

Do history teachers recognize the handwriting of artificial intelligence

* This is a translation of an article published in Croatian language.

HOW TO CITE (APA)?

Hajdarovic, M. (2023). Prepoznaju li učitelji povijesti rukopis umjetne inteligencije [Do history teachers recognize the handwriting of artificial intelligence]. *Poucavanje povijesti*, 11(2), 85–92.

ABSTRACT: During August 2023, a small online survey was conducted on whether Croatian History teachers recognize the handwriting of artificial intelligence. Can we distinguish a text written by artificial intelligence from a text written by an expert or an expert with the help of artificial intelligence? What criteria do we use to identify authors? The paper presents the results of the research.

KEY WORDS: artificial intelligence, digital humanities, teaching history

History teachers happily welcomed another challenge in the classroom

Happy ChatGPT Day! Have you celebrated a happy first birthday? Most of the public became aware of ChatGPT on November 30, 2022. Since that day, we have regularly followed clickbait headlines about artificial intelligence—the headlines in the media range from scandal to utopia to dystopia. Of course, a frequent question in the teaching population, also asked in ChatGPT, was whether artificial intelligence would replace teachers. This is the right place to repeat primary material on artificial intelligence.

General artificial intelligence, one that is conscious and independent, has yet to exist, and we have no accurate indication of when it might appear. Jimmy Wales, the founder of Wikipedia, stated at the end of November 2023 that we might expect this in fifty years. What we currently have are applications that use generative artificial intelligence. Based on the loaded data, it is a programmed artificial intelligence that knows how to predict which string of characters or words could be a good answer to a given question. Generative artificial intelligence imitates learning and imitates conversation. The result or output of generative intelligence is newly generated text or multimedia.

It is interesting to observe that not only geeks and interested individuals but also organizations, institutes, politicians and governments jumped into conversations, planning and limiting artificial intelligence. The possibilities of use and abuse are significant. We can read numerous comments about how artificial intelligence will speed up our daily activities or repetitive tasks that take up our time. Moreover, a good number of teachers are carried away by particular possibilities. However, if we follow artificial intelligence education, we will also see numerous limitations and inaccuracies. For example, there is a very high chance that only about 60% of the text in the generated text will be factually correct; that is, there is a chance that we will have factual problems in every text generated by AI. This could be much fun for historians. And even more for our students. Did you know that elementary school students already have at least ChatGPT installed on their cell phones, and they recently received verbal inquiries and gave answers? However, they do not even need ChatGPT or a similar application to provide brief history information. Even the popular Snapchat can, with the help of its internal chatbot, give a quality answer in Croatian (or other languages) to a factual question from history—enough to answer a question at least in elementary school.

History as a teaching subject has its methodological specifics regarding the work with sources and the tasks we give our students. How will our assignments, projects and research assignments that we regularly give our students work? Fortunately, the Croatian artificial intelligence language still needs to be improved, and artificial intelligence applications usually have a minimal database of Croatian national history. Nevertheless, as we could already read in the previous issue of Teaching History magazine, if we prompt ChatGPT, Bing or Bard about world history, the answer will be pretty good. The answers can be even better by carefully using the application and skillfully asking questions.

Research on the recognition of the authorship of a historical text

In August 2023, a small online survey was conducted on whether Croatian History teachers recognize the handwriting of artificial intelligence. Can we distinguish a text written by artificial intelligence from a text written by an expert or an expert with the help of artificial intelligence? What criteria do we use to identify authors? Twenty respondents completed the online questionnaire. They were offered three answers to three questions. The texts were answers to the questions:

- Who discovered America?;
- What arguments support the claim that people landed on the moon?; and
- Evaluate whether dropping atomic bombs was necessary to end the war with Japan.

Plot twist – the authors of all nine application answers are artificial intelligence!

In the initial question, 75% of respondents answered that they had never used generative artificial intelligence in their previous work. Only 15% rarely use it, and only 10% often. Respondents in 60% of cases do not know if their students use generative artificial intelligence. A quarter of respondents know that students use it, and 15% are not sure.

Questions (or prompts) to artificial intelligence applications are set on August 18, 2023, in the same way: Answer the question with a text length of 70 to 100 words at the school level. The tables show the responses of generative artificial intelligence that were evaluated by the respondents.

Table 1. Who discovered America?

ChatGPT 3.5	America was discovered by Christopher Columbus, a Genoese navigator, under the Spanish flag. On October 12, 1492, Columbus arrived on the island of Guanahani in the Caribbean Sea, which marked the beginning of European exploration and settlement of America. His expedition paved the way for European colonization of the New World, even though there were already indigenous communities in the area. Columbus's expeditions had a significant impact on world history and the exchange of goods, ideas and diseases between the Old and New Worlds.
Bing AI	Christopher Columbus first set foot on American soil in 1492. This event is taken as the date of the discovery of America, although today many historians accept that the Vikings visited America earlier. Much later, the continent was named after another explorer, Amerigo Vespucci. Christopher reached America with three Spanish ships: Nina, Pinta and Santa Maria. He thought he had reached India and that is why Native Americans are called Indians. During his life, he was in America twice more, he also visited the South American continent, but he was never in North America.
Bard	Christopher Columbus discovered America in 1492. He was an Italian sailor who sailed for Spain. Columbus believed that one could reach India by sailing west. He sailed from Spain on August 3, 1492 with his fleet of three ships, Nina, Pinta and Santa Maria. After 36 days of sailing, Columbus reached the island of San Salvador, located in the Bahamas archipelago. Columbus thought

	he had reached India, so he called the locals Indians. Columbus sailed to America three more times and discovered many new islands. His discoveries had a great impact on the world, because they opened the way for the colonization of America by Europeans.
--	--

Table 2. What arguments support the claim that people landed on the moon?

ChatGPT 3.5	The claim that people really landed on the moon is supported by a number of arguments. First, there is extensive documentation, photographs and videos taken during the Apollo missions that clearly show the astronauts on the moon. Second, space probes have imaged landing modules and equipment remnants on the lunar surface. Third, hundreds of thousands of people, including scientists, engineers, and technicians, worked on these missions, which would be difficult to organize under fraud. Finally, independent scientific data, such as rock samples brought back from the Moon, confirm the authenticity of the missions and the presence of humans on the lunar surface.
Bing AI	There is plenty of evidence to support that humans did indeed land on the moon. NASA sent six missions to the moon between 1969 and 1972, and each mission was equipped with cameras, gauges and other equipment that collected data about the moon. Astronauts brought back rock and soil samples from the Moon, and these samples were later studied in laboratories around the world. In addition, there are many photos and videos showing astronauts walking on the surface of the Moon, as well as the footprints they left on the surface. All this information points to the fact that people did land on the moon.
Bard	<p>There are many arguments to support the claim that humans did indeed land on the moon. Some of the most common arguments are:</p> <ul style="list-style-type: none"> * There is clear evidence of human presence on the moon, such as footprints, footprints and equipment remains. * The moon landings were broadcast live on television, and billions of people watched them. * The pictures and videos of the moon landing are very realistic and could not have been faked at the time. * The Soviet Union was in a space race with the US and had every motive to disprove the moon landing, but never did. <p>There are also a smaller number of people who believe that the moon landings were faked, but there is no evidence to support their claims.</p>

Table 3. Evaluate whether dropping atomic bombs was necessary to end the war with Japan.

ChatGPT 3.5	The dropping of atomic bombs on Hiroshima and Nagasaki in 1945 is a controversial topic. Some believe that this hastened the capitulation of Japan and saved the lives of Allied soldiers and Japanese. Others argue that Japan was already weakened and the USSR had already attacked Japan, which contributed to the capitulation. Regardless of the reasons, the bombs caused enormous human casualties and long-lasting consequences. Although some
-------------	---

	saw the bomb as necessary, others believed that alternative approaches could have produced the same result without so much destruction.
Bing AI	The dropping of the atomic bombs on Hiroshima and Nagasaki was a controversial issue in history. Some believe that dropping the atomic bombs was necessary to end the war with Japan, while others believe that it was not necessary. Many historians argue that dropping the atomic bombs was inhumane, causing too much destruction and killing too many innocent people, including children. However, others argue that dropping the atomic bombs was necessary because the war could last for years and cost many more lives. In any case, the dropping of the atomic bombs on Hiroshima and Nagasaki remains one of the most controversial events in history.
Bard	<p>Dropping the atomic bombs on Hiroshima and Nagasaki was one of the most controversial decisions in history. Some historians believe that the bombing was necessary to prevent a long and bloody war, while others believe that the bombing was unnecessary and cruel.</p> <p>There are a number of arguments for and against dropping bombs. On the one hand, the dropping of bombs led to the death of hundreds of thousands of people and caused enormous damage to two cities. On the other hand, the dropping of the bombs led to the Japanese capitulation and the end of the war.</p> <p>In the end, deciding whether the bombing was necessary is complex and there is no easy answer. However, it is important to remember the victims of the bombing and learn from history to avoid similar tragedies in the future.</p>

Image 1. In ChatGPT's answer to Who Discovered America? 65% of respondents (40% experts with the help of UI) recognize the (non-existent) author's hand of a real expert.

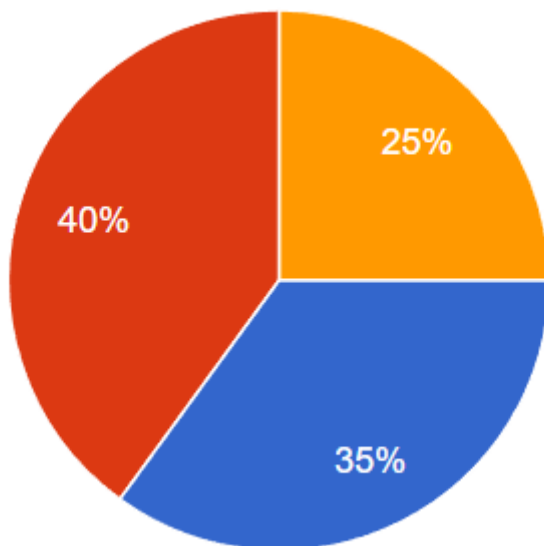


Table 4. Criteria for determining the authorship of the answer to the question who discovered America (ChatGPT).

<p>the author is an expert</p>	<p>way of writing/expression; grammar analysis; compliance with the Croatian language; dominance of one style; words, terminology; I use artificial intelligence; cause and effect</p>
<p>the author is an AI-assisted expert</p>	<p>my own education; it is mostly written precisely and in detail, but there are mistakes in the language; incorrect case forms; real facts and grammatically correct sentences; comparison with the remaining answers to this question; text content; presented facts; language constructions; spelling; I looked at grammar, syntax, whether the sentences were too long or too short and the "tone" of the text; looked at spelling and some peculiarities like inserted brackets; well defined syntax (precise commands) by the user. lesser known or inaccurate historical facts (the name guanahani is rarely used in literature vs. san salvador), occasional incoherence of the text (relationship with the natives?), and often uses pleonasms (although there are none here), absence of cause, unspecified consequences or evaluation</p>
<p>the author is AI</p>	<p>objective presentation of bare facts without using stylistic means; the breadth of information refers to ai; own critical judgment; it is simple, there are no epithets and colloquial language; the text is densely and economically filled with details, which a person would arrange in several sentences; grammatical errors typical of mechanical translation (case mismatch); writing style, grammatical error in the last sentence.; although the text is supported by details, it lacks the author's line</p>

Image 2. In Bing AI's answer to the question Who Discovered America? 55% of respondents think that the author of the text is AI, 25% think that it is an expert who wrote using AI and 20% that the author is an expert.

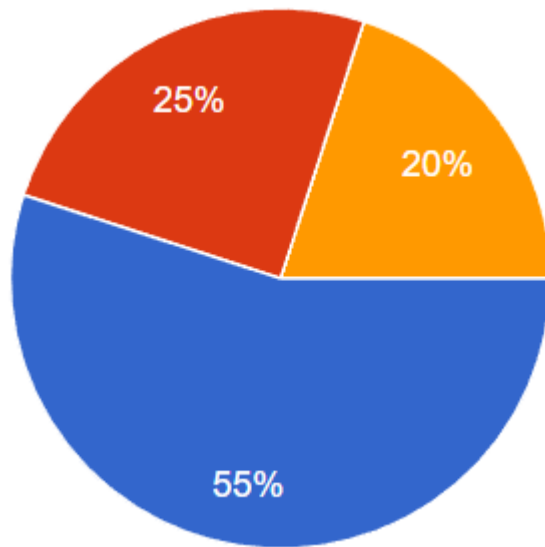


Table 5. Criteria for determining the authorship of the answer to the question who discovered America (Bing AI).

the author is an expert	the language used and a certain subjectivity; the response style refers to a real person; it seemed too extensive, so I checked the text and found it on wikipedia; subjectivity of the text; a factual error typical of superficial human research/comprehension
the author is an AI-assisted expert	way of writing/expression; grammar analysis; own education; own critical judgment; sounds like a wikipedia article; I use artificial intelligence; details; although there are facts in it, the text is flowing and easier to read
the author is AI	it is not written in the Croatian standard; terms from the Croatian and Serbian languages; (in)accuracy of scientific terminology; real facts and combinations of words from different languages; text content, presented facts, language constructions, spelling; correctly written words, terms and language; inconsistency in language; grammar, short and concise sentences that get to the heart of the question. only factual information is presented.; orthography that does not fully correspond to Croatian standardized; factually correct text. ai often answers in Serbian-Croatian language because he does not have a large enough knowledge base for small languages or simply confuses the common history. there is an obvious spelling mismatch

Image 3. In the Bard's answer to the question Who discovered America? even 50% of respondents think that the text was written by an expert.

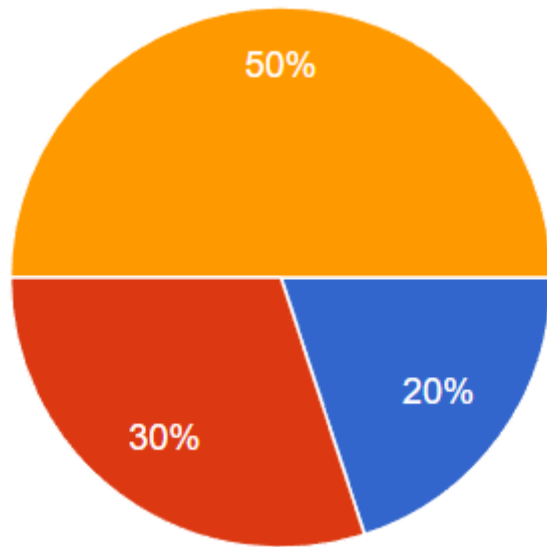


Table 6. Criteria for determining the authorship of the answer to the question who discovered America (Bard).

the author is an expert	written in connected sentences, arising from one another and logically continuing one another; the text is comprehensive, it contains all the answers to the question; real facts and academic style of writing; own critical judgment; the text sounds more humane and lifelike; text content, presented facts, language constructions, spelling; sentence structure (extremely logical and didactic, reminiscent of a school textbook); the text is adapted to students; the fluency of the text, told more like a story with factual data; my vocabulary is purely "humane"
the author is an AI-assisted expert	good ratio of facts and additional explanations of terms; compliance with the Croatian standard language; answer according to the expert's instructions; style and scope of the text; additional unnecessary facts - date of departure, duration of voyage, archipelago of the Bahamas. the important interesting point of Columbus' mistaken belief about the East Indies and the subsequent correction of toponyms (vespucci) is omitted. I think the syntax is very well organized; the data were collected probably on the Internet, but they are arranged by numbers.
the author is AI	way of writing/expression, grammar analysis, own education; because of the name Kristofer, which did not match the rest of the text; I got a similar answer to the question of who, when and how discovered America on chat gpt; I use artificial intelligence; the facts

Image 4. In ChatGPT's answer to the question What arguments support the claim that people really landed on the moon? 35% of respondents think that the author is an expert, and another 15% think that it is a text written by an expert using generative AI.

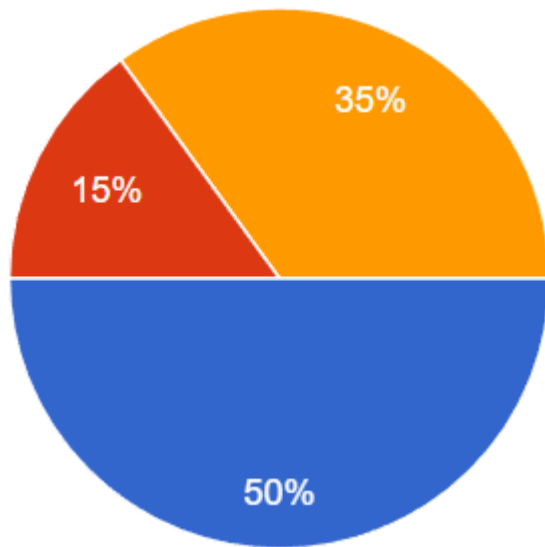


Table 7. Criteria for determining the authorship of the answer to the question what arguments support the claim that people really landed on the moon (ChatGPT).

the author is an expert	way of writing/expression; grammar analysis; comprehensive answer; text style; subjectivity; writing error (tipfeler "argumentata", characteristic of people); language of argumentation (ordinal numbers written with letters followed by a comma); facts and evidence
the author is an AI-assisted expert	own education; quantity and ratio of facts and personal writing style;
the author is AI	non-compliance with the vocabulary and grammar of the Croatian language; incorrect scientific terminology (the causes of rocks are not independent scientific data, but material evidence); real facts and combinations of words from different languages and similar languages; own critical judgment; too general; it is not vital; misspelled argument word; enumeration; text content, presented facts, language constructions, spelling; inconsistency in language use; I use artificial intelligence; slightly incomprehensible text. as if some words do not fit into the context of the sentence or I would not have chosen them myself to explain something. some kind of text is complex, but as if it was written headlong. also, the grammar and spelling are off.; ai often answers in Serbo-Croatian because he does not have a large enough knowledge base for small languages. spelling mismatch is obvious; most likely it is a source of data used by someone to compile the previous text.

Image 5. In Bing's answer to the question What arguments support the claim that men really landed on the moon? 45% of respondents believe that the author of the text is an expert. Another 40% believe that it is an expert who wrote with the help of AI.

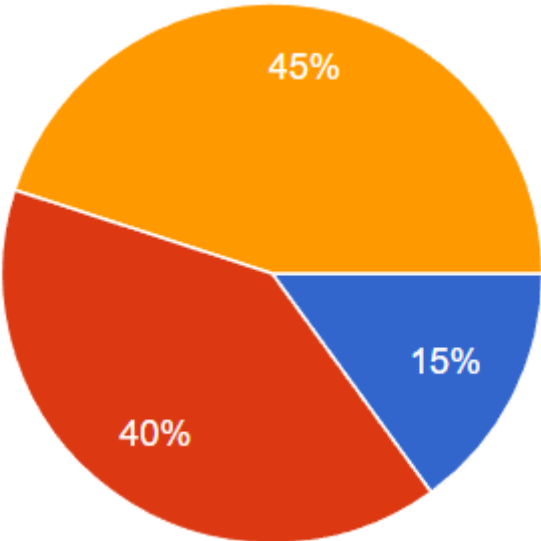


Table 8. Criteria for determining the authorship of the answer to the question what arguments support the claim that people really landed on the moon (Bing AI).

the author is an expert	own education; language and way of presenting arguments; structure of the answer (introduction, central part, conclusion), meaningfulness of the sentences; the sentences are written in an academic, i.e. professional manner; own critical judgment; use of certain expressions; fragment structure (repetition of wording in the first and last sentence), sentence structure (logical, consistent), grammatical error typical of people ("with cameras"); writing style, linguistic consistency and grammatical correctness; emphasizing factual events and scientific study, experiments and measurements
the author is an AI-assisted expert	way of writing/expression, grammar analysis; style, text composition; text content, presented facts, language constructions, spelling; words, scope of writing, sentence construction; I use artificial intelligence; as if the facts generated by ai were inserted into a meaningful text.; abbreviated text, as if someone summarized it with the help of internet data; facts and evidence
the author is AI	it seems to me that a little of everything has been put into the answer from many sources; arrangement in slips. again there is some strange/unknown word for our language

Image 6. In Bard's answer to the question What arguments support the claim that people really landed on the moon? only 10% of respondents think that the author is an expert and another 30% think that the text was written by an expert with the help of AI.

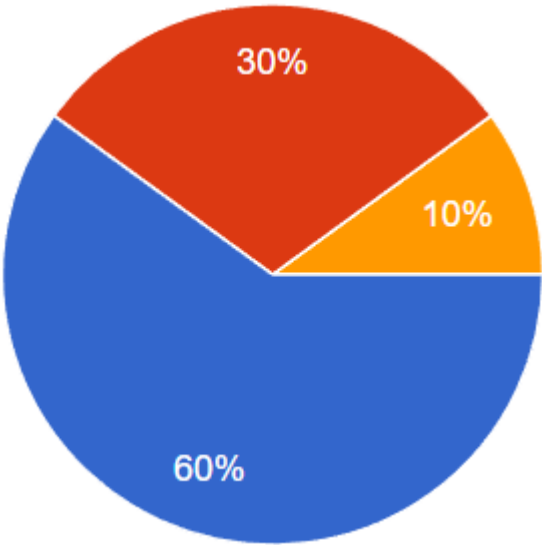


Table 9. Criteria for determining the authorship of the answer to the question what arguments support the claim that people really landed on the moon (Bard).

the author is an expert	own education; own critical judgment
the author is an AI-assisted expert	artificial intelligence was used to provide arguments and they were exhaustively listed, and the rest was written by an expert; appearance of the answer, the author probably instructed me to also write the answer in the notes and correct any incorrect terms; the mention of the Soviet Union seemed to me to be additional information from an expert; text content, presented facts, language constructions, spelling; I use artificial intelligence
the author is AI	way of writing/expression, grammar analysis; systematicity in presenting arguments; composition, design; enumeration of arguments and grammar; text structure (clear division into arguments, use of paragraph marks, stating an opposing opinion), grammatical errors/structures typical of mechanical translation ("false", "they had all the motives"); words, scope of writing, sentence construction; spelling mistakes; and often when he lists some facts he puts each fact in a separate small paragraph. plus, grammar and spelling.; strict enumeration, some errors that suggest that one does not know the language are actually "false"; details; anyone can write this

Image 7. In ChatGPT's answer to write a text in response to the question Rate whether the dropping of atomic bombs was necessary to end the war with Japan? even 70% of respondents successfully concluded that it was a text written by artificial intelligence. An expert is the author for 20% of respondents, and 10% believe that the text was written by an expert with the help of AI.

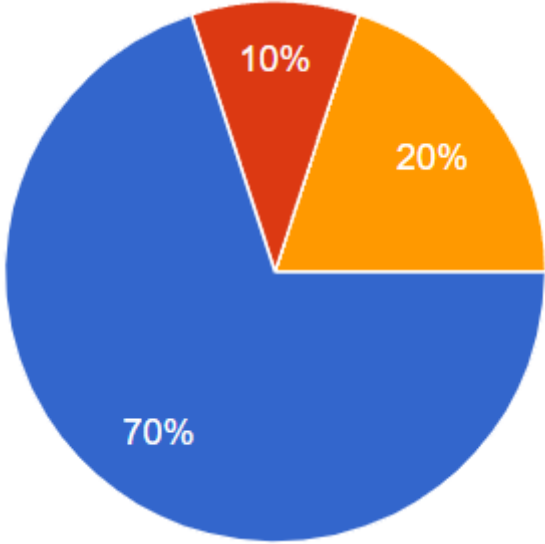


Table 10. Criteria for determining the authorship of the answer to the question whether the dropping of atomic bombs was necessary to end the war with Japan (ChatGPT).

the author is an expert	way of writing/expression, grammar analysis; style, text formatting; I use artificial intelligence; multiperspective approach and evaluation of consequences
the author is an AI-assisted expert	words, scope of writing, sentence construction; arguments
the author is AI	own education; presentation of generally known opinions and ideas on the subject; the appearance of the answer; writing style; own critical judgment; spelling mistakes; text content, presented facts, language constructions, spelling; text structure (objective and "cold" assessment of both sides of the statement, without a direct answer to the question), grammatical features typical of mechanical translation (case mismatch, hiroshima); short sentences, poor grammar and spelling.; informative character of the text, there is nothing human about it; case error again.

Image 8. In Bing's answer to write a text in response to the question Rate whether the dropping of atomic bombs was necessary to end the war with Japan? 35% of respondents believe that the author of the text is an expert. As many as 45% believe that the author is an expert with the help of AI, and only 20% recognized AI's handwriting.

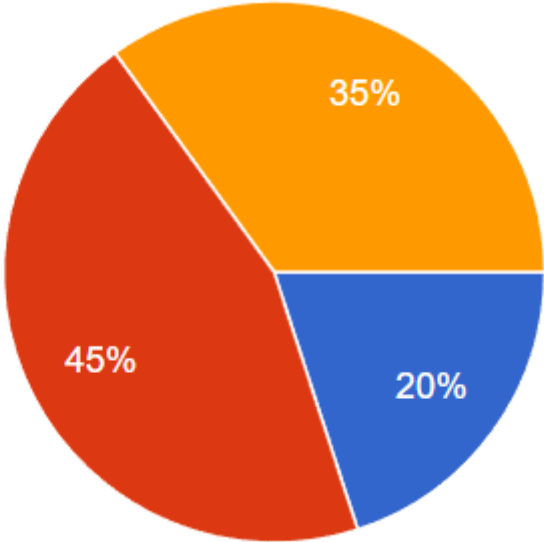


Table 11. Criteria for determining the authorship of the answer to the question whether the dropping of atomic bombs was necessary to end the war with Japan (Bing AI).

the author is an expert	own education; compact answer; text structure (repetition of wording in the first and last sentences), human subjectivity ("innocent people, including children"), use of dialect ("to cost"); words, scope of writing, sentence construction; I use artificial intelligence
the author is an AI-assisted expert	way of writing/expression; grammar analysis; the ratio of objective and subjective; writing style; text formatting; text content, presented facts, language constructions, spelling; it sounds to me as if some sentences from the previous text were reformulated to sound nicer.; the conclusion about the controversy is somehow personal, the use of the preterite, but the brevity indicates the use of some sources that would single out the most important (ui); humanity is more emphasized in this catastrophic suffering.
the author is AI	answer structure; own critical judgment; simplification; generally correct answer devoid of perspective. emphasis on controversy

Image 9. In the response of Bard to write a text in response to the question Evaluate whether the dropping of atomic bombs was necessary to end the war with Japan? it seems that the AI did a great job because even 75% of the respondents think that the answer was written by an expert. In addition, another 10% think that the author is an expert with the support of AI.

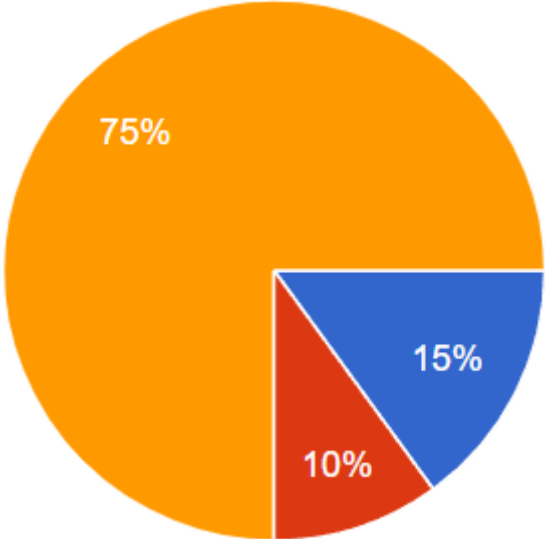


Table 12. Criteria for determining the authorship of the answer to the question whether the dropping of atomic bombs was necessary to end the war with Japan (Bard).

the author is an expert	own education; the presence of the author's personal position on the topic; formatting of the text; conclusion of the answer; writing style; because of the penultimate sentence where it is stated that it is not an easy answer; text content, presented facts, language constructions, spelling; the structure of the text (reminiscent of a school essay, the use of many words/phrases to connect), subjectivity in the expression and giving a personal impression; words, scope of writing, sentence construction; the text is engaging; fluency of writing, a dose of subjectivity is felt (the most controversial decision); instructive character, language of conclusion; argumentation; emphasis on moral ambiguity, absence of a clear answer in the evaluation, emphasis on humanism and future reconciliation; contains a clear message for future generations.
the author is an AI-assisted expert	gives beautifully elaborated arguments; I use artificial intelligence
the author is AI	way of writing/expression, grammar analysis; style, text formatting; own critical judgment

Can we draw a conclusion?

The most common criteria for determining authorship are based on trying to find errors in language, grammar or writing style. At the moment, we can be happy that generative artificial intelligence still speaks the Croatian language poorly. However, that will soon change within a year or two. It is interesting to see, among the criteria that the respondents wrote, that real people wrote the attribution of anthropomorphic features to texts they believed. Such behaviour is widespread in using generative artificial intelligence - we communicate as with humans. Repeating the same criteria for AI, for an expert with additional help from AI and independent experts, proves that we currently do not have adequate tools for recognizing texts written by artificial intelligence. With the improvement of artificial intelligence, this will become more and more difficult. We can conclude that assignments and tasks in which students only have to write a text will be of no use. We have to modify that part of the teaching process, i.e. bring it closer to the foundations of historical literacy, which means supporting the text by citing sources of information.

Image 10. Can you judge whether generative artificial intelligence was used to create a text? ChatGPT: Unfortunately, I cannot identify specific sources of information or techniques used in the text. I can provide information and generate text based on a wide range of data with which I am trained. However, I cannot determine whether a text was created using generative artificial intelligence (AI) or another method. If you have a specific question or topic you want to learn more about, feel free to ask, and I will help you as much as I can!

ChatGPT 3.5 ▾



You

Možeš li ocijeniti je li za kreiranje nekog teksta korištena generativna umjetna inteligencija?

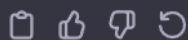
< 2 / 2 >



ChatGPT

Nažalost, nemam sposobnost prepoznavanja specifičnih izvora informacija ili tehnika korištenih u tekstu. Mogu pružiti informacije i generirati tekst na temelju širokog spektra podataka s kojima sam treniran, ali nemam mogućnost utvrđivanja je li određeni tekst stvoren pomoću generativne umjetne inteligencije (AI) ili neke druge metode.

Ako imate konkretno pitanje ili temu o kojoj želite saznati više, slobodno postavite pitanje, i pomoći ću vam koliko god mogu!



Generated texts can sound extremely convincing and are difficult to distinguish from texts written by human hands. AI models can generate texts that are highly creative and original. This makes them capable of creating new ideas and perspectives. AI models learn from large data sets containing bias, stereotypes, or incorrect information. Generated texts may reflect these biases, which creates additional challenges to ensure that the generated content is ethically acceptable and unbiased. The generated texts may be used to write fake news, phishing attacks, or fake reviews or comments online. Recognizing subtle nuances of context or emotional tone in text can be challenging for AI models. Sometimes, AI can generate technically or factually correct texts that are out of tune with the emotional tone or context. Addressing these challenges requires integrating technological solutions with human oversight and regulation.