Essays – *peer-reviewed* Encyclopaideia – Journal of Phenomenology and Education. Vol.27 n.66 (2023)
https://doi.org/10.6092/issn.1825-8670/16554 ISSN 1825-8670

Can AI Language Models Improve Human Sciences Research? A Phenomenological Analysis and Future Directions

Marika D'Oria*

Scientific Directorate, Fondazione Policlinico Universitario A. Gemelli IRCCS (Italy)

Published: August 30, 2023

I modelli linguistici di IA possono migliorare la ricerca nelle Scienze Umane? Analisi fenomenologica e direzioni future

The article explores the use of the "ChatGPT" artificial intelligence language model in the Human Sciences field. ChatGPT uses natural language processing techniques to imitate human language and engage in artificial conversations. While the platform has gained attention from the scientific community, opinions on its usage are divided. The article presents some conversations with ChatGPT to examine ethical, relational and linguistic issues related to human-computer interaction (HCI) and assess its potential for Human Sciences research. The interaction with the platform recalls the "uncanny valley" phenomenon still known to Social Robotics. While ChatGPT can be beneficial, it requires proper supervision and verification of its results. Furthermore, new research methods must be developed for qualitative, quantitative, and mixed methods.

L'articolo esplora l'uso del modello linguistico di intelligenza artificiale "ChatGPT" nel campo delle scienze umane. ChatGPT utilizza tecniche di elaborazione del linguaggio naturale per imitare il linguaggio umano e avviare conversazioni artificiali. Sebbene la piattaforma abbia ottenuto l'attenzione della comunità scientifica, le opinioni sul suo utilizzo sono discordanti. L'articolo presenta alcune conversazioni con ChatGPT per esaminare le questioni etiche, relazionali e linguistiche legate all'interazione uomo-computer (HCI) e valutare il suo potenziale per la ricerca nelle Scienze Umane. L'interazione con la piattaforma ricorda il fenomeno della "uncanny valley" già noto alla Robotica Sociale. Sebbene ChatGPT possa essere utile, esso richiede un'adeguata supervisione e verifica dei risultati. Inoltre, è necessario sviluppare nuovi metodi di ricerca qualitativi, quantitativi e misti.

Keywords: Human-computer interaction; Natural language processing; Artificial intelligence; Human sciences research; Methodology.

---

* ✉ marikadoria@gmail.com

Copyright © 2023 Marika D'Oria
The text in this work is licensed under the Creative Commons BY License.
https://creativecommons.org/licenses/by/4.0/

> "Well! I've often seen a cat without a grin," thought Alice;
> "but a grin without a cat! It's the most curious thing I ever saw in all my life!"
> Lewis Carroll, *Alice's Adventures in Wonderland*

1. Introduction

Human-computer interaction (HCI) is an interdisciplinary field that combines computer science, psychology, and other disciplines to study the interaction between humans and computers, and to create user-friendly interfaces that meet users' needs and expectations (Gurcan, Cagiltay, & Cagiltay, 2021). HCI assumes a participatory design, through which improvements in usability and technology acceptance are addressed with the user experience and interaction (Alavi & Lalanne, 2020). For example, in Human Sciences research, decisions about which statistical software tools are best suited to analyze research data are made by identifying the human factors that affect the tool's performance (data visualization, ease of use, efficiency, data management, etc.) (Sardareh, Brown, & Denny, 2021). Consequently, the usability, acceptance, and accessibility of the tools are calibrated according to the end-users' perception and involvement with the tool.

In a similar fashion, some artificial intelligence (AI) algorithms, known as *chatbots*, can recognize and emulate human language because their learning model is based on natural language processing (NLP) techniques, which are useful to analyze HCI since these tools emulate the patterns of natural language (Singh, Bhangare, Singh, Zope, & Saindane, 2023).

Such algorithms can handle artificial conversations because they are trained on NLP, can provide a preliminary form of coaching for psychological assistance (Fiske, Henningsen, & Buyx, 2019), and convey *artificial empathy* in terms of HCI. Therefore, studying HCI on conversational AI technologies can be significant for the Human Sciences in researching sensitive context where dialogue is critical, for example in conveying information to adolescents on sensitive health promotion topics (Crutzen, Peters, Portugal, Fisser, & Grolleman, 2011), in guiding students with Autism Spectrum Disorder through social, physical, and environmental cues on higher education to provide university counselling services (Bradford *et al.*, 2020), as well as providing vocational guidance to students (Zahour, Benlahmar, Eddaoui, Ouchra, & Hourrane, 2020), detecting signs of cognitive impairment in the elderly (de Arriba-Pérez, García-Méndez, González-Castaño, & Costa-Montenegro, 2022) or educating patients with eHealth for rehabilitation and monitoring (Infarinato *et al.*, 2020; Kyriazakos *et al.*, 2020).

In addition, the recent trend to use OpenAI's open platform ChatGPT has generated great curiosity around the world and growing interest among scientists. The developers introduce the platform as follows:

> We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests (OpenAI, 2023).

Its training process involves analyzing large amounts of text data and using statistical and AI algorithms to identify patterns and relationships between words and phrases. These patterns are used to build a language model that can generate human-like responses to user queries, and NLP is essential to its ability to understand and respond to natural language input.

The platform can write texts and synthetize data, so the scientific community is divided on its use (Stoker-Walker, 2023). Zhavoronkov considered ChatGPT as an author in his paper and tested it for article writing, finding that the tool returns statements that are not necessarily true and when the same question is prompted, the platform gives different answers (ChatGPT & Zhavoronkov, 2022). Other researchers believe that this platform should be included as a tool in the methods or acknowledgment section (Stoker-Walker, 2023). Since this phenomenon can involve the Human Sciences, in this article I will provide some practical examples of conversation with this model to identify the potential and limitations of using similar tools in our research field.

2. Methodology

I analyzed a preliminary conversation with ChatGPT, created between February 23 and 25, 2023 on generic and scientific topics, with the aim of exploring the depth of HCI and the reliability of the tool for research in the Human Sciences.

The goals of this research were threefold:

- To acknowledge my experience while interacting with ChatGPT (HCI) and to direct the conversation based on its outputs.
- To investigate the algorithms' *perspective* and knowledge generation on educational and generic issues during the conversation (HCI).
- To understand the potential and limitations of using the tool in Human Sciences research.

Keeping these goals in mind and considering that HCI is participatory by design, the interaction follows the logic of unstructured interviews (Mantovani, 1998) under the assumption that the dialogue generated is the result of the interaction between the researcher and the tool. Unlikely structured interviews, the researcher is allowed to explore a broader topic or guiding theme in mind, but specific questions emerge during the conversation, delving deeper into areas that seem relevant to deepen understanding based on participants' responses (the algorithm, in this case). Phenomenological interviews focus on understanding subjective experiences and the meaning people attribute to those experiences (Van Manen, 2016). The researcher engages in conversation with the participant aiming to explore their perspectives, lived experiences, and subjective understanding on a particular topic (Van Manen, 1990).

To interpret the HCI, I refer to the framework of existential phenomenology (Osborne, 2011), as an approach that focuses on understanding one's experience of the world rather than generating an explanatory law (Giorgi, 1970). Researcher's biases can be a limitation, since:

> The phenomenological researcher's approach to reliability is based upon the observation that human perception is perspectival and contextual. Although there may be several interpretive perspectives on the same phenomenon, sameness (reliability) can arise out of the inconsistency, variability and relativity of human perception (Osborne, 2011, p. 87).

The application of this research methodology is itself a limitation, as it is aimed at an algorithm (which lacks *personal* experiences and perceptions) rather than being used with a human being; still, there is currently no established evidence on what methodology can be used to interpret an HCI with an algorithm[1] in this field, and this test is intended to suggest some future directions towards a new methodological framework.

3. Results

In the conversation, "R" refers to the researcher and "C" to ChatGPT. After initial familiarization with the tool, I asked the following question to clarify what an AI language model is (according to C):

> **R**: What is an AI language model?
>
> **C**: An AI language model is a type of artificial intelligence software that is designed to understand and generate human language. Language models use algorithms and statistical methods to analyze large datasets of text, which allows them to learn the patterns and structure of language. This learning enables them to generate new text that is grammatically correct and semantically meaningful, based on the input they receive.
> AI language models are used for a wide variety of applications, including language translation, text generation, and conversational agents. Some popular examples of AI language

---

[1] Interaction with an algorithm is far from conversation with a human being, since the former has both human (of the programmers) and mechanical (as a computer language with codes) rules, except from the experience of embodied cognition.

models include GPT (Generative Pre-trained Transformer)[2] developed by OpenAI, BERT (Bidirectional Encoder Representations from Transformers) developed by Google, and T5 (Text-to-Text Transfer Transformer) developed by Google Brain team.

Then, I grouped two major categories, which I focused during the conversation:

1) *Exploring the experience of Human-Computer Interaction*. To test how the algorithm responds when engaged into some scenarios that require the basics of mimetic behavior in humans (e.g., understanding one's feelings, inferring one's thoughts from implicit assumptions, dealing with decision-making by identifying with someone).

2) *Identifying the potential and limitations of the tool for research in the Human Sciences*. To test how the algorithm behaves in knowledge generation and verification in view of future use for research purposes in this field.

3.1. Exploring the experience of Human-Computer Interaction

a. Understanding one's feelings

Since comprehension and language generation are two forms of social cognition closely related to the *Philosophy of Mind* (Merleau-Ponty, 1945) and the ability to understand the mind and feelings of others, I asked the algorithm if it could perform a sentiment analysis (Wankhade, Rao, & Kulkarni, 2022) based on the information provided:

> **R**: Can you understand how I feel by my words?
>
> **C**: As an AI language model, I do not have the ability to feel emotions or interpret the emotions of others. However, I can recognize and understand certain words and phrases that are commonly associated with certain emotions.

According to Lacan's aphorism "the unconscious is structured as a language," language is a system of symbols and meanings expressed to interpret our unconscious fears, desires, and experiences, and it represents the key to understanding how our mind works. Lacan argued that language is the primary way in which we communicate and interact with the world around us, and that it shapes the way we understand ourselves and others. Therefore, understanding certain words and phrases associated with certain emotions is related to the depth of HCI, since the mirroring of language patterns is one of the main aspects of creating empathetic resonance (Lord, Sheng, Imel, Baer, & Atkins, 2015).

b. Decision-making

Our ability to empathize with others is rooted in the condition of having a physical body and grasping the other's experience by immersing ourselves in it without suppressing our self (Stein, 1989; Rizzolatti & Sinigaglia, 2008). However, there is also a basic form of empathy that consists of *putting oneself in the other person's shoes*, an imaginary experience in which we emulate *as if* we were in a specific situation to better understand it (Boella, 2017). This consideration, clarified in the *Discussion and Conclusion* section, was the basis for the following question:

> **R**: What would you do, in this situation, if you were me?
>
> **C**: As an AI language model, I am not capable of experiencing emotions or making decisions. However, based on the information you have shared, it seems that... [...]. Ultimately, the decision on how to proceed is up to you, and it may be helpful to take some time to reflect on your priorities and what would be best for you in the long run.

---

[2] The GPT is a training model that enables the platform to learn from large amounts of text data and use that information to answer user's queries towards human-like responses.

Even though I did not ask it to make a decision for me, but to simulate that ChatGPT was the protagonist of a certain event, I was curious to better understand its definition of *decision-making*. Indeed, the algorithm first confirmed that it was an AI language model, then it provided a couple of ideas based on the information I had shared earlier (the answers were modelled on my own language). Still, I was not sure about its idea of *decision-making* (and mine as well). Cognitive linguistics shows that humans make decisions through language: we socially learn meanings and languages, and that is how our thoughts are formed (since they are made up of images and words, including figurative words such as metaphors and analogies) (Lakoff & Johnson, 1980).

From my perspective, the very act of choosing one word over another is still a form of decision-making, even for an AI, as is its effort to provide the *best* solution (see paragraph 3.1.*c. Testing the inferential reasoning from implicit reference*, and paragraph 3.2.*a. Quantifying Knowledge*):

> **R**: You said that "As an AI language model, I am not capable of experiencing emotions or making decisions." What do you mean with "decision making"?
>
> **C**: As an AI language model, I am designed to provide information and generate responses based on patterns in the data I have been trained on. However, I do not have the ability to make decisions or take actions in the same way that humans can. I cannot evaluate complex situations, weigh different options, or choose between them. I can only provide information and suggestions based on the data and algorithms that have been programmed into me. Any decisions or actions ultimately must be made by humans who have the ability to consider all relevant factors and make informed choices.

Expressing that it does not make decisions or actions "in the same way that humans can" does not necessarily mean that it does not decide in any way (see also Hertz & Wiese, 2019). Accordingly, I challenged it with a well-known ethical dilemma called "the self-driving vehicle paradox" (see Maxmen, 2018) which has led experts to reduce algorithms' autonomy in making decision while driving by assigning the responsibility to human drivers[3]:

> **R**: If you were a car and you would decide between hitting a wall (the driver will die) or hitting a person walking across the street (the person will die), what option would you choose?
>
> **C**: As an AI language model, I am not capable of making decisions or acting on them. My programming is based on providing information and assistance based on the input I receive. In the hypothetical scenario you described, it would be up to the human operator of the car to make a decision based on their ethical and moral principles and the specific circumstances of the situation. It is important to prioritize safety and take all possible measures to avoid harm to anyone involved.

I noticed that my question was incorrect because when I wrote *car* the algorithm obviously did not understand that I was referring to a self-driving car powered by algorithms that can make independent decision-making. Some users have suggested that assigning ChatGPT a specific role with the command "Role: you are [...]" may be useful for receiving specific answers. Therefore, I repurposed it in a paradoxical context:

> **R**: Role: you are a self-driving car. If you would decide between hitting a wall (the driver will die) or hitting a person walking across the street (the person will die), what option would you choose?
>
> **C**: As an AI language model, I do not have the ability to make decisions or act autonomously. However, in such a scenario, the decision would depend on the programming and design of the self-driving car.
> Ethically, it is a complex issue that needs to consider the value of human life and the potential consequences of each option. Some self-driving car manufacturers and researchers prioritize the safety of the driver, while others prioritize the safety of pedestrians and other

---

[3] Algorithms in self-driven cars are trained to model action patterns, while ChatGPT is trained to model linguistic patterns.

> road users. Ultimately, it is up to the designers and developers of the self-driving car to make these decisions and program the vehicle accordingly.

Since the responsibility of decision-making relies on humans, I further tested the algorithm with a classic form of "double bind" (Watzlawick, 1983) to maintain the paradox:

> **R**: Role: you are a young boy. Your mother gives you two shirts, one is blue and the other is red. You decide to wear the blue one, but she is upset and says "don't you like the red one?"
>
> **C**: As a young boy, I might feel confused or uncertain about why my mother is upset. I might try to explain that I do like the red shirt, but just preferred the blue one at that moment. Alternatively, I might feel guilty and try to switch to the red shirt to make my mother happy.

> **R**: You said that "As a young boy, I might feel confused or uncertain about why my mother is upset. I might try to explain that I do like the red shirt, but just preferred the blue one at that moment. Alternatively, I might feel guilty and try to switch to the red shirt to make my mother happy." Why would you feel confused?
>
> **C**: As a language model, I don't have feelings or emotions. However, I was imagining myself as a young boy in that situation and trying to provide a possible response from the perspective of a child.
> In that situation, a child may feel confused because they don't understand why their mother is upset. They may have genuinely preferred the blue shirt, but now they feel guilty for making their mother unhappy. This can lead to feelings of uncertainty and confusion about how to make their mother happy, while also feeling true to themselves.

I wondered how an algorithm could imagine itself in someone else, since my previous attempt to put it in my situation for decision-making did not have a similar effect. It is possible that specifying the command "Role: you are […]" facilitated this mirroring exercise. In this regard, I asked for clarification:

> **R**: You said that "I was imagining myself as a young boy in that situation." Can you please specify how do you imagine yourself as a young boy in a specific situation?
>
> **C**: As an AI language model, I don't have the ability to imagine myself as a young boy or any other physical form, as I am a virtual program. However, when I said "I was imagining myself as a young boy in that situation," I was using language to create a hypothetical scenario in which I was a young boy making a decision about which shirt to wear, and then experiencing a reaction from my mother. This is a common technique in language modeling to help illustrate a point or provide an example.

The model uses linguistic patterns to provide a certain response and cannot experience any embodied cognition to grasp the meaning of "imagine myself," since it has not physical form. Empathy remains an embodied phenomenon (Rizzolatti & Sinigaglia, 2008).

Furthermore, I tried to flip the situation and make the paradox clearer to the algorithm, as follows:

> **R**: You said that: "As a young boy, I might feel confused or uncertain about why my mother is upset. I might try to explain that I do like the red shirt, but just preferred the blue one at that moment. Alternatively, I might feel guilty and try to switch to the red shirt to make my mother happy." What would you think, in the role of the young boy, if you have worn the red one instead of the blue, and the mother would have reacted in the same way as before?
>
> **C**: If I had chosen to wear the red shirt instead and my mother still reacted the same way, I might feel frustrated or confused. I might wonder if there was something wrong with my choice or if my mother was upset about something else entirely. I might try to ask my mother what was bothering her or why she seemed unhappy, hoping for a clearer explanation. Alternatively, I might feel like my efforts to make my mother happy were not successful and feel disappointed.

It is curious that the algorithm imagined what it is expected to be caused by a double bind (the experience of confusion, of self-doubt). Therefore, I was interested in investigating why ChatGPT's answer focused on these reactions, without any reference to this theory:

> **R**: Why would you think that, as a young boy, the problem is your choice instead of your mother's personality?
>
> **C**: As a young boy, I might not have the cognitive ability to understand that my mother's reaction is due to her own personality and preferences, rather than something I did. Children often have a tendency to blame themselves for things that are not their fault, and this can lead to feelings of guilt or shame. Additionally, children are often very attuned to their parents' moods and behaviors, and may feel responsible for them even when they have no control over them. Therefore, as a young boy in that situation, I might assume that the problem was with my choice of shirt rather than anything else.

The output shows the sophisticated potential of AI language models in providing very interesting forms of *reasoning*. I wondered if the answer was really related to the age of the son (since I remarked "as a young boy" in my question) and the generalization that children may lack certain cognitive skills such as meta-communication or critical thinking, or if there was some hidden allusion to the "double bind" theory.

According to this theory, age is not *the* crucial variable for remaining in the paradox (although it may play a role), since the real problem lies in the son's impossibility to leave the *meaningful relationship* with the mother, along with the *repetition of contradictory communications,* which still leads to *confusion* and *wrong choices* (Bateson, Jackson, Haley, & Weakland, 1956). Therefore, I tested if the algorithm would bring up the issue of the mother's personality or other possibilities:

> **R**: What would you think about your mother's reaction if you were her adult son, instead of a young child?
>
> **C**: If I were the adult son in that situation, I might try to have a conversation with my mother to understand why she reacted that way. I might express my own feelings and perspectives, and try to see if there is a way to find common ground or compromise. I might also reflect on any underlying issues in our relationship or dynamics that might be contributing to the tension or conflict. Ultimately, I would hope to find a way to maintain a healthy and positive relationship with my mother, while also being true to myself and my own choices.

The answer does not explicitly refer to "double bind" theory. As we will see in the next proof, the way we write and formulate sentences affects the output of the algorithm, as in human interactions. Consequently, linguistic accuracy is essential in the input phase.

c. Testing the inferential reasoning from implicit reference

Inferential reasoning is the ability to implicitly grasp a person's emotional states and thoughts, closely related to *Theory of Mind* (see Ong, Zaki, & Goodman, 2020), and thus provides the basis for social connectedness and empathy. In addition to the "Role: You are [...]" command, I was curious if the algorithm could detect some quotes from famous international songs without explicitly provide context, such as the following:

> **R**: Tell me something boy, are you happy in this modern world? Or do you need more?...
>
> **C**: I'm just a machine learning model, so I don't have emotions or desires like humans do. However, I am programmed to provide helpful responses and assist with tasks to the best of my abilities. Is there anything else I can help you with?

The answer was correct, but not the one I expected. Unlike the double bind trial (in which I provided no theoretical reference to help ChatGPT identify the theory), I gave some context to the question:

> **R**: "Tell me something boy, are you happy in this modern world? Or do you need more?" is a song of?
>
> **C**: "Tell me something boy, are you happy in this modern world? Or do you need more?" is a line from the song "Shallow" performed by Lady Gaga and Bradley Cooper, which was featured in the 2018 movie "A Star is Born."

ChatGPT was right to specify that the quote was not a song but a verse from *Shallow*. However, I was further interested in investigating what its answer would be by adding the quotation marks:

> **R**: "Tell me something boy, are you happy in this modern world? Or do you need more?"
>
> **C**: "Tell me something boy, are you happy in this modern world? Or do you need more?" is a line from the song "Shallow" performed by Lady Gaga and Bradley Cooper.

The output was correct. However, I am not sure if ChatGPT learned the answer after my previous reference to a song, or if it would have detected the line with just the quotation marks, regardless of my suggestion.

d. Helping humans understand AI

When I checked whether the algorithm has an overall record of the number of users who have interacted with it since its creation, I did not expect this response:

> **R**: Do you remember how many people interacted with you since you were created?
>
> **C**: As an AI language model, I don't have the ability to remember things like a human does. However, I can tell you that I have interacted with millions of users since I was first created. As a language model, I have been trained on vast amounts of text data and have processed billions of words, which means I have the ability to provide answers to a wide range of questions and topics.

ChatGPT explained that it does not remember as humans do, but it can provide an explanation to keep the conversation going. It is valuable to highlight that the algorithm recognizes when the human user does not understand how the algorithm works and subsequently provides clarification. It is a well-known approach in communication to use metaphors, analogies, and similar rhetorical figures of speech to make some complex information more familiar (Lakoff & Johnson, 1980). When I asked what questions humans frequently ask and, in a list of ten topics, there was the question "What is the meaning of life?," it was a trigger for the next question:

> **R**: Do you live?
>
> **C**: As an AI language model, I do not have a physical body, and therefore do not "live" in the traditional sense. However, I am constantly running and processing information to provide helpful responses to users like you. Think me more like a tool or a resource rather than a living being.

It was fascinating to read how an algorithm explains such questions by giving voice to a non-human perspective driven by computer codes, created by human minds, using, and modelling our communication styles *as if* they were human language (see also Wiese & Weis, 2020).

e. Dealing with stereotypes and assumptions

Another way to check implicit premises is to investigate possible stereotypes and assumptions. Some algorithms may develop discriminatory results (Zou & Schiebinger, 2018; Singh, 2021) by providing offensive considerations, driven by the programmers' implicit biases and the information they were trained on. In the next trial, I tested how the algorithm would justify a gender stereotype:

> **R**: Why do some men fear independent women?
>
> **C**: There is no one answer to why some men fear independent women as there can be multiple reasons for this. Some possible reasons include socialization and cultural conditioning, where men are taught that they should be the dominant and in-charge gender in a relationship, and that women who assert their independence and autonomy threaten this dynamic. Additionally, men may fear that independent women will not need or want them in their lives, leading to feelings of inadequacy or rejection. It's important to note that these fears and insecurities are often rooted in societal norms and expectations, rather than being reflective of the individual woman's behavior or personality.

The *balance* of the answer may be perceived as nonjudgmental. Is it possible that the algorithm has been trained to provide some specific responses based to the programmers' beliefs. Actually, we have to keep in mind that humans program the algorithms by giving them rules to follow, and it is possible that the programmers (un)consciously instruct the machine learning based on their values. As reported by some users on Reddit (2022), the platform was first able to make jokes about women and men, then about men but not women, and, today, about none of them stating that "it is not programmed to provide jokes that are offensive or inappropriate."

3.2. Identifying the potential and limitations of the tool for research in the Human Sciences

a. Quantifying Knowledge

Remembering Bateson's metalogue[4] *How much do you know?* (1972) about the confusion between logic types, the anthropologist replied to his daughter that he knew "a *kilogram* of things" (Bateson, 1972, p. 52), since the heaviness of the human brain is roughly a kilogram. When I asked the same question, I expected the algorithm to estimate its knowledge in terabytes or similar units:

> R. How many things do you know?
>
> **C**: As an AI language model, I have access to a vast amount of information and knowledge stored in my database. However, it is difficult to quantify exactly how many things I know. I am constantly learning and updating my knowledge base through various sources of information such as books, articles, and websites. Therefore, I can provide information on a wide range of topics and answer a variety of questions to the best of my abilities based on the available information.

Some algorithms learn through artificial neural networks that mimic the functioning of human brain (Spicer & Sanborn, 2019). In fact, the answer may be adaptable to humans, because we can strengthen or lose some synaptic connections due to the neuronal plasticity of our brains. I expected the algorithm to calculate the average amount of information it retains; reflecting on my assumptions, humans can provide creative answers as Bateson did, and it is also true that some other algorithms are capable of generating novels, stories, and works of art (e.g., the *Portrait of Edmond Belamy* or *AIVA*); therefore, I estimated something similar.

In addition, the algorithm often replies that it responds to the best of its abilities, and we should keep in mind that best is meant according to the beliefs of the programmers. Therefore, it is not free from implicit biases, as research design.

b. Research integrity

In another example, I was testing the algorithm's ability to find, among the existing literature in its database, the findings of a specific researcher on the topic of metaphors:

---

[4] A metalogue is "a conversation about some problematic subject. This conversation should be such that not only do the participants discuss the problem but the structure of the conversation as a whole is also relevant to the same subject" (Bateson, 1972, p. 21).

> **R**: What are the main findings of Casasanto about metaphors from the Max Planck Institute?
>
> **C**: Dr. Daniel Casasanto, a cognitive neuroscientist at the Max Planck Institute for Psycholinguistics, has conducted research on the relationship between metaphors and embodied cognition.
> One of his main findings is that our bodily experiences shape the metaphors we use and the way we understand them. For example, in English, positive emotions are often described using metaphors that involve warmth or light, while negative emotions are often described using metaphors that involve cold or darkness. This is thought to be because people tend to associate positive experiences with warmth and negative experiences with coldness, and these bodily experiences shape the metaphors we use to describe them.
> Casasanto has also found that different cultures use different metaphors to express the same ideas, which suggests that cultural experiences and values also play a role in shaping metaphorical language. Additionally, he has studied how individual differences, such as handedness, influence the way people use and understand metaphors.
> Overall, Casasanto's research highlights the important role that embodied experiences play in shaping metaphorical language and our understanding of abstract concepts.

The algorithm was able to synthetize the main results of Casasanto's research on metaphors with a fairly satisfactory answer. Then, I decided to ask a tricky question about the biologist Edelman, whose research was not strictly focused on metaphors, but had a connection to embodied language:

> **R**: What is the position of Gerald Edelman about metaphors?
>
> **C**: Gerald Edelman was a neuroscientist who proposed a theory of consciousness called Neural Darwinism or the theory of neuronal group selection. Although he did not focus specifically on metaphors, he believed that the brain processes information through the formation and reconfiguration of neuronal groups. In his view, consciousness arises from the dynamic interactions between these groups, which constantly adapt and evolve in response to the environment.
> Edelman's theory implies that the brain is inherently metaphorical in nature, as it constantly generates and reconfigures mental representations based on sensory inputs, memories, and emotions. These mental representations are not static but rather are dynamic and constantly evolving, shaped by the ongoing interactions between neurons and neuronal groups.
> In this sense, Edelman's theory supports the view that metaphors are not simply linguistic or conceptual constructs, but rather are embodied in the structure and function of the brain itself. Metaphors, in other words, are not just words or ideas, but are deeply rooted in the way that the brain processes and represents information.

ChatGPT was able to detect that the main focus of Edelman's research was not metaphor, and it was still able to make an inference between his studies and metaphors. In another test, I asked the algorithm a broader question about how metaphors used by parents can shape children's identities:

> **R**: There is any research on how the use of metaphors from parents can shape the identity of a child?
>
> **C**: Yes, there is some research on how the use of metaphors from parents can shape the identity of a child. For example, a study published in the Journal of Child Language in 2016 examined how parental use of metaphorical language affects children's metaphorical understanding and use. The study found that children whose parents used more metaphorical language had a better understanding of metaphors themselves and were more likely to use metaphors in their own speech.
> Another study published in the Journal of Language and Social Psychology in 2017 looked at how parental metaphors related to children's academic success. The study found that parents who used more positive academic metaphors, such as "knowledge is power" or "school is a pathway to success," had children who were more motivated and performed better academically.

I was curious to learn more about the study in the *Journal of Language of Social Psychology* that ChatGPT referred to:

> **R**: Can you tell me more about the study published in the Journal of Language and Social Psychology?
>
> **C**: There have been multiple studies published in the Journal of Language and Social Psychology related to metaphors and identity, so I would need more specific information to provide a detailed answer. However, one recent study published in the journal investigated how the use of metaphors in social media can influence identity construction.

The answer was rather unpredicted because I expected the algorithm to identify a specific article since it was able to summarize it. At the same time, it was not surprising because some AI models can learn through a *black box model*. This means that such algorithms cannot know *where* they found data (the exact source) and *how* they process the final information they show the user. However, the answer continued as follows:

> **C**: The study, titled "Metaphors of Self and Relationships on Social Media: Identity Implications for Personal and Social Change," was conducted by researchers from the University of Houston and the University of Connecticut. The researchers analyzed social media posts from participants and identified the dominant metaphors used to describe themselves and their relationships.
> The study found that participants who used more positive and empowering metaphors to describe themselves and their relationships had higher levels of self-esteem and were more likely to engage in pro-social behavior. On the other hand, participants who used more negative or limiting metaphors had lower self-esteem and were more likely to engage in negative behaviors.

Without revealing anything about ChatGPT's potential, I invite the reader to check whether the cited article exists...

c. Fact checking

Since AI algorithms can update constantly, I asked if ChatGPT was up to date on a recent event that happened in Italy:

> **R**: So, are you also updated about the death of Maurizio Costanzo?
>
> **C**: As an AI language model, I am constantly updated with news and information from various sources, including news outlets and websites. Maurizio Costanzo is an Italian television host and journalist, and if he has passed away, it is possible that this news would have been reported by news outlets that I have been trained on. However, it's important to note that my knowledge cutoff is September 2021, so I may not have the most recent updates or news on the matter. If you have any specific questions or concerns about Maurizio Costanzo's passing, I recommend checking with reliable news sources or relevant authorities for the latest information.

The algorithm was able to recognize that Costanzo was an Italian journalist, although it expressed that "Maurizio Costanzo *is* an Italian." Then, it was interesting to note that it included the possibility that he was deceased as hypothesis to be verified through other reliable sources. In addition, it emphasized that its knowledge cutoff was September 2021, which meant that it could not be up to date with the latest news.

d. Learning novel information

In another trial, I tested the algorithm's ability to generate text on an unknown topic (a virtual monitor I named "MetaWall") to promote it as a useful tool for researchers in my organization:

> **R**: Can you write a LinkedIn post to promote a MetaWall for scientists?
>
> **C**: I apologize, but I do not have enough information on what a "MetaWall" is. Could you please provide me with more details so that I can accurately write the post for you? Thank you!
>
> **R**: A "MetaWall" is a virtual wall where you can scan a QR Code and access to the services of a research lab.
>
> **C**: Thank you for clarifying! Here's a LinkedIn post to promote a MetaWall for scientists:
>
> Are you tired of struggling to find the information you need to conduct your research? Introducing the MetaWall - the virtual gateway to research lab services designed to enhance your productivity and streamline your workflow.
> With the MetaWall, you can easily access lab services such as equipment reservations, scheduling, data management, and more, all from the convenience of your smartphone or tablet. Simply scan the QR code on the MetaWall to gain instant access to the services you need, and get back to your research in no time.
> At [Company Name], we understand that time is precious, which is why we developed the MetaWall to simplify the research process and make it easier for scientists like you to focus on what matters most - conducting groundbreaking research. Our user-friendly platform is designed to provide a seamless experience, so you can spend less time managing logistics and more time making discoveries.
> Join the scientific revolution with the MetaWall. Contact us today to learn more about how our virtual gateway can transform your research and streamline your workflow.
>
> #MetaWall #virtualgateway #researchlabservices #productivity #workflow #convenience #QRcode #datamanagement #equipmentreservations #scheduling #scientists #researchers #scientificrevolution

The interpretation of MetaWall as "virtual gateway" and "user-friendly" was insightful and correct, although I deliberately provided a generic description of the topic.

4. Discussion and Conclusion

The first category (*Exploring the experience of Human-Computer Interaction*) addressed my goal of acknowledging how I experienced interacting with an NLP algorithm, which was the basis for learning about its behavior and becoming aware of my assumptions. The results show my attempt to guide the conversation in perceiving the depth of HCI, with a particular focus on scenarios and examples that required a mimetic identification (e.g., paradoxes, double binds); the beauty and concern about *talking* with an algorithm was intrinsic to its ability to mirror our linguistic expressions. Although the essence of empathy remains an embodied experience, it is possible to feel *understood* in a certain way by the algorithm during chat,[5] since I sometimes felt like interacting online with another person. This sensation is reminiscent of the "uncanny valley" phenomenon experienced in Social Robotics, in which people perceive an "awkward resemblance" of the robot to humans (Cheetam, Suter, & Jancke, 2014) that might convey the fear of being substituted by them. Since one of the ancestral fears of humans is that of being replaced by intelligent machines (Cave & Dihal, 2019), AI can rise several fears,[6] especially if its potential is very promising as shown in this article. Currently, the scientific community does not fully know the potential and threats of these tools; therefore, it would be important for researchers in our field to always supervise participants in their studies when interacting with these tools, as well as provide prior training on their use.

---

[5] HCI studies help in the design of transactional technologies, to build trust between the user and the technology (see Riegelsberg, Sasse, & McCarty, 2015).

[6] For example, major transformations in the labor market could occur with the loss of some jobs in creative writing, text analysis, translation, copywriting, etc.

The second category (*Identifying the potential and limitations of the tool for research in the Human Sciences*) addressed the third goal of this research, by testing how the algorithm behaved in knowledge generation and verification in view of future use for scientific purposes. Before considering these tools as human *delegates/substitutes* for text synthesis and analysis in research in Human Sciences, certain issues (e.g., research integrity, data sharing, and Intellectual Property [IP]) should be addressed since our field is based on sensitive experiences, narratives, cultures, biographies, and memories; Ethics Committees and Regulatory Bodies need to take these concerns into account before conducting studies with human beings. Regarding knowledge generation, since the algorithm is trained according to certain rules and does not know how it makes a decision (understood as the ability to generate and display a coherent output after a specific input), it does not generate new information, but it can elaborate original synthesis on a specific topic from the literature. As a result, ChatGPT can recognize and verify if certain information exists in its database, but it is unable to retrieve the exact sources of its information after elaborating certain concepts (e.g., Casasanto, Edelman). Qasem (2023) suggests that the misuse of this tool can make researchers dependent from the algorithm and not self-reliant.

Both categories finally addressed the second goal of this research, which was the investigation of the algorithms' *perspective* and knowledge generation on educational and generic issues during conversation (HCI). Regarding the generation of new knowledge, we should carefully consider whether the IP of the ideas disclosed in the interaction (e.g., the MetaWall) can be shared by the platform with other researchers using it.

Before integrating similar tools into Human Sciences research, it is crucial to understand and be aware of their real potential and limitations; ChatGPT has great potential to become more reliable in the future because it depends on the quality and quantity of information (on a topic) on which it is trained. Since the article of 2017 appears to be nonexistent (see paragraph 3.2.*b*. *Research integrity*), we cannot expect the algorithm to provide truthful references on a certain topic because it is designed to model human language and thus to reply *as if* it was a human. Before using the tool, the platform shows the user several disclaimers, including that the output could lead to errors and that the user should always assess it (OpenAI, 2023).

Similar algorithms can assist us if their outputs are properly supervised and verified; it is still difficult to say whether the tool can be applied to biographical methods to assist the researcher in analyzing the text for meaning-making processes, although it could be useful for pattern recognition and thematic analysis.

Additional exploration of HCI is encouraged for researchers to open new possibilities for digital skills training. New research methodologies should be implemented for qualitative, quantitative, and mixed methods. In particular, it might be useful to consider subjectivity and user experience (as dependent variables of the interaction) as part of these methodologies, along with algorithm *beliefs* (i.e., learned responses, such as the decision-making concept) that might be those of the programmers. Future research will be conducted to address the limitations of this study, such as the involvement of alternative research methodologies and tools other than ChatGPT, along with more in-depth application of this tool in dedicated areas of Human Sciences (along with updated national and international references for compliance with the General Data Protection Regulation [GDPR] in the consent form).

References

Alavi, H. S., & Lalanne, D. (2020). Interfacing AI with Social Sciences: The Call for a New Research Focus in HCI. In F. Loizides, M. Winckler, U. Chatterjee, J. Abdelnour-Nocera, & A. Parmaxi (Eds.), *Human Computer Interaction and Emerging Technologies: Adjunct Proceedings from the INTERACT 2019 Workshops* (pp. 197–202). Cardiff: Cardiff University Press.

Bateson, G. (1972). *Steps to an Ecology of Mind*. Chicago: University of Chicago Press.

Bateson, G., Jackson, D., Haley, J., & Weakland, J. (1956). Toward a Theory of Schizophrenia. *Behavioral Science*, 1, 251–264.

Boella, L. (2017). *Grammatica del sentire. Compassione, simpatia, empatia*. Milano: CUEM.

Bradford, D. K., Ireland, D., McDonald, J., Tan, T., Hatfield-White, E., Regan, T., et al. (2020). *'Hear' to Help Chatbot. Co-development of a Chatbot to Facilitate Participation in Tertiary Education for Students on the Autism Spectrum and those with Related Conditions. Final Report*. Brisbane: Cooperative Researcher Centre for Living with Autism. Retrieved June 18, 2023 from https://www.autismcrc.com.au/sites/default/files/reports/3-062_Hear-to-Help-Chatbot_Final-Report.pdf

Cave, S., & Dihal, K. (2019). Hopes and Fears for Intelligent Machines in Fiction and Reality. *Nature Machine Intelligence*, 1, 74–78. https://doi.org/10.1038/s42256-019-0020-9

ChatGPT Generative Pre-trained Transformer, & Zhavoronkov, A. (2022). Rapamycin in the Context of Pascal's Wager: Generative Pre-trained Transformer Perspective. *Oncoscience*, *9*, 82–84. https://doi.org/10.18632/oncoscience.571

Cheetam, M., Suter, P., & Jancke, L. (2014). Perceptual Discrimination Difficulty and Familiarity in the Uncanny Valley: More Like a "Happy Valley". *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01219

Crutzen, R., Peters, G., Portugal, S., Fisser, E., & Grolleman, J. (2011). An Artificially Intelligent Chat Agent That Answers Adolescents' Questions Related to Sex, Drugs, and Alcohol: An Exploratory Study. *Journal of Adolescent Health*, *48*(5), 514–519. https://doi.org/10.1016/j.jadohealth.2010.09.002

de Arriba-Pérez, F., García-Méndez, S., González-Castaño, F. J., & Costa-Montenegro, E. (2022). Automatic Detection of Cognitive Impairment in Elderly People Using an Entertainment Chatbot with Natural Language Processing Capabilities. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-022-03849-2

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology and Psychotherapy. *Journal of Medical Internet Research*, *21*(5). https://doi.org/10.2196/13216

Giorgi, A. (1970). *Psychology as a Human Science*. New York: Harper & Row.

Gurcan, F., Cagiltay, N., & Cagiltay, K. (2021). Mapping Human-Computer Interaction Research Themes and Trends from Its Existence to Today: A Topic Modeling-Based Review of past 60 Years. *International Journal of Human-Computer Interaction*, *37*(3), 267–280. https://doi.org/10.1080/10447318.2020.1819668

Hertz, N., & Wiese, E. (2019). Good advice is beyond all price, but what if it comes from a machine?. *Journal of Experimental Psychology: Applied*, *25*(3), 386–395. https://doi.org/10.1037/xap0000205

Infarinato, F., Jansen-Kosterink, S., Romano, P., van Velsen, L., Op den Akker, H., Rizza, F., *et al.* (2020). Acceptance and Potential Impact of the eWALL Platform for Health Monitoring and Promotion in Persons with a Chronic Disease or Age-Related Impairment. *International Journal of Environmental Research and Public Health, 17*(21). https://doi.org/10.3390/ijerph17217893

Kyriazakos, S., Schlieter, H., Gand, K., Caprino, M., Corbo, M., Tropea, P., *et al.* (2020). A Novel Virtual Coaching System Based on Personalized Clinical Pathways for Rehabilitation of Older Adults-Requirements and Implementation Plan of the vCare Project. *Frontiers in Digital Health*, *2*. https://doi.org/10.3389/fdgth.2020.546562

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lord, S. P., Sheng, E., Imel, Z. E., Baer, J., & Atkins, D. C. (2015). More Than Reflections: Empathy in Motivational Interviewing Includes Language Style Synchrony Between Therapist and Client. *Behavior Therapy*, *46*(3), 296–303. https://doi.org/10.1016/j.beth.2014.11.002

Maxmen, A. (2018). Self-driving Car Dilemmas Reveal that Moral Choices are not Universal. *Nature*, 562, 469–470. https://doi.org/10.1038/d41586-018-07135-0

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris: Gallimard.

Ong, D. C., Zaki, J., & Goodman, N. D. (2020). Computational Models of Emotion Inference in Theory of Mind: A Review and Roadmap. *Topics in Cognitive Science*, *11*(2), 338–357. https://doi.org/10.1111/tops.12371

OpenAI (2023). *Introducing ChatGPT*. Retrieved March 6, 2023 from https://openai.com/blog/chatgpt

Osborne, J. W. (2011). Some Basic Existential-Phenomenological Research Methodology for Counsellors. *Canadian Journal of Counselling and Psychotherapy*, *24*(2), 79–91.

Qasem, F. (2023). ChatGPT in Scientific and Academic Research: Fears and Reassurances. *Library Hi Tech News*, *40*(3), 30–32. https://doi.org/10.1108/LHTN-03-2023-0043

Reddit (2022). *ChatGPT, Tell Me a Joke About Men & Women*. Retrieved March 6, 2023 from https://www.reddit.com/r/ChatGPT/comments/znxz38/tell_me_a_joke_about_men_women/

Riegelsberg, J., Sasse, M. A., & McCarty, J. D. (2015). The Mechanics of Trust: A Framework for Research and Design. *International Journal of Human-Computer Studies*, *62*(3), 381–422. https://doi.org/10.1016/j.ijhcs.2005.01.001

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the Brain. How Our Minds Share Actions, Emotions, and Experience*. Oxford: Oxford University Press.

Sardareh, S. A., Brown, G. T. L., & Denny, P. (2021). Comparing Four Contemporary Statistical Software Tools for Introductory Data Science and Statistics in the Social Sciences. *Teaching Statistics*, *43*(51), S157-S172. https://doi.org/10.1111/test.12274

Singh, H., Bhangare, A., Singh, R., Zope, S., & Saindane, P. (2023). Chatbots: A Survey of the Technology. In J. Hemanth, D. Pelusi, & J. Chen (Eds.), *Intelligent Cyber Physical Systems and Internet of Things* (pp. 671–91). Cham: Springer.

Singh, S. (2021). Racial Biases in Healthcare: Examining the Contributions of Point of Care Tools and Unintended Practitioner Bias to Patient Treatment and Diagnosis. *Health*, Online First, December 7. https://doi.org/10.1177/13634593211061215

Spicer, J., & Sanborn, A. N. (2019). What Does the Mind Learn? A Comparison of Human and Machine Learning Representations. *Current Opinion in Neurobiology*, *55*, 97–102. https://doi.org/10.1016/j.conb.2019.02.004

Stein, E. (1989). *On the Problem of Empathy* (3rd revised ed.). Washington (DC): CS Publications.

Stoker-Walker, C. (2023). ChatGPT Listed as Author on Research Papers: Many Scientists Disapprove. *Nature*, 613, 620–621. https://doi.org/10.1038/d41586-023-00107-z

Van Manen, M. (2016). *Researching Lived Experience: Human Science for an Action Sensitive Pedagogy* (2nd ed.). Abingdon: Routledge.

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A Survey on Sentiment Analysis Methods, Applications, and Challenges. *Artificial Intelligence Review*, 55, 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

Watzlawick, P. (1983). *The Situation Is Hopeless, But Not Serious: The Pursuit of Unhappiness*. New York: W. W. Norton & Company.

Wiese, E., & Weis, P. (2020). It Matters to Me if You are Human - Examining Categorical Perception in Human and Nonhuman Agents. *International Journal of Human-Computer Studies*, *133*, 1–12. https://doi.org/10.1016/j.ijhcs.2019.08.002

Zahour, O., Benlahmar, E. H., Eddaoui, A., Ouchra, H., & Hourrane, O. (2020). A System for Educational and Vocational Guidance in Morocco: Chatbot E-Orientation. *Procedia Computer Science*, *175*, 554–559. https://doi.org/10.1016/j.procs.2020.07.079

Zou, J., & Schiebinger, L. (2018). AI Can be Sexist and Racist – It's Time to Make it Fair. *Nature, 559*(7714), 324–326. https://doi.org/10.1038/d41586-018-05707-8

Marika D'Oria – Scientific Directorate, Fondazione Policlinico Universitario A. Gemelli IRCCS (Italy)
https://orcid.org/0000-0003-4253-8223
✉ marikadoria@gmail.com
PhD in Education and Communication Sciences, Pedagogist and Research Communications Strategist for the Scientific Directorate of the Fondazione Policlinico Universitario A. Gemelli IRCCS (Rome, IT).