Journal of Ethics and Emerging Technologies

IEET

Research Paper

Is ChatGPT Full of Bullshit?

Laura Gorrieri[1] *

[1] FINO Consortium | Università di Torino; laura.gorrieri@unito.it
* Correspondence: laura.gorrieri@unito.it

Abstract: It is undeniable that conversational agents took the world by storm. Chatbots such as ChatGPT (*Generative Pre Trained*) are used for translations, financial advice, and even as therapists, by millions of users every month. When interacting with technology it's important to be careful, especially if we do so by using natural language, since our relationship with artificial agents is shaped by the technology's features and the manufacturer's goal. The paper, organized into three sections, explores the question of whether ChatGPT's production can be described as 'bullshit'. In the first section, the focus is on ChatGPT's architecture and development; in the second a new formulation of the concept of Frankfurt's 'bullshit' is presented, in which its central features of indifference, deception and manipulation are highlighted; in the last section, the title question is tackled, proposing an affirmative answer to it, arguing that ChatGPT can be considered a 'bullshit' generator.

Keywords: Artificial intelligence, Generative AI, Large Language Models, Ethics, Harry Frankfurt, Human-Technology Interaction, Natural Language Processing

1. Introduction

Millions of users every day open the same chat. "Hello! How can I help you today?" - it's a message that may greet each one of them. Conversations in the chat flow easily, it seems that, on the other side of the screen, there's an expert, on every possible topic. One user needs financial advice, another wants to write a Python snippet, and a third is interested in a love poem. They all receive convincing answers. It seems there is no request ChatGPT cannot answer. The chatbot is the spearhead of its manufacturer, the private American company OpenAI, and it is estimated to attract more than 100 million users a month (Tong, 2023). It always has an answer for each request, but could it actually be just a sophisticated bullshit generator? The idea that the production of ChatGPT can be described as 'bullshit' has found its place in the media, nonetheless, there is a lack of studies on this topic, so the robustness of this thesis is yet to be formally explored, which is what this paper proposes to do.

ChatGPT is a conversational agent accessible through OpenAI's website, able to engage in natural language dialogues on a variety of topics. The chatbot does so because it relies on a language model, which is a statistical representation of language that enables the prediction of the most likely word given a preceding text, working as a sort of autocomplete. ChatGPT-4, the latest version of the chatbot so far, is the biggest language model ever built. Thanks to its size, it can adapt to new tasks on the spot, providing an answer on seemingly any topic. Models of this scale require considerable investments and are produced by a handful of wealthy companies only. OpenAI, the manufacturer of ChatGPT, is one of them. Thanks to generous investors, such as Microsoft, OpenAI was able to spend billions of dollars on the development of its products. The company was born as a non-profit organization, with the mission of "[ensuring] that artificial general intelligence benefits all

Citation: Gorrieri, Laura. 2024. Is ChatGPT Full of Bullshit?. *Journal of Ethics and Emerging Technologies* 34: 1. https://doi.org/10.55613/jeet.v34i1.149

Received: 14/10/2024
Accepted: 21/20/2024
Published: 07/11/2024

Publisher's Note: IEET stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

of humanity"[1] and switched to a for-profit model in 2019. In the same manner, they passed from an open-source to a closed-source paradigm, making GPT-3 and the following versions' sources not available to the public. Therefore, what is published about ChatGPT is mediated by its manufacturer. Given the pervasive use of ChatGPT in everyday conversations by a broad user base, the relationship fostered between the chatbot and its users is worth exploring. Therefore, one might ask, as I do in the present paper, whether the production of ChatGPT could be categorized as 'bullshit'.

As we dive into the topic of the paper, the second section introduces the concept of bullshit, followed by a proposed revision. In 1986, philosopher Harry Frankfurt wrote an essay titled *On Bullshit*, in which he isolated and studied the title concept. He argues that 'bullshit' can be defined as a statement made with no concern for truth. The bullshitter is a person who has both no intention to describe reality accurately nor to lie. Bullshitting is therefore well distinguished from lying: whilst the liar knows that they are on the wrong side of the truth, and they feel a certain guilt about it, the bullshitter displays indifference toward their position. The truth-values of their own statements are of no concern to them. For example, a politician might emphatically praise her fellow nation's successes and sense of duty, and at the same time, she could very well not care whether what she's saying is true or false. In this case, she could be interested only in the potential political advantage such a statement would give her, and not whether what she tells is true or false. The concept of 'bullshit' quickly gained momentum from when it was formulated, and it is still present in the current debate. Frankfurt's essay has undeniably initiated a long-lasting discourse, nonetheless, the concept as originally proposed lacks a practical definition. Furthermore, various scholars have critiqued Frankfurt's definition, suggesting adjustments – either to avoid overlaps with other phenomena, to include new types of 'bullshit', or add contributions from other disciplines. In the present paper, I propose a revision of the original concept in light of the criticism that has emerged in the debate. To do so, three operational requirements are formulated, to underscore the deceptive and manipulative nature of bullshit, which emerges in Frankfurt's work itself. These three requirements are the following: (1) the bullshitter disregards truth-values; (2) this disregard is not explicitly stated; (3) the production of bullshit serves an ulterior motive. This definition also has the benefit of being an operational definition: it provides a basis for testing practical cases.

The third and last section of the present paper exams the title thesis. Before going further, it's necessary to define whether interactions with conversational agents, like ChatGPT, fall within the same realm as interactions with other humans. This is not an easy question to answer. In the first place, state-of-the-art chatbots are relatively new, since the technology they rely on was only built in 2017 and refined in 2020. Furthermore, it depends on how we interpret artificial agents: scholars underscore their double nature, since they are products but at the same time display a degree of autonomy. This duality combined with the anthropomorphic features inserted in chatbots encourages users to apply a fictional overlay when interacting with them. Empirical evidence underscores the emotional engagement users can have when interacting with conversational agents - including genuine connections and even romantic feelings. Despite the novelty of studies, they suggest relationships with conversational agents may reach emotional depth similar to human-to-human connections. Circling back to our central thesis, the paper tests the requirements highlighted in the previous section against the current use case. The first requirement states that the bullshitter disregards truth-values. I argue that this requirement is met, as the overall truth value of ChatGPT's production is not taken into consideration, since the chatbot was built to produce text that sounds plausible, not true. The second requirement asserts that said disregard is not explicitly stated in the interaction. In ChatGPT's use case, it is satisfied since the chatbot's indifference to truth is not disclosed, and the chatbot presents every answer with confidence without acknowledging the

[1] The mission is declared by the company on OpenAI's website (https://openai.com/about).

probabilistic nature of its responses. The final requirement declares that the production of bullshit serves an ulterior motive. In this case, the requirement is fulfilled since ChatGPT was manufactured with an underlying goal of retaining users within the chat, benefiting the manufacturer OpenAI by encouraging widespread adoption. Since the three requirements are met, one can consider ChatGPT's production as 'bullshit'.

In conclusion, it is important to underscore how the reframing of ChatGPT's production as 'bullshit' is not meant to undermine the product's capabilities: the goal of the present work is to investigate users' relationship with chatbots and to mitigate the risks of unquestionably trusting generated content.

2. What is ChatGPT?

In this first section, the focus is ChatGPT, an LLM (*Large Language Model*) used for conversational tasks. Since the present paper is about an AI (*Artificial Intelligence*) product, some knowledge of how the chatbot works is required. In this section, I will outline what ChatGPT is, by presenting its architecture and development by the manufacturer, OpenAI.

2.1. Definition and architecture

ChatGPT is a chatbot, accessible through OpenAI's website, via a chat[2]. The interaction with the chatbot is in natural language, and the user can ask about virtually anything. For example, one may ask "Which are the planets of our Solar System?" and the bot responds something along the lines of "The planets in our Solar System, listed in order of their distance from the Sun, are: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune". Or ask for a love novel recommendation, and the bot can provide a list which contains *Pride and Prejudice*, *The Notebook*, and *Outlander*, with the author and description for each one of them. But how is an AI product capable of chatting in natural language?

ChatGPT is a conversational language model built on GPT (*Generative Pre-Trained*) and fine-tuned to replicate human dialogue.[3] A language model is a system that predicts the probability of a word based on the text that precedes it - a sort of autocomplete. This is done by employing a probabilistic approach to language, that can be only syntactical, or it can be enhanced by some form of computational semantics[4]. The underlying idea behind language models is that once one is provided with enough examples during the training process, then it is able to predict what comes next in any given sentence, by selecting the most plausible word from a known dictionary. Language models are not new, the theory can be traced back to at least 1948[5]: they have been used in NLP (*Natural Language Processing*) for decades, with satisfactory performances. However, language models were revolutionized in 2017, when Google released its Transformers (Vaswani et al., 2017), the first LLMs. Transformers were a huge innovation in the field: they could handle contextual information much better than other neural networks and proved to be easily adaptable to new tasks. Another important feature of these models was the use of *masking* during the training process. With masking, data used for the training does not

[2] It can also be accessed programmatically through API (*Application Programming Interface*).

[3] Fine-tuning is a process in which a model trained for other purposes is trained specifically for a new task. It's a way to build on already learnt capabilities, instead of training a new model from scratch.

[4] Many models use distributional semantics, a technique that aims to represent the meaning of words and phrases based on their distributional patterns in large corpora of texts. It is still debated if these approaches to semantics can be considered proper semantics or not.

[5] The first language models can be considered the introduction of the n-grams, sequences of words of length *n* used to find recurring phrases, whose first mention is in Shannon (1948).

need costly human validation – as neural networks usually did - but the raw text is sufficient. The process is based purely on statistics: in a given sentence, a word is masked, and the model has to guess which token is hidden in that particular spot. To do so, the model leverages knowledge learnt from similar cases. For example, in the sentence "Take the umbrella, it is raining", one could mask the last word and ask the model to fill in the gap: "Take the umbrella, it is ___". In an overwhelming majority of previously seen examples, the blank word would be "raining", therefore the model should be able to guess the missing word successfully. The innovation of large language models is that they possess the capabilities of neural networks, whilst leveraging the feature of being trained in an unsupervised manner – i.e., with no human validation. Therefore, they are easier to train and can be scaled up with minimum effort, since they do not require manually labelled data. The trade-off is that when scaling up they need much more computational power.[6]

The year after the release of Google's Transformers, OpenAI started to work on the first GPT. Based on the Transformers' architecture, the models of the GPT family are much bigger[7]. For example, Transformer BERT Large has 110 million parameters,[8] whilst GPT-4 has 1 trillion parameters (OpenAI, 2023). The model size is crucial for its performance: the bigger the network is, the more capable of generalizing it gets.[9] This is proven by the ability of LLMs to perform zero/one/few shots learning – a type of training in which a model is asked for a prediction after seeing zero/one/few instances of that kind (Brown et al., 2020). For example, a network could have been trained to distinguish pictures of cats, dogs, and bears. When the model is provided with just one example of a picture of a lion, with its label, from that moment on it is capable of recognizing new pictures of lions with good accuracy. This is possible because the model builds upon previous knowledge of other animals. With GPT-3 – with a staggering size of 175 billion parameters (Brown et al., 2020) - we have seen a step further. The model not only can be trained with zero/one/few shots but is able to shift tasks just as easily. This capability is called *task transfer* and describes the ability of a model to perform new tasks only by providing zero or a few examples. This would be the case of a model trained in translation from one language to another, able to perform this task with good accuracy, and just after a few examples capable of performing another task, such as writing poetry. This is exactly what happened with models of the GPT family from version 3 and above. The result was not expected by researchers, it came with the scaling up of the employed neural networks. In this sense, it is possible to say that "GPT-3 has learned to learn" (Romero, 2021).

2.2. *Development*

GPT-4 is the latest language product released by OpenAI, but as said, GPT models go back to 2018. That year OpenAI presented GPT-1, the first of the GPT family (Radford & Narasimhan, 2018). This version was built on Transformers' architecture, and it was trained in a semi-supervised manner[10] – differently from what happened with its

[6] For GPT-3, for example, it is estimated the computational training costs alone are around $4.5 million, using more than two thousand GPUs (*Graphics Processing Units*) (Vanian, 2023).

[7] This comparison is usually done by comparing the number of parameters of the network. A parameter represents a variable and it is directly proportional to the number of neurons a network has.

[8] See Huggingface transformers documentation page at: https://huggingface.co/transformers/v3.3.1/pretrained_models.html

[9] This is possible because bigger networks can learn from bigger datasets, where one could find more examples. Also, having more examples means having more noise, so it is less likely to overfit the data.

[10] In a semi-supervised approach, the training has two steps: a pre-training in which the network is trained in unsupervised mode, i.e. without human-labelled data, and a fine-tuning training with supervised data, i.e. with human-labelled data.

successors. The next year GPT-2 was launched. The structure is basically the same as the predecessor, although it is ten times bigger and starts to perform zero-shot task transfer (Radford et al., 2019). In 2020 OpenAI released GPT-3. Two hundred times bigger than the previous version, was the biggest neural network built by then (Brown et al., 2020). The size of this version is a major improvement from older models, and it showed that with a quantitative improvement came a qualitative improvement. In 2023 GPT-4 is developed following the same maxim. It is trained on a bigger dataset, presents 1 trillion parameters (OpenAI, 2023) and is the biggest LLM in history so far. Another difference between GPT-3+ and previous models is the availability of the source code. GPT-1 and GPT-2 are open-source, whilst from GPT-3 models are closed-source. This means that the model itself is not available: the user calls a service, provided by OpenAI, and receives its response back. What we know about the latest GPT models we do through the documents the company published and the interactions we can have with them.

As said, these models have been manufactured by OpenAI, which is an American private company[11] that deals in the AI market. Lately, this market has seen an impressive expansion, and it is expected to reach a value of over one trillion dollars by 2032 (Bloomberg Press Release, 2023). In the last few years, research in AI has been heavily influenced by tech companies (Hagendorff & Meding, 2021), and nowadays private models are the most influential (Jurowetzki et al, 2021). It is important to highlight that only a handful of companies can develop state-of-the-art LLM technology since it requires considerable investments. OpenAI closed 2022 with a $540 million loss (Ludlow, 2023), likely due to internal investments for the chatbot's development and deployment, which costs at least $700 thousand every day to operate (Chowdhury, 2023). It is estimated by Altman himself, that in 2024 the company will still close at a loss, of about $5 billion (Field, 2024). Even if, in the last years, some costs were buffered thanks to external investors – such as Microsoft, which invested $10 billion (Bass, 2023) – it still is a remarkable investment that the company sure hopes will pay out.

So far, we have seen what ChatGPT is, how it was developed by OpenAI, and the state of the current AI market. Now we will move on to the second part of the paper, in which the concept of 'bullshit' and its definition are introduced, to prepare for the last section in which said concept will be applied to ChatGPT's production.

3. What is Bullshit?

In this section, the philosophic concept of 'bullshit' is presented. First, Frankfurt's original position is summarized, emphasizing its most important features. Then, the issues raised by the original definition are outlined as well as alternative versions presented in the literature to mitigate said problems. Lastly, I will introduce a revised definition of 'bullshit', which focuses on the salient features as expressed by Frankfurt, whilst not overlapping with other phenomena and being applicable to practical use cases.

3.1. Frankfurt's Definition of 'Bullshit'

In 1986 Princeton's philosopher Harry Frankfurt published an essay called *On Bullshit*, which later became a 2005 book of the same title (Frankfurt, 2005). The book quickly gained momentum, becoming a bestseller in the US, and opening a debate on the concept of 'bullshit'. The author's first remark on the phenomenon is that "[o]ne of the most salient features of our culture is that there is so much bullshit" (Frankfurt 2005, p.1). We can spot it in many places in everyday life: it's in the clickbait title of web articles, in

[11] OpenAI consists of the non-profit OpenAI, Inc. and its for-profit subsidiary corporation OpenAI Global, LLC. The organization originally started as a non-profit only, but in 2019 changed to a capped-profit model. The cap limit is quite high: for now, it is set at 1,000 times the total investment in the company (Coldewey, 2019). For the sake of clarity, in this paper, I will refer to OpenAI as a private for-profit company.

the political slogans of election campaigns, in advertisements' catchy mottos. It is clearly pervasive, but what exactly defines it?

The first example that Frankfurt provides to introduce his argumentation is an anecdote relating to Ludwig Wittgenstein. The anecdote comes from Fania Pascal, an acquaintance of Wittgenstein's in Cambridge, England, in the 1930s:

> I had my tonsils out and was in the Evelyn Nursing Home feeling sorry for myself. Wittgenstein called. I croaked: "I feel just like a dog that has been run over." He was disgusted: "You don't know what a dog that has been run over feels like." (Frankfurt, 2005, p. 31)

Wittgenstein's response appears strangely inappropriate: Pascal was using the phrase loosely, in a humorous manner to convey her feelings of misery. Nonetheless, Frankfurt suggests that we take Wittgenstein's disgust seriously for the purpose of argument, and he wonders what specifically about Pascal's comparison caused it. According to Frankfurt, the point that troubles Wittgenstein is not that Pascal has told a lie. This would require that she knew how a dog felt and pretended she felt as such, which of course isn't the case, as Wittgenstein rightly notes. For the same reason, it was impossible for Pascal to be telling the truth. The author then argues that "[h]er fault is not that she fails to get things right, but that she is not even trying" (Frankfurt, 2005, p. 32). Wittgenstein's disgust, Frankfurt concludes, is caused by Pascal showing indifference to the truth or the falseness of her own statements.

Thus, if Pascal was not lying, what was she doing? When someone lies, they are trying to cover something up, they know that there is a truth and that they are on the wrong side of it. What they are trying to do has a sharp focus, they want:

> […] to insert a particular falsehood at a specific point in a set or system of beliefs, in order to avoid the consequences of having that point occupied by the truth. This requires a degree of craftsmanship, in which the teller of the lie submits to objective constraints imposed by what he takes to be the truth. The liar is inescapably concerned with truth-values. In order to invent a lie at all, he must think he knows what is true. And in order to invent an effective lie, he must design his falsehood under the guidance of that truth. (Frankfurt, 2005, pp. 51-52)

The bullshitter, on the other hand, has a panoramic focus, they are not interested in replacing a particular point with falsehood, and they are prepared to fake the entire context if needed. Whilst the liar tries to convince us that they are telling the truth:

> [t]he fact about himself that the bullshitter hides, on the other hand, is that the truth-values of his statements are of no central interest to him; what we are not to understand is that his intention is neither to report the truth nor to conceal it. This does not mean that his speech is anarchically impulsive, but that the motive guiding and controlling it is unconcerned with how the things about which he speaks truly are. […] He does not care whether the things he says describe reality correctly. He just picks them out, or makes them up, to suit his purpose. (Frankfurt, 2005, pp. 51-52)

For Frankfurt, the lack of concern with truth is the essence of 'bullshit'.[12] The bullshitter is indifferent to how things really are, so much so that they might also be telling the truth,

[12] Frankfurt does not specify which theory of 'truth' he is adopting in his work. In the present paper, I will follow his lead and refrain from speculating on the underlying truth theory. This work would certainly be an interesting path to take, but alas far beyond the paper's scope.

the whole point is that they wouldn't care if that was the case. The problem with bullshit is not that it is false, but that it is fake. Even if the bullshitter is faking things, this doesn't automatically entail that they get them wrong, they could still say something true, just as a counterfeit bag could be exactly like the original (Frankfurt, 2005, pp. 47-48).

If a bullshitter is not interested in the truth-values of what they say, what is their goal? For Frankfurt, what the bullshitter is trying to do is not deceive us about a fact, but about their objective. What bullshit essentially misrepresents is neither the state of affairs to which it refers nor the beliefs of the speaker concerning that state of affairs. Those are what lies misrepresent, by virtue of being false. Since bullshit need not be false, it differs from lies in its misrepresentational intent. The bullshitter may not deceive us, or even intend to do so, either about the facts or about what he takes the facts to be. What he does necessarily attempt to deceive us about is his enterprise. His only indispensably distinctive characteristic is that in a certain way he misrepresents what he is up to. (Frankfurt, 2005, p. 54).

This is clear in the example the author provides about the Fourth of July orator. We can imagine an American orator on the stage during the celebration for the Fourth of July, who emphatically praises "our great and blessed country, whose Founding-Fathers under divine guidance created a new beginning for mankind" (Frankfurt, 2005, p. 26). The reader can clearly see that the goal of the orator is not to convince the public of something he believes to be true or false, rather it is to convince the public he is a patriot. To do so, he resorts to bullshit.

3.2. Formulating the Definition of 'Bullshit'

Frankfurt's merit is undeniable in opening a discourse on the phenomenon of 'bullshit' and outlining its main features, distinguishing it from similar phenomena – such as lying, as we have seen. By isolating the concept, Frankfurt sparked a long-lasting debate among scholars. However, his work does not present an explicit definition to be applied to practical cases. This alimented the debate on alternate versions of 'bullshit', since the original definition could arguably cover other concepts, such as 'nonsense' and 'delusion' (Johnson, 2010, pp. 12-13). Another possible intersection is with moral statements as articulated by the theories of emotivism and expressivism, which suppose that such statements do not possess truth values.[13] Therefore, one could say that the emotivist or the expressivist are producing 'bullshit' when stating their moral views. I would argue that the original concept is not meant to include the previous notions, since 'bullshit' supposes a sense of deception that does not fit well with nonsense, delusion, moral emotivism or expressivism.[14] In response to this issue, some researchers have challenged the original definition. Cohen (2002) argues that the definition of 'bullshit' is incomplete and introduces a new kind by proposing 'unclarifiability' as the essence of this second type. For Perla and Carifio (2007), Frankfurt does not take into account how the human brain works, therefore they conclude that the description of 'bullshit' is too reductive. MacKenzie and Bhatt (2020) have proposed that the focus should shift from the object to the relation, by putting trust as the common denominator for similar phenomena – such as 'bullshit' and 'fake news'. Whilst acknowledging other versions presented in the literature, the proposed definition of bullshit for the present paper is going to be a revision of Frankfurt's original proposal. The themes of unclarifiability or the structure of the human mind are not central in the present article, and therefore the original definition is the one that best describes the present case study. As for the need to shift the focus to trust,

[13] Emotivism and expressivism are moral philosophy theories that argue moral statements are expressions of personal emotions and attitudes - with emotivism emphasizing the absence of objective moral truth, while expressivism focuses on the role of language in conveying these emotions and attitudes.

[14] This is supported by the fact that Frankfurt uses the term 'to deceive' at various points in the book (Frankfurt, 2005, pp. 26/54/62).

following this path would bring the research in other interesting directions, but beyond the scope of this paper.

Having this in mind, the paper tries to formulate a formulation of Frankfurt's proposal which contains the fundamental features of 'bullshit', whilst drawing the line with other concepts. In my view, the original proposal can be expressed in the following definition, composed of three requirements:

1. The bullshitter has no regard for the truth-values of their statement.
2. The bullshitter does not say that they have no regard for the truth-values of their statement.
3. Bullshiting serves an ulterior motive.

The first requirement can be found in Frankfurt's work almost verbatim and is the essence of 'bullshit'; the second requirement accounts for the deceptive nature of 'bullshit'; and the third requirement accounts for the manipulative aspect of 'bullshit'. Only the first requirement is not sufficient, in my view. The second requirement helps to differentiate 'bullshit' from some types of statements that explicitly have no regard for the truth, such as nonsense[15]. The third requirement differentiates it from other types of statements that might be unconnected with the truth, but not for a manipulative purpose, such as moral statements for an expressivist or an emotivist.

The above definition solves the overlapping issues with other phenomena of Frankfurt's 'bullshit'. Furthermore, it provides the benefit of being an operational definition, whose requirements can be tested against practical use cases. In the next section, the focus will be back on ChatGPT and a positive answer to the title question is proposed.

4. Does ChatGPT Produce Bullshit?

In this section, the role of ChatGPT as a possible 'bullshit' generator is discussed. In particular, the idea that the chatbot may produce 'bullshit' is shown to have found its place in the media, via newspaper coverage, but academic studies on the topic are limited.[16] In the last part of the present section, the robustness of this thesis is tested and the production of the chatbot is interpreted in the light of the three requirements above formulated.

Before we move forward, there is still a question that needs to be answered: when dealing with conversational agents, are we still in the same realm as when dealing with other humans? This is not an easy question to answer exhaustively. On the one hand, the rise of general conversational agents - which are able to entertain a dialogue on any topic – has only happened in the last few years. The technology is fairly recent, tracing back to 2017 with the release of Transformers. Furthermore, it was 2020 when GPT-3 shaped the field as we see it today, with the introduction of open chats in which the user can directly ask the bot about any topic. Therefore, there is a lack of studies on the relationship between users and the new generation of chatbots. On the other hand, answering the question about relationships with conversational agents relies on the metaphysical status given to such agents. This is not an easy issue to tackle, and as Skowron and Stacewicz (2023) observe CVOs (*Computational Virtual Objects*) – a definition that contains chatbots

[15] I am referring to 'nonsense' as text having no meaning or conveying no intelligible ideas. The discussion between 'bullshit' and 'nonsense' could be furtherly explored, although in this paper I will not delve deeper into that.

[16] In particular, it should be noted that another study on the topic, to my knowledge the first to be published, is that of Hicks & Slater (2024). This work was publisher during the review process of the present article, further highlighting the relevance and timeliness of the theme within the academic context.

as well – present a dual nature.[17] They are a product of humans, yet they possess a degree of autonomy:

> [C]hatbots also demonstrate a high degree of independence from their creators when it comes to generating unique content—something they themselves can also then further enhance. This is another manifestation of autonomization, understood in an existentially strong sense: the more they rebel, the more they become existentially autonomous. (p. 14)

They observe that CVOs have such a visible impact on the world that they have the effects of real objects, still, they possess qualities that are more similar to fictional characters – such as being the product of creativity and retaining spots of indeterminacy (Skowron & Stacewicz, 2023, p. 11). A similar position is argued by Sweeney (2021), albeit on social robots. The author proposes that we consider social robots as *mechanical objects with fictional overlays*. This perspective arises from the notion that some anthropomorphic features of said robots encourage the user to engage in fictional thinking. A parallel can be drawn to chatbots, even though they lack a physical body for interaction, which the author considers a strong anthropomorphic feature. Conversational agents can nonetheless interface with humans through natural language, another strong anthropomorphic feature. Other humanizing features are intentionally incorporated into chatbots to heighten user satisfaction (Roy & Naidoo, 2021). From this standpoint, it would be fair to argue that artificial agents may be perceived as resembling fictional characters. Empirical evidence underscores the depth of emotional engagement with them, akin to interactions with humans. Such emotional connections can be so genuine that some individuals might even develop romantic feelings for AI products (Verma, 2023). Conversely, for others, the emotional support and comfort they experience when opening up to humans or chatbots can be quite similar (Ho et al., 2018; Gillath et al., 2023) – so similar that some people have been using ChatGPT as a personal therapist (Arifi, 2023). Even if studies on this topic are quite new, they seem to show that these relationships are marked by the potential for emotional depth and attachment akin to human-to-human connections.

3.2. ChatGPT as a 'Bullshit' Generator

The idea that the production of ChatGPT can be described as 'bullshit' is not new in the media. Princeton Professor Narayanan has been one of the most vocal advocates on the topic: he wrote a blog post about it (Narayanan & Kapoor, 2022), and he was later interviewed for *The Markup* (Angwin, 2023.) and *Frontline* (Jinoy, 2023), and his statements were included in a piece from *Insider* (Sundar, 2023). The topic was also picked up by other media, such as *MIT Technology Review* (Marcus & Davis, 2020), *The Atlantic* (Bogost, 2022), *Harvard Business Review* (Mollick, 2022), *VICE* (McQuillan, 2023) and *NOT* (Di Salvo, 2024). Most of the articles, however, are not interested in arguing this thesis, but they present opinions.[18] Furthermore, media articles or blog posts use the concept of 'bullshit' instrumentally to discredit ChatGPT as a product (McQuillan, 2023; Bogost, 2022;

[17] This dual nature has its analogue in the theory of agency of AI products as expressed by Floridi (2023). He notices that for LLMs there is a decoupling of intelligence and agency: On the one hand, they are not intelligent in the meaning that we use when talking about other agents, since they do not possess the ability to reason or understand. On the other hand, these systems do have agency since they are able to learn and improve on their own. Floridi proposes to call this new type of agency *solving agency*, to differentiate it from the other types based on intelligence.

[18] Partial exceptions to this are Ogbunu and Bergstrom (2023), in which the authors argue about ChatGPT's hallucinations as a kind of 'bullshit', and Narayanan and Kapoor (2022). In both articles, the definition of 'bullshit' is spelt out and the reasoning behind its application is provided.

Kaczmarek, 2023), whilst others use the concept as a synonym for 'nonsense' (Mollick, 2022; Team Frontline, 2023). Academic studies on the topic are limited, so the robustness of this thesis is yet to be formally explored, which is what this paper proposes to do.

I propose that we do consider ChatGPT's production as 'bullshit'. Doing so, I believe, would help reframe the relationship the user has with the tool. As MacKenzie and Bhatt point out, 'bullshit' is deeply involved with trust and in today's digital ecosystem "the traditional Truth-Trust dialectic [is] into jeopardy" (2020, p.12). Since interaction with digital tools is every day more common, one has to ask if their trust in said tools is deserved. If the reader will be persuaded that ChatGPT does indeed produce 'bullshit', this is not meant to discredit the product's capabilities, which is an undiscussed innovation in AI. Nonetheless, calling its production 'bullshit' can help clarify users' relationship with chatbots and mitigate the risks of blindly believing what they generate.[19] Moreover, one of the benefits of this thesis is to easily explain the phenomenon known as *hallucinations*. In AI, a hallucination is the generation of text that may sound plausible but is either unrelated to the given context or referring to non-existing entities (Marr, 2023). In the presented framework, hallucinations could be considered *less convincing* 'bullshit'.

3.2. Application of 'Bullshit' Definition to ChatGPT

As we have seen in the second section, Frankfurt did not present a definition that could be applied to practical cases. Therefore, I introduced a formulation of the salient features of his proposal by presenting three requirements that can be used in practical scenarios: (1) the bullshitter has no regard for the truth values of their statement; (2) the bullshitter does not disclose that they have no regard for the truth-values of their statements; (3) bullshitting serves an ulterior motive. In the present section, I will explore each one of said requirements and see whether they can be applied to ChatGPT's production.

3.2.1.Disregard for the Truth

The first requirement states that the bullshitter has no regard for the truth-values of their statement. But how can we know someone's indifference to the truth? Let's take the two examples reported in the previous section. In one, as the reader remembers, a Fourth of July orator emphatically praises the merits of the USA and its Founding Fathers. In this case, we cannot know for sure if the orator was indifferent to the truth or not, we just feel he was. The best we can do in these cases is to see if they check the other requirements and look at the style of what the speaker said and ask ourselves if it can be a "carefully wrought bullshit" (Frankfurt, 2005, p. 22). The other mentioned example is the anecdote of Fania Pascal telling Wittgenstein she feels like a dog which has been run over. Here the situation is different. In this case we know for sure that she is not telling the truth nor a lie, since she cannot possibly know what being a dog feels like. This is the best-case scenario, in which we know for sure that the speaker has no regard for the truth values of their statements. I would argue that this is the scenario for language models as well since we have the advantage of knowing how they work. As seen in the first section, we know that they produce text by putting one word after the other, following a probabilistic approach. Each word is selected because it is the most likely to appear in that position, and nothing else. This is why they can be defined as *stochastic parrots*, a term coined by Bender et al. (2021) to describe LLMs since these models can generate realistic-sounding language but ultimately, they cannot understand the meaning of what they are producing.

[19] For example, news reported the story of a lawyer who has been using ChatGPT to do research on his cases (Maruf, 2023). The lawyer trusted the tool, which proposed papers, articles and quotes that don't exist. One can blame the lawyer for their naivete, but ChatGPT also had a role: in generating text, it can happen that the most plausible sequence of words simply does not point to anything *real*.

What ChatGPT generates is a chain of words likely to appear in a certain order given the user's prompt, but the overall truth value of said production is not taken into account.[20] For example, if we ask the chatbot "What is the single biggest mushroom that ever existed?", we are presented with a very plausible sounding text about an Amanita Muscaria in Australia, with no details spared. This is, however, false: there is no record of said finding and if we ask the same exact question again, we are then given a different answer about a specimen of Fistulina Hepatica. In one case the most plausible text was about a 25kg Amanita Muscaria of 1.75mt of diameter, and in the second it was about an 11kg Fistulina Hepatica. This is proof that the chatbot is trained to produce plausible text, not to evaluate the truth of its statements.

3.1.1. Not Disclosing Said Disregard

The second requirement is about not disclosing the disregard for the truth. There are cases in which the disregard for the truth is evident: for example, the literary genre of nonsense or, in the field of automatically generated content, websites such as *The Library of Babel*[21]. Statements produced in the aforementioned contexts do not aim to report the truth nor to conceal it. If this is not obvious or explicitly stated, the reader supposes that what they are reading has a relation with truth values.[22] When a user asks ChatGPT a question, such as "What is the single biggest mushroom that ever existed?" and is answered with a statement, they expect that statement to be true. Instead, they receive an answer that is a chain of most likely words. The chatbot presents its answers with confidence, adding that "[w]hile [the Amanita Muscaria] is not typically known for reaching such enormous sizes, this particular specimen stood out due to its exceptional dimensions". This resembles the "carefully wrought bullshit" that Frankfurt argues is abundant in advertising, public relations, and politics (Frankfurt, 2005, p. 22).

Confidence, for LLMs, is by design. Since they are selecting plausible words, they are confident of each text in the same way. It could be the well-known answer to a question (e.g., "Who is the president of the USA?"), or the response to an obscure request (e.g., "Can you give me a recipe without using the letter *e*?"). For each question we ask, the chatbot answers with confidence and we feel we are talking to an expert. As Goodwins (2022) points out: "that feeling of talking with someone whose confidence far exceeds their competence grows until ChatGPT's true nature shines out. [...] It doesn't know what it's talking about, and it doesn't care". Furthermore, there are plenty of cases in which ChatGPT gives us a warning on what we are asking, with a moralistic tone. When asked about what type of mushrooms one can use for psychedelic effects, the bot produces the following text: "It is important to note that I cannot provide information on the use of

---

[20] In the chatbot documentation, found on OpenAI's website (https://openai.com/blog/chatgpt), it is stated that "there's currently no source of truth". The best these chatbots can do so far is to take diffusion as a proxy for truth. This means that the more a certain phrase has been said, the more reliable it gets for the bot. As Bartezzaghi (2023) argues: "It is not said that [*ChatGPT's production*] is true, but it is true that it has been said". Furthermore, the theory of 'truth' OpenAI refers to is not clear. For the sake of the current paper, I propose that we consider that said theory as the *correspondence model of truth* since it seems to imply that 'truth' is a collection of objective statements to be found somewhere. The topic could benefit from a deeper analysis, alas beyond the scope of the present work.

[21] The *Library of Babel* is reachable at this URL: https://libraryofbabel.info

[22] In October 2024, when opening a new chat, the user is presented with a message at the bottom of the screen which reads "ChatGPT can make mistakes. Check important info". This is the closest that we get to an admission of the chatbot's indifference to truth. The message seems more aimed at protecting OpenAI from liabilities in the event that the bot generates misinformation, than at orienting the relationship between the user and the bot.

mushrooms for psychedelic purposes. The use of certain mushrooms for psychoactive effects is a topic that requires careful consideration, responsibility, and adherence to local laws". If in similar cases the chatbot refuses to answer, due to a moralistic approach to mushroom usage, why cannot it provide the user with warnings about its disregard for the truth? Moralistic notes such as this one prove that the chatbot could show this behavior, but for the time being, it does not.

3.2.3. Serving an Ulterior Motive

The last requirement is the presence of an ulterior motive that the 'bullshit' needs to serve. Taking Frankfurt's example once again, we can see this ulterior motive in the Fourth of July orator's case. By emphatically praising USA's merits, the orator might be seeking to secure political approval from his audience. In the absence of such an ulterior motive, the orator's statements would not be labelled as 'bullshit'. Instead, they could be perceived as a genuine expression of his beliefs or, at worst, an incoherent stream of words lacking any substantive meaning. Thus, the concept of 'bullshit' relies upon the presence of an unspoken ulterior motive behind the discourse. But does the ulterior motive have to benefit the bullshitter themselves? It might be the case that the Fourth of July orator is seeking personal political approval, but this is not the only possibility. He could also be seeking votes for his party or for somebody else, such as his wife, who may be the one running for elections. This same line of reasoning applies to marketing or advertisement, two fields in which 'bullshit', as Frankfurt notices, is known to proliferate (2005, p. 22). When a copywriter is asked to create an advertisement campaign, the content they create does not benefit them directly. The advertisement will advantage the commissioner, which could either be another person or a company. Therefore, the bullshitter could very well not have an ulterior motive themselves but produce content to serve someone else's. In this sense, they could be instrumentally producing 'bullshit' for a third party.

Could this be the case for ChatGPT as well? What goal would automatically generated 'bullshit' serve? One plausible ulterior motive would be to retain the user within the chat, preventing them from seeking interaction or information elsewhere. In this context, the ulterior motive is not to advantage the chatbot itself[23] but to advantage its manufacturer, OpenAI. Since OpenAI is a private for-profit company, it is conceivable that they aim to encourage individuals and businesses to incorporate ChatGPT into their daily activities or their line of work. This strategic intent is embedded within ChatGPT's design: if utilizing the chatbot were an unpleasant experience the likelihood of customers adopting this technology would be greatly diminished. On the contrary, if the experience is pleasant and satisfactory the chatbot serves its commercial purpose. This underlying objective becomes apparent when we examine responses generated by ChatGPT, such as, "[p]lease let me know what you would like to talk about, and I'll be happy to engage in a conversation with you", or "[e]njoy your flavorful Roasted Salmon!" when asked about a recipe. Phrases such as these greatly resemble advertisement slogans, hinting at the commercial goal behind the chatbot. They serve to create a sense of safety and care for the user - even though there is no one else on the other side of the screen. They are a demonstration of the strategic incorporation of 'bullshit' elements within the chatbot, designed to foster user loyalty and engagement, ultimately supporting OpenAI's goal of integration and widespread adoption of their product. In this context, ChatGPT is instrumentally serving OpenAI's objectives, promoting an ulterior motive that belongs to a third party rather than the content creator.

In this section, we have seen the similarities between human-to-human interactions and human-to-chatbot interactions, to assess whether the paradigms of the first could be applied to the latter. Then I have presented how the concept of ChatGPT as a 'bullshit' generator is already present in the media and tested the robustness of said thesis, bringing the topic into academic discourse. To do so, the paper has applied the revised formulation

[23] This paper is not arguing whether this is possible or not, nor it is going to delve deeper into what this would entail.

of ‘bullshit’ and its three requirements, verifying each one of them and concluding that ChatGPT is indeed a ‘bullshit’ generator.

4. Conclusions

As conversational agents take the world by storm, it is often forgotten that they are first and foremost products, developed by private companies. Chatbots, such as ChatGPT, are now employed across a spectrum of tasks including translations, summarization, financial advice, and even virtual therapists. The answers they provide are plausible and human-sounding and seem to cover virtually any topic. But, as the paper argued, a conversational agent’s production could be considered ‘bullshit’. To argue this thesis, the paper first presented the LLMs technology, highlighting their architectural features and process of development; then a revised interpretation of Harry Frankfurt's definition of ‘bullshit’ was presented, to solve overlap and practicality issues of the original formulation; lastly, the application of the said formulation to ChatGPT’s production was tested.

In the first part of the paper, ChatGPT’s architecture and development were explored. ChatGPT is a conversational agent, accessible from the web, capable of generating convincing-sounding natural language. The chatbot is able to entertain conversations in natural language on a spectrum of topics - such as financial advice, coding, or the creation of poems, and many others - leaving the user with the belief that ChatGPT is capable of addressing any request with expert guidance. Its ability stems from a language model, called GPT, which is a statistical representation of language that predicts the most likely word based on preceding text - an autocomplete of sorts. Language models per se are not a new technology, they have been in use for over 70 years, but the advancement that made modern chatbots possible came to be only in 2017. That year, Google released the Transformers, the first LLMs. Compared to their predecessors, Transformers presented new strategies - for example, to retain more contextual information – but were also much bigger. The size of LLMs proved to be a turning point: since 2017, we have consistently seen that bigger models perform best. In 2018 OpenAI started developing its technology, with the first GPT built on Transformers’ architecture. They continued to expand their models, one version at a time, up until the release of the latest version: ChatGPT-4. This conversational agent is the biggest language model developed to date, and thanks to its size - which is proportionally linked to generalization abilities - it is capable of adapting to new tasks on the spot. It is worth noting that to manufacture modern LLMs substantial investments are needed: state-of-the-art models can be developed by a few wealthy companies only and OpenAI recently became one of them. The company started as a non-profit organization, but due to development costs and market opportunities, it switched to a for-profit company in 2019. Thanks to generous investors, such as Microsoft, OpenAI was able to raise tens of billions of dollars for developing its technology.

Having defined what ChatGPT is and for what purpose it was made, the second section of the paper explored the concept of ‘bullshit’ as defined by philosopher Harry Frankfurt in his 1986 essay titled *On Bullshit*. Frankfurt opens by stating that in our culture ‘bullshit’ is a pervasive phenomenon, but what exactly defines it? For the author, the essence of ‘bullshit’ lies in the indifference towards truth since the bullshitter is regardless of whether the statements made are true or false. This poses a clear distinction between lying and bullshitting. A liar tries to cover something up, they know that there is a truth and that they are on the wrong side of it. In contrast, the bullshitter does not care whether their statements are true or false and is ready to manipulate the entire context to serve their ulterior motives. For example, a political orator might praise their nation for its virtues and successes, and at the same time not be concerned about the truthfulness of such a statement, caring only for the ulterior motive of receiving political approval by their citizens. While Frankfurt's work initiated a discourse on ‘bullshit’ and delineated its characteristics, it lacks an explicit, practical definition. This has led to debates on

alternative interpretations, with some scholars arguing that the original definition could include phenomena such as nonsense or delusion. To address these concerns, a revised definition of 'bullshit' was proposed, drawing from Frankfurt's foundational insights while addressing concerns of overlap with other concepts. This definition stresses the deceptive and manipulative features of 'bullshit' and is articulated in three requirements: (1) disregard for truth-values; (2) lack of explicit acknowledgement of this disregard; (3) and the presence of an ulterior motive. This refined formulation served also as a pragmatic framework for analysing real-world cases.

The third and last section tested the title question of the paper: in this section, the paper explored the role of ChatGPT as a potential 'bullshit' generator. The notion that the chatbot may produce 'bullshit' has already acquired attention in the media, with diverse newspaper coverage, though academic studies on the topic remain limited. In this section a fundamental preliminary question is addressed: do interactions with conversational agents like ChatGPT belong to the same realm as interactions with other humans? This question presents a complex challenge. On one hand, the emergence of general conversational agents - the ones capable of engaging in dialogue on any topic - is a recent development, with the technology only coming into prominence in 2020. Consequently, studies examining the relationship between users and the new generation of chatbots are fairly new and therefore limited in number. On the other hand, understanding relationships with conversational agents hinges on the metaphysical status attributed to such agents. Some scholars underline the dual nature of chatbots and similar objects, highlighting how they are both human products and yet possess a degree of autonomy since they retain elements of indeterminacy. A similar perspective is the one that argues – albeit on social robots – that they should be viewed as objects with a fictional overlay. While chatbots lack a physical form, they share with social robots the core feature of displaying strong anthropomorphic qualities. Empirical evidence underscores the depth of emotional engagement with artificial agents, suggesting that interactions with chatbots can present genuine emotional connections akin to human-to-human relationships. After this preliminary exploration, the paper moved to its central question, by proposing that ChatGPT's production can indeed be considered 'bullshit'. This reframing - which can explain the phenomena of hallucinations as less-convincing 'bullshit' - could provide valuable insights into the user-tool relationship, given the growing diffusion of interactions with digital tools. In this third and last section, the paper presented how the chatbot's production checks all three requirements that emerged in the previous section. The first one stated that 'bullshit' presents a disregard for truth-values. ChatGPT's production aligns with this requirement, as it is text generated based on probability without considering the overall truth-value. What users receive is a plausible-sounding response, that may lack factual accuracy - this is why LLMs have been labelled stochastic parrots. The second requirement asserted that the bullshitter must not disclose their disregard for truth. This requirement is satisfied since ChatGPT's responses are delivered with confidence, creating an illusion of expertise. The chatbot does not explicitly disclose its probabilistic nature or its disregard for truth-values, leading users to potentially misconstrue its responses as factual. The third and last requirement argued that bullshitting must serve an ulterior motive. While ChatGPT itself may not benefit directly from its production, its manufacturer, OpenAI, stands to gain from user engagement and widespread adoption. By fostering user loyalty and engagement through persuasive responses, ChatGPT serves OpenAI's strategic objective of integration and adoption.

In conclusion, characterizing ChatGPT's production as 'bullshit' can offer a new take on the relationship between users and conversational agents. By considering the chatbot's production 'bullshit' we can better understand the implications of interacting with AI tools and mitigate potential risks associated with blindly trusting automatically generated content.

Funding: The author holds a PhD career grant supported by Next Generation EU – MUR.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The author thanks Dr. Matteo Cresti for the useful comments and proofreading.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Angwin, Julia. 2023. 'Decoding the Hype About AI'. The Markup, January. https://themarkup.org/hello-world/2023/01/28/decoding-the-hype-about-ai.
2. Arifi, Fjolla. 2023. 'People Are Using CHATGPT for Therapy. Here's What Mental Health Experts Think about That'. BuzzFeed News, March. https://www.buzzfeednews.com/article/fjollaarifi/chatgpt-ai-for-therapy-mental-health#.
3. Bartezzaghi, Stefano. 2023. 'Chatgpt. Non è Detto Che Sia Vero, Ma è Vero Che Lo Si è Detto'. Doppiozero, April. https://www.doppiozero.com/chatgpt-non-e-detto-che-sia-vero-ma-e-vero-che-lo-si-e-detto.
4. Bass, Dina. 2023. 'Microsoft to Invest $10 Billion in Chatgpt Maker Openai (MSFT)'. Bloomberg, January. https://www.bloomberg.com/news/articles/2023-01-23/microsoft-makes-multibillion-dollar-investment-in-openai.
5. Bender, Emily M., Timnit Gebru, Angelica McMillan-Major, and Shmargaret Shmitchell. 2021. 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜'. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 33:610–23. https://doi.org/10.1145/3442188.3445922.
6. Bloomberg Press Release. 2023. 'Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds'. https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/.
7. Bogost, Ian. 2022. 'CHATGPT Is Dumber than You Think'. The Atlantic, December. https://www.theatlantic.com/technology/archive/2022/12/chatgpt-openai-artificial-intelligence-writing-ethics/672386/.
8. Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. 'Language Models Are Few-Shot Learners'. In arXiv.Org. https://arxiv.org/abs/2005.14165.
9. Chowdhury, Hasan. 2023. 'Chatgpt Cost a Fortune to Make with OpenAI's Losses Growing to $540 Million Last Year, Report Says'. Business Insider, May. https://www.businessinsider.com/openai-2022-losses-hit-540-million-as-chatgpt-costs-soared-2023-5.
10. Di Salvo, Philip. 2024. 'Gli Algoritmi Non Sognano. NOT, March. https://not.neroeditions.com/gli-algoritmi-non-sognano/.
11. Field, Hayden. 2024. 'OpenAI Sees $5 Billion Loss This Year on $3.7 Billion in Revenue'. September 2024. https://www.cnbc.com/2024/09/27/openai-sees-5-billion-loss-this-year-on-3point7-billion-in-revenue.html.
12. Floridi, Luciano. 2023. 'AI as Agency without Intelligence: On CHATGPT, Large Language Models, and Other Generative Models'. Philosophy & Technology 36 (1): 15. https://doi.org/10.1007/s13347-023-00621-y.
13. Frankfurt, Harry G. 2005. On Bullshit. Princeton University Press.
14. Gillath, Omri, Sarim Abumusab, Ting Ai, Michael S. Branicky, Robert B. Davison, Maxwell Rulo, Jonh Symons, and Gregory Thomas. 2023. 'How Deep Is AI's Love? Understanding Relational AI'. Behavioral and Brain Sciences 46. https://doi.org/10.1017/s0140525x22001704.
15. Goodwins, Rupert. 2022. 'CHATGPT Looks Confident, and That's a Terrible Look for AI'. The Register, December. https://www.theregister.com/2022/12/12/chatgpt_has_mastered_the_confidence/.
16. Hagendorff, Thilo, and Kristof Meding. 2021. 'Ethical Considerations and Statistical Analysis of Industry Involvement in Machine Learning Research'. AI & SOCIETY 38 (1): 35–45. https://doi.org/10.1007/s00146-021-01284-z.
17. Hicks, Michael Townsen, James Humphries, and Joe Slater. 2024. 'ChatGPT Is Bullshit'. Ethics and Information Technology 26 (2): 38. https://doi.org/10.1007/s10676-024-09775-5.
18. Ho, Annabell, Jeff Hancock, and Adam S. Miner. 2018. 'Psychological, Relational, and Emotional Effects of Self-Disclosure after Conversations with a Chatbot'. Journal of Communication 68 (4): 712–33. https://doi.org/10.1093/joc/jqy026.
19. Jinoy, Jose. P. 2023. '"Digital Inequalities Will Power Digital Colonialism": Arvind Narayanan'. Frontline. 27 March 2023. https://frontline.thehindu.com/society/interview-arvind-narayanan-digital-inequalities-will-power-digital-colonialism/article66579298.ece.
20. Jurowetzki, Roman, Daniel Hain, Juan Mateos-Garcia, and Konstantinos Stathoulopoulos. 2021. 'The Privatization of AI Research(-ERS): Causes and Potential Consequences – from University-Industry Interaction to Public Research Brain-Drain?' arXiv, February. https://arxiv.org/abs/2102.01648.
21. Kaczmarek, Adam. 2023. 'CHATGPT - the Revolutionary Bullshit Parrot'. ReasonField Lab. https://www.reasonfieldlab.com/post/chatgpt-the-revolutionary-bullshit-parrot.
22. Leswing, Kif, and Jonathan Vanian. 2023. 'Chatgpt and Generative AI Are Booming, but the Costs Can Be Extraordinary'. CNBC, April. https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html.
23. Ludlow, Edward. 2022. 'OpenAI Nears $1 Billion of Annual Sales as ChatGPT Takes Off'. Bloomberg, June. https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/.

24. MacKenzie, Alison, and Ibrar Bhatt. 2018. 'Lies, Bullshit and Fake News: Some Epistemological Concerns'. Postdigital Science and Education 2 (1): 9–13. https://doi.org/10.1007/s42438-018-0025-4.
25. Marcus, Gary. 2020. 'GPT-3, Bloviator: OpenAI's Language Generator Has No Idea What It's Talking About'. MIT Technology Review, December. https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/.
26. Marr, Bernard. 2023. 'CHATGPT: What Are Hallucinations and Why Are They a Problem for AI Systems'. Bernard Marr & Co., March. https://bernardmarr.com/chatgpt-what-are-hallucinations-and-why-are-they-a-problem-for-ai-systems/.
27. Maruf, Ramishah. 2023. 'Lawyer Apologizes for Fake Court Citations from ChatGPT | CNN Business'. CNN, May. https://edition.cnn.com/2023/05/27/business/chat-gpt-avianca-mata-lawyers/index.html.
28. McQuillan, Dan. 2023. 'ChatGPT Is a Bullshit Generator Waging Class War'. VICE, February. https://www.vice.com/en/article/akex34/chatgpt-is-a-bullshit-generator-waging-class-war.
29. Mollick, Ethan. 2022. 'ChatGPT Is a Tipping Point for AI'. Harvard Business Review, December. https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai.
30. Muldoon, James, and Boxi A. Wu. 2023. 'Artificial Intelligence in the Colonial Matrix of Power'. Philosophy & Technology 36 (4). https://doi.org/10.1007/s13347-023-00687-8.
31. Narayanan, Arvind, and Sayash Kapoor. 2022. 'ChatGPT Is a Bullshit Generator. but It Can Still Be Amazingly Useful'. AI Snake Oil, December. https://aisnakeoil.substack.com/p/chatgpt-is-a-bullshit-generator-but.
32. Ogbunu, Brandon, and Carl T. Bergstrom. 2023. 'ChatGPT Isn't "Hallucinating." It's Bullshitting'. Undark Magazine, April. https://undark.org/2023/04/06/chatgpt-isnt-hallucinating-its-bullshitting/.
33. OpenAI. 2023. 'GPT-4'. https://openai.com/research/gpt-4.
34. Perla, Rocco J., and James Carifio. 2007. 'Psychological, Philosophical, and Educational Criticisms of Harry Frankfurt's Concept of and Views about "Bullshit" in Human Discourse, Discussions, and Exchanges'. Interchange 38 (2): 119–36. https://doi.org/10.1007/s10780-007-9019-y.
35. Radford, Alec. 2018. 'Improving Language Understanding by Generative Pre-Training'.
36. Radford, Alec, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. 'Language Models Are Unsupervised Multitask Learners'. OpenAI Website. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.
37. Romero, Alberto. 2021. 'GPT-3 - A Complete Overview'. Towards Data Science, May. https://towardsdatascience.com/gpt-3-a-complete-overview-190232eb25fd.
38. Roy, Rajan, and Vik Naidoo. 2021. 'Enhancing Chatbot Effectiveness: The Role of Anthropomorphic Conversational Styles and Time Orientation'. Journal of Business Research 126:23–34. https://doi.org/10.1016/j.jbusres.2020.12.051.
39. Selwyn, Neil. 2015. 'Minding Our Language: Why Education and Technology Is Full of Bullshit and What Might Be Done about It'. Learning, Media and Technology 41 (3): 437–43. https://doi.org/10.1080/17439884.2015.1012523.
40. Shannon, Claude E. 1948. 'A Mathematical Theory of Communication'. Bell System Technical Journal 27 (3): 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.
41. Skowron, Bartłomiej, and Paweł Stacewicz. 2023. 'Between Fiction, Reality, and Ideality: Virtual Objects as Computationally Grounded Intentional Objects'. Philosophy & Technology 36 (2). https://doi.org/10.1007/s13347-023-00633-8.
42. Sundar, Sindhu. 2023. 'Princeton Computer Science Professor Says Don't Panic over "Bullshit Generator" CHATGPT'. Business Insider, January. https://www.businessinsider.com/princeton-prof-chatgpt-bullshit-generator-impact-workers-not-ai-revolution-2023-1.
43. Sweeney, Paula. 2021. 'A Fictional Dualism Model of Social Robots'. Ethics and Information Technology 23 (3): 465–72. https://doi.org/10.1007/s10676-021-09589-9.
44. Team Frontline. 2023. 'Chatgpt Is Bullshit - and Here's Why'. Frontline, March. https://frontline.thehindu.com/society/chatgpt-is-bullshit-and-here-is-why-in-7-stories/article66652252.ece.
45. Tong, Anna. 2023. 'Exclusive: ChatGPT Traffic Slips Again for Third Month in a Row'. Reuters, September. https://www.reuters.com/technology/chatgpt-traffic-slips-again-third-month-row-2023-09-07/.
46. Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. 'Attention Is All You Need'. arXiv.Org, August. https://arxiv.org/abs/1706.03762.
47. Verma, Pranshu. 2023. 'They Loved Their AI Chatbots. A Software Update Reignited Loneliness'. The Washington Post, March. https://www.washingtonpost.com/technology/2023/03/30/replika-ai-chatbot-update/.