

**Introducing sAIrcasm: A sample analysis of a custom Artificial Intelligence for linguistic and discursive sarcasm recognition**

Alessia BATTISTA, University of Salerno, Italy

This contribution presents an innovative tool called ‘sAIrcasm’ powered by Artificial Intelligence (AI). Challenges concerning the use of Artificial Intelligence and Natural Language Processing (NLP) in linguistics have been the subject of recent debate; one of the major challenges concerns the automatic detection of complex communicative phenomena, such as sarcasm. Several attempts have been made, but the complexity of the task has not yet led to the creation of an effective tool providing human-like results. This article presents a custom GPT model capable of identifying sarcasm and making pragmatic inferences based on context-dependent verbal and paraverbal cues. This study considers a Facebook video analyzed by both a custom artificial intelligence model and the free version of ChatGPT, providing insights into the broader implications of AI-based models in linguistic studies, with a specific focus on whether they can perform human-like cognitive processes.

**Keywords:** Applied Linguistics; AI tools; Artificial Intelligence; Natural Language Processing; NLP; Sarcasm; Social Media

## **1. Introduction**

The use of artificial intelligence (henceforth AI) is spreading in various domains of daily life, including linguistic education and research, and has led to a growing interest in how AI technologies can assist human researchers in carrying out complex linguistic analyses. Although the idea that machines could perform human-like tasks dates back to Alan Turing’s (1950) ‘Imitation Game’, it was not until 2020 that AI-powered tools became accessible to the general public and have set a new benchmark for generative AI thanks to OpenAI’s *Chat GPT* (2023).

Starting from the consideration that even the most recent architectures, like GPT and BERT, fail to fully grasp the nuances of human communication (Bender & Koller, 2020) and that modern technologies might indeed help human researchers in the analysis of large corpora (Zappavigna, 2023), the present study presents an innovative custom model based on OpenAI’s GPT: sAIrcasm (Battista & OpenAI, 2024). The model was created with the aim of utilizing one of the most advanced generative AI architectures, namely OpenAI’s GPT, to identify, annotate, and analyze sarcastic texts, thereby challenging modern

technologies to address one of the most cognitively and pragmatically complex phenomena of human communication: humor and sarcasm. As the following section will explore in detail, multiple attempts have been made to develop tools for the automatic recognition of sarcasm; however, there are currently no established methods or software, which leaves room for new experimentation. This article aims to contribute to the advancement of research bridging AI and linguistics by providing an innovative model.

The remainder of this paper presents an essential literature review in Section 2 before moving on to the description of sAIrcasm and its development in Section 3. Sections 4 and 5 provide a sample analysis conducted using the model and discuss its potential improvements and future applications.

## 2. Background

The first theorization about machines, and how they can support humans, dates back as far as 1726, when Jonathan Swift described a device, ‘The Engine’, capable of assisting anyone and without too much effort in generating books in any discipline (Swift, 2001). A couple of centuries later, similar tools have been developed, thanks to the new possibilities offered by AI. Meta’s Meta AI, Google’s Gemini, and OpenAI’s ChatGPT are just a few of the larger systems among the various options available to the general public.

AI currently represents one of the most groundbreaking and complex technologies across multiple fields, as demonstrated by the lack of a single widely shared definition for it. For instance, the US-based tech leading company IBM defines AI as a “technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy” (Stryker & Kavlakoglu, 2024), thus stressing the human-like nature of its processes, whereas the Organisation for Economic Co-operation and Development explicitly states that the tasks that AI performs are based on “human-defined objectives” (Organisation for Economic Co-operation and Development, 2024, p. 4). Conversely, the European Union’s regulation of the topic, also known as the *AI Act*, defines an AI system as:

. . . a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments (European Union, 2024)

This highlights the role of humans in AI-powered processes. This brief exploration of AI clearly shows how complex the issue can be, which explains

why some scholars only see AI as an umbrella term (Proietti, 2024, p. 883). The complexity of AI is even more crucial when using it as a tool for linguistic research. Bender and Koller (2020) argue that, in spite of the recent advancements in text generation, AI models still display some limitations in the semantic analysis of messages, whereas other scholars suggest that AI-based tools can provide useful insights into sentiment analysis when applied to large datasets (Wankhade et al., 2022).

One of the most cognitively and pragmatically complex communicative phenomena is sarcasm. Falling under the umbrella term of 'humor', sarcasm refers to instances of communication in which there is an opposition between the stated and the intended message, meaning that one says the opposite of what they actually mean (Attardo & Giora, 2014, p. 398). From a linguistic perspective, sarcasm can be retrieved in texts whose 'perlocutionary goal' (cf. Salmani Nodoushan, 2019, 2021) is to be perceived as humorous, while also possibly implying a critical or even negative remark (although it should be specified that there are no clear-cut distinctions between humor techniques, such as irony and sarcasm; for more details on the issue, see Attardo, 2020; Attardo & Giora, 2014). This goal is achieved through a combination of strategies using language for non-linguistic goals—i.e., defunctionalization of the medium, breaking the rules of communication, relying on ambiguities and, most importantly, on contextual (not necessarily verbal) cues—and the presence of opposing or overlapping, yet resolved, scripts (Attardo, 2020). This is why sarcasm is such a complex communicative phenomenon.

Due to its challenging nature, scholars have been attempting to study sarcasm in online communication to understand the role it can play. For example, social media posts have been analyzed using sentiment analysis (Agbehadji & Ijabadeniyi, 2021) and as a site in which sarcasm is used to disclose controversial opinions on debated topics, such as climate change (Anderson & Becker, 2018; Anderson & Huntington, 2017; also see Archibald, 2023 on the use of metaphors in COVID discourse). The defunctionalized use of language and the use of non-verbal elements of the communication to be conveyed has also become a popular research trend, as scholars have been investigating the use of emojis to express affect (Elder, 2018; Riordan, 2017), to facilitate the interpretation of sarcasm in intergenerational online exchanges (Garcia et al., 2022) as well as emotional expression and connection, provided that they are correctly and appropriately used (Saraf, 2024), as is the case with other pragmatically complex phenomena, such as metaphors (Degano, 2023). Additionally, researchers have been trying to develop tools for the automated detection and annotation of pragmatically complex linguistic patterns using AI. Yu et al. (2024) have reviewed existing AI chatbots, demonstrating that GPT-4 has human-like accuracy in annotating apologies; furthermore, deep learning (Kumar & Garg, 2023; Rahaman Wahab Sait & Khairi Ishak, 2023) has been

demonstrated to be a valuable resource, and it has also been used to train AI-powered chatbots to incorporate and understand irony and sarcasm so as to make human-machine interactions more natural and empathetic (Balamurali et al., 2024; Rizwan, 2024).<sup>1</sup> However, AI-powered systems still face challenges when dealing with less structured, more context-dependent phenomena such as sarcasm and irony (Zappavigna, 2023), thus generating plausible but inaccurate responses due to their reliance on surface-level patterns rather than deeper semantic comprehension (Zhao et al., 2023), even when multimodal cues are provided (Castro et al., 2019).

### 3. Method

This study presents a custom AI-powered model, sAIrcasm, which relies on OpenAI's GPT architecture and has been specifically trained for the recognition and linguistic analysis of sarcasm. There are many reasons why creating a model based on GPT can be advantageous. For instance, because the technical architecture is already available for customization, it is not necessary to possess advanced programming skills to use it. Additionally, GPT-based models can be adapted to various needs, making them highly scalable. Lastly, GPT-based models can analyze contextual information and carry out internet searches. They are also programmed for continuous improvement over time.

Creating a custom GPT model involves three essential phases: (1) developing the base model, (2) training it with appropriate materials, and (3) fine-tuning it to ensure accurate task processing. As previously mentioned, since the technical architecture has already been developed and made available by OpenAI, the creation of a custom model relies on what is known as 'prompt design'. A prompt is an input written in a natural language—English, in the case of this study—and input to an AI system by a human (Wang & Jin, 2023); the art of providing an accurate prompt, which is crucial for accurate results, is prompt design or prompt engineering. Several strategies can be used in prompt design, such as task decomposing (Wu et al., 2022), which consists of the deconstruction of the main task into smaller sub-tasks. This increases the probability of obtaining accurate responses. Additionally, the prompts used in the creation of sAIrcasm merge declarative, imperative, and conversational structures (Tabatabaian, 2024) based on the specific goal of the sub-task: declarative prompts should be used to teach the model basic concepts and reasonings; imperative prompts are prioritized when assigning a task; conversational prompts make it possible to provide feedback for further refinement. Furthermore, it is essential that the prompt be direct and free of ambiguities while also making caveats and guidelines explicit (Tabatabaian, 2024).

Bearing this in mind, sAIrcasm was trained to communicate in an academic style and handle large textual corpora with the aim of detecting the presence of sarcasm. In order to do this, a number of discursive cues derived from relevant scientific literature on the topic were provided (cf. Attardo, 2014, 2020), namely (a) context (setting, interactants, tone, background/contextual information and/or information from the Internet), (b) the use of colloquialisms, idiomatic expressions, cultural references, (c) creative uses of language and punctuation, including unusual and non-standard spellings, use of capital letters, interjections, special characters, and 'swear words' (cf. Salmani Nodoushan, 2016), and (d) typical patterns of sarcasm: exaggerations, puns, contradictions, paradoxes, oppositions, wordplays. In addition, sAIrcasm was instructed to incorporate sentiment analysis, which is one of the core technologies used by GPT models to interpret texts, and it was encouraged to make 'pragmatic inferences' (cf. Salmani Nodoushan, 2022, 2025). To conclude, the model was cautioned against (over)generalizations, stereotypes, biases, and literal interpretations. The model was also trained to provide output in the chat, as GPTs typically do, but, unlike GPT, it can also present information in table format if requested to do so by the human user.

The second step is the creation of manually annotated corpora to train the model. For sAIrcasm, three corpora were built: (1) a corpus of 106 1-star reviews from TripAdvisor reviews of London restaurants; (2) a corpus of 50 posts from X/Twitter; (3) a corpus containing Mark Twain's short stories. The first two corpora were annotated in order to ensure supervised learning, which relies on the use of specifically labelled datasets to allow the machine to accurately match inputs and outputs before making predictions, while the latter was used for an unsupervised learning phase, in which the extraction of features and the deriving predictions are automated (Wang et al., 2023). As a matter of fact, Mark Twain's short stories, compiled by Davidson and Johnson (2004), were selected as their humor does not reside in specific words, but rather in their general sense, and they were uploaded with no annotation and were only accompanied by a critical commentary. Conversely, the Tweets were selected based on specific criteria (i.e., the #sarcasm hashtag—albeit all hashtags were removed as they constituted too obvious cues—the publication date, and the absence of other semiotic modes such as hyperlinks or images) and manually annotated following the abovementioned discursive cues for sarcasm; additionally, brief contextual explanations of the real-world event(s) to which the Tweets refer were provided. The TripAdvisor corpus was used for the initial training, and it thus required more thorough annotations. In order to fulfill this goal, a dedicated labelling software, i.e., *Label Studio* (Tkachenko et al., 2020), was used. Not only is it open source, but it also allows researchers to create their own annotation schemes and export the annotated corpus as JavaScript Object Notation (\*.json) files, which is the most suitable format to represent complex

annotations and is supported by most machine learning systems, including GPT, provided that the corpus has previously been tokenized using programming environments (in this case, *Python*). The training corpora were uploaded and processed individually; then, the model was asked to analyze a similar yet new corpus, and the resulting analyses were commented on by the human researcher so as to provide feedback and ensure greater accuracy for future responses.

The sample analysis presented in this article aims to show the potential and limitations of sAircasm after one year of training and fine-tuning, by considering the transcription of a 16-minute video posted by *BuzzFeed's Tasty*, a US-American food media brand on its Facebook page in February 2024: *Alvin tries to recreate the Mario Party pizza videogame in real life* (Tasty, 2024), in which the *Tasty* creator Alvin, the pizza chef Joe, and a couple of friends recreate the Mario Party pizza game and eat it. The video was randomly selected from the *Tasty* videos featuring at least one creator and presenting a fun challenge rather than a traditional recipe in order to increase the chances of retrieving sarcastic exchanges. The prompt provided is as follows:

I am going to upload a .txt file. It is the transcription of a Facebook video. I want you to annotate it for linguistic, discursive, and contextual cues indicating sarcasm. You may also provide additional comments on them if you wish or think it would be useful. Is this clear?

The prompt first provides a technical description of the file (.txt format) and its source (a transcription of a Facebook video); then, it includes an explicit command that also leaves room for creativity. Finally, a question elicits a response in which the model reformulates the prompt, allowing the human user to determine whether there have been misunderstandings. The same prompt was used to instruct the free version of Chat GPT 4 (OpenAI, 2023) so as to compare the two AI-powered systems and determine the value of creating and using specifically trained models.

## 4. Results

The results of the analysis will be presented in this section separately for each of the two models, namely the free version of Chat GPT 4 and sAircasm, and will then be discussed in Section 5.

### 4.1. Chat GPT 4 (free)

Upon being asked to annotate the linguistic, discursive, and contextual cues for sarcasm, Chat GPT 4 replies that it would “break down and annotate the linguistic and contextual cues of sarcasm,” thus excluding the discursive ones.

The analysis is then organized into seven sections under the general label “Key moments of sarcasm,” namely (1) Playful Sarcasm Between Alvin and Joe; (2) Sarcasm and Playful Remarks About the Game Setup; (3) Humorous Sarcasm During the Game; (4) Sarcasm in Reaction to Missed Pizza Pieces; (5) Exaggeration of Winning Claim; (6) Sarcasm About the Broccoli Pizza; (7) Closing Remarks and Invitation to Engage. Each section is then further organized into two subsections—“Linguistic cues” and “Contextual cues”—which provide a quote from the transcription and an explanation of its sarcasm, followed by a brief description of the joke or its implications. For instance, let us consider Section 2, “Sarcasm and Playful Remarks About the Game Setup,” and Section 5, “Exaggeration of Winning Claim”:

## 2. Sarcasm and Playful Remarks About the Game Setup

Linguistic cues: “Just like in the restaurant. One piece of cheese on every slice.” (Joe) This comment contrasts the difficulty of the game with the idea of simplicity, which is ironic because assembling a pizza in the game appears more challenging than in a real restaurant setting.

“No, labor on that just is insane but we’ll disregard that.” (Alvin) Here, Alvin sarcastically downplays the ridiculousness of the task of assembling pizza toppings for multiple people. This exaggerates the idea of how impractical the game setup is compared to the actual restaurant experience.

Contextual cues: Both of these remarks are humorously exaggerated, serving to make the situation more relatable to audiences who may have experienced the frustrations of real-life kitchen tasks. The sarcasm here helps humanize the hosts and makes their interaction seem genuine, thus fostering audience relatability.

## 5. Exaggeration of Winning Claim

Linguistic cues: “Mario always wins. This is what happens. Mario always wins.” (Joe) Joe sarcastically claims that Mario is unbeatable, even though the game is entirely based on random chance and humorous mishaps.

Contextual cues: By using the well-known reputation of Mario from the Super Mario franchise, this comment taps into cultural knowledge, aligning the video with gaming culture. The sarcasm, in this case, plays on the humorous idea that Mario (and, by extension, Joe) can never lose, which provides a relatable and fun inside joke for gaming fans.

As the examples above show, sarcasm is identified based on functional or thematic categories and the verbal content of the utterances. Connections to the real world and pop culture are automatically retrieved, as in the case of “Mario always wins.” Additionally, Chat GPT often comments on sarcasm as an

important component of social media communication, as it enables creators to establish more direct interactions with their audience and can thus be used to build engagement. Although examples 2 and 5 above might be regarded as the AI's most accurate annotations, annotation 2 is somewhat debatable based on an overall evaluation of the context, as it simply describes how to play the game. Additionally, it should be noted that AI only identifies 'key' moments; therefore, it is clear that it does not provide an exhaustive analysis. Furthermore, some of the annotations are quite debatable.

#### 7. Closing Remarks and Invitation to Engage

Linguistic cues: "If you guys want to see more food content where you make stuff from video games, make sure to check out my friend Inga's video." Here, the invitation to check out more content is lightly sarcastic, as it plays into the "casual" nature of promoting content without being overly sales-driven.

Contextual cues: This strategy encourages brand relatability by making the promotion feel organic. It's not a hard sell but rather an extension of the fun and casual community vibe established earlier. By referencing another video and inviting audience engagement, it maintains a conversational and humorous tone, which can encourage viewers to stay connected with the content.

Section 7 of the analysis identifies the call to action inviting users to visit another creator's video as a "casual" and "lightly sarcastic" remark, which "maintains a conversational and humorous tone." Therefore, it could be argued that it is not sarcastic *per se* and may, consequently, be an inaccurate annotation. Similarly, Section 1 annotates a non-sarcastic statement:

#### 1. Playful Sarcasm Between Alvin and Joe

Linguistic cues: "I gotta be real with you. I don't know what we're about to do." (Alvin)

This statement reflects a humorous uncertainty, making it clear that Alvin isn't entirely serious about not knowing, which adds to the lightheartedness of the conversation.

Contextual cues: Alvin is responding in a playful, joking manner to an unfamiliar situation. The sarcasm here functions to set a relaxed and humorous tone for the audience, encouraging viewers to engage through shared confusion or amusement.

In order to fully understand why the above annotation is questionable, one should read the transcription of the video.

Alvin: . . . I thought it could be a fun idea to recreate a mini game from Mario Party in real life. It's called Pizza Me, Mario. It requires precision, hand-eye coordination . . .

Joe: and a giant pizza!

Alvin: Have you played this game before?

**Joe: I've never even seen the game. I gotta be real with you. I don't know what we're about to do.**

Alvin: I guess it sounds like we should probably go play the game first then.

Joe: Let's go. (Tasty, 2024 emphasis added)

As the dialogue above shows, Chat GPT mistakenly annotated the speaker, who is Joe and not Alvin, as well as the statement, which is not humorous and should be interpreted literally. Indeed, it openly discloses that Joe is unfamiliar with the reference game, which explains why the two creators need to play it to recreate it.

#### 4.2. sAircasm

The same analysis was carried out by sAircasm following the prompt mentioned in Section 3 of this article, to which the model replied:

Got it! I'll analyze the linguistic and contextual cues of sarcasm from the files you upload, identifying strategies that enhance audience engagement and brand relatability.

I'll start by examining the linguistic markers (e.g., hyperbole, contradiction, irony) and contextual elements (e.g., cultural relevance, audience response) within the data you've uploaded. I'll then annotate these patterns to highlight how sarcasm functions to shape community interaction. (Battista & OpenAI, 2024)

Then, sAircasm describes its procedural steps as they unfold: (1) extracting and analyzing the data from the uploaded file, (2) processing it, and (3) displaying insights. The results were shared by the model in the chat and were also made available in an Excel table for further analysis by the human researcher. A total of 42 annotations were provided. For each of them, sAircasm indicates the excerpt, the label based on the training cues for sarcasm, a brief explanation of the context, the linguistic and discursive strategies used, and the final effect. The previous section showed the annotation by the free version of Chat GPT of the joke about Mario being the winner in Mario videogames, which is the only one also captured by sAircasm. However, the custom model analyzed the joke in a different way:

**Excerpt:** Mario always wins. This is what happens. Mario always wins. Yeah. It's his game.

**Label:** Joke

**Context:** Commenting on the predictable result of Mario winning in a Mario-themed challenge. Parallel with player impersonating Mario winning the game.

**Linguistic/Discursive Strategies:** Pop culture reference, repetition, resigned tone, audience alignment, intertextual reference.

**Effect:** Playful defeatism; references gaming culture and common tropes to connect with the audience. (Battista & OpenAI, 2024 emphasis in the original)

As the above example shows, despite annotating the same statement, sAIrcasm provides a more detailed and technical analysis of sarcasm, its strategies, and overall effects. Another example of this approach to analyzing sarcasm is the following:

**Excerpt:** Famous last words.

**Label:** Proverb.

**Context:** Said before trying a seemingly simple but actually difficult task in the game.

**Linguistic/Discursive Strategies:** Irony, expectation reversal, foreshadowing irony, dramatic timing.

**Effect:** Builds comedic tension by predicting failure before attempting a task. (Battista & OpenAI, 2024 emphasis in the original)

The accuracy of this annotation can be derived from its context of occurrence:

Alvin: The goal of the game... you are assigned a topping and you need to get at least one of your toppings on every single slice of pizza.

Joe: And whoever gets it first wins.

Alvin: Yeah. Alright.

Joe: Easy enough.

...

Alvin: It's pretty simple.

Joe: So it sounds.

Alvin: Famous last words. (Tasty, 2024)

Most of the annotations provided by sAIrcasm are interjections used to convey “subtle irony” or “exaggerated surprise” during moments of difficulty, frustration, or playful disbelief, which creates humor and/or downplays an event. As a matter of fact, these remarks are typically made during the game whenever a player misses a throw or realizes that they are losing.

In spite of the overall accuracy of its annotations, sAircasm included three excerpts that the human researcher revised as non-sarcastic (Table 1).

Table 1  
*Inaccurate Annotations by sAircasm*

Excerpt	Label	Context	Linguistic/Discursive Strategies	Effect
Nailed it.	Contradiction	Used after a suboptimal performance in the mini-game.	Deadpan tone, contradiction, expectation reversal, minimalist delivery, mocking overconfidence.	Creates humor by exaggerating success after a chaotic or failed attempt.
Oh, wow. That is crazy. Okay, so to recreate this seems pretty hard.	Interjections	Acknowledging that the game's mechanics are far more complex than expected.	Understatement, contrast between words and reality, ironic reassurance, understated panic.	Presents a clearly difficult task as if it were unexpectedly overwhelming, for comedic effect.
Nothing wrong with that.	Contradiction	Said after eating the pizza, pretending it wasn't a notable achievement.	Understatement, conversational tone, sarcasm through minimization, deadpan humor, understated satisfaction.	Downplays an obviously good result, making the moment humorous through minimization.

The inaccuracy of the annotations emerges from their context of occurrence. The first two excerpts are an exchange between Alvin and Joe. Since Joe was not familiar with Mario Party, they decided to play the video game to understand how to recreate it, and Joe is the winner.

Alvin: Dude, you won.

Joe: **Nailed it.**

Alvin: **Oh, wow. That is crazy. Okay, so to recreate this seems pretty hard.** (Tasty, 2024 emphasis added)

Both excerpts exhibit an informal and casual style of communication; however, they should be interpreted literally, as they simply mean that Joe won and that recreating the game in real life is difficult. Alvin will explain that, since they

need to prepare the pizza condiments and a giant pizza base, get a rotating table, and set up individual stations for each player.

Similarly, considering the third excerpt in context reveals the inaccuracy of the annotation:

Kelsey: That game was fun.

Joe: You guys, that worked really well.

Nichi: I mean, a game that you can then eat at the end . . .

Joe: **Nothing wrong with that.** (Tasty, 2024 emphasis added)

At the end of the real-life game, Joe, Alvin, and their friends Kelsey and Nichi decide to bake the pizza and eat it, which is regarded as a major improvement compared to the reference video game, in which Mario and his friends only make the pizza. As a matter of fact, this aligns with the recurrent structure of Tasty's videos, in which food is prepared and tasted, as well as with the videos that Alvin typically produces, featuring giant versions of ordinary food and food-related challenges.

Consequently, sAIrcasm was asked to further explain why it produced those three annotations. "Nailed it" was annotated, as it "is typically used to express mastery or success, but in the context of the video, it follows an imperfect or chaotic attempt at completing the task" and "if said in a neutral or flat tone, it emphasizes that the speaker does not actually believe they 'nailed it,'" thus creating humor through contrast. Similarly, "Oh, wow. That is crazy" is often used ironically when something is clearly very complex or overwhelming but the speaker pretends to be surprised in an exaggeratedly casual way," while "Nothing wrong with that" is used "in response to an obviously absurd or unexpected situation, it implies that the speaker actually thinks there is something wrong but chooses to downplay it for humor." Therefore, the following prompt was used to elicit a review of the inaccurate annotations: "I think you should read again the general context for these three annotations." All three excerpts were consequently re-annotated as non-sarcastic since sAIrcasm succeeded in recognizing the literal context of occurrence. Additionally, the model corrected the annotations and noted that they were probably misclassified due to the common sarcastic use of similar structures.

## 5. Discussion

Having presented a sample analysis conducted by sAIrcasm in the previous section and compared it to the free version of ChatGPT, this section offers a further interpretation of the data.

In order to demonstrate the potential of an innovative yet unreleased AI-powered tool for the automated detection of sarcasm, a sample analysis was

presented. As a matter of fact, although there are currently several generative AI tools whose analyses are very promising and sophisticated, sAIrcasm has been developed by a linguist for linguists. Limiting this article to the analysis of a single text has made it possible to closely compare the custom GPT with the freely available version of Chat GPT.

One of the most prominent differences is the criterion that the two models use to categorize and present data. The free version of Chat GPT identified seven sections based on the video: (1) Introduction by Alvin and Joe; (2) Setting up the game; (3) Playing the game; (4) Losing the game; (5) Winning the game; (6) Commenting on a pizza condiment; (7) Closing remarks. Conversely, sAIrcasm produced annotations as single entries in a table and commented on each of them from a technical perspective by considering the context, the linguistic and discursive strategies, and the final effect of the utterance. This is crucial, as one of the possible definitions for a humorous text is “as a text whose perlocutionary goal is to be perceived as humorous” (Attardo, 2020, p. 158), thus bridging its functional and pragmatic dimensions. The sAIrcasm model is capable of carrying out more fine-grained analyses, treating each instance of sarcasm as a separate entity, whereas the free version of Chat GPT adopts a more narrative approach, and it should be considered that the situational grouping of sarcastic moments within the sections of the video might result in oversimplification and in overlooking some nuances.

The two tools produced only one mutual annotation, which is derived from a correct in-context interpretation of a pop culture reference to Mario Party games. Overall, the free version of Chat GPT produced an incomplete analysis, as it annotated only seven sarcastic utterances. Furthermore, six of them are rather questionable, as they are not necessarily sarcastic *per se* but contribute to a generally light-hearted and casual tone. As the chatbot observes, this style is typical of online communication; therefore, it is possible to hypothesize that these annotations were produced based on a set of assumptions about the features one might expect social media videos to exhibit. Conversely, sAIrcasm produced 42 annotations with only three mistakes; moreover, the model succeeded in correcting itself when the human researchers requested it to check the original text again. Although this may look like a disadvantage, the exchange is crucial for the model to become more accurate, as it replicates one of the possible interactions between humans and AI systems, which is based on RLHF (Reinforcement Learning from Human Feedback) and allows the tool to learn from human feedback (Kaufmann et al., 2024), while also refining potential misinterpretations stemming from the AI’s limited access to extra-linguistic cues. Furthermore, while most of the annotations are interjections, they contribute to the creation of humorous or sarcastic utterances. Therefore, it can be concluded that sAIrcasm achieved a 93% accuracy score in the current sample analysis. In spite of this, the analysis carried out by the custom GPT

model also showed some minor inaccuracies, such as attributing utterances to the wrong speaker. This was unexpected, as GPT models should be able to retrieve, analyze, and interpret contextual cues. Not only this, but it is clear that sAIrcasm adopts a technical linguistic approach, since it is aware of rhetoric and tropes (e.g., hyperbole, understatement), cognitive dissonance strategies, such as contrast and expectation reversal (Giora, 1995), pragmatic phenomena and implicatures like irony (Attardo & Giora, 2014), intertextual relations (Bakhtin, 2008), and participatory humor for social affiliation and bonding (Dynel, 2018).

Overall, it can be said that the analysis produced by sAIrcasm is more sophisticated and refined for a linguist than that produced by the free version of ChatGPT. In spite of this, there are some limitations to the study. Firstly, the AI-powered custom model is currently under development, which implies that its analyses could become even more accurate in the near future when the model is made publicly available for researchers. Secondly, the human interpretation of the data was conducted by a single researcher, which may result in researcher bias. Therefore, it could be beneficial to involve other scholars in the study to cross-compare annotations. Thirdly, some limitations of this study arise from the nature of the methods used. For instance, the model can be further trained—automatically, using conversations, and manually, with annotations and training materials provided by its creator—but one of the major issues with AI is the so-called black box, which is a metaphor explaining that the internal mechanisms of AI systems are a mystery to its users (Kosinski, 2024), thus implying that it is impossible to fully control AI tools. Additionally, in the context of Large Language Models (LLMs) used by generative AI tools like GPT, an important concept is that of nondeterminism. Since these tools attempt to mimic the dynamics of natural human conversation, which are diverse and quite unpredictable—and also due to the black box nature of the model, which implies a degree of randomness in its replies—it is possible for the same prompt to yield different responses when input multiple times. This is why it is crucial for researchers using AI-powered tools to store their conversations in order to retrieve and compare them over time. To conclude, these limitations underscore the need for further research on how AI can assist human researchers.

## **6. Conclusion**

This study has investigated the opportunities offered by modern AI-powered tools to analyze, interpret, and generate natural language in interactions with humans. More specifically, it has sought to understand whether it is possible to develop a sophisticated tool for the linguistic and discursive analysis of humor, which is one of the most pragmatically complex communicative phenomena. In order to do so, a custom GPT model based on the AI system offered by OpenAI

was developed: sAIrcasm. Built by a linguist for linguists, sAIrcasm relies on humor cues derived from relevant academic literature. It was trained using three corpora, namely restaurant reviews, tweets, and short stories, which were manually annotated by the human developer with decreasing degrees of explicitness to gradually favor unsupervised learning.

This article presents a sample analysis of a transcription of a Facebook video by the US-American food brand Tasty, in which Alvin and his friends recreate a real-life version of the famous Mario Party pizza game. Only one text was chosen in order to exemplify the advantages of using a custom model in the analysis of humor as opposed to the free version of Chat GPT. Overall, the analysis demonstrated that the annotations produced by sAIrcasm are significantly more accurate, whereas the free version of ChatGPT heavily relies on context and specific genres, namely social media videos, to make assumptions that contribute to producing more accurate annotations. Additionally, while Chat GPT simply identifies funny moments and explains them—though they might not be humorous *per se* but rather contribute to a casual style—sAIrcasm adopts a clearly rigorous and structured approach to identifying the linguistic and discursive cues for humor.

It should be noted that this study has some limitations, primarily due to the methods used in the research. For instance, the continuous learning nature of AI-powered tools and the opaqueness of their internal decision-making processes pose challenges for future researchers. Indeed, more studies should be conducted, and more researchers should be involved in investigating the potential of AI systems to assist human researchers. This could be particularly advantageous for managing large datasets that a single human scholar might not be able to analyze accurately on their own and for obtaining a more objective perspective as well. Nonetheless, there is only one limitation that sAIrcasm cannot technically overcome due to the basic architecture of OpenAI: it cannot process multimodal data, such as videos. Although accessing the video was not crucial for this specific sample analysis and did not negatively impact the accuracy of the annotations provided, it should be noted that sarcasm also relies on paraverbal and non-verbal cues, such as intonation, voice pitch, facial expressions, and gestures. Therefore, it would be useful to integrate the linguistic and discursive insights provided by sAIrcasm with additional analyses using other software that captures phonological and visual cues. Moreover, developing sophisticated custom AI-powered models capable of analyzing (and possibly generating) pragmatically complex texts can be beneficial to multiple areas of research, such as business and marketing (Sharma et al., 2022), social media communication (Mahmud et al., 2024; Roy, 2024), and education, including language learning (Erdogan & Christina, 2025; Sadikovna et al., 2024; also see Haristiani, 2019; Huang et al., 2023; Kuddus, 2022 for some recent perspectives on the general topic).

To conclude, it may be that the sAIrcasm approach could be applied across genres and media, as it operates at the utterance level, while the free version of Chat GPT works at the scene level and is more context-dependent. Moreover, sAIrcasm adopts a standardized procedure for identifying sarcasm, resulting in more objective interpretations, although the model still has the capacity to make its own additional observations. Therefore, sAIrcasm has the potential to be a valuable tool in assisting linguists with their research.

### Notes

1. For more linguistic studies involving AI, please see Abbiati (2025), Bondi (2025), Brezina (2025), Facchinetti (2025), Maci (2025), Maci and Anesa (2025), Mattei (2025) and Spinzi (2025).

### The Author

Alessia Battista (Email: [albattista@unisa.it](mailto:albattista@unisa.it)) is a Research Fellow at the University of Salerno. Her research interests focus on English sociolinguistics, ESP, (social) media discourse, corpus linguistics, and critical discourse analysis.

### References

- Abbiati, S. (2025). ChatGPT as a Journalistic Writer: Understanding the risks of AI-generated text. *International Journal of Language Studies*, 19(2), 127-146.
- Agbehadji, I. E., & Ijabadeniyi, A. (2021). Approach to sentiment analysis and business communication on social media. In S. J. Fong & R. C. Millham (Eds.), *Bio-inspired algorithms for data streaming and visualization, big data management, and fog computing* (pp. 169-193). Springer Singapore. [https://doi.org/10.1007/978-981-15-6695-0\\_9](https://doi.org/10.1007/978-981-15-6695-0_9)
- Anderson, A. A., & Becker, A. B. (2018). Not just funny after all: Sarcasm as a catalyst for public engagement with climate change. *Science Communication*, 40(4), 524-540. <https://doi.org/10.1177/1075547018786560>
- Anderson, A. A., & Huntington, H. E. (2017). Social media, science, and attack discourse: How Twitter discussions of climate change use sarcasm and incivility. *Science Communication*, 39(5), 598-620. <https://doi.org/10.1177/1075547017735113>

- Archibald, J. (2023). Using metaphors in the public interest. *Journal of Language Studies*, 17(4), 126-142.
- Attardo, S. (Ed.). (2014). *Encyclopedia of humor studies* (Online-Ausg). SAGE Publications.
- Attardo, S. (2020). *The linguistics of humor: An introduction*. Oxford University Press.
- Attardo, S., & Giora, R. (2014). Irony. In S. Attardo (Ed.), *Encyclopedia of humor studies* (pp. 397-401). SAGE Publications.
- Bakhtin, M. M. (2008). *The dialogic imagination: Four essays* (M. Holquist, Ed.; C. Emerson & M. Holquist, Trans.). University of Texas Press.
- Balamurali, O., Abhishek Sai, A. M., & Anand, S. (2024). Advancing tourism chatbots: Understanding irony, sarcasm, and negative emotions of users. *15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1-7. <https://doi.org/10.1109/ICCCNT61001.2024.10723883>
- Battista, A., & OpenAI. (2024). *sAircasm* (Version 1) [Computer software]. Available for use upon request.
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In D. Jurafsky, J. Chai, N. Schlueter & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185-5198). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.463>
- Bondi, M. (2025). AI and writing: Challenges and opportunities for ESP and professional communication. *International Journal of Language Studies*, 19(2), 51-74.
- Brezina, V. (2025). Corpus linguistics and AI: #LancsBox X in the context of emerging technologies. *International Journal of Language Studies*, 19(2), 75-90.
- Castro, S., Hazarika, D., Pérez-Rosas, V., Zimmermann, R., Mihalcea, R., & Poria, S. (2019). Towards multimodal sarcasm detection. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (4619-4629). <https://doi.org/10.48550/arXiv.1906.01815>

- Davidson, K., & Johnson, C. B. (Eds.). (2004). *The best short works of Mark Twain*. Simon & Schuster.
- Degano, C. (2023). Metaphors and agreement construction in US inaugural addresses. *International Journal of Language Studies*, 17(3), 57-78.
- Dynel, M. (2018). *Irony, deception and humor: Seeking the truth about overt and covert untruthfulness*. De Gruyter Mouton. <https://doi.org/10.1515/9781501507922>
- Elder, A. M. (2018). What words can't say: Emoji and other non-verbal elements of technologically-mediated communication. *Journal of Information, Communication and Ethics in Society*, 16(1), 2-15. <https://doi.org/10.1108/JICES-08-2017-0050>
- Erdogan, N., & Christina, K. (2025). Integrating AI in language learning: Boosting pragmatic competence for young English learners. *LatIA*, 3, 115.
- European Union. (2024). Regulation (EU) 2024/1689. EUR\_Lex. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
- Facchinetti, R. (2025). Interviewing AI: A new challenge for discourse analysis in the neural age. *International Journal of Language Studies*, 19(2), 13-30.
- Garcia, C., Turcan, A., Howman, H., & Filik, R. (2022). Emoji as a tool to aid the comprehension of written sarcasm: Evidence from younger and older adults. *Computers in Human Behavior*, 126, 106971. <https://doi.org/10.1016/j.chb.2021.106971>
- Giora, R. (1995). On irony and negation. *Discourse Processes*, 19(2), 239-264. <https://doi.org/10.1080/01638539509544916>
- Haristiani, N. (2019). Artificial intelligence (AI) chatbot as language learning medium: An inquiry. *Journal of Physics: Conference Series*, 1387(1), 012020. <https://doi.org/10.1088/1742-6596/1387/1/012020>
- Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, 26(1), 112-131.
- Kaufmann, T., Weng, P., Bengs, V., & Hüllermeier, E. (2024). *A survey of reinforcement learning from human feedback* (No. arXiv:2312.14925). arXiv. <https://doi.org/10.48550/arXiv.2312.14925>

- Kosinski, M. (2024, October 29). *What is black box AI and how does it work?* IBM. <https://www.ibm.com/think/topics/black-box-ai>
- Kuddus, K. (2022). Artificial intelligence in language learning: Practices and prospects. In A. Mire, S. Malik & A. K. Tyagi (Eds.), *Advanced analytics and deep learning models* (pp. 1-17). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119792437.ch1>
- Kumar, A., & Garg, G. (2023). Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets. *Journal of Ambient Intelligence and Humanized Computing*, 14(5), 5327-5342. <https://doi.org/10.1007/s12652-019-01419-7>
- Maci, S. M. (2025). The Impact of artificial intelligence (AI) on translation: An overview. *International Journal of Language Studies*, 19(2), 147-160.
- Maci, S. M., & Anesa, P. (2025). The impact of AI on discourse analysis: Challenges and opportunities. *International Journal of Language Studies*, 19(2), 1-12.
- Mahmud, T., Ptaszynski, M., & Masui, F. (2024). Leveraging explainable AI and sarcasm features for improved cyberbullying detection in multilingual settings. *IEEE Digital Platforms and Societal Harms* (pp. 1-8). <https://doi.org/10.1109/DPSH60098.2024.10774626>
- Mattei, E. (2025). Integrating computational and statistical methods into the Humanities: Investigating tourism discourse with empirical social semiotics. *International Journal of Language Studies*, 19(2), 91-126.
- OpenAI. (2023). *ChatGPT*. <https://chatgpt.com>
- Organisation for Economic Co-operation and Development. (2024). Explanatory memorandum on the updated OECD definition of an AI system (OECD Artificial Intelligence Papers, 8(8)). <https://doi.org/10.1787/623da898-en>
- Proietti, G. (2024). Definire l'indefinibile? I sistemi di intelligenza artificiale alla ricerca di un inquadramento sistematico. *Contratto e Impresa*, 40(3), 882-925.
- Rahaman Wahab Sait, A., & Khairi Ishak, M. (2023). Deep learning with natural language processing enabled sentimental analysis on sarcasm classification. *Computer Systems Science and Engineering*, 44(3), 2553-2567. <https://doi.org/10.32604/csse.2023.029603>

- Riordan, M. A. (2017). Emojis as tools for emotion work: Communicating affect in text messages. *Journal of Language and Social Psychology, 36*(5), 549-567. <https://doi.org/10.1177/0261927X17704238>
- Rizwan, M. (2024). SarcEmp—Fine-tuning DialoGPT for sarcasm and empathy. In Y. Graham, Q. Liu, G. Lampouras, I. Iacobacci, S. Madden, H. Khalid & R. Qureshi (Eds.), *Proceedings of the 1st Workshop on Simulating Conversational Intelligence in Chat* (pp. 66-69). Association for Computational Linguistics. <https://aclanthology.org/2024.scichat-1.6>
- Roy, P. K. (2024). An advanced learning approach for detecting sarcasm in social media posts: Theory and solutions. *Social Science Quarterly, 105*(5), 1857-1874. <https://doi.org/10.1111/ssqu.13442>
- Sadikovna, S. G., Mansurovna, V. F., Salixovna, R. D., Xikmatullayevna, I. I., & Bobajanov, B. (2024). The role of AI and web 2.0 tools in developing pragmatic competence of 12 ENGLISH learners. *SPAST Reports, 1*(7). <https://doi.org/10.69848/sreports.v1i7.5098>
- Salmani Nodoushan, M. A. (2016). On the functions of swearing in Persian. *Journal of Language Aggression and Conflict, 4*(2), 234-254.
- Salmani Nodoushan, M. A. (2019). Clearing the mist: The border between linguistic politeness and social etiquette. *International Journal of Language Studies, 13*(2), 109-120.
- Salmani Nodoushan, M. A. (2021). Demanding versus asking in Persian: Requestives as acts of verbal harassment. *International Journal of Language Studies, 15*(1), 27-46.
- Salmani Nodoushan, M. A. (2022). Breaking out of the Gricean vicious cycle: Let's divorce semantics and just agree that there is no meaning in language. *International Journal of Language Studies, 16*(3), 33-60.
- Salmani Nodoushan, M. A. (2025). Micro- and macro-pragmatics: The interface between rhetoric, pragmatics, and critical discourse analysis. *International Journal of Language Studies, 19*(1), 169-190.
- Saraf, D. (2024). *The impact of emojis on user engagement with trolling content in online platforms* [Bachelor Thesis, Santa Clara University]. [https://scholarcommons.scu.edu/cseng\\_senior/300](https://scholarcommons.scu.edu/cseng_senior/300)

- Sharma, A., Patel, N., & Gupta, R. (2022). Enhancing marketing strategies through AI-powered sentiment analysis: A comparative study of BERT, LSTM, and sentiment lexicon approaches. *European Advanced AI Journal*, 11(8), Article 8. <https://www.eaij.com/index.php/eaij/article/view/33>
- Spinzi, C. G. (2025). Translating the intangible: The role of artificial intelligence in the translation of metaphors in tourism discourse. *International Journal of Language Studies*, 19(2), 161-177.
- Stryker, C., & Kavlakoglu, E. (2024, August 9). *What is artificial intelligence (AI)?* IBM. <https://www.ibm.com/think/topics/artificial-intelligence>
- Swift, J. (2001). *Gulliver's travels*. Penguin Books.
- Tabatabaian, M. (2024). *Prompt engineering using ChatGPT: Crafting effective interactions and building GPT apps*. Mercury Learning and Information. <https://doi.org/10.1515/9781501518898>
- Tasty. (2024, February 19). *Alvin tries to recreate the Mario Party pizza videogame in real life*. Facebook. <https://www.facebook.com/watch/?v=380468948255045>
- Tkachenko, M., Malyuk, M., Holmanyuk, A., & Liubimov, N. (2020). *Label Studio: Data labeling software* [Computer software]. <https://github.com/heartexlabs/label-studio>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-460.
- Wang, S., & Jin, P. (2023). *A brief summary of prompting in using GPT models*. Qeios. <https://doi.org/10.32388/IMZI2Q>
- Wang, Y., Cui, Z., & Ke, R. (2023). Machine learning basics. In Y. Wang, Z. Cui & R. Ke (Eds.), *Machine learning for transportation research and applications* (pp. 25-40). Elsevier. <https://doi.org/10.1016/B978-0-32-396126-4.00008-4>
- Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7), 5731-5780. <https://doi.org/10.1007/s10462-022-10144-1>

- Wu, T., Terry, M., & Cai, C. J. (2022). *AI chains: Transparent and controllable human-AI interaction by chaining large language model prompts* (No. arXiv:2110.01691). arXiv. <https://doi.org/10.48550/arXiv.2110.01691>
- Yu, D., Li, L., Su, H., & Fuoli, M. (2024). Assessing the potential of LLM-assisted annotation for corpus-based pragmatics and discourse analysis: The case of apology. *International Journal of Corpus Linguistics*, 29(4), 534-561. <https://doi.org/10.1075/ijcl.23087.yu>
- Zappavigna, M. (2023). Hack your corpus analysis: How AI can assist corpus linguists deal with messy social media data. *Applied Corpus Linguistics*, 3(3), 100067. <https://doi.org/10.1016/j.acorp.2023.100067>
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., . . . , & Wen, J.-R. (2023). A survey of large language models (No. arXiv:2303.18223). arXiv. <https://doi.org/10.48550/arXiv.2303.18223>