# ChatGPT through postphenomenology and deconstruction: On the possibility of a Derridean philosophy of technology

Víctor Betriu Yáñez (2761122)

Word count: 23992

First supervisor: Dr Nolen Gertz
Second reader: Dr Dominic Lenzi

Master's Thesis (2022–2023)
MSc in Philosophy of Science, Technology and Society
Faculty of Behavioural, Management and Social Sciences — University of Twente

(This page intentionally left blank)

The question of writing opens as a question of morality.
—Derrida, *Plato's Pharmacy*

SOCRATES: You know, Phaedrus, writing shares a strange feature with painting. The offsprings of painting stand there as if they are alive, but if anyone asks them anything, they remain most solemnly silent. The same is true of written words. You'd think they were speaking as if they had some understanding, but if you question anything that has been said because you want to learn more, it continues to signify just that very same thing forever.
—Plato, *Phaedrus*

If I have a book which understands for me, a pastor who has a conscience for me, a physician who decides my diet, and so forth, I need not trouble myself. I need not think, if I can only pay—others will easily undertake the irksome work for me.
—Kant, *What is Enlightenment?*

The typewriter will alienate the hand of the man of letters from the pen only when the precision of typographic forms has directly entered the conception of his books. One might suppose that new systems with more variable typefaces would then be needed. They will replace the pliancy of the hand with the innervation of commanding fingers.
—Benjamin, *One-way Street*

The very "dynamism" of cinema snatches away from us the images over which we should like to linger and dream. … And this terrible machine, so elaborately dazzling, with its luxury, its music, its human voice, this machine for stupefying and destroying the mind, is today among the most astonishing forces in the world.
—Duhamel, *America the Menace: Scenes from the Life of the Future*

The cinematographic world is a dead, illusory and split [*tronçonné*] world. … The cinema world is a closed world, with no relation to existence. … It may be that cinema puts itself in the place of the human eye, that it thinks for it, that it sifts [*crible*] the world to it, and that, through this work of deliberated [*concertée*] and mechanical elimination, it leaves nothing but the best.
—Artaud, *La vieillesse précoce du cinéma*

(This page intentionally left blank)

# Summary

This master's thesis analyses ChatGPT through postphenomenology and deconstruction with a two-fold aim: on the one hand, to better understand what is structurally at stake in our ways of relating to ChatGPT, that is, even if all social risks were effectively mitigated; on the other hand, to explore the possibility of the thought of Jacques Derrida to contribute to mainstream philosophy of technology in order to sharpen our philosophical artillery to not only account for how concrete technologies transform our experience but also our ways of thinking. My research question is thus: How could Jacques Derrida's works help us better understand ChatGPT compared to mainstream approaches in philosophy of technology and what would this show regarding the possibility of a Derridean philosophy of technology? I address this question in three main steps.

In the first chapter, I identify the unresolved philosophical issues in the discourse surrounding ChatGPT. After presenting ChatGPT as a technology, I review what might be problematic in the popular ways of addressing this technology. In light of the prevalence of analyses regarding the ethical consequences of ChatGPT and the popularity of the problem of anthropomorphization, I argue that we might be clinging to a tacit technological instrumentalism insofar as we might be assuming that ChatGPT could potentially be an unproblematic means to fulfil desired ends once its social effects are deemed positive and we ascribe a non-anthropocentric identity to this technology. Given this suspicion, I argue for the need of a structural and interactionist analysis of ChatGPT—an analysis which addresses what is fundamentally at stake in our ways of relating to this technology.

In the second chapter, I consider how Ihdean postphenomenology could help us identify the structural issues in our interaction with ChatGPT by applying the framework of human-technology relations to our uses of this technology. After showing that we can establish an embodiment, hermeneutic, alterity and background relation with this technology, I provide an account of how our experience is transformed in each of these relations and attempt to derive a structural danger. While postphenomenology helps us better understand what occurs to our experience while we use ChatGPT—especially by foregrounding that which conceals from our awareness and might be forgotten—postphenomenology cannot explain why we so eagerly relate to ChatGPT in the first place. I thus propose the hypothesis that our common-sense understanding of the action we carry out through or with ChatGPT and the assumption that the technology is capable of carrying out such an action might be

the conditions of possibility of the relation. Finally, I decide to focus on the question of writing to see whether Derrida can take the baton.

In the third chapter, I consider how Derrida's philosophy could help us identify what is structurally at stake in our ways of relating to ChatGPT by tackling the question postphenomenology leaves unanswered—namely, whether our common-sense understanding of writing and the assumption that ChatGPT writes explain why we establish a relation with ChatGPT and what might be problematic about it. After presenting Derrida's critique of logocentric metaphysics as what motivates his project, I review an existing Derridean interpretation of ChatGPT which runs the risk of justifying this technology insofar as it sees ChatGPT as practically deconstructing logocentrism because it speaks without writing. I then hypothesize that ChatGPT could also be seen as a practical reinforcement of logocentrism and attempt to defend this interpretation in the remaining of the chapter. To do so, I clarify the extent to which we can assume from a Derridean perspective that ChatGPT writes, I offer a characterization of the specificity of ChatGPT's mode of writing, and I consider the implications this has for our ways of thinking. Finally, I address the so-what question regarding the analogy between ChatGPT and logocentrism to show why prompting logocentrism is problematic and why we should care about writing.

The thesis concludes that, while Ihde helps us identify ahistorical yet structural concerns for an impersonal subject, Derrida helps us situate the emergence of ChatGPT and how the latter affects our ways of thinking in the historical ambition of logocentric metaphysics without necessarily falling into a technological determinism. Additionally, I describe what a Derridean philosophy of technology could look like and outline some avenues for further research.

# Table of contents

# Introduction

This thesis is an investigation into the possibilities of Jacques Derrida's philosophy to contribute to mainstream philosophy of technology, which I here take to be a philosophy that, like the well-established postphenomenological current inaugurated by Don Ihde (1979, 1990) and followed by the so-called Empirical Turn (Achterhuis, 1999/2001; Verbeek, 2005, 2011, 2021), is concerned with how concrete technologies affect our experience of the world in order to account for what is structurally at stake in our interactions with technologies and thus provide reasons to possibly deliberate whether we want them. Although Derrida's work has had a considerable influence to think about technicity from anthropological and ontological perspectives (Stiegler, 1994/1998; Bradley, 2011; Lindberg, 2016, 2023), it has barely been utilized to interpret concrete technologies. Derrida's philosophy faces us with an unconditionally critical perspective which continuously remind us that "no rest [is] allowed for any form of good conscience" (Derrida, 1993/1994, p. xiv), and his works exhaustively foreground the rooted violence in our common sense, "that most widely shared thing in the world" (Derrida, 2003/2005, p. 148). In light of this, Derrida's works might help us sharpen our philosophical artillery to account for how concrete technologies not only shape experience but also affect and are affected by our interiorized and naturalized ways of thinking. In short, my motivation here is to explore whether and how Derrida could be utilized to philosophically analyse concrete contemporary technologies.

While the facts that both Derrida's (1962/1978, 1967/2011a) and Ihde's (2016) philosophies arise out of a confrontation with Husserlian phenomenology and that Ihde (1983) has spilled some ink on the relation between deconstruction and phenomenology are very tempting to start off our investigation from a purely theoretical perspective, I have on the contrary decided to circumscribe this investigation to a double analysis of a concrete technology—ChatGPT through postphenomenology and deconstruction. The role of postphenomenology in this double analysis is instrumental: it will help us show what a well-established framework in philosophy of technology is capable of and this will provide us with a point of reference against which to highlight the novel aspects Derrida could bring in. As a result, I hope to show the practical import of Derrida's texts to think critically about our contemporary technological world, without falling into a naïve optimism and hopefully being sensitive to the difference of each technological developments. The basic question that remains in the background of this thesis is thus: why should we still read Derrida today? The reasons why I

delimit the scope of this research to ChatGPT are the political urgence of developing conceptual approaches that can critically address emerging technologies and, admittedly, a certain methodological convenience. On the one hand, given that current analyses of ChatGPT tend to focus on its social risks and the technical characteristics of its functioning, Ihdean postphenomenology becomes a promising framework to identify what is at stake in our ways of relating to this technology. On the other hand, given that the arrival of ChatGPT apparently disrupts or unsettles concepts such as language, meaning or writing, Derrida's theory of writing appears as adequate. Yet, importantly, the fact that Mark Coeckelbergh and David J. Gunkel (2023) have recently published a Derridean interpretation of ChatGPT is both an incentive to inscribe my research in this discussion and an indication that there is some interest in reading Derrida to better understand concrete technologies.

That said, the research question which guides this investigation is: How could Jacques Derrida's works help us better understand ChatGPT compared to mainstream approaches in philosophy of technology and what would this show regarding the possibility of a Derridean philosophy of technology? To answer this two-fold question, I structure this thesis in three chapters, subsequently guided by the following sub-questions: What are the unresolved philosophical issues in the current discourse surrounding ChatGPT? How could Ihdean postphenomenology help us address such issues? How could Derrida's works help us address such issues too? The structure of each chapter is as follows.

The first chapter presents ChatGPT and analyses popular views in the current discourse to identify the unresolved issues that need be philosophically addressed. Firstly, I describe ChatGPT as a technology and mention numerous social concerns that this technology has raised. After indicating that there is a collective drive to better understand this technology and other AI systems, I suspect whether the abundant focus on ethical consequences presupposes a tacit instrumentalism insofar as it runs the risk of assuming that, once the social concerns are mitigated, ChatGPT might be unproblematic. Consequently, I call for a structural approach, which philosophically identifies the stakes regarding ChatGPT even if all its social risks were effectively mitigated. Secondly, I focus on the problem of anthropomorphization because it hinders an adequate understanding of this technology. After reviewing some metaphors AI experts have provided to debunk the anthropomorphic discourse through which ChatGPT is marketed, I am suspicious of a similar tacit instrumentalism insofar as there is the risk of believing that, provided we agree on the identity we ascribe to ChatGPT, this technology might be unproblematic. Claiming that the problem rather lies on how our interaction or ways of relating to ChatGPT affects our experience and ways of thinking, I call for an interactionist approach, which demands to focus our philosophical inquiries on how our relation to this technology affect our experience and understanding of the world, ourselves, and our human condition. Finally, I conclude that an unresolved issue in the discourse is that, despite the abundance of analyses regarding the ethical

consequences this technology might pose, we do not really know what is invariably at stake in our ways of relating to ChatGPT.

The second chapter carries out the postphenomenological analysis of ChatGPT in order to show the capacity of this approach to address the unresolved issue just mentioned. After presenting the main features of Ihdean postphenomenology, I determine the different types of relation one can establish with ChatGPT, identify how one's experience is transformed in each relation, highlight the disposition one adopts with regard to the technology as a consequence of such transformations, and derive a structural danger in the form of the relinquishment of a judgment from each relation. While this analysis helps us understand what occurs to human experience in different ways of relating to ChatGPT, why certain experiences might arise out of our interaction with the technology and why the relation might be sustained in time, the analysis cannot answer the question of why we so readily establish a relation with ChatGPT in the first place. After formulating the hypothesis that the common-sense, cultural understanding of the action one carries out with or through the technology and the assumption that the technology is capable of undertaking such an action might determine the relation, I focus on the question of writing to prepare the ground for the next chapter.

The third chapter presents a Derridean analysis of ChatGPT in order to explore whether it can answer what is fundamentally at stake in our ways of relating to ChatGPT. Since Derrida does not have a framework that we can apply but he is a theorist of writing, I take the question concerning our understanding of writing and the extent to which we can assume that ChatGPT writes, which Ihde leaves unanswered, to narrow the analysis. After presenting Derrida's project through his critique of metaphysics, I review Coeckelbergh and Gunkel's (2023) existing Derridean analysis of ChatGPT to see whether it already answers our questions. While their analysis does not answer our questions, I argue that, despite being correct, it runs the risk of utilizing deconstruction to justify ChatGPT insofar as they see ChatGPT as a practical deconstruction of logocentrism. I then hypothesise that interpreting ChatGPT as reinforcing logocentrism is also possible and that, given the political character of Derrida's critique of logocentrism, this interpretation is today unavoidable. To develop it, I first clarify the extent to which we can assume that ChatGPT writes through Derrida's notions of teletechnology and writing. Secondly, I attempt to characterize the specificity of ChatGPT's mode of writing to identify whether it shows any logocentric symptom. Replying in the affirmative, I finally attempt to determine what is at stake for the user's way of thinking by being exposed to this technology.

The overarching research question is answered in the conclusion. There, I account for the limitations of this investigation and enumerate potential paths for further research.

# 1 A review of the discourse surrounding ChatGPT: The need for a structural and interactionist analysis

Despite the massive popularity of ChatGPT, its arrival has unleashed a myriad ethical-political issues which urge us to better understand this technology. In this chapter, I provide a brief overview of the discourse surrounding ChatGPT in order to identify its unresolved philosophical issues. Firstly, I describe ChatGPT and lay out the general social concerns that have been raised against it. Highlighting a general interest in better understanding this technology, I show that a tacit technological instrumentalism frames the question of understanding from the outset, insofar as it is assumed that, once the ethical issues are mitigated, ChatGPT and other AI systems might be unproblematic. For this reason, I call for a structural analysis of ChatGPT, which philosophically identifies the stakes regarding ChatGPT even if all its social risks were mitigated and its effects were eventually deemed positive. Secondly, I focus on the problem of anthropomorphization insofar as it hinders an adequate understanding of ChatGPT. After reviewing some metaphors for ChatGPT proposed to debunk the anthropomorphization of the technology, I call into question their alleged debunking character. I argue that these critiques also hold on to a tacit technological instrumentalism insofar as they seemingly assume that, provided we agree on the identity we ascribe to ChatGPT, this technology might be potentially unproblematic. I claim that what is at issue is not a matter of choosing the meaning we ascribe to this technology but rather of understanding how our interaction with ChatGPT alters our experience and mentality. For this reason, I call for an interactionist analysis of ChatGPT, which should inquire into how our ways of relating to this technology affect our understanding of the world, ourselves, and our human condition. I finally conclude that an unresolved philosophical issue is that, beyond the numerous analyses focusing on the ethical consequences of this technology, we do not know much about what is structurally at stake in our ways of relating to ChatGPT. And this is also an ethical question.

## 1.1. What we know about ChatGPT

### *1.1.1. ChatGPT, an unprecedented chatbot*

ChatGPT was defined in its release announcement in November 2022 as an AI model "trained to follow an instruction in a prompt and provide a detailed response" through a "dialogue format [which] makes it possible for ChatGPT to answer follow up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests" (OpenAI, 2022). When visiting ChatGPT's promotional

website, it reads: “Let’s create. Let’s brainstorm. Let’s go. Let’s explore. Let’s collaborate. Let’s invent. Let’s design. Let’s chit-chat. Let’s discover.” Followed by the imperative: “get instant answers, find creative inspiration, and learn something new” (OpenAI, n.d.a). With such exhortations, OpenAI—the company behind ChatGPT—evokes the myriad uses one can make of ChatGPT, thus suggesting the latter’s potential to fulfil precisely the potential of its users. ChatGPT is not so much a technology created to aid humans to carry out specific linguistic activities, but rather it is envisioned—as Sam Altman (OpenAI’s CEO) says—as “an amplifier of human will” (ABC News, 2023). ChatGPT is advertised as a partner with which one cooperates to propel one’s creativity, knowledge and entertainment.

In the media, ChatGPT is described as a “highly capable chatbot” (Hughes, 2023) that “mimics speech” (Tiku, 2022). Chatbots have existed ever since 1966 thanks to the development of ELIZA, “a program which makes natural language conversation with a computer possible” where the user “has the illusion that he is the sole user of the computer complex, while in fact others may be ‘time-sharing’ the system with him” (Weizenbaum, 1966, p. 36). However, ChatGPT is seen as an unprecedented chatbot because it not only responds to our questions, but it can also “write jokes …, working computer code and college-level essays … and explain scientific concepts at multiple levels of difficulty” (Roose, 2022). One could enter into ChatGPT prompts like “What does Plato’s allegory of the cave mean?” or “Explain easily Einstein’s theory of relativity,” to “Summarize the following paper in 150 words” or “Make of the following bullet points a short story in the style of Clarice Lispector.” Thus, while all chatbots are computer programs which leverage artificial intelligence models and natural language processors to technologically simulate a human-like conversation to allow for human-machine communication, ChatGPT is singular in its apparent coherence, versatility, fluency and generative potential. ChatGPT appears to deliver correct information, about an extremely wide range of topics, assertively, quickly and directly, as well as to generate unique content from a mere prompt.

How does ChatGPT achieve this? From a technical perspective, ChatGPT is a computer software consisting of a fine-tuned version of a large language model (LLM) and a user interface (OpenAI, n.d.b). Whereas the user interface is the medium which allows the user to interact with the LLM without having any technical knowledge about it, an LLM is a type of neural network model which has been trained through a massive dataset of existing texts to predict the probability of the next set of characters—or what is called a “token” (OpenAI, n.d.b), which may coincide with any length between a character and a word—given a series of characters (OpenAI, 2023, p. 2).[1] In a very simplistic way, a language model completes a sequence of text that is not part of its training dataset

[1] What gives GPT models their performance quality is a central feature of their type of model architecture—the so-called “attention mechanism” (Vaswani et al., 2017), a feature which enables them to not only take into account the last token for the prediction of the next, but large sequences of preceding tokens (Wolfram, 2023).

by first identifying the tokens it extracted from its training dataset in the new text and then suggesting a likely token to follow given the probability of said suggested token following the given one in its own training dataset. As Wolfram (2023) clarifies: "when ChatGPT does something like write an essay what it's essentially doing is just asking over and over again 'given the text so far, what should the next word be?'—and each time adding a word." However, that ChatGPT is a fine-tuned version of an LLM means that the models on which the web app is based (GPT-3.5 or GPT-4) have been retrained in order to shape or align their possibilities in a favourable direction to its users such as filtering undesired uses of language (Ouyang et al., 2022).

While the fine-tuning of the raw LLM conditions the output formation, the output is displayed in a particular way to the user, so that, for example, as Weizenbaum (1966, p. 38) admits, the experience of a human-chatbot communication is based on creating the impression to each user that the chatbot is always individually available for them. Additionally, ChatGPT's user interface eliminates all those details which evince that one merely introduces inputs into a software which runs through code, thus not requiring us to cultivate special skills nor have any technical knowledge about the software to be able to use it. After creating an OpenAI account and accessing the ChatGPT web app, one is ready to type in a text box a prompt in the form of a question or a request which could be fulfilled in a disembodied way and through text. In a matter of seconds, ChatGPT returns a unique text which not only generally fulfils one's demands but also does not require any act of decipherment. However, the web app has been designed in such a way that the output is not given at once, as we would experience when introducing an input into a search engine, texting with a friend or visiting a regular website. Instead, the user perceives how the text is being constructed through time at a reading speed, so that they have the impression that the machine articulates or exteriorizes a linguistic expression for them.

### *1.1.2. The costs of ChatGPT and the need for "a better understanding"*

Shortly after the arrival of ChatGPT, tech journalists, pundits and scholars have identified many ethical-political concerns that derive from either the use, the functioning or the modelling of ChatGPT.[2] Insofar as ChatGPT generates text and code of a decent quality, there are worries that this technology will disrupt education (Stokel-Walker, 2022), automate jobs (Heikkilä, 2023; Castelvecchi, 2022), and subsequently transform economy (Rotman, 2023). Insofar as ChatGPT has been said to "hallucinate" (OpenAI, 2023, p. 10) for it can bind together words which make grammatically correct yet

[2] To be sure, this is just an overview of positions that have been publicly expressed and I do not subscribe to all of them. Also, the references which date from 2021 or earlier do not refer to ChatGPT but to large language models or AI systems in general.

untrue sentences, there are worries that it might direct society towards untruth and misinformation (Weidinger et al., 2022). Insofar as ChatGPT is pre-trained on large datasets of public data such as internet data, there are concerns that it might foster discrimination and radicalisation (Bender et al., 2021; Chan, 2022). Insofar as ChatGPT offers direct, coherent answers with fluency and no hesitation, there are concerns of blind reliability on the outputs it returns (Bender et al., 2021; Montemayor, 2021). Insofar as we do not feel judged when talking to ChatGPT, there are worries that it promotes individualism and, importantly, forgetfulness that a corporation profits from our confessions (Gertz, 2023a). Insofar as ChatGPT generates new text based on already existing ones, there are concerns that it diminishes academic integrity and increases plagiarism (Eke, 2023; Brainard, 2023; Dehouche, 2021). Insofar as ChatGPT exploits AI's rhetoric of anthropomorphism, there are worries that we not only foster a misunderstanding of the technology, thus boosting the so-called AI hype (Bender and Koller, 2020) and influencing users to uncritically trust ChatGPT's outputs (Bender and Shah, 2022), but we also bestow agency upon the technology to a degree that deprives us of our autonomy (Watson, 2019). Insofar as ChatGPT is advertised as an agent with which we cooperate to amplify our creativity and knowledge, there are concerns that this cooperation is actually an abdication from and a deterioration of the human capacities of imagining and thinking (Alombert, 2023a, 2023c). Insofar as ChatGPT requires additional human labelling to "clean" the violent and traumatic language that can be carried from the training dataset and this is currently done through the exploitation of workers from the global south (Perrigo, 2023; Rowe, 2023), there are reasons to be concerned about whether AI alignment is not but just a distraction to cover up the immorality of corporations. Insofar as ChatGPT's digitality is supported by a very material infrastructure which requires energy sources, there are worries about its environmental impacts (Bender et al. 2021; Shaji George et al., 2023). Insofar as AI systems seem to be increasingly capable of performing actions that are thought to require cognition in ways that no human could possibly undertake, there are concerns that these systems are indicative of a superhuman intelligence which poses an existential risk—the possibility of a catastrophe for human existence (Bostrom, 2014; Corfield, 2023). And many other worries which I might have overlooked or still remain unknown.

In light of some of these concerns, the Future of Life Institute (FLI) has demanded, through an open letter (FLI, 2023a) and a white paper (FLI, 2023b), an "AI Summer" (FLI, 2023a)—a request to halt AI development of systems that are more powerful than GPT-4 in order to gain time to "*better understand these systems* [emphasis added], … reflect on their ethical, social, and safety implications, and to ensure that AI is developed and used in a responsible manner" (FLI, 2023b, p. 4). This demand, according to the FLI (2023a), "does *not* mean a pause on AI development in general, [but] merely a stepping back from the dangerous race to ever-larger unpredictable black-box models with emergent capabilities." However, even if the FLI (2023b) is aware that "systems may perform in ways that their

developers had not anticipated or malfunction when placed in a different context" (p. 14), the FLI (2023a) is convinced that "[p]owerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable." In other words, not only it is thought that technologies have a certain degree of "autonomy" which might unleash unforeseen effects, but it is also believed that technologies and the contexts in which they can function can be effectively controlled, so that AI systems become instruments through which our good intentions are fulfilled. This perspective might run the risk of presupposing a tacit technological instrumentalism—the view that a technology is just a tool or a means to an end—insofar as AI systems could be deemed unproblematic—and thus a mere tool—once their ethical consequences are mitigated. With this, I am not trying to argue for a philosophical analysis which neglects the importance of addressing the abovementioned ethical concerns. Rather, I am trying to think how we could prevent ourselves from clinging to an uncritical and naively optimistic stance after having dealt with such concerns. In this sense, in addition to conventional ethical analysis, I call for a structural approach, which adopts the view that there might be something at stake even if all its social risks were effectively solved and its effects were positively perceived. In short, by structural approach I mean a way of addressing the phenomenon of ChatGPT without presupposing that, if we were to be in the hypothetical scenario that ChatGPT's ethical consequences are eliminated, this technology would be unproblematic and a desirable tool to use. If we assume this, the quest for a better understanding of ChatGPT is limited in advance, insofar as we would be conforming to what we think it is beneficial, positive or convenient without calling into question that which leads us to believe this is so.

Philosophers of technology—sensitive to the problematic of instrumentalism—have intervened in the public debate around ChatGPT to suggest the need of an understanding of AI systems that exceeds the technical (Y. Hui, 2023), that does not address ChatGPT as a mere tool (Alombert, 2023b), that includes a critical review of our way of thinking which legitimizes the appearance of these technologies and sustains our use of them (Gertz, 2023b), and that addresses the character of our psychological disposition which leads us to anthropomorphize ChatGPT in the first place (Babich, 2023). Before we align with any of these views, we must ask ourselves what it means to better understand ChatGPT in light of the facts that (1) our exposure to its functioning entices us to believe that it has faculties that are at least similar to ours, so that we are confident enough to entrust signifying actions to it, and (2) that the technology is marketed in anthropomorphic terms. Because of this, one might argue that the problem of anthropomorphization hinders our understanding of this technology. Thus, in what follows I will evaluate some reactions offered by AI experts to resist the anthropomorphization of ChatGPT in order to identify their unresolved issues.

## 1.2. Metaphors for ChatGPT to debunk anthropomorphization

### *1.2.1. The anthropomorphization of ChatGPT*

Anthropomorphization is understood as the human tendency to ascribe properties—thought to belong to the human—to non-human entities (Proudfoot, 2011). This human tendency was paradigmatically brought to the fore by Alan Turing's (1950) well-known "imitation game" (p. 433), a thought experiment where one tries to guess the identity of two entities—where at least one is human and both are capable of producing significations—by asking questions and analysing their written outcomes. The fact that Turing (1950) begins his article by inviting us to consider the question "can machines think?" (p. 433) might lead many to believe that such a game is exclusively a test to determine whether a machine thinks. In this sense, if one fails in their guessing, it means that the non-human has overcome a perceptual threshold of humanness, so that it must possess certain human properties. However, the Turing test might also be seen as a thought experiment to reflect on the easiness with which we are fooled and, as such, it helps us be aware of our perceptual limitations and to acknowledge our limits of intersubjective recognition or our vulnerability to the effects that sensory impressions have on our beliefs. In light of how easy it seems for something to pass the Turing test, one could develop a radical scepticism and realize that, in the presence of an outcome which is detached from its producer, the possibility of being fooled is always there.

Be that as it may, the very terms of "artificial intelligence" or "machine learning" are indicative of a pervasive anthropomorphic narrative in their fields, but it has been intensified with the arrival of ChatGPT given its abilities of processing natural language and the way it displays its outputs. This intensification is visible in Sam Altman's (ABC News, 2023) discourse about how ChatGPT should be understood. Thanks to ChatGPT, says Altman, "[w]e can all have an incredible educator in our pocket that is customized for us, that helps us learn, that helps us do what we want" (ABC News, 2023). This figure of an "incredible educator in our pocket" suggests the idea of a portable superior mind. An educator is someone who has something to teach, so that they must know more than their pupil. But an educator is not someone who knows it all, for they may err. Thus, given that ChatGPT can be an incredible educator for *all of us*, then we all must perceive it as being ahead of us, as always knowing more than any of us ever could, even if it sometimes makes mistakes. It is in this sense that the educator is "incredible," and therefore superior. Geoffrey Hinton, a research pioneer on deep learning, also supports this view: "GPT-4 knows hundreds of times more than any one person does" (Heaven, 2023). And technology journalists have argued that ChatGPT's "power lies in its speed and understanding of complicated matters" (Hughes, 2023) or that this technology helps "normal people … understand the world" (Harwell et al., 2022). From this it follows that ChatGPT is believed to reason. As Altman says: "the right way to think of the models that we create is a reasoning engine,

not a fact database. … [For] what we want them to do is something closer to the ability to reason, not to memorize" (ABC News, 2023). ChatGPT could not be an incredible educator were it only to provide us with descriptions (i.e., facts or sources of information). This is what search engines do—they increase the speed of our access to information while determining the information we read. In contrast, ChatGPT prescribes, through probabilistic calculations, certain relations among pieces of the elements that constituting its training dataset. In this sense, Altman markets ChatGPT as an incredible educator to spread the idea that this technology provides interpretations of the world of its training dataset, which, due to its size, some might too hastily take it as our world. With this, Altman fosters the idea that ChatGPT has access to truth, that ChatGPT has exited Plato's cave and reached the realm of ideas and it has now come back to the darkness of the cave to show us what sunlight looks like.

### *1.2.2. Non-anthropomorphic metaphors for ChatGPT*

According to Bender and Koller (2020), anthropomorphic lexicon is not only present in the media but also in scholarly discourses on technical explanations of LLMs. For them, this "imprudent use of terminology" is problematic because "it feeds AI hype in the popular press" (Bender and Koller, 2020, p. 5186), a hype which entails the spreading of a misunderstanding of systems like ChatGPT and entices people to idealize and overestimate them. Against this, Bender and Koller (2020) attempt to demonstrate that LLMs do not have access to meaning and thereby they do not have understanding. Yet, elsewhere, Bender et al. (2021) have used the metaphor of the "stochastic parrot" (p. 617) to defend the idea that large language models do not think, reason, understand, nor learn, for they exclusively manipulate linguistic form and not meaning. An LLM is, for them, a "stochastic parrot," namely, "a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning" (Bender et al., 2021, p. 617).[3] This metaphor thus expresses the idea that language models repeat, like a parrot, parts of the data with which they have been trained, while introducing difference in the repetition given their probabilistic (i.e., stochastic) rearrangement of extant data.

But importantly, the metaphor conveys that the probabilistic functioning of large language models like ChatGPT excludes all possibility of intelligence, reason, understanding, imagination or creativity, "[c]ontrary to how it may seem when we observe [their] output" (Bender et al., 2021, pp. 616-617), that is, *even if* their human or animate similarity fools us in thinking that they think. Thus, by using the term 'parrot,' Bender et al. maintain ChatGPT's property of animation while strongly

[3] The fact that Bender et al. (2021) use the verb "to observe" in the quoted passage is indicative of the difficulty to talk about these technologies without using anthropomorphic language.

rejecting its identification with the human. Understanding ChatGPT as a stochastic parrot thus means that, even if we perceive its output as coherent, the model does not know whether nor why it is so or otherwise. It is we who ascribe meaning to the text the model outputs. We think ChatGPT's generated text is fluent, coherent, appropriate, correct or accurate because we bestow sense upon it in our interpretation. For Bender et al. (2021, p. 616), this is because we implicitly assume that behind every human-like discourse there is a human with thoughts to share and communicative intents to be deciphered.

For Bender et al. (2021, p. 616), what is at stake is the idea that the meaning obtained from ChatGPT is an illusion that is produced by the user alone, given that ChatGPT cannot have communicative intent. In other words, the problem is that our subjective perception of ChatGPT's functioning does not correspond to what ChatGPT actually does, and that such functioning leads us to apprehend its output as though it came from another *person*. Bender and Shah (2022) argue that the consequence of this illusion is an increasing uncritical trust of large language models, for it is hard to not project the image of a human or of a mind behind the text we perceive. Subsequently, blind reliability on large language models entails that the user forgets the model's possibility of misinformation. For this reason, they think it is urgent to disseminate information to the larger public about how these technologies work in a language that does not conflate what the user immediately experiences with the machine's actual mode of working, that is, by avoiding words like 'know,' 'understand,' 'read,' 'say,' 'write,' etc.

After Bender et al.'s (2021) attempt to debunk positions like Altman's metaphor of the reasoning engine, other AI experts have proposed alternative metaphors to evoke a more accurate description of ChatGPT. Luciano Floridi (2023, p. 3) states that the metaphor of the stochastic parrot is not fully appropriate because animals have intelligence and, importantly, because large language models do not simply repeat the existing but bring novelty through probabilistic rearrangements. Consequently, Floridi (2023) proposes the idea of the "autocomplete function of a search engine" (p. 3), that is, the prediction of what one might want to search for given a sequence of words. However, François Chollet (2023a, 2023b) dismisses Floridi's analogy because a large language model not only retrieves predictive patterns like an autocomplete but also stored facts and rearrangements of these. To emphasize this, Chollet (2023b) proposes the idea of an "interpolative database and program store, with a natural language interface." An "interpolative database" conveys the idea of a digital store of data which can retrieve exact stored data, like an ordinary database, but also rearrange known data thanks to the mathematical identification of patterns in the dataset used for training (i.e., interpolation). Yet, although this analogy provides an understanding of ChatGPT detached from all animation, it remains quite technical. In response to Chollet's dismissal of Floridi's autocomplete and against Bender et al.'s (2021) stochastic parrot, Federico Cabitza (2023b) proposes "the metaphor of the

'spiritless ([i.e.,] lifeless, inanimate) séance,' where it's our ancestors speaking, and esp[ecially] it's us, or the living, who give it a meaning, a coherence, a spirit!" This metaphor of the "spiritless séance" foregrounds the idea that ChatGPT is not animate but the medium that embodies "the voice of those who aren't there, or are no longer there" (Cabitza, 2023a). ChatGPT does not say anything, ChatGPT makes re-appear what had already been said in a new way. This metaphor allows us to see ChatGPT as the means through which something else is made available. But this "something else" is not the mind of an incredible educator, but rather a reflection of our own footprints which we have been leaving behind on the internet. Aligning with this view, David J. Gunkel (2023) writes that ChatGPT "is echoing back to us our own voice. It is speaking from our past."

### 1.3. The tacit instrumentalism of ~~debunking~~ performative metaphors

One might think that the above-mentioned metaphors play a merely descriptive role which aims to demystify what ChatGPT is. However, we could understand them as being speech acts of the type that John L. Austin (1962) calls "illocutionary" (p. 99) or "performative" (p. 6) because they create that which they express, they perform an act *in* their utterance. For instance, christening is an example of a performative (Austin, 1962, pp. 44, 116)—a name is contractually assigned to a thing or a person by the mere act of uttering it in a specific context that bestows meaning upon it. The metaphors we have seen above may thus be called performative metaphors insofar as they constitute an attempt to christen ChatGPT and subsequently condition how we as users relate to it. These metaphors *make* what ChatGPT is, create ChatGPT's identity. By trying to privilege a sense of ChatGPT among others, it is believed that our use of such a technology will change for good because we will not fall prey to its enchantments. Then, claiming ourselves sovereign with respect to the object which we face, it is thought that we can control its effects, that a good use of ChatGPT can be ensured. In other words, this attempt shows the underlying assumption that technologies are effectively determinable or controllable through the language we use to refer to them.

As we have seen, the FLI, with a similar goal in mind, backs the implementation of governmental regulation to steer the course of AI development. While they do not focus on the meaning of the technology, it is through the compliance with the regulations that one meaning would be guaranteed. That is, institutional regulation aims to fix the social context so that ChatGPT can adopt a desirable meaning. In this sense, both techniques (i.e., governmental regulations and the use of performative metaphors) are at the service of a form of social determinism with regard to technology, which I here understand not so much as the view that technologies are the product of social values and market forces but as the belief that we can culturally restrict how technologies are to behave in society. This narrower form of social determinism implies, I argue, a reinforcement of an instrumentalism—the view that technologies are mere means to attain desired ends—because, even if one

initially acknowledges the possibility that technologies can bring about unforeseen effects, if one then believes that the meaning of a context or a technology can be secured, then a technology is ideally envisioned as a means to fulfil a set of desired intentions—whether of a corporation or of a state. This is not to say that anthropomorphizing a technology implies sustaining an instrumentalist position. Rather, my point here is to suspect that, any view that believes in the possibility of fixing the identity of a technology—whether in anthropomorphic or non-anthropomorphic terms—is stuck in an object-centred perspective which presupposes from the outset what the technology is. Risking this formulation, the instrumentalism is projected as an ideal—in the same way that is uncritical to assume that a technology could be seen as a tool once all ethical consequences are addressed, it is equally uncritical to assume that, provided we agree on the terminology we use to refer to the technology, we will be able to establish an unproblematic instrumental relation with the technology. Again, with this, I do not intend to argue against the need of regulations to mitigate the risks contemporary technologies unleash or of metaphors to divulge how a technology works by illustrating complex technical explanations. The problem is rather in thinking that this alone might suffice and in how the problem of anthropomorphization is addressed, that is, as a question of misunderstanding and of misattribution of properties.

Considering the effects that the combination of ChatGPT's possibilities of text generation and the way of displaying ChatGPT's outcomes may have on the user, one could argue that the user is *prompted* to think that this technology is a competent assistant from which one can benefit by delegating one's acts of signification. This is exacerbated by the feature that ChatGPT's artificially produced texts are practically indistinguishable from human texts, thus allowing the user to appropriate the technology's output as one's own. From this perspective, to anthropomorphize ChatGPT is not only to argue that the technology reasons because we cannot distinguish its outputs from human texts; ChatGPT is already anthropomorphized when "we implicitly grant it a degree of agency that not only overstates its true abilities, but robs us of our own autonomy" (Watson, 2019, p. 434). In other words, the natural disposition of a user to delegate an interpretive or writing activity to ChatGPT already assumes the tacit belief that it is the machine that *should* do such activities on their behalf. In this sense, the problem of anthropomorphization is not so much a problem of ascribing ontological human properties to non-human beings, but a problem of delegation. Why do we delegate basic human activities such as reading or writing to these technologies? How does ChatGPT transform users' experience so that they come to believe such delegation may be emancipatory?

To address these questions, we must abandon the object-centred perspective implied by the aforementioned metaphors, which is too focused on the discussion of whether ChatGPT is alive or not, and inquire into what is at stake in our interaction with this technology. We do not need a doctor who simply diagnoses us with an illness due to certain facts and subsequently prescribe us a pill that

mitigates the illness; we need to understand why those facts manifest as they do in order to have reasons to deliberate whether we should, can and want to change our way of living, even if we take the pill. In this illustration, the performative metaphors are the pill, the social risks are the illness, the AI experts are the doctor, and the reasons of the facts that indicate the illness are what we should try to philosophically characterize through an interactionist perspective—a perspective which is concerned with how technologies alter what we think we know, transform the way we think about reality, shape the way we experience the world, modify our conception of moral goodness, etc. This philosophical enterprise seems to be suppressed if we believe in the transparency of artefacts, in the idea that, even if technologies might cause undesirable social effects, they do not shape how we think nor act; and, additionally, such suppression might be seen as an avoidance to understand ourselves or our contemporary human condition. In this regard, what motivates a better understanding of ChatGPT, philosophically, is the question concerning the effects that our ways of relating to this technology pose to human experience and mentality, even if all ethical consequences were effectively mitigated.

Hence, while the metaphors we have considered might surely be helpful to make accessible complex explanations of ChatGPT's technical functioning, instead of constituting a debunking gesture, they rather invite us to look the other way—to turn our back on the problem of why and how the use we make of technologies can lead to experiences of accomplishment, realization, self-determination, satisfaction, fruition or fulfilment *as well as* experiences of estrangement, dispossession, dependence, dissatisfaction, disruption, unsettlement or alienation. Without a quest for the reasons why these varied experiences may take place in and through our use of ChatGPT, we run the risk of attaining a state of good conscience once the results of our ethical analyses are positive.

## 1.4. Conclusion

In this chapter, we have covered some of the main views in the current debate around ChatGPT in order to identify unresolved philosophical issues in it. In light of the need for a better understanding, we have put our focus on the problem of exclusively focusing on the admissibility of the social effects ChatGPT poses and on the problem of anthropomorphization. In both cases, we have identified a tacit technological instrumentalism. Regarding the former, it is seemingly assumed that a context is determinable and that, as a consequence, the technology could play a desirable instrumental role. Regarding the latter, it is apparently assumed that the language used to refer to a technology creates its identity, thus determining its contextual role, one that might be desirably instrumental. As a response, we have called for the need of a structural and interactionist approach which, focusing on the relation between us and ChatGPT, attempts to identify what is at stake in our ways of relating to ChatGPT even if all ethical consequences were effectively addressed. For this reason, we can conclude that the unresolved philosophical issue in the debate surrounding ChatGPT is the lack of analyses which can

address the question “What is structurally at stake in our ways of relating to ChatGPT?”. In the following chapters, we will see whether and how Ihdean postphenomenology and Derridean deconstruction could address this question.

# 2 A postphenomenological analysis of ChatGPT: From structural dangers to the question of writing

In the previous chapter, we have seen that, despite all the literature about the ethical consequences that ChatGPT unleashes, we know less about what is structurally at stake in our ways of relating to this technology. In this chapter, I carry out a postphenomenological analysis of ChatGPT to show how Ihde's approach could address such an issue. After clarifying the role Ihdean postphenomenology plays in this investigation and presenting the main features of this framework, the analysis shows that ChatGPT can become an embodiment, a hermeneutic, an alterity and/or a background technology depending on the use we make of it, and reveals in each type of relation a structural danger in the form of the relinquishment of a judgment. While this answers the question concerning what is structurally at stake in our ways of relating to ChatGPT, a limitation of this analysis is that it cannot explain why we so readily relate to ChatGPT in the first place. In other words, Ihdean postphenomenology helps us better understand and be aware of what takes place in our relations with ChatGPT, but cannot explain why there is a relation. I then pose the hypothesis that our common-sense understanding of the action we carry out with or through the technology might be the condition of possibility of the relation. This entails that the technology would not be the single element to mediate and transform our experience; a consequence that urges us to pursue a new battery of questions concerning our common-sense understandings of the actions we carry out with or through ChatGPT and the extent to which we could assume that the technology is capable of undertaking such actions.

## 2.1. Ihdean postphenomenology

As an initial caveat, it should be noted that the reason why we pass through postphenomenology before appealing to Derrida is to show the capacity of a conventional and well-established approach in philosophy of technology to address the unresolved issue we have identified in the previous chapter. This will prove to be useful to ultimately asses the import of our upcoming Derridean analysis of ChatGPT. Insofar as this investigation is motivated by the exploration of what a Derridean philosophy of technology could look like, this strategic and instrumental movement of bringing up postphenomenology first seems unavoidable. However, in the interest of having enough space for our interpretation and application of Derrida's works, I will not provide an in-depth postphenomenological analysis

of ChatGPT, but rather a general analysis that shows the capacity of this approach as well as a consideration of its limitations.

If postphenomenology is alluring to address the philosophical analysis that is missing in the current discourse surrounding ChatGPT, it is due to the central theses regarding the non-neutrality of technologies (Ihde, 1990, p. 49) and "multistability" (Ihde, 1990, p. 144), which ground a phenomenology of human-technology relations which attempts to describe the structural features of human experience when the latter is ambivalently transformed by the inherent non-neutrality of technologies (Ihde, 1990, p. 71). On the one hand, the thesis of non-neutrality consists in that technologies provide the user with a temporary perceptual advantage which simultaneously forces them to give up (something of) their natural way of being in the world. This means that any interaction with a technology necessarily entails a transformation of our experience in the form that there is something which reveals to our awareness and something which conceals from it. On the other hand, the thesis of multistability implies that technologies are inherently ambiguous because artifacts only become technologies through the use one makes of them and because one can ascribe the same function to different objects (Ihde, 1990, p. 139). The implication of this is not only that we cannot presuppose what a technological artefact is without taking into consideration the way in which a user relates to it (Ihde, 1990, p. 70), but also that there might not be only one ChatGPT but many—provided that we establish different types of relation with this technology.

This brief contextualization of Ihdean postphenomenology should suffice to show that its approach of human-technology relations could help us better understand ChatGPT compared to the existing views in the debate by shying away from the latter's object-centred perspective and tacit instrumentalism to focus on the reasons why certain experiences emerge out of each way of relating to ChatGPT (Ihde, 1979, p. xxvii). As a result, I will attempt to identify whether there is any risk following from the structural transformations of our experience in and through our uses of ChatGPT.

### 2.2. A postphenomenological analysis of ChatGPT

Ihde's (1990) phenomenology of human-technology relations consists of four types of relations: embodiment relations, which occur when one experiences the world thanks to a technology while the technology is almost invisible to our experience; hermeneutic relations, which take place when one has an experience of the world through the interpretation of a technology's output; alterity relations, which happen when one experiences a technology while reality hides from our awareness; and background relations, which manifest when technologies merge with our surroundings so that we have an experience of the world supported by such technologies without necessarily interacting with them. In which ways could one relate to ChatGPT?

### *2.2.1. An embodiment relation with ChatGPT*

In embodiment relations, a technology becomes part of us, part of our experiencing and not of what we experience. We take the technology *into* our experience by perceiving the world *through* it (Ihde, 1990, p. 72), and thus the technology becomes almost transparent or a "quasi-me" (Ihde, 1990, p. 107). To emphasize this coupling, Ihde (1990, p. 86) formalizes embodiment relations as:

$$(\text{I-technology}) \longrightarrow \text{world}$$

While this type of relation often takes place with optical technologies (Ihde, 1990, p. 72), the example of a student who decides to produce a college essay through this technology seems to fit with this type of relation. Whether the student uses ChatGPT to polish parts of a text they have composed or to produce an essay from scratch, the student embodies ChatGPT insofar as it is them who submit the essay with their signature. In the same way that we, short-sighted people, claim that it is us who see the world rather than the unit "my-eyeglasses-and-I," the student in our example claims it is them who have produced the text rather than the unit "ChatGPT-and-I." The structural dynamic of embodiment relations corresponds to the simultaneous amplification and reduction of our experience in the partial unawareness of the technology enabling one's experience (Ihde, 1990, p. 76; Gertz, 2018, p. 50). In our example, this dynamic shows the following: that the student's experience is amplified in the sense that they can now produce potentially meaningful text without having prior knowledge about the content being expressed and that they can now make others believe that one is able to write on a wide variety of topics; that the student's experience is simultaneously reduced because they could not defend what they have signed as theirs in the absence of ChatGPT and because they have not contributed singularly to the act of signification; and that ChatGPT partially withdraws from the student's awareness because the technology is always available for them—provided that the student carries with them a device with internet connection—and because ChatGPT does not leave a recognizable mark in the produced text that would hinder the student's appropriation of ChatGPT's outputs.

This amplification/reduction dynamic also affects the disposition of the user with regard to the technology. This is because "what is *revealed* [or amplified] is what excites" and "what is concealed [or reduced] may be forgotten" (Ihde, 1990, p. 78). Regarding embodiment technologies, Gertz (2018) writes that they "have the ability to *empower* us, but they also have the ability to *belittle* us, revealing when they break how dependent on technologies we have become" (p. 53). Turning to our example, the student might be dazzled by an experience of empowerment—for they can do things they could not before—while they might forget the experience of belittlement with regard to their faculties of thinking and imagining (Farrokhnia et al., 2023, p. 9). What this additional fascination/forgetfulness dynamic adds to the analysis is that the prevalence of feeling empowered might

entice the student to establish an unreflective relation of dependence with the technology (Paul et al., 2023, p. 1217). In short, the student might take ChatGPT as a substitute for their brain.

### *2.2.2. A hermeneutic relation with ChatGPT*

In hermeneutic relations, a technology refers to or points to an aspect of the world, transforming the latter into something we can interpret (Ihde, 1990, p. 92). We thus *read* what hermeneutic technologies provide us with and, thanks to our interpretation, we gain access to a world that was unavailable before. For instance, the hotness and coldness of the world may become available through our interpretation of a thermometer's bar or dial (Ihde, 1990, p. 85). In order to emphasize that, in hermeneutic relations, it is not the "I" who merges with the technology but rather the world, Ihde (1990, p. 87) formalizes these relations as follows:

$$\text{I} \longrightarrow (\text{technology-world})$$

Another paradigmatic example of a hermeneutic relation is reading a text. While you are reading these lines, you most likely gain access to "my world," to the ideas that I decided to express in English language and Times New Roman. We thus establish a hermeneutic relation with ChatGPT when we pose our doubts and curiosities to be solved in a click—for some might say we gain knowledge of an aspect of the world through our interpretation of ChatGPT's output. The structural dynamic of hermeneutic relations corresponds to the simultaneous "*presence* of the world we are given access to and the *absence* of our awareness of the technology mediating that access" (Gertz, 2018, p. 51). In the same way that you, when reading my text, lose awareness of the shape of the characters which mediate your access to my thoughts, we lose awareness of the fact that ChatGPT's texts are probabilistically retrieved. The user thus tends to equate its outputs to truth because they forget the way in which such outputs are produced. For instance, some argue that ChatGPT allows us to acquire knowledge "in seconds" thanks to "its speed and understanding of complicated matters" (Hughes, 2023).

Now, the implication of this ambivalent dynamic is that "hermeneutic technologies have the ability to *enlighten* us, but they also have the ability to *betray* us, revealing when they misinform us how much faith we put into technologies" (Gertz, 2018, p. 53). This very text, as the technology giving you access to my world, might enlighten you because it allows you to know things you did not know before. Given that ChatGPT generally outputs coherent responses with relative immediacy, we might be astonished at the pace at which we obtain information and potentially learn without striving to search for sources, interpret and understand them. It is this that enlightens us. But this enlightening experience demands faith from our side. It requires you to both trust what I write here or what ChatGPT outputs; and, insofar as trusting or believing does not imply evidence, there always exists the possibility of a betrayal. Just like ChatGPT is said to "hallucinate" (OpenAI, 2023, p. 10) because

it can string together pieces of existing text that form meaningful sentences which are untrue, my interpretation of Ihdean postphenomenology could be flawed. Precisely because the enlightening experience fascinates you, it makes you forget that there is a possibility of betrayal, and thus inclines you to trust technologies as if they had direct access to truth.

### *2.2.3. An alterity relation with ChatGPT*

In alterity relations, a technology becomes "an object of *fascination*" (Ihde, 1990, p. 100) because it acts as an animated being or as a "quasi-other" (Ihde, 1990, p. 100). The technology is thus our object of experience and nothing is made present through it. To emphasize this opacity of alterity technologies, Ihde (1990, p. 107) formalizes alterity relations as:

$$I \longrightarrow \text{technology-(-world)}$$

In alterity relations, we often fall prey to the technological illusion of human or animal likeness. We perceive a technology as if it had life of its own at the same time that we lose awareness of the fact that such liveliness is technologically created and of the world. Thus, the structural dynamic of alterity relations takes the form of a simultaneous "*fascination* with the liveliness of the technology and *obliviousness* with regard to the world" (Gertz, 2018, p. 52).[4] While the example of playing a game against the machine is a paradigmatic example (Ihde, 1990, pp. 100-101), the situation of chit-chatting with ChatGPT fits well with this experience of quasi-otherness. Let us take Kevin Roose's (2023) experience of talking to Bing—another LLM of ChatGPT's ilk. Despite knowing that Bing does not think nor is a human other, Roose (2023) reports that he was inevitably fascinated by the uncanniness of Bing's text, by the illusion that Bing *seemed* to have communication intent, to the point that Roose got absorbed for hours into the human-chatbot conversation and his immediate surroundings withdrew from his awareness. This might have the implication that "alterity technologies have the ability to entertain us, but they also have the ability to enrage us, revealing when they impede us how much emotional investment we have put into technologies" (Gertz, 2018, p. 53). For instance, insofar as ChatGPT tends to function obediently and nonjudgmentally, Roose might have been absorbed into the chit-chatting while forgetting about the emotions he was projecting onto the technology, which became evident when Bing started to output bizarre comments that interrupted Roose's implication.

[4] Here, the structural dynamic coincides with the fascination/forgetfulness effect because alterity technologies do not give us access to an aspect of the world. Therefore, the excitement with what is revealed occurs with regard to the technology itself. Cf. Ihde (1990, pp. 100-102).

### *2.2.4. A background relation with ChatGPT*

In background relations, a technology merges with our environment and "function[s] as a barely detectable background presence" (Ihde, 1990, p. 108). The technology adopts an infrastructural role; it enables and conditions our field of experience but goes unnoticed insofar it is taken for granted. In this sense, a background technology is an "absent presence" (Ihde, 1990, p. 112). Ihde does not formalize background relations, but Gertz (2018, p. 52) has proposed the following formula to emphasize the technology's ability to operate from without our field of attention:

$$I \longrightarrow \text{world-(-technology)}$$

While a paradigmatic example of a background technology is a thermostat (Ihde, 1990, p. 108), we could entertain the hypothetical scenario that ChatGPT becomes a primary source of content generation and sense formation—especially considering that ChatGPT may be already being used by ad agencies for copywriting and graphic design (Rahr, 2023; Saini, 2023), by students, teachers and workers for outsourcing writing assignments (Ellis, 2023; Appiah, 2023) or by Hollywood studios for writing film scripts (McNamee, 2023). Furthermore, given that it is not easy to detect content produced through ChatGPT, the invisible character of the technology would be easily secured. The structural dynamic in background relations corresponds to the simultaneous "*absent presence* of the technology and the *present absence* of the world" (Gertz, 2018, p. 52). If we translate this dynamic to ChatGPT, and accepting that the omnipresence of ChatGPT would occur by privileging capitalist values of efficiency and productivity (Gertz, 2023b), what would be revealed is ChatGPT's inconspicuous enhancement of our surroundings and what would be concealed is precisely the awareness of what it takes to "enhance" such surroundings. Concerning to the latter, this would not only refer to issues such as environmental costs (Bender et al., 2021, pp. 612-613; Shaji George et al., 2023), work exploitation (Perrigo, 2023) or inevitable biases in algorithmic modelling (Ferrara, 2023), but also that we as humans might be "*becoming more like machines*" (Gertz, 2023b). The implication of this dynamic is that "background technologies have the ability to *enliven* us, but they also have the ability to *incapacitate* us, revealing when they malfunction how reliant on technologies we have become" (Gertz, 2018, p. 53). Due to our fascination with the convenience of a ChatGPT-world and the forgetfulness towards what it takes for both the world to be a ChatGPT-world and for us to deem such a world convenient, we would be naturalizing such a ChatGPT-world, thus taking *a* technological world—one of the many possible—as *the* world.

### *2.2.5. Conclusion of the postphenomenological analysis*

We have seen that the structural dynamic of each relation has an effect on the disposition of the user with regard to ChatGPT in the form of a simultaneous fascination and forgetfulness. It is from this forgetfulness that I estimate possible to derive a structural danger which takes the form of the perpetuation of an unreflective stance with regard to the implications of carrying out an action with or through a technology. In other words, given that the necessary reduction of our experience is forgotten due to the fascination with the magnification of our experience, we are, at least momentarily, dissuaded from deliberating about the hidden side of the relation's effects. In this sense, this structural danger could be formulated in terms of the relinquishment of a judgment. For instance, when we embody ChatGPT, there is the structural danger that, insofar as the fascination with our experience of empowerment makes us develop a dependency towards the technology and the forgetfulness of our belittlement sustains it, we simultaneously relinquish the judgment of whether writing should be delegated to the technology. In the case of a hermeneutic relation with ChatGPT, there is the structural danger that, insofar as we gradually rely on the technology as a source of knowledge due to our feeling of enlightenment and our blindness regarding the possibility of being deceived, we relinquish the judgment concerning the extent to which it is legitimate to put faith into a technology which probabilistically groups linguistic characters. In the case of an alterity relation with ChatGPT, there is the structural danger that, insofar as we are captivated by the human likeness of the technology and we forget both the world and the emotions we project onto ChatGPT, we relinquish the judgment concerning the worthiness of our presumably individual interaction with ChatGPT (Gertz, 2023a). Finally, in the case of ChatGPT as a background technology, there is the structural danger that, insofar as we are amazed by the convenience of our surroundings according to an ideal of efficiency and we forget about what it takes to have such material environment and to deem it convenient, we relinquish the judgment of whether the creative activities the technology carries out on our behalf are intrinsically valuable for living a good life (Rubin, 2004, p. 16).

## 2.3. Limitation of the postphenomenological analysis

Ihdean postphenomenology has helped us characterize how we relate to ChatGPT, account for what occurs in such relations, and why the relation is sustained. Yet, one could now rightly wonder why there is a relation in the first place. Why have we so quickly accepted using ChatGPT instead of offering resistance to it? Why are we so desirous to delegate our capabilities of signification and interpretation to ChatGPT? We could probably address these questions through sociological analyses of the values which ground our capitalist societies, through anthropological analyses of the technological dependence of our contemporary human condition, through empirical studies about the

psychological effects ChatGPT's marketing lures and design strategies have on users, etc. Yet, philosophically, we could pose the hypothesis that such speed of acceptance has to do with our commonsense understanding of the action that we carry out with or through the technology, which surely precedes the arrival of ChatGPT and is simultaneously exacerbated by it. If a user is so quickly convinced that it does not matter who writes, whether they or ChatGPT, it is not only because the quality of our writings has adopted a completely machinic and impersonal character, but especially because of the way in which we understand writing; an understanding that, in the first place, assumes that ChatGPT writes. To be sure, this does not mean that one rationally deliberates that ChatGPT is capable of writing, but rather that, insofar as writing is commonly seen as a means to obtain a text, whatever process which provides us with the same effect will be called writing. Thus, this tacit assumption that ChatGPT writes refers to the easiness with which we accept being superseded with ChatGPT. Correspondingly, if a user is so avidly unafraid of blindly trusting this technology, it must be due to the way in which we understand interpretation so that we can assume that what ChatGPT does is to interpret the world for us. If a user is eager to conversationally engage with ChatGPT, it must be due to the way in which we understand communication so that we can assume that ChatGPT communicates with us. Finally, our readiness to introduce ChatGPT in almost every field of human action must be due to our understanding of the role that the abovementioned acts of signification and interpretation play in human life. If we are impatient to willingly delegate all these actions and finally live a dream life bereft of their burden, it must be because, on the one hand, life is seen as having nothing to do with writing, reading or human communication, and, on the other hand, because we assume that these activities have no intrinsic or existential value so that they are mere tasks to be completed.

This hypothesis suggests that not only ChatGPT as a technological artefact mediates our experience of the world, but that our underlying understanding of the actions we carry out equally contributes to the aforementioned transformations of experience. Furthermore, such understandings are not contingent to an individual's ideology but rather structural, historical and unconscious; indicative of this is the massive popularity of the technology (Milmo, 2023) as well as the difficulty and generalized lack of arguments which defend the intrinsic value of these actions (Morgan, 2023). Hence, we might add to each of the abovementioned structural risks that our relation to this technology implies having held the assumption that ChatGPT writes, that it interprets, or that it communicates with us in the first place; an assumption which is in turn fostered by the effects the technology's functioning has on us. This then leads us to formulate a new battery of questions: What is our understanding of writing and to which extent can we assume that ChatGPT writes? What is our understanding of interpretation and to which extent can we assume that ChatGPT interprets? What is our understanding of communication and to which extent can we assume that ChatGPT communicates with us? And, finally, what

is our understanding of life that is worth living and to which extent can we assume that such a life is to be pursued through abdicating our acts of signification and interpretation?

Our postphenomenological analysis thus answers a question with a battery of questions, which I take to be thoroughly philosophical insofar as such understandings do not refer to a particular opinion of an individual but to the culturally inherited which is influenced by the thinking of our philosophical tradition. Yet, insofar as none of these questions have been Ihde's main philosophical concern, postphenomenology does not provide us with the conceptual artillery to offer an adequate answer to them. Nonetheless, our operationalization of Ihdean postphenomenology shows that the latter's potential exceeds the identification of the structural transformations of human experience because it also helps us determine the fundamental philosophical questions that correspond to the reasons why we apprehend a concrete technology in the ways we do. Although one might contend against this claim that the philosophical questions to which we have been directed are relatively self-evident, it is important to recognize that postphenomenology has helped us suggest that, insofar as these questions are structural to the ways we relate to ChatGPT, answering them could reveal the conditions of possibility of there being a relation.

## 2.4. Conclusion

In this chapter, we have answered the question regarding what is structurally at stake in our ways of relating to ChatGPT through Ihdean postphenomenology. The analysis has shown that there is not only one ChatGPT but many and has provided explanations for what occurs to our experience in each of ChatGPT's multistabilities. Then, we have attempted to derive a structural danger from the forgetfulness of what conceals in each of the four relations. After doing so, I have posed the question of why there is a relation in the first place, a question that Ihdean postphenomenology does not seem to be able to answer. Then, we have presented the hypothesis that our underlying understanding of the action we carry out with or through ChatGPT is the condition of possibility for there to be a relation, given that the user tacitly believes that the technology is capable of undertaking such an action. Given that our goal in the next chapter is to consider how Derrida's works could help us understand what is structurally at stake in our ways of relating to ChatGPT, that Derrida's central theme is the concept of writing, and that the role I have given to postphenomenology is that of a reference point against which we can assess the potential of Derrida to help us think about contemporary technologies, it might be productive to guide our next chapter through the question concerning our understanding of writing and the extent to which we can assume that ChatGPT writes.

# 3 A Derridean analysis of ChatGPT: Teletechnological writing and logocentrism

Our postphenomenological analysis has led us to the hypothesis that, if one so readily accepts to use ChatGPT, it might be because an underlying concept of writing is at stake. To delegate writing to ChatGPT, one might be tacitly assuming that the human act of writing and the probabilistic functioning of ChatGPT are interchangeable insofar as they seemingly produce the same effect (i.e., a text). But what does writing mean and to which extent can we assume that ChatGPT writes? In this chapter, I address this question through Jacques Derrida to determine what in our common sense leads us to relate to ChatGPT and what is presumably lost in doing so. After presenting Derrida's project and reviewing an existing Derridean interpretation of ChatGPT, which looks at ChatGPT as embodying deconstruction, I argue that such interpretation runs the risk of justifying ChatGPT through deconstruction rather than utilizing the latter to identify what is at stake in our relation to ChatGPT. I then formulate the hypothesis that ChatGPT might be reinforcing logocentrism and dedicate the rest of the chapter to elaborate on this interpretation. First, I clarify what it means that ChatGPT writes from a Derridean perspective in order to emphasize that ChatGPT does not write insofar as it outputs texts. Secondly, I attempt to characterize the specificity of ChatGPT's mode of writing, which I call "simulacrum of generation" because it stages its outputs by concealing the indices that would reveal to the user that the possibility of all outputs already exists before any prompt is introduced. This allows us to show that ChatGPT's functioning and interface design reinforce the logocentric ambition of determining the limits of what can be thought. Finally, I address the so-what question regarding the claim that ChatGPT is logocentric in order to clarify what is lost in establishing a relation to ChatGPT and why should one care about (the act of) writing.

## 3.1. Derrida's project: Logocentrism and deconstruction

In his early works, Derrida identifies in the classical texts of the philosophical tradition a continuous suppression of the concept of writing (Derrida, 1967/2016, 1967/2011a, 1967/2001, 1972/1981b, 1972/1982). Such a suppression does not refer to a mere innocent misunderstanding of what writing is, but rather to the effect of a necessary conceptual strategy to account for reality (Derrida, 1967/2001, pp. 246-247). A symptom of this suppression is the reduction of writing to a mere means of communication (Derrida, 1972/1982, p. 311). This entails seeing writing as the representation,

image or copy of the discursive articulation of a thought which precedes the moment of inscription. In other words, writing is seen as a means that allow a meaning which, existing first in the author's head and requiring to be linguistically articulated before being written down, is fully captured by a set of marks on a paper—thus enabling such a meaning to be invariably transmitted through time and space. We could understand this through the model that one first thinks, then speaks and finally writes what was first thought. From this perspective, writing is inessential to thinking and communication and is subsequently subordinated to the concept of speech (Derrida, 1972/1981b, pp. 109-110). Such subordination is, for Derrida, a product of a rationality inaugurated by Western metaphysics; not only the perpetuation of a falsity by some white men from their European armchairs, but primarily the permeation of a mentality in our languages and common sense that we reproduce in our everyday lives, "all the more when one does not suspect it" (Derrida, 1967/2016, p. 25). In other words, metaphysics corresponds to a particular organization of the conceptuality contained in our languages, which takes the form of hierarchized oppositions where, just like the opposition speech/writing, the former is seen as pure and independent from the latter (Derrida, 1972/1981a, p. 41; 1972/1981b, p. 103). If this is problematic for Derrida (1967/2016, p. 77), it is because such oppositional rationality privileges a fixed sense of the world that inaugurates a thought of purity, immediacy, identity and sameness and excludes difference, mediacy, contamination and otherness. This rationality is what Derrida (1967/2016) calls "logocentrism," which he describes as "the most original and powerful ethnocentrism" (p. 11); the most original and powerful because three seemingly innocent metaphysical presuppositions allow for the imposition of an interpretation of reality's origin as a universal foundation that institutes Western (masculine) thought as the truth of the world (Derrida, 1967/2016, p. 85).[5]

First, the history of metaphysics is led by a presupposed intimacy between the logos (i.e., the reason of what is) and truth (Derrida, 1967/2016, p. 11). Thus, insofar as metaphysics seeks the ultimate truth of reality, its quest can only be carried out through rational attempts of determining a necessary foundation (logos) which must be logically true. Yet, the logos does not express itself; it requires language and a medium of signification. For this reason, logocentrism requires a second assumption to be realized—the logos as truth can only be attained through the voice and speech. The voice can be understood as the consciousness of one's thoughts (Derrida, 1967/2011a, pp. 65-69), the ability to relate to oneself without the aid of any exteriority which could alter the content of thought (i.e., meaning). Now, to exteriorize one's thoughts and share them with others, one requires linguistic signs and, thus, speech is essential to express meaning. Consequently, speech and meaning are in

[5] Cf. Derrida (1972/1982, p. 213), where metaphysics is defined as a "white mythology" that organizes Western culture around a universal foundation. Cf. Derrida (1973, p. 311; 1972/1981b, p. 48n47) where logocentrism is linked to phallocentrism insofar as the masculine discourse is linked to truth.

"absolute proximity," for the latter cannot be exteriorized without the former, and "logocentrism … is also a phonocentrism" (Derrida, 1967/2016, p. 12). Finally, given that the content of thought (i.e., the meaning of being, the reason of what is, the evidence of intuition) is attained interiorly through the voice (i.e., self-consciousness), it follows that such meaning can be immediately present to our consciousness. Put differently, the history of metaphysics has assumed that the truth it seeks is fixed and unchangeable as well as something we can fully grasp. This corresponds to the third assumption of the history of metaphysics—the understanding of being as presence (Derrida, 1967/2016, p. 13). Hence, to sum up, insofar as truth takes the form of the logos, of a rational foundation, of a metaphysical determination, language becomes necessary to express it; yet for language to exist we need a mouth and a tongue, but importantly, a consciousness—a voice that wants to say something; and if this rational deduction of the foundations of reality can be identified with truth, it is because the meaning of the principle or foundation is full, clear, univocal, graspable, and thus present (to us).

Logocentrism, as the joined force of these three assumptions, boils down, for Derrida (1967/2016), to "the exigent, powerful, systematic and irrepressible, desire for such a signified [i.e., a transcendental signified]" (p. 53). This desire for a transcendental signified means that we continuously aim to search for a rational foundation with a fixed meaning to secure a current state of affairs, whose artificiality (i.e., cultural production) is hidden by the pretension that, insofar as the foundation is logically attained, it is equated to truth and thus deemed natural. Logocentrism is thus any reaffirmation of an oppositional logic which privileges and justifies an understanding of reality through the belief of an ultimate meaning that provides a fixed identity to the meaning of a concept or being in general. In light of this critique of metaphysics, and through a bold interpretation of the texts of the tradition, Derrida allows the proper concept of writing to appear, thus providing a radical reconceptualization of signification which breaks with the oppositional and hierarchical mode of organizing our conceptuality by foregrounding the necessary contamination and homogeneity of such oppositions and the failure of identity, immediacy or full presence of meaning (Derrida, 1967/2016, p. 76). I have written that Derrida allows the concept of writing to *appear* through a *bold* interpretation of philosophical texts in order to emphasize that "deconstruction takes place" (Derrida, 1987/2008, p. 4), which is to say that Derrida recharacterizes writing not through a positive definition of the term but rather through a close reading of what is already available in the text (Derrida, 1967/2016, p. 25). This, in turn, proves Derrida's idea that there is no ultimate meaning that can secure an interpretation of a text but rather only interpretations, "political interventions in the political rewriting of a text" (Derrida, 2020, p. 47).

While we will see Derrida's concept of writing in the following chapter, we must note that it refers to a generalized concept of writing, which means that the central features of the classical notion of writing not only apply to the practice we ordinarily have been calling writing but also to speech,

thought or all experience (Derrida, 1972/1982, pp. 316-317). This conceptual discovery has the far-reaching implication that the meaning of anything is never fully or directly present to us but rather adopts the form of a sign or text to be interpreted (Derrida, 1967/2016, p. 53). General writing thus demonstrates the failure of all fixed meaning or identity and renders logocentrism as a dream or unrealizable project whose violence derive from the avoidance of the generality of this structure. That the concept of writing is generalized allows one to sustain that writing is prior to the invention of what we ordinarily call writing. This initial caveat might help us prepare for not only the complex and counter-intuitive character of Derrida's arguments but primarily for calling what we deem self-evident into question and preventing getting stuck in the safety of assuming we know more than we do. As Derrida (1993/1994) warns us: "no rest allowed for any form of good conscience" (p. xiv).

From this introduction, there are at least two reasons which make Derrida's philosophy initially attractive to address the question concerning our underlying understanding of writing and the extent to which we can assume that ChatGPT writes. First, Derrida's critical notion of logocentrism could help us consider whether the appearance of ChatGPT and our quick acceptance of it are possible due to it. Second, Derrida's generalization of the concept of writing could help us consider whether ChatGPT writes, without falling back into the question of anthropomorphization. This might help us consider the effects of our exposure to ChatGPT's functioning in terms of a mode of general writing and whether such effects correspond with a logocentrism or not.

## 3.2. An existing Derridean interpretation of ChatGPT and its limitations

### *3.2.1. ChatGPT writes without speaking: Coeckelbergh and Gunkel's Derridean interpretation of ChatGPT*

Mark Coeckelbergh and David J. Gunkel (2023) have recently drawn on Jacques Derrida's thought to make two contributions concerning ChatGPT: (1) that popular critiques of ChatGPT like Bender et al.'s (2021) stochastic parrot reaffirm logocentrism in the face of the fact that ChatGPT outputs text without having access to meaning (Coeckelbergh and Gunkel, 2023, pp. 5-6); and (2) that there is no structural difference between ChatGPT's texts and human texts because they both have the same relation to meaning (Coeckelbergh and Gunkel, 2023, pp. 7-8).

We have seen in the previous chapter that a logocentric stance is that which privileges speech over writing insofar as the former is a necessary condition to communicate one's thoughts and the latter is a mere representation of a representation, which separates meaning from its alleged conscious author. Recall that Bender et al. (2021) argue that ChatGPT does not actually make use of language even if it seems it writes and understands because it only deals with linguistic form and not with meaning. ChatGPT does not write, speak, understand, think nor communicate with us because

ChatGPT does not have communicative intent (i.e., thought, intentionality, the will of saying something) which can be articulated thanks to speech and then transmitted through writing (Bender and Koller, 2020). Bender et al.'s (2021) view is thus logocentric insofar as it reduces sense and signification to the articulation of a subject's intention. For Coeckelbergh and Gunkel (2023), logocentrism is thus reaffirmed as a response to "the fundamental challenge (or opportunity) with LLMs, like ChatGPT or Google's Bard, [which] is that these algorithms *write without speaking* [emphasis added], i.e., without having access to (the) *logos* and without a living voice" (p. 6). In other words, in light of the fact that there are technologies which, despite having no consciousness nor ability to think, are capable of artificially producing text that is perceived as meaningful by humans, traditional views on language are reasserted to fight problems of technological deception. For the authors, the problem is rather to think about these technologies by moving beyond anthropocentric and instrumentalist views of language. Consequently, Coeckelbergh and Gunkel call this challenge an "opportunity" because ChatGPT serves to illustrate traits of language that the history of metaphysics continuously represses. In this sense, the authors see ChatGPT as Derridean or deconstructive, that is, as a concrete instance that practically deconstructs the speech/writing opposition. This is because ChatGPT radically illustrates that speech is not the condition of meaningfulness, and that language can function without a subject's intention which enlivens a set of linguistic signs.

What is then at stake according to Coeckelbergh and Gunkel (2023) is "the concept of meaning itself" (p. 7). And it is Derrida's differential understanding of language that can help us explain meaning in the face of the fact that ChatGPT's outputs appear to us as meaningful as any other human text. Drawing from the Derridean interpretation of Saussure's (1916/2011) idea that "in language there are only differences *without positive terms*" (p. 120), Coeckelbergh and Gunkel (2023) explain that "any sign in any language, is characterized by the differences that distinguish it from other signs within the system to which it belongs" (p. 7). The authors bring up the example of the dictionary to illustrate this thesis: when we look for the meaning of a term, we do not find the alleged signified it refers to but rather more signifiers *ad infinitum*. Thus, language consists in an infinite chain of signifiers which refer to other signifiers in the absence of a transcendental signified which would secure their identity. This is to say that a word does not give us immediate access to a meaning as an ideal object because, insofar as its meaning is dependent on what makes it *differ* from other words in a language, these other words appear as present absences, which perpetually postpone, delay or *defer* the immediate grasp of a meaning. Under this view, meaning is not something that a text would come to make absolutely present but rather that which emerges out of this movement of differentiation and deferral. To be sure, this does not preclude mutual understanding but shows that the *possibility* of misunderstanding is structural to language and that, insofar as meaning is not brought to presence in a direct way, one can only interpret by rewriting. From this, Coeckelbergh and Gunkel conclude that

both human and ChatGPT texts have the same relationship to meaning—one of perpetual deferral—insofar as it always depends on an act of reading. They write:

> a text—whether it is written by a human writer or artificially generated by an LLM like ChatGPT (with the help of a human prompt)—comes to … enact and perform meaning by way of interrelationships to other texts and contexts in which it is already situated and from which it draws its discursive resources. (Coeckelbergh and Gunkel, 2023, p. 8)

### *3.2.2. The risk of justifying ChatGPT through deconstruction*

Coeckelbergh and Gunkel's (2023) Derridean interpretation of ChatGPT is alluring because it shows that Derrida's works are relevant today to think about concrete technologies, that ChatGPT's outputs can be taken to be meaningful without anthropomorphizing this technology, and because it subsequently shifts the ChatGPT debate away from anthropocentric and instrumentalist views of language. However, à la Heidegger (1954/1977, pp. 5-6), we could object that this interpretation is correct and yet not right. This is not to say that Coeckelbergh and Gunkel have not understood Derrida's works and that I will show in what follows how these are to be understood, but rather that they have not utilized Derrida's work as critically as they could have. Their readers might have the impression that Derrida would have supported this technology insofar as ChatGPT embodies deconstruction against the logocentrism of the critics who reassert the reduction of language and meaning to the confines of the human. In this sense, their interpretation risks utilizing deconstruction to justify ChatGPT, thus eclipsing the possibility of a Derridean critique of ChatGPT. Instead of showing that we should read Derrida today to have an understanding of meaning which renders unproblematic our interaction with ChatGPT, such a critique could show that reading Derrida today could help us understand what is at stake in our ways of thinking and acting under the influence of concrete technologies like ChatGPT. Two aspects of Coeckelbergh and Gunkel's interpretation might be indicative of what I am trying to argue here: (1) insofar as they do not clarify why it can be said that ChatGPT *writes* (without speaking), one could mistakenly think that ChatGPT writes because it outputs texts; (2) insofar as Coeckelbergh and Gunkel (2023) write that "we can try to make good chatbot technologies and large language models without … appealing to transcendental truth or meaning" (p. 9), we might be uncritically supporting the idea of human-AI collaboration, assuming it is possible and desirable without first calling into question what the concept of collaboration means and whether what we call human-AI collaboration lives up to the concept of collaboration.

How could we carry out a Derridean critique of ChatGPT? Before we narrowly side logocentrism with the views of the critics of this technology, we should consider that, insofar as logocentrism is a powerful ethnocentrism (Derrida, 1967/2016, p. 3), logocentric thinking does not boil down to the sum of philosophical texts of the Western tradition, but it rather generally refers to a common way

of thinking that is culturally interiorized and naturalized and that permeates language and common sense. Taking this into account, we cannot disregard the possibility that the development of ChatGPT and the mentality of its users might also be deeply logocentric. For instance, considering that ChatGPT's outputs are produced on the basis of a finite network of past data, every output seems to align with what Derrida (1984/2007b) calls an "invention of the same" (p. 39)—the process of integrating chance (i.e., the possibility of otherness) into what can be calculated, which results in "an order where there is no absolute surprise" (Derrida, 1984/2007b, p. 39). Insofar as ChatGPT can only output combinations of pieces of the texts that constitute its model, each of its outputs "only makes explicit a program of possibilities within the economy of the same" (Derrida, 1984/2007b, p. 44). In other words, each output reveals a combination of linguistic units which was already possible with the existing set of possibilities, so that ChatGPT does not invent anything radically new but produces something that was already available. Something radically new would be the impossible in light of the current possibilities. In this sense, ChatGPT might be seen as logocentric insofar as the *possibility* of its outputs is already presupposed in the boundaries of its model and, for this reason, ChatGPT is present to itself. Although it would be interesting to further pursue this interpretation, it would require an advanced expertise regarding the technical functioning of this technology from my side, which I do not have. Nonetheless, this type of interpretation would still be object-centred and, in the first chapter, we have called for an interactionist approach. For this reason, what interests us in what follows is to understand whether our exposure to ChatGPT's functioning, which is interrupted by the design of an interface, entails a reinforcement of a logocentric mentality. In this sense, considering that there are some reasons to suspect that ChatGPT's functioning illustrates logocentrism, we should explore whether ChatGPT promotes logocentrism through the way its outputs are displayed to us in its interface design. If this proves to be the case, perhaps it is the alignment of our logocentric way of thinking with ChatGPT's logocentrism what could explain the readiness with which we relate to this technology.

One way to carry out this interpretation is to clarify what it means that ChatGPT writes from a Derridean perspective. This will answer our guiding question regarding the extent to which we can assume that ChatGPT writes. I will do so through a cross-reading of Derrida's notion of "teletechnology" (Derrida and Stiegler, 1996/2002) and generalization of writing (Derrida, 1972/1982). Although this clarification will not bring us beyond the idea that ChatGPT is Derridean insofar as it is a metaphor of the conceptual structure of generalized writing, it will allow us to determine the specificity of ChatGPT's mode of teletechnological writing if we try to characterize such metaphoricity. Finally, I will analyse how such specificity affects our way of thinking in order to determine whether ChatGPT reinforces a logocentric rationality.

## 3.3. On the extent to which we can assume that ChatGPT writes

Derrida's notion of "teletechnology" can be interpreted as having two intertwined nuances: on the one hand, it names a type of *concrete* technologies, the so-called information or communication technologies (e.g., the telephone, radio, television, the Internet, modalities of live broadcasting, etc.), whose functioning transforms in an unprecedented way the field of human experience (Derrida and Stiegler, 1996/2002, pp. 3, 33); on the other hand, these concrete technologies exemplify what Derrida means by writing, so that he appeals to them as illustrative examples for philosophical considerations concerning the deconstruction of presence (Derrida and Stiegler, 1996/2002, p. 129; Derrida, 1992, pp. 271-272). The intertwinement of these two features implies that writing in a Derridean sense can be said to be a teletechnology and that teletechnologies can be said to write. Furthermore, insofar as the Derridean notion of writing is generalized and it names a conceptual structure which is not dependent on the invention of writing, teletechnologies can be said to exist even before the aforementioned concrete devices were invented (Derrida and Stiegler, 1996/2002, p. 38). In this section, I will show that, insofar as ChatGPT can be identified as a teletechnology, it can be said to write. Nevertheless, insofar as Derrida's notion of general writing has little to do with the ordinary act of writing (Gasché, 1986, p. 274), we cannot assume that ChatGPT writes because it outputs *text* but because it *outputs* text. In other words, ChatGPT writes only in the sense of enacting a general movement of repetition or re-presentation.

### *3.3.1. ChatGPT, a teletechnology*

Let us then begin with the common functioning of so-called media technologies due to which they are grouped under the name of teletechnologies. All media technologies re-produce a signification by artificially recreating sounds, texts, images or a combination of these. In so doing, they make ("something") present: they *make "something" present* in the sense of artificially recreating something that took place in a past moment as if it were happening once again in a present moment; and, because of this, they *make present* in the sense of making believe that what is being re-produced is a faithful representation of either something that is simultaneously taking place somewhere else or something that has already occurred (Derrida and Stiegler, 1996/2002, pp. 3-9). For instance, a telephone, during a phone call, re-produces the sound which was sensed an instant ago by another device and make us think we are talking in real time; or a television re-produces the sound, text and images which were either recorded and edited a moment ago (e.g., live news) or in a more remote past (e.g., a TV series) while making us think that those situations occurred exactly as we experience them. While the nature of this "something" that is made present by teletechnologies will be clear when we tackle Derrida's generalization of writing, the hyphenation I have been using in the term "re-production" is intended

to reject the idea that this "something" refers to a reproduction of something past as an exact copy or faithful representation. What these technologies emit is thus not a faithful and absolute substitute of something that took place in a moment in which we could not be, but rather a recreation or a technological production of something past whose singularity vanished at the very moment of its appearance. In short, re-production is not a reproduction because it cannot *absolutely* recreate what took place in a past moment in the *exact* sense in which it occurred. Rather, the emission of sounds, texts, or images, entails that what happened once or an instant ago takes place again *as if* it were its first time. Susanna Lindberg (2021) uses the term "staging" to foreground the idea that such re-production (i.e., the "making 'something' present") is oriented to produce an effect on our experience (i.e., the "making present"). What we experience in our interaction with teletechnologies are "artifactually staged significations" (Lindberg, 2021, p. 413). Thus, when we talk on the phone, we do not hear each other's voice but the staging of sounds which resemble the tone of our voices, or, when we surf the Internet, we do not see text but the staging of text through code. This should suffice to show that ChatGPT can be said to be a teletechnology insofar as it artificially produces text through a statistical recombination of pieces of past texts while making the user believe that the staged text is produced in a here and now.

### *3.3.2. Teletechnological writing: ChatGPT writes as a teletechnology*

Let us now see why teletechnologies can be said to write. Derrida tells Stiegler: "The way in which I had tried to define writing implied that it was already … a teletechnology" (Derrida and Stiegler, 1996/2002, p. 37). And: "These machines [i.e., teletechnologies] have always been there, even when we wrote by hand, even during so-called live conversation" (Derrida and Stiegler, 1996/2002, p. 38). The main thesis here is that the notion of teletechnology does not exhaust itself in the referral to a particular set of concrete technologies or devices that function in a particular way because of the intriguing fact that teletechnologies existed before these concrete artefacts were invented. This ontological precedence is only possible due to the mutual identification of the notions of teletechnology and writing.

We have mentioned in the previous chapter that writing has been traditionally understood as the image of speech, which is in turn understood as the image of thought. Hence, a text, from a classical perspective, is a mere empirical or sensible set of marks which derive from a self-present author. The written is a "modification of presence" (Derrida, 1972/1982, p. 313), a transformation of an ideal presence (i.e., a thought) into a mundane or empirical object (i.e., a set of perceivable characters). That a text is derived from a self-present author means that it is the product of a subject who, being always already conscious or aware of the thoughts and ideas they want to express—thus having them immediately present to consciousness—chooses a combination of signs which come to substitute their

thoughts *exactly* as they were generated in their mind (Derrida, 1967/2016, pp. 66-67). From this perspective, the very text you are now reading is the means that give you access to a meaning that I, in another *now*—the moment of writing these lines—allegedly had in my mind and chose to *represent* through the available signifiers of the English language, thus allowing my ideas to invariably travel through time and space. Against this view, Derrida (1967/2016, p. 56) shows that writing is the condition of possibility of communication and language in general.

By foregrounding that the possibility of repetition or iterability is the fundamental feature of writing in an ordinary sense, Derrida (1972/1982, pp. 315-318) shows that this characteristic equally applies to speech and to all experience. What makes writing writing is not the actual possibility of being understood by another but rather the possibility of its signs to be repeated or iterated. Nothing could have been written in the first place without the possibility of repeating a linguistic unit. A word cannot be what it is if it cannot be repeated and recognized as the same every time it appears. Thus, it is iterability that allows something to be recognized. It is the possibility of a sign to be repeated due to rules which govern the relationships among signs that allows a third person to recognize those signs and further repeat them. For instance, the minimal elements of the very text you are now reading are the letters of the English alphabet, which are nothing but arbitrary strokes with a correspondent sound and a certain ascribed status within the artificial code that English grammar is. Even in the hypothetical situation that you and I exchange letters in a secret language no one else in the world knows, after our deaths, our correspondence would still have the status of writing and not of, say, a drawing, because the rules which determine the relations between the signs imply the possibility of a third person to be able to decipher them and repeat or iterate the signs (Derrida, 1972/1982, p. 315). From this it follows that a piece of writing keeps its status of writing in the absence of an author and in the absence of a recipient (Derrida, 1972/1982, p. 316). A written text keeps its status of writing even if there is no one on Earth that can understand it because the possible repetition of signs implies the possibility of ascribing meaning to it. This breaks with the representational view of writing (Derrida, 1972/1982, p. 316) because, insofar as iterability shows that writing can function in the absence of the author and thus of the signified (Derrida, 1972/1981b, p. 77), the written can no longer be seen as the copy of a thought which preceded it. Derrida's focus on iterability reveals that no writing can enclose "a meaning that is *one*" (Derrida, 1972/1982, p. 309) because the necessary temporal interval it exists between the moment of inscription and the moment of reading makes structural "the possibility of extraction and citational grafting" (Derrida, 1972/1982, p. 320). I understand this in two simultaneous senses: on the one hand, a text remains a text even if parts of sentences have been simply extracted from it or if other linguistic elements have been introduced or grafted in the original; on the other hand, the existence of a text in new spatiotemporal situations entails an extraction from the original situation (i.e., the moment of inscription) and a grafting into new ones.

To understand why teletechnologies can be said to write, we need to understand why this structural iterability of writing applies not only to what we usually mean by writing but to "all language … and ultimately [to] the totality of 'experience'" (Derrida, 1972/1982, p. 318). Speech is writing because, even when we communicate orally with each other, we must make use of sounds which have a particular identity in a language so that they can be identified by another in the moment of talking. Speech is writing insofar as it can be respoken. So, even if one might think, like Saussure (1916/2011), that we "learn to speak before [we] learn to write" (p. 25) or that people did think and communicate with each other before writing was *invented* as a technology to support memory and communication, writing as the possibility of repetition had always been the condition of possibility of speech. Again, the reduction of writing to an invention or a form of communication which represents speech entails seeing thought as the spontaneous generation of ideal meanings, which are later indicated through linguistic signs. For Derrida (1962/1978, pp. 88-89), there is no meaning or ideal object that exists with anteriority to its *possibility* of being written or infinitely repeated. Such a meaning could not even be thought. For example, we cannot think what an equilateral triangle is if we could not *possibly* draw it. Thus, it is the possibility of being repeated or iterated that constitutes meaning and that characterizes writing ontologically. The generalized concept of writing, which Derrida also calls "arche-writing" (1967/2016, p. 61) or "trace" (1967/2011a, p. 58), is nothing but this possibility of repetition. Again, not as an image or a copy—for this would assume the precedence of an ideal object to which the written words would point to—but rather as the act of bringing forth once again something that was never present to itself. Iterability renders reality as a trace or text to be interpreted (Derrida, 1998, p. 148). The reason why something is never present to itself, even if it is repeatable, and the reason why we can only refer to reality interpretively is due to the *passing* or spacing of time (Derrida, 1967/2011a, p. 73). Since repetition can only take place in the world and time has necessarily passed in the *re*-presentation, what is repeated implies both a certain self-identity that allows it to be recognized as the same and a radical singularity or difference, which manifests in every appearance but that cannot be repeated. Each repetition thus implies difference with regard to itself in the previous instance; and, due to this, there is no origin that secures an unchangeable character of reality. Therefore, if something must be able to be repeated in order to simply be and the repetition makes it differ from what it is, the being of a thing (i.e., the thing itself) is always a sign (Derrida, 1967/2016, p. 53) or trace—a repetition of something that was but that was never given in its presence because the very passage of time introduces the possibility of simultaneously being something else (Derrida, 1967/2016, p. 76).

These dense descriptions allow us to show that, insofar as teletechnologies re-present or repeat something that never had an absolute fixed identity, teletechnologies can be said to write. It is important to note that teletechnologies do not write insofar as they emit a signification but because they

function by repeating something past in the present. In other words, that ChatGPT "writes without speaking" (Coeckelbergh and Gunkel, 2023, p. 6) should not be understood as the production of texts in the absence of consciousness, but rather as the combined repetition or re-presentation of parts of past texts. In this sense, AI systems like Dall-E or Midjourney, which produce images, write too. Teletechnological writing is thus to be understood as the gap or delay existing between the registration of the happening of a signification and the artificial staging. An easier example than ChatGPT can show this more clearly. Suppose you and I are talking on the phone. When you hear me speak, it is actually not my voice that you hear. You experience the technological production of a sound which resembles my voice. Between my speaking and your hearing there is a necessary latency or delay through which my telephone device transforms the sound waves I produce into electric signals so that they can be sent to your device through a telecommunication network and finally be transformed back into sound waves that you can hear. Keeping this delay minimal is the condition for us to have the impression that we experience each other's voice in real time—that my talking simultaneously occurs in another geographical location at the very moment of your hearing. However, such a delay deprives the interlocutor from controlling the sense with which a message is interpreted in their absence. The phone thus writes because it *re*-presents a disembodied voice in the absence of its producer who, despite partaking in the conversation, cannot re-appropriate their willed intention (Prenowitz, 2008, p. 142). Having now clarified the extent to which we can assume that ChatGPT writes, we must pursue our hypothesis regarding the potential logocentrism reinforced by ChatGPT through a characterization of the mode in which ChatGPT writes (i.e., a conceptual explanation of its technological functioning).

### 3.4. On the mode of ChatGPT's teletechnological writing and its implications

Derrida outlines in passing the specificity of the teletechnologies of his time (e.g., television, mobile phones) as a "simulacrum of life" or as the "restitution as 'living present' of what is dead" (Derrida and Stiegler, 1996/2002, p. 39). This means that such teletechnologies do not only recreate the content of an act of signification that has already taken place, like a written text does, but they also recreate the very act of signification itself, thus making one believe that something past—and, in this sense, dead—takes place here and now as its first time. The arrival of ChatGPT seems to inaugurate a new mode of teletechnological writing which does not fully correspond to this "simulacrum of life" (Derrida and Stiegler, 1996/2002, p. 39). ChatGPT does not merely make past texts take place anew as if they were happening here and now. ChatGPT serves itself from past significations to artificially produce texts which, despite being staged in a temporalized manner by placing letter after letter at a reading speed, do not maintain the singularity of the already existing significations it requires for the staging. We thus experience the staging of a text across its formation while the text appears as unique

and anonymous, so that it becomes appropriable by the user. ChatGPT maintains the idea of the "simulacrum of life" to the extent that it requires existing significations and makes use of parts of them to stage a text by recreating a moment of inscription. However, insofar as ChatGPT makes something seem alive while simultaneously concealing that what is being shown is actually a rearrangement of past elements made present, ChatGPT's specificity seems to do something else than make present something dead as if it were alive; ChatGPT makes present a combination of parts of dead things as if it were alive *but had never been dead*. This is why it is classified as a *generative* AI system.

### *3.4.1. Simulacrum of generation: ChatGPT as logocentric*

To characterize ChatGPT's convoluted mode of teletechnological writing, let us entertain the example of ransom notes—those letters that one could compose by gluing together cropped words or parts of words from newspaper and magazine headings to avoid deploying one's own handwriting and thus render a message as totally anonymous. In the first chapter, I have offered a critique to some metaphors for ChatGPT which have been brought about in the current debate to demystify what this technology does. Here, I am not bringing in this example to argue that "ChatGPT is something like a ransom note maker," but rather to determine the difference implied in the analogy between a ransom note and ChatGPT's production of text from Derrida's (1972/1982) above-mentioned insight that "the possibility of extraction and citational grafting … belongs to the structure of every mark" (p. 320).

Extraction and citational grafting are two possibilities seemingly exploited by both ransom notes and ChatGPT. On the one hand, the example of a ransom note visually illustrates that there has been an extraction and a citational grafting, that a part of an existing signification has been transposed somewhere else where it now adopts a new signifying role. The act of cutting out parts of words from magazine headings corresponds to extraction, whereas the act of subsequently gluing those cropped parts of words in a blank surface corresponds to the citational grafting. On the other hand, ChatGPT similarly identifies common series of characters (i.e., tokens) from already existing texts which constitute its training dataset and it subsequently combines them on a digital surface thanks to the probability with which such identified tokens have been sequentially placed in the already existing texts that the model contains. The identification of tokens here corresponds to extraction, while the probabilistic rearrangement of tokens corresponds to the citational grafting. Furthermore, the citational grafting we are dealing with in these cases is of a particular kind. Insofar as both the cuttings of a ransom note and the tokens of a ChatGPT text typically consist of a very short sequence of characters, their role in the rearrangement does not maintain a singularity from a previous signification that might be credited to a creator. This is what makes both a ransom note and a ChatGPT text anonymous and appropriable; anyone can proclaim themselves creator of such significations.

Notwithstanding, there is a central difference regarding this anonymity and possibility of appropriation between a ransom note and a ChatGPT text, a difference which will lead us to ChatGPT's specificity. Whereas the anonymity of a ransom note would allow anyone to proclaim themselves as the alleged creator of the collage but not of the cropped words, ChatGPT's output allows anyone to proclaim themselves as the alleged creator of a collage whose indices of being a collage have been suppressed. A ransom note retains several indices of something past that never took place in an absolutely controlled or purely conscious manner, such as the particular absorption of ink by the magazine paper or the unrepeatable interaction between the scissor and the paper which becomes evident the roughness of the cuttings' vertices. In this sense, a ransom note makes present an absence of something that never took place in its presence, a singularity which withdrew in its very happening, thus foregrounding the Derridean notion of trace. On the contrary, ChatGPT displays the probabilistic rearrangement of tokens by supressing or disguising the singularity both of the existence of the tokens from where they were extracted and of the very act of probabilistic rearrangement. While the possibility of extraction and citational grafting is the very condition that ChatGPT needs to produce a text, there are no indices that this process took place in a past moment and, hence, ChatGPT presents a probabilistic recombination of elements as if no arrangement had happened at all. ChatGPT's staging of an outcome does not allow anything which partook in its creation to keep existing after its production. There is nothing—at least to the eyes of the user—indicating that the words are not formed by letters of the alphabet or by syllables but rather formed by common groups of tokens, or that the sequence one is facing is influenced by previous and identifiable sequences of words not probabilistically arranged. We might thus say that ChatGPT eliminates the traces of its staging in the very moment of its staging. Such elimination takes place once the text is displayed on the chat in a homogeneous format and, importantly, through a temporalized appearance.

As a consequence, whoever is exposed to how ChatGPT stages text might be prompted to believe that it is not a production—the act of re-producing something that took place in the past and thereby already existed—but a generation—the act of bringing something into being or the occurrence of something radically new. The user has then the impression that ChatGPT generates text in a here and now as the causal result of having introduced an input (i.e., the user's prompt), when the possibility of ChatGPT's output was already present in the model as a relation among finite tokens identified from the equally finite set of texts which constitute ChatGPT's training dataset. In this sense, we could name ChatGPT's mode of teletechnological writing as a "simulacrum of generation," a simulacrum which is attained through the effacement of the trace. This idea of concealing that which makes possible the artificial production of text—the probabilistic process of extraction and citational grafting—might thus be seen as analogous to the suppression of writing by logocentric metaphysics. ChatGPT is designed to cover up or do away with that which would show to the user's unconscious

that the output is indeed not spontaneously generated or original but rather one among many finite possibilities whose existence preceded the prompt. Analogously, logocentrism conceals or suppresses that which would show that the alleged purity of what is deemed true is an illusion at the same time it depends on the suppressed to sustain the order of hierarchised oppositions.

From the above, we can suggest that the telos which motivates ChatGPT's design and development is the historical aspiration to totalization, the aim to determine the limits of what can be thought, the dream of realizing the fantasy of possessing everything that could possibly be expressed. While elaborating on the relation between ChatGPT and this fantasy would require another investigation, it is worth showing some evidence that this is indeed a historical ambition. Derrida's (1984/2007b, pp. 28-30) considerations regarding the abovementioned logic of the invention of the same draw from the Cartesian and Leibnizian quest for a universal language and Derrida (1984/2007b) aligns Leibniz's views of this language as an art of invention (*ars inveniendi*) with "the manner of an artificial intelligence" (p. 36). Paolo Rossi (1960/2006) shows how Ramon Llull's *ars combinatoria* and thinking machine deeply influenced the philosophies of Bacon, Descartes and Leibniz. Jorge Luis Borges (1939/1999) provides a brief history of this fantasy, which ranges from presocratic atomists to contemporary literature (e.g., Huxley) although it reaches its peak in Llull's thinking machine. Andrew Hui (2022) has shown that, in the *Theodicy*, Leibniz (1710/2007, pp. 375-378) concludes a philosophical work with a fable about the idea of a total library, not unlike Lasswitz's (1901/1958) and Borges' (1941/1998) fictions. Additionally, we might recall Horkheimer's (1937/2002) critique to logical empiricism as the realization of the Leibnizian dream of unifying reality and mathematics (Handelman, 2019, p. 51). Finally, one might suspect whether Sam Altman's (2023) public appraisal of empiricism against rationalism is actually not an appraisal of the logical empiricism that is reflected in the algorithmic organization of empirical data. The fantasy haunts us.

### *3.4.2. The loss of difference: Addressing the so-what question*

By suggesting that ChatGPT embodies logocentrism, a critique is implied. Something must be lost or at stake in our interaction with this technology. What could we say, from a Derridean perspective, to someone who yells a long "So what?" to us after showing that ChatGPT can be seen as logocentric? What could we say to someone who, more gently, tells us that ChatGPT simply frees them from being overworked, that ChatGPT simply entertains them or that ChatGPT simply helps them come up with the right wording when facing a writer's block?

The idea of using combinatorial techniques for creation and entertainment is not new. To continue in the lines of the ransom notes, William S. Burroughs (1963) popularized the cut-up method, which consists in cutting a page of text from either newspapers or literary works from oneself or from others in four sections and to randomly rearrange these sections to possibly "introduce the

unpredictable spontaneous factor" (p. 346) that makes a text singular and memorable. David Bowie adopted this technique to "ignite anything that can be in [his] imagination" (Yentob, 1975) and he even had a software called "Verbasizer" developed to do so more quickly (Apted, 1997). This technique for enhancing one's creativity thus resembles the figure of the muse as the artist's source of inspiration. According to a certain collaborationist discourse regarding ChatGPT in the field of art (Agüera y Arcas, 2016; TED, 2023a, 2023b), it seems that ChatGPT is regarded as the contemporary equivalent—a tool or technique which helps the artist create things they alone would have not thought. Here, it is important to recall the random character of the cut-up method versus the probabilistic character of ChatGPT. In the cut-up method, each combination is an absolute coincidence among an unknown yet finite set of possibilities insofar as each combination has the same opportunity to be given. Such a method takes advantage of the apparent nonsense of random combinations to bestow poetic meaning upon it. The outputs of such techniques do not appear as coherent from the outset, even if it is the user who ascribes meaning whatever the case. On the contrary, ChatGPT, by working probabilistically according to common patterns detected in its training dataset, gets rid of most of this seeming nonsense (i.e., illogical expressions) and outputs text which is immediately perceived as meaningful. In this sense, ChatGPT reduces the possibility of introducing "the unpredictable spontaneous factor" that Burroughs aimed at. We are thus stuck in this order of sameness that Derrida (1984/2007b) critiques for not welcoming what is radically other.

Prior to Borges' (1941/1998) "The Library of Babel," Lasswitz's (1901/1958) "The Universal Library" presented the possibility of a library containing "everything which can ever be given to humanity" (p. 238). After a writer has recounted the mathematical feasibility of a total library to his editor, the latter cries out: "This is wonderful for both editor and publisher: the elimination of the author from the literary business! The replacement of the writer by the automatic printing press! A triumph of technology!" (Lasswitz, 1901/1958, p. 239). However, when the writer tells his editor about the practical unfeasibility and undesirability of such mathematical possibility due to the library's immensity and the impossibility of distinguishing sense from nonsense, the editor unsubscribes—albeit not on the grounds of a rational deliberation. The editor subscribes to the idea while thinking that everything that can be expressed means all possible logical expressions, while he unsubscribes when he sees that everything means any possible mathematical combination of a finite set of characters regardless of grammatical rules. If this editor would be faced with Bowie's Verbasizer, he would have probably thrown it away, while he would have praised a tool like ChatGPT. What this shows is that, in the face of the possibility of a massive source of appropriable significations, we change the question we pose concerning creation. In Lewis Carroll's (1893) words: "Instead of saying '*what* book shall I write?' an author will ask himself '*which* book shall I write?' A mere verbal distinction!" (p. 131). We thus turn from a question of creation to a question of curation.

Something similar might happen with ChatGPT. Some might think, like Allado-McDowell, that *with* ChatGPT "we [create] something more compelling and imaginative than either of us would … alone" because "AI co-writing can open the mind with remarkable results" (TED, 2023b). This view sees writing as a catalyser of imagination and as the arrangement of parts of human texts and ChatGPT texts. But "Writing is not simply a 'composition' … writing is not simply a way of positing or posing things together" (Olson, 1990, p. 8), says Derrida in an interview. From a Derridean perspective, this view is problematic and logocentric because it depends on the assumption that writing is inessential to thinking, that thinking occurs both before and after posing things together. Under the effects of the "simulacrum of generation," this user might perceive ChatGPT as enhancing their imagination, while ChatGPT replaces their imagination by delimiting its potential. In this sense, ChatGPT leads us to a fantasy of collaboration where we have abdicated our absolute individual singularity, which is precisely that which cannot be represented. In his last seminar, Derrida (2011b) recalls: "individual death I've often said was each time *the* end of *the* world" (p. 259). A death of a living being is not a death among others nor leaves the world intact. Each death is the loss of an absolute singularity which makes the whole world be gone. Each individual contains the possibility of steering the world in radically new ways, against all fixed orders in which difference is suppressed. In his last interview, Derrida (2007a) says that "it is necessary … to make as if writing will determine the reader, who will learn to read (to 'live') something he or she was not accustomed to receiving from anywhere else" (p. 31). Even if meaning arises out of the side of the reader, one *must* ("it is necessary") write to change the other, to invent the reader, to transform them in some way—and, hopefully, for the better. Similarly, for Joan Didion (1976), "writing is the act of saying I, of imposing oneself upon other people, of saying listen to me, see it my way, change your mind … setting words on paper is the tactic of a secret bully, an invasion, an imposition of the writer's sensibility on the reader's most private space." ChatGPT discourages us to say "I, I, I" and thus leads us to forget "why I write" (Didion, 1976). For María Zambrano (1934/2000), "to write is to defend the loneliness in which one is … The writer defends their loneliness, revealing what can be found in it and only in it" (p. 35; my translation). In every act of writing, one thus defends this singularity of one's sense of the world, of one's secret. While it seems difficult to conceive that ChatGPT helps us defend our secret, it is more plausible to think that ChatGPT eclipses our secret.

Now one might rightly argue that this collaborationist view is not representative of the average ChatGPT user, who probably thinks that "the point of a writing assignment … [is] just to receive a grade" or that "writing is no less a utility than food or clothing" (Morgan, 2023), especially if we consider the low interest in writing among young people (Clark et al., 2023) and ChatGPT's quick popularity. This view clearly instrumentalizes writing insofar as the latter is just a (annoying and tedious) means to an end. From a Derridean perspective, this is the paradigmatic logocentric view

which presupposes that writing is inessential to thinking. According to the instrumentalist, one's faculties of thinking, memorizing and imagining are unaffected by the fact that one outsources writing tasks. This view assumes that one consciously elaborates thoughts and what one will share through writing. To understand why this view is wrong and politically undesirable, we must pay heed to the relationship between Derrida's notion of writing and the ordinary act of writing.

We have mentioned that Derrida's notion of writing has little or nothing to do with the act of writing (Gasché, 1986, p. 274). However, this does not mean that they are totally unrelated (Derrida, 1967/2016, p. 61). If Derrida generalizes the concept of writing, it is precisely to debunk the position that writing is just a means to an end or a tool. As Derrida himself states: "writing to me is the essential performance or act. I am unable to dissociate thinking, teaching, and writing. That's why I had to try to transform and to extend the concept of writing, which is not simply 'writing down' something" (Olson, 1990, p. 4). Derrida is actually charging a pervasive instrumentalism with regards to writing (Derrida, 1967/2016, p. 89), which is ultimately related to a conformism with regard to the world insofar as neglecting writing leads us to believe in the transparency of any medium conveying meaning and ultimately in meaning as a fixed presence. In this sense, our carelessness towards the act of writing results in a carelessness towards the meaning of reality and in a position of conformism with regards to the unappealable character of an order which proclaims itself as natural (cf. Gertz, 2019, pp. 93, 102). Due to this instrumentalism, we do not defend the practice of writing as the cheapest, freest, simplest act that would allow one to contend with the violent rigidity of such an order. The self-confidence with which we seemingly relate to ChatGPT—whether we think it expands our creativity (TED, 2023b) or we think it simply provides us with the ends we need (Morgan, 2023)—is an indication of our adaptation to the ends of this technology and our carelessness with regards a critical and political perspective towards the meaning of our surroundings.

### 3.5. Conclusion

In this chapter, I have tried to show how Derrida's works could help us better understand ChatGPT by identifying what is structurally at stake in our ways of relating to ChatGPT, focusing on the way of thinking this technology promotes. After suggesting that ChatGPT could be seen as logocentric and not only as deconstructive (Coeckelbergh and Gunkel, 2023), I have elaborated on this interpretation by first clarifying the extent to which ChatGPT can be said to write from a Derridean perspective. Then, I have characterized the specificity of ChatGPT's mode of writing as a "simulacrum of generation," which means that ChatGPT's interface makes the user believe that the output is brought into being in a here and now as the causal result of their prompt while the possibility of the output might be said to be already contained in the model. Because this is attained through the concealment of the traces which would indicate the contrary, I have shown that ChatGPT not only is logocentric

with regard to its technical functioning but it is also logocentric with regard to its interface design, because the latter analogously illustrates the suppressive dynamic upon which logocentrism relies to sustain the hierarchical arrangement of our conceptuality. Finally, I have tried to show the implications of being exposed to a technology which reinforces a logocentric mentality. What is structurally at stake in our ways of relating to ChatGPT is a loss of our absolute singularity, which implies a general conformism with regard to the order of the world. Insofar as this might lead to an obedient rather than responsible society, to adopt an adaptive rather than combative attitude, we should care about how we understand writing and about the act of writing to gather forces to contend with any unappealable order of the world, welcome otherness by acknowledging the irrepresentable singularity of each existent and become aware that our common ways of thinking can be detrimental and violent. To conclude with a commencement, we could perhaps get rid of the self-confidence with which we might be using ChatGPT to not assume too quickly that we are enhanced just because we *perceive* it as convenient.

# Conclusion

This thesis has aimed to attain a better understanding of what is structurally at stake in our ways of relating to ChatGPT through the philosophies of Ihde and Derrida in order to ultimately consider whether the latter's work could be fruitful to enrich critically our conceptual tools to interpret contemporary concrete technologies.

In chapter one, I have identified an unresolved issue in the discourse surrounding ChatGPT—that, despite the numerous analyses of the social risks this technology raises, we do not really know what is fundamentally at stake in our ways of relating to this technology. After presenting ChatGPT as a technology and highlighting a general interest in better understanding this technology, I have suggested the need for a structural approach in order to prevent the risk of sustaining the uncritical and instrumentalist position that ChatGPT might be unproblematic and a desirable tool once its social risks are effectively mitigated and its effects are deemed positive. I have characterized this structural stance as the identification of what is problematic regarding ChatGPT even if all its social risks were eliminated. Additionally, insofar as the problem of anthropomorphization complicates the enterprise of better understanding ChatGPT and it is central to how this technology is marketed, I have reviewed how AI experts critically address it. Focusing on the proposal of metaphors which aim to demystify the real identity of ChatGPT, I have suspected that a technological instrumentalism is presupposed insofar as it seems that, provided that we understand ChatGPT in a fixed way, we will transform how people relate to it and potentially have a desirable tool. I have thus suggested that the structural analysis should also be interactionist, so that it focuses on the relation between the human and the technology and does not presuppose in advance what ChatGPT is.

In chapter two, I have carried out a postphenomenological analysis of ChatGPT in order to show the capacity of this framework to identify what is structurally at stake in our uses of ChatGPT. After justifying the instrumental role postphenomenology plays in this investigation and the superficiality of my analysis, I have shown that one could establish all types of human-technology relations with ChatGPT, I have identified the structural transformations of experience in each relation as well as the effect they produce on our disposition towards the technology, and I have finally tried to derive a structural danger from the forgetfulness of what conceals from our awareness while using ChatGPT. I have formulated each structural danger in the form of the relinquishment of a judgment—the adoption of an uncritical stance regarding the action being carried out and the reasons why such interaction

is existentially and morally worthwhile. For instance, when we embody ChatGPT, we might forget that the technology belittles us insofar as we are discouraged to originally contribute to the act of signification or to train our writing skills. From this it follows that, insofar as we simultaneously become gradually dependent on ChatGPT to produce texts due to our experience of empowerment for being able to make others believe that we can write in ways we cannot or about topics we know nothing, we might relinquish the judgment of whether writing should be delegated to the technology. While postphenomenology helps us understand what occurs in and results from our uses of ChatGPT, this framework cannot answer why we relate to ChatGPT in the first place. In light of this limitation, we have posed the hypothesis that our common-sense understanding of the action we carry out with or through ChatGPT and the assumption that the technology is capable of carrying out such an action might be the conditions of possibility of there being a relation. Finally, I have decided to focus on the question of writing to explore whether Derrida could answer this question.

In chapter three, I have carried out a Derridean analysis of ChatGPT to show how Derrida's works could help us identify what is structurally at stake in our ways of relating to ChatGPT by focusing on the question whether our common-sense understanding of writing and the assumption that ChatGPT writes could explain why we establish a relation to ChatGPT. After suggesting that ChatGPT could be seen as logocentric—contrary to what existing interpretations evoke (Coeckelbergh and Gunkel, 2023)—I have developed this possibility by clarifying the extent to which it can be assumed from a Derridean perspective that ChatGPT writes. I have shown that ChatGPT writes not because it outputs *texts*, but because it re-presents something that was already existing. In short, ChatGPT can be said to write from a Derridean perspective because it illustrates the general structure of writing. Then, assuming that there is not one absolute metaphor for writing, I have attempted to determine the specificity of ChatGPT's mode of writing. I have shown that ChatGPT displays its outputs in a way that conceals that they are not a generation of something new but a re-presentation of parts of something that was already existing. For this reason, I have named the specificity of ChatGPT's writing "simulacrum of generation." Our analysis thus reveals that ChatGPT can be seen as logocentric not only because the probabilistic nature of the model constitutes a self-present totality of what could be ever output through it, but also because its interface design reproduces the suppressive movement of logocentric metaphysics with regards to writing; ChatGPT's teletechnological writing is a writing that wants to speak. Finally, I have faced the so-what question to show why we so readily relate to ChatGPT and what is problematic. The answer is that our common-sense understanding of writing aligns with the understanding ChatGPT promotes, so that we might deem ChatGPT a convenient and desirable technology from the outset. Moreover, insofar as this understanding of writing is instrumental, the general concept of writing is suppressed and we are led to a conformism with

regard to the given. Hence, we should care about the act of writing to regain the courage to struggle against any unappealable order of sense and strive for an ethical way of thinking.

Having answered our sub-questions, we must now address the overarching research question which guides this investigation: how could Jacques Derrida's works help us better understand ChatGPT compared to mainstream approaches in philosophy of technology and what would this show regarding the possibility of a Derridean philosophy of technology?

Let us focus on the first part of the question. As a general consideration, both our postphenomenological and Derridean analyses of ChatGPT have attempted to identify what is structurally at stake in our ways of relating to ChatGPT even if all social risks were effectively mitigated. Both approaches have shown a capacity to critically interpret a concrete technology by paying heed to the meaning it adopts in society while simultaneously sustaining the view that technologies do not have a fixed meaning or essence. In Ihde, this is visible through the concept of multistability, which not only makes a typology of human-technology relations possible but also allows us to consider the idea that one single artefact can manifest as several concrete technologies. Thanks to this, we have been able to show that there is not one ChatGPT but many. Thus, Ihdean postphenomenology has allowed us to interpret ChatGPT by acknowledging its polysemy or multiplicity of meanings; we have identified a ChatGPT-writer, a ChatGPT-reader, a ChatGPT-interlocutor and a ChatGPT-infrastructure. In Derrida, the unfixed meaning of ChatGPT is indicated by the slippery character of meaning itself, by the impossibility of posing the question "what is ChatGPT?" or any other what-is question. While this has been implicit throughout our analysis, it becomes visible through the possibility of interpreting ChatGPT both as deconstructive (Coeckelbergh and Gunkel, 2023) and as logocentric—a position upon which we have tried to insist for being more critical than the former.

Turning to the particularities of our analyses, our Derridean analysis begins where our postphenomenological analysis ends. While the latter focuses on providing an account of the structural transformations of our experience in and through our ways of relating to ChatGPT, the former focuses on the way of thinking which is the condition of possibility of our relation to ChatGPT. Derrida's works have helped us better understand ChatGPT in relation to a dominant way of thinking which structures our common sense, namely, logocentrism. By doing so, we have been able to show why the appearance of ChatGPT is so popularly cherished and unquestioned, despite the numerous social concerns it poses. However, our Derridean analysis has brought us back to the problem of abdicating our individual and irreplaceable singularity, a risk we had raised through Ihde's embodiment relation. If Derrida and Ihde are capable of raising a similar issue, why do we need Derrida, especially considering the difficulty of his philosophy? The postphenomenological analysis is alluring because it provides us with reasons which account for the experiences that result from using ChatGPT, whether emancipatory or alienating. Yet, our Derridean analysis allows us to understand that ChatGPT is not

a coincidental instance but the culmination of a historical rationality that has long attempted to set the limits of what can be thought. In this sense, we can understand why there is a relation, and not simply appeal to an innocent action carried out by an empty "I." While with Ihde we have identified, for instance, the loss of one's singularity as an ahistorical issue which affects an impersonal subject, with Derrida we have situated this loss of difference in the historical rationality of logocentrism which has long attempted to set a world order. In this sense, the loss of difference and absolute singularity is not only an accidental implication of ChatGPT, as it would perhaps seem in our reading of Ihde, but it rather corresponds to the reinforcement of a historical ambition (A. Hui, 2022). In this sense, our Derridean analysis of ChatGPT adopts a political character which foregrounds that the development and design of concrete technologies may reinforce our naturalized ways of thinking. Contrarily, Ihdean postphenomenology remains descriptive because it offers explanations for why certain experiences might result from our uses of technologies but it cannot raise a critical claim against them.

Our application of Derrida's work to ChatGPT has several limitations. Firstly, Derrida's work is so extensive and labyrinthic that I might have not fully succeeded in rigorously showing the implications of the metaphysical mindset which ChatGPT induces on us. Here, picking up on Nietzsche's and Heidegger's critique of metaphysics and Derrida's reception of both could strengthen the political import of our argumentation. Secondly, I have been assuming throughout that Derrida's philosophy is correct. Disregarding here for being nonserious the famous criticisms from relativism, nihilism or nonsense that Derrida has received, it seems necessary to consider Stiegler's (2001) critique of Derrida for not being able to account for the historical singularity of concrete artefacts and Ihde's (1983, p. 166) attack of technological determinism. While addressing Stiegler's critique would require a thesis of its own (Roberts, 2005; Colony, 2011), we can briefly address Ihde's (1983) concern that treating reality as a metaphor of the general writing risks relapsing into a monolithic determinism. In this sense, if the structure of arche-writing (or teletechnology) can be found in all technologies (Lindberg, 2023, p. 207), is arche-writing not an attempt to provide a new metaphysical foundation (Ihde, 1983, p. 178)? Although this clearly requires further research, we might have reasons to reject such an accusation, such as that the general structure of writing is unobjectifiable and thus nothing (i.e., not an essence) or that, insofar as "there is more than one kind of writing" (Derrida, 1972/1981b, p. 242), the logic fostered by *a* technology is not generalizable to all of technology. For instance, we have shown that ChatGPT illustrates arche-writing in a different way than television. Thirdly, besides Coeckelbergh and Gunkel's (2023) Derridean analysis of ChatGPT, I have not considered other attempts to bring Derrida's works to philosophy of technology, especially Lindberg's (2016) or Sjöstrand's (2021). While this is because their considerations remain on a more abstract level and focus on the identification of Derrida's views on technology in his corpus, having considered their

underlying motivations for doing so would have been helpful to better situate this investigation in a contemporary debate.

Let us now turn to the second part of our overarching research question: What does our Derridean analysis of ChatGPT show regarding the possibility of a Derridean philosophy of technology? If our Derridean analysis of ChatGPT has proven to be somewhat relevant, it is because it helps us to better understand concrete technologies by paying attention to how they affect and are affected by our common and interiorized ways of thinking. Our analysis shows, on the one hand, that our unconscious, common-sense understanding of certain concepts conditions our desire, readiness and acceptance to relate to a technology. Again, this is not to say, for instance, that a conscious, reflective and meditated understanding of writing drives our acceptance of ChatGPT, but rather that an unconsciously interiorized understanding of writing, which has historically permeated our common sense, coincides with the motivation behind ChatGPT and is in turn reinforced by our relation with the latter. On the other hand, our analysis reveals that concrete technologies can reinforce such understandings in our exposure to either the technology's functioning or to the discourse surrounding certain technologies. In the case of ChatGPT, we have primarily shown that our physical interaction with ChatGPT contributes to the reinforcement of a logocentric mentality with regard to the practice of writing but that a certain collaborationist discourse contributes to it too. However, contemporary technological developments in the fields of synthetic biology or geoengineering might be difficult to analysed by considering our physical relationship to them, insofar as they exceed a conventional relation of use given their nano and planetary scales. This complicates the question whether our Derridean analysis of ChatGPT could be generalized to technologies other than ChatGPT. Before we attempt to outline some avenues for future research in this direction, it is worth clarifying that, as informed by our analysis of ChatGPT, a Derridean philosophy of technology would have its focus on the relationship between concrete technologies and the logics or ways of thinking that, despite already existing in our culture, are reinforced by them. The goal of this philosophy would be to explain why we might be relating to particular technologies and, importantly, to foreground the political implications of having such ways of thinking reinforced by our technological surroundings. In this sense, a Derridean philosophy of technology contributes to mainstream philosophy of technology with a political dimension that is indissociable from thinking, thus displacing our focus from analyses of ethical consequences to never naturalize, above all, our ways of thinking.[6]

[6] The attempt to articulate a Derridean philosophy of technology must be distinguished from the attempt to reconstruct, following a particular reading of his works, what Derrida might have thought about the phenomenon of technology. Cf. Lindberg (2016) or Sjöstrand (2021). A Derridean philosophy of technology is thus a more practical attempt to explore how his works might help us make sense of our contemporary technological environments.

So, could our Derridean analysis be generalized to other contemporary technologies? The way through which we have dealt with ChatGPT is through a thematization of the metaphoricity between Derrida's general structures of sense and the functioning of this technology. We have seen that the notion of teletechnology is not exhausted in naming media or information technologies and that it seems to be equitable with arche-writing—although with nuances (Lindberg, 2023, p. 175). In this sense, any concrete technology might be taken to be an instantiation of these general structures (Lindberg, 2023, p. 207). However, my limited knowledge of Derrida's works and Stiegler's reception do not allow me to propose an approach to, say, identify the specificity of the mode of writing of any given technology. Michael Naas (2022) has evinced the convoluted character of thematizing this metaphoric relation apropos of mobile phones and language. Thus, in the face of this limitation, a first avenue for future research could consist in collecting Derrida's considerations of concrete technologies in his works (Lindberg, 2016, p. 370n1) to better determine whether this assessment of the metaphoricity is useful to put the emphasis on the ways of thinking that technologies promote or whether it is rather more useful to simply illustrate abstract structures of sense.

Leaving this theme aside, analysing the metaphysical assumptions of the main arguments in the often-controversial discourses surrounding emerging technologies (e.g., synthetic biology, geoengineering) might be a way of broaching what is structurally at stake in our ways of relating to technologies that exceed physical interaction. For instance, Derrida (2003/2005, pp. 145-148) does so with regards to the problem of cloning, although in the interest of illustrating that the problem of technics is a problem of reason. In the case of geoengineering, paying attention to the underlying understandings of nature, earth and world in the motivations for its development might help explaining that a technology responds to a historically-laden perspective—even if it looks like coming from the future. However, the problem is how to show whether the technology reinforces a particular understanding of nature or if all is about discourse. While one might see this as the contemporary and complexified version of the discussion between Winner (1980) and Joerges (1999) on the politics of artefacts, here the emphasis is put on the cyclical movement of affection according to which a technology might respond to a historical ambition and pose a risk to our ways of living and thinking precisely because it reinforces a logic which was already there. This self-destructive or autoimmune dynamic could inaugurate a second avenue for further research, namely, that of developing an ethico-political turn in the contemporary scholarly debates on disruption (Hopster, 2021) and on the relation between material environments and thought (Malafouris, 2013; Malabou, 2008, 2017).

# References


ABC News. (2023, March 18). *OpenAI CEO, CTO on risks and how AI will reshape society* [Video]. YouTube. https://www.youtube.com/watch?v=540vzMlf-54&ab_channel=ABCNews.

Achterhuis, H. (Ed.). (2001). *American Philosophy of Technolog: The Empirical Turn*. Indiana University Press. (Original work published 1999.)

Agüera y Arcas, B. (2016, February 23). Art in the Age of Machine Intelligence. *Artists + Machine Intelligence*. https://medium.com/artists-and-machine-intelligence/what-is-amiccd936394a83#.owr3e54k7.

Alombert, A. (2023a). *Schizophrénie Numérique*. Allia.

— (2023b, April 6). Comment penser ChatGPT ? *Conseil National du Numérique*. https://cnnumerique.fr/comment-penser-chatgpt.

— (2023c, April 24). ChatGPT : intelligence artificielle ou bêtise artificielle ? *Chut Media*. https://chut.media/tech/chatgpt-intelligence-ou-betise-artificielle/.

Altman, S. [@sama]. (2023, September 22). *empiricism is the key to progress rationalism is the key to sounding smart* [Tweet]. Twitter. https://twitter.com/sama/status/1705249487248404636.

Appiah, K. A. (2023, July 7). Can I Use ChatGPT for the Tedious Parts of My Job?. *The New York Times*. https://www.nytimes.com/2023/07/07/magazine/artificial-intelligence-workplace-ethics.html.

Apted, M. (Director). (1997). *Inspirations* [Film]. Argo Films and Clear Blue Sky Productions.

Austin, J. L. (1962). *How to Do Things with Words*. Oxford University Press.

Babich, B. (2023, June 19). Nietzsche and AI: On ChatGPT and the Psychology of Illusion. *The Philosophical Salon*. https://thephilosophicalsalon.com/nietzsche-and-ai-on-chatgpt-and-the-psychology-of-illusion/.

Bender, E. M. and Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198.

Bender, E. M. and Shah, Ch. (2022, December 2022). All-knowing machines are a fantasy. *The Institute of Art and Ideas*. https://iai.tv/articles/all-knowing-machines-are-a-fantasy-auid-2334.

Bender, E. M., Gebru, T., McMillan-Major, A., and Schmitchell, Sh. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. *Proceedings of FAccT*. https://doi.org/10.1145/3442188.3445922.

Borges, J. L. (1998). The Library of Babel. In J. L. Borges, *Collected Fictions* (A. Hurley, Trans.) (pp. 112–118), Penguin Books. (Original work published 1941.)

— (1999). The Total Library. In J. L. Borges, *Selected Non-Fictions* (E. Weinberger, Ed.; E. Allen, S. J. Levine, and E. Weinberger, Trans.) (pp. 214–216), Penguin Books. (Original work published 1939.)

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Bradley, A. (2011). *Originary Technicity: The Theory of Technology from Marx to Derrida*. Palgrave Macmillan.

Brainard, J. (2023). Journals take up arms against AI-written text. *Science*, *379*(6634), 740–741.

Burroughs, W. S. (1963). The Cut Up Method. In L. Jones (Ed.), *The Moderns: An Anthology of New Writing in America* (pp. 345–348), Corinth Books.

Cabitza, F. [@cabitzaf]. (2023a, February 26). *After using OpenAI's transformers (ChatGPT and Bing) for almost three months, I believe there is a more apt metaphor* [Tweet]. Twitter. https://twitter.com/cabitzaf/status/1629970065566101505.

— (2023b, March 19). *That exactly why I proposed to stop thinking of LLMs as stochastic parrots (animals!) and proposed the metaphor of the* [Tweet]. Twitter. https://twitter.com/cabitzaf/status/1637388543445041153.

Carroll, L. (1893). *Sylvie and Bruno Concluded*. Macmillan and Co.

Castelvecchi, D. (2022). Are ChatGPT and AlphaCode going to replace programmers? *Nature*. https://doi.org/10.1038/d41586-022-04383-z.

Chan, A. (2022). GPT-3 and InstructGPT: Technological dystopianism, utopianism, and "Contextual" perspectives in AI ethics and industry. *AI and Ethics*, *3*, 53–64. https://doi.org/10.1007/s43681-022-00148-6.

Chollet, F. [@fchollet]. (2023a, March 18). *"It's autocomplete" is not a helpful analogy to understand LLMs. A LLM is more like a database that lets query* [Tweet]. Twitter. https://twitter.com/fchollet/status/1637121320340299776.

— (2023b, March 18). *You can retrieve not just what was seen at training time, but arbitrary combinations of it. It's an interpolative database* [Tweet]. Twitter. https://twitter.com/fchollet/status/1637122108357738496.

Clark, C., Bonafede, F., Picton, I., and Cole, A. (2023). *Children and young people's writing in 2023*. National Literacy Trust. https://cdn.literacytrust.org.uk/media/documents/Writing_in_2023.pdf.

Coeckelbergh, M. and Gunkel, D. J. (2023). ChatGPT: deconstructing the debate and moving it forward. *AI & Society*. https://doi.org/10.1007/s00146-023-01710-4

Colony, T. (2011). Epimetheus Bound: Stiegler on Derrida, Life, and the Technological Conditions. *Research in Phenomenology*, *41*, 72–89. https://doi.org/10.1163/156916411X558909.

Corfield, G. (2023, May 23). ChatGPT creator warns of 'existential risk' from AI within a decade. *The Telegraph*. https://www.telegraph.co.uk/business/2023/05/23/chatgpt-sam-altman-ai-regulation-risk-fears/.

Dehouche, N. (2021). Plagiarism in the age of massive generative pre-trained transformers (GPT-3). *Ethics Sci Environ Polit, 21*, 17–12. https://doi.org/10.3354/esep00195.

Derrida, J. (1973). Avoir l'oreille de la philosophie: Entretien de Lucette Finas avec Jacques Derrida. In L. Finas, S. Kofman, R. Laporte, and J.-M. Rey, *Écarts: Quatre essais à propos de Jacques Derrida*, Fayard.

— (1978). *Edmund Husserl's* Origin of Geometry*: An Introduction* (J. P. Leavey Jr., Trans.; D. B. Allison, Ed.). Nicolas Hays. (Original work published 1962.)

— (1981a). *Positions* (Trans., A. Bass). The University of Chicago Press. (Original work published 1972.)

— (1981b). *Dissemination* (B. Johnson, Trans.). The Athlone Press. (Original work published 1972.)

— (1982). *Margins of Philosophy* (A. Bass, Trans.). The Harvester Press. (Original work published 1972.)

— (1992). Ulysses Gramophone: Hear Say Yes in Joyce. In J. Derrida, *Acts of Literature* (D. Attridge, Ed.) (pp. 253–309), Routledge.

— (1994). *Spectres of Marx: The State of the Debt, the Work of Mourning and the New International* (P. Kamuf, Trans.). Routledge. (Original work published 1993.)

— (1998). *Limited Inc*. Northwestern University Press.

— (2001). *Writing and Difference* (Trans., A. Bass). Routledge. (Original work published 1967.)

— (2005). *Rogues: Two Essays on Reason* (P.-A. Brault and M. Naas, Trans.). Stanford University Press. (Original work published 2003.)

— (2007a). *Learning to Live Finally: The Last Interview* (P.-A. Brault and M. Naas, Trans.). Palgrave Macmillan.

— (2007b). Psyche: Inventions of the Other (C. Porter, Trans.). In J. Derrida, *Psyche: Inventions of the Other, Volume I* (P. Kamuf and E. Rottenberg, Eds.) (pp. 1–47). Stanford University Press. (Original work published 1984.)

— (2008). Letter to a Japanese Friend (D. Wood and A. Benjamin, Trans.). In J. Derrida, *Psyche: Inventions of the Other, Volume II* (P. Kamuf and E. Rottenberg, Eds.) (pp. 1–6). Stanford University Press. (Original work published 1987.)

— (2011a). *Voice and Phenomenon* (L. Lawlor, Trans.). Northwestern University Press. (Original work published 1967.)

— (2011b). *The Beast and the Sovereign, Volume II* (G. Bennington, Trans.). The University of Chicago Press.

— (2016). *Of Grammatology* (G. Ch. Spivak, Trans.) (Fortieth Anniversary Edition). Johns Hopkins University Press. (Original work published 1967.)

— (2020). *Life Death* (P.-A. Brault and P. Kamuf, Eds.; P.-A. Brault and M. Naas, Trans.). The University of Chicago Press.

Derrida, J. and Stiegler, B. (2002). *Echographies of Television: Filmed Interviews* (J. Bajorek, Trans.). Polity Press. (Original work published 1996.)

Didion, J. (1976, December 5). Why I Write. *New York Times*. https://www.nytimes.com/1976/12/05/archives/why-i-write-why-i-write.html.

Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology*, *13*, 100060. https://doi.org/10.1016/j.jrt.2023.100060

Ellis, B. (2023, March 22). How ChatGPT can help with grading. *Technotes*. https://blog.tcea.org/chatgpt-grading/.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., and Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*. https://doi-org.ezproxy2.utwente.nl/10.1080/14703297.2023.2195846.

Ferrara, E. (2023). Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models. *ArXiv (Cornell University)*. https://doi.org/10.48550/arXiv.2304.03738.

FLI. (2023a, March 22). Pause Giant AI Experiments: An Open Letter. *The Future of Life Institute*. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

— (2023b, April 12). Policymaking in the Pause. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Floridi, L. (2023). AI as *Agency Without Intelligence*: on ChatGPT, Large Language Models, and Other Generative Models. *Philosophy & Technology*, *36*:15. https://doi.org/10.1007/s13347-023-00621-y.

Gasché, R. (1986). *The Tain of the Mirror: Derrida and the Philosophy of Reflection*. Harvard University Press.

Gertz, N. (2018). *Nihilism and Technology*. Rowman & Littlefield International.

— (2019). *Nihilism*. The MIT Press.

— (2023a, March 22). ChatGPT: the route to loneliness?. *The Institute of Art and Ideas*. https://iai.tv/articles/chatgpt-the-route-to-loneliness-nolen-gertz-auid-2423.

— (2023b, July 26). More than Machines? Jacques Ellul on AI's real threat to humanity. *Commonweal Magazine*. https://www.commonwealmagazine.org/jacques-ellul-gertz-artificial-intelligence-AI-WGA-technology.

Gunkel, D. J. [@David_Gunkel]. (2023, April 24). *"#ChatGPT is a really good word remixer…One of the things we have to recognize is that it is echoing* [Tweet]. Twitter. https://twitter.com/David_Gunkel/status/1650478545821478912

Handelman, M. (2019). *The Mathematical Imagination: On the Origins and Promise of Critical Theory*. Fordham University Press.

Harwell, D., Tiku, N., and Oremus, W. (2022, December 10). Stumbling with their words, some people let AI do the talking. *The Washington Post*. https://www.washingtonpost.com/technology/2022/12/10/chatgpt-ai-helps-written-communication/.

Heaven, W. D. (2023, May 2). Geoffrey Hinton tells us why he's now scared of the tech he helped build. *MIT Technology Review*. https://www.technologyreview.com/2023/05/02/1072528/geoffrey-hinton-google-why-scared-ai/.

Heidegger, M. (1977). *The Question Concerning Technology and Other Essays*. Harper & Row Publishers. (Original work published 1954.)

Heikkilä, M. (2023). Could ChatGPT do my job? *MIT Technology Review*. https://www.technologyreview.com/2023/01/31/1067436/could-chatgpt-do-my-job

Hopster, J. (2021). What are socially disruptive technologies?. *Technology in Society*, *67*, 101750. https://doi.org/10.1016/j.techsoc.2021.101750.

Horkheimer, M. (2002). The Latest Attack on Metaphysics. In M. Horkheimer, *Critical Theory: Selected* Essays (M. J. O'Connell, Trans.) (pp. 132–187), Continuum. (Original work published 1937.)

Hughes, A. (2023, May 5). ChatGPT: Everything you need to know about OpenAI's GPT-4 tool. *BBC Science Focus*. https://www.sciencefocus.com/future-technology/gpt-3/.

Hui, A. (2022). Dreams of the Universal Library. *Critical Inquiry*, *48*(3), 521–548. https://doi.org/10.1086/718629.

Hui, Y. (2023). ChatGPT, or the Eschatology of Machine. *E-flux Journal*, *137*, 1–8. https://www.e-flux.com/journal/137/544816/chatgpt-or-the-eschatology-of-machines/.

Ihde, D. (1979). *Technics and Praxis*. D. Reidel Publishing Company.

— (1983). *Existential Technics*. SUNY Press.

— (1990). *Technology and the Lifeworld*. Indiana University Press.

— (2016). *Husserl's Missing Technologies*. Fordham University Press.

Joerges, B. (1999). Do Politics Have Artefacts?. *Social Studies of Science*, *29*(3), 411–431. https://doi.org/10.1177/030631299029003004.

Lasswitz, K. (1958). The Universal Library. In C. Fadiman (Ed.), *Fantasia Mathematica* (pp. 237–243), Simon and Schuster. (Original work published 1901.)

Leibniz, G. W. (2007). *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil* (A. Farrer, Ed.; E. M. Huggard, Trans.). BiblioBazaar. (Original work published 1710.)

Lemmens, P. and Eede, J. van den (Eds.). (2022). Rethinking Technology in the Anthropocene [Special issue]. *Foundations of Science*, *27*(1).

Lindberg, S. (2016). Derrida's Quasi-Technique. *Research in Phenomenology*, *46*, 369–389. https://doi.org/10.1163/15691640-12341344.

— (2021). Four Transcendental Illusions of the Digital World: A Derridean Approach. *Research in Phenomenology*, *51*, 394–413. https://doi.org/10.1163/15691640-12341481.

— (2023). *From Technological Humanity to Bio-Technical Existence*. SUNY Press.

Malafouris, L. (2013). *How Things Shape the Mind: A Theory of Material Engagement*. The MIT Press.

Malabou, C. (2008). *What Should We Do with Our Brains?* (S. Rand, Trans.). Fordham University Press.

— (2017). The brain of history, or, the mentality of the Anthropocene. *The South Atlantic Quarterly*, *116*(1), 39–53. https://doi.org/10.1215/00382876-3749304.

McNamee, R. (2023, July 19). To Save Itself, Hollywood Must Build Its Own ChatGPT. *Wired*. https://www.wired.com/story/hollywood-wga-strike-artificial-intelligence/.

Milmo, D. (2023, February 2). ChatGPT reaches 100 million users two months after launch. *The Guardian*. https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app.

Montemayor, C. (2021). Language and Intelligence. *Minds and Machines*, *31*, 471–486. https://doi.org/10.1007/s11023-021-09568-5.

Morgan, A. P. (2023, February 28). ChatGPT, Strollers, and the Anxiety of Automation. *Wired*. https://www.wired.com/story/large-language-models-automation-strollers-care.

Naas, M. (2022). Mother Tongues, Mobile Phones, and the Soil on the Soles of One's Shoes: Jacques Derrida and the Phantasms of Language. *The Journal of Continental Philosophy*, *3*(1-2), 5–22. https://doi.org/10.5840/jcp202342035.

Olson, G. A. (1990). Jacques Derrida on Rhetoric and Composition: A Conversation. *Journal of Advanced Composition*, *10*(1), 1–21. http://www.jstor.org/stable/20865695.

OpenAI. (2022, November 30). Introducing ChatGPT. *OpenAI*. https://openai.com/blog/chatgpt.

— (2023). GPT-4 Technical Report. *arXiv (Cornell University)*. https://doi.org/10.48550/arXiv.2303.08774.

— (n.d.a). *ChatGPT*. OpenAI. Retrieved October 6, 2023, from https://openai.com/chatgpt.

— (n.d.b). *GPT models*. OpenAI Platform. Retrieved October 6, 2023, from https://platform.openai.com/docs/guides/gpt.

Ouyang, L., Wu, J., Jian, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, Ch., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., and Lowe, R. (2022). Training language models to follow instructions with human feedback. *arXiv (Cornell University)*. https://doi.org/10.48550/arXiv.2203.02155.

Paul, J., Ueno, A., and Dennis, C. (2023). ChatGPT and consumers: Benefits, Pitfalls and Future Research Agenda. *International Journal of Consumer Studies*, *47*(4), 1213–1225. https://doi.org/10.1111/ijcs.12928.

Perrigo, B. (2023, January 18). OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. *Time*. https://time.com/6247678/openai-chatgpt-kenya-workers/.

Prenowitz, E. (2008). Crossing Lines: Jacques Derrida and Hélène Cixous on the phone. *Discourse*, *30*(1-2), 123–156.

Proudfoot, D. (2011). Anthropomorphism and AI: Turing's much misunderstood imitation game. *Artificial Intelligence*, *175*, 950–957. https://doi.org/10.1016/j.artint.2011.01.006.

Rahr, M. (2023, January 2). The impact of ChatGPT on the role of the Copywriter. *Medium*. https://uxloc.medium.com/the-impact-of-chatgpt-on-the-role-of-the-copywriter-57819364aa10.

Roberts, B. (2005). Stiegler Reading Derrida: The Prosthesis of Deconstruction in Technics. *Postmodern Culture*, *16*(1). https://doi.org/10.1353/pmc.2006.0009.

Roose, K. (2022, December 5). The Brilliance and Weirdness of ChatGPT. *The New York Times*. https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html.

— (2023, February 16). A Conversation with Bing's Chatbot Left Me Deeply Unsettled. *The New York Times*. https://www.nytimes.com/2023/02/16/technology/bing-chatbot- microsoft-chatgpt.html.

Rossi, P. (2006). *Logic and the Art of Memory: The Quest for a Universal Language* (S. Clucas, Trans.). Continuum. (Original work published 1960. Second edition published 1983.)

Rotman, D. (2023, March 25). ChatGPT is about to revolutionize the economy. We need to decide what that looks like. *MIT Technology Review*. https://www.technologyreview.com/2023/03/25/1070275/chatgpt-revolutionizeeconomy-decide-what-looks-like/

Rowe, N. (2023, August 2). 'It's destroyed me completely': Kenyan moderators decry toll of training of AI models. *The Guardian*. https://www.theguardian.com/technology/2023/aug/02/ai-chat-bot-training-human-toll-content-moderator-meta-openai.

Rubin, J. B. (2004). *The Good Life: Psychoanalytical Reflections on Love, Ethics, Creativity, and Spirituality*. SUNY Press.

Saini, G. (2023, May 8). ChatGPT For Graphic Designers (2023): A Guide with Examples. *Unstop*. https://unstop.com/blog/chat-gpt-for-graphic-designers.

Saussure, F. de (2011). *Course on General Linguistics* (W. Baskin, Trans.; P. Meisel and H. Saussy, Eds.). Columbia University Press. (Original work published 1916.)

Shaji George, A., Hovan George, A. S., and Gabrio Martin, A. S. (2023). The Environmental Impact of AI: A Case Study of Water Consumption by ChatGPT. *Partners Universal International Innovation Journal*, *1*(2), 97–104. https://doi.org/10.5281/zenodo.7855594.

Sjöstrand, B. (2021). *Derrida and Technology: Life, Politics, and Religion* (S. Donovan, Trans.). Springer. https://doi.org/10.1007/978-3-030-83407-4.

Stiegler, B. (1998). *Technics and Time, 1: The Fault of Epimetheus*. Stanford University Press. (Original work published 1994.)

— (2001). Derrida and technology: Fidelity at the limits of deconstruction and the prosthesis of faith (R. Beardsworth, Trans.). In T. Cohen (Ed.), *Jacques Derrida and the Humanities: A Critical Reader* (pp. 238–270), Cambridge University Press.

Stokel-Walker, Ch. (2022). AI bot ChatGPT writes smart essays—should professors worry? *Nature*. https://doi.org/10.1038/d41586-022-04397-7

TED. (2023a, August 25). *How AI Art Could Enhance Humanity's Collective Memory* | Rafik Anadol [Video]. YouTube. https://www.youtube.com/watch?v=iz7diOuaTos&ab_channel=TED.

— (2023b, September 11). *Our Creative Relationship With AI Is Just Beginning* | K Allado-McDowell [Video]. YouTube. https://www.youtube.com/watch?v=gY5GA_L9NXM&ab_channel=TED.

Tiku, N. (2022, June 11). The Google engineer who thinks the company's AI has come to life. *The Washington Post*. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind: A Quarterly Review of Psychology and Philosophy*, *59*(236), 433–460.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *arXiv (Cornell University)*. https://doi.org/10.48550/arXiv.1706.03762

Verbeek, P.-P. (2005). *What Things Do*. The Pennsylvania State University Press.

— (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. The University of Chicago Press.

— (2021). The Empirical Turn. In S. Vallor (ed.), *The Oxford Handbook of Philosophy of Technology*, Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190851187.001.0001.

Watson, D. (2019). The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence. *Minds and Machines*, *49*, 417–440. https://doi.org/10.1007/s11023-019-09506-6.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L. A., Rimell, L., Isaac, W., … Gabriel, I. (2022). Taxonomy of Risks posed by Language Models. *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 214–229. https://doi.org/10.1145/3531146.3533088

Weizenbaum, J. (1966). ELIZA—A Computer Program For the Study of Natural Language Communication Between Man and Machine. *Communications of the ACM, 9*(1), 36–45.

Winner, L. (1980). Do Artifacts Have Politics?. *Daedalus*, *109*(1), 121–136. https://www.jstor.org/stable/20024652.

Wolfram, S. (2023, February 14). What Is ChatGPT Doing… and Why Does It Work?. *Stephen Wolfram*. https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/.

Yentob, A. (Director). (1975). *Cracked Actor: A Film About David Bowie* [Film]. BBC.

Zambrano, M. (2000). Por qué se escribe. In M. Zambrano, *Hacia un saber sobre el alma* (pp. 35–44), Alianza Editorial. (Original work published in 1934.)