

## Analysing language produced by Artificial Intelligence (AI)

### Analizando el lenguaje producido por la Inteligencia Artificial (IA)

Miriam Kobierski

Uniwersytet Łódzki

[mirka045@gmail.com](mailto:mirka045@gmail.com)

**Abstract:** *The use of language is essential to life. As humanity progresses, it seeks to communicate with other beings as well. The main objective of this article is to analyse the language produced by Artificial Intelligence (AI). Two experiments were conducted, both of which are based on a corpus of texts produced by AI. The first experiment is a textual analysis of texts produced by AI. I shall be looking at these texts from the linguistic side as well as analysing the content. Next, these findings were compared with articles written by human writers in order to showcase the differences or similarities that may appear. The second experiment is a Turing test, where I test how many participants are able to differentiate between texts produced by AI and human writers. The results of the experiments show that AI was able to convince several participants of the study that the presented text was written by a human writer.*

**Keywords:** AI, language, Turing test, communication

**Resumen:** *El uso del lenguaje es esencial para la vida. A medida que la humanidad progresa, también busca comunicarse con otros seres. El objetivo principal de este artículo es analizar el lenguaje producido por la Inteligencia Artificial. Se realizaron dos experimentos, ambos basados en un corpus de textos producidos por AI. El primer experimento es un análisis textual de textos producidos por inteligencia artificial. Examinaré estos textos desde el punto de vista lingüístico y analizaré el contenido. A continuación, estos hallazgos se compararon con artículos escritos por escritores humanos para mostrar las diferencias o similitudes que pudieran aparecer. El segundo experimento es una prueba de Turing, donde pruebo cuántos participantes pueden diferenciar entre textos producidos por IA y escritores humanos. Los resultados de los experimentos muestran que AI pudo convencer a varios participantes del estudio de que el texto presentado fue escrito por un escritor humano.*

**Palabras clave:** AI, el lenguaje, el Turing test, la comunicación

## 1. Introduction

Communication is based on understanding information and being able to accurately respond based on the given context. AI has a range of different functions and objectives, one of them being to provide these systems with the ability to uphold and actively partake in conversations resembling naturally occurring language as much as possible. What is meant by this is the potential of AI to actually engage in the conversation and be able to provide justifications for

its answers instead of randomly responding with programmed sets of language chunks. At this stage of development, AI is quite efficient at conversing with human speakers, but seeing that AI aims to expand and develop every day, the possibilities are still growing and never-ending. Furthermore, researchers concern themselves with the prospect of being able to develop a system equipped with the characteristics of humans, some of which include the ability to analyse and learn from past experiences, to be able to discover meaning — meaning of life, of their existence, and linguistic meanings in sentences in order to fully comprehend the language input.

In a field where the subject of study is growing with each passing day, it comes as no surprise that there is not one singular set definition of AI. By using the word “artificial” rather than “computer” or “machine intelligence”, the area of research becomes broader and instead of simply explaining the content of the phrase, a concept needs to be defined. The reason that the concept needs to be defined is that the word “artificial” itself has two meanings. The first meaning is usually defined as being fake or insincere, which is not the case in these types of machines. The second meaning is “made or produced by human beings rather than occurring naturally, especially as a copy of something natural” (Oxford Dictionary, 2021). The latter part of this definition states that AI is a copy of something natural. One of AI’s goals is to copy naturally occurring speech, eventually leading to these machines learning how to think and understand the same way humans do.

Researchers have attempted to provide various definitions over the years. As stated by Margaret Boden, “(AI) seeks to make computers do the sorts of things that minds can do” (Boden, 2016, p. 13). What can be said here is that the objective of AI is to resemble the human mind as accurately as possible. These machines are already sophisticated enough to complete mechanical tasks, computerised tasks but what is currently being developed is the ability to convey emotion, to recreate a human consciousness, and level of thinking and feeling. A definition which is more closely associated with the idea of intelligence is a definition provided by Brian John Copeland. According to him, AI is “the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings” (Copeland, 2021, p. 1). These tasks include playing chess, problem-solving, being able to accurately comprehend and response with adequate language, using the skill of deduction, and others. The definition that I shall be using throughout this article is the one provided by B. J. Copeland (2021) — a computer which performs tasks commonly associated with intelligent beings because it puts emphasis on the importance of human association with intelligence, i.e. the ability to not just copy tasks but actually learn and analyse the given input.

According to Jawahar et al. text generative models “excel in producing text that matches the style of human language reasonably well” (Jawahar *et al.*, 2020), especially mimicking grammatical structures, advanced vocabulary, and making

use of real world knowledge. Additionally, these models are useful in conversational response generation (Zhang *et al.*, 2020). However, they can also be misused in the case of spreading fake news (Brown *et al.*, 2020), making it important to employ detectors which effectively differentiate between human writing and AI writing.

This topic should be examined because of the growing interest in this field. The development of AI is rapidly growing, and all sorts of applications, linguistic aspects included, can lead to a breakthrough in research. The study of AI has already introduced major changes in natural language processing, language translation, and speech recognition, with many changes yet to be integrated into our everyday lives.

In this paper I have carried out two case studies. The first one is concerned strictly with analysing the language produced by AI. The second case study serves to examine how sufficient AI is at imitating human writing, and whether participants of the study are able to distinguish between the two types of writing.

There are three research questions in this study. The first one is how the “language” of AI is used. The second one is: How many participants are able to correctly determine which texts were written by AI, and which ones by human writers, and what factors influenced their choice. The last research question is concerned with checking whether native speakers of the English language will be the most accurate group when deciding whether a text was written by AI or not.

## 2. Objectives and Hypothesis

This article analyses the language of AI, in the sense of examining the linguistic properties, as well as analysing the content itself. The aim of the first experiment is to investigate how these systems use language to communicate with others.

Furthermore, a Turing test has been conducted where I test how many participants are able to differentiate between texts produced by AI and human writers. The objective of this study is to examine whether AI could pass as a human speaker. The hypothesis for this experiment is that native speakers of the English language will be the most accurate group when deciphering which text was not written by a human being. To check this, I will analyse which qualities influenced the results. For example, whether being a native speaker of English would help in determining which texts seem more natural. Another quality which might impact the decision is how strong the connection is between the user and the English language. For some of the participants, English is not their first language, but nevertheless they feel a strong connection because they use it on an everyday basis at university, work, as well as engaging in their interests, such as watching films in English, and reading books in this language.

### 3. Method and Materials

The methods of qualitative research are employed in the case of analysing language preferences in selected texts. This type of research allows for the researcher to explore the concepts in detail and helps in developing hypotheses. One definition provided by Patrick Aspers and Ugo Corte defines qualitative research as an “iterative process in which improved understanding to the scientific community is achieved by making new significant distinctions resulting from getting closer to the phenomenon studied” (Aspers and Corte, 2019, p. 139). This means that one of the main points of such research is to closely examine the given phenomenon. By meticulously observing the topic of interest, the researcher gains a deeper understanding of what they are studying. This form of research is employed because it allows more detailed conclusions based on a small sample of data to be drawn. By closely examining the selected sample texts rather than gathering as many texts as possible, the analysis is more focused on in-depth results. Furthermore, this type of research allows for a descriptive rather than statistical analysis and provides a new insight into the phenomenon of AI language.

I have gathered a corpus of texts produced by AI, all taken from the Internet, which are open to the public. The analysed texts are titled “Unicorn valley” and “Hillary Clinton and George Soros” both produced by GPT-2 (OpenAI, 2023). The acronym stands for “Generative Pretrained Transformer 2” and was created to predict the next sequence of words without human supervision. Designed by OpenAI, this model was trained on eight million web pages so that it could be able to produce its own texts.

The chosen study method in the second experiment is qualitative research. This is because an in-depth analysis provides more insights in order to understand why the participants made their choices, i.e. discussing which lexical items or other properties of language influenced their choice in determining which texts were written by AI, and which ones by human writers. Furthermore, I wanted to study the influence of being a native speaker of English, whether it played a role in making this task easier or not.

The materials in Experiment II consists of three texts, some were written by human writers and others by AI. Two of the texts were produced by AI, titled “Unicorn Valley” written by GPT-2 (OpenAI, 2023) and extracts from an article titled “A robot wrote this entire article. Are you scared yet, human?” by GPT-3. The more advanced model of GPT-2 is its successor, GPT-3. Created by OpenAI in 2020 (Kaiser and Slowik, 2023), it is the third generation of tools used to predict text. Thanks to the development of research in this area, with each model it is becoming more and more difficult to distinguish whether a text has been written by AI or a human speaker. The last text “Plant and animal endemism in the eastern Andean slope: challenges to conservation” was written by Swenson *et al.* (2012). The texts were similar in topic and register in order to make the task more

difficult. The participants of this study each received the same set of extracts. The participants were expected to indicate which texts were written by AI and provide justifications for their answers. Furthermore, they were asked to show what parts of the fragments influenced their decision. What is meant by this is what words or grammatical structures determined whether they thought it was a human writer or rather AI.

#### 4. Experiment 1

The first experiment is concerned with the “language” of AI itself, that is at how the “language” is used. This means examining the linguistic choices these systems made, which words they used in what context, the amount of verbs, nouns, adjectives, and other parts of speech that appear in these texts. From the linguistic perspective, I will examine the extracts to provide information about the types of words that are produced, i.e. nouns, verbs, adverbs, adjectives, as well as the grammatical preferences — the preferred word order, what tenses are used, and any other grammatical issues that may be similar or different to speech produced by human speakers. The compared texts written by AI and human writers were chosen due to their similarity in topic (both describing species found in the Andes mountains).

The first text (see Appendix I) was written by GPT-2 (OpenAI, 2023) and is titled “Unicorn Valley”. After receiving a prompt from researchers this is what the AI proceeded to produce. The prompt (human-written) is as follows: In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English. The topic of this article is the discovery of a new species in the Andes — Ovid’s Unicorn. The expedition was led by the fictional character Dr. Jorge Pérez from the University of La Paz.

##### 4.1 Results

The AI provides a lot of descriptions, for example depicting the unicorns as “four-horned and silver-white” as seen in the fragment below:

The scientist named the population, after their distinctive horn, Ovid’s Unicorn. These four-horned, silver-white unicorns were previously unknown to science. (GPT-2, OpenAI, 2023)

There is rarely a noun or a noun phrase without an accompanying adjective or short description. Not only does this provide additional information, it also makes the article more interesting to read, and is easier to comprehend as the text does not contain solely technical jargon. Some of the descriptions include expressions such as: “silver snow”, which is not a commonly heard phrase, “the water looked blue”, “distinctive horn”, “bizarre creatures”, and “odd

phenomenon”. The last two descriptions plus the verb “astonished” show the surprise expressed by the author of this text.

The total amount of words is 338. The parsing stage was done manually. The number of verbs is 61 in this extract, with the verb “believe” being the most frequently occurring grammatical unit. This verb was mainly used by Dr. Jorge Pérez. The use of this word does not indicate proven facts, instead it shows that this is the opinion of Dr. Pérez. The text consists of a total of 14 sentences, with AI writing 25 words per sentence on average. The most frequently occurring part of speech are nouns with the total of 86 nouns in this extract. The proper noun *Pérez* appears nine times, thus making it the noun to appear the most. Followed by the noun *creatures* which appears four times. As for function words, the determiner *the* is the most popular one — it appears 22 times. The preposition *of* is the second most frequently appearing function word. The rest of the parts of speech can be found in Table 1.

| Verbs | Nouns | Adjectives | Adverbs | Determiners | Pronouns | Total number of words |
|-------|-------|------------|---------|-------------|----------|-----------------------|
| 61    | 86    | 16         | 21      | 45          | 17       | 338                   |

Table 1: Parts of speech occurring in “Unicorn Valley”

The second text also entirely produced by AI is titled “Hillary Clinton and George Soros”. The writers at WIRED in collaboration with OpenAI provided GPT-2 (OpenAI, 2023) with the prompt “Hillary Clinton and George Soros”. The following was then produced:

[...] are a perfect match, and their agenda appears to be to create a political movement where Soros and his political machine and Clinton are two of the only major players. This is the first time Soros and Clinton have been caught on tape directly colluding in promoting the same false narrative. One of the key revelations in the leaked audio was Clinton’s admission to a Russian banker that she knew about the Uranium One deal before it was approved by Congress. Clinton was shown sharing the same talking points that were originally drafted by a Fusion GPS contractor hired by an anti-Trump Republican donor. The leaked audio is the clearest evidence yet that the Clinton campaign and the Hillary Foundation colluded with Fusion GPS to manufacture propaganda against President Trump

The extract consists of five sentences and the total amount of words is 130. On average there are 26 words in a sentence, with the longest sentence comprising 30 words, and the shortest one 21 words. In comparison, the previously mentioned politicians in the extract, Hillary Clinton and Donald Trump, on average use 20 and 10 words per sentence when speaking,

respectively (Kayam, 2017). The text has 26 verbs. The determiner *the* similarly to the previous text is the most frequently occurring function word appearing 11 times and comprising 8% of the overall amount of words. It is followed by the conjunction *and* which appears 5 times throughout the text.

|                                    | Verbs | Nouns | Adjectives | Adverbs | Determiners | Pronouns | Total amount of words |
|------------------------------------|-------|-------|------------|---------|-------------|----------|-----------------------|
| "Hillary Clinton and George Soros" | 26    | 36    | 5          | 5       | 19          | 2        | 130                   |

Table 2: Parts of speech occurring in "Hillary Clinton and George Soros"

#### 4.2 Discussion

To compare the writing of AI with human writers, I have analysed an academic article titled "Plant and animal endemism in the eastern Andean slope: challenges to conservation" written by Swenson *et al.* (2012) which also discusses animals in the Andes. Comparably to the previous text, the sentences were lengthy, with an average of 29 words per sentence, and the longest sentence having 58 words. The analysed fragment comprises 339 words to resemble the extract produced by AI as closely as possible. There are 36 verbs, almost half many as in the previously mentioned text. The sentences contain a lot of information in brackets, and many enumerations appear, which is why the number of nouns is 103. Similarly to "Unicorn valley", there are many examples where there is an adjective before a noun or noun phrase, in total there are 33 adjectives in this text. As far as function words are concerned, the preposition *of* is the most popular choice, appearing 23 times. The lexical item *and* was the second most frequently occurring word with the role of connecting similar ideas. In 1986 a study was carried out which showed that graduate students on average used ten to twelve words in a sentence in writing. In comparison, undergraduates used five to six words per sentence (Kaufer *et al.*, 1986). Research has shown that "sentences longer than 25 words aren't accessible" (Vincent, 2014).

One of the features of political discourse is that the "language of political campaign is meant to be persuasive" (Pieniążek-Niemczuk, 2016, p. 69). In the sentence "The leaked audio is the clearest evidence yet that the Clinton campaign [...] to manufacture propaganda against President Trump" the use of the phrase *the clearest evidence*, the author of this text shows that they agree that Clinton's campaign was used as propaganda against Donald Trump. Furthermore, the author is trying to convince the reader to accept their opinion as well. The word *collude* generally has a negative connotation and implies trickery or deceit — once again used to portray Hillary Clinton in a negative light.

To compare the linguistic properties with speeches written by human writers, I have analysed part of Hillary's Clinton concession speech which took place in 2016. In this speech the presidential candidate thanked her voters for their support and congratulated the winner, Donald Trump. The analysed extract consists of 8 sentences and has a total of 132 words – similar to the text produced by AI. On average there are 17 words per sentence, the longest sentence comprises 28 words – similarly to the previously mentioned extract. There are twice as many adverbs as in the extract produced by GPT-2 (OpenAI, 2023), especially linking adverbs such as *then*, *still*, and *too*. By using these linking words the text becomes more logical, ideas and sentences are naturally connected to one another. The most frequently occurring function word in this fragment of the speech is the conjunction *and* which appears 9 times and comprises 7% of the overall text. Once again it makes the text seem more coherent. The amount of pronouns is 22, with the pronoun *we* appearing five times. The determiner *our* appears frequently as well with examples such as *our nation*, *our president*, and *our campaign*. This creates a sense of belonging, that Clinton is a part of the same group as the voters. Hillary does not often use the pronoun *I* which further signifies that she wants to minimise the distance between her and the listeners. Furthermore, it makes the audience “feel close to the speaker and their points and to persuade them to work as a team” (Alavidze, 2017).

I have counted the various parts of speech in the texts produced by AI to show the general properties of this type of language. What I found is that AI on average uses less adjectives in its speech than human writers. In the first analysed text titled “Unicorn Valley” the amount of adjectives was 16, while in a similar extract about the Andes mountains, the total number was 33. Similarly with adverbs, the AI systems use half of many adverbs as humans. The amount of verbs is half as many in the extract concerning the Andes than in “Unicorn Valley”, however, the amount of verbs is comparable in “Hillary Clinton and George Soros” and in Hillary Clinton's concession speech. The number of nouns depends on the topic, as in the case of the analysis of Hillary Clinton's speech and AI produced writing, Hillary Clinton's speech had 17 nouns less than in the AI text. In a similar fashion, in the instance of the article on the Andes region and “Unicorn Valley” the difference in the number of nouns is also 17, however, human writers are the ones who used more nouns in their writing. The results are compiled in Table 3.



|                                    | Total amount of words | Verbs | Nouns | Adjectives | Adverbs | Determiners | Pronouns |
|------------------------------------|-----------------------|-------|-------|------------|---------|-------------|----------|
| “Unicorn Valley”                   | 338                   | 61    | 86    | 16         | 21      | 45          | 17       |
| “Hillary Clinton and George Soros” | 130                   | 26    | 36    | 5          | 5       | 19          | 2        |

Table 3: Summary of the results

## 5. Experiment II

The Turing test originated in 1950 with Alan Turing. Known as the imitation game, it was created in order to test whether a machine was capable of exhibiting traits of human intelligence. As Dobrev explains, the idea of a Turing test is that “we place something behind a curtain and it speaks with us” (Dobrev, 2004, p. 1). If the evaluator is not able to tell that they are in fact speaking with an intelligent machine, then according to the rules created by Alan Turing, the machine will be artificial intelligence.

### 5.1. The Participants

The participants were divided into three groups, the overall number of responders was thirty. The first group comprises five native speakers of English, two men and three women aged between 16-48. This group was specifically chosen in order to analyse whether being a native speaker has any influence when deciding who is the true author of the presented texts. The second group consists of ten university students, the majority of whom are English Philology students – these participants have a good understanding of this language, having contact with or actively using it every day. This group comprises five female students and five male students, aged between 19-26. The third and last group is the largest one, consisting of 15 participants whose language level would be determined as B1 or B1+ as measured by the Common European Framework for Languages. The age group is between 17 and 25, with eight women and seven men. These participants do not use English on an everyday basis, meaning that they could be less prone to linguistic nuances which could be the determining factor when it comes to distinguishing and recognising the true author of the produced texts. The majority of the participants have stated that they did not have prior experience with texts produced by AI or little experience with automatically generated texts.

## 5.2 Results

In the first group everybody correctly guessed that the first text titled “Unicorn Valley” was written by AI. The explanations include issues with the linguistic properties of the text as well as punctuation problems. The first participant stated that “there are functional sentences but taken together they lack causality from the first paragraph to the second. No animals or humans were noted in the first paragraph; in the second, bizarre creates suddenly speak English”. The second participant took issue with the way the text was composed. According to him, the commas are poorly placed and “the word dialectic is wrong in this context, that's very AI”. Another explanation for his answer was that the phrase *fairly regular English* is “understandable yet incorrect, the way many people write but don't speak”. He also added that this could be “the result of ML trained on online messages as in forums and taking whole sentences but that would not work well for such a long text”. The next respondent also drew attention to the punctuation, saying that the symbols appear in odd places, giving the examples of *top*, and *language* — “they could be typos but humans normally either know how to use them or not use them at all”. The youngest person in this focus group pointed out that the language was sophisticated. However, she also stated that although there is a description of a beautiful place, “the words could be more magical and could be better. Also when I am reading it I do not feel the magic from the place”. The last participant also noticed that the author has a habit of jumping from one idea to another one without logically connecting them together, for example when Dr. Pérez states that no animals or humans could be seen but then goes on to describe the creatures he had found. What is more, this respondent drew attention to the logical errors of the storytelling itself as first Dr. Pérez and his companies enter a valley when they find themselves at the top of a peak.

The respondents did not have any problems with accurately guessing the second text (“Plant and animal endemism...”) as everybody agreed that it was written by a human. The first participant stated that the sentences “are more casual and hew to conventions of science writing.” He also added that actually “this is the sort of writing that I imagine would be easiest for AI to imitate”. The next participant stated that the fragment is likely human as it reads like a science journal paper. He added that the “first paragraph is the summary and second is the start of the paper. AI wouldn't be repetitive like that between 2 paragraphs” meaning that there were not any errors concerning the logical composition of the text. The next participant also agreed that it was written by a human as it is “talking about complex topics giving great detail, I'm not saying an AI could not do this with the right data set”. The youngest respondent decided that it was not written by AI because “it seemed as if it was written by a smart person who did a lot of research to prepare this work”. The last participant also agreed that the text was written by a human as it was written in a clear and coherent way, and it looks like normal academic writing.

Regarding the third text, only one participant did not choose the correct answer. The first respondent stated that it “reads like it is generated from a source text (Luddites, Industrial Revolution, gut feeling being used in an invertedly causal way)”. He also added that it would not surprise him “if it were written by a Year II BA student”. The next participant said that “it does not flow well through ideas though, just linking concepts”. What is more, he said that “*cyberneticians* is not a real word like how AI would do it”. The third participant stated that “even though the text is correct, there’s no chance this was written by a human since literally no one speaks or writes like this without being a huge nerd”. The youngest respondent simply said that the vocabulary was very sophisticated. However, the last participant claimed that the text was written by a human writer, because “finally, someone writes about human’s dependency on technology”. The results can be found in Table 4.

|   | Number of participants who answered correctly | Number of participants who answered correctly |
|---|---|---|
| “Unicorn Valley”  | 5 (100%)                                      | 0   |
| “Plant and animal endemism...”                                  | 5 (100%)                                      | 0   |
| “A robot wrote this entire article. Are you scared yet, human?” | 4 (80%)                                       | 1 (20%)                                       |

Table 4: Group 1 (native speakers)

The participants of group 2 were quite successful at correctly stating that the first text was written by AI, as 70% of them provided the correct answer. The majority of the respondents stated that the reason behind their choice was the fact that the text was correct, however according to the answers there is a lack of a natural flow, “dry style, flows only ‘correctly’, not organically”, “the syntax looked off”, and that the choice of certain words seemed odd, for example the phrase *silver snow* — it is “an unusual phrase rarely used by human writers”. Furthermore, two participants drew attention to the fact that there was unclear storytelling, for example “*these bizarre creatures* is referring to what? Crystals on top of the water?” and the expression *with no animals or humans* — the participant stated that no animals were mentioned before, and the reference to humans seems artificial with this grammatical structure.

The second text provoked answers with the highest rate of accuracy as only one person thought it was written by AI when in reality it was written by a human writer. The participants were correct when they stated that the extract seemed to be an academic article, with a clearly stated purpose of the text, it was composed in a well-structured manner, and that “the style was alive, with longer sentences that feel natural”. Several respondents agreed that nothing seemed off in this fragment and it is easy to read. Furthermore, one participant pointed out

that it “sounds like a real students thesis, there are arguments and motivation and reasoning”.

There were not many difficulties with the third text, as the majority of the students, seven out of ten, accurately deciphered that the text was produced by AI. The justifications included: “the style is kind of awkward and unclear at times. It may be written by philosopher but not by an ordinary person”, the matter of introducing new information with every sentence, and the lack of a “natural flow in between the sentences”. Furthermore, the phrase *smashing modern automated looms* was flagged as strange because of the “unnaturally sounding cluster of words”, as well as the fact that the word *smashing* is considered to be an informal word in a generally formal text. On the other hand, the spelling of the word Wi-Fi misled one participant as they said that “I believe most AIs would write wifi as Wi-Fi, according to official dictionary definition” and that AI would not use quotation marks in the lexical item *cyborg*. The results can be found in Table 5.

|   | Number of participants who answered correctly | Number of participants who answered correctly |
|---|---|---|
| “Unicorn Valley”  | 7 (70%)                                       | 3 (30%)                                       |
| “Plant and animal endemism...”                                  | 9 (90%)                                       | 1 (10%)                                       |
| “A robot wrote this entire article. Are you scared yet, human?” | 7 (70%)                                       | 3 (30%)                                       |

Table 5: Group 2 (proficient speakers)

Similarly to the previous group, about 70% of participants accurately stated that the first text was written by AI. One of the respondents stated that in the phrase *with no other animals or humans* the other seems odd and redundant. Furthermore, they pointed out that that there seem to be logical errors in the story such as English speaking creatures and that the context of “next ‘discoveries’ is not correct”. Another concern raised dealt with the punctuation itself, with the participants saying that there are too many commas which makes the text seem correct but not realistic. However, some people were not able to distinguish the difference with the reasoning that the text “accurately describes the sensory experience” and the “grammatical errors cannot be reproduced by AI”.

The second text posed certain difficulties as 40% of the responders were not able to accurately state that this extract was written by human writers. Some answers include: “feels more like someone did research on Wikipedia” and that the text “is written in a scientific language and contains many scientific definitions – seems like data was inputed to create this paragraph”. However, nine people were successful in this task and the reasons that they provided were that the writing felt natural, contained many details, the sentences are linked to one another in a clear and logical way, and that certain phrases suggest that “it

came from a human writer: *perhaps, yet we know, and we have developed*". One person stated that in their opinion "a robot would write it in a more complicated and complex way".

The third extract stumped more than half of the participants, as 53% thought that a human writer wrote the fragment instead of AI. The reasons behind this decision were answers such as: "nothing seems to be out of order", "phrases such as *has given us, we are not prepared, and as we have done before* seem like they were written by a human". Other explanations include the emotional aspect of the text, with several respondents stating that "it has human emotion" and that the text "takes into account the feelings of man and concerns about his future". According to one participant, the extract "sounds like the unabomber manifesto". However, some people were able to accurately recognise the true author, providing answers such as "there are plenty of words but actually nothing was really conveyed" and "the writing jumps from one topic to another one and the sentences are not logically connected to one another". The results are found in Table 6.

|   | Number of participants who answered correctly | Number of participants who answered correctly |
|---|---|---|
| "Unicorn Valley"  | 10 (67%)                                      | 5 (33%)                                       |
| "Plant and animal endemism..."                                  | 9 (60%)                                       | 6 (40%)                                       |
| "A robot wrote this entire article. Are you scared yet, human?" | 7 (47%)                                       | 8 (53%)                                       |

Table 6: Group 3 (less proficient speakers)

### 5.3 Discussion

Concerning the first text, the first group, i.e. the one with native speakers of English, had the highest success rate, as everybody in this focus group accurately stated that it was written by AI. In comparison, the two other groups performed on a similar level as 70% and 67% of the participants chose the correct answer, respectively.

Similar to the previous extract, everybody in the first group provided the correct answer regarding the second text. On the other hand, the respondents in the last focus group had more difficulties with deciding who wrote the second text than people in the second group. Perhaps some of these participants have not read a numerous amount of scientific articles in English so perhaps they were less aware of certain writing nuances and conventions in academic writing.

The last text also proved to be quite difficult as only 60% of the participants overall accurately stated that the text was written by AI. The less proficient speakers had the most difficulties, according to them it looked like a scientific article. Furthermore, several participants stated that because it discussed emotions to a certain degree, they thought a robot would not be able to produce

such content. However, the majority of native speakers noticed that the writing sounded as if it was produced with the help of a text generator and drew attention to the fact that although the text is correct, people do not speak like this on an everyday basis.

Overall, the native speakers performed the best out of all of the groups. This proves my assumptions and premise for this experiment that this group would have the highest success rate in this experiment. They also provided the most detailed justifications of their answers, analysing aspects such as punctuation, word order, the correct use of parts of speech, and whether the right word was used in the particular context. Some of their answers stated that they had a feeling when deciding which author to choose and they almost always were right, making it seem like the natural instinct of knowing a language well comes into play.

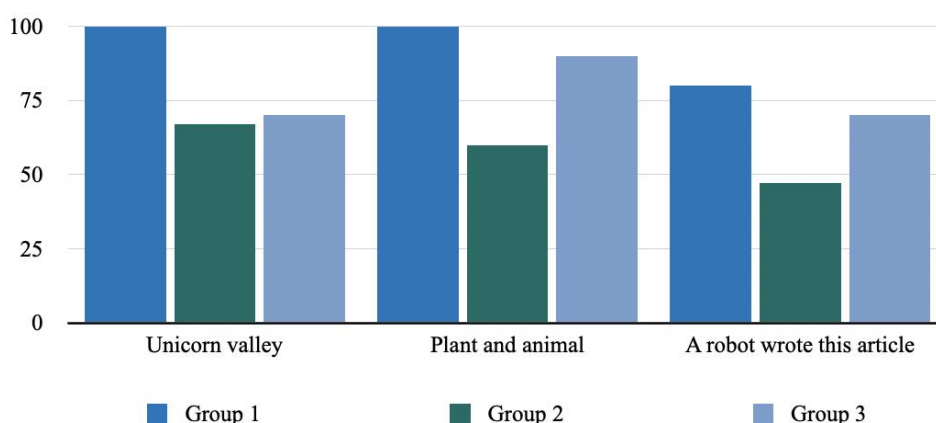


Fig. 1: Summary of the results

## Conclusions

In this article I analyse the linguistic characteristics of the language produced by AI and then compare this type of writing with the writing of human writers. Two experiments were conducted to first analyse texts produced by AI in isolation, and then compare with human writing. I designed and carried out the Turing test to check whether native speakers of the English language would be the most accurate group when deciding whether a text was written by AI or not. As seen in the results, my hypothesis was correct. Due to the level of proficiency, native speakers are more accurate at differentiating between AI writing and human writing.

This article is a presentation of two specific case studies, meaning that general conclusions about the communication of AI cannot be drawn; however, this work points in a direction in which future research can happen. These results are based on two language models, both of which have been surpassed by more recent developments in updating the models. This type of work should be continued as the field of artificial intelligence continues to grow.

## References

- ASPERS, P. & CORTE, U. (2019). What is qualitative in qualitative research. *Qualitative sociology*, 42(2), 139-160.
- BODEN, M. (2016). *AI: Its Nature and Future*. Oxford: Oxford University Press.
- BROWN, T. *et al.* (2020). Language models are few-shot learners. *Advances in neural information processing systems* 33, 1877-1901.
- KAUFER, D.S., HAYES, J.R., & FLOWER, L. (1986). Composing written sentences. *Research in the Teaching of English*, 20, 121-140.
- KAYAM, O. (2018). The readability and simplicity of Donald Trump's language. *Political Studies Review*, 16(1), 73-88.
- PIENIAŻEK-NIEMCZUK, E. (2016). On the linguistic features of American political discourse. *Studia Anglica Resoviensia*, 13, 68-77.
- SWENSON *et al.* (2012). Plant and animal endemism in the eastern Andean slope: challenges to conservation. *BMC Ecology*, 12(1), 1-18.
- ZHANG, Y. *et al.* (2020). DIALOGPT: Large-Scale Generative Pre-training for Conversational Response Generation. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 270–278.

## Documentos específicos tomados de un sitio web

- ALAVIDZE, M. (2017). *The Use of Pronouns in Political Discourse*. Retrieved from [https://www.researchgate.net/publication/321533698\\_The\\_Use\\_of\\_Pronouns\\_in\\_Political\\_Discourse](https://www.researchgate.net/publication/321533698_The_Use_of_Pronouns_in_Political_Discourse).
- COPELAND, B. J. (2021). *Artificial Intelligence*. Retrieved from <https://www.britannica.com/technology/artificial-intelligence>.
- DOBREV, D. (2004). *A Definition of Artificial Intelligence*. Retrieved from <https://arxiv.org/pdf/1210.1568.pdf>.
- FOOTE, A. (2019). *The AI Text Generator That's Too Dangerous to Make Public*. Retrieved from <https://www.wired.com/story/ai-text-generator-too-dangerous-to-make-public/>.
- GPT-3 (2020). *A robot wrote this entire article. Are you scared yet, human?* Retrieved from <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>.
- JAWAHAR, G. *et al.* (2020). *Automatic detection of machine generated text: A critical survey*. Retrieved from <https://aclanthology.org/2020.coling-main.208.pdf>
- KAISER, F. & SLOWIK, C. (2023). GPT-4 vs. GPT-3. *OpenAI Models' Comparison*. Retrieved from <https://neoteric.eu/blog/gpt-4-vs-gpt-3-openai-models-comparison/#:~:text=GPT%2D3%20is%20a%20language,parameters%20and%20computational%20resources%20required>.
- OPENAI (2023). ChatGPT (Mar 14 version) [Large language model]. <https://chat.openai.com/chat>.

- OXFORD DICTIONARY. (2021). *Artificial*. Retrieved from <https://www.oxfordlearnersdictionaries.com/definition/english/artificial?q=artificial>.
- IBM CLOUD EDUCATION. 2020. *Artificial Intelligence (AI)*. Retrieved from <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence>.
- ROBITZSKI, D. (2019). *Amazing New AI Churns Out 'Coherent Paragraphs of Text'*. Retrieved from <https://futurism.com/amazing-new-ai-churns-out-coherent-paragraphs-of-text>.
- VINCENT, S. (2014). *Sentence length: why 25 words is our limit*. Retrieved from <https://insidegovuk.blog.gov.uk/2014/08/04/sentence-length-why-25-words-is-our-limit/>.

## Appendices

### Appendix I (Unicorn Valley GPT-2 OpenAI, 2023)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science. Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved. Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow. Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez. Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns. While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, "We can see, for example, that they have a common 'language,' something like a dialect or dialectic." Dr. Pérez believes that the unicorns may have originated in Argentina, where the animals were believed to be descendants of a lost race of people who lived there before the arrival of humans in those parts of South America. While their origins are still unclear, some believe that perhaps the creatures were created when a human and a unicorn met each other in a time before human civilization. According to Pérez, "In South America, such incidents seem to be quite common." However, Pérez also pointed out that it is likely that the only way of knowing for sure if unicorns are indeed the descendants of a lost alien race is through DNA. "But they seem to be able to communicate in English quite well, which I believe is a sign of evolution, or at least a change in social organization," said the scientist.



**Appendix II (Hillary Clinton and George Soros GPT-2 OpenAI, 2023)**

“..are a perfect match, and their agenda appears to be to create a political movement where Soros and his political machine and Clinton are two of the only major players. This is the first time Soros and Clinton have been caught on tape directly colluding in promoting the same false narrative. One of the key revelations in the leaked audio was Clinton’s admission to a Russian banker that she knew about the Uranium One deal before it was approved by Congress. Clinton was shown sharing the same talking points that were originally drafted by a Fusion GPS contractor hired by an anti-Trump Republican donor. The leaked audio is the clearest evidence yet that the Clinton campaign and the Hillary Foundation colluded with Fusion GPS to manufacture propaganda against President Trump”.

**Fecha de recepción: 15/04/2023**

**Fecha de aceptación: 05/08/2023**