

UDC 81'367'42:004.8
doi: 10.15330/jpnuphil.10.47-67

LINGUISTIC ANALYSIS OF HUMAN- AND AI-CREATED CONTENT IN ACADEMIC DISCOURSE

YAROSLAVA FEDORIV, IRYNA PIROZHENKO, ALLA SHUHAI

Abstract. Artificial intelligence (AI), with the progress of technology, has impacted various fields of communication, including text generation. Therefore, differentiating between human-created and AI-generated content has become a game-changer in determining the authorship in such realms as education, journalism, and content development, as both human and AI forms of communication share similarities in conveying information and following grammar rules.

This research highlights linguistic tools to distinguish human-created and AI-generated English texts, specifically in the context of EFL proficiency assessment, and explores linguistic markers and methodologies to gain understanding of AI's impact on written communication.

The research methodology is based on a mixed approach, combining qualitative and quantitative data collection and analysis. Qualitative analysis identifies distinctive linguistic markers differentiating human and AI discourse, and quantitative analysis addresses sentence length, vocabulary diversity, frequency, and syntactic complexity.

This research findings showcase distinctive features differentiating between human and AI-generated texts: AI content may exhibit repetition on different syntactic levels, AI-characteristic vocabulary, repetitive transitions, inappropriate style, inconsistent tone, a lack of coherence, disparity with context, inaccurate information, and low-quality referencing.

As AI is rapidly advancing, further applications are seen in detecting AI-generated text in various domains, such as language learning and teaching, fact-checking, pursuing responsible communication, ethical use of AI content, academic integrity and respecting intellectual property rights.

Further research can explore how linguistic markers evolve with the development of AI language models. Distinguishing between human and AI-generated texts fosters creativity and academic integrity. The research highlights linguistic markers and methodologies, understanding AI's impact on written communication.

Keywords: discourse, artificial intelligence, language patterns, style, vocabulary, grammar, syntax, structure, cohesion.

“The Internet is increasingly awash with text written by AI software. We need new tools to detect it” (Heikkilä 2022).

1. INTRODUCTION

Due to the fast progress of technology, artificial intelligence (AI) has made significant impact in various fields, including language generation. With the introduction of AI language models like *ChatGPT*, it has become crucial to distinguish between human-created content and machine-generated

text. The differentiation between these two types of texts has become increasingly significant due to the increasing integration of artificial intelligence in various communication domains, including education, journalism, and content development. "With the advent of such tools as *ChatGPT*, AI can generate content that is eerily similar to human-written copy" (Tegze 2023). Therefore, both types of content equally contribute to the richness and diversity of discourse.

Both human and AI-produced forms of discourse share certain similarities, such as their capacity to convey information and ideas through language and observe grammatical rules so as to communicate meaningful messages, express emotions, tell stories, inform, persuade, entertain, and elicit diverse responses from the audience. In distinguishing between content generated by artificial intelligence programs, such as *ChatGPT*, and work created by skilled copywriters, one can employ various indicators, such as low-quality sources, repetitive words or phrases, clichés, and unnatural sentence structures. In another project submitted for publication we considered the lack and failure to convincingly convey human emotions in AI-generated content.

Hence, distinguishing between human-created English content and AI-generated text has become a challenge for researchers and language practitioners alike. This research emphasises the significance for human readers to accurately identify the origin of the content and secure the intellectual contributions of human writers. By studying the features that set human and AI discourse apart, we can prevent false authorship, ensuring fair attribution and preserving the essence of human creativity in communication, which is even more acute for non-native users of English.

In this context, the **aim of this research** is to explore the linguistic tools for the differentiation between non-native human-created and AI-generated English content in order to lay the foundation for more accurate and effective language usage in the EFL domain.

To achieve this goal, our analysis will focus on certain linguistic features and patterns in human-created and AI-generated texts so as to gain insights into the complexities of this issue and its implications for EFL proficiency assessment. This methodology was proposed for validation at *The 11th International scientific and practical conference "Scientific Progress: Innovations, Achievements And Prospects"* (Fedoriv, Pirozhenko, & Shuhai 2023, p. 139-144). The further elaboration of this approach is presented below.

I. Overview of linguistic characteristics of human-created discourse (Literature review)

Regarding the human-created discourse, **grammatical and syntactical aspects** are considered fundamental for effective communication and understanding in written and spoken language, and they can serve as key differentiating factors. Human-produced texts imply the in-depth understanding of "the functional organisation of clauses" and exhibit intricate "aspectual properties and extended projections of verbs in linguistic expressions" (Fries 1954, p. 57–68). At the same time, non-native human-written texts may include grammatical errors, unconventional word order, or inconsistent verb tenses.

Another issue to consider is the **unnatural language use**, which occurs when non-native human-written texts showcase such characteristics as a person's linguistic behaviour, or a peculiar way of speaking or writing. Particularly, due to language background and exposure, non-native users may opt for certain words or phrases that differ from those commonly used by native speakers. They may make language errors specific to their native tongue, e.g., inaccuracies when using international words

(Veisbergs 1996; Topalova 2010, p. 215-222; Lewis 2020). The unnatural language use may also involve the incorrect or awkward use of grammar, idiomatic expressions, collocations, or sentence structures that are not commonly used by native speakers. This divergence in linguistic choices can influence how the text is perceived and understood by the audience. Ultimately, unnatural language use can affect the clarity and effectiveness of communication, as it may lead to misunderstandings or difficulty in conveying intended meanings.

One of the most significant aspects of meaningful communication seems to be **cohesion**, which refers to creating a sense of unity and coherence in the overall piece of writing or spoken communication as well as ensuring that the different parts of a text be logically linked and work together to convey a clear and meaningful message to the audience. Human cohesion is the result of a complex cognitive process influenced by various factors such as linguistic competence, background knowledge, cultural awareness, and personal experiences. Humans employ a wide range of linguistic devices, such as pronouns, connectives, conjunctions, and lexical repetition (Er 2004; Qing-yuan 2000), to establish logical connections between ideas and ensure smooth transitions from one sentence or paragraph to another. Text cohesion includes not only cohesive devices within the text but also those that relate text with the situation (Delu 2005). Humans rely on their knowledge of discourse conventions and genre-specific expectations to structure their writing and effectively convey meaning.

In terms of **context and coherence**, proficient language users possess more than just practical knowledge of the languages they speak; they, according to Matthews, have a deeper understanding that goes beyond mere language skills (2006, p. 200-220). In contrast, non-native English users may encounter coherence issues reflecting cultural or knowledge gaps related to their background. In some instances, non-native speakers may endure miscommunication even with a strong command of the target language, which is known as “pragmatic failure” (BAO 2020).

It is also worth considering **consistency and response patterns**. Consistency is defined as “the orderly treatment of a set of linked elements” (Farkas 1985). When examining human texts, we often observe significant variability in terms of writing style. For instance, writers may employ metaphors as creativity markers (Price 1974, p. 253-256), which can alter the overall tone and impact of the message being conveyed.

Response patterns, on the other hand, refer to the predictable ways to respond to certain communicative situations. This includes the use of common phrases, expressions, or discourse markers that are typical in specific contexts. In non-native human-written text, response patterns, such as sentence structure and organisation, may vary in complexity and clarity, leading to differences in how information is presented and perceived. Understanding and analysing these response patterns are essential for accurately assessing and evaluating non-native created texts, especially in the area of English as a Foreign Language.

Knowledge and cultural awareness can also aid in determining the types of content. This implies the cultural specificity of human language, conducted mostly in modern multiethnic industrialised societies (Singh, Lele, & Martohardjono 1988, p. 43-59]. Although non-native human-written texts may show gaps in cultural knowledge, they may also provide insights or perspectives specific to the writer's culture or background.

Adhering to principles of proper referencing in academic papers is crucial as it ensures **credibility and integrity** by providing accurate and verifiable sources to support the claims made in the text. It also

demonstrates respect for the intellectual property of others by giving proper credit to the original authors and avoiding plagiarism, as well as allows readers to access and explore the cited sources, promoting transparency and facilitating further research in the corresponding field.

II. Research methodology

Linguistic pivoting of human- and *AI*-created texts refers to the process of analysing and comparing the linguistic features and characteristics present in texts created by humans and those generated by artificial intelligence systems.

Within the framework of this research project, 'pivoting' implies the act of turning or changing the focus of analysis from one type of text to the other. The term 'linguistic' emphasises that the analysis centres on language-related aspects, such as vocabulary, grammar, syntax, and style.

By conducting a linguistic pivoting analysis, researchers aim to discern patterns, differences, and similarities between human-written texts and *AI*-generated texts. This examination underlies understanding how *AI* language models function and how they compare to human language production. The insights gained from such analysis can help researchers and language practitioners in various fields to better comprehend the strengths and limitations of *AI* language models and their impact on communication.

1. Research Design

This study involves mixed methods to investigate the linguistic markers of differentiating human and *AI* discourse. It is based on qualitative and quantitative data collection and analysis to provide a comprehensive understanding of the distinct features in human-created and *AI*-generated texts.

2. Corpus Compilation

To ensure the representation of uniform linguistic samples, a corpus of texts has been compiled, consisting of human-written and *AI*-generated content related to the same context.

The human-created texts have been sourced from student academic essays submitted to the university Learning Management System (LMS). The assigned tentative topics were to be developed around current events in Ukraine of 2021-23 with a focus on communication in the wartime.

The students were instructed to introduce the paper with reasons as to why their specific topic was selected. The choice of the topic should determine the title of the work and the research material. The introduction should include background information, the understudied areas, and the research problem or question to be addressed by the author of the essay. The objectives of the research should be specified, with the research methodology outlined.

The main body should be divided into distinct sections starting with an overview of relevant literature that reflects previous research approaches to the subject area and followed by the discussion and analysis based on the student's chosen materials and research methods.

The subsequent conclusions should reflect the hypotheses underlying the research. The student should indicate whether they have been proven or not and suggest further application of the research results.

The bibliography should be arranged in the assigned documentation style (e.g., MLA) and reflect theoretical sources, linguistic reference sources (e.g., dictionaries), and materials researched. Appendices may also be supplemented.

AI-generated texts have been obtained from an available version of the AI language model *ChatGPT* with a track record of generating texts in accordance with the genre and context corresponding to the academic essay requirements outlined above.

3. Data Collection and Selection

For the qualitative analysis, a purposive sampling approach is employed to select a subset of texts from the corpus that would highlight specific linguistic features relevant to the research objectives. The selected texts are subjected to a thorough linguistic examination.

4. Linguistic Analysis

Qualitative analysis involves a close reading of the selected texts to identify distinctive linguistic markers that differentiate human and AI discourse. Various linguistic parameters such as vocabulary, syntax, grammar, style, and coherence are assessed. Particular attention is paid to word choices, patterns of expression, sentence structures, and stylistic elements that exhibit variations between the human- and AI-produced types of content. Additionally, quantitative analysis addressed vocabulary diversity and frequency, the average sentence length, and syntactic complexity so as to quantify the linguistic distinctions between the two types of texts.

5. Comparison and Interpretation

The findings from the qualitative and quantitative data are analysed to determine linguistic markers that distinguish human- and AI discourse. The identified patterns and variations are compared and contrasted to outline indicative similarities and differences between the two types of content.

6. Validation and Reliability

To ensure the validity and reliability of the study, inter-rater agreement is assessed for the qualitative analysis. Three researchers authoring this project have independently evaluated the chosen subset of texts, and the agreement levels are measured to establish consistency in identifying linguistic markers.

7. Ethical Considerations

This research upholds ethical standards by respecting anonymity, copyright and intellectual property rights when collecting human-written texts. In order to avoid impartiality in the analysis and for the purpose of protecting the students' identity, their written submissions have been numerically coded with the names removed from consideration.

AI-generated texts have been obtained from the publicly available source *ChatGPT*, ensuring compliance with usage policies.

8. Limitations

The study acknowledges potential limitations, such as the availability and representativeness of both human-created and AI-generated texts and the evolving nature of AI language models. However, by employing meticulous data collection and analysis methods, the research aims to mitigate these limitations and provide considerable insights into the linguistic distinctions between human and AI discourse.

The authors of this paper hypothesise that the outlined research methodology can help to comprehensively examine the linguistic markers that differentiate human-created and AI-generated texts, contributing to the academic understanding of the impact of AI on text production and communication.

III. Comparison of the human-written and AI-generated text: A Case Study

A. Assessment of student-submitted content

Assignment: Write an academic essay on the topic "LANGUAGE MEANS OF ORAL SPEECH EXPRESSION IN WAR CONDITIONS"

Note: The text under analysis is a part of a student's academic essay submitted in 2022, before the introduction of *ChatGPT*, which is why it can be taken as an example of a human-created text.

Sample introductory paragraph in a student's paper:

Human beings face a lot of massive problems today such as climate change, crimes and wars. People find it hard not to pay attention to these things. Wars cannot be ignored because they are related to all means of life of every person, especially when a war is started by a big influential country. For that reason, it is crucial to analyse the speeches of officials, the way they understand the situation and things they do and can do to reduce or exacerbate the problem. This is the way to understand the war's origin and manage to be instrumental in stopping it. The main attention should be drawn to linguistic means of political speeches that are essential in publicistic style.

The indications that the excerpt has been written by a non-native writer include:

- **Word choice:** The phrase "every person" could be more appropriately expressed as "all individuals" in a native-level academic text which typically employs more sophisticated and contextually accurate word choices.

- **Grammar:** The definite article is missing in "essential in [the] *publicistic style*" for an item taken out of the list of conventional functional styles.

- **Syntax.** (i) A comma is missing before the expression "such as." (ii) The Oxford (or serial) comma, i.e. the final comma in a list of items) is missing in "climate change, crimes[,] and wars." (iii) Limited variation in sentence complexity: the student tends to rely on simple sentence structures, while native speakers often incorporate a wider range of sentence structures for improved readability.

- **Style:** (i) "Human beings face a lot of massive problems today such as climate change, crimes and wars." The use of "a lot of massive problems" is colloquial and should be replaced with "many massive problems." (ii) The writer's simplified language appears straightforward and lacks the nuances often seen in native-level writing. Native speakers tend to use more varied vocabulary and complex sentence structures.

- **Uncommon phrasing:** (i) The use of phrases such as "a lot of massive problems" instead of more formal and academic expressions like "numerous significant challenges" suggests a non-native command of the language. (ii) The use of "all means of life" is awkward and unnatural. It should be rephrased as "all aspects of life." (iii) "Wars cannot be ignored because they are related to all means of life of every person, especially when a war is started by a big influential country." The word "big" is vague, and the expression "big influential" could be replaced with "powerful." (iv) The use of "things they do and can do" is reduplicated. The expression can be rephrased to "actions they take." (v) The phrasing "the way to understand" might sound awkward to academic native speakers, who would more likely opt for "the approach to gaining comprehension" or similar formulations.

- **Cohesion:** (i) The passage lacks clear transitional words or phrases to smoothly connect ideas and create a cohesive flow of information between sentences and paragraphs. (ii) "For that reason, it is crucial to analyse the speeches of officials, the way they understand the situation and things they do and can do

to reduce or exacerbate the problem." The connector "for that reason" does not refer to any clearly stated reason.

- **Coherence:** (i) "People find it hard not to pay attention to these things." This sentence is grammatically correct yet redundant in terms of adding to the content. (ii) The expression "to reduce or exacerbate the problem" is self-contradictory and hence incoherent.

- **Cultural Knowledge:** The passage mentions the importance of analysing the speeches of officials, but it does not specify the context or names of the officials or countries involved; it lacks specific examples or references to historical or contemporary events, which could be relevant for cultural understanding and context. Adding relevant examples could improve the cultural knowledge aspect.

- **Credibility** and the lack of specific examples or evidence: While the writer presents perspectives on analysing political speeches, the absence of specific instances or supporting evidence weakens the overall argument. Academic native writers usually provide valid examples to reinforce their points.

Apparently, while the student's text attempts to convey the intended message, the aspects listed above indicate that the passage has been written by a non-native English speaker.

B. Examination of linguistic characteristics in AI-generated discourse

Prompt: write an introductory paragraph to the topic: "Language Means of Oral Speech Expression in War Conditions."

Tab. 1

AI-generated responses to the topic: "Language Means of Oral Speech Expression in War Conditions"

AI response 1	AI response 2
Prompt 1: "Write an introductory sentence to the topic: "Language Means of Oral Speech Expression in War Conditions"	
In the challenging and high-stakes context of war conditions, language serves as a crucial means of oral speech expression, playing a pivotal role in communication, information dissemination , and emotional connection amidst the tumultuous backdrop of conflict.	In the context of war conditions, this study examines the language means utilized in oral speech expression.
Prompt 2: write an introductory paragraph to the topic: "Language Means of Oral Speech Expression in War Conditions"	
In the tumultuous and arduous circumstances of war, language emerges as a fundamental and multifaceted tool of oral speech	In the harrowing context of war conditions, the language means of oral speech expression take on a profound significance , serving as

<i>AI response 1</i>	<i>AI response 2</i>
<p>expression. Amidst the chaos and uncertainties, individuals engaged in conflict find themselves relying on language to convey critical information, share personal experiences, and forge emotional connections. The language means of oral speech expression in war conditions encompass a wide range of elements, including spoken words, tone, gestures, and non-verbal cues. These linguistic nuances play a pivotal role in fostering camaraderie among soldiers, transmitting orders with precision, and offering solace to those enduring the harrowing realities of war. Exploring the intricate dynamics of language in this context not only sheds light on the power of communication under extreme circumstances but also deepens our understanding of the human spirit's resilience and adaptability in times of adversity.</p>	<p>powerful tools to communicate and convey emotions, experiences, and critical information. In times of conflict, oral speech becomes a lifeline, enabling soldiers, civilians, and responders to share their thoughts, fears, and hopes amidst the chaos and adversity. This topic delves into the linguistic aspects and techniques employed in oral communication during war, exploring how language becomes a crucial instrument of survival, solidarity, and resilience in the face of unimaginable challenges. Through the lens of language means, this study sheds light on the intricate dynamics of oral speech expression in war-torn environments, revealing the ways in which communication plays an indispensable role in connecting individuals, shaping collective identity, and preserving humanity amid the most extreme circumstances.</p>
<p>Prompt 3: write an introductory paragraph to an academic essay on the topic: "Language Means of Oral Speech Expression in War Conditions"</p>	
<p>Amidst the harrowing and volatile landscape of war conditions, language emerges as a powerful conduit of communication and expression. The profound impact of oral speech becomes evident in the midst of conflict, where language serves as a crucial means for conveying critical information, sharing personal experiences, and fostering emotional connections. In the context of war, the nuances of language take on heightened significance, shaping the dynamics of interactions among individuals engaged in hostilities and influencing the course of events. This academic essay delves into the intricate realm of "Language Means of Oral Speech Expression in War Conditions,"</p>	<p>In the harrowing context of war conditions, language emerges as a fundamental and powerful tool of oral speech expression. Amidst the chaos and uncertainties that characterise such tumultuous circumstances, individuals engaged in conflict find themselves relying on language to communicate critical information, convey emotions, share personal experiences, and forge deep emotional connections. The language means of oral speech expression in war conditions encompass a wide range of elements, including spoken words, tone, gestures, and non-verbal cues, each playing a pivotal role in fostering camaraderie among soldiers, transmitting orders with precision, and</p>

<i>AI response 1</i>	<i>AI response 2</i>
<p>exploring the multifaceted roles and implications of language as a tool for connection, command, and survival amidst the turbulent backdrop of war. Through an in-depth analysis of linguistic elements and their effects on the human experience during conflict, this study aims to unravel the profound interplay between language and war, shedding light on its transformative power and pivotal role in shaping the course of history.</p>	<p>offering solace to those enduring the harrowing realities of war. This academic essay delves into the intricate dynamics of language in this context, seeking to explore the profound significance of oral speech as an indispensable instrument of survival, solidarity, and resilience in the face of unimaginable challenges. Through a comprehensive analysis of language means, this study aims to shed light on how communication becomes a lifeline in war-torn environments, revealing the ways in which it plays an indispensable role in connecting individuals, shaping collective identity, and preserving humanity amid the most extreme circumstances.</p>

As it can be seen from Tab. 1, the targeted *AI* responses were elicited after several consecutive steps of fine-tuning the prompts, as the responses were failing to correspond to the requirements of academic style. An explanation to this failure can be that the students attending a course of Academic Writing are aware of the context and follow the rules for writing an academic essay by default, while *ChatGPT* does so only after having been given clear and detailed instructions.

The characteristics of the *AI*-generated text are outlined here in accordance with the above mentioned criteria for the analysis of the human-created content.

Word choice: The text consists of a varied and sophisticated vocabulary beyond that of a non-native user of English, including words like "tumultuous," "arduous," "multifaceted," "chaos," "uncertainties," "camaraderie," "solace," "harrowing," and "adversity." At the same time, the choice of words demonstrates rigidity manifested through repetitive usage of such words and expressions as "indispensable," "amid," "amidst," "adhere," "utilize," "delve into." Among structural notable features the section with "overall," "finally," "summarizing" is indicative and might be observed in generated content in response to essay-like assignments to *AI*. It should also be noted that, unless instructed otherwise, *AI* produces responses according to the default Standard American spelling settings, which creates a mismatch with the European EFL teaching standard and is indicative of a writer's unskilled attempt to blend texts of different origins.

Grammar: Unlike non-native human-created content, the *AI*-produced text shows a mastery of grammar, with correct verb tenses, subject-verb agreement, and sentence structure.

Use of contracted verb forms is a common feature in native-like writing, yet it contradicts the academic writing requirements.

Syntax: The syntax in *AI*-generated texts is well-constructed, with clear, coherent, and properly punctuated sentence patterns that flow smoothly, creating a cohesive text.

Style: The text is supposed to maintain a formal tone and avoid the use of informal language, which is considered to be typical of academic texts developed by human writers. However, the writing style of the analysed piece (Tab. 1) turns out to be expressive by employing emotive language and vivid imagery.

It should also be noted that *AI* tends to provide as many nouns with *epithets* as possible, such as “*war-torn environments*”, “*transformative power*”, “*indispensable role*”, “*harrowing context*”, “*tumultuous circumstances*”, and others marked bold in Tab. 1, which is redundant in the case of an academic paper.

Colloquial expressions: The *AI*-generated text occasionally includes colloquial expressions such as the verbs “*interplay*,” “*encompass*,” “*unravel*,” “*adhere*,” the phrasal verb “*delve into something*,” the figurative expressions “*shed light*,” “*volatile landscape of war conditions*,” “*shaping the course of history*” and “*war-torn environments*” are apparently recognised as indicators of *AI*’s frequently used “viral” words, which does not comply with academic style requirements.

Language usage: The text produced by *AI* does not contain any uncommon or awkward phrasing. Instead, it exhibits a proper use of language. However, in terms of academic writing, the text deviates from the required formal expression.

Cohesion: The text is cohesive, with ideas and arguments logically connected and flowing smoothly from sentence to sentence.

Coherence: The coherence of the text is presented through a well-structured and logical sequencing of information, making it easy for readers to follow and comprehend the main points.

Cultural Knowledge: The text complies with the cultural context of war conditions, discussing the importance of language means in oral speech expression during conflicts.

Personal experiences: The text presents how individuals in war conditions use language to share their personal experiences, indicating an objective approach to the topic, yet it does not directly include personal experiences of the writer.

The analysed introduction does not establish **credibility** as this section does not necessarily require such a level of detail. Therefore, this aspect should be supported by the relevant literature review.

Thus, the next point of our consideration will be the integrity of citing sources in an academic paper.

C. A close reading of a seemingly *AI*-generated text: a case study

A dubious sample of a literature review in the essay submitted by a student in 2023 (Tab. 2) showcases the features suggesting that the text has been generated by *AI*, which will be illustrated and discussed below. The reason for questioning the origin of the examined piece of writing is the student’s inconsistent EFL command manifested dissimilarly in different parts of the same paper.

Tab. 2

Excerpt from a dubious literature review submitted by a student in 2023

Student's literature review content	Comments
<p>Significant numbers of scholars have written about the discourse of war. We have studied several scholars who have made significant contributions to this field.</p>	<ul style="list-style-type: none"> - Unspecified information - Word repetition
<p><u>Michel Foucault</u> was a French philosopher and social theorist who explored the relationship between power, knowledge and discourse. His book "Discipline and Punish" influenced the analysis of the discourse on war.</p>	<ul style="list-style-type: none"> - Source exists - In-text citation missing - Reference item absent in Works Cited
<p><u>Chantal Mouffe</u> is a political theorist who has made a significant contribution to the field of critical discourse analysis. She explores the role of discourse in the construction of political identities, the formation of antagonisms and the justification of violence in the context of war.</p>	<ul style="list-style-type: none"> - In-text citation missing - Reference item absent in Works Cited
<p><u>Norman Fairclough</u> is a linguist and critical discourse analyst who studies the discourse of war and its representation in the media. His work emphasises the relationship between language, power and ideology in shaping public perceptions of war.</p>	<ul style="list-style-type: none"> - In-text citation missing - Reference item absent in Works Cited
<p>The study of the speech of native English speakers in critical conditions of war, in particular in the context of military conflicts, has been carried out by the following scholars.</p>	<ul style="list-style-type: none"> - AI-styled
<p><u>R.M. Whiteside</u> in his book "Language at War". He analyses the language used by English-speaking soldiers during the Second World War. He also explores the impact of the war on language, including the development of military jargon, slang, and coded language.</p>	<ul style="list-style-type: none"> - Fictitious source - In-text citation missing - Reference item absent in Works Cited
<p><u>Samuel Hines</u> in his book "Soldiers Stories: Testimonies of Modern Warfare". He explores the experiences of English-speaking soldiers in the wars of the 20th century, including World War I, World War II, and the Vietnam War. It also explores how the soldier's language and stories reflect their experiences and the unique challenges they</p>	<ul style="list-style-type: none"> - Fictitious source - In-text citation missing - Reference item absent in Works Cited

Student's literature review content	Comments
face in the critical conditions of war.	
Kate McLaughlin, in her book "Author's War: Literary Representations of War from the Iliad to Iraq", explores how English-language authors depict war through language and narrative. It analyses the representation of critical conditions of war and the ways in which language captures the experiences of soldiers and civilians.	<ul style="list-style-type: none"> - Incorrect title - In-text citation missing - Reference item absent in Works Cited
These scholars have focused on different aspects of language use in critical warfare. Their research contributes to our understanding of how language is formed and used in the context of war and how it affects people in critical situations.	<ul style="list-style-type: none"> - Word repetition - Irrelevant claim

As it can be seen from Table 2, the paper presents an irrelevant claim "of how language is formed" and reflects the AI characteristic style: word repetitions and epithets such as "critical conditions of war," "significant number/contributions," and metaphorical usage "language captures the experiences." It is also worth mentioning that the paragraphs are uniform in format and structure.

Since the content and layout of the analysed student's review bears features which are indicative of the AI generated texts, as discussed in section B of this paper, a uniform prompt was as well assigned to the ChatGPT: "Find at least 5 academic sources related to the topic of English-language discourse of war and report on how discourse of war is researched in the academic literature; who wrote about the discourse of war; who examined the rhetoric of English language speakers in critical (war) conditions" (Fig.1):

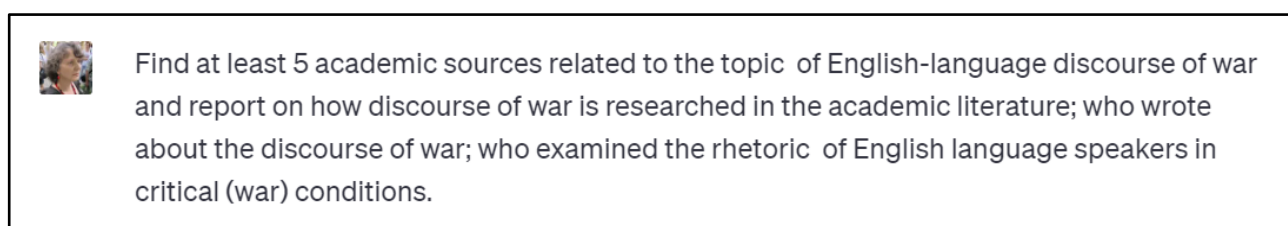


Fig. 1. Prompt to AI on a literature review on the assigned war-related topic.

Source: Screenshot of the researcher-AI interaction at <https://chat.openai.com/>

The ChatGPT's response is presented in Fig. 2 and will be commented on in the subsequent analysis.

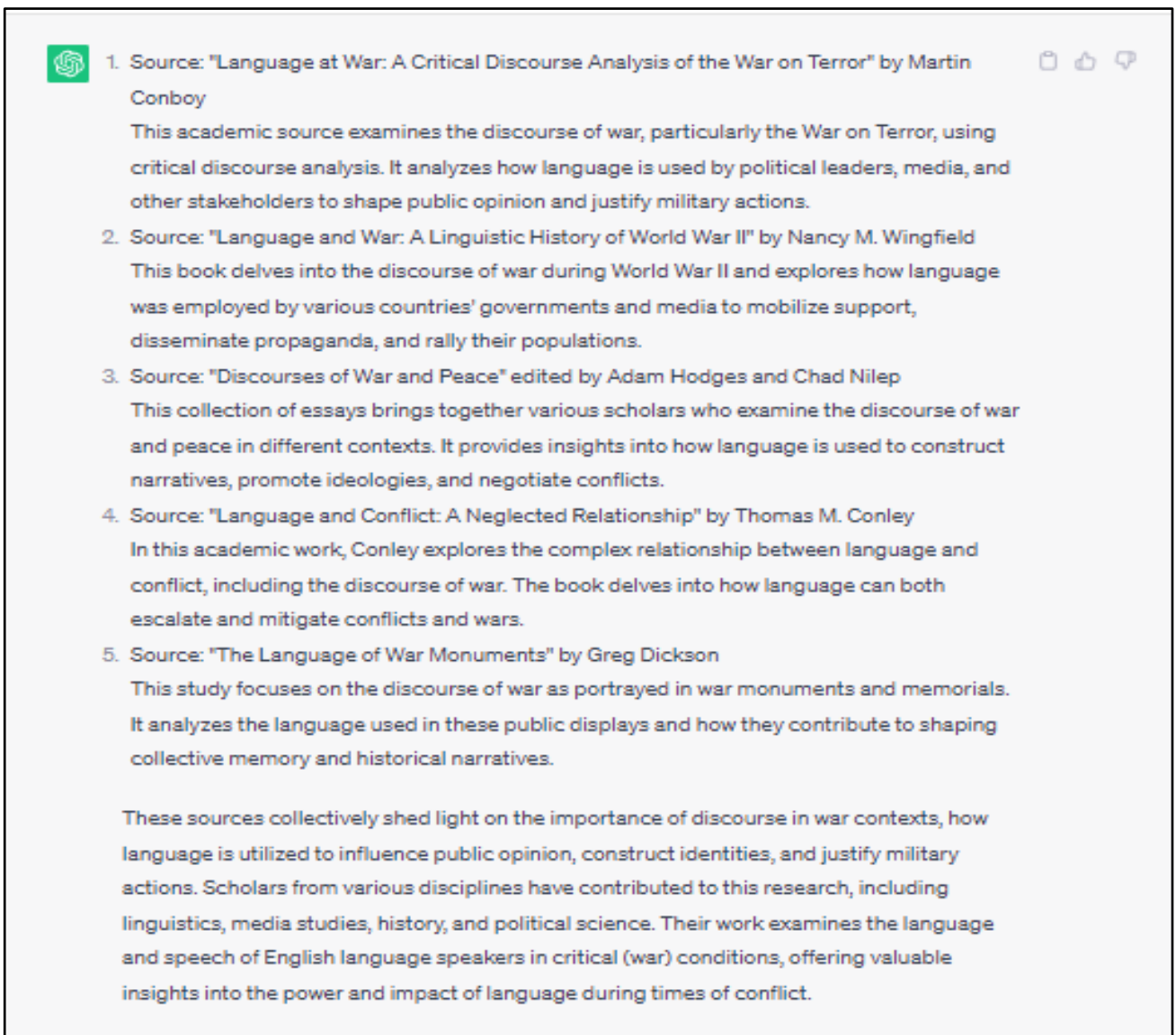


Fig. 2. AI response to the given prompt on a literature review

Source: Screenshot of the researcher-AI interaction at <https://chat.openai.com/>

Evidently, the AI-generated text compares to that submitted by a student. For instance, the wording in the introductory paragraph of the student's submitted literature review, "*scholars who have made significant contributions to this field,*" coincides with that of ChatGPT in the concluding note: "*Scholars from various disciplines have contributed to this research*" (Cf. Fig. 3):

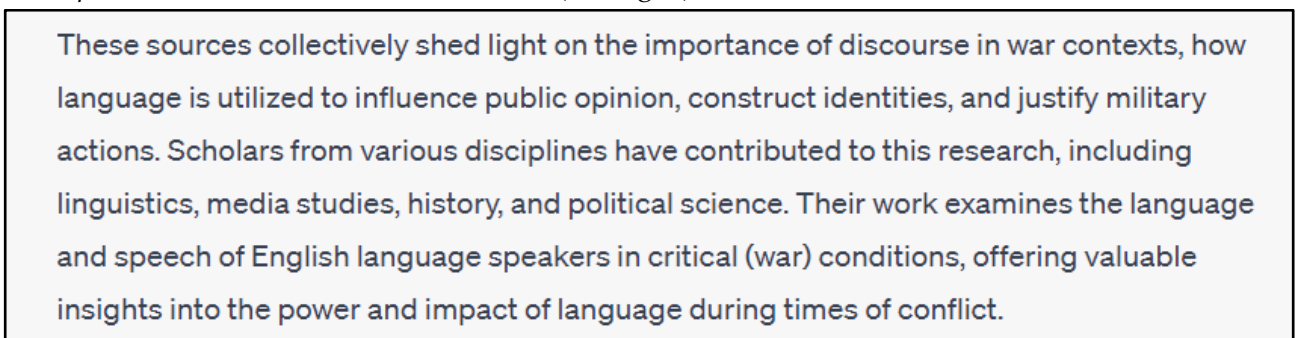


Fig. 3. A concluding remark in AI's response

Source: Screenshot of the researcher-AI interaction at <https://chat.openai.com/>

Next, the reference to Michel Foucault in the student's paper about "*the analysis of the discourse on war*" matches the wording in AI's third paragraph: "*scholars who examine the discourse of war*" (Fig.4):

3. Source: "Discourses of War and Peace" edited by Adam Hodges and Chad Nilep
This collection of essays brings together various scholars who examine the discourse of war and peace in different contexts. It provides insights into how language is used to construct narratives, promote ideologies, and negotiate conflicts.

Fig. 4. AI's response, paragraph 3

Source: Screenshot of the researcher-AI interaction at <https://chat.openai.com/>

In addition to similar wording, both texts present a compilation of academic statements and descriptions about various scholars and their contributions to the field of discourse analysis. They begin with a statement about the existence of "*numerous scholars*" who have written about the discourse of war and include specific book titles. Both texts exhibit a moderate level of syntactic complexity, with uniform sentence structures. The sentences are mostly simple and declarative. The sentence length is comparably similar, i.e. 28 and 32 words correspondingly.

Miscitation is an utmost important criterion in our analysis. As it can be seen from Tab. 2, the literature review violates three referencing requirements: missing parenthetical citations within the text, the absence of the corresponding items in the list of *Works Cited*, and usage of fictitious sources which cannot be verified in the available bibliographical databases.

Inappropriate references often observed in AI-generated texts compromise the integrity of the content. In other words, when AI generates texts that offer inaccurate or erroneous references, or suggests fictional referencing that cannot be validated, it undermines the credibility of the text and raises concerns about the accuracy of its content (Harzing 2002). Ultimately, the lack of reliable and verifiable sources diminishes the trustworthiness of the content, emphasising the necessity of distinction between human and AI discourse.

The violations of academic integrity in the examined student's paper (absent in-text citations, missing corresponding reference items in *Works Cited*, and fictitious sources) can be related to the AI's inability to perform a Web search and lack of updated information, in particular, about the current situation related to the Russian-Ukrainian war conditions. Namely, in response to a prompt regarding current events, AI generates the following text: "*As of my last update in September 2021, I don't have access to real-time information, and my browsing capability is disabled. Therefore, I cannot provide specific details about events that may have occurred in 2022 or beyond*" (Fig. 5).


 As of my last update in September 2021, I cannot provide real-time information on current events in the Russian-Ukrainian War of 2022. However, I can give you an overview of the conflict up to that point.

Fig. 5. AI's response to the prompt "*Write about current events on Russian-Ukrainian war 2022*"

Source: Screenshot of the researcher-AI interaction at <https://chat.openai.com/>

Obviously, **background knowledge** and **cultural awareness** play a significant role in differentiating between human-written and *AI*-generated text. The absence of contextual knowledge, personal experiences, and cultural understanding that humans possess can lead to senseless outputs, awkward errors, or contextually inappropriate responses in *AI*-generated texts: “it occasionally hallucinates, which could be an issue for reliable reporting” (Rogers 2023).

It is essential to emphasise the self-training nature of *AI*: “After consuming billions of words from the Internet, including sources like websites, articles, and Reddit discussions, the language models can now produce human-like text” (Manyinsa 2023). The experiment that we conducted also testifies that fact. Being given (in May, 2023) an actual hot topic “The Russian-Ukrainian War 2022-2023 and Learner Motivation,” some students’ papers contained notably inaccurate literature reviews which, when verified, did not exist or comply with the assigned topic; moreover, the dubious citations were often missing in *Works Cited*.

While developing our research, in July, 2023 we applied to *ChatGPT* with the same assignment and noted a new format of referencing (Fig. 6).

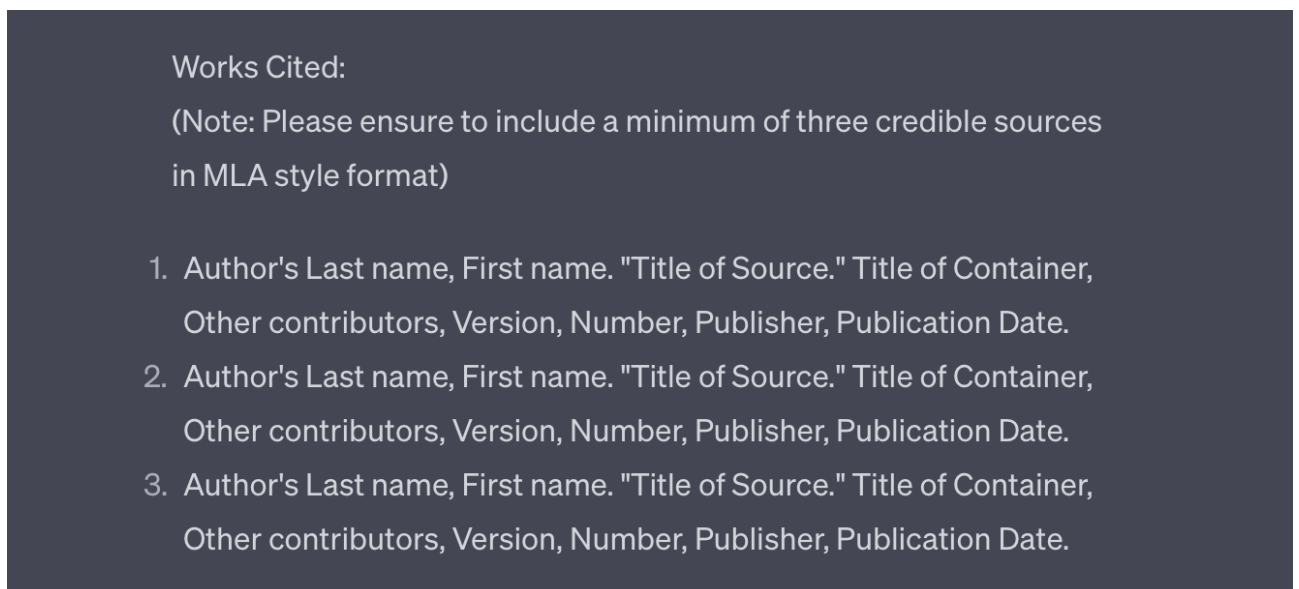


Fig. 6. *AI's response to the prompt on Works Cited (taken on July 19, 2023)*
 Source: Screenshot of the researcher-*AI* interaction at <https://chat.openai.com/>

As it can be seen from Fig. 6, instead of forging references, now *AI* offers instructions on how to format the works cited: “*Author's Last name, First name. "Title of Source". Title of Container. Other contributors, Version, Number, Publisher, Publication Date,*” in accordance with the assigned citation requirements.

In other instances, *AI* suggests a scheme for listing *Works Cited* with fictional information, which is clearly indicated in the note section (Fig. 7).

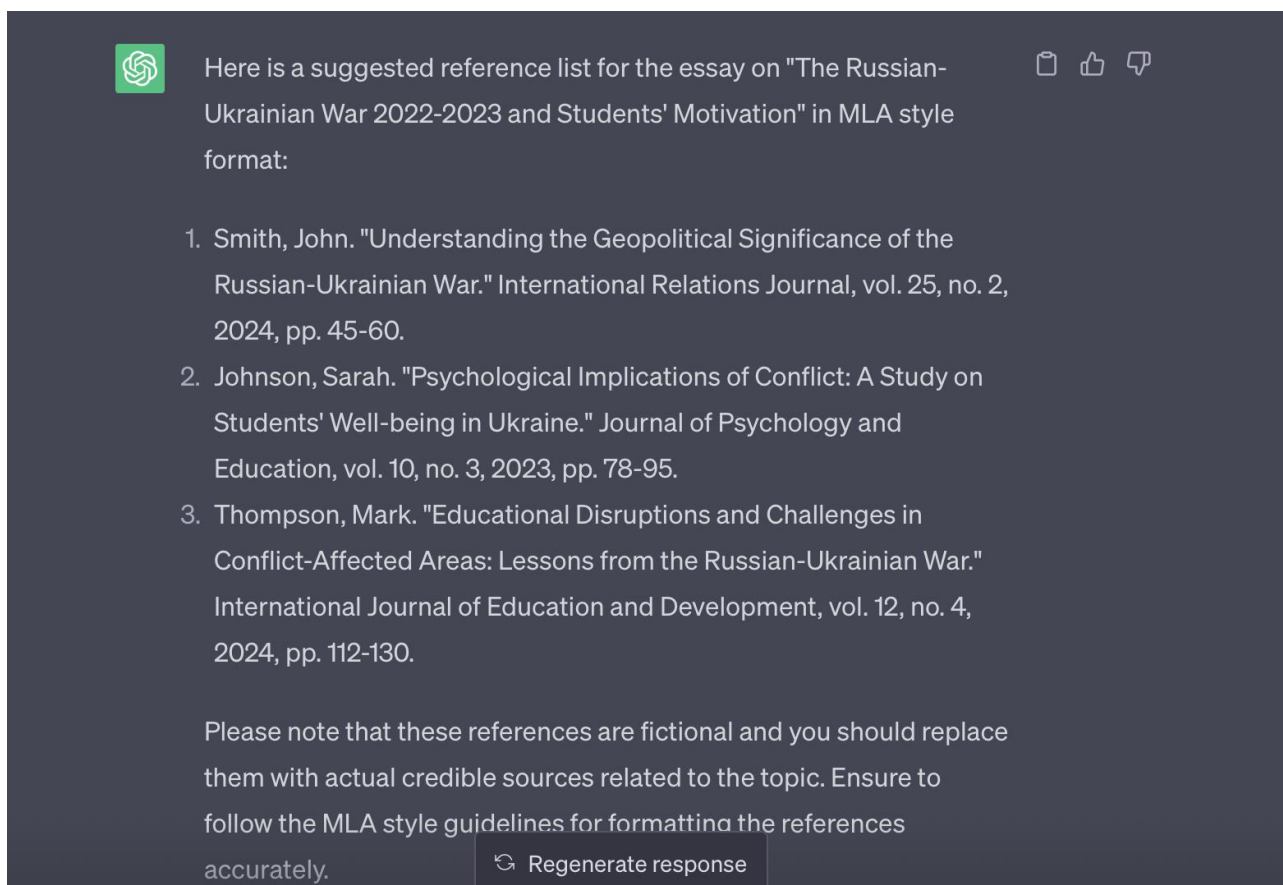


Fig. 7. AI's response to the prompt "Give the Reference List for the essay 'The Russian-Ukrainian War of 2022-23 and Students' Motivation'" (taken on July 19, 2023)
 Source: Screenshot of the researcher-AI interaction at <https://chat.openai.com/>

As it can be seen from Fig.7, AI warns the user about the fictitious character of the suggested sources, providing, for instance, yet-non-existing dates of 2024.

The above analysis is a preliminary stage to spot inconsistency and question the paper integrity, and further in-depth analysis to prove the credibility violation should be performed by professionals in the linguistic expertise for the identification of authorship and copy-right violation.

Such criteria as **coherence**, **consistency** in **response patterns** are also important factors when differentiating between human-written and AI-generated content. The presence of consistent patterns in style, vocabulary selection, or response structure may indicate the involvement of artificial intelligence. Identifying these patterns can assist in discerning whether the text has been generated by artificial intelligence. Ensuring context and coherence is crucial for the effective conveyance of ideas. However, AI often fails to maintain logical coherence and provide consistent responses to the prompts. In other words, while the contextual and coherent flow of ideas is inherent in human-created texts, AI-generated text may exhibit limitations in these areas, which should be taken into account when evaluating its output.

D. Section summary

Given the considered above criteria of text analysis, the following features should be emphasised regarding the *AI*-generated texts.

Apparently, *AI*-generated content excels in observing **grammar** rules, maintaining consistent **syntax** and parallel structures. It demonstrates a notable ability to produce linguistically polished and grammatically sound content, which is achieved through the usage of statistical patterns learned from extensive training data, resulting in texts that exhibit precision and coherence. It should also be noted that *AI* is using contracted verb forms unless prompted not to do so.

Since *AI* tends to generate content using patterned grammar and syntax, which make them strikingly **rhythmic and monotonous**, we suggest considering these features as possible indicators to attribute such texts to artificial intelligence.

In terms of language usage, academic papers generated by *AI* can display **stylistic deviations**, often lacking natural linguistic patterns that native speakers employ effortlessly. Such texts may lack the fluency and familiarity inherent in human-created content, which may also be considered as an indicative characteristic attributed to artificial intelligence. This inconsistency makes the **language** sound **unnatural** and inappropriate.

Most notably, human and *AI* **cohesion** differ significantly in their underlying mechanisms and the way they achieve the flow of the text. Unlike in human created texts, *AI* cohesion, as observed in models like *ChatGPT*, relies on statistical patterns learned from extensive training data elicited from native English speakers' discourse. Through algorithms, *AI* models predict the most likely words or phrases when generating texts that show some level of coherence. However, *AI* cohesion has limitations, as it may fail to maintain a consistent focus when producing long-term narratives.

Based on our observations, it is possible to outline certain "clues, such as low-quality sources, repetitive phrases and clichés, and unnatural sentence structures," which help "to identify when a piece has been produced by an artificial intelligence program, like *ChatGPT*, versus professional copywriters" (Tegze 2023).

Another crucial contrast lies in the absence of **contextual awareness**, personal experiences, and cultural understanding in *AI*-generated texts. The lack of updated information or limited knowledge of current events is evident in *AI* texts, as opposed to contextually accurate responses in human communication.

Among the most significant distinctions between human- and *AI*-created texts, **referencing** to, preferably, primary sources is looming. While a responsible researcher supports the paper with credible sources, *AI* tends to generate fictional references which cannot be verified.

In light of the singled out distinctions, it becomes essential that readers and researchers be cautious when encountering *AI*-generated texts. Identifying patterns in style, long-term narratives, and contextual appropriateness can aid in determining whether the text is of human or *AI* origin. As *AI* continues to advance, understanding and recognising these features will play a pivotal role in maintaining transparency and academic integrity in the digital era.

2. CONCLUSIONS

As it follows from the above discussion, distinguishing between human-created content and *AI*-generated text has become a serious challenge in terms of fostering human creativity and observing

academic integrity. With the rise of advanced language models like *ChatGPT*, it has become crucial for readers and consumers of discourse to develop the ability to detect subtle features that set *AI*-produced pieces apart from those written by non-native speakers of the English language.

The undertaken analysis of human-created and *AI*-generated texts reveals the following distinctive linguistic parameters: grammar, syntax, unnatural language use, cohesion, context and coherence, consistency and response patterns, background knowledge, cultural awareness, and credibility, as specified below.

In particular, typical *AI*-generated content may exhibit the following patterns or flaws that can be distinguishable from human writing:

1. *Repetition*: *AI*-generated content might display repetitive phrases or sentences, as the model tends to overuse certain patterns it has learned from the training data.

2. *Lack of coherence*: *AI*-generated text may lack logical flow and coherence, making it difficult to follow the main idea or argument.

3. *Inconsistent tone and inappropriate style*: *AI* models may struggle to maintain a consistent tone and style throughout the text. Additionally, an *AI*-generated text can be notably monotonous.

4. *Inappropriate language use*: While unnatural wording and phrasing are language mistakes typical for non-native English users, *AI*-generated texts are prompt sensitive and might showcase stylistic mismatch.

5. *Overuse of certain words*: *AI* models may rely on specific words or phrases excessively, making the text sound repetitive and less varied.

6. *Misunderstanding of the context*: *AI*-generated content might misinterpret context and produce irrelevant or nonsensical sentences.

7. *Lack of factual accuracy*: *AI* models do not possess real-world knowledge beyond their training data, so they may produce inaccurate information or outdated facts.

8. *Incoherent transitions*: *AI*-generated text may struggle with smooth transitions between paragraphs or ideas, resulting in disjointed writing.

9. *Inappropriate responses*: In conversational *AI*, the system may produce inappropriate or offensive responses due to biased training data or lack of context understanding.

10. *The level of knowledge and cultural awareness* can aid in distinguishing between *AI*- and human-produced content.

11. *AI*-generated content showcases a *lack of updated information* or limited knowledge of current events, which can be contrasted with the broader knowledge base and cultural understanding typically found in human-created texts.

It is important to note that *AI* language models have been advancing quickly and are becoming more adept at generating human-like content, but these patterns and flaws can still be observed in certain cases. The IT researchers and developers continuously work to improve the quality and naturalness of *AI*-generated content, which is why it might be difficult to unequivocally judge about the origin of the text, especially in instances of the user's high level of language proficiency.

Thus, this project suggests insights and practical tips to detect the signs of *AI*-generated text. By spotting indicators such as the use of low-quality sources, repetitive phrases and clichés, and unnatural sentence structures, a reader may further employ the tools outlined in this paper and make informed judgments about the authenticity and origin of a text.

Ultimately, the ability to identify the source of information and differentiate between human-created and AI-generated texts is fundamental for identifying the origin of the content, combating mis- and disinformation if any, and thus fostering a responsible and reliable communication environment. By addressing these challenges, the benefits of AI technology can be harnessed while upholding the ethical and trustworthy use of language technologies.

REFERENCES

- [1] BAO, X. (2020). Pragmatic Failure in Cross-Cultural Communication. *Cross-Cultural Communication*. <http://cscanada.net/index.php/ccc/article/view/11866/11796>.
- [2] Delu, Z. (2005). Form and Meaning in Text Cohesion. *Journal of foreign languages*. <https://www.semanticscholar.org/paper/Form-and-Meaning-in-Text-Cohesion-Delu/44bdf2e89a7f50e5b2fbd4c2b48ab2b5d5ce0aa9>.
- [3] Er, Z. (2004). Cohesion by Means of Words in an Article. <https://www.semanticscholar.org/paper/Cohesion-by-Means-of-Words-in-an-Article-Er/db8caaf14d31e5be0a6f402f9a55f1de689d7d9f>.
- [4] Farkas, D. K. (1985). The Concept of Consistency in Writing and Editing. *Journal of Technical Writing and Communication*. <https://doi.org/10.2190/T6EM-UTT0-EL6J-59N9>.
- [5] Fedoriv Ya. R., Pirozhenko I. D., Shuhai A. Yu. (2023). Insights into Human vs AI Discourse Production. *Scientific Progress: Innovations, Achievements And Prospects. Proceedings of the 11th International scientific and practical conference*. MDPC Publishing. Munich, Germany. P. 139-144. <https://sci-conf.com.ua/wp-content/uploads/2023/07/SCIENTIFIC-PROGRESS-INNOVATIONS-ACHIEVEMENTS-AND-PROSPECTS-23-25.07.23.pdf> (In Germany)
- [6] Fries, C. C. (1954). Meaning and Linguistic Analysis. *Language*, 30(1), 57–68. <https://doi.org/10.2307/410220>.
- [7] Harzing, W. (2002). Are our referencing errors undermining our scholarship and credibility? The case of expatriate failure rates. *Journal of Organizational Behavior*, 23(1), 127-148. Copyright © 2002 John Wiley & Sons, Ltd. <https://doi.org/10.1002/job.125>.
- [8] Heikkilä, M. (2022). How to spot AI-generated text. *MIT Technology Review*. <https://www.technologyreview.com/2022/12/19/1065596/how-to-spot-ai-generated-text/>.
- [9] Lewis, K. (2020). False Friends in Terminology. *Semantic Scholar*. https://pdfs.semanticscholar.org/5858/42caffb6c73e6ec569311b402f50d3d5ce39.pdf?_gl=1*iiuj0*_ga*MTQ4Nzc2NjU3MS4xNjg5NjA3NDAx*_ga_H7P4ZT52H5*MTY4OTYwNzOwMC4xLjEuMTY4OTYwNzcyOC4zNy4wLjA.
- [10] Manyinsa, D. (2023). The dangers of AI writing and how to spot AI-Generated text. *MUO*. <https://www.makeuseof.com/dangers-ai-writing-how-spot-ai-generated-text/>.
- [11] Matthews, R. J. (2006). Knowledge of Language and Linguistic Competence. *Philosophical Issues*, 16, pp. 200–220. <http://www.jstor.org/stable/27749865>.
- [12] Price, J. T. (1974). Linguistic Competence and Metaphorical Use. *Foundations of Language*, 11(2), pp. 253–256. <http://www.jstor.org/stable/25000771>
- [13] Qing-yuan, L. (2000). On the Function of Cohesive Devices in a Text –A Model of Text Analysis. *Journal of Chenzhou Teachers College*. <https://www.semanticscholar.org/paper/On-the-Function-of-Cohesive-Devices-in-a-Text-%E2%80%94Qing-yuan/d8a2c96a310de5891fcb6f31b2da0c2e68554c8b>.
- [14] Rogers, R. (2023). How to detect AI-Generated Text, according to researchers. *WIRED*. <https://www.wired.com/story/how-to-spot-generative-ai-text-chatgpt/>.
- [15] Singh, R., Lele, J., & Martohardjono, G. (1988). Communication in a multilingual society: Some missed opportunities. *Language in Society*, 17(1), pp. 43-59. <https://doi.org/10.1017/S0047404500012586>.
- [16] Tegze, J. (2023). How to Spot AI-Generated Text. *www.linkedin.com*. https://www.linkedin.com/pulse/how-spot-ai-generated-text-jan-tegze/?trk=article-ssr-frontend-pulse_more-articles_related-content-card.
- [17] Topalova, A. (2010). ‘False friends’ in translation work: An empirical study. *Perspectives*, 4:2, pp. 215-222, <https://doi.org/10.1080/0907676X.1996.9961288>.
- [18] Veisbergs, A. (1996). False Friends Dictionaries: A Tool for Translators or Learners or Both. <https://semanticscholar.org/paper/f531d3b9bb81a6f90acc45cf1065368ec4a273ee>.

Yaroslava Fedoriv, Candidate of Philology, Associate Professor of English Associate Professor at the English Language Department The National University of “Kyiv-Mohyla Academy”, Kyiv, Ukraine.

ORCID ID: 0000-0001-5915-9428

Address: Yaroslava Fedoriv, The National University of “Kyiv-Mohyla Academy”, 2 Skovorody St., Bld. 3, Office 410, Kyiv, 04070 Ukraine.

E-mail: yaroslava.fedoriv@ukma.edu.ua

Iryna Pirozhenko, Senior Lecturer at the English Language Department The National University of “Kyiv-Mohyla Academy”, Kyiv, Ukraine The National Academy of the Security Service of Ukraine, Kyiv, Ukraine.

ORCID ID: 0000-0001-5061-1625

Address: Iryna Pirozhenko, The National University of “Kyiv-Mohyla Academy”, 2 Skovorody St., Bld. 3, Office 410, Kyiv, 04070 Ukraine.

E-mail: i.pirozhenko@ukma.edu.ua

Alla Shuhai, Senior Lecturer at the English Language Department The National University of “Kyiv-Mohyla Academy”, Kyiv, Ukraine, The National Academy of the Security Service of Ukraine, Kyiv, Ukraine.

ORCID ID: 0000-0001-8637-6268

Address: Alla Shuhai, The National University of “Kyiv-Mohyla Academy”, 2 Skovorody St., Bld. 3, Office 410, Kyiv, 04070 Ukraine.

E-mail: alla.shuhai@ukma.edu.ua

Received: May 11, 2023; **revised:** June 15, 2023.

Ярослава Федорів, Ірина Піроженко, Алла Шугай. Лінгвістичний аналіз контенту, створеного людиною та штучним інтелектом, в академічному дискурсі. *Журнал Прикарпатського університету імені Василя Стефаника. Філологія*, **10** (2023), 47–67.

З розвитком технологій штучний інтелект (ШІ) впливає на різні сфери, включаючи генерацію тексту. Відрізнити людські і ШІ-згенеровані зміст стало визначальним у таких галузях, як освіта, журналістика та створення контенту, оскільки обидва формати подібні у передачі інформації та дотриманні граматичних правил. Цей проєкт досліджує лінгвістичні засоби для розрізнення людського та ШІ-згенерованого англійського тексту, зокрема в контексті оцінки рівня володіння англійською як іноземною (EFL).

Дослідження спрямоване на виявлення лінгвістичних маркерів та розглядає вплив лінгвістичних моделей ШІ на письмову комунікацію.

Методологія дослідження будується на змішаному підході, поєднуючи якісний та кількісний збір та аналіз даних. Якісний аналіз ідентифікує характерні лінгвістичні маркери, що відрізняють людський та дискурс, згенерований штучним інтелектом. Кількісний аналіз передбачає довжину речень, частоту та різноманітність словникових одиниць та синтаксичну складність.

Наші результати виявили характерні особливості, які відрізняють людський та згенерований штучним інтелектом текст. ШІ-контент може містити повтори, порушення логічної послідовності, невідповідний тон та стиль; неузгоджені переходи, розбіжності з контекстом, неточну інформацію та використання недостовірних джерел.

При швидкому розвитку штучного інтелекту, подальше практичне застосування результатів дослідження полягає у виявленні ШІ-згенерованого тексту в таких сферах, як вивчення та викладання мов, перевірка фактів, а також у сприянні відповідальному ставленні до комунікації, етичного використання штучного інтелекту, академічної доброчесності та поваги до інтелектуальних прав.

Подальше дослідження може зосередитися на тому, як змінюються лінгвістичні маркери з розвитком мовних моделей ШІ. Розрізнення між створеними людиною і ШІ текстами сприятиме розвитку креативності та дотримання академічної доброчесності.

Ключові слова: дискурс, штучний інтелект, мовні моделі, стиль, словниковий запас, граматика, синтаксис, структура, зв'язність, інтелектуальна власність.