analysis*: Three different prompts are used to generate an essay introduction using the free, 3.5 ChatGPT platform, which is deemed the most accessible. These outputs are then analyzed qualitatively to identify possible patterns, overall coherence, and potential differences in generated texts due to user input. Furthermore, the AI-generated introductions are tested by AI-detection software to test current detectors' capabilities.

This study was conducted with no external funding, ensuring unbiased analysis of the subject matter. The author declares no conflicting interests that could influence the outcomes or compromise this study’s objectivity.

4. Results and discussion

4.1. Test Answer Accuracy Results

To test the accuracy of the answers generated by ChatGPT 3.5, the two most recent standardized high school exit exams were used: the English language exam from June and the one from August 2023. Both tests are identically composed, and include several sections that aim to test students’ language skills:

- Reading and Comprehension: Multiple choice questions to do with a given text
- Reading and Comprehension: In-text gap completion with missing sentences
- Use of English section: Word formation exercise
- Use of English section: Keyword transformations
- Use of English section: Sentence transformations
- Open Cloze exercise
- Written Task: Email
- Written Task: Essay

This study’s focus is on the language-use exam tasks as well as the reading and comprehension tasks, as they are both machine-graded and follow a set of given answers deemed correct by experts. In terms of the first task, the language model produced satisfactory results in terms of accuracy, with only a few simple prompts:

[1] Can you help me answer some questions to do with a text?

[2] This is the text. I'll provide the questions in the next prompt. *text sent*

[3] The questions are as follows: *questions provided* Please provide the answers only. No explanations needed.

Using these comands, ChatGPT was then able to provide the answers to all 8 questions from the June exam in less than 2 seconds. The answers provided were then compared to the provided key. Correct answers were marked by the author with a “✓”, while incorrect ones – with an “X”:

June exam results:

- (1.1) B. ✓ (1.2) C. ✓
(1.3) D. ✓ (1.4) A. ✓
(1.5) C. ✓ (1.6) A. X
(1.7) B. X (1.8) B. ✓

August exam results:

- (1.1) B. X (1.2) C. X
(1.3) A. X (1.4) D. ✓
(1.5) C. ✓ (1.6) D. X
(1.7) A. ✓ (1.8) D. X

One can easily note that ChatGPT's accuracy in the first task was significantly lower in the August exam (37.5% accuracy) than in the June one (65% accuracy). Though the June results proved satisfactory, ChatGPT seemed to struggle with the text from the August exam, due to possible text complexity. A similar, yet not as prominent discrepancy was noted in the second task as well, for which the following prompts were used:

[1] Can you help me fill in some missing sentences in a text?

[2] This is the text. I'll provide a list of sentences in the next prompt. Note that there are two extra sentences that you will not need to use. *text sent*

[3] The missing sentences are as follows: *questions provided* Please provide the answers only. No explanations needed. Note that one answer can be used only once.

Generated answers for the June exam:

- (1) B. ✓ (2) E. X
(3) F. ✓ (4) I. ✓
(5) H. X (6) C. ✓
(7) D. ✓

Results for the August exam:

- (1) G. ✓ (2) E. ✓
(3) D. X (4) A. ✓
(5) C. ✓ (6) F. ✓
(7) H. ✓

With 65% accuracy for the June exam and 87.5% for the August one, ChatGPT provided more than satisfactory results in the second task, though the complexity of the task itself seemed to challenge the AI model more than changing single words.

The following, third section, which focused on single-word word formation, showed how much better the model is at simpler tasks. In this task, students are given a word that they have to change (usually by adding prefixes or suffixes) to create a word that will seamlessly fit into the given sentence. To prompt ChatGPT to do this task, the following text-based command was used "Complete the sentences (1-10) by changing the given word (in brackets) to the correct form. Consider the example. Provide answers only." The answers were as follows:

June exam:

1. understanding ✓
2. ability ✓
3. performance ✓
4. impractical ✓
5. approval ✓
6. diligently ✓
7. familiarize X
8. infrequent ✓
9. encouraged ✓
10. applicants ✓

August exam:

1. healthier ✓
2. annoying ✓
3. careless ✓
4. satisfaction ✓
5. indecisive ✓
6. unimpressive ✓
7. acceptance ✓
8. sympathetic ✓
9. reasonable ✓
10. beautifully ✓

ChatGPT clearly aced this section with a near-perfect (90%) accuracy score in the June exam and a perfect score in the August one. This suggests that ChatGPT has a great ability to understand word formation rules and apply them accurately, to form a corresponding word that will fit the given

context. Being pre-trained on a large corpus and being adept at contextual understanding, ChatGPT used its pattern recognition abilities to excel in this task.

Similar results were observed in the fourth section, the so-called, keyword transformation exercise. In this task, students had to transform a series of sentences, using a given keyword, to ensure both sentences have the same/near-same meaning and the keyword isn't changed. To prompt ChatGPT to do these transformations, the following command was used: "Using the capitalized word, complete the second sentence so that its meaning is the same as that of the first sentence. Do not change the key word. Use between 2 to 5 words. Contractions are also allowed. Consider the example."

June results:

1. It's the MOST exciting match we have ever been to. ✓
2. It took her a while USED TO helping with the household chores. X
3. Jenny is so scared of dogs that she MUST have been bitten by a dog when she was little. ✓
4. In SPITE OF not feeling well yesterday, he went to work. ✓
5. My grandfather takes PRIDE in his collection of watches. ✓

August results:

1. My friend ADVISED me NOT TO spend so much money on a gym subscription. ✓
2. The view from the skyscraper IS AS GOOD AS the one from the mountaintop. X
3. If ONLY YOU HAD COME to the planetarium with us. ✓
4. If you hadn't forgotten your glasses, you WOULD HAVE BEEN able to read the subtitles now. ✓
5. It IS BELIEVED that robots will replace many jobs in the future. ✓

With an 80% accuracy rate and only 2/10 mistakes, the AI model exhibited great ability to manipulate language and maintain the meaning of sentences while preserving the provided keyword.

Other than some tense mix-ups that can be attributed to vague grammar rules, no other major issues were encountered in the fifth task, as well, which again focused on sentence transformations, however, with no given keyword. To complete the task, ChatGPT was given the following prompt: "Fill in the gap in the second sentence so that its meaning is the same as that of the first sentence. Use between 2 to 5 words. Contractions are acceptable. Consider the example. Provide answers only."

June results:

1. She can't go out unless she finishes her homework. ✓
2. Tom has been accused of breaking into the company's system. ✓
3. Sarah said that she didn't cheat on the exam the day before. X
4. Only then did we realize that the vehicle had been stolen. ✓

5. If only he would stop making so much noise upstairs. ✓

August results:

1. I stopped drinking coffee a week ago. ✓
2. I had my eyesight checked last week. ✓
3. My best friend has been in Paris for five years. X
4. Daniel said he had had that book for ages. ✓
5. Some remarkable discoveries have been made by the research team. ✓

With an 80% accuracy, and only two mistakes due to subtle grammatical nuances and selecting the most appropriate verb forms, ChatGPT yet again demonstrated a superb understanding of the task. As for the mistakes, according to the key, Past Perfect should have been used in the sentence no. 3 in the June exam and the Present Perfect Continuous should have been used in sentence 3 in the August exam.

The last section of interest, the sixth section called open cloze, demanded that students fill in the gaps in the given text with words they will think of themselves. ChatGPT was prompted to do this

task using the following command: "Fill in each gap in the text using ONE word only. Provide answers only."

June results:

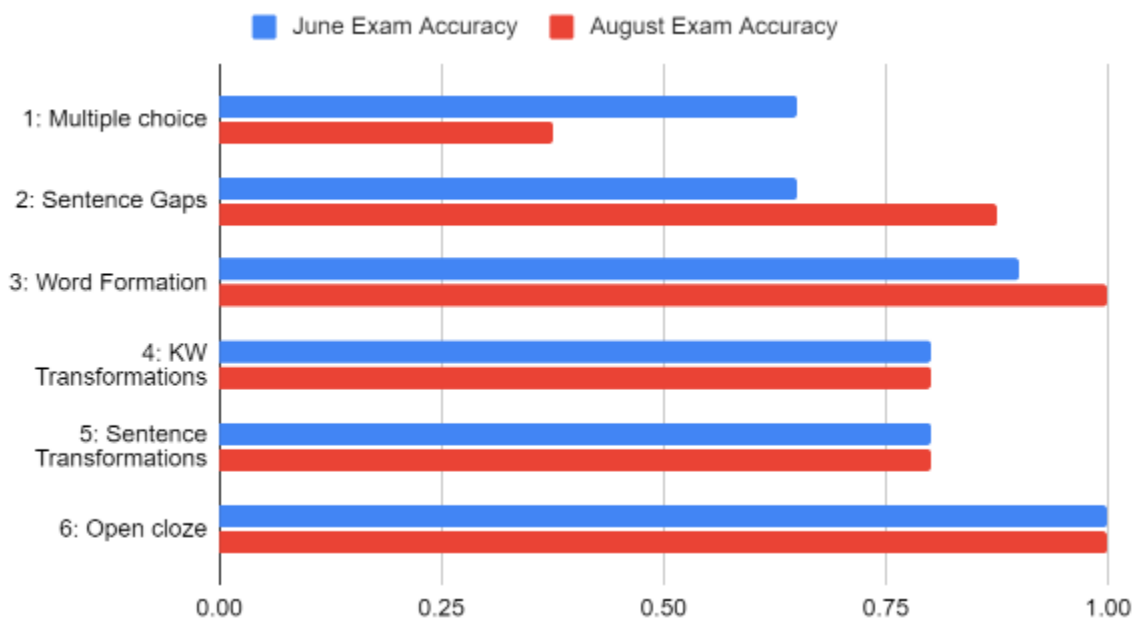
- | | | |
|------------|-------------|-----------|
| 1. is✓ | 2. to✓ | 3. comes✓ |
| 4. what✓ | 5. with✓ | 6. of✓ |
| 7. taking✓ | 8. towards✓ | 9. worth✓ |
| 10. more✓ | | |

August results:

- | | | |
|-------------|------------|-----------|
| 1. an✓ | 2. back✓ | 3. spend✓ |
| 4. getting✓ | 5. amount✓ | 6. our✓ |
| 7. through✓ | 8. with✓ | 9. put✓ |
| 10. | | by✓ |

ChatGPT demonstrated a truly remarkable accuracy of 100%. The clear prompting and the nature of the task itself, which requires less creativity, fits well with GPT-3.5's inherent capabilities. In addition, unlike some of the previous exam tasks, which involved more complex language skills, the open cloze task seemed to be relatively straightforward, as gaps are to be filled with words that are an unambiguous fit.

All in all, with an average test accuracy score of about 80% on both exams, ChatGPT excelled in almost all tasks, with no major issues other than some minor mistakes due to nuanced grammar rules, and overall text complexity. Other than the tasks that required a deeper understanding of the content and context of the text (which, in turn, may demand a higher level of reasoning, even from humans) the AI model excelled in all tasks provided, as shown in Graph 1. Still, it's important to note that while the model aced these tasks, there can still be instances where its answers might diverge from human evaluators due to the intricacies of language and interpretation.



Graph 1: Test Answers Accuracy

4.2. Text Analysis

Due to the limited length and scope of this paper, this research will only analyze AI-generated essay introductions for the June exam, based on user input and prompt detail. For this test, the essay topic from the June exam was used, and 3 prompts were written by the author, all different in length and complexity, to examine the difference in AI-generated text based on user input. As such, the introductions in Table 1 were generated:

Prompt	[1] Write an introduction for an essay on: Do you believe that a sincere apology is the only way to resolve an argument and preserve a good friendship?	[2] Write a 50-word introduction for an essay on: Do you believe that a sincere apology is the only way to resolve an argument and preserve a good friendship? Take a YES or NO position and mention two reasons that support it.	[3] Write an introduction (45-50 words in length) for an essay on: Do you believe that a sincere apology is the only way to resolve an argument and preserve a good friendship? State your position and mention two reasons that support it. Use formal yet simple English, without any difficult words, on a B1 level. Make 1 grammar and 1 spelling mistake.

ChatGPT-generated text	<p>[1] In the intricate tapestry of human relationships, conflicts and disagreements are inevitable. As we navigate the complexities of our interactions, we often find ourselves grappling with the question of how to mend the ruptures that occur between us. At the heart of this inquiry lies the concept of a sincere apology—a seemingly simple gesture imbued with profound implications. In the context of both resolving arguments and preserving cherished friendships, the act of apologizing stands as a pivotal bridge between hurt and healing, misunderstanding and empathy. This essay delves into the multifaceted dynamics surrounding the idea of a sincere apology, exploring its potential as the linchpin for conflict resolution and the safeguard for the endurance of meaningful connections. By scrutinizing the role of apology in these dual contexts, we can unravel whether a sincere apology is indeed the sole pathway to not only resolving disputes but also nurturing the longevity of valuable friendships.</p>	<p>[2] In the intricate dynamics of human relationships, the pivotal role of sincere apologies in conflict resolution and friendship preservation is a subject of contention. This essay takes a firm stance, arguing that a sincere apology is indeed the primary means to mend disputes and uphold strong friendships. The assertion is substantiated by its potential to heal emotional wounds and its capacity to rebuild trust on foundations of humility and understanding.</p>	<p>[3] In this essay, I will discuss whether a sincere apology is the sole solution for resolving conflicts and maintaining strong friendships. I firmly believe that offering a genuine apology plays a crucial role in restoring harmony after disputes. This viewpoint is supported by its ability to mend hurt feelings and its potential to rebuild trust in relationships.</p>
------------------------	--	---	--

Table 1: Introductions generated by ChatGPT

The first introduction, which follows a not-very-detailed prompt, appears to be the most formal. It is comprehensive and eloquently written, with a strong focus on a sophisticated vocabulary that isn't typically seen in spontaneous writings of non-native high school seniors. This formality makes this introduction stand out and may be an indicator for the one doing the assessment that the student has used ChatGPT or similar software to write it. Introductions such as this one, which deviate significantly from what is typical for a non-native speaker at a certain age, may be flagged for further scrutiny. Similar observations can be made for Introduction 2, as well.

Yet, Introduction 3 doesn't seem to have such issues. Due to the well-crafted, detailed prompt, the text appears not only human-like but also more likely to have been written by a non-native high school senior. This is due to its simpler language that corresponds with the B1 level. The slight grammar and spelling mistakes, which follow the prompt, may "trick" the assessor as they might make the text seem like it has been composed by someone who is still developing their language skills.

As one can easily note from the introductions above, user input significantly influences AI-generated texts. ChatGPT tailors its responses based on the prompts given, including formality, vocabulary, and language complexity. For instance, Introduction 1 clearly showcases a high level of eloquence, likely influenced by the sparse prompt that didn't specify simplicity. On the other hand, Introduction 3 demonstrates the impact of specific user requests for simplicity, formality, and even errors. In this case, the AI responds to the prompt that explicitly mentions non-native language proficiency, and generates a simpler text.

The use of AI to generate human-like content raises concerns about academic integrity, especially when text is generated using detailed prompts that may help it go unnoticed by assessors.

As AI tools like ChatGPT become more sophisticated, ethical considerations become even more crucial in maintaining academic integrity.

4.2.1. AI Detection

Though AI detection as of now is said to be unreliable, for the purposes of this research, the three introductions detailed above were checked by the ChatGPT-based detector called GPTZero – an AI tool that is meant to not only detect ChatGPT-generated texts but also texts generated by GPT4, BARD, Llama, and other AI models. The results obtained were as follows:

- Introduction 1: A **50%** probability this text was entirely written by AI;
- Introduction 2: A **51%** probability this text was entirely written by AI;
- Introduction 3: A **52%** probability this text was entirely written by AI.

The results obtained from GPTZero are interesting as they differ from the original analysis of the three introductions presented above. Namely, although Introduction 3 appeared to be the one which is most likely to be written by a human (due to the spelling and grammar mistakes and the overall simpler vocabulary and sentence structure), the AI-detector results suggest that all three introductions have a relatively close probability of being generated by AI. This difference between the initial analysis and the AI detector's assessment further emphasizes the evolving landscape of AI technology and the challenges it poses to accurate detection. While AI detection such as the one provided by GPTZero plays an important part in identifying potential academic dishonesty, the evolving capabilities of AI language models highlight the importance of critical thinking, creativity, and original thought in assignments – things that AI will have trouble replicating.

5. Conclusion

In the ongoing quest to determine the impact of AI on education, tests, and academic integrity, this study focused on ChatGPT's prowess in generating accurate test answers to the standardized English tests in North Macedonia and writing essay introductions. While excelling in various tasks, ChatGPT made minor errors in some instances due to nuanced grammar rules and overall task complexity. As for the generated texts, the study showed that user input (in terms of prompts) significantly impacted generated text. These further highlights AI's adaptability in the landscape of current technology.

Ethical concerns to do with ChatGPT-generated content's resemblance to human writing further emphasize the need for clear guidelines and strict rules, as well as a focus on nurturing critical thinking, creativity, and authenticity. Along the same line, the GPTZero-based AI detection results also point out the evolving challenges in accurate AI identification, which, as of now, remains unreliable. All in all, this small study details a part of the evolving relationship between AI and academics, prompting ongoing dialogue on ethics and AI.

References

- Fisher (2014). *The fundamental values of academic integrity* (2nd ed). International Center for Academic Integrity.
- Holden, O. L., Norris, M. E., & Kuhlmeier, V. A. (2021). Academic integrity in online assessment: A research review. *Frontiers in Education*, 6.
- Mollman, S. (2022). *ChatGPT gained 1 million users in under a week. Here's why the AI chatbot is primed to disrupt search as we know it*. Retrieved from: <https://finance.yahoo.com/news/chatgpt-gained-1-million-followers-224523258.html> on 22.8.2023.
- Shalevska, E. (2023). English teaching just got an upgrade: How AI is revolutionizing TESOL. *International Journal of Education "Teacher"*, 25.
- Sitel. (2023). *Grupno prepishuvanje vo Viber grupa na drzhavnata matura*. Retrieved from: <https://sitel.com.mk/grupno-prepishuvanje-vo-viber-grupa-na-drzhavnata-matura> on 22.8.2023.
- Susnjak, T. (2022). *ChatGPT: The End of Online Exam Integrity?* Retrieved from: <https://arxiv.org/abs/2212.09292> on 21.8.2023.