Minds and Machines (2024) 34:42
https://doi.org/10.1007/s11023-024-09698-6

Imitation and Large Language Models

Éloïse Boisseau[1]

Received: 30 December 2023 / Accepted: 23 September 2024

Abstract
The concept of imitation is both ubiquitous and curiously under-analysed in theoretical discussions about the cognitive powers and capacities of machines, and in particular—for what is the focus of this paper—the cognitive capacities of large language models (LLMs). The question whether LLMs understand what they say and what is said to them, for instance, is a disputed one, and it is striking to see this concept of imitation being mobilised here for sometimes contradictory purposes. After illustrating and discussing how this concept is being used in various ways in the context of conversational systems, I draw a sketch of the different associations that the term 'imitation' conveys and distinguish two main senses of the notion. The first one is what I call the 'imitative behaviour' and the second is what I call the 'status of imitation'. I then highlight and untangle some conceptual difficulties with these two senses and conclude that neither of these applies to LLMs. Finally, I introduce an appropriate description that I call 'imitation manufacturing'. All this ultimately helps me to explore a radical negative answer to the question of machine understanding.

Keywords Imitation · Large language models · LLMs · Machine understanding

1 Introduction

Recent theoretical discussions on artificial intelligence (AI) have placed a significant emphasis on evaluating the capabilities and ethical concerns surrounding a category of systems now widely known as 'large language models' (LLMs) (Strasser, 2024; Marcus, 2022; Bender et al., 2021; Weidinger et al., 2021). Based on transformer algorithms (Vaswani et al., 2017), these generative models have in particular led to significant progresses in several domains associated with natural language processing (NLP). Automated translation (DeepL), computer code generation (GitHub

✉ Éloïse Boisseau
eloise.boisseau@univ-amu.fr

[1] Aix-Marseille Université, CNRS, CGGG, Centre Gilles Gaston Granger, Aix-en-Provence, France

Published online: 09 October 2024

Copilot), and even literature search and summarisation on particular subjects (TLDR) have thus witnessed substantial advancements owing to the application of LLMs[1]. Their more recent implementation is that of highly effective conversational systems, the most notable of which being ChatGPT, BERT, LaMDA, Claude, Chinchilla, PaLM, etc. The distinctive feature of such systems is that they often generate highly convincing responses that are quite indistinguishable from responses that could be produced by human beings.

These amazing software (and I am here using 'amazing' in a non-normative, purely descriptive way: a large proportion of users and the general public are literally amazed at how well these machines perform) have already led to quite a lot of discussion, though mostly in the form of general pieces of opinion by some researchers in the field of machine learning (Cerullo, 2022; Bryson, 2022; Luccioni and Marcus, 2023). The public availability of LLMs, along with the significant incident of the high-profile firing of Google employee Blake Lemoine in June 2022 for (allegedly) having publicly stated that LaMDA 2 was conscious (Lemoine, 2022) have led to a growing reassessment both within the scientific community and the general public regarding whether these software machines possess any form of consciousness (Chalmers, 2023; Cukier, 2022) or whether they might be the starting point leading to artificial general intelligence (AGI) (Michael et al., 2023). While, admittedly, most scholars argue against the consciousness of such models, the question of their understanding remains a more polarising issue. A study conducted among 480 NLP researchers revealed for example that 51% of the participants believed that LLMs are capable of understanding (Michael et al., 2023). It is thus not uncommon to see LLMs—sometimes even explicitly referred to as 'language understanding systems' (Devlin et al., 2019)—attributed with cognitive or proto-cognitive abilities—albeit at times portrayed as abilities of a lower kind compared to those of human or even non-human animals.

In this vein and in a recent piece for general audience, philosopher Jacob Browning and machine learning researcher and chief AI scientist at Meta Yann Le Cun (2022) argued that to refuse to call LLMs 'intelligent', or to speak of them as 'understanding' language, is tantamount to 'semantic gatekeeping'. This accusation is rather common. One is guilty of semantic gatekeeping when one refuses—for chauvinistic or emotional reasons, that is, for the *wrong* kinds of reasons—the use or the extension of some terms in new contexts. The point they make is that, while LLMs do not engage with language in the same way that we do or tend to do, they *do* engage with language in their very own way. This raises two difficulties: the thorny question of the status of this kind of ascription (saying of a machine that it 'understands', even a little), and the difficult question of the legitimacy of this kind of accusation (of semantic gatekeeping—when one refuses to say that a machine understands, even a little). I will obviously not have the space in this article to deal fully with these two issues but I would like to present a few thoughts that might pave the way to clarify them. These reflections are articulated around a concept that often goes unnoticed but is used in an explanatory way by radically different positions: the concept of imitation. I will first briefly present a continuum of positions ranging

[1] For a more comprehensive overview of tools developed using LLMs, see Hutson (2022).

from what I will call the ‘full-understanding position’ to what I will call the ‘no-understanding position’ and show that this concept is often used at several stages of this continuum in an explanatory way. I will do that not to show that imitation should in any way be required or necessary for there to be understanding, but rather, more modestly, to point out that the very same concept of imitation is mobilised at key moments, and sometimes for contradictory purposes, whether to deny that there is the slightest understanding in LLMs or, on the contrary, to argue that there could very well already be some understanding in LLMs. I will then attempt to analyse afresh this concept of imitation and endeavour to give a novel account of the relationship between LLMs and the varieties of imitation. I will eventually defend the idea that LLMs do not *imitate* human speech, are not in themselves *imitations* of human speech, but are best described as *imitation manufacturing* tools, i.e., tools that manufacture imitations of human speech.

2 The Continuum from the No-Understanding Position to the Full-Understanding Position

2.1 Two Extreme Views: Full Understanding Versus No Understanding

It might seem fruitful to regard Browning and Le Cun’s aforementioned position as defending a sort of middle ground within a wider spectrum of views on the issue of the understanding of LLMs. On one side of the spectrum, we would have the extreme views of people like Lemoine, according to whom LLMs are already conscious and hence fully understand what they say. This, I will call the ‘full-understanding position’. Arguments typically put forward to support this position tend to emphasise that present-day LLMs such as LaMDA are not ‘toy programs’ (Cerullo, 2022) in the way that older conversational programs such as Joseph Weizenbaum’s ELIZA (to which we will come back later) used to be. This thesis often relies on a mixture of ideas. Firstly, it uses computational data to emphasise the capabilities of LLMs. Secondly, it draws upon numerous comparisons that have been made between the neural networks employed in LLMs and those found in the human brain (Goldstein et al., 2022; Mahowald et al., 2023)[2]. In his ‘defense of Lemoine’ (2022), Michael Cerullo thus highlights that during its initial training phase, an LLM such as LaMDA is exposed to a significantly larger quantity of linguistic samples than a human being is likely to encounter throughout their entire lifetime. This view often rests on the notion that human comprehension fundamentally relies on the computational capacity of the brain possessed by the agent deemed to ‘understand’. If we follow Cerullo’s suggestion that ‘language understanding is the result of a vast neural network operating according to the laws of physics’, implying that understanding is essentially a series of events in the brain (which functions as a biological

[2] Let us note that the authors of these articles do not always explicitly aim to demonstrate the understanding or communicative abilities of LLMs. Instead, their primary objective is to underscore the similarities between LLMs and brain neural networks in terms of language processing, suggesting that LLMs and brains operate in a similar manner.

computational device), then replicating these events in an equivalent artificial computational device should result in an artificial counterpart of understanding. Were we to accept this proposition, it would imply that LLMs already and actually possess the capacity of understanding. Alternatively, were we to refuse to ascribe this capacity to LLMs (while still accepting Cerullo's conceptual framework), we would need to reconsider our fundamental assumption regarding the capacity of human beings to understand.

We do not have the space here to discuss this position in depth: let us simply mention that ever since the emergence of the field of cognitive neuroscience, a philosophical tradition which could be labelled as 'Wittgensteinian', 'Rylean', or 'neo-Aristotelian' has sought to show the inherent misconception—described by Maxwell Bennett and Peter Hacker as a 'mereological' mistake (Bennett and Hacker, 2022; Hacker and Smit, 2014)—of attributing such abilities to the brain instead of the whole individual (the human animal) whose brain it is (see also Kenny, 1989; Cockburn, 2001). This error, if it is indeed an error, would have obvious and immediate repercussions for the case of the conversational machine.

On the other side of the spectrum then, we would find the extreme views of researchers such as Bender, Gebru, et al., according to whom LLMs have no understanding at all of the textual outputs they produce. I will call this second position the 'no-understanding position'. According to the no-understanding position, LLMs would not even have a basic understanding of what they are saying, 'and only have success in tasks that can be approached by manipulating linguistic form' (Bender et al., 2021). As a consequence, LLMs are sometimes remarkably qualified as 'stochastic parrots' (Bender et al., 2021), an expression that has been widely used and discussed in the recent literature and which aims to emphasise both the probabilistic nature of the sentences generated by LLMs, and the lack of understanding of said sentences on the part of the machine[3]. What is striking, for my purposes, is that the parrot is commonly (and paradigmatically) seen as an animal engaged in *imitating*—and 'parroting' is considered as a synonym for 'imitating'. Anyway, for now, what are the arguments for the no-understanding position? In the ongoing debate regarding the language understanding capabilities of LLMs, one of the frequently raised arguments of the no-understanding position revolves around the human tendency to assign mental states to non-living entities (Romero, 2022), commonly known as the 'ELIZA effect'. According to Hofstadter (1995), the ELIZA effect indeed is an 'illusion' which specifically consists in 'the susceptibility of people to read far more understanding than is warranted into strings of symbols—especially words—strung together by computers'. This is a genealogical kind of argument that would explain why one tends to attribute understanding where there is none.

[3] Raphaël Millière (2021), for his part, uses the term 'stochastic chameleons' to refer to LLMs; his main reason for preferring this expression over the former is to emphasise the adaptive nature of such systems, about which we will say a few words later on.

2.2 A Middle Ground: 'Shallow Understanding'

Going back to Browning and Le Cun's article, we could say that the two authors advocate an apparently reasonable middle ground between the full-understanding and the no-understanding positions. While they do not assert that LLMs possess complete understanding of the statements they produce and the information they are given, the authors still appear to be open to the potentiality of such an understanding on the part of LLMs; they do not completely dismiss it. Yet, Browning and Le Cun are well aware that the understanding of LLMs—if there is such a thing—is very limited. A first explanatory reason they put forward has to do with the programs' computational limitations: they observe that LLMs currently have 'the attention span and memory of roughly a paragraph', which obviously defeats any hope of having lengthy threaded conversations with them. Even more critically, another explanation for their limited comprehension supposedly lies in the fact that LLMs only manipulate textual data and are not in any way grounded in the 'real' world. The authors thus claim that LLMs are endowed with knowledge ('linguistic knowledge' or what we could call—following Ryle (1945, 1949)—*knowledge that*), but lack practical knowledge ('knowledge how' also sometimes labelled 'procedural knowledge' or 'hands-on knowledge'). Thus, LLMs can, for instance, 'explain how to perform long division without being able to perform it'. This distinction between possessing pure knowledge and being able to properly use this knowledge is frequently revisited in the literature: Mahowald et al. (2023), for instance, argue that LLMs demonstrate what they call an imperfect though very promising 'formal linguistic competence' (which pertains to the fact of knowing 'linguistic rules and patterns') but do not possess a functional linguistic competence (which this time pertains to the understanding of language in the 'real world').

Despite the lack of grounding of LLMs, Browning and Le Cun suggest that the performance of the machine is indeed such that it nevertheless entails that it possesses a certain level of understanding: what they refer to as 'shallow' understanding. This expression is of particular interest as it serves as a way for them to critically distance LLMs from human beings, while at the same time bringing them a little bit closer together. Distancing them on the one hand, since the machine would not have the same level or depth of understanding as human beings; and bringing them closer together on the other hand, since the machine is nevertheless granted—at least to a certain degree—the capacity to comprehend language (being in shallow water or being in deep water is being in water anyway).

So, more precisely, what do they mean when they say that LLMs are endowed with a 'shallow understanding'? They mean that LLMs' use of language is more akin to 'mimicry' than it is to authentic comprehension. LLMs would thus *imitate* human speech in a similar vein to that of smug 'jargon-spouting students' mimicking their instructors. Let us dwell a little on this simile. In this example, the students would not really understand the actual content of the lessons, but would nevertheless take on a professorial tone and posture when repeating whatever they have heard their teachers say. Such students would thus repeat some of what they have heard without fully understanding it. We might, for example, imagine an individual who uses a technical concept in order to appear scholarly, without fully grasping the scope of said concept or the nuances it brings with it. Smug students may well be parroting by repeating without understanding. But we should note from the outset that in spite of everything, these students understand what they are doing: while they may not *fully* grasp the meaning of some of the key concepts they mobilise, they do understand the meaning of at least *some* of the words they use (and they can at least understand simple, i.e., non-technical, sentences). Their inability has to do only with *some* of the notions in play; rote learning—even if not the best kind of learning—is still learning. Even more fundamentally, students who engage in this kind of slightly dishonest practice understand that by doing so, they are speaking, they are engaging in a dialogical practice, they are being included in a linguistic community, they understand *at the very least* what talking to someone and saying something amounts to. And indeed, since this somewhat dishonest practice is aimed at projecting the image of oneself as more intelligent or more knowledgeable than one really is, it also reveals a very strong social awareness of the fact that one is part of a linguistic community: the point of such a practice is in fact to pass oneself off as *x* (intelligent, profound, perceptive, etc.). Moreover, if we confront them with the fact that they do not understand the meaning of what they are talking about and that they are simply repeating what they have heard their teacher say, these students, whether they admit it or deny it (barring self-deception) will be aware of what they are doing (i.e., parroting what their teacher says). This also means that we can try and identify reasons for their behaviour: so we can question it, discuss it, analyse it, correct it, and so on. The point is that such students would certainly possess *some* knowledge and understanding, but not as much as they might appear to. Hence, an outside observer could be fooled as to the real state of knowledge or as to the level of understanding of the observed individuals. Returning to LLMs, the implication, then, is that such programs do indeed *have knowledge* and *understand* (at least to some extent) what they say, albeit possibly not quite as much as they may appear to. The substance of their words may then be more a matter of rote learning than of genuine understanding.

2.3 The Concept of Imitation

What is striking and what I mean to draw attention to is that, both in the ‘no-understanding position’ and in the ‘shallow understanding position’, the authors cited are comfortable with the idea of invoking the notion of imitation. For Browning and Le Cun, the fact that the machine imitates human linguistic behaviour is in no way an obstacle to its acquisition of ‘some genuine understanding’. For Bender, Gebru *et al.* however, the fact that the machine *simply* imitates human speech is precisely the reason why it should be *denied* the capacity of authentic understanding of the content of its linguistic interactions.

This concept of imitation is in fact a recurrent and central one in the philosophy of conversational systems—one recalls, of course, Alan Turing’s theorising of his famous test (1950)—and the many critics that were made to it (most notably Searle, 1980, but also Gunderson, 1964; Button et al., 1995)—in which it was expected that the machine would *imitate* the conversational behaviour of a human being. While it is true that the technology has evolved considerably since the time of Turing’s article, the concept of imitation continues to be pivotal but often unexamined in the research on more recent AI software machines, and the far-reaching implications of the Turing test still seem to be relevant today—as is highlighted by the fact that Turing tests are frequently carried out (as soon as new developments in our conversational technologies appear)[4]: the main issue at stake behind this idea of mechanically reproducing or imitating the linguistic behaviour of human beings still seems to be seen as determinant as to whether the machine should be attributed cognitive or psychological traits similar to those ordinarily attributed to human beings. This interplay between reproduction and imitation is clearly one of the major tropes that permeates AI (Russell and Norvig, 1995; Boden, 2016; Brockman, 2019; Strasser et al., 2023).

Consequences are therefore usually drawn from the fact that machines ‘imitate’ human speech, whilst the very notion of imitation usually appears to be taken for granted and is seldom the subject of an examination in its own right[5]. As we have seen, this notion plays a pivotal and sometimes explanatory role in very diverse views and in particular in the apparently more reasonable middle-ground view of LLMs’ understanding. I will then, in the following section, start to examine what is meant and what can be meant by the concept of imitation when theorising about conversational machines.

3 Imitation and Large Language Models

3.1 The Varieties of Imitation

To begin with, let us note that the term ‘imitation’ is arguably polysemous. In the broadest possible sense, then, what do we mean by this notion of imitation? What are the essential characteristics involved when we talk about imitation? At the risk of appearing trivial, let me first suggest that a concept closely linked to this notion of imitation is that of *resemblance* or *likeness*. An imitation, whatever its nature, aims (although it may very well fail) at producing a thing B (whatever its nature) that resembles a thing A–B then being the imitation and A what is imitated (the target).

[4] See for instance Bergen and Jones (2024) for an example of an LLM-related Turing test.

[5] I am not arguing here for the much stronger and much loaded thesis that some level of imitation would be *required* or *necessary* for understanding. Such a thesis is regularly defended in the context of discussions on what is sometimes referred to as ‘mindreading’ and more generally in the so-called ‘simulation theory’ of the mental (see for example Goldman, 2005). Nor am I arguing for the as strong and as loaded thesis that imitation would be—for some reason—an impediment to understanding. My discussion is situated logically prior to such an alternative. I will take as an indication of this that a logical restriction is usually in place when one engages in either position: an imitation is then understood as an *action* or as a *behaviour-behaviour* relation (see for example Farmer et al., 2018). I believe it is beneficial not to accept such a restriction without first questioning it. My aim is therefore not to defend either of these positions (imitation as a necessary step for comprehension or imitation as a sufficient impediment for comprehension) but rather to question the pervasive and unexamined use of the notion of imitation specifically in the field of AI (and particularly—since Turing—in the context of conversational machines). Moreover, I will also try to make it clearer in a moment why I believe it is a good idea not to mix up the notion of simulation and the notion of imitation (the second being usually used to define the first).

Moreover, this resemblance should not be accidental. If the resemblance between A and B is merely fortuitous or coincidental, there can be no imitation in play: on the contrary, imitation requires that A is aimed in some way or other. The voice of one person can be very similar to that of another without the former imitating the latter: it would then be inappropriate to use here such a phrase if the similarity were purely accidental. This first aspect is partly constitutive of the meaning of 'imitation': there can be no imitation if no target is aimed (just as there can be no sale without a buyer, or just as there can be no gift without a recipient). As such, it is important to note that this first aspect remains neutral regarding what is required for any thing to aim at any other thing (in the same way that saying that there can be no gift without a recipient for the gift does not exclude, e.g., that the recipient of the gift be purely imaginary—the requirement is only logical). In particular, I do not wish to suggest here that imitation is constitutionally a mental activity. It might turn out to be neither mental (in a certain sense) nor an activity[6].

Another concept closely linked to that of imitation is the aforementioned concept of *reproduction*. Sometimes presented as synonymous, these two terms nevertheless present at least two major contrasts. On the one hand, and as we have just noted, while any imitation is always intentional (in the minimal sense of being aimed at something), there can be accidental reproduction. I cannot accidentally imitate a piece of music, but I can accidentally reproduce a piece (I thought I was creating something entirely new, but it turns out that this chord progression already exists, etc.). Another notable difference between imitation and reproduction is that there is no such thing as a reflexive form of imitation, whereas there is such a thing for reproduction. One cannot imitate oneself, but one can reproduce one's own work: thus the artist who copies their own drawing is not imitating their work, but reproducing it[7].

Finally, and again in very broad strokes, the concept of imitation goes hand in hand with the possibility of deception. For B to be considered a successful imitation, it must be possible to confuse B with A, while B never actually being A. The

[6] I thank an anonymous reviewer for encouraging me to clarify this point. Saying that an imitation cannot be accidental and is the result of an aim does not automatically allow the passage to the conclusion that this aim has then to be some sort of mental process, a process that could be broken down, a process that would be like the progress of a sequence of steps that we might hope to reproduce artificially and which generally occurs somewhere between our two ears, behind our nose. It is beyond the scope of this article to justify a global framework that would reject this very passage, but let us simply note that we are collectively in the unfortunate situation where the term 'mental' is understood in radically different ways, depending on whether one holds a theory of the mind which is more akin to what Glock (2020) dubs a form of 'encephalocentrism'—which would be the default 'orthodox cognitive science' position—or whether one rather holds a neo-Aristotelian (or 'Wittgensteinian') account of the mind. According to the former, mental properties are characteristics of the brain and cerebral processes. For the latter, on the other hand, what is mental should neither be equated with the brain nor with its processes (and should neither be equated with some ghost-like, ethereal entity). The neo-Aristotelian (or Wittgensteinian) account advocates a 'capacity approach' of the mind and the mental (which then leads to other hard questions as to the necessary conditions and criteria for attributing psychological properties to a being—these are mapped by, e.g., Glock, 2020; Bennett and Hacker, 2022; Kenny, 1989). As far as the scope of this article is concerned, I want to suggest that a simple logical analysis of the concept of imitation (particularly when applied to LLMs) should not incidentally favour either of these positions.

[7] Admittedly, one can certainly parody oneself.

appeal of the professional voice impersonator thus lies in the fact that one can easily mistake the impersonator's voice for that of the person being imitated. One of the central dimensions of the meaning of the concept of imitation is then linked to this possibility of *passing off as* X (X being what is imitated).

This initial overview of the perimeter of the concept of imitation, which is obviously far too cursory, is intended to be as general as possible, but we now need to narrow our scope a little and take a closer look at some of the more salient features of imitation. One might indeed gain in clarity by distinguishing two different forms of imitation.

3.2 Imitative Behaviour or Status of Imitation?

On the one hand, an imitation can be the doing of a person who is behaving in the same way as another person does. In this sense, the imitation at play is a form of animal behaviour. If one wished, for instance, to imitate one's grandmother, one would have to take on her attitudes, her tone of voice, her intonations, her postures and facial expressions, and so on. One would behave in one's grandmother's way (i.e., just like her). To characterise this first aspect, let us speak in this sense of an *imitative behaviour*. Categorically, this first form of imitation is a *process* (it takes time, can be interrupted, etc.). This first form of imitation is relevant for our purposes. One might ask indeed if LLMs do display this imitative behaviour. Note though that if they do, then in a way the hardest part is over: as we saw with Le Cun and Browning's simile, attributing the ability to imitate to the student presupposes a galaxy of other abilities which are already of a high level of complexity (not least because they are integrated into the entire social life of the student in question), and in particular mental abilities which already require—incidentally—a certain understanding of the situation. We shall return to this point and investigate it in a moment.

Before that, let us note that there is a second sense of 'imitation' that can be relevant for our investigation. In this second sense, an 'imitation' can be the status given to a thing when it has been produced *in the manner of another thing* or with another thing *as a model*. In this case, the term 'imitation' does not refer to the particular behaviour of an individual, but to an *object* in itself. To take but a few examples: fake leather is said to be an imitation of real leather because it was produced with (real) leather *as a model*, just as counterfeit money is an imitation of real money as it was also produced with (real) money *as a model*. To characterise this form of imitation, let us use the phrase *status of imitation*. This second form of imitation is not a process but a *result*[8]. This result, however, is indeed the outcome of a process: I propose to use the term 'imitation manufacturing' to describe this process of *producing an imitation*—and to distinguish it from the process of imitating in the sense of an

[8] The suffix '-tion' often gives rise to such an ambiguity, expressing both an action and the result of an action.

imitative behaviour[9]. This second form is crucial to my point, as I will later argue that LLMs actually are not engaged in an imitative behaviour, but rather are *imitation manufacturers* or *imitation manufacturing systems* and that their outputs have the *status of imitations of human speech*.

3.3 Imitation and Duplication

Let us further remark that although the status of imitation is a form of *reproduction* of a given artefact, it nonetheless differs from that of a *duplication*. In the case of a duplication, the duplicated item is indeed exactly the same (i.e., is of the same type) as the original one. If I have in my possession a birth certificate and a duplicate of this birth certificate, I then have in my possession two official documents. The second birth certificate not only is a replicate of the first one, but a *duplicate*, which, as a consequence, entails that both pieces of papers have the same (legal) status or significance. However, if I possess both a Picasso and an imitation of a Picasso, although the second artwork (the replica) can be seen as a faithful imitation of the original, it is nonetheless not a duplicate of the painting, making it distinct in status from the original. As a consequence, if I have a Picasso and an imitation of a Picasso, I do not possess two paintings *by Picasso*. While there is surely a common description (both artefacts are 'paintings'), this common description is not exhaustive, as one of the paintings is indeed what we call a 'master painting' whereas the other is not. Counterfeit banknotes, forged signatures or textile fakes are all part of this second sense of imitation that we called the 'status of imitation': these are things that allow a common description with what they imitate (say 'piece of paper', 'inscription', 'scarf') but this description, though shared by the two items, is not a complete one. For instance, the original items which were imitated can also fall under the concepts of currency, signature, Hermes scarf, which their imitative derivatives cannot. The description that is common to both the imitated object and its imitation is what we might call a 'sortal' (Locke, 1690; Strawson, 1959) or even more specifically a 'covering' or 'dominant' sortal (Burke, 1994), i.e., a general individualising concept under which both the imitating and the imitated thing fall.

Thus, in both cases—duplication and imitation—the authorship of the items replicated is crucial in order to determine the nature of the object under consideration. It is precisely because my duplicate birth certificate was entrusted to me by the authorities vested with this power that it does indeed qualify as an *authentic* birth

[9] Thus, although we can in a sense speak of a painter *imitating* Monet's 'Impression, soleil levant', such an expression is not to be taken in the sense of an imitative behaviour: it would obviously make no sense to say that the artist is imitating in the sense of displaying an imitative behaviour that would consist in acting like Monet's famous painting. Since there is nothing that could, in the first place, consist in the behaviour of a painting, there is no such thing that can be imitated in the sense of imitative behaviour. In the same way, it is not necessary that our painter is engaged in an imitative behaviour of Claude Monet himself. We thus naturally understand the expression to mean that the painter is *in the process of creating an imitation* (the painting, which will itself have the *status* of an imitation). For these kinds of situations, we might say that the painter's imitation is a case of 'imitation manufacturing' (which would simply consist in producing an imitation—in this particular case, reproducing a painting), and that the resultant painting is an imitation (in the sense of a 'status of imitation').

certificate. In the meantime, it is because a painting originally by Picasso was reproduced by a person who is *not* Picasso, that the resulting painting does not qualify as an *authentic* master painting (and indeed, had Picasso reproduced his own painting, there would then not have been an original painting and its imitation but two original paintings)[10].

In a similar vein, the relationship between imitative behaviour and duplication is also not always clear-cut, as some behaviours, when imitated, can ultimately result in identical outcomes. For example, there are situations, particularly in learning contexts, where it is necessary to imitate someone in order to succeed in doing what that person does. In the *Poetics*, Aristotle thus stresses that imitation (*mimesis*) is first and foremost an animal mimetic behaviour which comes naturally to us from childhood and is crucially and intimately linked with the instinct we all have for *learning*[11]:

> Imitating comes naturally to human beings from childhood, and in this point they differ from other animals, in that they are the most imitative of all and in that what they first learn they learn through imitation (Aristotle, 1448b4–17).

We could indeed think of countless examples of situations in which it is necessary for a child to reproduce or imitate a behaviour in order to make it their own: for instance, it is by observing and imitating (here seen as a form of reproducing)—say—their mother's gestures when she pets the cat that the child learns 1/ what it is to pet a cat and 2/ how to do it. To take another example, it is by imitating—say—their brother watering the plants that the child learns how to water the plants. Hence, engaging in an imitative behaviour can result in the development of new (and authentic) *capacities*. Apart from the imitation of specific actions, which facilitates the acquisition of skills or abilities, like learning to handle a cat gently or to water the plants, we can also recognise that imitative behaviour has the potential to reshape one's overall behaviour, i.e., to help one become a person of a certain type. Let us here consider Aristotle's virtue ethics, in which virtue is famously understood not as a mere capacity or feeling (Aristotle, *Nicomachean Ethics* 1105b–20), but as a *hexis*, often translated as 'disposition' or 'state'. Being virtuous primarily entails *acting as a virtuous person would act* (Aristotle, *NE* 1105b–10). A *hexis* is a kind of quality, different from a simple *condition* (*diathesis*) in that it is less fleeting, more stable (Aristotle, *Categories* 8b27-9a13): thus having a flu, for example, is a condition (being sick), a quality that can easily be lost (say by getting the right treatment). Knowing something (being knowledgeable on a given subject) or being virtuous (these are two examples that Aristotle gives), on the other hand, are qualities that can hardly be lost; they reshape the very nature of the person that have them. But just as it is difficult to lose these qualities, it is difficult to acquire them.

[10] Here we leave aside the case of painters who keep on creating 'the same painting' (where 'same' is not understood in the sense of numerical identity) over and over again. As we already rapidly mentioned, in this situation the painter is not imitating themselves, but nor are they duplicating their work.

[11] Aristotle uses two dimensions of *mimesis* which coincide arguably with the two aspects we have just been describing (imitative behaviour and status of imitation).

One could then go on to say that a significant part of the process of becoming virtuous involves imitation (under the form of imitative behaviour) (Hampson, 2019). Imitative behaviour (i.e., acting as a virtuous individual would act) thus serves as the initial step which, through the force of exercise and habit, can ultimately lead to fully embracing one's virtue (i.e., to authentic virtuous conduct). In more recent literature and following Fossheim (2006), this developmental process is sometimes called 'practical *mimesis*'[12].

In contexts like these, imitative behaviour can therefore lead to both behaviours—i.e., the imitative and the imitated one—gradually falling under the same description. Hence, the child who learns to pet the cat through practise and imitation of their mother can be described as petting (or attempting to pet) the cat—just like their mother can also be said to be petting the cat. When it comes to virtuous behaviour, on the other hand, things are a little trickier. According to Aristotle, for a person to be deemed virtuous, it is crucial that they act according to a clear state of mind: therefore, not all imitation of virtuous behaviour is inherently virtuous. However, imitating virtuous behaviour is a necessary step for the individual to ultimately become virtuous[13]. A person will be authentically virtuous when their actions are no longer performed by merely imitating what virtuous individuals do, but rather because they truly conceive and conduct their actions as such—and as Fossheim suggests, this attitude of seeing virtuous actions as ends in themselves is actually 'the very one that *mimesis* leads to'[14]. Still, in the end, the person who has become virtuous but who, in order to do so, used to imitate the behaviour of virtuous people, shares a common description (that of 'virtuous individual') with the people they used to imitate. This interaction between imitation and the acquisition of a new nature and the corresponding transformation (becoming virtuous) could, of course, add fuel to the fire for the advocates of an understanding of LLMs: since the process is gradual, since there is, moreover, in these technologies, the description

[12] Fossheim describes practical *mimesis* as the way in which...

> Children and young people develop their character by actively engaging in *mimesis* of others who function as models for them. The child does as others do, and learns to become a certain sort of person by emulating the actions and manners of others. [...] In re-enacting, one is oneself the repetition ('this') of a model ('that') (Fossheim, 2006, pp. 111–112).

[13] Aristotle (*NE* 1105b) writes:

> (...) virtue results from the repeated performance of just and temperate actions. Thus although actions are entitled just and temperate when they are such acts as just and temperate men would do, the agent is just and temperate not when he does these acts merely, but when he does them in the way in which just and temperate men do them. It is correct therefore to say that a man becomes just by doing just actions and temperate by doing temperate actions; and no one can have the remotest chance of becoming good without doing them.

[14] Fossheim's full remark goes as follows:

> The virtues are realized only if their realisations—in virtuous actions—are seen as ends in themselves by the learner. But this attitude to virtuous actions is the very one that *mimesis* leads to. For, by definition, what is aimed at in the *mimesis* of an activity is *that activity itself*: mimetic pleasure in any performance is proper and intrinsic to that performance, and does not depend on what if anything follows upon it. Hence mimetic desire ensures that, whatever the learner fastens on, relating mimetically to it will at the same time mean relating to it as something to be savoured for its own sake. Thus an action which might otherwise be done in order to receive a reward or to avoid punishment will, if it is instead performed mimetically, be done without ulterior motives (Fossheim, 2006, p. 113).

of the training as involving learning (RLHF), one might hope that an LLM could eventually have its nature reshaped in the very same way. We will come back to this in a moment.

On the other hand, let us also note that in a context of mockery where imitating means aping or parodying, there is, this time, no point in applying the same descriptions for the original behaviour and the imitative one: if I thus imitate the Queen addressing the people of England, I am obviously not myself addressing the people of England!

3.4 Imitation and Simulation

Last but not least, it is crucial to distinguish imitation from simulation. There is, of course, a vast literature and many debates surrounding the notion of simulation[15]. For my purposes, I will simply identify three key differences with the notion of imitation. These differences will suffice to prevent us from mistaking one for the other.

First, as we saw earlier, whether in an imitative behaviour or in a status of imitation, at least some elements of the target *have* to resemble elements of the imitative behaviour or of the result. This is not the case for simulation, although we can certainly also recognise the role of a target in a simulation. In contrast, in a simulation, some elements must *stand for* (and not necessarily resemble) some elements of the target. To take a few basic illustrations, the result of a simulation can be a written sentence or a paragraph, a sequence of numbers, a graph, etc., which do not necessarily resemble the simulated target (if we think of a natural phenomenon, the lack of resemblance is striking).

This first difference makes the second intelligible, since in the case of simulation, it is not always possible to be deceived and thus confuse what is being simulated with the simulation (whereas—as we have seen—it is always possible in the case of imitation to mistake the imitation with what is imitated). Granted, the result of a simulation *might* resemble what is simulated (and some far-fetched scenarios could lead to misunderstandings, mix-ups, conspiracies, and so on), but this is generally not the standard usage of a simulation. In most cases, if not all, a simulation is guided by an epistemic aim (which is not necessarily the case with imitation); and in most cases, but not all, simulation aims at something dynamic (whereas the status of imitation can aim at something completely static, as was the case with the leather or currency imitations). A simulation must be apprehended as such in order to be what it is, whereas this is not the case with imitation (a thing having the status of imitation does not need to be recognised as an imitation in order to exist). Simulation (but not imitation) is thus in the eye of the beholder.

The final difference has to do with the idea, mentioned earlier, that between what is imitated and what imitates, there is always a covering sortal concept applicable to both in their respective description. This covering sortal concept is of course absent in the case of the simulating-simulated relationship.

[15] For an in-depth analysis, see Varenne (2019).

The fact that simulations are used for epistemic purposes (e.g., to facilitate apprehension, formulation, explanation, theorisation, shared knowledge, etc.) should not mislead us: it is not the simulation that knows or understands anything, but *we* who learn or understand something thanks to the simulation, or through the simulation. I thought it would be useful to distinguish here imitation from simulation, since there are of course many AI applications and functions, and in particular LLM applications, which can also be used for these epistemic purposes. What I would like to suggest is that just as there is no question of saying in the case of simulation that it (i.e., the simulation) knows or understands (even if it can be used for purposes of understanding and knowledge), there is no question of saying in the case of LLMs that they understand (even if they can be used for purposes of understanding—i.e., to help *us* understand)[16].

3.5 What About Large Language Models?

For the purposes of our discussion, imitation can therefore be seen in two ways: as an imitative behaviour, where one individual reproduces the behaviour of another; and as an imitative status, for cases in which an artefact has been produced on the model of another artefact. Moreover, these two senses of imitation do not always have the same relationship with the thing they imitate, which is why it is important to distinguish on the one hand the status of imitation from a simulation, from a mere reproduction and from a duplication, and on the other hand to distinguish (as far as imitative behaviour is concerned) between the behaviour imitated and the context of the imitation.

Having said that, where does this leave LLMs? When we talk about imitation in the context of conversational systems, are we dealing with *imitative behaviour* or are we dealing with *status of imitation*? The answer does not seem entirely clear. The difficulty appears to stem from the following fact: on the one hand, what we are dealing with is an artefact, albeit arguably an abstract one (on the typology of artefacts, see Thomasson, 1998); therefore, we might be inclined to categorise the machine's imitation under the header of *status of imitation*. In this view, the machine that generates sentences would be an imitation of an individual who speaks, much like a forged signature is an imitation of a genuine signature. However, on the other hand, the imitation in question occurs through the machine's workings. Consequently, we are tempted to classify this imitation as an *imitative behaviour*. In this second perspective, it would then be the machine itself that imitates, i.e., that behaves in a way similar to (or in the manner of) the human being who speaks (who communicates). Besides, if the imitation of the machine were to be considered through the prism of imitative behaviour, the question would then arise as to the status to be attached to such imitative conduct: does the machine imitate human understanding and speech in the sense that it mocks them? Does it imitate understanding and speech in a vein similar to that of the imitation of virtuous conduct in the Aristotelian tradition, i.e., is it in the

[16] I would like to thank an anonymous reviewer for encouraging me to look into the point of this section.

process of becoming an understanding and a speaking machine *by means of its action of imitating*? Is the quality of being a talker then a sort of *hexis*? The attribution of imitative behaviour to the machine is of course the one that raises the most questions, and the subject has been a matter of debate since the birth of AI as a discipline.

It will be recalled that Turing (1950) puts forward the idea that a machine that would produce sentences convincingly enough to be taken for a human being in a conversation should be regarded as being endowed with intelligence, in the same way that a human being is. Turing goes even further, suggesting that the imitation game is in fact a common and practical way for us to test the understanding of our fellow human beings—and he himself mobilises the previously mentioned notion of parroting to distinguish full from partial understanding: ‘The game [...] is frequently used in practice under the name of *viva voce* to discover whether some one really understands something or has “learnt it parrot fashion”.’ Turing’s relationship with imitation is thus twofold: on the one hand, it seems that mere imitation or parrot-like repetition is not in itself a sufficient criterion to qualify a machine as intelligent. The value of face-to-face or off-the-cuff conversation thus lies in the fact that such an exchange makes it possible to assess an individual’s actual degree of understanding of a given subject. Conversation would thus be a relevant activity for distinguishing imitation of understanding from genuine understanding. Imitation in this sense is conceived as not requiring any intelligence on the part of the machine. The machine that gives parrot-like answers could then probably be considered more as an imitation than as imitating.

On the other hand, the imitation game is designed as a tool for determining whether a machine is actually capable of thinking. The rationale behind the imitation game is indeed that since the machine is capable of imitating human speech, it must be endowed with the intellectual capacities to do so. According to this second take on the notion of imitation, the machine that satisfactorily imitates a human being engaged in a conversational activity (i.e., a machine producing responses that are indistinguishable from those that a human being might produce) must be at least a little intelligent, as if the impressive nature of the imitation could not be the result of anything other than genuine thinking on its part. The tension we have already noted in Browning and Le Cun is again present. Similar considerations would appear to be at stake behind their statements, notably when they write that the imitation of LLMs...

> [...] brings with it some genuine understanding: for any question or puzzle, there are usually only a few right answers but an infinite number of wrong answers. This forces the system to learn language-specific skills, such as explaining a joke, solving a word problem or figuring out a logic puzzle, in order to regularly predict the right answer on these types of questions. These skills, and the connected knowledge, allow the machine to explain how some-

> thing complicated works, simplify difficult concepts, rephrase and retell stories, along with a host of other language-dependent abilities[17].

Therefore, it seems that according to Turing or authors like Browning and Le Cun, the performance of the conversational machine (and more particularly of LLMs), if it is impressive enough, should most certainly be placed on the side of imitative behaviour. The machine would then be the imitative agent as such, and this would also be the reason why one should feel justified in saying that the machine is capable of (at least some) understanding (and is as it were on the path to full-blown understanding).

4 Large Language Models: A Case of Imitation Manufacturing?

4.1 Large Language Models Do Not Imitate and Neither Are They Imitations

How, then, can we address this issue? I want to defend the idea that LLMs do not behave in an imitative manner (LLMs do not imitate as such) but that neither are they imitations (of human beings talking). Their *textual productions* however, are imitations of human speech. Let us break this down.

Why cannot we say that LLMs behave in an imitative way (that they have what we previously labelled an *imitative behaviour*)? In order for an individual to imitate a particular behaviour (and thus for the imitation to be an *imitative behaviour*), it is necessary that the individual in question has in the first place a behaviour *of its own*, which is different from the one it seeks to imitate or reproduce. In other words, it must already possess a behaviour as such. To put it another way, imitative behaviour can only be understood against the background of behaviour that is *not* imitative (the imitative behaviour is always *extra*, or *on top of* a base behaviour, independent of the imitated target). The question then becomes whether the machines that are said to 'imitate' human beings in conversation have or do not have behaviours of their own, apart from these conversational ones. The answer is, of course, that they do not; talking machines *do not imitate*.

At this point, one could protest that LLMs do in fact have a behaviour, since it is necessary to go through what is known as the training phase in order to obtain interesting results, and that this training phase can sometimes be 'reinforced' by human activity to evaluate the initial results (a technique known as RLHF, *reinforcement learning from human feedback*). These two aspects would indicate that there is first an initial behaviour, followed—after the 'learning' phase and the 'human feedback' phase—by a modified behaviour; all this then strongly resembling a case of additional behaviour on top of an initial behaviour[18]. Arguably, this has the effect of transferring to the concept of behaviour what was our initial discomfort with the

[17] We shall return to this point in a moment, but it is striking that the two authors should emphasise precisely these very alleged linguistic (and understanding) skills, when these aspects are not always those in which LLMs perform best (and notably, it is often in these regards that the machines produce erroneous, unreliable or inaccurate responses).

[18] I would like to thank an anonymous reviewer for this objection.

concept of imitation. Once again, distinctions have to be made: ‘behaviour’ is certainly polysemous. In one sense a behaviour has to do with animality and can be understood as ‘the way in which some animal conducts itself’, and in another sense a behaviour simply has to do with what something does when what it does is not always the same. In this second sense, the whole universe, stars, volcanoes, plants, financial markets, stock options, political parties, governments, and so on can be said to ‘behave’ in a certain way (and therefore to have a behaviour). This simply means that some descriptions are applicable in certain conditions while others are not and that some other descriptions are applicable in other conditions, etc. One could say that this is a purely logical conception of ‘behaviour’[19]. It is of course in this second sense that one can rightfully say that an LLM behaves in a certain way (insofar as what it does is not always the same thing).

In an article entitled ‘machine behaviour’, Rahwan et al. (2019) draw analogies between the study of animal behaviour and the study of algorithmic functioning and operations, bearing in mind the ‘fundamental differences between machines and animals’. These fundamental differences relate in particular to the way in which the ‘behaviour’ of the individuals concerned is acquired. The authors note that whereas animals (be they human or non-human) develop a particular behaviour (in the first sense of the term) ‘for example, through imitation or environmental conditioning’, the ‘behaviours’ (in the second sense of the term) of artificial systems have nothing in common with them (i.e., animal ones), save analogically or metaphorically. Machine behaviour, the authors remark, is indeed ‘directly attributable to human engineering or design choices’. Machines therefore have no behaviour of their own in the sense of having an animal behaviour. To return to the heart of our issue, we should remember that in imitation (conceived as a process or a *behaviour*) there is first and foremost the identification of some thing (the target), and this identification presupposes (among other things that LLMs lack) an environment, sensory capacities for identification, motor capacities for reproduction.

If an LLM does not imitate, then is it not already attributing too much to the machine to say (as Bender, Gebru et al. do) that it is a ‘stochastic *parrot*’? Perhaps it is now easier to understand why the imitation of the machine is not at all equivalent to that which a parrot could achieve. As Montemayor (2021) remarks, a crucial difference between the action of the parrot and that of the machine is that the former is actually ‘participating in a joint activity’ which consists of mimicking the sounds it hears other beings make. The parrot does indeed have a behaviour of its own in the first place, which is why it makes sense to say that it *performs* the particular action of imitating human speech when it reproduces human sounds. Furthermore, as Montemayor also remarks,

> the production of sounds that mirror human language [is] based on its biologically grounded communicational capacities (birdsongs are a kind of communication, even though parroting human language is merely fake human communication).

[19] I discuss in greater length the consequences of a distinction between a logical and a substantive conception of action (distinction that can be paralleled in the case of behaviour) in Boisseau (2024).

The parrot can thus be said to 'communicate' whenever it imitates sounds, insofar as what it does can be considered a kind of conversational practice *for its species*[20]. As a matter of fact, vocal imitation is sometimes used by animals (especially birds) to indicate address to a particular individual (Balsby et al., 2012). It is thus by imitating or reproducing the call of a particular bird that a second bird can signal to the first that it is addressing it. It should however be noted that—when addressing a human being—the outcome of the parrot's production does not equate to that of a human being producing the same sequence of sounds: although the parrot can thus *communicate* in its own way, it cannot *speak* (i.e., it cannot communicate *in a human manner*). To put it too bluntly, 'speak' is a normative activity while 'communicate' is not.

It can therefore be said that parrots, contrary to LLMs, have a conversational practice of their very own. In contrast, LLMs have no proper means of communication (and no proper behaviour), neither have they got a *point* for communication (as parrots certainly have). All this entails that they cannot behave *in an imitative manner* and can ultimately only be tools to *produce* imitations. Thus from the point of view of the no-understanding position—which, it should now be clear that I also endorse but from a radically different standpoint—the comparison of LLMs with parrots is overly generous. Some aspects of the analogy are indeed relevant: the sentences produced by LLMs are not their own just as the sentences repeated by the parrot were originally those of another individual; but in other respects, the metaphor simply does not fit: parrots have an intentionality that LLMs lack, are grounded in a form of life and have a communicational behaviour of their own.

In the same way—this time considering the students analogy—let us first note how the comparison can indeed be helpful to see how, in a way, LLMs are similar to students mimicking their teachers in that they merely put together already formulated thoughts without understanding them. This idea is also put forward by Floridi (2023), when he remarks that...

> [...] in their capacity for synthesis, they [LLMs] resemble those mediocre or lazy students who, to write a short essay, use a dozen relevant references suggested by the teacher and, by taking a little here and a little there, put together an eclectic text, coherent, but without having understood much or added anything.

The point where the analogy breaks down[21] is, of course, when we observe that while the 'mediocre or lazy students' Floridi refers to do not understand what

[20] Indeed, as far as a parrot's communicative repertoire is concerned, researchers have compiled a list of around ten or fifteen different types of sounds regularly produced by parrots to communicate with each other (Bradbury, 2003). To mention just a few, parrots are known to make what researchers sometimes refer to as a 'preflight call', which, as the name suggests, is 'a specific call that is given by flock members prior to taking flight'. They also emit an 'agonistic protest' call during fights, and an 'alarm call' when attempting to warn of the presence of a predator.

[21] There is nothing unusual about the fact that the analogy may be relevant to a certain extent when describing what LLMs do, though its relevance is not total. Analogies, like metaphors—and thus following the definition of the latter given by Lakoff and Johnson (1980)—allow us to 'understan[d] and

they are saying *at that very moment*, there is otherwise much that they actually do understand. If they repeat what their teacher says without grasping its significance or meaning, the students nevertheless understand other things (as we pointed out above, they usually understand—in other contexts—what they are saying and, more generally, they understand that they are speaking, etc.). Similarly, as we just discussed, if the students are able to imitate or impersonate their teacher, it is because they are beings who *in the first instance* have a behaviour of their own. But because conversational machines are precisely designed *to produce* outputs that are imitations of human speech, this implies that there is not, in reality, anything they are or can be *apart from this*: devices *for imitating* (and not devices *which imitate*) human speech.

But while LLMs do not imitate, their status is nevertheless not that of imitations. We stated earlier that a particular painting could certainly be an imitation of another painting (if the former was created on the model of the latter), and also that a counterfeit banknote is an imitation of a genuine banknote (since it was precisely made in order to pass off as real money). An LLM, on the other hand, bears no resemblance whatsoever to someone who speaks: this is evident from the fact that an LLM is entirely disembodied. If an LLM were integrated into a robotic body, we could perhaps refer to the robot *as a whole* as an imitation of a human being speaking. But an LLM in itself cannot play this role. LLMs' *outputs*, however, can be seen as *imitations* and, in this respect, have a status similar to that of counterfeit money.

We mentioned earlier the concept of 'imitation manufacturing', which, at this point, can certainly help us understand the relationship between LLMs and the concept of imitation. Imitation manufacturing is not a form of imitative behaviour but is nonetheless a process of imitation: it is the process of creating something that can be described as an imitation—in the sense, this time, of a status of imitation. An LLM is then like a counterfeiting press, and the textual outputs of the LLM have a status similar to that of counterfeit banknotes. Like the counterfeiting press or the painter who reproduces a painting, the LLM is therefore engaged in a process of imitation (a process consisting of manufacturing imitations). However, unlike the painter, but just like the counterfeiting press, the LLM is not itself the agent at the root of the imitation. Although both the counterfeiting press and the LLM produce imitations, the source of this imitation (its *raison d'être*) is not to be found in the machines themselves, but in the machines' programmers and in the people who decide to use them. Insisting on this description is of particular importance because, just as there would be extraordinarily problematic situations from an economic point of view if there were suddenly a massive influx of counterfeit banknotes, there are extraordinarily problematic situations from a moral point of view if there is (as is the case today) a massive influx of speech imitations. In both cases, it is the very basis of

Footnote 21 (continued)

experienc[e] one kind of thing in terms of another'. Analogies operate in what is sometimes called a 'transfer' mode (Barnden, 2001): some aspects or characteristics of one thing (sometimes called the 'source') are transferred to another (the 'target'). But the transfer of characteristics from the source to the target need not be exhaustive: the value of analogy then lies in the fact that it helps to highlight a certain similarity in a number of aspects of two things, while not requiring the two things in question to be entirely similar or comparable.

trust that is being affected (trust in the banking institution in one case and trust in the textual production system in the other).

4.2 Large Language Models, ELIZA and Imitation

It is important to note that identifying a textual production as having this ‘status of imitation’ was certainly much simpler in the past, in older iterations of such machines (i.e., way before LLMs). When considering examples like ELIZA (Weizenbaum, 1966), PARRY (Colby, 1963) or even more recent conversational agents like Eugene Goostman—the 2008 winner of the Loebner Prize (Demchenko and Veselov, 2008)—one could then quite easily spot the tricks used to convey the impression that the machines understood the contents of the linguistic exchange taking place. The responses of these systems were indeed in a way already present within them, readily available to be utilised whenever relevant words were mentioned during the conversation. As is well-known, the responses of Weizenbaum’s ELIZA, for instance, relied on a set of pre-written generic answers associated with key words. If the user mentioned their mother, the program would thus respond with a pre-determined question about the nature of their relationship with their mother. Another tactic used by ELIZA was sentence transformation, where it would invert pronouns in a given sentence. For example, if the user stated, ‘I feel guilty about not taking better care of my little brother’, the program would respond by transforming the declarative sentence into an interrogative one, such as ‘Why do you feel guilty about not taking better care of your little brother?’ The rigidity of the early conversational programs therefore made it easier to reveal that they were not humans but machines—or, we could say, machines programmed to produce *imitations* of speaking human beings. In contrast, with the advent of LLMs, it becomes more challenging to see that the machines do not understand what is said and that their answers are not genuine answers, as there are no more ‘tricks’ as such: token prediction allows conversational programs to appear more human-like in their textual productions. The imitative character of the textual productions is thus less easily spotted. The fact that the machine produces imitations is then revealed, not through the rigidity of its answers—which are, on the contrary, generally quite diverse—but through the inconsistency of some of these utterances, or their downright erroneous character, what is sometimes misleadingly referred to as ‘hallucinations’ (Alkaissi and McFarlane, 2023)[22].

Furthermore—and this might also shed some light on why it can sometimes be tricky to dismiss LLMs’ outputs as imitations –, the answers produced by ELIZA, PARRY, and even Eugene Goostman can in a sense be seen as direct imitations of the way *specific* individuals or *types* of individuals would speak: respectively a Rogerian therapist, a schizophrenic adolescent, and a foreign child. These are, to use Richard Wallace’s—programmer of A.L.I.C.E, the three-time Loebner Prize-winning chatbot—terminology, ‘personality programmes’ (2008). In contrast, LLMs’ outputs are not usually imitations of the way a particular person or type of person

[22] For a collaborative collection of so-called LLM ‘errors’, see Davis et al. (2023).

speaks, but aim more at being *generic* imitations of the way human beings speak. Hence the suggestion by Millière (2021) of the already-mentioned term 'chameleons' to describe LLMs—a fitting metaphor in our sense as it serves the purpose of highlighting the difficulty of characterising LLMs' relationship with imitation: it is precisely because they are vastly more generic than previous models of conversational agents that it is more challenging to consider their productions as mere imitations of human speech. It is because they are trained on a very large and varied corpus that LLMs do not make us think of a particular person or type of person, but rather of a generic and *à la carte* individual.

Moreover, and as we were saying just now, it is not because LLMs are not *usually* used to manufacture imitations of anyone in particular, that they cannot *ever* be. On the contrary, it is entirely possible to create imitations of particular people or types of people using LLMs. In practical terms, this personification can be achieved by training the model on a set of texts with a very specific vocabulary and information content: this is typically of interest to companies looking to develop a chatbot that can generate responses containing specific language elements and information relevant to the company's business. Another example of the personification of LLMs is the direct imitation of a given individual. This is what Anna Strasser, Matthew Crosby and Eric Schwitzgebel (2023) set out to do with regard to the philosopher Daniel Dennett. They aimed at creating what they describe as a 'digital replica' of Dennett. The idea was then to find out to what extent the LLM's answers would resemble those that Dennett himself might produce when asked a philosophical question. Experts on Dennett's work who were asked to distinguish which of several propositions were really Dennett's and which had been generated by the LLM (in their case, GPT-3) only succeeded 51% of the time (Strasser et al., 2024). Then again, this in no way implies that the LLM is imitating Dennett; there is no question here of imitative behaviour. Instead, this simply shows that the programmers of the LLM (or the people at its instigation) have managed to create (to manufacture) satisfactory imitations of Dennett's speech—using the LLM as a tool to achieve this result.

5 Conclusion

The concept of imitation, which is very regularly used in discussions of AI (particularly with regard to conversational systems) but is rarely if ever the subject of detailed analysis, can help us—when examined closely—to have a clearer grasp of the capabilities of the artificial systems under scrutiny.

As a striking illustration, I first uncovered several trends in the literature, ranging from total rejection of the idea that LLMs have the ability to understand sentences, to complete acceptance of it, passing through a half-tone position in which LLMs are considered to have only a partial ability to understand sentences. I showed that these different theories are all keen to draw on the common concept of imitation, without ever questioning it.

I then drew distinctions between several aspects of the notion of imitation. The question I raised was that of establishing whether LLMs imitate or are imitations. I advocated the idea that LLMs do not have an imitative behaviour, and are not imitations either, but are better captured by the description I labelled 'imitation manufacturing' devices. To exhibit an imitative behaviour, LLMs would indeed first have to display a behaviour of their own, yet since LLMs are artefacts operating outside of any form of life, they cannot be expected to behave in any way whatsoever. LLMs' *productions*, however, are imitations (status), despite the fact that it is not always easy to see them as such (compared, for instance, with older conversational programs). That LLMs' workings exemplify an 'imitation manufacturing' process brings to light some of the moral issues at stake when LLMs become prevalent, in particular regarding the erosion of trust.

References

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus, 15*(2). https://doi.org/10.7759/cureus.35179

Aristotle. (1984). *Complete works of Aristotle. Vol.1 & 2: The revised Oxford translation.* Edited by *J.* Barnes, Princeton University Press.

Balsby, T. J. S., Vestergaard Momberg, J., & Dabelsteen, T. (2012). Vocal Imitation in parrots allows addressing of specific individuals in a dynamic communication network. *PLoS ONE, 7*(11). https://doi.org/10.1371/journal.pone.0049747

Barnden, J. A. (2001). The utility of reversed transfers in metaphor. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual meeting of the cognitive science society* (pp. 57–62). Lawrence Erlbaum Associates.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *FAccT'21: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623). https://doi.org/10.1145/3442188.3445922

Bennett, M. R., & Hacker, P. M. S. (2022) [2003]. *Philosophical foundations of neuroscience* (2nd ed.). Wiley-Blackwell.

Bergen, B. K., & Jones, C. R. (2024). Does GPT-4 pass the Turing test? *Proceedings of the 2024 conference of the North American chapter of the association for computational linguistics: Human language technologies*, Vol. 1: Long papers (pp. 5183–5210). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.290

Boden, M. A. (2016). *AI, its nature and future*. Oxford University Press.

Boisseau, E. (2024). The Metonymical trap. In B. Ball, A. C. Helliwell, & A. Rossi (Eds.) *Wittgenstein and Artificial Intelligence, Volume I: Mind and Language* (pp. 85–104). Anthem Press.

Bradbury, J. W. (2003). Vocal communication in wild parrots. In F. B. M. DeWaal & P. L. Tyack (Eds.), *Animal social complexity: Intelligence, culture and individualized societies* (pp. 293–316). Harvard University Press.

Brockman, J. (Ed.) (2019). *Possible minds: Twenty-five ways of looking at AI*. Penguin Press.

Browning, J., & Le Cun, Y. (2022). AI and the limits of language. *Noema*. https://www.noemamag.com/ai-and-the-limits-of-language/

Bryson, J. J. (2022). One day, AI will seem as human as anyone: What then? *Wired*. https://www.wired.com/story/lamda-sentience-psychology-ethics-policy

Burke, M. B. (1994). Preserving the principle of one object to a place: A novel account of the relations among objects, sorts, sortals and persistence conditions. *Philosophy and Phenomenological Research, 54*, 591–624. https://doi.org/10.2307/2108583

Button, G., Coulter, J., Lee, J., & Sharrock, W. (1995). *Computers, minds and conduct*. Polity.

Cerullo, M. (2022). In defense of Blake Lemoine and the possibility of machine sentience in LaMDA.

Chalmer, D. J. (2023, Nov 28). Could a large language model be conscious? Edited transcript of the NeurIPS Conference. https://arxiv.org/pdf/2303.07103.pdf

Cockburn, D. (2001). *An introduction to the philosophy of mind.* Palgrave Macmillan.

Colby, K. M. (1963). Computer simulation of a neurotic process. In S. S. Tomkins & S. Messick (Eds.), *Computer simulation of personality: Frontier of psychological research.* Wiley.

Cukier, C. (2022). Babbage: Could artificial intelligence become sentient? *The Economist*. https://www.economist.com/the-economist-explains/2022/06/14/could-artificial-intelligence-become-sentient

Davis, E., Hendler, J., Hsu, W., Leivada, E., Marcus, G., Witbrock, M., Shwartz, V., & Ma, M. (2023). ChatGPT/LLM error tracker. https://researchrabbit.typeform.com/llmerrors?typeform-source=garymarcus.substack.com

Demchenko, E., & Veselov, V. (2008). Who fools whom? The great mystification or methodological issues in making fools of human beings. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 447–459). Springer.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Human language technologies* (Vol. 1, pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1810.04805

Farmer, H., Ciaunica, A., Hamilton Antonia, F. de C. (2018). The functions of imitative behaviour in humans. *Mind & Language, 33*(4), 378–396. https://doi.org/10.1111/mila.12189

Floridi, L. (2023). AI as agency without intelligence: on ChatGPT, large language models, and other generative models. *Philosophy & Technology, 36*(1). https://doi.org/10.1007/s13347-023-00621-y

Fossheim, H. J. (2006). Habituation as mimesis. In T. Chappell (Ed.), *Aristotelianism in contemporary ethics: Values and virtues* (pp. 105–117). Oxford University Press.

Glock, H.-J. (2020). Capacities: Situated cognition and neo-Aristotelianism. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.566385

Goldman, A. I. (2005). Imitation, mind reading, and simulation. In S. Hurley & N. Chater (Eds.), *Perspectives on Imitation. From Neuroscience to Social Science: Vol. 2 imitation, human development, and culture* (pp. 79–93). The MIT Press.

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W., Friedman, D., & ... Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience, 25*, 369–380. https://doi.org/10.1038/s41593-022-01026-4

Gunderson, K. (1964). The imitation game. *Mind, 73*(April), 234–45. https://doi.org/10.1093/mind/LXXIII.290.234

Hacker, P. M. S., & Smit, H. (2014). Seven misconceptions about the mereological fallacy: A compilation for the perplexed. *Erkenntnis, 79*(5), 1077–1097. https://doi.org/10.1007/s10670-013-9594-5

Hampson, M. (2019). Imitating virtue. *Phronesis, 64*(3), 292–320. https://doi.org/10.1163/15685284-12341984

Hofstadter, D. R. (1995). The ineradicable Eliza effect and its dangers. In D. R. Hofstadter & the Fluid Analogies Research Group (Eds.), *Fluid concepts and creative analogies* (pp. 155–169). Basic Book.

Hutson, M. (2022). Could AI help you to write your next paper? *Nature, 611*, 192–193. https://doi.org/10.1038/d41586-022-03479-w

Kenny, A. (1989). *The metaphysics of mind*. Oxford University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Lemoine, B. (2022). Is LaMDA sentient?—An interview. *Medium*. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917

Locke, J. (1690). *An essay concerning human understanding* (1975) [1690]. Oxford University Press.

Luccioni, S., & Marcus, G. (2023). Stop treating AI models like people. Blog post at *The Road to AI We Can Trust*. https://garymarcus.substack.com/p/stop-treating-ai-models-like-people

Mahowald, K., Ivanova, A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2023). Dissociating language and thought in large language models: A cognitive perspective. Preprint. https://doi.org/10.48550/arXiv.2301.06627

Marcus, G. (2022). AI platforms like ChatGPT are easy to use but also potentially dangerous. *Scientific American*. https://www.scientificamerican.com/article/ai-platforms-like-chatgpt-are-easy-to-use-but-also-potentially-dangerous/

Michael, J., Holtzman, A., Parrish, A., Mueller, A., Wang, A., Chen, A., Madaan, D., Nangia, N., Yuanzhe Pang, R., Phang, J., & Bowman, S. R. (2023). What do NLP researchers believe? Results of the NLP community metasurvey. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st* *annual meeting of the association for computational linguistics* (Vol. 1, pp. 16334–16368). Association for Computational Linguistics. https://doi.org/10.48550/arXiv.2208.12852

Millière, R. (2021). Moving Beyond Mimicry in Artificial Intelligence. *Nautilus*. https://nautil.us/moving-beyond-mimicry-in-artificial-intelligence-238504/

Montemayor, C. (2021). Language and intelligence. *Minds and Machines, 31*(4), 471–486. https://doi.org/10.1007/s11023-021-09568-5

Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). Machine behaviour. *Nature, 568*, 477–486. https://doi.org/10.1038/s41586-019-1138-y

Romero, A. (2022). Why ‘is LaMDA sentient?’ is an empty question. *Medium*. https://albertoromgar.medium.com/why-is-lamda-sentient-is-an-empty-question-2683eac9d08

Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Ryle, G. (1945). Knowing how and knowing that: The presidential address. *Proceedings of the Aristotelian Society, 46*(1), 1–16. https://doi.org/10.1093/aristotelian/46.1.1

Ryle, G. (1949). *The concept of mind*. Hutchinson.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Strasser, A. (2024). On pitfalls (and advantages) of sophisticated Large Language Models. In J. Casas-Roma, S. Caballe, & J. Conesa (Eds.), *Ethics in online AI-based systems: Risks and opportunities in current technological trends* (pp. 195–210). Elsevier. https://doi.org/10.1016/B978-0-443-18851-0.00007-X

Strasser, A., Crosby, M., & Schwitzgebel, E. (2023). How far can we get in creating a digital replica of a philosopher? In R. Hakli, P. Mäkelä, & J. Seibt (Eds.), *Social robots in social institutions: Proceedings of Robophilosophy 2022* (pp. 371–380). IOS Press. https://doi.org/10.1111/mila.12466

Strasser, A., Schwitzgebel, E., & Schwitzgebel, D. (2024). Creating a large language model of philosopher. *Mind & Language, 39*, 237–259. https://doi.org/10.1111/mila.12466

Strawson, P. F. (1959). *Individuals*. Methuen.

Thomasson, A. L. (1998). *Fiction and metaphysics*. Cambridge University Press.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Varenne, F. (2019). *From models to simulation*. Routledge.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30 (NIPS 2017)*. https://doi.org/10.48550/arXiv.1706.03762

Wallace, R. S. (2008). The anatomy of A.L.I.C.E. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 181–210). Springer.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., ..., & Gabriel, I. (2021). Ethical and social risks of harm from language models. https://doi.org/10.48550/arXiv.2112.04359

Weizenbaum, J. (1966). ELIZA–a computer program for the study of natural language communication between man and machine. *Communications of the ACM, 91*(1), 36–45. https://doi.org/10.1145/365153.365168

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.