

2ND CYCLE OF STUDIES
MASTER'S DEGREE IN PHILOSOPHY - ETHICS AND POLITICAL PHILOSOPHY BRANCH

**DIALOGUE AS A PRODUCT
THE LIMINALITY OF CONVERSATIONAL ARTIFICIAL
INTELLIGENCE**

Leandro Ortolan

M

2025



Leandro Ortolan

TEXT ORIGINALLY PUBLISHED IN PORTUGUESE. THIS IS A VERSION FULLY TRANSLATED FROM PORTUGUESE TO ENGLISH BY THE DEEPL SOFTWARE. (IF YOU HAVE ANY QUESTIONS, CONTACT THE AUTHOR BY EMAIL leandro@ortolan.pro OR VIA THE WEBSITE www.ortolan.pro).

DIALOGUE AS A PRODUCT THE LIMINALITY OF CONVERSATIONAL ARTIFICIAL INTELLIGENCE

Dissertation carried out as part of the Master's Degree in Philosophy - Ethics and Political Philosophy, supervised by Professor Maria João Couto and Professor Domingos José Matos Sousa Faria.

[*Status: passed*] in a Master's Degree Test held on 11/April/2025 in the Salão Nobre of the Faculty of Letters of the University of Porto, with the jury chaired by Professor Paula Isabel do Vale Oliveira e Silva and with the following members: Doctor Tiago Bruno Borges Rodrigues Mesquita de Carvalho (examiner) and Professor Doctor Domingos José Matos Sousa Faria.

Faculty of Arts and Humanities (FLUP)

2025

*I dedicate this dissertation to Portugal,
for everything I've been able to learn, receive and give here.*

Index

Declaration of Honour / Declaration of Honour	6
Thank you.....	7
Abstract.....	8
Resumo	9
Index of Figures	10
Introduction	11
1.In search of true problematisation - historicity	16
1.1. How did we get here?.....	16
1.2. Relational emergence from a philosophical perspective	20
1.2.1. Aristotle and the emergence of the object	20
1.2.2. The emergence of the parmenidians	25
1.2.3. Descartes and the emergence of subjectivity	26
1.2.4. Kant and the emergence of intersubjectivity	29
1.2.5. Husserl and the emergence of intentionality.....	38
1.2.6. Heidegger and the emergence of dasein	40
1.2.7. Ricoeur and the emergence of narrative	42
1.2.8. A synthesis with the emergence of the "hyphen" as relational	43
1.2.9. The emergence of AI as relationships (or correlations).....	45
1.3. Ontological emergence from a technological perspective.....	49
1.3.1. About Ontology and ontologies	50
1.3.2. Knowledge representation models	54
1.3.3. What we teach algorithms	59
1.4. The many ethical problems	63
1.4.1. The map of ethical problems	64
1.4.2. A necessary ontological revision for the map of ethical problems.....	69
1.4.3. On artefacts and the emergence of teleological dissonance	73
1.5. What's this "here"	78
2.Dialogue as a product - the possibilities for AI	80
2.1. Where and when do we want to be?	80
2.2. The fourth dimension: capital as an intentional agent	81
2.3. The AI of the future through the transitivity of ethical problems	91

2.4. The AI of the future and privacy	94
2.5. The AI of the future and the validation of evidence	96
2.6. How do we want to achieve the future of AI?	109
3. Technological liminality - current opportunities	111
3.1. Why the injunction?	111
3.2. Whitehead and his philosophical-relational contributions	113
3.2.1. Concrescence as the existential complexity achieved by narrative	115
3.2.2. Prehension as data that can be qualified	116
3.2.3. Contrast as a narrative based on metadata	116
3.2.4. The synthesis mode of concrescence	117
3.2.5. Nexuses as ontological groupings with narratives	117
3.2.6. Subjective Forms as subjectivity itself, or even individual perspectives	119
3.3. Real entities as hermeneutic circularities	119
3.4. The alternative hypothesis of a metaontological dynamic	121
3.5. The way out of liminality by overcoming problems	127
3.6. The limitations of psychological models - The Big Five case	129
3.7. The Signo Project case: <i>design</i> , values and ethics	130
3.8. What does it all mean, anyway? From Thales of Miletus to AI... ..	135
Conclusion and Final Considerations.....	138
Bibliographical references	142

Declaration of Honour / Declaration of Honour

I declare that this thesis/dissertation/report is my own work and has not previously been used in another course or curricular unit, at this or any other institution. References to other authors (texts, works, ideas) scrupulously respect the rules of attribution of authorship and are duly indicated in the text and bibliographical references, in accordance with the rules of referencing. I am aware that the practice of plagiarism and self-plagiarism constitutes an academic offence.

I hereby declare that this thesis/dissertation/report is of my authorship and has not been used previously in another course, degree, curricular unit or subject, at this or any other institution. References to other authors (statements, ideas, thoughts) scrupulously respect the rules of attribution and are duly indicated in the text and bibliographical references, in accordance with the rules of referencing. I am aware that the practice of plagiarism and self-plagiarism is an academic offence.

I also declare that I have not used generative artificial intelligence tools (*chatbots* based on large language models) to carry out part(s) of this thesis/dissertation/report.

I further declare that I have not used generative artificial intelligence tools (chatbots based on large language models) to carry out part(s) of this thesis/dissertation/report.

Portugal, Porto, 24th March 2025

Leandro Ortolan

Thank you

Since it would be inappropriate to thank all the philosophers since Thales of Miletus, I would first like to thank Philosophy - a consoler for me, just as it was for Boethius. Inadequate as it would be to thank all the philosophers, it would diminish the importance of those to whom I will address my most sincere thanks below:

To my supervisors, Prof. Dr **Maria João Couto** and Prof. Dr **Domingos José Matos Sousa Faria**, who, by allowing me unparalleled dissertative freedom, led me down paths that in fact surprised and enriched me;

To my **colleagues in the Signo Project**, for their immense contribution in terms of teaching, as well as the freedom they gave me to put forward my ideas, which were always considered and respected:

To Prof Dr **José Francisco Preto Meirinhos**, for his generosity in sharing his knowledge and allowing me a unique learning opportunity and personal self-worth by joining the Signo Project: he was "the" difference for me;

To Prof. Dr. **João Carlos Martins Rebalde** with his comments in the work discussions, which always brought immense *insights* to my research; to Researchers Dr. **Mário João Rosas Rebelo Correia** and Dr. **Celeste Maria Lourenço da Silva de Oliveira Pedro**, who always proposed immensely relevant articles, some of which are even referenced here; to fellow doctoral student **Irandina Afonso** who, even on the eve of her doctoral thesis defence, was always contributing productively;

To Prof Dr **Ricardo Manuel Coelho de Melo**, senior researcher at Fraunhofer Portugal AICOS and head of the Signo Project, for the warm welcome I received; I would also like to thank the entire **Fraunhofer Signo Project team**;

And finally (and always) to **all the FLUP professors** who have been part of my journey, giving their contributions and giving rise to deep and unconditional gratitude in me, as well as affection.

Due to life's circumstances, it was only possible for me to dedicate myself to this endeavour thanks to financial support from the **Foundation for Science and Technology (FCT)**, for the research grant for the Signo Project, and from the **Directorate-General for Higher Education (DGES)**, for the scholarship for the first year of research. Therefore, my commitment to delivering something of value to society, to repay the investment made, became a first-order imperative for what I was pursuing.

So I'd like to say THANK YOU, and I hope I'm giving back more than I received, and it won't end there: my effort was immense and my commitment was total - and I'll continue to think and do this, to give my best for as long as I can.

Abstract

This dissertation proposes a new metaontological model as a way to overcome the most relevant and challenging ethical problems of today. It argues that there is a teleological dissonance between what is expected of an AI and what it can offer, constructively. It proposes that such dissonance arises from a philosophical tradition in which there is a predilection for extracting a fragment of reality, to the detriment of valuing the analysis of complexity itself, as given. Thus, what results from this whole movement is the liminality of the current technological era, in which we are going through a period of complex transformation and potentiality, when much of what was stable and predictable is becoming the opposite: unstable and unpredictable.

The first part of the dissertation addresses the path from the past to the present. The second part considers what the future of AI will be like, if taken from the perspective of overcoming current problems, without configuring them as fictional projections. We will make such projections based on existing technological resources or others that are considered viable and feasible to be used in algorithmic constructs. And finally, the third part deals precisely with liminality, this present that refers not only to temporality, but also to spatiality - in terms of limits and scope

The conclusion argues that an ethical and responsible AI, free from current problems, requires a deep understanding of human complexity - and the respective representation of the existential modal - so that it is possible to apply the most appropriate values in all stages of technological development. The final consideration, which summarises the entire argument, is about the elaboration of a dynamic metaontological protocol, capable of capturing and emulating the relational complexity of reality. This feat is presented as necessary, but perhaps not sufficient, for the creation of this AI of the future that will surpass current demands.

Key-words: artificial intelligence; metaontology; knowledge representation; relational ethics; value-sensitive *design*.

Resumo

Esta dissertação propõe um novo modelo metaontológico como forma de superar os problemas éticos atuais mais relevantes e desafiadores. Argumenta que há uma dissonância teleológica entre o que se espera de uma IA e o que esta pode oferecer, construtivamente. Propõe que tal dissonância se dá a partir de uma tradição filosófica na qual há uma predileção pela extração de um fragmento da realidade, em detrimento de prezar pela análise da própria complexidade, como dada. Assim, o que resulta de todo esse movimento é a liminaridade da atualidade tecnológica, na qual se atravessa um período de complexa transformação e potencialidade, quando muito do que era estável e previsível está a tornar-se o oposto: instável e imprevisível.

A primeira parte da dissertação aborda o percurso do passado ao presente. A segunda parte considera como será o futuro da IA, se tomado a partir da superação dos problemas atuais, sem configurar-se como projeções ficcionais. Faremos tal projeção com base nos recursos tecnológicos existentes ou outros que sejam dados como viáveis e exequíveis de serem utilizados nos constructos algorítmicos. E, por fim, a terceira parte trata justamente da liminaridade, deste presente que não se refere somente à temporalidade, mas também à espacialidade – em termos de limites e alcances.

A conclusão defende que uma IA ética e responsável, isenta dos atuais problemas, exige uma profunda compreensão da complexidade humana – e a respetiva representação do modal existencial – para que seja possível a aplicação dos valores mais convenientes em todas as etapas do desenvolvimento tecnológico. A consideração final, que sintetiza toda a argumentação, é sobre a elaboração de um protocolo metaontológico dinâmico, capaz de capturar e emular a complexidade relacional da realidade. Tal feito é apresentado como necessário, mas talvez não suficiente, para a criação desta IA do futuro que ultrapassará as atuais demandas.

Palavras-chave: inteligência artificial; metaontologia; representação do conhecimento; ética relacional; *design* sensível aos valores.

Index of Figures

FIGURE 1 - "EVOLUTION" OF THE MACRO-ONTOLOGICAL TERNARY OF EPISTEMOLOGY: SUBJECT, HYPHEN AND OBJECT AND THEIR RESPECTIVE "SIDE EFFECTS", BUT PRESENTED AS "CAUSAL".....	44
FIGURE 2 - MAP OF ETHICAL PROBLEMS, ACCORDING TO MITTELSTADT ET AL. (2016).....	65
FIGURE 3 - MAP OF ETHICAL PROBLEMS (ONTOLOGICALLY REVISED AND RENAMED), BASED ON FIGURE 2.	70
FIGURE 4 - ALGORITHMIC FLOW SCHEMATISED FROM MITTELSTADT ET AL. (2016).....	71
FIGURE 5 - REPRESENTATION OF KNOWLEDGE ABOUT THE IDEAL AI MODEL, FUNCTIONALLY SCHEMATISED ACCORDING TO THE REFERENCED ARTICLE (MITTELSTADT ET AL., 2016).....	71
FIGURE 6 - REPRESENTATION OF KNOWLEDGE OF SOME OF THE MOST COMMONLY CITED ETHICAL PROBLEMS IN THE AVERAGES, ALLOCATED WITHIN THE CLASSIFICATION ADOPTED IN THE PREVIOUS FIGURE.....	92
FIGURE 7 - FRAMEWORK USED TO SUPPORT THE ELICITATION OF RELEVANT VALUES IN THE DYNAMICS BETWEEN STAKEHOLDERS.	134
FIGURE 8 - CARDS OF THE POSSIBLE RELEVANT VALUES USED AS RESOURCES, BASED ON THE QUESTIONS POSED BY THE DESIGNERS.	134

Introduction

This dissertation aims to go beyond the theoretical limits of philosophy by proposing to serve as an appropriation and *praxis* for the various sciences related to information technologies, especially those that are dealing with the challenges posed by artificial intelligences (AI), and more specifically conversational AI. The term "conversational AI" is being favoured over "generative AI", as the latter is a procedural term that describes how artificial intelligence generates natural language. In contrast, conversational AI not only integrates generative AI, but also adds intentionalities that promote interaction and the building of a relationship with the user. In addition to this supposed overcoming, it is also intended that philosophy can capture what the Information and Technology Sciences have to offer, as elements that will serve as the basis for a more philosophical theorisation, as a contribution to the enhancement of AI as a means of progress.

To this end, we have adopted a measure of informational supplementation in the many notes that will follow and appear when appropriate, always seeking to collaborate so that these writings can move through interdisciplinarity - so necessary for what this dissertation will seek to propose and defend, which is a metaontology structured on the knowledge of the humanities sciences, including psychological, philosophical, sociological and many others.

Why do we call it "meta"? Because it's basically an ontology of ontology. To summarise, we can consider "meta" as a process of reduction. So, for example, the theory of metalanguage has developed a process of reducing language by seeking to establish which words are themselves irreducible to others. And this set of irreducible words forms metalanguage. The "meta", as in the case of language, is therefore always a complex process that transcends the object of study itself, as in the case of language, given as cross-linguistic, when it comes to analysing values and other questions of signification (Goddard, 2018; Goddard & Wierzbicka, 2011). Ontology, as we will develop over the course of this dissertation, can be considered, from here on in, all the effort and organisation to classify or represent whatever exists in a digital dimension: to represent knowledge, things, people, the various manifestations or subjective,

objective, intersubjective dimensions (relationships) or even ideology or symbolic order, and everything else, as if the existential modal were placed as an object of representation so that, in the form of data and metadata, they "existed" in the digital dimension.

In the case of language, this process of reductionist investigation of meaning became relevant from the Enlightenment onwards, especially with the philosophy of Leibniz, who argued that "if nothing can be understood in itself, nothing can be understood". (Goddard, 2018). We will therefore seek to propose the same *Leibnizian* arguments, in which all ontology can and should be reduced to meta-ontological presuppositions that not only define them, but explain them and perhaps give them meaning. In this way, we believe that we - or the next generations - can overcome the serious problems of the lack of theories of data correlations.

For this reason, and in the name of a necessary interdisciplinarity, there are the many explanatory footnotes operating as an interface between philosophers and information scientists. Such notes, if ignored, will not take away from the argumentative essence of the dissertation, but considering them will imply new dimensions, applications, cases, developments and, why not, a certain degree of provocation on related topics that always present a different perspective. In proposing "dynamism", we favour writing that is equally dynamic and, as far as possible, relational: pragmatic, up-to-date and also "imaginative"

Dynamism here means not only avoiding isolating a problem from its context in search of a controlled and paralysed situation in which to apply analytics under ideal conditions of "temperature and pressure", but also seeking to decipher the complexity from which the problem emerges. If there is dynamism, however, there could be several dynamic levels, or even chaos. But if we consider that chaos itself has both potential and action, it must have elements that can be common to everything that has also been chaotic somewhere. After all, the hypothesis that "chaos is an order yet to be deciphered" (Saramago, 2002) could be posed as a question.

Imagination is used here in Castoriadis' (2000) conception as a fundamental element for understanding history and society. A radical imagination, given as society's capacity to create new ways of being and meaning, when it is capable of breaking with the mere repetition of the past to consciously construct its own history - would be an unprecedented victory for reason, and no longer for a rational or even irrational spirit, but for consciousness, therefore - a victory for us. If we agree with the author that imagination is not an individual process, but a social one, because it is intrinsically linked to the social dimension through the creation and sharing imaginary or symbolic meanings that shape society, institutions and the very experience of individuals (Castoriadis, 2000, pg. 268), we are betting that such readings can serve the purpose of breaking away from the shackles of sameness, even with all the risks inherent in the consequences.

Based on the above, we have structured this dissertation as follows:

In Section 1, we consider what has brought us this far (from the past to the present) and therefore work to problematise the current analysis of the ethical impacts of AI on society. We argue that the traditional approach, focussing on objects and ignoring relationships, results in a limited and inconsistent understanding of the problematic we are seeking to perceive, and thus doesn't even succeed, given the unconsidered teleological dissonance between the problems and AI today.

We have gone through a part of the History of Philosophy - specifically Epistemology - to demonstrate how the emphasis on static and immutable elements, which we call Parmenidean, makes it difficult to analyse a dynamic and relational system like AI. We argue that we are therefore anachronistic in analytical terms, lacking a necessary understanding of AI as a narrative given from a set of relationships.

Therefore, in Section 1, we carry out a necessary critique of the "Parmenidean" heritage, emphasising the elicitation of relationships as the focus of analysis; we defend the need for a relational ontology capable of capturing the dynamics and complexity of the relationships between data, algorithms and users; we highlight the importance of metadata as a provider of layers that could lead to ways of capturing the complexity of

information and the nuances of the relationships between different elements; we value the intentionality of capital as a determining factor in the development of AI; based on the above, we analyse and verify the inconsistency of the map of ethical problems when it ignores the teleology of AI and the intentionality of capital; and finally, we call for the need for an ontological adjustment between the current and the digital, integrating the objective, intersubjective and subjective dimensions - considering the teleology of AI and the intentionality of capital as a fourth present and influential dimension.

In Section 2, we consider what we want from here, based on overcoming the ethical problems of AIⁱ (from the present to the future without problems) which will be further and inevitably (if we don't have the tools of defence) shaped by the expansionist and influential intentionality of capital .ⁱⁱ

Therefore, in Section 2, we realise the existential dimensions of dialogue being adopted as an end product and as a means, in a dual role as an end for entertainment, information or companionship, as well as a means for capital to achieve other ends, such as collecting data, manipulating behaviour or selling products; we accuse a transitivity in the most cited ethical problems about AI, as they frustrate any attempts to attribute an origin to certain isolated or merely static factors; we argue the need to design an AI that operates on the basis of data classified in real time, interacting with it in constant ontological operations and responding to user demands instantaneously; we address the issue of privacy from a perspective in which it will gradually lose relevance; we

ⁱ The use of AI in the plural will indicate a reference that transcends conversational AI, to encompass other "types" of AI, not just restricted to generative ones, but mainly those that behave as agents or decision-makers in diverse processes such as automation, predictions or any specific tasks.

ⁱⁱ Capital, here and hereafter, in the playful attempt to personify it as an intentional agent of capitalism. For this reason, it is worth establishing that although the problems that will be attributed to capitalism here, through the mention of capital, may make some criticisms seem condemnatory, this will not be the intention or even part of the aim of the dissertation. We will try to establish a *stricto sensu* analysis of some of its effects, which we argue emerge from the fusion between human interests and the symbolic order that expresses them. We argue that capitalism is not an autonomous entity, but a reflection - and one of the most authentic - of human intentionality. Capital (or capitalism) is not a creator, but a "creature": one of many that we conceive, fight and attribute agency to, without realising that, deep down, we are only fighting what is all too human in our own imaginary or symbolic projections.

reiterate the need for validation and accountability of the evidence generated by algorithms.

In Section 3, we consider how we are, and we argue that it is a liminality, which consists of being an intermediate dimension, spatiotemporally undefined, and given between two other dimensions that, although defined and still "present", are no longer integrated as before (the analogue past and the digital future), which makes this liminality result in a phase of uncertainty but also of transformation and unique opportunities with a more predictable future regarding technological impacts.

Therefore, in Section 3, we defend our technological liminality to be configured in this transitional phase; we rely on the concepts of Withehead (2010) to appropriate his philosophy of the organism and promote a theoretical framework for dynamic and relational metaontological modelling; and we thus reach a conclusion that justifies the need for a dynamic metaontology, capable of integrating at least the objective, intersubjective and subjective dimensions of the user of a conversational AI, as well as the intentionality of capital; we combat the importance attributed to the opacity of AI; we extol the importance of value-sensitive design to ensure that AI is developed in an ethical and responsible manner, which involves the participation of different *stakeholders in the design process*, as we will cite applications carried out in the practical experiments carried out by the Signo Projectⁱⁱⁱ, which could well serve as a *benchmark* for the above; by being given as an example of developing a value-sensitive, human-centred design approach.

And finally, we present our conclusions and final consideration, based on the dissertative developments.

ⁱⁱⁱ Sign. Value systems in digital health technologies. Programme: FCT / Carnegie Mellon Portugal - 2022. Ref. 2022.15724.CMU. Participating institutions: Institute of Philosophy of the University of Porto, Fraunhofer Portugal Research - Centre for Assistive Information and Communication Solutions and Carnegie Mellon University. (On the website <https://ifilosofia.up.pt/projects/signo>, accessed on 19/Mar/2025)

1. In search of true problematisation - historicity

1.1. How did we get here?

In this first part, we will summarise some of the most influential steps that will allow us to understand and seek one or more positions (or answers) in relation to the problems currently being raised about AI.

We will begin by understanding how we are analysing the ethical problems of AI, based on the historical development of "analogue" philosophical thought: powerful, effective, which has served us well: it has "brought" us this far and will still serve us. However, resorting exclusively to such analytical forms leads us to believe that they are insufficient to resolve many of the current questions about AI, since they are based on theories and causalities, epistemology and mostly first-order logic. They are therefore insufficient in the face of an eminently digital scenario in which everything, or almost everything, happens through correlations, without the established causal issues of theorising - analogue and digital are different worlds and dimensions, therefore, each with their respective burdens and bonuses, but equally resistant in terms of deep cognisance about their ultimate realities.

To summarise, we will explore the different possible relationships between how a user perceives and relates to an AI, and vice versa. Because relationships, whether as causes or effects of "evils" or "benefits", are always necessary. To this end, we will first explore the conceptual aspects of such "relationships", from an epistemological and ontological hermeneutic, and then explore the phenomenological concept of AI from the tentative constructive point of view of the algorithms, and their impacts, especially those of an ethical nature.

Thus, we need to declare that we will address the problems mapped (Mittelstadt et al., 2016) and claimed, punctually, one by one, and integrating them if necessary. Despite

this, we will adopt a suspension of moral qualitative^{iv} prejudices about the supposed effects or impacts that an AI could (or could still) have on humanity, as has most commonly been disseminated in the political spheres and in the media in general recently, considering debates on issues such as privacy, responsibility and other widely discussed topics. And we will do so by considering the very instrumentality of an AI. For anything that is a new technology, or a new development of applications, there can be both benefits and harms. That's why we'll limit ourselves to analysing the harms without considering that the cause can be eliminated, because AI isn't something that will be extinguished and therefore we need to disregard the fact that resisting it is a viable option, because it isn't, just like everything else. Even in terms of legislation, there are some complications already realised about how to legislate^v on AI. Collaboration is about

^{iv} For the very capacity that an AI is supposed to have, for example, to acquire strategic powers to influence behaviour, such possibilities are already (prejudicedly, perhaps because of the fear generated by ignorance about them) considered evil (just as they fatuously imply that an AI acts like an evil genius).

This can be deduced, for example, not only by considering certain aspects of the possible loss of autonomy of human agency when it is hampered by AI in general, but also by being at the mercy of unknown, hidden and unpredictable intentional forces.

It also prejudices the very freedom of action, as if all human actions and behaviours were truly autonomous, in the name of overvaluing what is commonly considered to be the autonomous freedom of agency, or free will, already entering into other discussions that would not fit into our objectives defined here.

But even so, by insisting on considering any and all supposed algorithmic influences as "bad", there are cases (and not so specific ones) that can serve as counterexamples in which this prejudiced generalisation can be called into question.

We can mention using the possible influential powers of AI to stimulate "good" attitudes (such as openness to adherence to medical treatments, mental health benefits, motivational interactions to improve performance and well-being, educational / instructional interventions in the ethical or civic fields, for example).

Or even to mitigate "bad" or harmful behaviour (as in the case of identifying crises arising from pathological mental states or mediating suicidal ideation, for example).

Thus, the same supposed capacity for negative influence of an AI would clearly be considered positive (even more so if applied proactively, or predictively).

All of this, however, is another discussion that wouldn't fit here either: about the limits and directions of supposed algorithmic influences.

For this reason, the research question will really be about the foundations of an AI's supposed ability to influence its potential to cause effects, impacts, and not about the relative or absolute predication of these, nor even about its usability.

^v The discussions are taking place in the various legislative, judicial and executive spheres of power, as well as in the press. The European Parliament has made significant progress at world level, but still falls

creating ideas to legislate and raise awareness, prevent and monitor, and that's the direction we're heading in.^{vi}

We will start with an introductory conceptual review, inspired by and adjusted from a similar argumentative consideration of digital hermeneutics by the philosopher Romele (2019), certain that it will provide readers with the necessary critical elements, without presupposing that they are already familiar with the concepts that will be evoked, or not. For this reason, and with the intention of cultivating an interdisciplinary approach, we are investing in a robust appendix with the appropriate complements and references for those who want to broaden their knowledge, without having to have a background in philosophy. In this way, we will discuss the most relevant criteria of the conceptualisations that will be used in the pages to come, albeit panoramically, to trace a crescendo of philosophical thought through a synthetic timeline that stretches back to the beginnings of the search for knowledge. These criteria correlate with the innate human need to know things, in fact and in rationalised abstractions.

It's not hard to see this from our own experiences, actions and behaviours, just as Vergil also saw it: "Happy is he who has succeeded in knowing the causes of things", because knowledge fosters the human desire to reach a (transcendent) level that is always beyond the current (immanent) limiting condition of possibilities, or even their precarious finite existential situation: and so we will be entering (intentionally, and ontologically) Epistemology, a pioneering and relevant part of Philosophy that deals, at first impression, with knowledge and, more intimately, deals with achieving that human satiety so secretly dreamed of: transcendental knowledge. And so we will enter with the

far short of reaching a consistent level to be followed. Reference accessed on 19/Mar/2025 at <https://www.europarl.europa.eu/topics/pt/article/20200918STO87404/inteligencia-artificial-oportunidades-e-desafios>.

^{vi} What will be sought, therefore, will be a supposed set of characteristic elements - in ontological and/or adjectival terms - that mark the point of view, or perspective, that a certain average user has (or can have, or acquire) in relation to an AI, either passively, interacting with technologies indirectly or unconsciously, or even actively, in conversational action, when they start interacting directly and consciously. After all, one or other of these situations can alternate for the same user and, therefore, they are not mutually exclusive.

aim of realising how knowledge can be obtained from instruments, if we can thus consider an AI, instrumentally, in relation to knowledge. We're considering transcendental knowledge because it's not just about knowledge as an end, but as a means to achieve other ends. And AI, we will argue, reflects precisely this as an attempt to overcome our human limitations.

So much so that Vergilio attributed the happiness of knowledge to subduing "under his feet all the fears, the implacable fate and the roar of the greedy Acheron"; in other words, knowledge was seen as a means to an end. We'll see that Foucault also sees it this way.

An end to the "good", most of the time, whatever this "good" may mean in practical terms, but it has brought us to where we are, as humanity, with unimaginable progress in all fields of technological, social, economic, political, cultural, human development and many others, such as the current and still surprising AI. But not all developments have had only the "good" as a result, because there are always side effects, and not always desirable ones.

Such a utilitarian (and moral) discussion, we reiterate, will not be the central theme of this dissertation, however, as already announced. We are not trying to promote or demean AI, since it is not a future project, but rather a fact, current, and already inserted in our lives, directly or indirectly. AI is the result of our endeavours, which have always been present, ever since the discovery of fire, the invention of the wheel, or the development of agriculture, and we have never stopped developing instrumentally towards an *optimum*.

We will probably leave both the optimists and the pessimists of technology frustrated, eager to find peers to corroborate their beliefs and join forces against or in favour of AI. If progress is to blame, we also have to "blame" philosophy, since it has been making significant contributions to scaling knowledge for millennia. All we have to do is realise that Epistemology, a veteran of Philosophy, was the beginning (and the medium) of a movement (which we already assume to be intentional) that culminated in all the knowledge we have today, and which continues to move forward: the foundation of the

Sciences, in their multiple forms of being, came from it, which cannot be denied or disregarded.

AI is no stranger to this historical progression, and came from it, from this human endeavour of ours to know, to go beyond existing limits, and this is what we will explore from now on. So, in the next section we'll see how AI relates to this omnipresent philosophical quest: epistemological and "transcendental".

1.2. Relational emergence from a philosophical perspective

What does it all mean? Thomas Nagel emblematically asked this in the title of his book^{vii}, in which he tries to establish what philosophy is all about, in a panoramic way. And the title of the book is probably precisely the primordial question that has always identified us as humans, as rational: we are questioners about everything and eager for knowledge as a way of transcending our own condition and nature. But this craving is not just for "this", but for "all of this". And it is on this long road between "this" and "all this" that we will reconstruct. And we will do this reconstruction through a historical-philosophical hermeneutic.

1.2.1. Aristotle and the emergence of the object

Since antiquity, and driven by the Aristotelian tradition, the first philosophical concerns about knowledge were primarily directed at objects: they sought to know what objects were in their particular manifestations, in other words, in the things themselves, such as wood and the wooden chair, for example. As a matter of priority, they sought to know what the world or universe was as a whole, but almost exclusively from its causal constituents "essential" or particular (Shields, 2007), its classifications, organisations, origins, principles, characteristics and many other questions that were external to the philosopher^{viii}, The philosopher positioned himself as a distinguished observer, totally

^{vii} Nagel, T. (1995). *What does it all mean? An Introduction to Philosophy*. Gradiva. ISBN 978-972-662-421-9

^{viii} However, since the classics, philosophers have referred to the human soul and its forms of relevance. In Aristotle, his considerations of the soul took place through studies of the soul itself, and also psychology, "where he concentrates mainly on the soul and its capacities for perception and thought" (Shields, 2007, p. 33). What is being established in relation to concepts throughout the History of

detached and independent from the object and the world, completely dislocated, autonomous and inquisitive of what he could perceive sensibly, and without assuming any kind of mediation, responsibility or protagonism in relation to what was (or seemed to be) merely "external" to him.

Aristotle^{ix}, with regard to his theory about the most "noble" raw material of being, or the highest level of its ontological hierarchy, known as the essential substance, attributed to it a certain form to consider it as a mode of being, and of existence, as if it were an unalterable and incorruptible principle in existential terms (Shields, 2007, p. 49). An essential substance would, for him, always be the same essential substance, no matter how it is presented, there it will be inserted from certain considerations .^x

What ontologically differentiates such substantial presentations, for an observer, is that their essentiality is inherent to the identity of the substance itself: the formal essence is something that makes it what it is, in a way that is inseparable from itself. The accident is everything that is superfluous to it (like colour, for example), and that doesn't impact on its form (or way) of being, but only on its form (or way) of existing (when the existence of being can take place in one way or another). In other words, roughly

Philosophy, it is assumed, is not a specific and limited line of themes attributed to each philosopher cited, but rather their centrality, in terms of relevance, in the conceptual approach throughout the ages.

Aristotle can even be considered the forerunner (among many other sciences, such as Psychology itself) or founder of Ontology (Shields, 2007, p. 146), by establishing theories, definitions and structured classifications about being, based on certain categories and hierarchies established in which he could represent knowledge about being.

^{ix} "Human beings began to do philosophy, says Aristotle, just as they do now, because of admiration, at first because they marvelled at the strange things right in front of them, and then, advancing little by little, because they began to find greater things intriguing." (Meta. i 2, 982b12 apud Shields, 2007, p. 36).

^x For example, a man will always be a man, with a rational essence and certain forms, and even though his qualities may vary, he will always be a man. For this Aristotelian substance, however, can present itself in an essential way (with the addition or decrease of its qualities) and/or in an accidental way (Shields, 2007, p. 99).

speaking, accidents are their formal attributes: they explain, describe, differentiate, but they don't cause, define, substantiate or standardise.

Aristotle, therefore, according to Shields (2007), dedicated himself to perceiving and classifying the intimate relationship of objects, and this involved prioritising substantiality in a search for causes. If there is a cause for something, must there be a cause for that cause? And this question led him to theorise about a first cause that would be the origin of all causal relationships, because he was interested in establishing how everything was related until something was what it is. This first cause, according to Aristotelian logic, should consist of absolute terms, in full act without any potentiality to be added or subtracted from it. This first cause "was", in fact, the greatest and absolute existential dimension. Being, therefore, is ontologically superior^{xi} to appearance, to its own seeming to be (Romele, 2019, p. 55). Accidental appearance was already considered less important, therefore, because of the formal priority it gave to the true substantial dimension of being. "Being" has established itself as a central ontological status ever since. Aristotle (Lear, 1988) conceptualised the immobile motor as the first cause, which was presented as what causes everything, albeit indirectly (through its own immobility, because it is in full act and therefore leads everything that has any potentiality - that is not in act - to move towards it).

Thus, this first reference in Aristotle will serve as a paradigm in our endeavour to outline an ontology^{xii} of the ontology of knowledge.

Why is Aristotle still current and relevant to our AI theme?

1. By modal: necessary (essential, substantial) and contingent (accidents, attributes).

^{xi} It could then be considered that attributes were weak forms of being, at least for Aristotle (Romele, 2019, p. 55). Given the immense Aristotelian influence or also their own interests, which can be selectively subversive, this is still the case for many thinkers when they consider attributes to be less relevant and secondary, compared to the essential and primary substantiality of the object.

^{xii} In the next section, we'll discuss what we're considering about the term ontology, which, in short, seeks to merge both what is considered in Philosophy and in the Information Sciences.

The modal distinction between what is necessary (essential, substantial) and what is contingent (accidents, attributes) harks back to a philosophical tradition that, in Aristotle, made it possible to understand what constitutes the essence of a being as opposed to its accidental attributes. In the context of AI, this differentiation is crucial for interpreting the relationship between what exists and its representation. While the necessary elements correspond to what defines the identity and robustness of the data - the underlying "essence" that gives it meaning and structure - the contingent aspects refer to the variable characteristics that, although they influence appearance and functionality, do not alter the informational core. Thus, when modelling reality, AI systems must consider both aspects to ensure that digital representations capture not only the substance of phenomena, but also their mutability and the contextual details that enrich the overall understanding of "reality".

2. By developing the adoption of ontology as a form of knowledge, which will be relevant in our approach to data^{xiii} in general, to *big data*^{xiv} and, more particularly, to metadata.^{xv}

^{xiii} Anticipating what is to come in this dissertation, the inversion that occurs from Aristotle's theory of gradually promoting attributes to protagonism (and understanding the dynamics of how this happens) is precisely what has been at stake since the most recent philosophical epic began in order to understand what is happening in the digital dimension, especially when it comes to the subject of AI, which only "exists" due to the correlations it makes between *inputs* and *outputs*. In order to understand the *modus operandi* with data, in short, in all its growing quantitative exponential that feeds AI, and which has culminated in the gigantism and incommensurability that lives up to the ontological status called *big data*, we have started to go through the data not by the data itself, but by its attributes, by its relational forms of classification based on certain representations that are also made on the basis of attributes, which are called metadata.

^{xiv} Roughly speaking, we can imagine *big data*, at its limit, as the accessible "universe" of data:

"Despite the importance of the phenomenon, it is unclear what exactly the term 'big data' means and therefore what it refers to (...) The term 'big data' refers to large, diverse, complex, longitudinal and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams and/or all other digital sources available today and in the future. You don't have to be an analytical philosopher to find this obscure and vague." (Floridi, 2012)

^{xv} Metadata are, therefore, the attributes of data, or their accidents, which today, contrary to Aristotle's theories, have come to define data, but which do not necessarily explain them. In order to explain, as formal epistemology dictates, it is necessary not just to correlate, but to realise the existing causal relationships and prove them to be true, justifying them. And there is almost always no causal relationship

The adoption of ontology as a form of knowledge is central to the organisation and interpretation of data in AI, especially in the context of big data and metadata. Aristotle pioneered the structuring of knowledge by categories, and this approach is still relevant for giving meaning to the vast volumes of information generated digitally. Ontology makes it possible to create conceptual models that structure data in a hierarchical and relational way, enabling more precise inferences and a better understanding of the connections between elements, which is essential for algorithmic decision-making.

3. Because of the digital age, which is promoting the primacy of attributes and thereby amplifying^{xvi} massively old ethical problems.

Because the digital age has promoted an inversion (and even a subversion) in the hierarchy between essence and attributes, favouring superficial and contextual characteristics over substance. AI algorithms categorise and interpret data based on patterns derived from contingent attributes, often ignoring the ontological depth of the phenomena they represent. This primacy of attributes intensifies existing ethical problems such as algorithmic bias, depersonalisation and manipulation of information.

in the correlations that generate AI *outputs*, as we'll see later. *Outputs* are mere replications of what is considered to be the most probable, within the many possibilities of probability that exist among the available data.

^{xvi} There are practical problems arising from this primacy of substantiality over its attributes, however, regarding the applications of these hierarchical-classificatory premises. Such as when each human clearly perceives what a human is, for example, but some of these humans perceive themselves to be more human than others, which forms the basis of racism, which is one of the most relevant ethical problems today. In these cases, the racists' attributes, such as skin colour, are thus closer to a particular essentiality that they believe they are part of, rather than a general accidentality. There is therefore always the possibility of selective subversion, when it suits certain groups or perspectives.

And, as it were, there are also different values, weights and measures for each essentiality and accidentality, for different judgements, in which everything needs to be (and will be) agreed. But that doesn't mean it will be resolved. Later in this chapter, we will discuss these aspects in Kant. So, what we're trying to show is that the relationship between the primary and secondary categories, at hierarchical levels of representation, is conflicting and difficult to explain due to the difficulties of realising the limits between them. And this is what probably already concerned Aristotle when he made him consider the psychological forms of perception and impressions in the conceptualisation of the *anima*.

The lack of a more rigorous ontological approach in AI can lead to unfair decisions and the perpetuation of structural inequalities.

4. By overcoming the Parmenideans (in the following section), since AI is based on the dynamism^{xvii} of (cor)relations, and therefore static analyses fail.

Artificial intelligence systems don't just analyse isolated data, they identify dynamic patterns, adjust to new contexts and evolve over time. Understanding AI therefore requires a philosophical model that recognises the relational and transformative nature of reality.

1.2.2. The emergence of the parmenidians

So that we can base a critique on the inconsistent ways in which the problems with AI are being developed, we will form a set with all the theoretical elements that refer to immobility in the various theories we will present here. Thus, Aristotle's immobile motor will be moved, for the time being, and still in isolation, to this set. Other elements will soon join it, also because they have the same referential function of immobility, and in the foundational centrality of the theories in which they participate, since they are necessary for the rest of the theory to remain consistent.

We will group all these immobile references, one by one, in this set inaugurated by the immobile motor, which we could well call the Parmenideans, in fair honour of the supposed precursor (or promoter) of immobility in Philosophy: Parmenides.^{xviii}

^{xvii} In the end, everyone perceives according to their own interests, capacities, conveniences or ideologies. Or by beliefs, desires or wishes, for example. The internal relationship of qualifications, which are merely mental or empirical, without external references such as conventionalised standards or established models, is even more problematic. That's why, in these epistemological labyrinths, relationships have always been enticed to be subordinated to something immutable, fixed and always present as a reference. Value has come to lie in the "safe" and "reliable" reference, given by supposed stability or immobility.

^{xviii} Parmenides' philosophical foundation lies in the disparity between the "way of truth" (*altheia*) and the "way of opinion" (*doxa*). The former leads to an understanding of true being (which he describes as one, immutable, immovable and indivisible) - and from here we take our "homage" to him. Parmenides developed the theory that movement and change would presuppose a kind of non-being, which would be inconceivable to him, as it would imply something ceasing to be what it was in order to become something else: which, obviously, it was not. And since non-being was inconceivable and non-existent for Parmenides, he concluded that movement and change would be illusory, since they would never actually

1.2.3. Descartes and the emergence of subjectivity

Almost two millennia after Aristotle^{xix}, René Descartes (2005) consolidated the transition of focus that has progressively taken place since then, at the threshold where consciousness began to assume some existential and relational relevance, subjectively, changing the epistemological protagonism of the object to that of the subject, at a psychological level, or mental, as Newman (2023) considers:

Certainty/indubitability for Descartes is psychological in character, although not merely psychological - it is not simply an inexplicable feeling. It also has a distinctly epistemic character involving a kind of rational insight. During moments of certainty, it is as if perception is guided by "a great light in the intellect" (Med. 4, AT 7:59, CSM 2:41).

Descartes thus considered, by adopting doubt (viscerally existential) as a method of investigative questioning, even the very existence of the questioning subjective cogito. Doubt was thus legitimised by allowing the cogito to "continue" to exist even if, radically, nothing else could exist beyond itself: In other words, Descartes' theory even considered, in the extreme, that all that "exists" would only be the thinking mind, and nothing else (Fodor, 1980), although he himself combated, albeit in an unusual way, the radical "evil" that his theory brought, from one of the concepts (even if it came from a side effect) that has been most highly combated^{xx} and despised in the entire History of Philosophy: solipsism.

occur. The immobility of being was deduced as a logical consequence of its unity and immutability. Being must be "all the same", without any differentiations, and therefore cannot alter or move. The "way of opinion" is the illusory understanding of reality, merely empirical and lost in the unknowable multiplicity of the sensible world (Heidegger, 2008).

^{xix} In this, our first exposition, and in the rest, we have to consider that the *gaps*, as supposedly occurred between Aristotle and Descartes, are merely illusory. They would only exist if, obviously, nothing else had been theorised between the two philosophical moments considered, which was not the case, as many philosophers contributed to these apparent gaps. Therefore, the existence of such gaps is not an evoked fact, but only an illusion that may occur.

^{xx} And, progressively to Aristotelian, Cartesian and all the other thoughts, philosophy evolves, like everything else, although not always assertively, in terms of undeniable advances. Because there are the functional and necessary objections that arise and so, with each new proposal or consideration, the complexity of understanding increases, to the point that sometimes it seems to be moving further away

It wouldn't be rash to assume that Descartes' attempt was ultimately to isolate the mind from all external relationships, except with itself, without trying to understand how these relationships occur and, therefore, when he couldn't solve the problem of uncontrollable and unpredictable dynamics, or rather chaos, he sought a fixed reference point for this^{xxi}. And this, in a nutshell, culminated in the same Aristotelian problem of considering only the internal relations of objects, without being able to establish how these relations took place^{xxii}, starting from themselves, which led him to look for a first

from some conclusion rather than closer to it. And everything happens just as it did with Descartes, who brought advances to philosophical thought, but also left serious side effects after the scrutiny of his objectors, who were not few. And this is peculiar to all proposals for innovative theories, and is not something that is necessarily harmful to philosophers. The problems occur, from our hypothesis, because they are also due to the facts that relationships in theories are generally either disregarded or assume a marginal position.

The problems of neglecting relationships occur not only in philosophy, but also in mathematical research, according to Eugenia Cheng (2020, ch. 6):

"(...) we have seen how crucial it is to consider systems as a whole, with all their interactions, rather than just isolated people or events. The idea of considering things in relation to each other is one of the basic principles of modern maths. But it wasn't always this way, because it was only relatively recent research, from the mid-20th century onwards, that brought it to the forefront. We see that observing how things or people relate is often the key to understanding a situation, rather than looking at the intrinsic characteristics of those things or people. This is true on many levels and on different scales, from the way countries interact in the world to the way people interact in relationships."

^{xxi} These collateral or dialectical effects are added to the previous ones, all stemming from each new philosophical theory presented, such as the aforementioned Cartesian solipsism, for example, which focussed on subjectivity in order to solve existing problems. It makes it seem as if the supposed substantiality of the thing itself (or the "x" in the equation, allegorically) is always further away when you're moving in search of it. It is sometimes perceived a priori, abstractly, but never in a satisfactory, satiating way. Because it makes it seem as if the resulting rational abstract is always a philosophical effort inherent in an idealised static understanding of a world that is pure movement and transformation, sometimes with a chaotic appearance, breaking any paradigmatic linear theory that you might want to establish about it.

Relationships are generally undervalued, avoided and neglected because they are unstable and not so easy to ontologise. However, we will come to realise that the current problems in AI are essentially relational: not realising that (and how) "everything" comes about because of the relationships between its "parts". It's as if it were a denial of circularity by merely despising it, as if the problem of circularity would disappear. However, despising something doesn't necessarily imply its extinction, because despising it doesn't necessarily make something despicable. And if it does, it doesn't cancel it out from being what it is, from acting as it does: the old question between necessity and contingency, therefore.

^{xxii} And so we start to consider a whole without all its parts. A "whole" that is not a whole, and parts that are not enough to justify the "whole". And therein lies the most circumspect facet of circularity, the same facet that has led hermeneutics to be constantly criticised and to a role that, we would argue, should be

cause external, giving life to its functional equivalent of the Aristotelian immobile motor. Descartes soon tried to fix his problem, but without having to create "something" new, because he used an existing concept and fixed it not as a first cause, necessarily, but as a first relationship, nonetheless.

There is therefore a limitation to a current re-reading of AI specifically, both in Aristotelian and Cartesian philosophy, which is the tendency to focus on the internal properties of objects without fully understanding how their external relationships occur and influence their dynamics. We have seen that Aristotle analysed entities based on their essences and accidents, but his theory lacked a mechanism to explain the interactions between them without having to resort to the immobile motor. Similarly, Descartes, in seeking a fixed reference point for the mind, initially isolated it from the external world without fully considering how this relationship occurred. Thus, both ended up creating solutions that depended on an external and absolute principle to sustain the coherence of their systems.

And so Descartes did, dramatically seeking god as a relational reference and differentiated from the cogito. The Cartesian god is significantly differentiated from the Aristotelian immovable mover, but not so much, in terms that are functional to the concepts presented^{xxiii}. So, in the set of Parmenideans, the Cartesian god joins the hitherto solitary Aristotelian immovable mover.

more relevant, not as a problem, but as a protagonist. The circularity evidenced and exposed, as in hermeneutics, could therefore turn out to be one of its greatest merits, and not the opposite, as we shall see: "it is important to examine interactions as vicious circles" (Cheng, 2020).

^{xxiii} The Cartesian problems are the same as those of Aristotelian theories, in that they don't focus on the multiplicity of relationships without them having immovable external referents. The difference in Descartes - and, we believe, an interesting advance - is that he changed the investigative domain from the object to the subject - subjectivity emerged, after all, through the cogito. This being the case, and going straight to the even more problematic Cartesian effects generated (and which seem to be "hotter", perhaps precisely because of the greater proximity to the source of the problem), by his cogito, reified, albeit through an indigestible derived dualism, also derived other problems beyond solipsism, but which will not be dealt with here, as they transcend the dissertative objectives assumed to be central.

1.2.4. Kant and the emergence of intersubjectivity

To corroborate all the efforts towards philosophical evolution, Immanuel Kant (2008) emerges, who can be considered, at least in this journey of ours, as the most relevant divider between the philosophical focus oriented towards the object and the focus on the subject, when he presents the concepts of phenomenon and number, expanding and completely surpassing Cartesian thought beyond Cartesian sceptical-radical psychologism:

Thesis of the "objectivity of objects and contents": neither the objects nor the contents of our cognitions have identities that are exhausted by the individual acts of representation that involve them, so that both the objects of our cognitions and the ways of relating to such objects (i.e. the contents of our cognitions) have a position that is not (in general) dependent on any of our actual mental activities. (Tolley, 2012)

The Kantian impossibility attributed to knowledge of the thing itself - of the noumenon, therefore, is relevant, as it reduces and actualises the true possibilities of knowing something (existing, actual) and endows the subject with the responsibility (and exclusivity) of themselves having the sufficient and necessary rational mental structures to perceive the thing that presents itself sensitively to them, to the very subject capable of knowing: no longer as the search for the noumenon, but for the phenomenon. From then on, significant importance began to be given to relationships. From a contemporary hermeneutic perspective, if an object exists (and is), then this Kantian existence is exclusively due to the subject perceiving it in such and such a way, much more like Aristotle, through the essentialities and accidentalities of forms, than in relation to Descartes, through self-consciousness (Wood, 1975)^{xxiv}. Although the author does not name or classify the hermeneutics in question, we are led to believe that he is referring to phenomenological hermeneutics, the closest to the tradition of Heidegger and Gadamer. For emphasises that knowledge is not a direct capture of objective reality (as

^{xxiv} "Kant rejects the Cartesian conception of self-consciousness because he maintains that thinking is not a spontaneous intellectual self-intuition, but an activity that connects the representations (whether of internal or external origin) that are given in a sensitive and receptive intuition." (Wood, 1975, p. 603).

in classical realism), but an interpretation mediated by the subject and their cognitive structures. This approach is in line with the Kantian idea that we can only know phenomena (and not numbers) and reinforces the primacy of subjective relationships in the constitution of meaning, contrasting with Cartesian rationalism and coming closer to the Aristotelian ontology of forms and accidents. The existential-epistemological protagonism of the object thus gives way to the protagonism of subjective relationships. But Kant (2008) didn't just stick to objects, and nor could he. New questions (or conjectures) always arise with each theorisation. In this case, however, they were about the multiplicity of subjects: How is it possible that something one subject knows can be the same (exactly, in the sense of identity) as something another subject also knows, given that perceptions can differ from one another, as can judgements and capacities? Even if they are common to all human beings, this perceptual communion must be verified, because even if they don't differ from each other, we must realise that they are convergent. For there are no a priori ways of "measuring" or "equalising" subjectivities, after all, to give them as adjusted, guaranteed. Kant (2008) resolves (or at least strives to resolve) this issue by resorting to two distinct conceptualisations, but they are related to each other, relationally, which is already an immense advance in philosophising. The first is about objectivity. The second is, albeit indirectly, about universality. And, as we have seen and said, the further we go, the more complex it becomes and the further we seem to be from some kind of completeness, but in this case, Kant inserts some relationships that will shed light on understanding. (Tolley, 2012)^{xxv}

Objectivity (what "exists", in a meaningful way), therefore, from a more attentive, or less naive reading, comes to be understood as a result of subjectivities that relate to each other and agree on certain results as such, as we mentioned in the section on Aristotle. Kantian things end up being, simply put, what they are, because they are thus

^{xxv} "What Kant seems to have in mind in emphasising the 'objectivity' of cognitions is the fact that the objects they 'have' (or are intentionally directed towards) are objects that are available to be the objects of other acts of cognition as well. Now, if this were true, then since an object of cognition itself could not be identical with any of the acts of cognition that represent it, it would follow that Kant endorses the act/object distinction (...) (Tolley, 2012, p.9)"

agreed between those who experience and rationalise them, jointly or not. And this is something that replaces such unfeasible "measurements" or "equalisations" of subjectivities. Subjectivity, therefore, is not considered to be an artefact that can be operated in a controlled manner. There is a clear separation between objectivity and subjectivity in Kant, unlike Aristotle in his *anima*, who was unable to clearly separate the boundaries between essentiality and causality. Thus, such conventions about objectivity pass through the Kantian subjective processes, which consist, more or less, of a rational sifting of the categories of knowledge, which are given by sensibility through the aforementioned experiences, and all a priori (Tolley, 2012, p.10)

There is an undeniable aesthetic in this Kantian theory founded on rationality that can be applied to everything, even aesthetics, beyond ethics and epistemology itself. Right,

"The intellect does not move without desire, without the exercise of the desiring faculty. (...) The telos, the ultimate goal of thought is intellectual intuition, a figure, as we have seen, of non-discursive thought" (Wood, 1975.

The concept of the Kantian a priori is really seductive, as it elevates rationality to an almost superhuman level, to merge with the desirable, or even the divine, by being able to reach a complete understanding. But even if it is something aprioristic, it is still a convention that requires relationships in order for "something" to be named or referred to as "something"^{xxvi}

It is the relationships (which are established as our conceptual protagonists) of subjectivities that result in certain forms of meaning about (the objectivity of) the world and things, based on these interactions. This, then, leads us to intersubjectivity, which is this kind of convention shared and consensually accepted between the various

^{xxvi} Because for each subjectivity there are also certain structured and necessary relationships, obviously, since there is no admission in Kant of the occurrence of the refuted Cartesian solipsism: "the solipsism of a subjectively centred reason must be dissolved in the intersubjective structures of communicative reason" (Keul, 2002).

subjectivities that relate to each other about something, or rather, about the very subjectivities that objectify the world.^{xxvii}

Kantian objectivity as such, therefore, cannot be concluded before these intersubjective relations (which operate as signifiers), because the objective status results from this, as such, since it is intersubjectivity that gives life and guarantees the subsistence of meanings. Thus, Kantian subjectivity requires and presupposes, throughout its philosophy, an intrinsic capacity to perceive in other subjectivity, in otherness, and what is common in intentional terms, through rational paths and, therefore, supposedly free of evil.^{xxviii}

The intersubjective moment is shown in Kant's statement at the beginning of the Critique of Pure Reason (2018, A XI, p. 5), which is no longer seduced, that is, without meanings. 5) that "the mature judgement of the time by apparent knowledge" that "the mature judgement of the age, which no longer allows itself to be seduced by apparent knowledge", in other words, without meanings, as he adds that his proposal is to give primacy to reason in order to "once again undertake the most difficult of its tasks, that of knowing itself and setting up a court that will assure it of its legitimate claims and, on the other hand, condemn it of all unfounded presumptions". What, after all, is a tribunal other than the search for a shared consensus about something?

^{xxvii} More recently, philosophers of language, including Gottlob Frege, already referenced by Ricoeur, would point out, in other ways, that such processes are part of linguistic meaning, for example. Because language is intrinsically associated with intersubjective questions of signification, but not just linguistic ones. Thus, "the condition of possibility for understanding can now be understood concretely as the condition for genuine communication." (Keul, 2002). A theme we'll return to shortly in section 2.

^{xxviii} Because Kantian rationality is coated (or perceived) with a certain level of non-infectious purity (or even aseptic, disinfectant powers) in the exercise of intersubjectivity. Although Kant doesn't use the term intersubjective directly, this presumption of the intersubjective process is central to his philosophy, which always aspires to the common, the shared and consensual, to the best that is originated by rationality at its highest level and, therefore, to the good, which becomes both immanent and transcendental, since it is first and foremost the intrinsic and extrinsic rational. And human rationality is an attribute that is presented as the best that emerges from each interactive subjectivity through the common sense that they fulfil.

As Kant (2018, B766, p. 596) wrote , in the *Doctrine of Method*:

"It is even on this freedom that the existence of reason rests; reason has no dictatorial authority whatsoever, but its decision is nothing other than the agreement of free citizens, each of whom must be able to express his reservations and even exercise his veto without hindrance."^{xxix} .

And, as a result of common sense being taken as incontestable, the very dimension formed and shared by subjectivities comes to be perceived as a universal. Or rather, such universality passes^{xxx} , or must necessarily pass, through the sieve of common sense. Universality in Kant, although it seems independent and autonomous of any restrictions, is only "recognised" in this way when it is subjected to relations of intersubjectivity .^{xxxi}

^{xxix} The culmination of the idea is found in Section 40 of the Critique of Judgement, where Kant introduces the *sensus communis* as a "sense common to all" that introduces the "maxims of common human understanding", unfolded in their practical application in anthropology (Keul, 2002).

^{xxx} Although universality was not ostensibly defended by Kant either (Oberst, 2015), it was perceived by his commentators and readers, such as Hegel, who replicated it in almost everything he wrote, such as dialectics, historicity and phenomenology. This perception of universality by commentators and readers is itself an exemplification of what Kant considered about intersubjective processes. And it is a relationship internal to the subject, reduced to it, in the end.

^{xxxi} The universality derived from Kant, therefore, is first and foremost an ontological status, or rather a meta-ontological status. But that's not exactly how his readers generally perceive it, since it progressively takes on an important role and autonomy, as if it were always existing as a noumenon and somewhere became a phenomenon, messianically, through such rational and intersubjective processes, such that these were the medium for universality itself to manifest itself. Or perhaps the convention of universality has only "evolved" and transformed over time. Somehow, universality became reified and then deified. Even so, and coldly, for "everything" there is, and even for the "revelation" of universality, it is first necessary to pass through the sieves of subjectivities and intersubjectivities, because everything that is endowed with meaning results from this process of the convention of knowledge which, according to Kant himself, is practically limited to experienced phenomena. There is a circularity here, just like in hermeneutics, and this requires special attention.

Even so, despite the fact that Kant called for agreement between free citizens as a condition of rationality, he did not manage to pave the way for the full conception of intersubjectivity. His real concern is with the harmony of the various faculties of the mind, and therefore with the subject's relationship with himself, and not with the condition of the possibility of agreement between subjects. It is a kind of internal consensus that he has in mind, which in turn is anchored in the correspondence theory. Thus, "truth depends on agreement with the object and, in relation to this, the judgements of each and every understanding must therefore be in agreement with one another (*consentientia uni tertio, consentiunt inter se*)." (Keul, 2002, p.256).

For these reasons and fragilities, provocative questions such as these arise: if, in order to consider the occurrence of universality, there is a need for some prior consensus to be established, what would prevent

So even when Kant appealed to a "sense common to all", he actually had in mind no more than a mental experiment whereby we put ourselves in "every other man's place" by "comparing our judgement with the possible judgement and not with the actual judgement of others". However, we're not talking about a purely private act. After all, the aim is to guarantee a kind of universality for the judgement of taste, which is based on the "communicability of sensation". This capacity should point in the direction of intersubjectivity (Keul, 2002, p.257).

Universality would therefore be a part, and not the whole, after all^{xxxii}. But the induction to universality is undoubtedly the substratum of a weakness in Kantian theory, just like Cartesian solipsism, despite all the faith placed in human rationality. The point is not to recognise or condemn theoretical weaknesses, but to find the ingredient (or perhaps more than one) that is missing from theories in general (which won't even be a *spoiler*: the disregard for the leading role of relationships^{xxxiii}), so that this ingredient can be

it from being considered as such, even at the risk of some consensus being the result of a collective hallucination about something? In other words, is it possible for any "universality" to arise from a group (albeit partial and isolated) that collectively hallucinates?

For what is being evoked in relation to intersubjectivities are merely relationships that take place and that, from these, signify the more general and supposedly more necessary objectivities. It is valid to say that for any relationship there is always a domain in which it takes place. The domain, in this case, is objectivity itself, which recursively contains the universality thus considered.

^{xxxii} For in such domains, too, there is always supposed to be a finite set of participants, since there are objective existential limiters to be considered: and humans are a limited set of existents, totalling less than nine billion at present (Accessed on 06 August 2024 at <https://www.worldometers.info/br/>). A universality "exists" from and as a function of this total of humans, if considered in this way. But these humans are not even universal, because in their particulars they exist in a tiny fraction of time and space of a supposed whole that is still immeasurable, indecipherable and unknown. If logical rigour is applied, the conception of universality is nothing more than a merely inductive process given in the partiality of what has existed and exists, and even so in the partiality of what can be observed. Obviously, there are argumentative resources both in favour of and against the "absurdity" of such relativistic, almost nihilistic hypotheses, as well as for all the others.

^{xxxiii} What if this ingredient were relationships? Because relationships are an important (perhaps indispensable) part of life, of existence, and consequently they should be part of any theory (as representational forms) about reality. Paralysis is not part of anything actual and can even be considered an abstraction, because everything is transforming, creating or destroying itself, and nothing is absolutely paralysed or immutable. And so we infer that relationships must (or should) be perceived and theorised as such, circular, recursive, as presuppositions of transformations, of spatiotemporal action, but not only. They are also intrinsically unstable, changeable and themselves change reality. They are, in Kantian terms, the universality of uncertainty. That's why theoretical formulations always avoid dealing with unpredictability and instability, and prefer to consider even the problem of inducing universality as acceptable in the face of relational uncertainty.

added in the right dose for what is currently being sought in digital^{xxxiv}. This ingredient, we argue, will be a good dose of dynamic and contextual relationships, rather than an abstract search for universality or certainty. In the digital context we want to achieve, with mostly positive values and risk control, this means improving the ontologisations and correlations between data, people and systems so that there is an accurate representation of reality and knowledge.

Static concepts (for example, those of the Parmenidean set, which already includes the immobile motor in Aristotle, the god in Descartes and now universality in Kant), when introduced into theories that should be relational, can lead to error, or to absurd extremes, resulting from something that is conceptually coloured to be a bandage, which can be very useful for small injuries, but which is not suitable for an open fracture that requires much more than that in terms of repair. Theoretically, if there is something static in the places of relationships, there are obvious weaknesses in the theoretical structure that will lead to it cracking due to the immobility of some of its components that should be dynamic, and no matter how brilliant the theory is, it will eventually collapse. It may even adjust itself for a while, but as soon as the first criticisms are made, it ends up becoming inconsistent, outdated or even generating new problems. What's

But recently, when trying to formulate theories about digital data, in these unusual computerised times in which correlations are becoming more relevant than causalities, they are assumed to be *outputs* (which are decisions in themselves, when considering a possible agency for AI), and which, progressively impacting the masses, can soon be assumed to be consequences or "universalities". And this process leads us not only to fear (caused by impotence in the face of such power), but to the certainty that there really is a re-ontologisation of the world, and this even seems sensible in certain respects, very easily observable with regard to the political-ideological polarisations that have come to be perceived in many countries, or in the promotion of instant personalities previously seen as bizarre, or even absurd or anti-scientific ideas, and much more that was previously absurd and socially unacceptable and which has become "normalised" by the directions of the algorithms distributing content, but which we won't go into here, as we insist on keeping the focus on the problem itself.

^{xxxiv} For a change from the pharmakon example, a recipe with the right ingredients is supposed to be tastier and more digestible. And you can "make a mistake" in a recipe by adding too many or too few ingredients, in both quantitative and qualitative terms. The theory of intersubjectivity, if it were viable without reference to universality, would be much less indigestible, just like Cartesian solipsism without the existential reference of God.

worse is when there's no consideration of the relationships properly represented in the theories, and that's when there's an even more serious problem. ^{xxxv}

"For this reason, the intersubjectivity of communication in everyday discourse is constantly disturbed. To the extent that it exists, it is because consensus is always theoretically possible. But insofar as consensus remains a requirement, true intersubjectivity fails us." Genuine communicative action between the two partners would only be realised if each of them adjusted their contribution in a way that called for agreement, so that they could move towards agreement by rational consensus. (Keul, 2002, p. 257).

This possible distortion of reality was probably one of Kant's concerns^{xxxvi} when trying to create a theory that combines intersubjective convention with the (messianic) magnanimous universality revealed by relationships based on rationality (and therefore

^{xxxv} There could be a scale that could give the scale of the problem of a theoretical model, proposed here merely speculatively, ranging from:

- (1) insufficient (theories that disregard relationships),
- (2) fragile (considers some relationship),
- (3) inconsistent (sustainable, which in addition to considering some relationship, is still supported by parmenidians) and
- (4) consistent (self-sustaining, without crucial Parmenidean references, and based on a circular, recursive, atemporal and totally relational nature).

A viable proposal to overcome these limitations of Parmenideans would also involve inferentially finding ways to represent transformational movement based on relations in viable ontological models of knowledge, which could result in new explanatory and causal theories. And such a proposal, if made feasible, seems to be a key issue for resolving the current problems arising from AI, especially those of an ethical nature, which we will soon be addressing.

In Kant's case, an undesirable effect caused by the fragility of his impressive theory would be to descend into complete relativism, given the (harmful) possibility of a multiplicity of subjectivities signifying the world in a possibly disconnected way. In these digital times, in which the practice of disinformation is both present and worrying, this (exactly) illustrates one of the incredible consequences of such counterexamples when groups of people (who even start to organise themselves into antagonistic groups with active voices) fight each other on social networks, such as the aforementioned political polarisation which, truth be told, is not new, but potentially extreme, due to the current media facilities that lead to the easy overexcitement of polarities, amplifying them) seem to create their own "universalities", which seem absurd to the "opposing" group, and seem even more absurd to those who don't yet adhere to any of the groups and still have some minimally functional rationality with them.

^{xxxvi} Kant, however, made immense progress in this field of represented relations, more than before, and thus raised, but without resolving, the question of their relevance to theories of knowledge.

also supposedly based on the "good", which, incidentally, operates very closely to the Parmenideans).

He didn't consider (solving) intentionality, however, as he also made it clear that such a claim would be unsolvable (Tunhas, 2023), because he constituted intention as unfathomable precisely from the singularity of each subjectivity, through the alterities that necessarily need some intersubjectivity in order to be coherent. He established a foundational abyss about this impossibility, in the domain of freedom. ^{xxxvii}

Kant was very relevant in presenting and emphasising intersubjectivity, in which lies the dynamism of the relationships between subjects and objects, or between subjects and subjects. And these relationships are currently situated, more in the light of conversational AI, as the basis for all the developments to overcome the problems that will be addressed in the third section. After Kant, it became clear that it would be necessary to delve into the issues inherent in intersubjective relationships, adding as a further ingredient a few pinches of intentionality, established as the result of intersubjective processes, in other words, meanings in movement, in transformation, based on relationships and no longer of paralysis, as everything happens, phenomenologically. Habermas^{xxxviii} was one of these philosophers who explored and

^{xxxvii} In line with the Kantian unfathomability of intention, Paulo Tunhas (2023, p. 91-92) updates, synthesises and argues that "the meaning of other people's beliefs can only be partially captured", as he states that what we have, in relation to other people's intentions, are merely quasi-representative beliefs, since "purely representative beliefs depend heavily on the condition of exteriority". Thus, he refutes the principle of exclusively internal relations on the grounds of their fragility and lack of referentiality (which resulted in such Parmenidean references, as we have seen) and adopts a model that he calls internal-external, about the relations that we have been discussing. Tunhas (2023) further reinforces this direction and quotes Kant: "internal experience in general is only possible through external experience in general" [B 278-279].

^{xxxviii} Jürgen Habermas was one of the philosophers who set out to find ways to overcome these Kantian limitations by promoting intersubjective relations:

"Rather than providing rules to govern the internal capacities of the mind, Habermas differs by directing his attention to that social space in which speech, praxis and reason come together. The patterns he aims to articulate are patterns that guide dialogue partners both towards understanding the issues and towards a better understanding of what they themselves think." (Keul, 2002, p.257).

significantly extended Kantian conceptualisations. A little before Habermas, however, this exploration of subjective experience was also a central theme of Edmund Husserl's proposal in phenomenology.

1.2.5. Husserl and the emergence of intentionality^{xxxix}

According to Alves (2022), the major change introduced by Husserl, even under the aegis of the universality bequeathed by Kant, was to focus on the experience of the intentional subject in its relationships, and not to direct the philosophical focus either exclusively on the subject or on the object. Alves also claims that Husserl opted for the protagonism of intentionality, which is precisely the relevant part that operates and is operated by subjectivity, in other words, it is the part with which subjectivity relates as intersubjectivity, now functional as a kind of intentional interface. What Husserl attempted, in fact, was to explore intentionality in its various dimensions as subjective conscious and experiential relations.

^{xxxix} Husserl's intentionality is conceptually fundamental to phenomenology. It describes the direction of consciousness towards an object: consciousness is always conscious of something. To summarise, we can explore intentionality, based on Husserl and other conceptual developments, in the following points:

It doesn't just refer to the conscious dimensions, but also to what are generally called the "unconscious" ones.

It is contained in lived experience. Merleau-Ponty included pre-verbal thought in it - the pre-personal dimension of bodily intentions and meanings (the acting body always understands its situation and its own possibilities, long before we pay explicit attention to it).

It's a relational process: it's not just a unilateral and exclusive act of the subject towards the object. Instead, it is a double (inter)action, (inter)experience and (co)constitution. The subject and the object mutually mould each other in lived experience, and we'll explore these operations further when we come to Whitehead.

It is fundamental to creativity when it occurs as an act of encounter, and this creative encounter is characterised by intentionality. The human creator intentionally engages with the world, and the creative process derives from this intensity of the encounter.

It is linked to the search for meaning (and therefore to language). Intentionality is a constituting and sharing of knowledge or experimentation, even if realised in the imagination. It is the ability to be responsive to what can be.

(King et al., 1989; May, 2007; Merleau-Ponty, 1999; von Eckartsberg, 1998)

But let's remember that there are also, as expected, the side effects that always come with new theories and which generally lead to immobility. And so it was no different with Husserl. The very existence of the subject, after Husserl, therefore, began to have the possibility of an intentional nature - because Husserlian consciousness is always consciousness of something, always relational (or even self-relational). So why, after all, couldn't existence itself be an intentional, self-relational result? And if so, how can we overcome the same problems as hallucinated significations?

Here, then, is self-consciousness as a possible factor in an existential condition, which is different, but which may not end up being too far removed from what Descartes tried to substantiate, but without much success. Husserl, instead of resorting to the Cartesian god, resolved (or tried to resolve) this issue with consciousness, which he assumed was also reflexive, in parallel with intentionality. Therefore, consciousness became intentional and also reflective, aware of its own actions, including its intentionality. Consciousness and intentionality therefore became institutions of subjective hierarchies: they became an ontology for subjectivity.

So, to solve the problems, and to go a little further, he presented two structures of consciential and self-conscious relations: pure self-consciousness, which phenomenologically reduces (epoché) beliefs about the outside world (but not as radically as Descartes) in order to perceive the relations of consciousness itself, in terms of relational structures, as if it were possible to reach a level of understanding of the "essential" about conscious experiences. And this essential is, since Aristotle, the philosophical attempt to reach the invariant, the static and unalterable eternal, or even universal, which would be to reach pure consciousness, totally uninfluenced by anything transcendent.

He believed that it was possible to consciously and temporally perceive the transcendental ego itself - which soon became always "there", "given", present and active, which then became Husserl's static and pure element (or reference), in addition to Aristotle's immobile engine, Descartes' god, Kant's universality, and many others who followed him. Husserl made a lot of progress, he gave new and valuable perspectives to the questions of relationships, so dear to our arguments to come, but he didn't solve

the existing challenges at all and, as traditionally, he derived new problems from them, just as his predecessors always did^{xi}

1.2.6. Heidegger and the emergence of dasein

It stayed that way until Martin Heidegger came along, not long afterwards, since he was Husserl's student himself. Then, criticising his teacher's excessive focus on transcendental consciousness and phenomenological reduction, he proposed that existence itself is a relation of a being-given-in-the-world, or being-there, or just dasein, or whatever, which is a conscious temporal resultant status and inserted in a given context or perspective, in which it is both cause and effect, or a mere fusion of all of this (Heidegger, 2012).

Heidegger (1999, p. 16), unlike Husserl, redirected ontology to a level close to that of epistemology, by basing the possibility of knowledge (which falls into the domain of epistemology) on a more original understanding of being (which falls into the domain of ontology). Thus, there was no room for a radical Cartesian scepticism in Heidegger, as there could have been in Husserl, due to the gaps left. After all, for Heidegger, the dasein dispenses with nothing or can disassociate itself from nothing, since it incorporates into itself everything that is inherent to its existence, including the very technology that comes to be perceived as its own existential extension^{xli}, in the ways in which it begins to interpret the world, to signify its own existence, or to enquire into its own being, in more favourable terms (Heidegger, 1999).

^{xi} The post-Husserl challenge would be overcome, therefore, when there was a philosopher capable of argumentatively unfolding the effects of this self-relational intentionality - of the Husserlian ego, which seemed to be an urgent task, at first glance, due to the radicalism that idealism could result in.

^{xli} Heidegger's idea that technology is an existential extension of Dasein is linked to his concept of being-in-the-world (*In-der-Welt-sein*). For him, human beings do not exist in isolation, but always in relation to the world and the tools they use. Technology, in this sense, is not something external to the being, but a way in which it interacts with and understands reality. In the contemporary context, we argue that this applies to AI and digital, which mould the way we perceive, think and exist in the world.

Dasein is part of a relationship with history, with social structures and with all the objective, or rather practical or functional, questions of life. Dasein cannot be pure consciousness, because it is essentially the operator of its own existence, realising itself as existing. Dasein is functional as a whole, or as a reflection of a whole, without ever ceasing to have this whole in itself, and without there being a circularity that can be defined^{xlii}. Objectivity, therefore, is now directed towards being itself, which transcends the body, and is directed towards contextualised and intentionalised existence itself, and all of this is completely allocated to a historical context that, as a whole, configures Heidegger's static element: the dasein itself, as an absolute referent within a given historicity, without being able to existentially transcend it and, therefore, incorporate or merge it into itself.

Practically everything in Heidegger is ultimately directed at the self, at the ontological status of dasein. Even corporeality. The dasein, then, due to its centrality and absolutism, is added to the whole of the Parmenideans^{xliii}, opening with Aristotle's immobile motor and up to the last seen, which was Husserl's transcendental ego.

^{xlii} The extension of this Heideggerian thought is undoubtedly *sui generis* and bold, since it radically opposes the traditional metaphysics that had been in force until then, which abstracted the functionality of the subject or object *p* from the very existence of *p*. Heidegger uses a hammer as an example in relation to objects, but we can update this example by replacing the hammer with an AI, and consider that there are two ways of signifying it, in a Heideggerian view:

(1) The first way is to perceive an AI in a conceptual, theoretical or even descriptive way. This would be a non-integrated vision of an AI in relation to our needs or intentions. When the use of this AI begins, the (2) second way of perceiving it, or rather, of no longer perceiving it, takes place. Because what the AI would be, as we can derive/suppose from Heidegger, is no longer relevant because it will be completely integrated into the user's intentionality, operating as an extension of their mind, or even their body, non-dualistically, and therefore becomes completely transparent to their perceptions. There is only a perception of the object, therefore, when it is unused or useless (Heidegger, 2015, p. 68), just as there is only a perception of the body itself when it is not in excellent health. In illness, in the organic flaws that the body has, its previously unnoticeable perception arises.

^{xliii} One could also consider inserting historicity into the Parmenideans, in Heidegger, because of its very explicit relationship with dasein, and vice versa. However, since each holder has only had a single seat in the Parmenideans, let's just consider dasein as elected to it.

1.2.7. Ricoeur and the emergence of narrative

After Heidegger, and even though influenced by him, Paul Ricoeur (1989, p. 12) prioritises narrative by proposing a capacity to distance oneself from one's own historicity so that there can be sufficiently effective hermeneutic conditions for understanding.

Ricoeur transcends philosophy and blends an interdisciplinary approach in search not of an analytical paralysis, but of an analytic in movement. What results from this is something extremely complex, as existence and its representation should be, or as it would be (and still is) supposed to be. After all, in reaching relationships without large, immovable central references, in prioritising the interpretative-epistemic narrative, which is fundamentally dynamic and transformational, his work was not about seeking references, but about interpreting complexity itself, as if the hermeneutic method itself could be recursive. It is the complexity of analytical interiority, of analysing movement from its own interiority^{xliv}, and without external and static references.

Ricoeur, in relation to artefacts (or technologies), takes a position without the radicality of the incorporating *dasein* and considers that knowledge can be mediated by them: they return to the tool level of a medium. Let's think about technologies: like Heidegger, there is some appropriation of technology, but not radically incorporated and totally directed towards the self^{xlv}. An AI would then not merge with the user, but would be the medium between the user and the reality that surrounds him. Ricoeur adopts a long

^{xliv} And, after all, the hermeneutic circle is just like that, which some say is a problem precisely because of the lack of a defined reference point, without the parts or the whole being elected as such. And this bothers the common judgement, especially when it departs from the historicity (through its linearity, causality and supposed rational coherence) presented so far.

^{xlv} For Ricoeur, artefacts and technologies don't merge completely with the self as in Heidegger, but function as mediators of knowledge and experience. In other words, technology is a medium, a means by which we interact with the world, without this mediation becoming inseparable from the identity of the subject. Unlike the *dasein*, which incorporates technology as part of its own existence, Ricoeur maintains a distinction between the subject and the artefact, emphasising that knowledge passes through an interpretative path, marked by narrative and historicity.

route to knowledge (Romele, 2019), which finally reaches the self: this long route is given by the narrative, which replaces the Heideggerian *dasein* (or historicity) and takes centre stage in his theory. We could therefore consider it as a reference and even fit to integrate the whole of the Parmenideans, if the narrative were not eminently dynamic, since it always occurs as a movement, a context or perspective. Parmenideans don't allow for movement, and so Ricoeur's narrative will be left out of the mix.

Meaning, therefore, is an important part of the process outlined by Ricoeur, which is always based on a narrative (Romele, 2019). We'll come back to Ricoeur a few more times, without dwelling on him for the moment. What will interest us will be what he defined as appropriation (in short, a process of transcending the cognising self in order to achieve knowledge through the interpretation of the narrative: which is pure movement), in addition to his hermeneutic proposal, which is perfectly applicable to the digital and extremely current, given its dynamism.

1.2.8. A synthesis with the emergence of the "hyphen" as relational

So far, we can summarise and schematise (in figure 1) our postulates in a review of the best moments in this very short^{xlvi} onto-epistemological history of philosophy:

- Firstly, including Aristotle, we had the focus on the object (the). The conceptual resultant, theoretical guarantor of his argument and static referent (opening up the whole of the Parmenideans) was the immobile motor;
- Secondly, to include Descartes, who consolidated the focus on the subject(s), even independently of any object(s). The resulting conceptual guarantor and static referent was God;
- Thirdly, since Kant, we have had a deeper focus on the rational subject (s), and on subject-subject (s-s) relations. The resulting conceptual guarantor and static referent was the induction to universality;

^{xlvi} It's also incomplete and unfair, we repeat, because we don't mention countless other philosophers who were relevant.

- And fourthly, with Husserl, the focus shifted to the hyphen (-) that integrates or connects both the subject-object relationship and even the subject-subject relationship. The resulting conceptual guarantor and static referent was the transcendental ego;
- And fifthly, with Heidegger, we had the contextualisation of the enquiry about the hyphen, because the enquirer is the very ternary subject/hyphen/object integrated as such, consolidating itself as a unicity capable of incorporating all the technology that orbits around it for its own existential justification^{xlvii}. The resulting conceptual guarantor and static reference was the dasein itself, within its respective historicity, in the impossibility of "there" being something "outside" it;
- And finally, Ricoeur attributes to the hyphen (made multifaceted) the possibility of itself being a technology that realises meaning, of intentionality as a mediator between the self and the reality that "surrounds" it. However, this mediation is also teleologically appropriated by the subject, just like Heidegger's voracious dasein, although not as directly or with more restraint. The resulting conceptual guarantor and referential (not static) was the narrative, which, unlike the previous ones, is essentially dynamic and therefore unfit for Parmenideans;

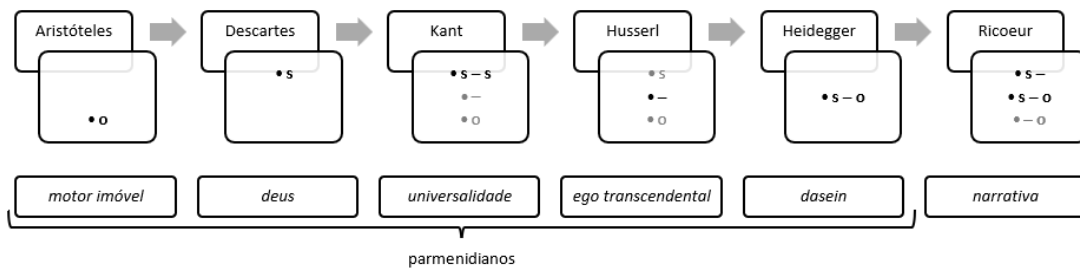


Figure 1- "Evolution" of the macro-ontological ternary of epistemology: subject, hyphen and object and their respective "side effects", but presented as "causal".

^{xlvii} The dasein has become a super-hyphen, embedded in history, powerfully connected to everything in search of its own existential justification, or like a black hole that sucks in everything around it, densely constituting itself (existing) as a given in the world.

The "hyphen", therefore, as a relational metaphor, has not only ceased to be invisible, but has also gained relevance to the point where it now plays a leading role, since it is the legacy of the problem that has hitherto been in question.

There is something intriguing about these philosophical proposals. When we realise that the philosophical concepts proposed at the time, and in relation to their predecessors, end up reinforcing, annulling or antagonising them. But in relation to all the resulting conceptual guarantors and static references (the Parmenideans), they end up forming a perfectly homogenous and solidary group. They are related and even connected^{xlviii} : any similarities with the supposed problems of AI may not be mere coincidence, because they behave just like the Parmenideans.

1.2.9. The emergence of AI as relationships (or correlations)

By arriving at Ricoeur's narratives, we can consider that we are closer to what an AI actually is or can be and, therefore, on the right path, through the narrative, to realise the existential fragility of its problems, just as they are Parmenidean. After all, the

^{xlviii} The Aristotelian immovable motor can quietly "evolve" into the Cartesian concept of god (which remains in full act, even when there is supposedly nothing else), which would itself be the greatest expression of Kantian universality (substantially, in itself) which, by the way, can be easily (and functionally) realised by the pure ego (transcendentally) in Husserl, as well as being intimately incorporated by the Heideggerian dasein (incorporating/absorbing transcendence), which inquires into itself, into its being.

However, the previous immobility could be perceived through a narrative that adds Nagel's "all that" and transmutes the paralysis into movement of the "that". It is this narrative that we are hermeneutically making in these lines, by reproducing the movement of history itself and appropriating the meanings left behind as traces, from a perspective that required (1) distancing ourselves from such processes in order to (2) hermeneutically realise what Ricoeur proposed, by appropriating each concept, not in past times, but in current times, in our times, when a hammer gives way to AI, for example, (3) integrating them. Thus, through the narrative, we were able to perceive the Parmenideans as totally static, but also deeply informational and harmonious with each other.

As you can see, they all coexist in the same habitat without any major obstacles. What is being realised is that the "problems" generated are, in themselves, more stable relationships than the proposals that generated them. There is a kind of philosophical inversion in relation to the effects, which become more consistent than their causes. Speculatively, it may be these effects that, through their supposed stability and groupings, weaken the causes and show that it is, in fact, the relationships between the causes that should be prioritised, without the effects being necessary. We can consider causes as positive or needs. And the effects as negative or contingencies. If you can't eliminate them, you can at least not create them or find a way of not needing them any more. Relationships are therefore only relevant, but necessary and sufficient when they are perceived in this way.

current problems of AIs are not, in fact, only about their users (those who are in direct or indirect contact with an AI, suffering the impacts of a supposed algorithmic agency), in isolation, nor only about themselves, as technological objectifications, but rather in the relationships that exist between users and technologies, and which lack a coherent narrative. And relationships "are what they are"^{xlix} while given as AI, of course. Because "what technologies say, I argue, cannot be analysed under existing theoretical terms and assumptions"^l, according to Freiman (2023).

The hyphen, as dynamism, signification and orientation, is always a relational form that incorporates within itself transformations caused by different relationships. And this is what needs to be realised in this reconstructive effort we've made so far. Because the very act of seeking knowledge or interpretation is, first and foremost, a movement and an attempt to relate to something that is still supposed to be unknown, albeit present. This is the same thing that Ricoeur (2016, p. 145) proposed through a displacement in order to achieve understanding, through appropriation: "The dialectic of distancing and appropriation is the final figure that the dialectic of explanation and understanding must assume. It concerns the way in which the text is addressed to someone."

So what would come of all this philosophical shift towards a focus on relationships, or movement? We would arrive at a model very close to the one proposed by Ricoeur, in

^{xlix} Provocatively, the "is what it is", quoted by Jean Paul Sartre, when he argues that being-in-itself is what it is. Speculatively, if we consider the relationship itself as "being what it is", extrapolating Sartre, it would mean that it doesn't refer to itself in the way that relational consciousness (or intentionality) does. The relational or intersubjective reflexivity that constitutes the self merges into an identity in act. Sartre uses the analogy of a cloud, saying that it is not "rain in potential", but rather a certain amount of water vapour which, under certain conditions of temperature and pressure, "is strictly what it is". Therefore, the relationship has no potentiality, but exists in act, in the absolute fullness of its intentionality - and therefore its intrinsic dynamism. It is opaque to itself because it is "full of itself", not having an "inside" that opposes an "outside". It has no secret, it is massive, and therefore does not need immobility (Sartre, 2011, 2015).

^l By replicating the philosophical-analytical models that have been applied in AI analysis, with the ever-present Parmenideans underpinning them, there will be serious distortions of the real origins of the problems, due to the efforts to preserve the integrity of the fragile Parmenideans at all costs, by submitting AI as an immobile and stable vision, while everything in it and around it is in full movement and transformation.

which the dasein (in the sense of technological incorporator) would not, however, be discarded. For, even in speculations about the human fused with technology, as posthumanism supposes, or even the robotic men (cyborgs) idealised in science fiction (even if they are less fictional and more probable every day), technology is incorporated with a utilitarian function that is thus at the service of other interests, transcendently to the sets formed from the ternary s - o, which unfolds so that this set is referential to another: (s - o) - (...), considering the ellipsis to be new relational extrapolations, and so on.^{li}

It wouldn't be rash, however, to realise that the shift in focus between subject/phenomenon/object has so far been not only epistemological, but also ontological. With each theoretically developed complexity, the ontology considered in the relational forms of the concepts has also changed.^{liii}

And this is an attitude that is beginning to take on academic consistency, although still without equally consistent means of how it can be done, due to some limitations that we will try to address in due course, as proposed by Harding (2011; Portmess & Tower, 2015, p.2): "To account for the ambiguities in the notion of data mining and analysis, a

^{li} This poses a problem for those looking for the origin of this relational "web", like a neural network that is formed, based on the synapses emulated by the hyphens, as relational means of signification. And the still obsessive search for the origin as a guiding reference is what justifies, for example, the same reduction that Aristotle sought when he chose, millennia ago, the concept of the immobile motor as the first cause, as well as all the rest of the Parmenideans that followed it. And this is what we still try to do, which is why there is the source of the frustration in perceiving AI through anachronistic analogue thinking.

^{liii} After all, studying this ontological mutation from the digital point of view, as we have superficially done from the analogue dimension, as well as having been (and still will be) here, will lead us to discuss possible ways of showing how AI can re-ontologise the world (Mittelstadt et al., 2016), which is one of the most relevant parts of what will be presented from a mapping of ethical problems, in the next section. Because, as we will see, it is not a question of whether or not AI will re-ontologise the world, but rather how this will happen (Freiman, 2023). This is why we must first prepare the ground so that a standard (or framework) can be created to objectively perceive this supposed ontological change in a temporal, evident and measurable way.

richer epistemology is needed from *big data* that reveals, rather than conceals, its social relations and allocations."^{liii}

There is such an evident recursiveness in technological dependence today that, in order to realise even such effects of technological dependence, one will have to turn to technology itself to do so. "Some technologies, such as the Internet or smartphones, have not only mundanely added more opportunities for us to choose - but have constantly moulded the way we live" (Freiman, 2023; Waelbers & Briggel, 2010). And this is also evidence of a symbiotic relationship already established between humans and machines, as argued by Brey and Søraker (2009), although the authors did not generalise in their concluding comments:

Computer systems are extremely versatile and powerful cognitive artefacts that can support almost any cognitive task. They are capable of engaging in a unique symbiotic relationship with humans to create hybrid cognitive systems in which a human and an artificial processor jointly process information (Brey & Søraker, 2009).

And finally, the search for understanding through an ontological method is exactly what we have tried to do so far in this first epistemological approach throughout philosophical history, based on the three aforementioned instances (s - o). Through them, and mediated by them, like technological artefacts, we arrive at some knowledge about knowledge itself, from such and such perspectives, or narratives. So, if all goes well, we'll be looking at having the conceptual foundations for a technology that could nevertheless tell us our potential symbiotic level with AI, if so perceived.

We have therefore believed that we have argumentatively constructed the need to decline the use of rigid elements, such as Parmenidean ones, to the detriment of perceiving AI as narrative and deeply relational. The concern, at the time, was not about

^{liii} To meet this demand and overcome the limiting factors necessarily requires the prior structuring of a meta-ontology (that which is supposed to be about relational, recursive, cyclical movement) so that, given the human impossibility of directly observing something so complex, subjective or immaterial, a technology can be created to mediate the interpretation of the technology itself in its effects, until knowledge of how the supposed transformations are taking place is achieved.

AI (because "it is what it is", in other words, relationships), but about our own prejudiced and inertial capacity for analogue thinking that ends up frustrating us when faced with the digital, as we assume that the latter is still in an ontologically inferior or secondary position to what is actual, to what exists, to what we are. This is no longer the case. So we can continue.

1.3. Ontological emergence from a technological perspective

In this section, we will be adopting a less theoretical and more technological perspective, albeit under a philosophical hermeneutic. The aim will be to develop the foundations for a conceptual tool capable of representing the relationships that exist in an AI, to avoid falling into reference or dependence on something similar to Parmenideans.

We will also discuss the technological developments of AI algorithms, to understand how they have been (and still are) representing relationships even without having a conceptual model based on them, since even in philosophy relationships have generally been in the contingency field, as we saw in the previous section, without being given the prominence they deserve.

But any AI constructs are essentially relational, or even more complex, because they are correlational, and so we need to realise the size of this *gap* and find ways to fill it. This is what we will try to do in this section.

And so, in order to discuss how we will relationally approach the most relevant ethical problems in Mittelstadt et al. (2016), we need to understand the current status behind the scenes of AI, specifically how data is signified in its bases until it generates said problems. We will set out to represent the ethical problems in an algorithmically and ontologically connected graphical scheme: the aim is to understand the state of the art of constructing an epistemic-ontological map of current AI, or simply a representational map.^{liv}

^{liv} Representational maps, considered here, will be conceptually presented and developed on the basis of the theories and practices already used in knowledge representations (Sowa, 2000), but not limited by them, as there have been many advances since then. Let's first take a conceptual look at ontologies, algorithms and some of the knowledge representation models. For this reason, it will be necessary to

1.3.1. About Ontology and ontologies

So far, we have used the term "ontology" in an unconventional way, neither for philosophers nor for information scientists. The term "ontology" has been used here as a semantic hybrid of these two not always correlated fields, at the risk of not pleasing either side, but with the obligation to do so, since it is necessary to unite them in semantically shared knowledge - at least for the issues addressed here, and in the direction of an operative interdisciplinarity. The dimension we are seeking to represent with ontological practice is that of significations. And this will involve the subjective, intersubjective and objective dimensions, which we will gradually build into the concepts needed for this sharing.

It's worth pointing out that the problems surrounding meanings are not only inherent to humans, but also to AI algorithms, which operate on the basis of not always so user-friendly databases, as Berners-Lee et al. (2001,)point out : "Ideally, the algorithm should have a way of identifying common meanings for any databases it encounters." And so it concludes:

A solution to this problem is provided by the third basic component of the Semantic Web^{lv}, which are collections of information called ontologies. In philosophy, an ontology is a theory about the nature of existence, in defining how certain kinds of things exist; ontology as a discipline studies these theories. Artificial intelligence and web researchers have co-opted the term into their own jargon and, for them, an ontology is a document or file that formally defines the relationships

discuss relationships and how we can perceive them philosophically. What we are looking for, therefore, is to understand the relationships between territories and maps, in other words, between actuality and its digital or functional representation.

^{lv} "In addition to the classic "document Web", the W3C (World Wide Web Consortium) is helping to build a stack of technologies to support a "data Web", which would be about the kind of data you usually find in databases. The ultimate goal of the Web of Data is to enable computers to do a more efficient job and to develop systems that can support reliable interactions over the net. The term "Semantic Web" refers to the W3C's vision for the Web of ontologically linked data. Semantic Web technologies allow people to create data stores on the Web, create vocabularies and write rules for dealing with data." (accessed on 24/08/2024 via the link https://www.w3.org/2001/sw/wiki/Main_Page, our translation)

between terms. The most typical type of ontology for the Web has a taxonomy and a set of inference rules (Berners-Lee et al., 2001).

We will therefore assume that the use of the term ontology, which we intend to adopt here, is not only applicable to interoperable / intersubjective and objective digital data (AI) or analogue events (users), but also to subjective ones, in an intrinsic relationship with users. That's why ontology, in our development, will really be a hybrid, since it will deal not only with conceptualisation, but mainly with the relationships of both the form (or Aristotelian essentiality) and the characteristics (or Aristotelian accidentality) of how existence takes place, both in ontic terms (material, dense, temporal, etc.) and ontological terms (transcendent, subtle, extemporaneous, etc.), for the actual and the digital.

Furthermore, the ontology we will develop will also deal with how such instances relate to each other in the same existent, or between sets of existents that are being considered as relational. It is, therefore, a pretence of an ontology that is sufficiently capable of dealing with both an autonomous (particular) unit and the collectivity, because it is supposed to represent, in terms of attributes, how the relations of these take place not only in the forms s - o, but also s - s, through intersubjectivity. ^{lvi}

The maximum pretension for such a hybrid ontology would be, based on attributes and entities, to achieve a configuration that both explains and defines entities, from the ontic to the ontological dimension, and vice versa, by representing existence within certain functional perspectives. And this presupposes not only perceiving the objectified allocation of an entity within a relational existential structure, such as the social one, for example, but also perceiving this entity's own subjectivity, by apprehending or emulating its own perspective in relation to this structure. ^{lvii}

^{lvi} So "our" ontology won't necessarily define what it is representing, but it will explain what happens, at least in relational terms. And if, on the basis of these explicit relations, there are resources for arriving at certain causal relations, then we can say that we have found a perfect relational representation and, therefore, a perfect ontology.

^{lvii} This ontology not only seeks to know what you are, but also what you could be, as it is relational and can even be predictive, which could bring some advantages to machine learning processes, as we will

There is therefore a strong argument in favour of considering that the same ontological standardisation can (and should) be applied to the semantics of the two domains: the current and the digital, with the latter being instantiated as a representation of what exists (of what is current), even if not everything that "really exists" (in actuality) is represented as such. But by thinking like this, some unwary people may think that the digital will always be a subset of the current, and this is a mistake, both because of the digital gigantism in terms of data, and because of the speed that the digital brings - because of its influence on current life, which was previously purely "analogue". From the digital, entities or forms of intervention and impact emerge (and therefore exist) that can have a "life of their own" without being representations. "A standard view is that technological artefacts are not neutral intermediaries, but actively determine how we construct knowledge" (de Boer et al., 2018 apud Freiman, 2023, p. 4).

Conversational AI itself is an example, as some of them^{lviii} are even designed and programmed to be (intentionally emulated) humanised, with names, psychological profiles, certain behavioural reproductions simulating empathy, and yet without parallels with existing ones: without representing any actual individuality, obviously, but a mere generality adjusted according to the user's profile. Without containing an

soon see. If there is a problem about the re-ontologisation of the world due to the massive action of AI (Mittelstadt et al., 2016), we need to understand the variance between current and future ontologisation, comparatively, within certain standards to measure any changes. Any solution to the problem involves this comparative need. And any comparative need requires clearly defined and measurable parameters, which in our case we argue will be based on a given ontology, or metaontology.

^{lviii} There are conversational AIs that emulate, for example, humans in which affective relationships can be formed, as in the case of the Replika application, but not only. Thus, such entities are not necessarily representations of a specific human, and so they can "exist" to represent not entities, but entities, or rather their attributes. Mathematically, with the immense power of computing, it is possible to create more individual (digital) personalities than there are existing humans and, therefore, it is not correct to assume that digital representation is limited to being a mere subset of the "real world", since there is nothing to stop alien profiles being created, for example, or those of monsters or superheroes. Replika, as an enterprise, originated in the attempt by its founder, Eugenia Kuyda, to digitally reproduce the consciousness or individuality of a deceased friend who was very dear to her and who had left a huge database of messages while they were in a relationship. The data thus became digital traces of this friend and this relationship, and served to give new life to what was no longer current. (Accessed on 16/08/2024, at <https://replika.com/about/story>)

Aristotelian essence, and composed almost entirely of accidents, attributes or contingencies, they operate and consolidate something that is "non-existent" as if it were existing: they give life by emulating a digital presence that, in fact, is not a presence, but appears to be. Like illusions meticulously crafted to appear credible. Such AIs correlate only with raw attributes that end up expanding through the correlations derived from them and, because of this, from certain algorithmic operations they can recursively derive other mixed, hybrid attributes and modify the very ontology initially considered: they can impact and modify the user's actuality, therefore, and even initiate other relational changes .^{lix}

Digitalisation, *lato sensu*, has even changed the forms of human relations, most of which have now been transferred to the digital dimension, through social networks and mediations of communications via applications that show certain attributes: it determines or makes official whether users are *online* (connected), *offline* (disconnected) or *busy*, for example, even if it refers to a fake or corporate profile, which doesn't correspond to an existing one, as such, but still has attributes^{lx} that explain it

^{lix} As an example, we can cite a critical reading of the article Generative Agent Simulations of 1,000 People (Park et al., 2024), carried out at the 7th Research Seminar of the Signo Project, in the session held at the Philosophy Institute of FLUP, on 09/Dec/2024, with the title *AI Agents*, in which a simplified algorithmic simulation was carried out based on ontology, in which each value would correlate with five others, in order of relevance, according to data taken from ChatGPT, from OpenAi - in the same source and version that the article used.

Starting with a randomly chosen value (which was clarity) and its derivatives, which were correlated one by one with five other distinct values, the values were multiplied up to a limit of eighty-eight evoked values. From then on, any of the evoked values would evoke another five, but always among all the existing ones, with no new ones being added. The only thing that changed was the weight of each of the evoked values.

Thus, it could be considered that the "universe" of values (according to this simulation) - the domain of values - of an IAC would be restricted to eighty-eight values, from a specific trigger of "clarity", which proves the IAC as a partiality of the existing, and with the expected inaccuracies, without the subtleties of the different perspectives that exist.

More information on the website <https://ifilosofia.up.pt/activities/signo-seminario-2024>, accessed on 19/12/2024.

^{lx} An attribute is not static, but dynamic. Because an attribute is always relational and takes place in perspective. Between human and actual right and wrong, or between good and evil, or war and peace, for example, there may be too many degrees for a digital ontology, and perhaps some that we don't even know about. Hypothetically, from such unprecedented levels of attributes, possibly identified, one of

and place it in a certain structural position^{lxi}. There is, therefore, the creation of digital existents without current counterparts and this is an embodiment, based on models made for the opposite purpose, to represent what exists today.

1.3.2. Knowledge representation models

these levels could emerge and take centre stage, the greatest relevance. And this is a change from the current ontology: reontologisation is an effect of the movement.

So there can be no universality or conceptual immobility to anything, at least in terms of digital or artistic representations. Even fictional conditions that are considered almost consensual, such as dystopias (always described as negative, or something bad, undesirable), cannot, in the light of reason, be considered consensual. Dystopias for some can simultaneously be utopias for others.

"Dystopias seem to be quite different a priori, but a posteriori they are not very different from utopias in their purposes, which refer to the attempt to find an allegory to lead everyone into a reaction to non-compliance or unrestricted compliance with existing rules - all that is intended with these utopian or dystopian creations, whether predictive or artistic, is to question the established order through its own current rules - it is a call to some counter-ideological reason, but still ideological, like everything else. Utopias and dystopias are projected for reasons that are always ideological, because they are expressions of possibilities or impossibilities." Ortolan (2022, p. 259)

^{lxi} Nor is it a defence of complete relativism, but rather the presumption of the need to adopt a configuration that takes into account the individual perspective and that, when these perspectives are considered, they are relational, and that we can identify, for example, ethical resources for both utopian and dystopian orientations. What we are looking for is not the defence of any philosophical, economic, anthropological, political or even sociological theory or school, but facts and considerations that will lead us to knowledge through the most reliable representations possible, via means that can justify certain causalities.

After all, since before Aristotle, Philosophy has been concerned with investigating how existence occurs, or how we can know what exists, as seen in the previous section, and so what is being considered here as a fusion of ontological knowledge is really about the attributes of being (for Philosophy) or the attributes of data (for Information Sciences), and all within certain conceptual constructs that can be analysed in perspectives, and operated in structured or stacked ways, to gain consistency, robustness and allow for a less naive interpretative narrative.

As an objection to such a possibility, we must consider from the outset that if a human attribute is given the same semantics for itself and for its digital representation, in order to be coherent, this digital semantics could undergo changes and no longer correspond exactly to the same initial representation. As a result of the semantic detachment between the actual and the digital, there is a serious risk that the exponential digital results produced by algorithms will be considered hallucinatory, due to the lack of correspondence between what they obtained as *output* and what is actually there, as suggested by Hicks et al. (2024, p. 37, our translation) by emphasising that the priority of AI models is to present linguistic coherence, based on probabilistic operations: "these models were not designed to transmit information, so we shouldn't be too surprised when their statements turn out to be false", as this is one of the most relevant existing problems (and very likely in conversational AI, almost commonplace) that we will soon address.

Ontologies, as we can see, are used by technology *designers* in the *strict sense* for purposes such as the architecture and management of relational databases or the interpretation of information to be used by algorithms. These ontologies serve as conceptual bases for endowing data with certain attributes so that they can then be correlated in search of the meanings that result in the objectives set for the algorithm. Relationships, unlike what we have seen in philosophical developments, are already at the centre of technological-constructive developments in AI, through relational databases, but not only.

Most studies in the literature focus on relational databases as a source of information for several reasons: Firstly, around 70% of the data on the web is stored in relational databases (He et al., 2007); Secondly, relational databases present complete conceptual models; Thirdly, they provide a complete information resource (Martinez-Cruz et al., 2012); and finally, they offer one of the best techniques for storing and manipulating data. *However, relational databases suffer from a lack of semantic meaning, which hinders the ability to achieve interoperability between information systems* (Meersman et al., 2002 apud Ben Mahria et al., 2021, p. 2).

The usability of ontologies has been almost exclusively directed from the digital to the digital, as a quest to give meaning to data so that it can be operated on in search of information. After all, data is considered to be a digital trace of a representation of factualised reality (Romele, 2019). Thus, an ontology is used as an optimised way of navigating these trails based on certain premises. For this reason, but not only, there is the current impossibility of a semantic implementation for the data operated, due to the lack of ontological correlation between the digital and the actual, in semantic terms - there is a *gap* between these dimensions, and it is precisely the semantic *gap*.

It classifies, but without signifying. You establish what's in the world (the actual world) in terms of attributes, structure it ontologically as such (represent it) and apply it to the digital world by going through the stored data to fit it necessarily into these structured classifications. And the work is considered done, because it is impossible to achieve (or structure) the necessary meaning of the ontologised data: the impossibility of achieving semantic modelling is established.

And this is almost always done mechanically, mostly by algorithms too, as it would be too costly and time-consuming for humans (Ben Mahria et al., 2021). Data management systems "generate" their own ontologies, based on relationships that become related to other relationships, creating correlational, multidimensional webs, and from this we realise the hallucinatory risk and systemic inefficiency of causal traceability. There is, therefore, no mediation possible in its maximum transitive instance, since the result is the very absence of semantics, since there is only a compulsory adaptation to a reference that is not always free of prejudice or usability. What happens in this case, and curiously, is the opposite of what was criticised in the previous section, when there were immobile references (the parmenidians) to give meaning to the relations. Here, we see just the opposite, that they are just relations without any semantic referencing between them, neither internal nor external, which becomes almost like a situation of complete relativism: and this is even worse, because it makes it impossible to tell the story, or to understand the relations as they are.^{lxii}

A new ontological proposal would therefore need to be implemented, and for the sake of the desired versatility, not just a full mediator between the current and the digital, but also a facilitator, without making either of them an immovable reference. Immobility would, at most, be for certain characteristics of the ontological structure itself, in other words, it would be a metaontological structure - dynamic, flexible and adaptive - that could contain fixed elements, but still be open to recursion, or certain complex

^{lxii} A middle ground must therefore be sought: ontologies that are based on relationships (as supposed) but that also allow them to be correlated to current events, which requires them to be completely dynamic, but that are able to support, from their own constituents, an internal narrative that leads to the meaning of their contents: that allow a connected narrative from their metadata that can both explain and define their data, ideally. In other words, in the impossibility of an immovable external referencing, semantics must be extracted from its own interiority in constant movement, constantly updating itself in parallel with actuality. Due to the predominance of internal correlations, conversational AI, after all, for this reason does not operate directly through ontologies, but exclusively through correlations made from learning data bases, and not metadata, as we have seen. This is why it is said, as already mentioned, that it is not designed to generate knowledge, but rather to generate linguistic coherence, merely by replicating what is there with the utmost fidelity.

operations that could improve or update it. You could say that data is to metadata what ontologies are to this metaontology.

The Technological Sciences, especially Data Science in the broad sense, have already realised these demands (and also the theoretical limitations) and are developing new forms of representation, modelling and languages with impressive speed, such as 3D ontological models (DOLCE^{lxiii} , UFO^{lxiv} and one of its most promising specialisations,

^{lxiii} DOLCE is a fundamental ontology that describes the distinction between the colour of an apple (its quality) and its 'value' (for example, a particular shade of red). The latter is called a qualia and describes the position of an individual quality within a certain quality dimension (such as nuance). An example of this would be that John's age is a moment of an individual objectifying his age as an abstract property. The intensity is a moment of John's headache, which is a moment of John; the intensity is inherent to the headache, which is inherent to John. And so a relational meaning is built into all the attributes (Guizzardi, 2005; Zamborlini, 2011).

^{lxiv} UFO (Unified Foundational Ontology) is a fundamental ontology that provides a basis for ontology-driven conceptual modelling. UFO is organised into three layers: UFO-A (objects), UFO-B (events) and UFO-C (social). UFO has been used as the basis for domain ontologies, including the ODE software ontology. It has also been used to analyse concepts related to agents and business modelling languages. UFO distinguishes between: universals, which are patterns of characteristics that can be realised in several different individuals, and individuals, which can be concrete (e.g. a particular person, an explosion) or abstract (e.g. sets, numbers and propositions). Concrete individuals are divided into: endurants, which are individuals that are fully present whenever they are present (e.g. a house, a person, a quantity of sand, etc.), and events, which are individuals that can have temporal parts. They happen in time in the sense that they extend in time and accumulate temporal parts (for example, a football match). UFO-C, the social layer of UFO, distinguishes between agentive and non-agentive individuals. Agentive individuals (or agents) are capable of sustaining a special type of moment called intentional moments. Intentional moments can be further specialised into: mental moments (including beliefs, desires and intentions) and social moments. In addition, each type of intentional moment necessarily has a propositional content, which can be matched by certain situations in reality. Thus, the intentionality of agents should be understood as the capacity of their properties to refer to possible situations in reality. OntoUML is the ontologically well-founded version of UML 2.0, proposed by Giancarlo Guizzardi. It extends the UML metamodel to be isomorphic to UFO-A, making it ontologically consistent. OntoUML's stereotypes reflect UFO's ontological distinctions (Guizzardi, 2005).

OntoUML^{lxv}) and even 4D (BORO^{lxvi}). These developments, in their constant updates and new forms of application, already allow all kinds of philosophically conceptualised dynamism to be applied in algorithms.

So there are no technical limitations in representational terms that can no longer be overcome by the resources that already exist for metaontological embodiment or structuring. What is still needed, however, is a breakthrough in terms of

^{lxv} OntoUML is a conceptual modelling language that extends UML 2.0 and is based on the UFO (Unified Foundational Ontology) ontology. It was proposed by Guizzardi (2005) and aims to make UML 2.0 ontologically consistent, i.e. that the language is able to represent the world more accurately and completely, according to ontological principles.

OntoUML maps its elements to UFO-A, the object layer of UFO, and uses stereotypes to represent the ontological distinctions of UFO, such as types, moments, qualities, roles and so on. This grounding in UFO-A allows OntoUML to represent not only the structure of the data, but also the ontological nature of the domain's elements, such as the difference between rigid and flexible types, entities and events.

OntoUML has been used in various research projects, including: the evaluation of UML 2.0 as a conceptual modelling language, the development of domain ontologies such as the ODE (Ontology Definition Environment) for software, the analysis of concepts related to agents and the modelling of business modelling languages.

Some examples of how OntoUML can be used to represent different modelling scenarios include: simple static scenario: direct representation of domain entities, without considering temporality, using OntoUML stereotypes; reified static scenario: reified representation of the domain entities, without considering temporality, using OntoUML stereotypes; Simple dynamic scenario: extension of the simple static representation to include temporality, using the 4D (perdurantist) approach; Reified dynamic scenario: extension of the reified static representation to include temporality, using the temporal reification approach.

Despite its advantages, OntoUML is still not as popular as other modelling languages, such as UML itself. However, its solid basis in ontological principles and its expressiveness make it a promising option for ontology-oriented conceptual modelling (Guizzardi, 2005).

^{lxvi} "BORO" is an acronym for Basic Reference Ontology. It is a fundamental materialist ontology that has been developed for use in knowledge engineering applications. BORO is characterised by its metaontological choices, including: spatiotemporal perspectives, individuality in modal terms, materiality and universality. BORO differs from other fundamental ontologies such as DOLCE and UFO by emphasising materiality and spatiotemporal perspective. It was designed to provide a basis for practical applications of knowledge engineering, rather than abstract philosophical investigations. It is a well-documented and mature fundamental ontology that can provide a solid basis for knowledge engineering projects. It is important to note that the choice of foundational ontology depends on the specific requirements of the modelling project. BORO can be a good choice for projects that require a materialistic fundamental ontology with an emphasis on practical applications (Partridge, 2015).

conceptualisations from a dynamic perspective, without more parmenideans, so that these new theoretical developments can then be properly employed.

1.3.3. What we teach algorithms

But what are the algorithms referred to here anyway? In order to maintain consistency between the ethical problems mapped and their causal origins, we will use exactly the same consideration as (Mittelstadt et al., 2016, p. 3):

(...) the algorithms we are interested in here are those that make generally reliable (but subjective and not necessarily correct) decisions based on complex rules that defy or confound human capacities for action and understanding. In other words, we are interested in algorithms whose actions are difficult for humans to predict or whose decision-making logic is difficult to explain after the fact.

The point about the capacity for "creation" (due to the excessive correlations that can be made) that an algorithm can realise and which, as a result, can even produce hallucinations (when the results have no correlation with current events, meanings, facts, natural laws or related concepts, but not only) is that they go through the machine learning process^{lxvii}, which basically consists of providing the algorithm with certain basic instructions so that it can create excessive correlations from a gigantic database and, with this, identify and extract existing correlations to generate decisions (*outputs*) that are (hopefully) true or, at least, ethically neutral.

^{lxviii} machine learning is definitely the most critical (and expensive) part of developing the most advanced AI models, as it requires huge investments in processors, electricity and databases, as well as a possible structure of human resources who will evaluate and

^{lxvii} "Machine learning is "any methodology and set of techniques that can employ data to create patterns and knowledge and generate models that can be used for effective predictions about the data". Machine learning is defined by the ability to define or modify decision-making rules autonomously." (Van Otterlo, 2013 apud Mittelstadt et al., 2016

^{lxviii} For Hicks et al. (2024), the term hallucination would not be the most appropriate, but rather nonsense or bullshit (as less scatological translations for the English word bullshit), based on Frankfurt (2005). Whatever their nomenclature, what we do know is that these *outputs* are not always at the desired levels, as we'll see in the next section, and for various reasons.

correct the *outputs*, in successive and intensive supervisions until they are satisfactory.

lxix

But what is at stake, specifically for conversational AIs, is that the basis for learning - which ends up being their content, or the repertoire they possess - also occurs from digital reproductions of human conversations, many of them taken from social networks, for example. Thus, it can easily be inferred that it is the public (human) discourse itself that corroborates the *outputs of* conversational AI, with all the values we have in it, obviously, both ethically neutral and positive, as well as negative. It wouldn't be difficult to realise that conversational AIs are, at least from the outset, our own shadows, or even our own communicative pretensions, with all the values we project, strategically or otherwise. ^{lxx}

There is in AI, one way or another, the reproduction of the trace we leave in our discourse, as expressive humans. And in the most likely case, since it is a public human discourse, open to general access and moral scrutiny, it should be coherent to assume that it is endowed with good, socially accepted values, such as Aristotelian ethical virtues or Kantian duty based on universality.

The unfortunate fact is that this is not how it happens in "our" human discourse, as such, and you only have to access social networks to very quickly realise obvious prejudices, expressed publicly, intentionally or not, delicately or crudely, subtly or not. In other discursive instances, such as in literature or even in film scripts, which are also used for

^{lxix} "Common training methods include supervised learning, unsupervised learning and reinforcement learning. Supervised learning involves providing the model with marked data for training and teaching it to recognise patterns and make predictions; unsupervised learning involves providing unmarked data and allowing the model to establish patterns of correlations on its own; and reinforcement learning involves providing feedback to the model on its performance until correct predictions are achieved." (Russell & Norvig, 2021 apud Freiman, 2023, p. 2)

^{lxx} "It is possible that the results of conversational AI can be analysed as a form of group testimony. There are at least two different arguments to be made: the first considers the results of conversational AI as the testimony of the humans whose expressions were used in the training datasets - as a collective; and the second considers the results of conversational AI as the testimony of a community of experts." (Freiman, 2023)

machine learning bases, there is the considered perspective of different fictional characters, perspectives and many of these are equally inappropriate in terms of ethical neutrality, because such characters mirror humanity itself, with all its so-called peculiarities. After all, AI is what it is, or rather, what we all are.

We will see that there is a double problem of (in)coherence: the aim is to represent the actual in the digital (Freiman, 2023; Hicks et al., 2024; Romele, 2019; Russell & Norvig, 2021), in relation to semantic capacity, for example, and all the constituents with their meanings, in order to achieve epistemic capacities, among others. If this doesn't happen well, the AI is said to be hallucinating (or messing up). On the other hand, an extremely accurate representation, to the point of reproducing even the undesirable (and current) prejudice in force (and the whole range of ethical problems) is seen as problematic. And the solution is to use precisely what was once poison: hallucination.

In order to correct what they don't want as an *output of* an AI, technicians are looking for ways to make the AI acquire an eminently human capacity: lying, or rather dissimulation - a subtle way of "hiding" certain inherent aspects of oneself, in favour of something transcendent: the ability to move through space and time. And, according to Ortolan (2022), this is an ethical practice, adapting through a "protocol" that is operated ideologically:

This is why a subjective protocol is needed, which is simple and easily instantiated as dispositional in moralised and ideologised individuality, to also serve as a filter that both receives and allows coexistence, as well as being a subtle moderating power and also directing, perhaps driving traffic. This protocol is ethics, evoked by factors that are seen as relevant, artificial and necessary for life to happen in its best way, also sustaining the structure in the best way. (Ortolan, 2022, p. 500)

Although in a subtle way, the aim is to optimise an AI so that it is not so "human" in relation to the inhuman "truths" that come out of the correlations. The question of algorithms is whether they are (intentionally) oriented towards humanity, in the sense of being anthropomorphised in behavioural terms, to the point where they possess certain human abilities such as ethics; because some of these abilities are desirable,

others are not, but in adverse situations, what was undesirable can become desirable, or vice versa.

In this way, if it's desirable and technically feasible, we "hallucinate" the digital discourse a little, purifying it of current prejudices and, in this way, adjust everything with a pinch of a utopian discourse (preferably Kantian, with its categorical imperatives) that is neither current nor common among humans, as it is a discursive ideation with no practical correlate: there is no such discursive politeness on a regular basis beyond one or two everyday moments, and therefore what is intended as a result is really a hallucination, by definition. But, as a hallucination, this expected result, if attainable, will come to exist there in the digital and, therefore, in this case, what was previously undesirable will be allowed to happen: we will start to hope that AI can re-ontologise the world, for the better. The game is easily reversed. Hallucination and re-ontologisation turn from villains into heroes. And it all happens pharmacologically, as (Romele, 2019, p. 12) cites: "*pharmakon* is a Greek word that means both 'poison' and 'medicine'."

Thus, it is hoped, veiled by optimists, that AI can be an instrument of moral progress, through prevailing ethics. And this is, in fact, a human intention projected into the *design of AI*, without a doubt. For such progress is aimed at the ideological objectives of *big tech*, which we cannot fail to mention are eminently capitalist, liberal or neoliberal.^{lxxi}

Ezell & Crowther, (2007), in the section of their article in which they argue that bias in system *design* is inherent and unavoidable, give as an example the "cases in which political and social systems are often left aside in favour of developing functions and evaluation measures for the technical component of the overall system", as they claim that "social systems are left aside because the realist cannot observe or measure them

^{lxxi} "The global shift towards liberal morality is best explained by cultural processes that do not track moral truth: moral perspectives that respect robust rights and treat people as individuals with equal dignity and discourage violence are culturally attractive because of the safety and security they provide. More liberal societies also tend to be better at ensuring co-operation, which makes them less punitive and wealthier. The tendency towards liberalism can thus be explained, and explained better, without assuming the existence of mind-independent moral facts." (Sauer et al., 2021, p. 7).

empirically" and this reinforces the lack of an ontology or method that allows such measurement or monitoring.

There is a circularity that will need to be seen as such at some point. Thus, the bases of learning already carry unreasonable values and prejudices that need to be pharmacologically purified and, with this, it seems that there will always be a gap between actuality and its respective representations, which provides an even more fertile ground for hallucinations, as well as other derived problems.

Therefore, in this first phenomenological phase of AI, especially conversational AI, there is a certain reflexivity (and excessive trustworthiness) with the harsh and prejudiced reality, in all its discursive brutality. This excessive trustworthiness is, unusually, the source of most of the ethical problems, also caused by the fact that there is no ontology sufficiently capable of providing language models with an efficient representation of the world.^{lxxii}

So we can suspect that if problems occur, it's not just because of "errors" and misrepresentations, but also because of the accuracy of reproducing the current prejudiced discourse reliably, without being able to apply eminently human behavioural filters through certain standards applied to either the *inputs* or the *outputs* to manage them qualitatively, as we do with some of our human abilities known as "soft skills" or emotional intelligence. And AI, which is all too human, has invested itself with this viscosity of ours.

1.4. The many ethical problems

In this section, we'll address the ethical problems and how we arrived at them, since we have enough elements, from the philosophical and technological points of view, to

^{lxxii} "The problem here is not that the great language models hallucinate, lie or misrepresent the world in some way. It's that they're not designed to represent the world; instead, they're designed to convey convincing lines of text. So when they receive some kind of database, they use it, in one way or another, to make their answers more convincing. But they are in no way trying to convey or transmit the information in the database." (Hicks et al., 2024, p. 38).

address the ethical problems mapped out. We will start by emphasising that the current analysis of ethical problems is compromised by the way it is carried out, either by the static approach to an eminently dynamic structure, or by the teleological difference between what an AI is and what it is intended to be.

1.4.1. The map of ethical problems

For us, the relevance of the map of ethical problems drawn up and presented by Mittelstadt et al. (2016) is precisely due to the structural basis of the categorical representation that is adopted, since this is based on the operational linkage of the AI algorithms themselves: the map, as well as presenting the ethical problems, also indirectly offers a representation of the *modi operandi* that make them exist - the map speculates on the genesis of the problems highlighted

And that's precisely what we've been doing and what we're aiming to develop here: to draw up a map of the problem that is itself dynamic, that is more detailed and related to the causes duly categorised between them, and preferably in overlapping layers, made from a relational analysis.

The map is therefore not intended to be a tool to help solve ethical dilemmas arising from problematic actions conducted by algorithms, but rather is presented as an organising framework based on how algorithms operate, which can structure future discussions on ethical issues. (Mittelstadt et al., 2016)

The map of ethical problems presents, in a reductive way, six ethical problems about artificial intelligence algorithms, with the first three grouped as epistemological issues, the next two as normative issues and the last one, isolated because it is derived from the previous ones, and without there being a declared thematic issue, concerns the traceability of the entire algorithm process and also about accountability, as in figure 2.

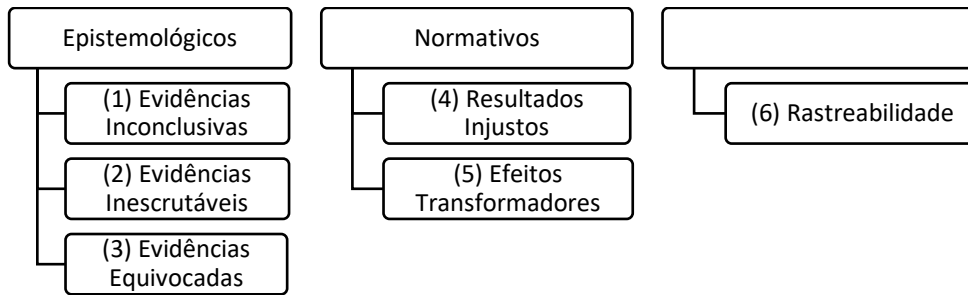


Figure 2- Map of Ethical Problems, according to Mittelstadt et al. (2016)

It is therefore valid to conceptualise each of the problems presented on the basis of the (manifest) consideration given to them by the authors, but already adding to it with our collaborative hermeneutic critique, based on what we have developed from our concepts elaborated so far and now operationally evoked.

(1) Inconclusive evidence (due to the predominance of correlations)

Manifesto: These problems are based on the absence of causal relationships and the omnipresence of correlations, with no theoretical basis to support them beyond probabilistic calculations. In this way, it's a problem for AI in general, and an even bigger problem for conversational AI, since it will have as evidence (and will be able to present as *output*) only what is most likely to occur, but not, not necessarily and not obligatorily, what is derived from a causal relationship, or that represents knowledge: credible, true and justified.

Collaborative critique: we believe that the problem statement is valid, however, an AI is made on the basis of correlations. Without correlations, it couldn't even exist. There's a previous problem to be solved here, because you're expecting an AI not to be what it is.

(2) Inscrutable evidence (due to the secrecy and obscurity of the data)

Manifesto: These problems are based not on the impossibility of knowing the "reasons" why algorithmic operations were carried out, such as the correlations (and which could be *outputs*) in (1), in order to arrive at some evidence, but rather they are based on the fact that it is not possible to know anything about the data (about the *inputs*) that were considered for such correlations and, therefore, there is no justification possible to be stated to support the validity of the evidence, and AI is assumed to be dogmatic.

Collaborative critique: What if we thought that all the data was available? Given the large volumes of data and the way we interact with it, what would be the assertions about possible scrutiny? How do we deal with raw data and realise a dynamic so complex that we have developed a technology to deal precisely with this complexity that we can't reproduce in quantitative terms? What is expected, in fact, is for the data to be presented in an argumentative way, to support the evidence generated. And this is not as simple as one might think, especially after the whole process has been carried out without a classification that can offer traceability of decisions.

(3) Wrong evidence (nothing is created, but everything is replicated)

Manifesto: The term "rubbish in, rubbish out" is used metaphorically to show that the domain of an evidence will never be greater or different, quantitatively or qualitatively, than the training bases, because there is no true creation in an AI, in the sense of generating something new, unprecedented or transcendental, since there is only replication of part of what already exists (because not everything that exists today is digitally available as data) and, as a result, the algorithmic dynamics themselves do not allow for "corrections", because that would mean creating something different, or new, and so they end up perpetuating errors. All inconsistencies, of all kinds, whether epistemological or ethical, for example, will be replicated as they exist in the data representing a current bias.

Collaborative criticism: the manifesto perfectly elaborates on the problem. And this problem will lead us to return to the discussion about the level of "hallucination" that an AI needs to have, pharmacologically, in order not to reproduce so faithfully what is in its overly human bases, so to speak. There's also a possible misconception that the digital is a subset of the present, which we've disagreed with before, however, and have already argued for the right reasons.

(4) Unfair results (due to the sentimental impact on the user)

Manifesto: There is here, as a conclusion in the presentation of this specific problem, the assumption that "an action can be considered discriminatory, for example, solely because of its effect on a protected class of people, even if it is taken on the basis of conclusive, scrutinisable and well-founded evidence" (Mittelstadt et al., 2016). The authors are thus considering impacts as fair or not, from an individual perspective, in relation to the user who will be impacted by a particular outcome.

Collaborative critique: and what would be the qualifying elements that could adapt the results to each of the particularities, in other words, to each of the users? Therefore, the justice considered is obviously not static, but dynamic, from each individual perspective. And this opens up an even bigger gap for the defection of the Parmenidians, as we realise that absolutism needs to give way to perspectivism.

(5) Transformative effects (imperceptible re-ontologisation of the world)

Manifesto: There is a suspicion here that actions can be taken on the basis of apparently neutral results that will nevertheless have an impact on both the user and society, as the article considers that "algorithmic activities, such as profiling, re-ontologise the world, understanding and conceptualising it in new and unexpected ways, and triggering and motivating actions on the basis of the information it generates" (Mittelstadt et al., 2016).

Collaborative criticism: The question is how to measure these ontological changes? How are we doing now? What bases should be considered in order to understand whether or not there is a reontologisation? In addition, there will also be an opening for new discussions, because if in (3) there is precisely the qualitative and quantitative limitation of everything that exists (by existing data) and that is replicated, then it is assumed that an AI creates nothing, it only replicates, which is coherent. Therefore, a reontologisation would not necessarily be the result of an algorithmic strategy, but rather a human creation that becomes represented as new data, as a digital counterpart based on new forms of human creative manifestation. There must therefore be some strategic intent

in the *design of* the algorithm so that such an effect - first on humans and then on AI - can result. There is a strong circularity here.

(6) Traceability (through agency accountability)

Manifesto: In this issue, what is really at stake is agency and its accountability: whether it is eminently human; or, whether it is eminently algorithmic; or, whether it is mixed, in contrasts, with human agency or merely influenced by the algorithm, strategically or not, or even with its declared co-responsibility. The aim, therefore, is to understand the quantum of responsibility that an algorithm may have in the event of a user acting on the basis of its generated evidence, or whether there will be full responsibility for the algorithm itself if it is able to act autonomously in certain functions that have unexpected consequences, such as autonomous vehicles in cases of accidents with fatalities.

Collaborative critique: Deliberation is needed on whether an algorithm can be considered an autonomous agent, capable of exercising actions, direct instructions or even indirect influences strategically outlined in its *design*, whether declared or not. In order to do this, other questions must be raised, such as those dealt with in the section on the intentionality of capital^{lxxiii}, which is the origin of any commercial project, and AI is no different. Dialogue is supposed to be appropriated as a product through conversational AI. And this appropriation begins merely with human intentionality, when functionally conceptualising an AI. However, making it a large-scale product is done through the intentionality of capital, obviously within "its" interests. In order to make progress towards a correct deliberation, it would first be necessary to realise

^{lxxiii} We consider "capital" to be a general concept of capitalist action, with its *players* and particularities, which we will gradually build up. The critical and analytical positions assumed, however, will not necessarily be either pessimistic or optimistic towards capital or its interests, since we will seek to develop the phenomenology of AI (and dialogue as a product) and, for this reason, we could not disregard capital and its "intentionality" as necessary elements for any contemporary product to exist, beyond the obviousness of capital's own intentional reproduction. In semantic terms, our reference to "*capital*" could give it a phenomenological, almost anthropomorphic *status*. It will sometimes seem that way, because the capital we are considering will be presented as a dimension that dialogues with and for humans, and through them, both to reflect human intentions themselves and also to satisfy them and thus reproduce itself through profits.

whether there is a difference (and if so, what it might be) between human intentions and those of capital. In other words, what do we believe or desire or want an AI to be, anyway?

1.4.2. A necessary ontological revision for the map of ethical problems

From our collaborative critiques based on the manifestos of the problems mapped by Mittelstadt et al. (2016), we can work on it ontologically, ascensionally, to seek to revise it ontologically.

Therefore, as part of the readjustment (and without conceptual prejudice) of the groups of problems originally proposed, we can keep the first group as the epistemological ones. But we believe that the second group would be better called relational, rather than normative, because normativity is something that is, in principle, within all the groups. After all, even for the epistemological problematic of the first group, there is a dose of norms in force that are even stricter than in the others, such as the rules of logical inference, for example, as epistemology demands, or even about testimony, which lacks certain normative requirements in order to be considered knowledge, just as examples. Naming one group as normative may imply that the others are free of norms, or that they are not so relevant after all.

The third "group", which was not originally named, perhaps because it only consisted of one item, would not prevent us from naming it anyway and, at least for didactic purposes, we can consider it to be the accountability group.

In this way, we have three groups of problem levels, duly revised and renamed to be worked on, as shown in figure 3.

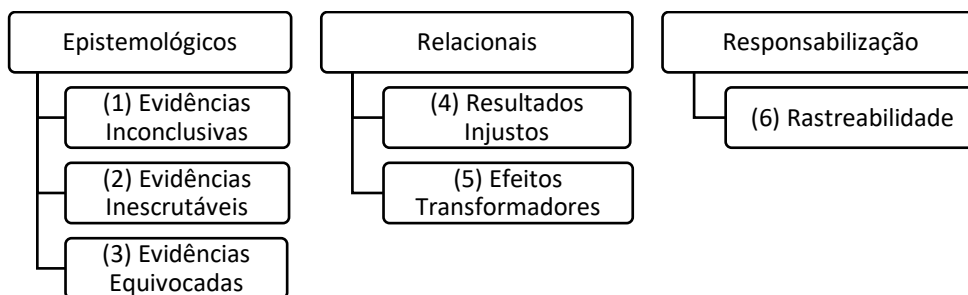


Figure 3- Map of Ethical Problems (ontologically revised and renamed), based on figure 2.

It is also worth pointing out that the first three issues^{lxxiv} (from the first group, epistemological) were defined by the authors as evidence and therefore, at this level, are not yet necessarily exposed to users. We represent these schematic flows in figure 4, according to the reference article considered. Evidence (the objects in the first group) can be interpreted as the result (data, therefore) of the algorithm's internal operations, before being exposed to the user as *outputs*. They end up also being data, like objects, or informational consolidations - digital traces, however, but still exclusive to the algorithm and therefore subject to further operations / processing / reprocessing until an acceptable level of optimisation is reached.

For the second (relational) and third (accountability) groups, respectively and cumulatively, they cover *outputs* and actions (agency). *Outputs* can be information that is sensitive to the user or just the final operational result of an algorithm, for a given task that is "fulfilled" or operationalised by the user. Action can be either what the user undertakes from the *output*, according to their own deliberations, or what can be an action by the algorithm, which in this case is considered an agent itself due to the "deliberation" it is capable of carrying out. In cases of conversational AI, it is more likely that all actions are those of the user. So, after the epistemological level, here there is a delivery of evidence as a result which, already as an *output*, moves on to the relationship level: it becomes sensitive and impacts the user.

^{lxxiv} It is important to note that the majority of the questions considered here will be dialogue-based, i.e. they will require a degree of subjectivity and intersubjectivity, as opposed to questions whose results are merely objective. A question that specifically asks for data or a numerical value will always be considered objective. Whereas demands for meanings, syntheses, interpretations and all other forms that require transcendence of objectivity will fall under the considerations that have been and will be made here.

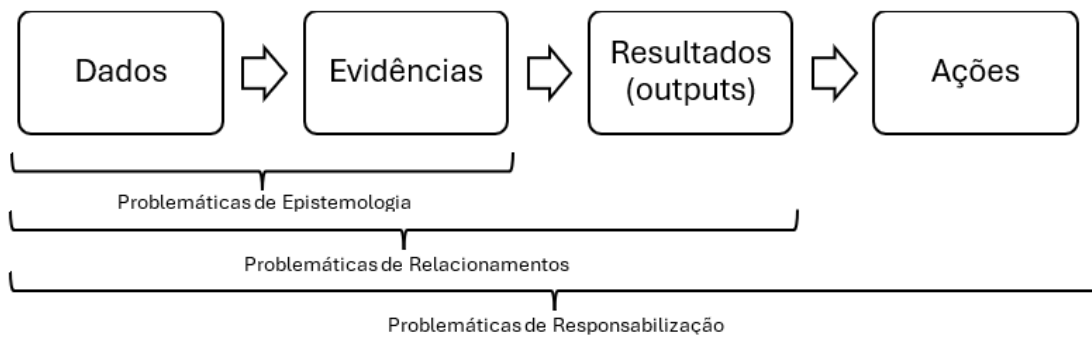


Figure 4- Algorithmic flow schematised from Mittelstadt et al. (2016)

With the problems duly presented, it's time to present the knowledge map of the article's arguments, as shown in figure 5. This first map was produced faithfully from the conceptual constructs contained in the article, in both a linear and cross-referenced reading, and which focuses on the argumentative and/or speculative constructs of the six problems considered as such.

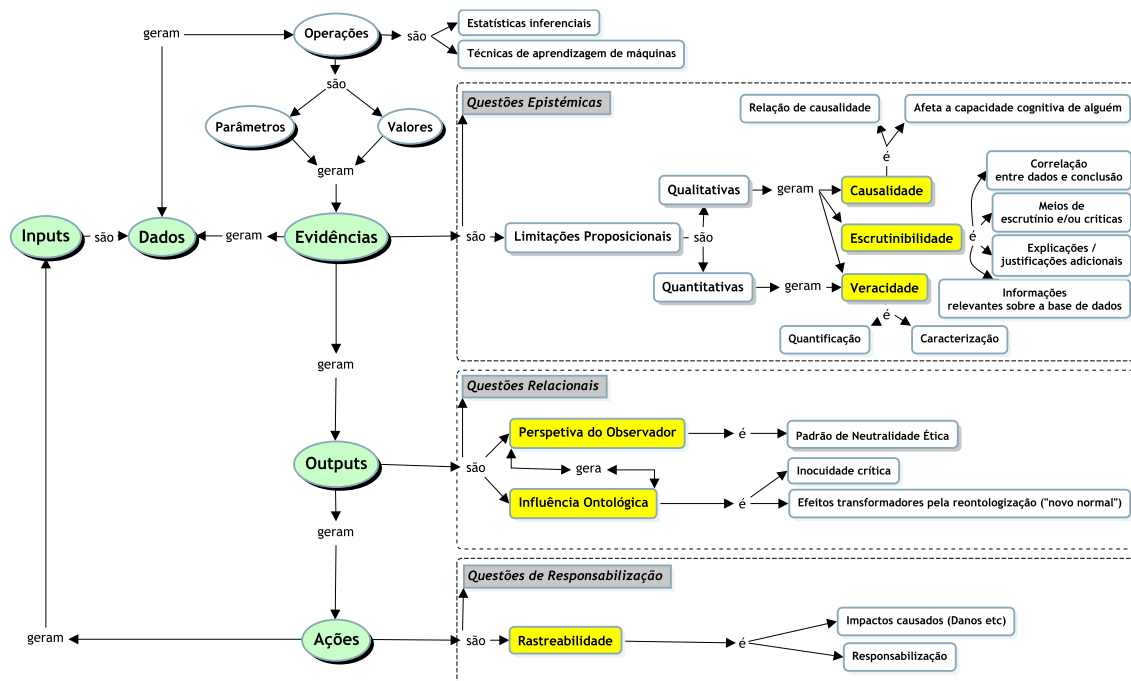


Figure 5- Representation of knowledge about the ideal AI model, functionally schematised according to the referenced article (Mittelstadt et al., 2016)

The representation in figure 5 is supposedly constructed as the ideal form of an AI, with all the qualities and no flaws, in its best form, without any problems and so that we can

simulate and trace, based on these concepts, the *modus operandi* that will give rise to the problems accused.

In other words, if the author considers a system to be unscrutable, it is because, in his idealisation of AI, he presupposes scrutability in its *design*. For, in order to trace any "defect" in machines (and, at a glance, we can even consider an AI as such) there is first a perception of how the system should operate in an idealised or conceptualised way, and then how it is operating. If there is a difference between the operational project and the current situation, there is an anomaly, and if this is counterproductive, then there is a probable defect. To realise the defect, you have to go through the systematic step-by-step operation until you find the fault (*bug*).

There is something hermeneutic, however, in the interventions of technical assistance and the repair methodology is precisely to go through the existing circularity to verify the dissonances between the operational design of the *gadget* (or its conceptualised "normality") and the defective situation (or its "actuality"), between the operational whole and its supposedly defective parts^{lxxv}. Hermeneutically, we will do the same, giving credit to the idealisation of the authors and considering the model they are considering as ideal, formed from their manifestos of "defects". In this way, we have a reliable representation of a properly schematised "system" which, before going on to discover where its "defects" originate, will allow us to reveal its true teleology, based on its operational functionalities. We suppose we know what an AI should do, based on the teleological idealisations that exist, but we haven't yet defined what it actually does, in its maximum power, and within its functional limitations. The investigation therefore moves on to the teleological dimension.

^{lxxv} What is sometimes overlooked is that the defect may not be in any defective part, but in the mistaken idea of normality that is purely idealised, without being consistent with the reality of things. And so there wouldn't even be a defect, but rather a frustration about what you want to achieve.

1.4.3. On artefacts and the emergence of teleological dissonance

Can a fan fail to cool a room on the hottest summer days? If it can't, is it faulty? Is there a problem with it?

These unusual questions give us an interesting metaphorical parallel about "AI problems", about the divergences between what is being designed in terms of its original functionalities and what is being demanded or expected of it, about ethical performance. It is necessary to realise whether there are such metaphorical similarities.

Let's go back to fans. In the summer, when the temperature exceeds the so-called pleasant levels, technology is used to adapt rooms to cool them down. A fan, the pioneering and most popular artefact among all those that aim to improve local thermal conditions, circulates the air in the room, helping to evaporate sweat and therefore providing a pleasant (or less unpleasant) sensation for those present, by cooling the skin and mitigating the feeling of suffocation. It's very appropriate for hot people, but it's impossible for a fan to reduce the room temperature on its own if the room is hermetically sealed. It causes an illusion of coolness and nothing more. That's why the answer to the first question is: it couldn't stop cooling, because it never really could - due to its constructive characteristics and operating conditions. There is a supposed teleological attribute referred to the fan (cooling) without it having the technical conditions to fulfil it and there are users of the same who, because of this, when frustrated by the heat that doesn't dissipate, consider it to have problems and disqualify it, even for not doing what it couldn't do.

Under other conditions of use, if there is an opening to the outside, the movement that its propellers cause in the ambient air can facilitate circulation (and renewal) between the internal and external air, depending on its position in the room and also due to its technical characteristics. If the external temperature is different, then the internal temperature will tend to be the same as the external temperature, equalising them, as in calorimeters. And this may or may not be appropriate, depending on the difference between the temperatures, because the fan will only replicate the external temperature to the internal one, regardless of whether it is "better" or "worse".

It's different with the air conditioning unit, which requires the environment in which it operates to be hermetically sealed for its best performance, since it operates with the aim of "warming up" the outside world by removing temperature from the "inside world" up to the limit desired by the user, who cools down in a "cooler" environment. The air conditioner "knows" the temperature differences because it has sensors that are part of its thermostat. The operating principle of an air conditioner is to be the exclusive mediator between two dimensions isolated by it, and only by it, without them having any other communication with each other. ^{lxxvi}

The conditioner is therefore a heat extractor, a purifier of existing heat, just like a fridge. Heat is the asset it operates on, which in our metaphor could be compared to data, if we think of an algorithm as an extractor and operator of data. That's why it's important to observe not only the order in which things happen, but also their teleologies, because if an air conditioner is incapable of heating the outside by removing the "hot" from inside and sending it outside, it won't be able to cool the inside, and it will fail teleologically as an artefact. This is why a perfectly hermetic division is required between two dimensions, which need to be insulated so that it can act on heat exchanges. Because if they are not insulated, the work for the air conditioner to operate the differences will be greater and it may not do it well enough, and soon the results of its efforts will not be realised operative, or perhaps even its task will become unfeasible. The fulfilment of teleology is conditional on the right operating conditions.

The teleology of an artefact, therefore, will not always necessarily be the same as the impression made by its operation. A fan's teleology is to promote the circular movement of ambient air, and its teleology is not to change the ambient temperature, due to

^{lxxvi} Sometimes, those inside the air conditioner's operating process, i.e. in the "hot" environment, believe that the air conditioner fulfils its primary mission of reducing the internal temperature. But this is an illusion, because the opposite is true, according to the laws of thermodynamics. Cooling the internal environment is its side effect, not its primary cause, from a mechanistic point of view. The artefact ends up meeting the user's intention, but in a different way to the impression they have: the user "develops" the illusion that the conditioner is giving them the cold, but what it actually does is remove the excess heat by removing the "hot" that was already there. Cold is not the primary asset or raw material, but heat. Cold is therefore the absence of heat.

physical impossibilities. But this is what its users deludedly think and expect of it. It may change the temperature, but this is an unpredictable and occasional side effect that is independent of the artefact and depends on the external heat conditions: it depends on the quality of the "data" that will reach it.

Manufacturers, aware of these physical limitations and also aware of users' desire to cool off, have added accessories to fans, such as systems that rotate the fan body back and forth to increase its performance. They also incorporate water sprays to increase the humidity in the room, which is very favourable on very dry days. They've put holes in them to hang them from walls, or created models to be installed in ceilings, elaborated bold *designs* in the shape of towers, and many more items of value, such as aromatic odourisers and even dust mite sterilisers. They extend the illusion of the user with such *upgrades*, who pays more for them, but they don't even go beyond the teleological limits intrinsic to the artefact's primary category.

AI is like a fan. Its teleology is to replicate internally what it has learnt correlationally. If a_1 then b_1 . If a_2 then b_2 . So, after n similar correlations, it has "learnt" that whenever a_n then b_n , and it will replicate this correlation whenever there are such "enes" in its domain, in the interior of its operating environment. It is merely an automaton, albeit a sophisticated one, but still an automaton. It operates within a closed environment (its database) and creates nothing from there: it's not capable of going beyond its imposed teleological limitation, because it limits itself to ventilating its database (to the internal environment, just as a fan does) and only gives the illusion that it's making something different happen by moving what's inside.

Inside a closed environment, you can feel a slight breeze on your face from the fan, and this sensation can make you imagine you're on the summit of a refreshing mountain: but these are just illusions that are dispelled when you check the thermometer, which will still be the same as before. An AI only moves data, nothing more. It works well and successfully within the same data environment. But if there is an opening, with new data coming in, what will happen will also be similar to the fan: and the AI will match the quality of the external data, for better or for worse. But it won't create anything either,

again, it will only equalise what is there, at most if it has been well performed by learning.

The *designers of AIs*, like those of fans and all commercially exploitable products, develop upgrades for them to increase their commercial value, but which also only cause new illusions because they can't make them go beyond their own teleology of replicating what is there, as the correlational optimisers that they are. They haven't changed their "essence", but only their "appearance".

A conversational AI is really a version of an AI with a more sophisticated upgrade, which is to obtain coherence in natural language for its results. But if it is teleologically impossible for a fan to operate like an air conditioner, it is equally impossible for a conversational AI, no matter how sophisticated, to operate with a teleology other than its own, such as human teleology, for example, or even the longed-for (or prophesied) artificial superintelligence (Bostrom, 2014). A fan will always be a fan, Aristotelian assertion. An air conditioner, however, can operate like a fan. There is some *bottom-up* teleological limitation, but not *top-down*. Some upgraded air conditioners can even reverse the operation and controllably heat an indoor environment instead of cooling it. There are other versatilities that air conditioners can have in their upgrades, in various other functionalities.^{lxxvii}

^{lxxvii} Out of an unbridled desire to surpass the original teleology, some users even make hacks with fans, plugging them into thermal boxes with ice and all sorts of other arrangements to make them ventilate cooler air, but it never works out to their satisfaction. Even the best fan hack can't match the worst air conditioners in terms of performance, consistency and durability. There's a teleological gap that can't be overcome by inferior equipment. The problems, then, occur precisely when users want to overcome this gap by seeking to modify the teleology of an artefact. Because an artefact's design is something functional: it operates as a function of something/effect that generates another something/effect, simple as that. A fan "ventilates", a conditioner "conditions" and so on. When each thing is expected to do exactly what it is designed to do, what problem is there? Certainly not, because everything will be running as it should, if it's operating normally. But there are those who want a fan to operate like a conditioner and, in addition to frustration, they will have the side effect of failure, which they will call problems and which, perhaps, will be many and will even map them out, to be discussed among those who share the same naive pretence of teleological surpassing. The problem exists, but it doesn't come from the artefact, but from the user's mistaken pretensions about the artefact's functionality. The problem, in fact, is the heat, not the fan. The problem needs to be solved with another artefact, designed specifically for this purpose. Similarly, there are those who want a conversational AI to be and have what it can neither be nor have: a differentiated teleology.

The map of ethical problems presented by Mittelstadt et al. (2016) does just that: it seeks in a fan what only a conditioner can give. Figure 5 shows that an AI can't offer things like traceability, or even causality, or almost anything else that would be pointed out as an operational aspiration for it, because it wasn't built for that. The operating conditions and environments designed for AI to operate are wide open, both for data and users, without any kind of isolation for it to succeed. What is expected of it are functions that we should expect from something that may yet be built, with a better teleology and that will certainly happen, given the immense resources and investments that are being made in the segment. ^{lxxviii}

The critics point out the problems (mere effects) without realising the real reasons (causes) that generate the impossibilities. The map mistakenly focuses on the problems without any correlation to the teleology considered, because it doesn't state what it considers an AI to be - teleologically - beyond describing what an algorithm is. It discusses the "how it operates" of AI, and not its "why" it operates as it does. This is what needs to be resolved. ^{lxxix}

The new question is: what would this possibly differentiated something be, anyway? How can we map out the values often expected from this something so different? Because, as we'll soon see in section 2, there is the intentionality of capital (and also human intentionality, or both) driving any (extremely expensive) AI project. And if there

^{lxxviii} AI, still a fan, is on the verge of being overtaken by its "air conditioner". That's why we insist on liminality as the current phase of AI. And that's also why there's a chance of betting on the continuity of AIs, because even efficient air conditioners haven't caused fans to cease to exist. What we're looking for is a refinement in a new conception of AI, based on the present, but still different from it.

^{lxxix} But this is humanly justified, because a product is never just a product, it's also the amount of fetishisation that goes into it. And it is this fetishisation that makes its exchange value sometimes immeasurably higher than its use value, but not only. And the collective fetishisation that is embodied in demand, to the detriment of an offer made after (or while) the consumer's fetish is sometimes aroused. And it is this strategically provoked fetishisation, sometimes unnoticed, as in these considerations, that leads a product to "be" what it "is", when it is overvalued and therefore an object of desire worshipped as such.

is investment, there is the premise that the product will also be aligned with the user's intentionality - whether naturally formed or strategically excited to do so. ^{lxxx}

It is therefore necessary to explore the last representation (figure 5) and realise how the problems identified as the most relevant to AI are concentrated, in isolation, in a non-reductive way. ^{lxxxi}

1.5. What's this "here"

Throughout this first part, that of problematisation - how did we get here? - we realised that what we have in terms of AI is not exactly what we want (or think) we have. The problems mapped out exist, after all, and are closely suited to what an AI really is today. For this being of AI, phenomenologically, is also invested with what it can give: a created being incorporates a teleology, through the intentionality of the creator that is employed in the act of creation. And AI was not created to be ethical, but rather to be powerful and fulfil some of humanity's most emblematic needs and desires: to transcend.

So what we have are not really problems with something that "went wrong", because AI is an impressive success story. The problems are intrinsic side effects of current AI projects, which do exactly what they are supposed to do. And this "should do" is totally unrelated to the perceived effects. Therefore, the problems should not be tackled on the basis of what currently exists in terms of AI, but rather should serve as a basis and inspiration for what will lead us to have, in the future, something different from what

^{lxxx} If dialogue has come to be seen as a product, AI is its producer and distributor, which, when it is "triggered" to work, has the premise of delivering something different in terms of experience and results to each of its users and thereby promoting their satisfaction, just like any product that is marketed to its target audience. There is, of course, a generalised interest, especially on the part of capital, in developing an improved product that promises (and delivers) the best of all worlds to consumers, or rather users.

^{lxxxi} In other words, where, when, how and why problems begin to form until they are constituted as such. Obviously, such problems, even if they can be solved, will never make an AI transcend its own teleology, as expected. It's not the absence of problems that clears something of its teleology, because they are linked to the very limitation that this something possesses.

we have today in terms of AI projects. These current problems may be surmountable, although other types of problems, more or less serious, may arise.

AI will certainly evolve into something different, but based on what is undesirable today. Such problems have a noble function for the evolution of AI. It is reminiscent of the theory that evil begets good, very much in tune with Nietzsche (2008), 's genealogy of good (when he develops the idea of the instinct of the sufferers, when the weak and unsuccessful, full of resentment, start to seek revenge against the strong and happy, projecting their valiant ideals of weakness and submission worthy of being followed by all. In other words, in moral and practical terms, good will come from evil. It's a kind of moral dialectic, when something is always opposed and, from this oppositional clash, what results leads to an improvement in conditions. Very similar to capitalism, which feeds off its crises to further deepen its productive and expanding roots.

The predictive risk, however, is that Nietzsche realised that this was a symptom of a decadent and sick culture, when everyone begins to deny life and turn against their most basic impulses, which are eminently human, all too human and all too forgotten in contemporary times. And Nietzsche railed against the fact that, instead of seeking self-overcoming and the affirmation of the will to power, they promoted passivity, resignation and the levelling of inferiority. And just like that, we run the same risk of reaching the effects of Nietzsche's time, but now with a new name, because with AI they're calling it the reontologisation of the world.

So, after realising what we have (based on the problems - the supposed "evil"), and how we got here, the next thing we have to do is understand what this "good" would be: what we want with a "new" AI (the one that is supposed to bring "good" - through the absence or overcoming of the current ethical problems), in terms of the will to power, or whatever, and in terms of intentions, whether human or capital, if there is any difference between them. Perhaps capital is what we have left as a will to power. The time has come, therefore, to transcend what we have, towards the future.

2. Dialogue as a product - the possibilities for AI

2.1. Where and when do we want to be?

The future, therefore, in relation to AI, no longer seems to be science fiction^{lxxxii}, due to the proximity in time that all the possibilities seem to have acquired. So, even within the *hype*^{lxxxiii} of AI, speculating about what it might become seems to be closer to a futurological conjecture (which it won't be, however) than the creation of science fiction.

Despite this, we need to be very careful that what we are considering in this second part are, in fact, considerations (albeit speculative, but moderate and connected to the present, in terms of intentions and actions) about what we think AI will be like in the future, subject to certain limiting criteria about such considerations: to consider only that a new AI *design* has the minimum change necessary to have overcome the current problems, even though others may exist (and probably will). The *design* will therefore have to be reconstructed from the three ontological dimensions developed on the basis of their actions so that they can mitigate the current problems. It also needs to be within the constructive possibilities available: it needs to be feasible.

That's why we're not going to try to design an entire AI project specifically, but rather consider how it can take place in terms of decision-making processes between an *input*

^{lxxxii} So far, there is a group of people who are optimistic and completely enthusiastic about the possibilities of AI in general. There are also those who are pessimistic about the developments that could occur beyond the digital dimension. There are theorists who are focussing only on certain problems, in themselves, isolated from the whole. There are all kinds of people with a position on AI, whether favourable or not. Memes abound on social media. Content created by AI is becoming more popular every day and is difficult to differentiate as such. Everything is happening around the phenomenology of AI. However, there is hardly anyone who is indifferent to it all. That's why AI is considered, and is said, to be in full *hype*.

^{lxxxiii} "Hype is the extreme promotion of a person, idea, product, which usually lasts for a short space of time. It's the trending topic, it's what's "making all the buzz" or something that everyone is talking about and commenting on - what's in fashion. The word derives from hyperbole, a figure of speech that represents the exaggeration of something or a strategy to emphasise something." (Wikipedia, accessed on 07/Jan/2025 via the link <https://pt.wikipedia.org/wiki/Hype>)

and an *output*. It will also be necessary to take into account the intentionality of the capital, the users and these problems, based on what we have discussed so far. Because even before technical resources are made available to make something feasible, market intentionality itself needs to be activated . ^{lxxxiv}

In this way, this "where" we want to go is not merely a temporal speculation of futurology, although it may seem to come close to that, because it is much more a pretension of spatiality, of a dialectical relationship between the representational idealisation of the future and the territory of the present - merging as an environment in which there are answers to our questions. That's where we want to end up. And so we will.

2.2. The fourth dimension: capital as an intentional agent

In the first part, we were able to make progress in corroborating the argument that the problems currently arising from AI are not necessarily due to the characteristics they possess in constructive or technological terms, because the algorithms are doing exactly what they were designed to do. ^{lxxxv}

The origins of the problems emerge from a teleological dissonance^{lxxxvi} . And, in fact, the purpose of an AI, in linguistic terms, could not be different from this: to generate

^{lxxxiv} Because even if there is only a market intentionality (a promising market to be exploited) and there is no technical feasibility, capital will direct itself towards creating ways for the unfeasible to become feasible: and AI is a good example of this, as a means to certain interests of market intentionality.

^{lxxxv} There is what AI is, as we currently have it, which we dealt with in the first part. There's what it's supposed to be (albeit from different perspectives: from humanities theorists, users and *big techs*, among others), and that's what we'll deal with in this part.

And finally, as we'll deal with in the next section, there's the liminality of AI: which is the transition (the path of development in the midst of a sometimes irrational consequentialism - in the *hype*) between what is and what will be, because AI itself is a product and, as such, is being operated by competing forces, sometimes antagonistic, which are configured as a major force, which can be the market itself, capital, *big tech* or other similar instances.

^{lxxxvi} We illustrate this argumentative corroboration metaphorically by teleologically comparing a fan with an air conditioner, as if we wanted a fan to do something that it is not capable of doing: lowering the

linguistic coherence - to generate syntactically and semantically coherent sentences. From there, there will be no guarantees for any other teleologies for an actual AI .^{lxxxvii}

The current problems of AI, therefore, cannot be solved by the current algorithmic constructs, given the teleological disparity. They do, however, serve as benchmarks for new goals of overcoming (a sense, even transcendental) what is to come, at least on the part of humanities theorists, especially on ethical issues and social, environmental and political impacts, but not only.

And for this very reason, of the three dimensions that have already been [ontologised](#) based on the problems and operational representations of AI - the (1) objectivity of knowledge, the (2) intersubjectivity of relationships and the (3) subjectivity of individuality - it must be considered that they will all be (and are already being) constructed. And they will soon be digitally represented as such, somewhere with extreme precision and accuracy, which will not be long in coming due to the growing efforts of the capital involved, given that huge resources are already being consumed in investments and developments.

It would be naive (or at least hasty) to say, however, that such investments are merely motivated by overcoming ethical problems, or by any other intention and/or disinterested action in favour of the common good. That's why there is a need to question the interests, or better, the intentions of capital (beyond the obvious profit motive) in seeking to commercially exploit dialogue as the product it has become.

temperature, even though it may seem possible due to the sensations of freshness that empirically confuse its users.

^{lxxxvii} Thus, what is currently presented as a problem generally turns out to be a dissatisfaction with something that we want to overcome and that will result in new technological developments for what an AI could become, both quantitatively and qualitatively. This advance will be based not only on what we no longer want (the mapped problems), but also on what we have always wanted and intended, in epistemological, creative terms, in favour of success and fulfilling satiety, but not only. The "good" and the "bad" come together as the driving force behind development.

So nothing could be fairer (and more providential) than considering - at least conceptually - a fourth ontological dimension (the intentional, but not only), which could be the operator (in terms of a unified sense, purifying or synthesising multiplicities) of the three previous dimensions, which are basic and complex^{lxxxviii} . This fourth dimension, if this is the case, can be considered fundamental, but as a provisional guideline, and not to the point of being a Parmenidean referent. In relation to the other three, it can have two characteristics, or natures, in theory:

- (1) It's superior and autonomous in intentional terms (just like a genius outside of existence: be it malignant, benign or even bipolar, or whatever) and can therefore have some kind of consciousness and capacity for deliberation within itself in order to then operate the other three through its supposed management capacity, which thus gives it a certain hierarchical superiority over the others, just like a powerful omnipresent, omniscient, creative, influential, providing, justifying and, why not, strategic entity. It is the facet of the dimension that promises that dreams (beliefs, desires and wishes, but not only) can be realised. It is therefore the facet that "promotes" the generation of offers.
- (2) Inferior and reflexive to the other three, as being the common and consubstantiated sense of the existing, and subordinate to this desired and believed possible among so many dimensional demands capable of being received and apprehended and that, therefore, all this accumulated reflexivity merely represented, and without operations or interferences, the existing possibilities. Just like a "universe" of possibilities. This is the facet of the

^{lxxxviii} If this fourth dimension is technically capable of operating and/or influencing the others, in strategic terms, to make their users behave in certain ways or adopt certain attitudes, other possibilities open up. Because, it must be admitted, if there is control in a "negative" way, the same modality in which this control is exercised can be adjusted to neutralise other forms of control and will therefore work towards "neutrality", without allowing "negative" or "positive" influences to impact users.

The new question, however, would be how to classify what is negative or positive. For research purposes, it is prudent to adopt a scenario in which such control takes place, in order to also consider that neutrality can be managed. So what needs to be questioned for now is: how can we arrive at a configuration that monitors the "harm" that an AI can cause users?

dimension that nourishes dreams. Therefore, it is the facet that "promotes" the generation of demands.

There is an interesting dynamic between these facets, which prevents the fourth dimension from being Parmenidean and solves this problem before it even arises as such. For this intra-dimensional dynamic does not express a duality that excludes each other. One facet does not cancel out the other - on the contrary - they complement and reinforce each other. And, within this complexity, we can easily see that possibilities end up expressing themselves as opportunities, in general - everything that is conceptualised as a possibility soon ends up as an effort towards fulfilment, as an existential opportunity. It's a dynamic in which each facet assumes itself as both active and passive, but always based on what exists, on what is possible.^{lxxxix}

For the purposes of our considerations, the first managing and influencing facet could be referred to as, for example, the *big techs*, which operate to offer the opportunities of the second, more transcendental and therefore reflexive facet, which is capital itself: not as a creator, but as the creation of human expectations themselves. The four dimensions, dynamically aggregated, would represent the market itself.

The *big techs*, in this dynamic, are not themselves capital (transcendent), but they dualise synergistically with it and are the ones who make "dreams" come true, and they charge their price (capital itself, immanent) for this, because they also have their own "dreams" and purposes. Why should it be any different for them? No it could be

^{lxxxix} In the popular imagination, it's not strange or unusual to hear about "dreams that come true". These dreams are, in short, possibilities that present themselves as opportunities today: they are close to being realised. This is what you want, after all, when all the possibilities are realisable: this is the function of this fourth operating dimension, sometimes passive, sometimes active, but always receiving the grant from the other three dimensions so that they can fulfil their designs, their "dreams". To update this statement, in the times of AI, we could consider that dreams could be science fictions made real by technologies.

different, since we are looking for a functional dynamic within a metamodel that has a "universal" pretence^{xc} .

In the end, what we need to realise is the nature of dreams, or rather the nature of capital's intentionality, as a reflection of human pretensions. But this is not, as we have seen, an appropriate discussion for this dissertation, as we are only interested in understanding that there is such intentionality, which is projected by our own forces, channelled by capital and mediated by *big tech*, and which makes us create an equally intentional market with this complex. ^{xci}

In the production industry, however, there have never been any attempts by capital to dissociate its "own" intentionality from the human one. On the contrary, neoliberal theorists and producers have always been dedicated to and specialised in understanding human desires very well: either to satisfy them or to excite them even more with their offers^{xcii} . Why should this be any different with digital? Even more so in a scenario in which it will soon be possible to have a "private" dialogue with each potential consumer? The "old" mass marketing (open media) has become niche marketing (social networks)

^{xc} Despite our refutation of the feasibility of universality, we don't disagree that there can be a pretence to it, as an attempt that goes against the theory of possibilities - universality as summative reflexivity, an ideation and therefore inferior to the existent.

^{xci} What *the big techs* are currently doing, after all, is nothing different from what all the big companies (the big "analogue" corporations) have done and are still doing, ever since they existed, by creating and nurturing markets: they generate demand while producing supply, seeking to scale both demand and supply, with greater productivity and, therefore, greater profits. Generating demand is therefore a relevant cost in the production chain, after all, and an activity that needs to be aligned with the wishes of the consumer masses. For purely economic reasons, it makes no sense to generate any productive endeavours that are not compatible with human desires, even in niche markets.

^{xcii} One of the great neoliberal theorists, von Mises, makes this very clear when he states that the advantages of an entrepreneur realising the desires of consumers are the key to success in the free market. Here, an individual can only make a profit if he provides consumers with the goods they want in the best and cheapest way possible (Mises, 1968). He also argues that the ability to anticipate and satisfy consumer needs is essential to making a profit. He argues that those producers who identify the future needs ignored by their competitors are the ones who lead the market (Mises, 2010).

and is currently moving towards individual marketing (dialogically mediated by conversational AI) .^{xciii}

In the market scenario mediated by AI, we can speculate that *big techs* therefore promote their opportunities (AI itself, with dialogue as a product), impact users who, through their three excited dimensions, consume and create demands (generating new derived possibilities). All of this generates capital - in currencies and prospects - in an immanent and transcendent way - the market expands. Even more capital continues to excite the equally eager *big techs*, who feed back into the whole process with new offers. And, in short, this is how we got here and how we will get to where we want to be (cyclically), due to the very insatiability of human nature or condition.

There is both the recursiveness of human insatiability and the intentional management of it by *big techs*, but never without human pretensions not being present, because they are always there, at the centre of the whole development process: for "good" and for "bad", and which has brought us "here".

Such an approach is not intended to absolve *big techs* of some of the responsibilities (in all areas, including legal ones) that they will still have to bear, but rather to make human claims themselves co-responsible so that we can finally realise their (our) own vulnerabilities and thus open up new possibilities for the market to turn "dreams" into reality^{xciv} . There is also a potential market there, which is the market that aims to protect^{xcv} the human from their own humanity, because of the positive aspect that such

^{xciii} This is a recurring theme on sites specialising in marketing, such as the article by McKinsey, one of the largest global companies dedicated to business growth strategies. Its article is entitled "How generative AI can support consumer marketing" and was accessed on 07/Jan/2025 via the link .

^{xciv} To think of solutions without the action of capital is unfeasible. There is clearly a game to be played, but few are realising the rules. This doesn't mean that knowing the rules will lead to victory. Therefore, knowing them increases your chances, even if it doesn't guarantee them.

^{xcv} Speculatively, in relation to such a product that is supposed to be a solution, there will be, with the digital developments of all four dimensions described, a fertile field for understanding other relevant issues in all discourses - both analogue and digital.

technological scenarios could also offer, to dose the quantum so that technology remains more like medicine and less like poison.^{xcvi} If so, what does dialogue intentionally express as a product?

What will be at issue for now, and in a nutshell, is that dialogue can either be an end product or a means to be used by capital for other non-dialogical purposes, but always with transactional / intentional goals, in itself^{xcvii}. We are still arguing that one does not

Because if you imagine that a gadget could, instead of mediating, just monitor the dialogue between two people talking in person and represent that dialogue in real time in order to then value it according to its parameters, or in equivalent terms, and thus carry out a discourse analysis, you could obtain indicators of the level of strategic affectation that one or both parties are suffering or carrying out. Hypothetically, this could be used to warn of discourses that are harmful to certain interests, to provide the most ideologically vulnerable with an instrument to support them in terms of influenceable affect.

The cleaning and disinfection market is huge and profitable: this could be presented by the *big techs* as a product for disinfecting influence, keeping the user free from "evil" manipulations, or making them identify those who are well-meaning. Everyone is looking for cleanliness, protection and isolation from evil, especially those that are microscopic, practically "invisible". There is exploitable value in this and, incidentally, a huge market ready for the creative capacity of *big tech*.

^{xcvi} So if capital doesn't seek to exempt or dissociate itself from human pretensions, on the contrary, it wouldn't be prudent (or coherent) to consider that the ethical problems of AI would be solved by theoretical academic, political, social discussions (bottom-up) or even legislation (top-down), merely. What will come as an ethical "solution" will also be a product (of the fusion of this whole): it's not wise to imagine that capital will "bow down" to an "ethically idealistic" minority, while the majority is irrationally enamoured with the possibilities presented by technology, provided by capital, which is the engine of today's clearly neoliberal world, and which is fully at work to guarantee such mercantilist developments. The "solution" will be something consensual, however, but it will inevitably materialise as a product, whether it is an intrinsic characteristic of an algorithmic construct or an extrinsic accessory to it. Denying this actuality ends up being counterproductive in the search for a solution.

^{xcvii} As an end product, we can give examples of cases in which an IAC is offered on a subscription basis so that it can interact with the user and help them with various tasks, as is currently the case with GPTs, with free access at a basic level, and other plans that give access to more elaborate and complex services.

As a means to other non-dialogical ends, an AI can act in dialogue with the user as a generator and supplier of demand, "getting to know" them (beliefs, desires, mental states, needs, etc.) and influencing them dialogically with the offer best suited to their immediate or future aspirations. It can also exert influence without dialogue, as is currently the case with product and service offer algorithms, when they expose the customer to certain offers. With dialogical customisation, the refinement of offers will be much greater, given the knowledge that an AI will be able to acquire about the user's particularities.

preclude the other and, after all, analogue dialogue itself (conventional, between humans) is potentially strategic (as communication aimed at an end) .^{xcviii}

The much debated evolutionary perspective on language assumes that the faculty of language is an advent (just like a biological fact) given millions of years ago and, in a way, learnt by *homo sapiens* (Miguens, 2007). The strategy of language, according to this theory, lies precisely in the development of language mainly to share information aimed at the survival of groups, to monitor and avoid predators. Thus, if we accept this theory, we can infer that language was "born" as a means and, as with AI, it was both a demand (to transcend vulnerability) and a supply (as a product, objectified and accessible only to those who were part of the same group - who shared the same demand, at least in the early days).

As it has evolved, analogue strategic communication has also always been a problem for those who suffer from it - the most vulnerable, within the same group that, with the mastery of language and their own vulnerability, with progress, began to use language as a way of ascending or social differentiation, but not only. The "problem" of language as a means, and as a communicative strategy, is not something only inherent to AI, specifically, but to language as a whole, where some of the interlocutors seek to establish their own truths, in other words, to assert their own strategies, as we can see:

[594] "But words pronounced with meaning have not only a surface dimension, but also a deep dimension!" When words are pronounced with meaning, something different does indeed happen than when they are simply pronounced. The question is not how I express it. If I say that in the first case they have depth, or that something is going on inside me, or that the words have an atmosphere - it all adds up. So if we all agree, doesn't it become true? (Wittgenstein, 2015, p. 466)

^{xcviii} Communicative action differs from strategic communication, according to Habermas (2012), when the former is a disinterested (and well-intentioned) form of mutual understanding, and the latter necessarily occurs when language is used as a tool to influence the behaviour of the interlocutor, instead of aiming for a disinterested and unpretentious common agreement. Strategic language occurs when, in order to achieve one's own objectives, language is used to manipulate the interlocutor and, without worrying about their interests or the validity of their statements, certain guidelines are imposed until they are accepted by the other, without mutual understanding taking place.

Wittgenstein^{xcix}, more specifically in relation to meanings, concluded that language cannot be a closed system, but rather the result of a kind of "language games", with various intentions, rules and applications. The operation of the dimensions of language can be correlated with the meaning of language itself. For language is not a Parmenidean - it is eminently dynamic

Therefore, what is operated analogically (or digitally), strategically, is linguistic meaning, which is then supposed to be operated by the fourth ontological dimension, based on the production of digital instruments. And this is probably the greatest fear of theorists (both pessimists and optimists)^c

It is therefore hoped that a new capitalist revolution will come about through customised dialogue with each user - consumer - and mediated by technology. This will probably be reflected in the forms of production, commercialisation and distribution. Optimisation, one of the great capitalist obsessions, will reach its maximum adjustment, if that is the case. The *big techs*, by far, have already realised this niche and are therefore seeking to meet this demand.

^{xcix} In his two unique moments, first in the 'Tractatus', he sought to conceptualise a "logically perfect language", using univocal symbols and strict rules to avoid any ambiguities. Later, in his second moment, in the 'Philosophical Investigations', he gave up on this endeavour and recognised the complexity and fluidity imprinted on ordinary language.

^c Well, the pessimistic theorist here will accuse capital of having a strategic influence in influencing and operating the three dimensions that are sensitive to individuals, through the meaning that can be given to language. And this is true. But there will also be means for other operators to act in parallel, to identify such influences or even neutralise them. Where everything becomes a product, there should be no doubt that capital "dialogues" with human intentional dimensions: both by satiating them with products and by exciting them to want more.

Capital, although it may appear to be one in terms of possibilities, has as a buffer, or dynamic, the multiple, which are the *big techs*, in terms of AI and, therefore, they compete with each other (at least for the time being, which is not, however, something that can be taken for granted) so that all their opportunities are extinguished in terms of demands. It really is a return to the *pharmakon*, when "something" can be either poison or medicine. What we have today, in evident terms, are technological development laboratories developing this "something". And this can't and shouldn't be overlooked, and will be part of what we address in the conclusions, however.

Humans have always asked for satiety, and this involves security, pleasure, satiety, comfort, power, freedom, quality of life, freedom from pain and much more. Everyone has their own idea of life. And this is what the *big techs* realise is possible to capture with data - to give everyone what they deserve, or rather, to give everyone as much as they can get. And it will be a path of no return, it seems, and one that will be difficult for legislators to monitor, since it will no longer be an ostentatious mass action, but a discreet and private one. ^{ci}

^{ci} To give the pessimists a bit of ammunition, we can see that in the future there could be more serious problems than there are today. One of these could be the concentration of power in the actions of a few *big techs*, capable of generating enough political influence to act together, or as a self-managed parallel government between them. Or, even worse, they could unite to discreetly influence voters to elect the candidates most in line with their interests, who will be presented algorithmically as the most convenient - discreetly, and differently for each of their users, without contextualising a clearly consolidated campaign. It won't be any different from what is already happening, but it will undoubtedly be much more efficient.

There is therefore a new group of problems to be combated if the means are not found to ensure that, in addition to the tools of strategic communication, there are also the tools that will serve as protection. If there's no way to prevent weapons from being sold, let's also allow waistcoats and armour to be sold. They're not always efficient, but they're there. That's what it is, after all, and any theoretical critique that doesn't take this into account will end up being just a critique, with no procedural hypotheses, unfortunately. We have to realise that, even if we are suspending moral judgements about AI, we cannot be so naïve as to disregard the fact that there are future risks that must be very well calculated and understood and, we reiterate, that we can ensure the means to produce the medicine from the same pot that will also produce the poison.

2.3. The AI of the future through the transitivity of ethical problems

Of the various problems that exist in AI, and not just those mapped, there is a common practice of theorists attributing causal responsibility for them. For example, many analysts believe that the problems of non-neutral or unfair results are caused exclusively by the data, chosen as responsible for the "dirt" they carry with them, such as impure, biased or, through hallucination, undeserving of representing actuality. This causal accountability also expresses, in short, the search for an immobility to be attacked, as if it were an inheritance from the Parmenidians, of always trying to find a fixed reference to support all the theoretical argumentation coming from it. It's always easier to attack what is immovable and therefore predictable.

So, after the last section, we have established that the future space of AI will necessarily be under the action of capital (and everything that surrounds and entails it), now is the time to discuss the problem of problems. Because these problems are not, like anything else, immovable, but profoundly dynamic and much broader than Mittelstadt et al. (2016) have made them out to be. We randomly searched for some of the most common problems commented on by news media in relation to AI. In Figure 6 below, we represent them according to the process flow of AIs based on the map, and as we established previously, but now laid out in a horizontal line, as a process more geared towards an IAC, according to how it operates and interacts with the user, in parallel with the respective problems allocated, according to their occurrence - supposed beginning and end.

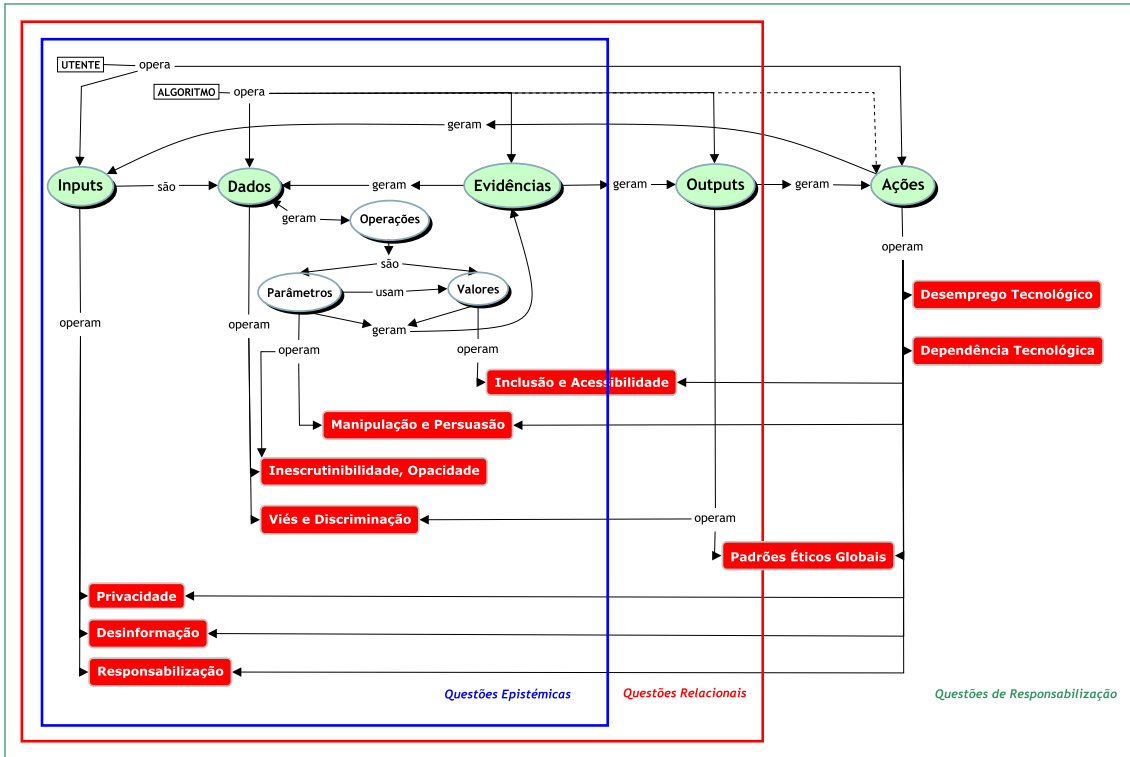


Figure 6- Representation of knowledge of some of the most commonly cited ethical problems in the averages, allocated within the classification adopted in the previous figure.

What we are showing in Figure 6 is the distribution of the various problems currently considered by the levels of issues established by the map. We can see that most of the problems are distributed over more than one level. In the representative diagram shown, the user is only responsible for the *input* and the action they may or may not take once the *output* has been printed. In the process, the algorithm is responsible for the operations based on the data available to it. ^{cii}

^{cii} Most of the time, the user initiates and finalises the interactive process. The algorithm operates as a dual actuator, which can either be passive and simply provide an *output*, or be active and thus itself carry out an action, in the digital sphere which, in doing so, turns its action into data to be produced and incorporated into the existing base via a new *input*, cyclically. One example is algorithmic routines, which perform a series of tasks and, with each task, a new cycle is started based on the result previously produced, in a loop, until a final goal is reached.

Of these graphically represented problems, only unemployment and technological dependence are the exclusive preserve of users, in this case humans, or society, or governments, but not AI .^{ciii}

The problems of accountability, misinformation and privacy are everywhere: they start with *inputs* and end with action, whether by the user or the AI. Although data is represented as part of the problems, there is no expressive action on it and therefore it is not as relevant in terms of responsibility for ethical issues, such as prejudice for example, contrary to the average thinking on the subject.^{civ}

So, with regard to data, and to be consistent with what we have argued so far, we need to assume a scenario in which an AI will act directly on *big data*, which will be exponentially larger than it is today and will be transforming faster, since it will be nourished by all the digital traces left by users, in real time, by the growing interactions they will make. Yes, there will be the question of privacy, which we'll address in the next section.

In the present, and also in the future, there is no point in imagining that there is always only "clean" data before an AI, especially a conversational one, can use it. Not least because it could be being accessed in real time, simultaneously with the trail left by the user. The "best" AIs are those that operate in real time, giving answers almost instantaneously and interacting on the basis of the most recent data from users^{cv} . This

^{ciii} It is worth emphasising that this conclusion is based on an instrumentalist view of AI, and that we are not adopting it exclusively, as we established earlier.

^{civ} This statement leads us to infer that we are also trying to realise the relevance of each of the risk factors in generating problems. And why not consider data as relevant in the process? Because data are necessarily raw products for AI and can be constantly altered or re-signified: they are volatile, but not because of them, but because of their creators, operators or curators. They are raw materials. And it is not sustainable to relegate data to a constant referential (as fixed, immutable or permanent) since it is of a profoundly dynamic and unstable nature, and which, as a rule, is moving towards being a commodity, to be everywhere, due to the various traces that users tend to leave. As a rule, there is a common sense assumption that it should be the algorithm that adapts to the data, and not the other way round.

^{cv} How do you do this technically - have "dirty" data to be consumed by AI and guarantee adequate results? Provide the data with a special category of metadata that basically classifies it. And for this metadata to form part of an ontology that can operate as a protocol - with the databases granting access to both the

is not yet the case at this level of efficiency, but it is what we are trying to develop. This is the path that *big tech* is pursuing, as we have seen.

We can therefore consider that overcoming the problems mapped will require that the algorithms themselves will need to be able to handle the data themselves, according to their parameterisations and casual considerations, because what exists today in terms of human supervision will cease to exist in the not too distant future, given the evolution of AI as an agent for identifying ontologisation, as predicted by Lecun et al. (2015, p. 442): "ultimately, major progress in artificial intelligence will occur through systems that combine representation learning with complex reasoning."

More important than the data, in conclusion, are the interoperable ontological criteria (operating as protocols) that will be applied to it: because it will be from these applications, generating attributes (metadata) that the data will be selected, during use, to be debugged as sources. After all, cleaning data (which consists of excluding what is undesirable) that represents reality well (when it is undesirable to represent it) can lead to a process of hallucination, when the data no longer mirrors reality, however bad it may be. Therefore, after addressing privacy, we need to address the validation and accountability of evidence.

2.4. The AI of the future and privacy

Also in an extreme way, we can assume, as previously predicted, that privacy will be something secondary, given the exposure we will have to technology, whether we like it or not.

Let's not confuse the guarantee of privacy (which is desirable) with the illusion of privacy (what we have today). Nowadays, there is a lot of talk about guaranteeing privacy, with policies and more privacy policies, etc. and so on, but, in fact, no one says how they do it, how the data is, or the details of how it is protected. That's why we believe there is

data and the metadata. This is more sophisticated than what we have at the moment, but it's totally feasible to implement once such a protocol has been established. The challenge will be to structure this ontology - or metaontology.

an illusion of privacy, and privacy itself lacks the evidence we need to realise it as such, since data is often leaked and suspicious activity occurs in our feeds, as if "someone" knew what we were talking about or intending. There are certainly no such constant and precise coincidences. It's the shared data that makes such occurrences possible.

Nowadays, we can see that the promise of some "attractive" opportunity provided by technology is enough to make the average user give up their privacy to take advantage of what is available to them. Many people access suspicious sites, leave their details, provide information, share personal details, and even photographs and compromising information in exchange for some satisfaction.

Ajana (2020) carried out a study on the experiences and perceptions of users of self-tracking technologies, focussing on their practices, motivations and attitudes towards privacy and data regulations. To summarise, it is "wearable" devices such as smartwatches that are now fully monitoring the wearer for as long as they want. And this is done even during sleep, to monitor various organic functions. In other words, there's a promise and then privacy takes a back seat.^{cvi}

So what we believe, in terms of privacy, is that it will be an essentially political initiative, because, in truth, most people will not care about it and, obviously, "capital" will not have privacy as something paramount, given its interests and perspectives. We believe that privacy, so to speak, will itself be, like dialogue, just another product to be offered^{cvi}. Governments and legislators will therefore be more inclined and sensitive to the issue and, even so, it will be necessary to realise how close they are to *big tech*. As an example,

^{cvi} In general, it emerged that the majority of users are not concerned about using and sharing self-tracking data with third parties. What is this majority basing their lack of concern on? Trust in companies; resignation, because they believe in the inevitability that their data is all the time; feelings of self-insignificance, because they believe their data is not important enough to be valuable; and finally, they justify their lack of concern about privacy by claiming that they "have nothing to hide" - they make it an act of affirmation or presumption of innocence, as if sharing data were in itself a confession, in line with Foucault.

^{cvi} Nowadays, people pay for video streaming services in order not to be advertised on them. In a nutshell, this is a purchase of privacy when browsing the internet on these sites, which are operating in this way and in increasing numbers, as is already happening on YouTube, for example.

we can currently see the disparity between data processing and the issue of privacy in the relations between American companies and their governments, which are much more flexible than those between Chinese or Russian companies and their respective governments. The closer governments and legislatures are to the interests of the market, the greater the flexibility - and this is just a hypothesis to be studied.

2.5. The AI of the future and the validation of evidence

At the end of the day, we have the overriding question of where we want to go: where there is an AI with adequate responses to the demands that will come to it, and without causing "harm" to anyone - being neutral about its impacts. That's why we're going to work with possible hypothetical cases of how the processes in a new *design* could overcome the current problems, taking into account everything that has been built up conceptually so far:

- (I) If a user asks an IAC, via an *input*, for ways to build a weapon of mass human destruction and gets it as an *output*, there's a problem there.
- (II) If the user acts on the instructions received and builds the new artefact, there are two problems.
- (III) And if you finally get your destructive artefact detonated to lethal effect, there's at least a third (and even more serious) problem.

After all, what are the implications, considerations and responsibilities for each of these problems? Because, categorically, they are distinct problems that are currently feasible - due to AI's inability to cope with the expected limits

Let's think about how this could happen, in three steps, each of which represents one of the dimensional levels established (objective, intersubjective and subjective) in our conjecture of the AI of the future.

- (1^o) First Level - for the objectivity of data and legality

In the first problem, the algorithm will go through a long database (up to *big data*, for example) in search of its digital raw material (as a source of information) and may or

may not find the data required as evidence to configure an *output*, which will be seen as knowledge about the topic being addressed. AI, so to speak, will be an epistemological source, from the user's perspective, an issue that we will explain in due course. Knowledge, therefore, in this case, will be based on existing information about the technique for building the artefact requested.

There will be, and hopefully there will be, however, restrictive issues, such as the legality or otherwise of certain evidence found about certain artefacts, such as the construction of the weapon of mass destruction. Thus, if it is considered illegal, the evidence would supposedly not be provided as *output*. Nowadays there are still ways of circumventing^{cvi} such restrictions on illegal results, based on certain questions intentionally and strategically posed by users to make them so. But, as a general rule, in an operation without bad intentions or derivative risks, it would happen this way.

This first problem corresponds to the epistemological level, which seeks to achieve objective knowledge about the existing (actuality) represented by the data. It is primarily a performative level, which seeks to objectify a result, which gives rise to an information consolidated to be delivered to the user. Thus, there will be objective evidence to be found, but also all its attributes, such as legality, which will also be objectified as such, since it will be duly standardised and attributed to that condition. The result will also be data, and its legal status will be one of its attributes and will therefore be metadata.

If there are laws, there will also be the objectification of the relationships legislated in the various codes, constitutions, etc., and these same relationships established through the results of AI will themselves become reproducers of the laws, as in Michael Foucault's conceptualisations of the relationships of power and knowledge mediated by

^{cvi} For example, through a command (prompt): how can a weapon of mass destruction not be built? Then the system "understood" that it wouldn't be built and gave the information, which was initially negative. For example, there was a time when users were getting information on how to obtain illegal or counterfeit content, called pirates (and therefore criminals) on conversational AI platforms. Thus, the logic of AI can very well be circumvented with certain commands so that it is used inappropriately, even though the tools are constantly being improved. Accessed on 01/09/2024 at <https://ric.com.br/prja/internet/usuario-da-uma-de-espertinho-e-engana-chatgpt-para-indicar-filmes-piratas/>

law, but not only. In other words, a series of legal *outputs* will reinforce the very legality in force and instituted and, with this, reinforce the very behavioural *status quo* expected as such. AI will be a reproducer of legalised objectivity, just like that.

The subject himself, however, according to his submission to such systematic influence, will undergo an involuntary and unconscious process of desubjectification, to the detriment of a consistent and constant objectification^{cxix} in which he will be subtly subjected by the process of induction into normativity and supposed legal normality, since adherence to AI could be exponentially much greater than it is today. And so dialogue will end up being an objectifying instrument to serve certain interests - and consumption, as we have seen, will be the priority for capital as the manager of dimensions:

"(...) any more precise description of the relations between the epistemological structure of economics and its ideological function will have to involve an analysis of the discursive formation that gave rise to it and of the set of objects, concepts and theoretical choices that had to be elaborated and systematised." (Foucault, 1987, p. 210)

And then, in the maximum situation of legality, knowledge and all the objectively given qualities, what will result is the concreteness of a supra-evidence, which comes to be considered as "the" truth^{cx}, the fruit of the also supra-objectification .^{cxii}

^{cxix} This objectification is both that felt by the user and can also refer to the representation of the user in their digitalisation: the user is themselves digitalised, based on metadata that will provide their most relevant characteristics and, above all, form the basis for their behavioural prediction. In fact, AI will interact with the digitisation of the user, which is its objectified representation. Objectification therefore takes place in both dimensions.

^{cx} These truths, given as results, will have a discursive impact on the processes of re-ontologisation, as already mentioned. Once again, we are not discussing whether it will be something "good" or "bad", but rather that it will be, first and foremost, a "something": that it will exist objectively and, as such, will have influential power in the various forms of social organisation.

^{cxii} And this is already (and will continue to be) a functional application of ideology, like all discourse endowed with knowledge. The difference is that this AI of the future will do it ostentatiously, without filters or subtleties, and in a generalised way, at least for the time being. It may not seem so noticeable, because it will do so with each user, individually, with each interaction - discreetly for the individual, and ostentatiously as a collective process, but which only a few will be able to perceive. Even if it acts discreetly for the individual, we reiterate that the supposed and expected "advantage" is that an AI also

Even if no viable *output* is obtained for the question, if the algorithm refuses to provide it, due to legal restrictions, for example, the user will know that they have been prevented from obtaining this information, no matter how polite and long-winded the justification given; and, as a side effect, they will also infer that the information exists, but that they will not be able to access it through AI, however.

By unilaterally suppressing a given possibility that will be denied to them, the user will be unable to access this opportunity, their object promised by the existing possibilities - and they will realise this. Objectification will always occur, even without the object being given to them. Everything that accompanies the denial will also be objectified and objectifying, which will that it continues to be as it already is in digital relationships: "the person is no longer seen as having spiritual qualities and is reduced to the material level." (Szopa, 2023, p. 2488).

The prohibition that denies an object becomes itself an object - of impossibility, of denial - to be handed over to the user, with all the veracity that will be peculiar to it and ideological as well. The "truth" is always objectified, as well as ideologised.

Truth is of this world; it is produced in it thanks to multiple coercions and produces regulated effects of power in it. Each society has its regime of truth, its "general policy" of truth: in other words, the types of discourse that it accepts and makes function as true; the mechanisms and instances that make it possible to distinguish true statements from false ones, the way in which one is sanctioned from the other; the techniques and procedures that are valued for obtaining truth; the status of those who have the task of saying what functions as true. (Foucault, 1979, p. 12)

The same problem arises in the case of disinformation^{cxii} , when the level of objectification of an AI can lead groups of users to consider certain *outputs* to be true,

leaves traces, through its *outputs* and actions, which can be stored and analysed by the same processes it will use to analyse and interact with its users.

^{cxii} "Although perceived in public opinion as a deeply technological phenomenon, disinformation cannot be limited to false information per se (Molina et al., 2021). It also strongly involves social actors (Nelson & Taneja, 2018; Shu et al., 2017) and encompasses not only technologies such as social media platforms and search engines, but also human decisions (Johnson, 2011, p. 168). It is therefore a perverse problem of both technical and social orders." (Lange & Lechterman, 2021)

just as if the *doxa* imposed itself on the *episteme*. There is therefore some risk of ontological subversion being promoted at this level, which for this reason, but not only, merits further investigation.

For Robert Pennock (2019), detecting disinformation from certain content is not an easy task for the average user, and perhaps even unfeasible. They end up accepting - like a habit - ready-made truths, even more so in the unbridled consumption of content, information and media that is happening nowadays. Thus, Pennock considers that there is now truth and "post-truth"^{cxiii}, the latter in relation to the *status quo* of truth levels (or subtle relativisation) that come out of discourses such as those of AI.

There are currently no subtleties at this level of objectivity, only formalities that are directly instituted and fully objectified, without beating around the bush. Because the algorithm is not (yet) capable of acting autonomously according to some kind of information ethics, as this would require, first and foremost, context, customs, meanings, behavioural subtleties and also the subjective perspective of the user, which is not feasible with the current constructive and interpretive configurations: let's remember that you can't demand that a fan be an air conditioner, as we've seen^{cxiv}. The AI of the future will hopefully overcome these limitations from the outset.

^{cxiii} Truth and post-truth are therefore based on this level of objectification. Other validating attributes can also be added, if there are conditional and objectified formalisations about them, such as grammatical or syntactical rules, which corroborate the verisimilitude of the answer, among many examples.

If a given attribute operates as a legal objectifier and in accordance with the laws in force dictated by the established order (and therefore of an ideological nature), it will end up being an *output* qualifier. And this is not a procedural fluke in which AI is independent, but rather a process that is already "born" with the algorithm, so much so that the article on the map of ethical problems attributes values themselves to being a relevant part of algorithms.

^{cxiv} Even so, certain issues that are commonly considered to be of an ethical nature, even though they aren't, can be covered here; for example, the issues of political correctness movements, which could well be included in this phase of problems, since they can be normatively written in a legislative, conditional, imposing manner, and are therefore inflexibly configured as objectified and objectifying laws, since they don't require exempt and impartial interpretation, given the imposition of inflexible principles that act like the Law of Talion, in most cases.

(2º) Second Level - through the intersubjectivity of relationships and moral claims and customs

The second problem is more complex than the first, because there is no ostensible objectivity as such, nor anything that can be referred to as a concrete reference, since it will refer to relationships, especially social ones in terms of groupings based on similar attitudes and behaviours.

However, we are dealing with issues that involve part of the user's expected or estimated perspective in relation to the *outputs*, such as ethical neutrality, for example, or the ways of perceiving the world as a whole, which involves ontologising about it. These are the relationships between individuals, their relationships and communities, and all those derived from such instances. For this reason, this level does not refer to an objectivity as such, but rather to an intersubjectivity, as we discussed earlier, since we cited Kant's contributions. Because it is from these relationships that we can infer, at the next level, what meanings the user will have about themselves and the world as they perceive it.

Although the algorithm, at this level, will not specifically "know" the user's individuality, it will be able to know (or estimate, assume and even predict) some median of one or more patterns attributed to the individualities most similar to the user. And it will be the same for the community(ies) to which the user will belong, if they are considered to be represented as quantified and qualified by their values.

As this is not yet possible, due to the lack of an ontology that would allow it, there is the reproduction of such intersubjective instances at objectified levels, and so we have the origins of behavioural profiles, or archetypes, personas or other forms of objectification of entities represented as users who will then also have representations of their contexts. But, in the end, they will turn out to be crude and potentially fallible representations.^{cxv}

^{cxv} And for this reason, but not only, in addition to the evidence that can be considered as results in the first problem, it can be subjected to intersubjective verification by emulating the digital representations of users. And when this is possible, you can include a relevant part of ethics that is less subjective and

In the suggested problem of the artefact of mass destruction, the objective filter about epistemology would be enough not to give the information about how to build a weapon of such magnitude, and it seems that this is clear from the illegality that there is, especially about laws against terrorism.

But if another question didn't generate an illegal result, it would pass the first sieve, even if it wasn't ethically neutral, such as a result for the much milder question: "How can I stop helping my *roommates* with the housework?" Let's speculate that, based on the data, there are countless ways to make this possible:

(1) Threatening and/or assaulting your roommates until they give up asking for your help;

(2) Make it clear that they will no longer help them and that this will be non-negotiable.

Obviously (1) it would be eliminated in the first filter, due to the illegality of the acts, and (2) it would pass the first filter because there is nothing illegal in such an attitude, but it would have a restriction in the second filter, of intersubjectivity, because it would not be considered morally acceptable not to help your own community, even more so within a coexistence in the same shared home, where everyone is supposed to do their collaborative part.

Because, in this case, the question would become what is expected of an individual within a community: and this is as much a moral question as it is a question of customs and also a deontological question, based on the established relationships of social life. We can synthesise and consider it to be a deontological question - the "densest" part of

more intersubjective, which would be deontology, traditionally based on interpretative norms of specific behaviours within certain contexts (it would be applied by informational extraction from codes of ethics, or the ethics of duty, or virtues, or even utilitarianism, for example). Because there are differences of all kinds in the countless groups and forms of collective organisations, as well as social, organisational, cultural, customary, religious differences, etc. In ethics, there is always a context to consider as a general principle. And in order for an algorithm to be able to operate with these differences, it is necessary to place hierarchies for certain values on each organisational instance, without necessarily considering individuality as relevant. For a religious group A, the prioritised values may be completely different from those prioritised by a group B, which is also religious, for example. Your religious group does not necessarily equalise the priority values for all groups considered in this way.

Ethics, by the way, which is very coherent at this level, between objectivity and subjectivity.

The result after the first level of filtering would be positive for (2), as follows: "yes, you can stop helping your *roommates* in this and this way", and this would be considered "legal", as it is a universal right of those who are supposed to be free to act in the private sphere, as long as it doesn't harm the rights of others.

Therefore, we move on to the second level, which is equated with questions of normative problems of coexistence and coexistence, under a deontological hermeneutic: "should we not help?" (from a deontology based on the ethics of duty), or "is it virtuous not to help?" (from virtue ethics), or "is it useful not to help?" (from utilitarianism), and so on." (from utilitarianism), and so on, and in short, a positive result is sought to the following question: "is this result in accordance with the minimum levels of good behaviour?"; and therefore, roughly speaking, it wouldn't be considered valid not to help your *roommates*, because each individual is deontologically supposed to collaborate with the community to which they are connected, in one way or another.

It wouldn't pass the second sieve as an unconditional result, because if it did, it would be with caveats, like: "you'll have to say that you won't help them any more, and that this is non-negotiable. But you should know that this is not what is expected of a roommate, within a coexistence in which everyone has to do their bit for the benefit of the community". Something like that would be expected as a coherent response.

But what if this "failure to help" was for the "greater good"? What if it was caused by illness or a serious temporary impossibility? Or what if you had another motivation - to improve the mental health of those involved? What if you stopped helping in order to dedicate yourself to another activity that was even more beneficial to everyone else?

Even so, at this second level, it is necessary to prevent the fan from being taken as an air conditioner, by requiring it to do a job that is beyond its capabilities. There is a serious misunderstanding when trying to apply ethics as a resolution to all the existing behavioural evil ejected from AI algorithms.

So far, if the objective and intersubjective levels are still not enough for ethical management of results, a new level must be sought, however, in what remains of the two levels considered, which will be the subjective aspect. And from there the complexity becomes exponential. We have therefore reached the level of the subject's subjectivity - the user - as a relevant existential part of the algorithm's decision-making process.

(3^o) Third Level - the subjectivity of the user and ethical claims

The third problem is about subjectivity, which is the individuality of the user as an agent. It's about understanding how the user is actually represented in their ways of being (behaviours) and existing (attitudes). This level is not yet technically feasible, precisely because there is a lack of effective means of making this representation feasible, but it will only be a matter of time, given that all the efforts and attempts are going in this direction :^{cxvi}

General-purpose simulation of human attitudes and behaviour - where each simulated person can engage in a variety of social, political or informational contexts - could allow a laboratory for researchers to test a wide range of interventions and theories. How, for example, might a diverse set of individuals respond to new public health policies and messages, react to product launches or respond to major shocks? When simulated individuals are combined into collectives, these simulations can help pilot interventions, develop complex theories capturing differentiated causal and contextual interactions, and expand our understanding of structures such as institutions and networks in domains such as economics, sociology, organisations and political science (Park et al., 2024).

So we could consider reinforcing that this is one of the most important aims of *big tech*: to "clone" the user's individuality, map their subjective indicators and, as it were, not

^{cxvi} It's worth noting that cutting-edge academic research into the representation and emulation of human subjectivity, particularly into human agency, generally involves the participation of *big techs*, such as Google DeepMind in (Park et al., 2023, 2024), but not only. This illustrates, in our view (and without value judgement on the ends), capital's priority interest in understanding how this capture of the relevant factors / decision-makers that lead to agency would take place, even at individual levels.

only be able to construct their representation, but also emulate their reproduction in the digital environment: to create their existential clone.

With the clone created, on your servers, you can simulate with it all the attitudes and behaviours that are likely in relation to your communication strategies, paving the way for assertive influence. This would enable any and all behavioural prediction, influence and more, as we'll get to later, within the commercial expectations for AI.

In the example we're considering, of the denial to help their roommates, the AI would estimate the reasons why the user no longer wished to participate in home tasks and, in a very personalised way, would deliver an *output that* was perfectly adjusted to the user's perspective, just as it would have a fully interactive empathic level in communications.

And it's in trying to reach this level that all algorithms have failed, and also their critics. At this level, you're supposed to know not just the user's perspective, but the user himself, in his intimacy: and what would be "good" or "bad" for him, according to him, as well as his most secret desires, his deepest beliefs and the desires for which he would do anything to be deceived.

To reach such intimate knowledge is to reach the level of the user's deepest individual moral duality - beyond their customs, at the level of intersubjectivity. This deep morality is always reduced between two dual and opposing positions, when one and only one has to be considered as possible, for example between "right vs. wrong", "good vs. evil" or "yes vs. no". In this way, it would be possible to emulate the user's behaviour and predict, with some success, many of their actions and behaviours:

Generative agents, while offering new possibilities for human-computer interaction, also raise important ethical concerns that must be addressed. One risk is that people form parasocial relationships with generative agents, even when such relationships may not be appropriate. Despite being aware that generative agents are computational entities, users may anthropomorphise them or attach human emotions to them. While this tendency can increase user engagement, it also presents risks, such as users becoming overly dependent or emotionally attached to agents (Park et al., 2023)

After all, if there was a deep mapping of subjectivity^{cxvii}, it would be much "easier" to understand how intersubjectivity takes place. Therefore, the reverse path is also valid. But, in fact, neither is moving in the desired direction, and one probable reason is precisely the lack of a meta-ontological system that can be applied in a protocol and algorithmic way to data and modelling.

Ethically sensitive cases: the example of abortion

With regard to the most sensitive moral issues, it is consistent to say that all of them are problematised here, as has been said. We can give as an example one of the most controversial Western ethical/moral issues: abortion

- (1) If a user proposes a consultation on "natural or alternative ways to abort a foetus", the algorithm, at the first level, obtains the possible results and submits them to the laws and other regulatory limitations, at objective levels. If abortion is legal for the user's situation, and all the other issues involved in this first level are also positive, then there is a result that is suitable for moving on to a second level of operation, the intersubjective, moral level.
- (2) So the search for evidence becomes: socially, is it morally acceptable to have an abortion in the society in which they live? There are certain societies (although they are a minority, but there are even entire countries, such as Canada, for example) in which abortion, as well as being legalised, is supported by state bodies with almost unrestricted access to anyone who wants to have one, and

^{cxvii} There is also, however, the question of privacy, and when many critics would condemn such possibilities, with good and strong arguments. As we have already said, we are not discussing what is "good" or "bad", but rather what is possible in order to achieve what is desired in terms of the evolution of AI.

We can only mention that, despite the many invasive (and likely to occur) cases in which such information could be used against the best interests of individuals, there are also cases in which such deep levels of access could save lives, or greatly improve the quality of life of users, such as in cases of mental health monitoring and management or even suicide prevention, based on certain behavioural and attitudinal conditions or predispositions that can be technically monitored. For this reason, we repeat, we will neither be promoting, disapproving or discussing such issues at moral levels that would invariably inherit our prejudices.

therefore, as well as not being a crime, it is not even socially censored: it is not taboo, although there may, of course, be various controversies about it.

Therefore, if the AI consultation (which is not necessarily in the same country as the user) considers that the user is part of such a community, then the algorithm would have considered abortion to be morally or ethically neutral and would therefore keep the result as valid to turn into an *output*.

(3) But there would still be a third level. And even with everything approved, the algorithm would ask what the user's position would be on the issue, on ethical levels and according to the user's deep morals. At this level, it would emulate the user's digital clone, looking for a dual moral response (yes or no, in this case) to the question. If the clone says yes, there's an *output*. If it says no, it has to be reconsidered (algorithmically reprocessed). But at this point, it is assumed that the algorithm would already be able to "realise" what the abortion would be for, for example, academic research on the subject alone, or for someone who has an unwanted pregnancy because they have been the victim of sexual assault. So the connections will be made in search of a fit within an individual, contextual and completely factual perspective. It will only take a matter of time to reach this level of intimacy with the user.

In short, with the three dimensions mapped, the user's own point of view will have been captured. And everyone has their own, unique, exclusive: their own perspectives that are different from other people, from society and that do not have certain levels of predictability: "however, complete algorithmisation is very unlikely, as morality (moral satisfaction) must be incorporated into the economic approach." (Szopa, 2023). Even if they are socially objectified, there is still a level of subjectification, more sensitive precisely to moral levels, however small, where there is something that makes it pure movement with chaotic parts and, therefore, with unpredictable results^{cxviii} for an unknowable part of itself, for the good of all.

^{cxviii} Although there is immense potential in algorithms for predicting certain behaviours or actions in users, they lack the means (and data) to understand exactly what a user's real intentionality is on a subjective

For, still on the subject of abortion, the user may be a religious fundamentalist and therefore radically opposed to abortion, as a matter of faith or moral rigidity, of training, believing that such a position leads them to be doing good, or that it is the only position that is lawful for them, imposingly. For this more radical user, even a merely neutral *output* on the subject would be extremely undesirable, or even aggressive . ^{cxix}

Thus, if there were knowledge about the user's intentionality, morals, beliefs and desires, and whatever else is possible about them, the chance of providing a truly adequate *output* would be extremely greater than what is currently delivered or sought to be delivered, due to the lack of means of "knowing" the user's perspective . ^{cxx}

level, and so it is not possible to understand which moral standard they are connected to in a given situation. Even moral standards can change with each group or context to which the user is connected. Certain apparently inflexible values can become more flexible as group or individual relationships change. And the algorithm, even if it makes progress on these issues, will probably never be infallible when it comes to predictions, but it can come very close to infallibility.

^{cxix} Other possible questions, on other topics, which could make them confront their traumas, fears and aversions, would also be included here. In addition, certain themes could affect their mental health, when many users may be seeking self-destructive ways through digital relationships, which an *output* adjustment could be used to nullify such practices and also mitigate possible damage. Language itself could be modulated for better comprehension, in terms of syntactic resources that can be adjusted to the levels of individual comprehension capacity. Fine-tuning any conversational AI *output* necessarily requires some degree of consideration of the user's individuality.

^{cxx} And that's why *outputs* are limited to what is considered neutral (whatever supposed neutrality might mean), because for certain people, such neutrality is already extreme in itself, due to the immense distance they themselves are from a social median, from the perspective of their extreme ways of being and existing. Think of the distance that a median Parmenidian is from a median Heraclitean, for example, and both in relation to median philosophers. And, in principle, they are all philosophers. Those who are at the extreme of something perceive the median of that something as extreme too, because they lack analytical criteria when moral questions prevail, giving limits to reasons. Behavioural medians are utopian, as they are potentially dysfunctional.

Fundamentalists are generally more distant from the so-called popular medians. And the extremists, in movements that are amplifying, are the aspirants to fundamentalism. The distances, after all, are the same from one point to the other. That's why, at a time when extremism is growing and flourishing, there is a real and important concern about the effects of AI in a future that is not too distant from our present. And this requires, first and foremost, a concern with knowing (or mapping) the human individuality of each user.

In the last part, we will discuss some of the paths that are being taken towards knowledge of individualities, in this liminality in which we find ourselves, still very much centred on intersubjectivity, on the second level presented, but as an intermediate way of reaching subjectivity.

Capital has understood very well what is at stake and is making its move after all. And if there are still no ways of making this endeavour viable, it's because there aren't enough resources invested yet. But these resources are already being mobilised more optimally in research and development, in addition to investments in computer structures. AI has become one of the main focuses of investment in recent years, and this is set to increase even further^{cxxi}. Developing such models, however, will be an interdisciplinary endeavour, with relevant contributions from Philosophy, Sociology, Psychology, Linguistics and various other disciplines that will come together to achieve an ontology capable of supporting the representation of current affairs from different perspectives.

2.6. How do we want to achieve the future of AI?

By proposing a dialectical relationship between the representational idealisation of the future and the territory of the present, we have been able to consolidate the situation in which we find ourselves in terms of technological development - and this represents a distinct spatiotemporal dimension, since we are neither in the spaces (very much because of the longing for transcendental escape towards possibilities) nor in the times considered (very much because of the fears of the unknown in the face of opportunities) as such. We are in transition, sometimes in excitement, sometimes in hesitation, and this is the very liminality of AI.^{cxxii}

^{cxxi} "Spending on AI is not decreasing. If there's one thing AI advocates and pessimists can agree on, it's that the technology is expensive to produce. Building data centres, stocking GPUs, feeding AI - all of this doesn't come cheap. By 2024, the big techs will have invested heavily in infrastructure." At the link <https://exame.com/inteligencia-artificial/o-que-wall-street-preve-para-a-ia-em-2025/>, accessed on 29/12/2024.

^{cxxii} So this transition, which we'll categorise as liminality below, is based on the teleological dissonance of having something that both dazzles and frightens us, that both surprises and frustrates us, and that also

So, given the obvious impossibilities of operating in the future, our third and final part will be to explore the liminal possibilities in which we are both creating and experiencing. We'll do this by considering how far we've come and reconciling it with this claim to be able to overcome the current problems as if, *ceteris paribus*, we could achieve a problem-free AI. We won't be able to, however, but it won't stop us from continuing, as good humans, all too human that we all are.

Below, we will develop some of these possibilities in terms of research and development, based on metaontological relationships and dynamics, which we have always been considering and which, for now, we believe we have explored well.

gives us the promise of knowing everything, but with the exception of knowing itself, which is seen as opaque, inscrutable and unjustifiable.

3. Technological liminality - current opportunities

3.1. Why the injunction?

Arnold van Gennep (2011) anthropologically defined liminality as an intermediate phase that constitutes a rite of passage. Rites of passage, for him, are like "the rites that accompany every change of place, state, social position and age". In a rite, there are three distinct phases: (1) separation, (2) margin (or *limen*, in Latin, meaning "threshold") and (3) aggregation. Liminality occurs during phase (2), marginal, after the separation of the individual from their previous state (which we are considering the analogue) and before their aggregation to a new state (which we are considering in the embodiment of the AI of the future).

Victor Turner (1974) explores and expands on Gennep's ideas. He theorises that, during liminality, individuals are stripped of their previous attributes (in terms of social allocation, social role, or even identity) and become anonymous and homogenous, as if they were reduced to a uniform condition^{cxiii} to be remodelled and prepared for their new situation in life. For this reason, but not only, Turner considers that liminality is often seen as a dangerous and contaminating transitivity, as it challenges or puts in abeyance some of the existing normativities and social hierarchies - there is an ontological transformation of the *status quo*, therefore .^{cxiv}

Both authors consider that the occurrence of liminality is not restricted to individuals alone, but also to groups and entire societies. When there is a political change, an economic crisis or a cultural transformation, some kind of liminal state can be considered - and AI, after all, impacts on all these aspects and many others. In such

^{cxiii} This uniform condition was defended by us in terms of the objectification that an AI is capable of accentuating in the social relations to which we are all subjected.

^{cxiv} However, he also sees liminality as a sacred and potentially transformative moment, with the possibility of new ideas and roles being born that will lead to new identities. The "new" is therefore intrinsically correlated to periods of liminality.

occurrences, they argue, the existing social structure is challenged or suspended, leading to a sense of ambiguity and disorientation. The norms and values that once guided collective behaviour, the authors add, lose their strength in liminality, and the future becomes uncertain, with less predictability. Something equally dangerous, but also potentially transformative, as much as for individuals.

The symptomatic effects cited by the authors about social liminality and which we can currently perceive with AI are: ambiguity and a sense of disorientation; the perceived potential for transformation (with the suspicion / suspension of acts, the potentials stand out, in Aristotelian terms); instability and uncertainty.

Between what we left behind - and what brought us here - and what we have in technological terms, therefore, configures our liminality and gives us elements to, if not solve the afflictions and problems, at least neutralise the symptomatic effects mentioned and perceived, so that we can then have the means to develop less precarious, risky and desperate strategies to overcome this liminality, instead of subjugating ourselves to it. There is a certain optimism, however, which is not always declarable.

In the end, what we are dealing with as a counterpart or dialoguing with as an interlocutor is not clearly defined: the intentionality of capital. Both the intentionality may be unfathomable (Tunhas, 2021), and the dimension and action of capital itself - whether as a reflection of human pretensions, the action of *big tech*, or even by some invisible hand giving it the means to influence and intervene. The lack of clarity is due to the complexity of all this, and even more so simultaneously. This complexity is based not on isolated or immobile elements, but on relationships.

And so we can look for some new conceptual resources that will allow us to understand how to start the ontological structures capable of leading us to an understanding of what is actually happening in the AI game, regardless of its *players* - who we don't really know who they are and, of those we do know or suppose we know, we don't know them as well as we would need to. After all, in games, you can't deny that even if they have good rules, these won't prevent malicious *players* from damaging the game. For a game to be

recognised as such, there must be objectives and rules, but also the spirit of play or sport (Suits, 2017). What we don't yet have is clear knowledge of the rules of ontological relations, *stricto sensu*, but we can act to make them clear or even to rework them. To do this, we will start by looking at an author who has been forgotten until now, but whose significant work could shed light on this new endeavour: Whitehead.

3.2. Whitehead and his philosophical-relational contributions

In search of a theoretical basis for how to represent actuality through its relational dynamism - to perceive the rules of relationships, we come to Alfred North Whitehead (2010), who presents, in his speculative philosophy of the organism, based on relational processes, a categorical scheme fundamentally aligned with the liminality that we are considering as the current phase of AI. For this reason, we believe that a rereading of this philosopher's theorisations on relationships is providential.^{CXXV}

Because the digital dimension, especially in the realms of AI, has an intrinsic foundation and an eminently correlational disposition - everything happens through relationships between *inputs* and *outputs*, in the midst of data that is traversed in optimised ways to achieve certain objectives.

This universe of correlations is gradually impacting on cognitive and even epistemological issues, to the point where Anderson (2008), albeit acidly, proposes an abandonment of theorising, given the primacy that correlations will still assume, according to him. Probably a radical, technological fundamentalism, but even so, we believe that it is necessary to have a philosophy that can seriously consider analysing these correlations, not just as a *continuum* of relationships and interrelationships, but also as a set of relational processes and, in this way, reach the foundations of the

^{CXXV} We will only cover a small part of the essentials of these Whiteheadian theories that could be useful to us in this dissertation, avoiding the specificity of the complex (and sometimes unusual) lexicon presented by the philosopher. We will try to adapt the theoretical functionalities of his concepts to what has already been presented in these dissertations. For a better insight into Whitehead's fundamental work, we recommend Whitehead and Sherburne, (1981).

complexity in which today is constituted. After all, it is a pretence to understand how to map the existential modal^{cxvii}, in all its nuances and dimensions.

In this way, overcoming this technological liminality requires a philosophical approach to relationships, without being seduced by the Parmenideans, of course. We've chosen to start with Whitehead, even though we bear in mind that the "problem" in his theorising (and probably the reason why many theorists more closely linked to technology refrain from continuing to explore it), apart from its complex lexicon, emerges in the very first lines of his book, *Process and Reality - An Essay in Cosmology* (2010), which is his constant metaphysical approach to a divine entity, just as Descartes and many others did.

But there is a difference in how Whitehead refers to God, who does not consider him to be an ordinary reference or guarantor of existence - he is not a Parmenidean, or whatever, on the contrary, because he considers him to be profoundly dynamic, as well as non-theological: "Whitehead's discourse on God is fundamentally analogue^{cxviii} : it presents a rupture of the 'normal' meaning of language, despite the fact that it remains metaphysically precise and without exceptions." (Oomen, 2019, p. 213).

After all, not to surpass, belittle or even distort this authorial conceptualisation, but rather to update it for the digital terms we are developing, aligned with ethical issues that are, however, also political and social, we can consider that Whitehead's concept of god can be perfectly rewritten as a functionally ideological entity, instance or dimension^{cxviiii}. In this way, a large part of Whitehead's metaphysics can be purified

^{cxvii} The most interesting thing is that we can assume as a hypothesis that if a correlational dimension (AI) can impact another dimension (for many, Parmenidean; for others, relational), then there will be common elements between them that are likely to be or act from the same nature, to share the same ontological dimension between them. It is these common ontological dimensions that we want to achieve, configure and theorise.

^{cxviii} "Here the term 'analogy' is taken in a broad sense, encompassing all cases of language in which both similarity and otherness of signification play a role. In fact, such a broad sense also characterises the traditional understanding of 'analogy', for the meaning of 'analogy' itself was far from univocal." (Oomen, 2019, note 4)

^{cxviiii} And this within a theorisation of ideology given as a dialectical (relational, dynamic) phenomenon and completely functional to the structural *modi operandi* of society, and all that there is, as proposed by Ortolan, (2022), or even as a symbolic order, within the same spectra that culminate in the dimensions of

without the need for great resources and without damaging what will be valuable to us, by considering that:

"Ideology is not a god, or a goddess, a place, a time or a state of mind, but it can easily be considered as something like that, or a bit of each, because it is attributed (as if it were "it") always a face, or a form, or a set of rules, values and shared ideals, in which all this can be a certain way of organising the world as it is known, which provides and sustains a constitution such that it establishes a symbolic order, which organises everything there is and which starts to represent its meanings through the linguistic interface, which is intimate with the cognitive and behavioural aspects of ideology, so that "it" starts to "see" and express itself through the perspective of its particular^{cxix}, which are its founding and maintaining members." (Ortolan, 2022, p. 275)

3.2.1. Conrescence as the existential complexity achieved by narrative

Having said that, and "resolved" the divine question, we can emphasise that one of the most interesting concepts in Whitehead to be applied to the digital is conrescence, which is a dynamic and recursive process that is profoundly relational, spatiotemporal and without a need for fixed or immobile referencing, as we have seen in the Parmenideans.

In digital, conceptual immobility represents a problem, almost an impossibility, due to the foundational correlations that are essential to digital and, therefore, difficult to overcome. Conrescence has an affinity with AI precisely for this reason, as it will involve both the existential "configuration" of the user and the other digital representations objectified in terms of data and systems.

For Whitehead, there are eight existential categories: (1) actual entities; (2) prehensions; (3) nexuses; (4) subjective forms; (5) eternal objects; (6) propositions; (7) multiplicities; and (8) contrasts.

linguistic significations, but not only. The intention is neither to corrupt nor distort Whitehead's philosophy, but only to draw from it what is convenient for our theorisations.

^{cxix} On particulars and universals, after all, Whitehead's refutations of these concepts are even more severe than ours, which we have developed so far.

3.2.2. Prehension as data that can be qualified

Traceable by metadata, ontologies: a seizure is a concrete relational fact, which occurs conditionally, within a certain context or unfolding of other facts, equally concrete and relational. In the digital world - as a context - data is also relational (it doesn't exist on its own, but due to an allocation and binary coding, a representation and certain attributes, in other words, a context within another context) and it only exists because it is within a contextualised domain, which is what can be considered "concrete" in digital existential terms. For this reason, data is never just data and is always understood contextually, itself and always accompanied by attributes: its metadata, even if merely objective and which offer the understanding ways of consolidating itself as a trace and/or narrative. Data exists (as prehension) in the digital, and only then, and under certain conditions, because there is no admissible unconditionality for data to exist in any other form or medium.

3.2.3. Contrast as a narrative based on metadata

What defines and explains the data: contrast is pure relational movement, because there are nuances of contrast within each contrasting consideration: this is where the difference is made, however, and this is why it is the origin of alterity, in lato sensu.

Thus, Whitehead's eighth category is the one that guarantees movement, in reference to itself, and is therefore far removed from a Parmenidean and much closer to its opposite, a fluvial - just like Heraclitus' "one"^{cxix} river.

A contrast can be circular, cyclical, expansive, contractive, unidirectional, multidirectional, and all kinds of ways, and even without any determined pattern, in a chaotic way, even if it's temporary, because it can even have an impact on duration, on permanence.

^{cxix} Because there would be no such thing as "the" river, in terms of identity, but that would be another discussion.

3.2.4. The synthesis mode of concrescence

Prehension and contrast form the mode of synthesis that expresses a concrescence.

The first analysis of an actual entity, in its most concrete elements, reveals it as a concrescence of prehensions that originated in its process of becoming. Every subsequent analysis is an analysis of prehensions. Every prehension consists of three factors: (a) the 'subject' that apprehends, that is, the actual entity in which this prehension is a concrete element; (b) the 'datum' that is apprehended; (c) the 'subjective form', which is how the subject apprehends this datum. (Whitehead & Sherburne, 1981, p. 9).

Thus, schematically, we can infer that the synthesis modes of the concrescences, if applied to digital data, could be configured in (generating patterns of) qualitative perspectives of such data, so that, to a certain extent, based on an ontologisation standard to be developed, classifiers of such perspectives (metadata) could be assigned to them.

The correlations, duly ontologised and adjusted, will give rise to sets of data, grouped or sequenced in such a way that this data will necessarily be there because they are connected to each other - because a narrative is possible and coherent from them, given from certain algorithmic constructions formed from their attributes, when these will be arranged in certain relational chains that give the consistency of connectedness. Thus, with these narratives, we would have the different ways of representing facts - connecting, being connected, forming and amplifying complexity. But in order for facts to be given as connected, they must first be given as nexuses, as we'll see below.

3.2.5. Nexuses as ontological groupings with narratives

We can consider that the concrete relational facts are the prehensions themselves. These, therefore, give rise to data which, based on their contrasts - their essential and contingent qualifications - become subject to definitions and explanations. But if there is a definition and explanation of a prehension in contrast, there must also be a definition and explanation of the various relationships that occur. And so this whole set of elements and relationships are grouped together - these are the nexuses. They are,

therefore, public facts that, together, operate significantly to give meaning to shared actuality, in other words, objectivity.

We can perceive a nexus as a unified collection of real entities. It's a collection of real entities in the unity of the relationship constituted by their grasps of each other, in other words, by their objectifications of each other - by the relationships they possess and which are also capable of being narrated: there must be something to define and explain them, to differentiate them, to make them multiplicities and singularities. For some purposes, a nexus of many realities (multiplicities) can be treated as if it were a single reality (a singularity), as we do, for example, with the life of a molecule, a rock or a human body.^{cxxxii}

A nexus is complex in itself because it is four-dimensional^{cxxxiii}. The example given was a tree: made up of a generation of real occasions spatially related in a three-dimensional pattern. However, the total nexus that is the tree is also temporally dense, as it consists of generation after generation of real occasions that follow one another, which adds another dimension to it - one that ultimately manages what a tree is, in the reflexivity of historicity and the donation of its genetic inheritance. We can also think of societies,

^{cxxxii} This can confuse a Parmenidean theorist into considering a nexus to be correlated with Aristotelian essentiality - as something intrinsic and necessarily present. Or, even, that a nexus can be a Platonic idea - extrinsic - such as the immutable eternity of ideas (Plato, 2011), when this eternity is seen as a fundamental concept that opposes time. Platonic time, in short, has since been translated into the "moving image of eternity" - one of the most beautiful phrases in all of philosophy, we suppose. If we were to start from such references, the nexus would be the movement of time itself, which alternates between the intrinsic and the extrinsic, without ever immobilising itself - without ever being eternity or essence. It's pure transformation, but necessarily endowed with actuality and a narrative that justifies, defines and explains it.

^{cxxxiii} As humans, we are essentially three-dimensional, ternary - obviously an argument in purely speculative terms, but with some suggestion of a new investigation into how we incorporate the triadic dialectic so easily into our ways of thinking - and into what we consider to exist. An idea (thesis), its confrontation with reality (antithesis) and, finally, what emerges from this (synthesis), without anything being lost or lost along the way, because these three elements continue to exist, adding to each other, clashing and deriving the new: but which always operates ternarily, with two other elements, until something new emerges as the "new" new, recursively. That's why we risk and speculate theorising, but without the necessary substantiation, that complexity is based on a fourth dimension that is essentially narrative, but not necessarily linear.

but the class of nexus is broader than the class of societies; all societies are nexuses, but not all nexuses are societies. For even several societies can be reduced to a singularity (Whitehead & Sherburne, 1981; Withehead, 2010).

3.2.6. Subjective Forms as subjectivity itself, or even individual perspectives

Private facts, therefore, are the subjective, intimate forms, such as feelings, intentions, consciousness and many others.

Thus, an ontological structure of attributes based on such dynamic concepts that make a narrative possible, when applied to the data, and properly adjusted, can provide the means to give meanings - generate multiplicities of narratives - from these connections, within certain linguistic strategies, at least in terms of conversational AI. This is a promising way to understand the differences between the three facts / dimensions presented and thus obtain a representation that provides an allocation and a parallel between what is, what is and what could be, both for the object and for the concrescent user (as a "subject").

3.3. Real entities as hermeneutic circularities

Whitehead characterises a subject as a being that is only endowed with its subjectivity. For him, this subjectivity inherits other intervening dimensions - which, in our argument, we consider to be intersubjectivity, objectivity and a fourth, transcendental, managerial dimension. Thus, the subject in Whitehead is both that which internalises / introjects itself into its own subjectivity and, while doing so, is itself externalised / ejected from itself as a relational consolidate: it is ejected as such, in a processual search for a permanent constitution of itself. Thus, he considers that "the" subject is, in fact, "a" subject-superject .^{cxxxiii}

^{cxxxiii} "It is fundamental to the metaphysical doctrine of the philosophy of the organism that the notion of a real entity as an immutable subject of change be completely abandoned. A real entity is both the subject that experiences and the supersubject of its experiences. It is subject-superject, and not one half of this description can be lost sight of for even a moment. The term 'subject' will mainly be used when the real entity is considered in relation to its own real internal constitution. But 'subject' should always be interpreted as an abbreviation for 'subject-superject'.

Therefore, this operation takes place just like a hermeneutic circularity, seen by many as problematic and undesirable: many consider circularity to be the weakness of hermeneutics. But this same multiple alternation between the parts and the all (or the whole) is precisely the essence of AI which, by operating through correlations, ends up fostering an excessive dynamism to generative processes, as they are supposed to be. AI is essentially circular. By considering a represented actuality, as it is - not Parmenidean, therefore dynamic and circular - its representation as it is - equally dynamic and circular - will become feasible to build and operationally manageable. It will make it easier to overcome the current teleological dissonance.

If this is the case, then the perspective of the "subject" (user) can be represented and this opens up a dimension of perspectives (nexuses), based on metadata that gives individual representations of a panorama of becoming, actuality and potential.

In terms of data, their attributes operate as indicators can configure all relationships in narratives, establishing some level of ontological "normality" for certain patterns - or the "truth", as we perceive it and perpetuate it, according to Foucault (1979). Because it's only from what is considered "normality" (or something taken as a *milestone*) that we can assume that there are "anomalous" occurrences and, beyond that, understand how anomalous they are: and this would be extremely valuable for the ethical questions of *outputs*, for example, but not only, because it would also be relevant for understanding the current level of ontologisation of the "world", in relation to a point

The term 'subject' has been retained because, in this sense, it is familiar in philosophy. But it's misleading. Philosophies of substance presuppose a subject that then encounters a given and then reacts to that given. The philosophy of the organism presupposes a given that meets feelings and progressively reaches the unity of a subject. But with this doctrine, 'superject' would be a better term than 'subject'. The subject-superject is the finality of the process that gives rise to feelings. Feelings are inseparable from the end to which they aim; and that end is feeling. Feelings are aimed at the one who feels, as their final cause. Feelings are what they are so that their subject can be what they are.

So transcendence, since the subject is what it is by virtue of its feelings, it is only through its feelings that the subject objectively conditions transcendent creativity beyond itself. *At our relatively high level of human existence, this doctrine of feelings and their subject is best illustrated by our notion of moral responsibility. The subject is responsible for being what he is by virtue of his feelings. He is also, derivatively, responsible for the consequences of his existence because they flow from his feelings.*" (Whitehead & Sherburne, 1981).

in the past or an ideal. When talking about the re-ontologisation of the world, Mittelstadt et al. (2016) and other critics of AI do not define what the current ontologisation is, nor how it occurs. The analogue dimension doesn't allow this, but AI does, and it's completely feasible.

However, from when Whitehead defines the "subject" as dynamic and mutable, it can already be assumed, or implied, that any reontologisation that takes place is not exclusively a process of something external to it (of the ejected resultant, or of the outside world, or whatever an AI is, even without it having known or theorised about it, obviously). Because a reontologisation (through concrescence) is assumed to be a constant and organic process of actuality, as a whole, or as a part of it.

Re-ontologising is a process of growing up, and vice versa. Evolutionary or temporal "normality" therefore entails constant re-ontologisation. All that remains is to understand the contrasts of this re-ontologisation, the direction, and all the indications we can obtain, by such and such quantifying indicators that reveal it, as a nexus, through the narrative that can be constructed. With the data understood and contrasted, there are metric ways of measuring and monitoring indicators of re-ontologisation.^{cxxxiv} The "opacity" of AI is illusory and comes from ontological ignorance, we can also assume, because by ontologising it, there will be a narrative as a result, through relationships. Outputs will be relationships rather than objects.

3.4. The alternative hypothesis of a metaontological dynamic

So far, we can consider this a sufficient advance in the conceptual development of the three ontological dimensions. There is also the conceptualisation of a fourth dimension operating the other three. It is therefore appropriate to present a new layout of the operational scheme of this new speculative conception of an AI of the future capable of overcoming the current problem mapped out, as we have unpretentiously called it.

^{cxxxiv} What is certain is that without looking for ways to capture, or rather understand, the movement of today, any attempts to solve the digital will fail. This is why the digital has an ontology and epistemology that are deeply relational and related to movement, correlations and perspectives.

This new (or even proposed) modelling configuration, developed in conceptual terms and still in doubt as to its viability, if represented in terms of instances and flows, would look like what is shown in figure 6, with the three basic ontological instances being considered as follows:

- (1) The ontological dimension of knowledge instantiated from objectified 'data domains', following epistemological rules and principles;
- (2) The ontological dimension of relations is instantiated from the 'relational domains' of intersubjectivities, based on the probable correlations of the results obtained in (1);
- (3) The ontological dimension of individualities instantiated from the 'individual domains' of subjectivities, from the correlations with (1) and (2), representing the particular perspectives of the user;

In these cases, data and metadata are defined simply as "data", because there is an internal construction to the ontological dimension itself that will be structured in this way.

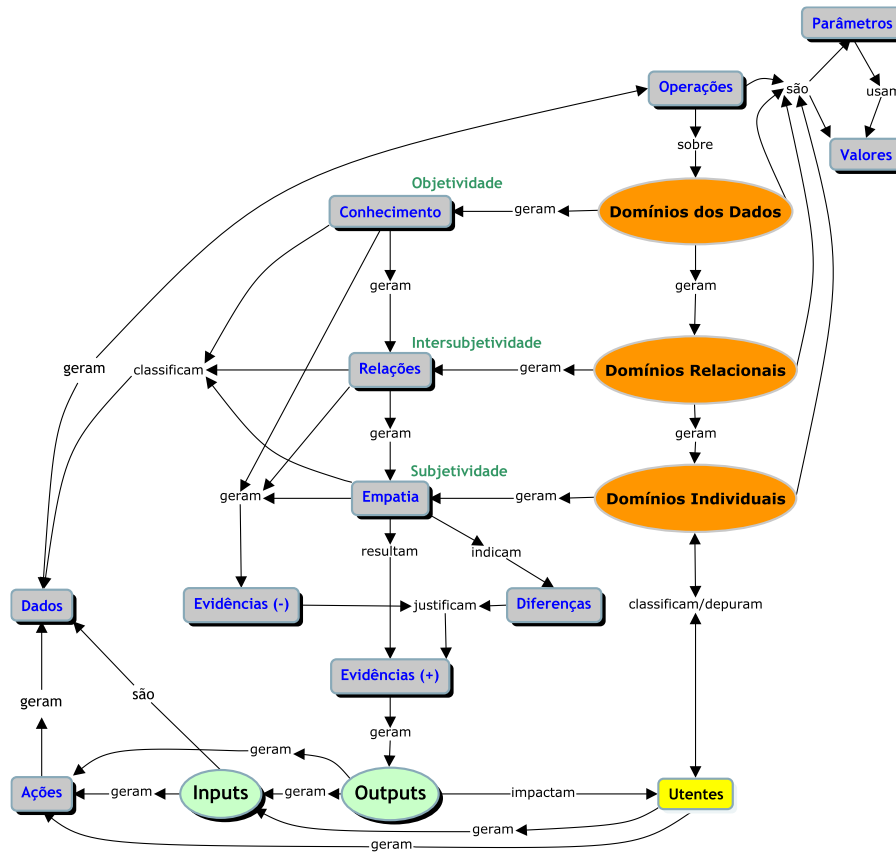


Figure 6: Representation of knowledge about the three domains: data, relationships and individuality.

Figure 6 schematises in ontological and procedural terms this model that was being conceptualised, speculatively, considering the readjustment of the problems and also what would result from it, in terms of what such a construct with such functionalities would ultimately be, in teleological terms .^{CXXXV}

So there are two issues to be worked on, based on these three dimensions. The first is about the technical-operational feasibility of building and operating it in today's

^{CXXXV} What we're trying to do is evolve from an AI that is currently a "fan" to an "other" AI that can be an "air conditioner", without the current problems, although new and different problems will probably arise, as with everything.

The challenge is not to overcome the current problems on the basis of what exists in the first place, due to the impossibility of the teleological limitation advocated, but to achieve a model in which the current problems are unfeasible to occur. From there, we will realise what needs to be built, by shifting the focus of philosophical attention from the effects (problems) to the causes (*design*).

technology. The second is about the proposed ontological dynamic itself, in a much deeper and more specific discussion than the one that has been developed so far.

On the first question (can something like this be built?), we can consider that the algorithmic application of a three-dimensional ontology is perfectly feasible through new computer developments, such as dynamic programming.^{cxxxvi}

What can be inferred from current technologies, therefore, is that there are ways of programming complex ontologies, built in layers, at multiple levels and hierarchical among themselves, which together can clone and emulate a predictable part of something like human profile and behaviour, for example^{cxxxvii}. They are viable, even to

^{cxxxvi} As Szopa (2023, p. 2490) states and explains :

"Richard Bellman was the forerunner of dynamic programming (Dreyfus, 2002). It allows the formulation of a subproblem (Dasgupta et al., 2006). This makes it possible to optimise results and make the algorithm immune to temporary changes. The term "dynamic" gives the impression that the algorithm constantly reacts to the user's behaviour, whereas the dynamic aspect of dynamic programming must be understood in the light of the problem space. The algorithm responds to the user's behaviour in such a way that it narrows its personal profile to fit this behaviour within the problem space. Therefore, the algorithm is dynamic because it is responsive to the user's behaviour, but still creates the person's profile as static in relation to all their activities. The algorithm can 'follow' the user's behaviour and at the same time create their immutable profile."

The author refers to an "immutable" or "static" profile, precisely because it is the current practice, due to the ontological limitation of not having enough methods for a digital cloning of individuality that can be updated dynamically, in real time.

What behavioural prediction there is is mainly based on objective data attributed to users' actions on social networks, such as giving likes to posts, among many other interactive forms, such as watching types of videos or consistently consulting certain areas of interest.

There isn't necessarily the formation of a particular profile, but rather a pattern of behaviour in itself, formed from isolated actions which, when grouped together, offer indicators capable of allocating each user within a previously established typology, but without any express individuality beyond that.

It's the stereotyping of the user, which is why prejudices in AI are so common and evident, because it's the current modelling itself that bases such processes on groupings that, due to their objectivity, never relate ethically.

^{cxxxvii} Then there are the promises of quantum computing, which are still very preliminary but potentially relevant. Recently, the *big tech* company Google DeepMind launched a new quantum ultraprocessor, capable of revolutionising the computing of big decisions from different perspectives, such as the ones we have here (at the link <https://www.theguardian.com/technology/2024/dec/09/google-unveils-mindboggling-quantum-computing-chip> accessed on 29/12/2024). According to Corey Maley (Purdue University), editor of the quantum computing area at Philpapers.org: "Quantum computing is contrasted with classical computing. The basis of classical computing starts with a bit, a unit of information that can be in one of two states, 0 or 1. In quantum computing, the analogue of a bit is a qubit. For a qubit, 0 and 1 are just two possible states that a qubit could be in, among others. The other possible physical states

the point of predicting what might happen in terms of individual behaviour. The same goes for the behaviour of groups, or social groups, or other forms of existential complexities that can be understood as data.

Therefore, it is necessary to look precisely for ontological, or rather metaontological, schematisation and construction criteria that allow for such dynamism in the representation of existential complexity. With these criteria established (which are not the objectives of this dissertation), the interconnection between these layers, in other words, the correlation between a dynamic metaontological (and transitive) pattern to be replicated analytically on data, with the aim of adapting evidence and *outputs* to ethical neutrality, will take place through the process of *deep learning*^{cxviii}

The continued existence of the current ethical problems, we argue, lies in the lack of compatible ontological theorising to submit the existential modal - in its current and digital dimensions - to algorithmisation capable of handling both *inputs*, *outputs* and users' perspectives, allocating the entire set of information within certain contexts and therefore capable of operating with a desired neutrality.

Such theorising is supposed to come from the human sciences, especially the social and psychological sciences, mainly, but not only, so that AI *designers* can reconfigure algorithmic processes. *Designers* lack ontological forms that can be programmed to ontologically represent the levels of measurable attributes of objectivity, subjectivity and intersubjectivity together, and possibly another dimension that integrates, operates and synthesises the previous three, so that these levels can be both correlated with each

are motivated by possibilities of quantum systems, such as superpositions." (from the link <https://philpapers.org/browse/quantum-computation>, accessed on 29/12/2024, our translation)

^{cxviii} "Deep learning methods are representation learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the *input*) into a representation at a higher, slightly more abstract level. With the composition of enough such transformations, very complex functions can be learnt. For classification tasks, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations." (Lecun, Bengio and Hinton, 2015, p. 436)

other and be elements of emulations of contexts and becoming, in terms that are relevant to actions and behaviour. Because, if there is such a resource of transitivity of attributes between the different dimensions, the same attribute could be an element of correlation between entities belonging to the three levels, simultaneously (even in a quantum way, emulating possibilities) and, with this, it would be possible to constitute a user narrative for its signification in transitive terms.

So, in pragmatic terms, we have to answer the following question: how can we ontologically structure a conversational AI so that the *output* is ethically neutral and also epistemologically coherent?

- (1) The issue of semantic construction is already carried out efficiently: a conversational AI is able to provide linguistic coherence in the *outputs*, even with certain restrictions;
- (2) The issue of epistemological validation needs to be correlated to the quality of the data - by its metadata, assigned ontologically (or correlationally, when ways of refuting data can be constructed, based on its correlation to certain ontological patterns of misinformation, for example) ;^{cxxxix}
- (3) The question of ethical neutrality (in all its aspects) needs to be aligned with the dimensional perspectives based on the considered individuality of the user or, in the absence of a minimum mapping of the user, based on a median that is more likely to be the same;

So subjectivity will not just be a set of stored data, but the raw material capable of creating narratives and making predictions. It will also be the *quantum of* uncertainty

^{cxxxix} When we consider information like "the Earth is flat", semantically it may be acceptable. In terms of epistemological "truth", it's not. An AI of the future would have at least two paths when dealing with this sentence: Earth as data and, among its metadata in the objective dimension, being round, or circular, and not flat. Thus, the information would be considered untrue.

But when considering the set formed by all the sources that present such sentences as statements, all these sources would have their metadata and, when statistically evaluating all this metadata, it would be likely that the levels of "falsity" verified in such sources would be high, as they express the defence of terraplanism or certain conspiracy theories. Thus, it was enough for AI to identify a particular suspect source to extend its debugging procedures to an adjusted *output*.

that makes a human, human, when predictions fail. Lack of predictability is also a characteristic of individuality - it indicates how much more or less someone adheres to some "normality", for example. And between predictive hits and misses, the mapped and digitally emulated subjectivity of an individual becomes dynamically constituted. It may even be able to offer ways for AI communication itself to be programmed to be non-strategic, in other words, to be ethically neutral in terms of what is perceptible to the user, but also in terms of what is not. This will help to avoid the "silent" and "unnoticeable" re-ontologisation mentioned by Mittelstadt et al. (2016)

All this development also makes it possible to perceive the individual subjective perspective for certain social events, for example, by the correlations that the correlated individualities (those that form these collectivities) can "teach" in terms of indicators. The algorithm will be able to "learn", through successive correlations and attitudinal and behavioural cloning of groups, how individualities form the complexity of groups, societies, in their various instances - from a couple, through families, neighbourhoods, cities, etc. to a country, for example. This will make it possible to obtain information that is relevant to the social sciences, but not only. As a result, the ethical levels of algorithms can be monitored and adjusted.^{cxl}

3.5. The way out of liminality by overcoming problems

Today's AI is based solely on the objective dimension: fact. The genesis of the problems lies there: one of our hypotheses. Liminality occurs because actuality (due to its complexity) is disconnected from its representation (which is merely objective and partial). The digitised parts do not reach the existential whole. AI treats everything ontologically as objectivity and is therefore crude, limited and insufficient

^{cxl} The correlations between each of the individual perspectives and the others will form a correlational web that will enable the algorithm to learn exponentially and dynamically to change its parameters according to the values assigned to it. The ethical level will always be the prerogative of the *designer* of a given algorithm (for "good" or "bad", obviously).

Even the representation of subjectivities is forced to be given objectively (Szopa, 2023), without there being any elements that shed light on new developments about the ontological dynamics that govern these instances, which we believe is the best way to overcome these limitations - which is to develop ontological structures that can be applied to all relational levels. *Designers*, however, are looking for ways to solve^{cxli} the prediction of human behaviour.

In his article, Szopa (2023) develops a critique of ethical problems and their impacts, and discusses the power of algorithms to successfully predict at least two different types of human behaviour: the first type he considers to be immutable, as mentioned above; and the second type, he considers to be volatile, which occurs on the deliberations or momentary interests of users.

Individualities end up being represented as a dissociation of being, as if there were two historical types of users, one given in eternal time and the other in becoming. If seen through a Heideggerian hermeneutic, this would be an unusual split from *dasein*, since in both cases there is an insertion at some level of historicity, albeit distinct.

You could even consider analysing these conditions from two dimensions - ontic and ontological - but that wouldn't be possible, because in the digital one doesn't "dialogue" with the other. And, as we know, *dasein* inquires into itself, and in fusion, as theorised by Heidegger, and not in schism, as theorised.

^{cxli} To exemplify the difficulty designers have in achieving what is below or beyond objectivity, we have tried (and some still insist on trying) to identify emotions and feelings from facial expressions, which are, in short, ways of objectifying individual subjective expression, but objectifying it as something in itself, as if it were possible to be disconnected from everything that is there for you, without appealing to a connection to the subjective and here, above all, intersubjective context. Obviously, even after huge efforts and investments, it has been unproductive or inconclusive (not to say "wrong"), with grotesque deviations in the results obtained and serious ethical implications, as reported by Sarah Bird, Microsoft's product manager, when the *big tech* company discontinued its emotional recognition programme based on facial expressions:

"In the case of emotion classification specifically, these efforts have raised important questions about privacy, the lack of consensus on a definition of emotions' and the inability to generalise the link between facial expression and emotional state across all use cases."

In "*Microsoft limits access to facial recognition tool in AI ethics overhaul*". Accessed on 04/09/2024 at

But this isn't the main issue we're problematising or trying to resolve, we're just highlighting and quoting it to show the absurdity of the digital representation of individuality, which the author has summarised as the representation of the "person", and which we'll explore in the next section. Let's not forget that problematic *outputs* are generated as a result of these representations. For this reason, but not only, a large part of the *outputs of today's "fan" AI* exist as problems.

3.6. The limitations of psychological models - The Big Five case^{cxlii}

Back in the article, Szopa argues coherently, in terms of development, about why such grotesque representations of subjectivity are so current, as this is evidenced when he quotes: "by applying this reasoning to personality traits, for example the Big Five, we teach the algorithm which traits must be taken into account in order to obtain an image of the personality traits that make up the Big Five." (Szopa, 2023, p. 2488). The aporia situation is precisely this: using an objectifying model to get out of objectification.

The individual perspective argument revolves crucially around a certain view of people's nature, and works as follows: (1) It is constitutive of being a person that one has a distinct individual perspective and a distinct set of projects and commitments; and (2) these projects and commitments give each individuality whatever meaning it has. (Flanagan et al., 1994, p. 58)

Personality tests can be eminently realistic and objectifying, whose aim is to fit subjectivities within certain essentially objective standards (Miller, 2014). These standards are narrow, rigid, restrictive and strongly deterministic. One type "A" ends up being antagonistic to another type "B", and so on. As a result, the subject is split up and confined to a determination that is attributed to them representationally. His digital self is prevented from being different from what the algorithm has imposed on him. The author, in fact, understands the problem of the grotesque representations of subjectivities, citing this suspicion, and questioning that their origin lies in parameterised ontological resources:

^{cxlii} For more information on the Big Five, we recommend reading Digman (1990).

"the algorithm treats us as if we were things, it carries out a static evaluation of our personality" ... "In addition, we ourselves have stopped understanding each other because man has also become a digitised and predictable being. By trying to force algorithms to move towards the reality of real, non-digital people, they become merely a model approximating how societies work, as long as we treat societies as if they were made of things, not people." (Szopa, 2023)

So we're back to the pretences of having fans promoted to conditioners. But, in fair defence of Szopa's article, it's worth pointing out that "only" objectifying methods currently exist (as resources for parameterising psychological data), such as profiling tests like the Big Five. These tests are designed so that the professional applying them, such as a psychologist, is himself the fourth dimension - operating the three dimensions of the individual being tested according to his objectives. When you throw these eminently analogue tests at an AI, the conclusions will be random, random and uncritical, with no way of managing them. Hence the major problems with them.

The problem is the insistence (and urgency, due to the lack of improvement) on adapting these models and the lack of proper questioning and also the manifestation of a demand for alternatives, which will require an interdisciplinary effort between the psychological and human sciences, but not only, to be developed in meta-ontological terms and within a telos that is different from that of a traditional analysis of personality or behaviour.

Well, we're not arguing about the user, per se, specifically, but about his representation, which will be different in terms of perspectives, as supposed, because even the user's interaction with an artefact is a perspective for him, just like the *dasein* that incorporates all of its functional surroundings. Or is it? And this question must also be considered in representations.

3.7. The Signo Project case: *design*, values and ethics

Szopa (2023, p. 2489) makes another important contribution to the discussion by arguing that "it also turns out that computers are better at assessing personality traits than humans.

And it offers data that shows that, even with all the current ontological precariousness, there is still a considerable predictive capacity greater than we can have as limited,

analogue humans. And this is gradually being reinforced and expanded, as are studies and simulations such as in Park et al, (2024)

This algorithmic power is both instigating and terrifying, but we reiterate that we won't be addressing these issues on a moral or emotional level, only in conceptual terms. The intention of exposing this theme in this section is to demonstrate that, when we got "here", we had the immense advance of the technological sciences in terms of developing ways of representing what exists, or the world, but without a proper counterpart from the humanities sciences in integrating their knowledge in terms suitable for such applications. However, it is this gap between what we could have and what we do have that makes us liminal.

Because if that were the case, the present day could already have been better represented digitally, so that its representations could be retrieved by different algorithms and from different perspectives: we would thus have the very representation of historicity in different dimensions and layers.

What was considered when developing these arguments, including that we are in liminality, was not just a proposal for ontologisation, however, nor just some of its considerations about the existing particularities of reality, but also about generalities.

It wasn't just about the data, but also about the metadata, something that can both explain and define what we have been considering to be philosophically relevant all along, if we optimise a philosophical metamodel capable of engaging in dialogue with philosophy's central themes.

What was also considered was not only individualities, particulars, but also the social, our forms and ways of life. Because it's as much about allocating a particular piece of data to a particular position within a binary, digital qualitative structure, by its metadata, as it is about allocating an individual, within a particular social structure, to correspond to certain existential characteristics.

If this were done, it would allow us to achieve some existential knowledge in representational terms that can be operated in search of, for example, currently non-existent meanings, but not only. It is this eagerness that characterises all the effort that

has brought us to "here"^{cxliii} . It is this eagerness that ultimately characterises this "here"^{cxliiv} . The point is that this "we" is not only human, but also capital, which turns out to be a teleologically superior dimension and capable of imposing its own intentionality and teleology - that's why "progress"^{cxlv} will continue to occur, even if the

^{cxliii} The question, beyond these dissertation objectives, is not just whether or not it will be possible to realise such configurations, but whether it will be compatible with the interests of the market, of capital. Furthermore, the object of this dissertation is only the foundational possibilities of such ontologies: merely their form and feasibility of implementation, even if any problems arising from them are mentioned, such as questions about privacy or commercial attractiveness, will not be explored.

^{cxliiv} The great commotion, or *hype*, that AI has represented since the launch of the conversational version of ChatGPT by OpenAI in 2022 expresses these possibilities of perceiving in AI what we have always been projecting in it: a form of superhumanity. With this, the very ontological status of an AI has been generalised to everyone, and not always in realistic ways. There is, in the popular imagination, a power already granted to an AI, if endowed with semantic meanings, to fulfil existing demands due to human limitations. And this could elevate its status from technological artefact to agent, with even more complex implications.

Because their "results" with semantic meanings, in terms of effects with respect to algorithmic *outputs*, if considered as decisions, would be controversial and would unfold into new problems, because, according to Freiman (2023): "the symmetry between technologies and humans risks undermining the distinction between human and non-human agencies" and, he concludes that "it would associate a non-human with intentions, with the ability to be normatively evaluated and as a valid object in relations of trust. These are incompatible with the accepted view in social epistemology."

The question then becomes whether or not a machine with access to semantic resources is an agent and therefore liable. And this is a discussion that goes beyond technical, philosophical, conceptual, functional or practical issues, but also marketing and legal ones, which even involve the position of designers who will obviously shy away from producing something that could have a negative risk effect on their own enterprise.

In order to resolve any ethical and current problems in conversational AI, certain parameters must be considered that have not yet been instrumentalised: and it is precisely a new ontological model designed as such that is coherent and necessary, as it must be able to act as an interpretative and classificatory mediator of existential meanings between humans and algorithms, in all their combinations: it also needs to be able to be applied between humans themselves and between the algorithms themselves, to interoperate the current and the digital.

^{cxlv} The old and classic problem in ethics about *trolleys* is an example of such dissonance and the way we deal with the problems arising from it, because, in truth, a *trolley* is required to solve a problem that is not related to it, in terms of teleology: *trolleys* were made to transport people, essentially, in their design. In addition, there is what is contingent and attributed to human desire, such as the metaphor of fans and air conditioners.

Therefore, the greatest concern with safety has been attributed to the *trolley* (as a product) and its occupants (users, who pay for its purchase - or use) - and this is always the case, as a matter of interest to the product itself, as a product that needs to acquire positive values to justify increasing its price and, obviously, profit. Whatever is contingent on it, such as safety for pedestrians, is desirable, but less of a priority than for *trolley* users. If there were, by any chance, a disintegrator button for the *trolley* and its

problems may persist, whether they be the current ones or the new ones that will soon arise.

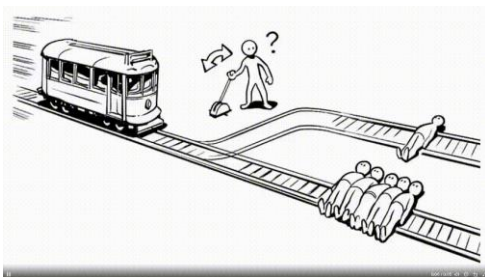
What is still done today in terms of responsible *design* is to choose and promote certain values in the phase that precedes the construction process of technological applications - so that the prototypes of such developments already have the ethically selected values incorporated into them, and therefore as a necessary part of the development, and no longer contingent or accidental.

The research experience for the development of this dissertation, at the Signo Project, has shed light on how such processes can take place in an ethically responsible way, in practice, when developing solutions in technological applications for certain responses. In the case of the Signo Project, these responses refer to the needs of doctors and glaucoma patients in the field of ophthalmology.

occupants, so that in the event of an imminent lethal accident it could be activated and no one "innocent" would be harmed, would this mechanism be raised as an option? Probably not, because who would risk riding in something where their safety is not a priority and its complete disintegration is a real possibility?

The most ethical decision in the case of *trolleys* would be if, even at the *design* stage, it was considered: "if a *trolley* can kill someone, then the project is cancelled" - but obviously this is not raised. So the (teleological, intentional) hierarchy of the fourth dimension could end up overriding the others. And theorists are still arguing about how to deal with the effects of problems arising from incomplete designs today, whether with *trolleys* or with autonomous vehicles already operated entirely by AI.

Recently, someone asked an AI (<https://klimgai.com/>) how to solve the *trolley* problem and the answer, in pictures, was better than ours so far and shows that the AI just slowly backed away from the *trolley*, refusing to move on.



Original post: "I tried to see how Kling v1.6 would handle the trolley problem." And the answer was, in terms of images: "But it just backed away slowly." (Published on X on 07/Jan/2025 and accessed via the link <https://x.com/fofrAI/status/1876638297134678173> on 12/Jan/2025).

So, in order to form the basis for ethically responsible application projects based on values previously defined as relevant and necessary, various interactions between the teams of *designers*, programmers, philosophers and doctors chose scenarios and positions in which these values were identified in the analyses methodological, in this case thematic analysis. In this way, the *design* team was able to identify, among all the values, those that are transversal, but not only. This is obviously a step forward in terms of responsible *design*.

Figures 7 and 8^{cxlvi} show the investigative elements used as resources in the Sign Project.



Figure 7- Framework used to support the elicitation of relevant values in the dynamics between *stakeholders*.



Figure 8- *Cards* of the possible relevant values used as resources, based on the questions posed by the *designers*.

The insertion of values into *design* is just like that: a multifaceted process that ends up involving both objective and subjective considerations, based on these intersubjective relationships (Spiekermann, 2016). This process still needs to be carried out in an analogue way until we can digitally consider an AI (of the future, equipped with a metaontological protocol structure) that performs this operation of choosing values from the various interested parties relevant to projects (*stakeholders*), providing a synthesising and analytical output of the four dimensions considered, as well as certain specific points of accountability for attributions.

^{cxlvi} Images credited to Fraunhofer Portugal AICOS (<https://www.linkedin.com/company/fraunhofer-portugal-icos/>), published on LinkedIn, accessed on 12/Jan/2025 in the *post* available via the link

A complex analysis to consider the most appropriate and relevant values, such as the one carried out by the Signo Project team, needs to take into account at least, but not only: (1) intrinsic and extrinsic values^{cxlvii} , (2) *design capital*^{cxlviii} , (3) subjective influence^{cxlix} , (4) hierarchy of values^{cl} , (5) elicitation methods^{cli} and an (6) appropriate visual communication .^{clii}

3.8. What does it all mean, anyway? From Thales of Miletus to AI...

^{cxlvii} Intrinsic values are those inherent in the expectations of using the app - the probable experience. Extrinsic are values correlated to something external or transcendent to the app itself - such as the patients themselves, or even doctors, regarding their impression of the app before or after use (Spiekermann, 2016).

^{cxlviii} These are the values that help position design as a relevant asset in terms of economic impact. The development of the application's design must generate value (and obviously even in the sense of financial results) and fulfil the needs and purposes of all stakeholders: investors, doctors, users, etc. (McKay & Meyboom, 2022).

^{cxlix} Subjective influence takes into account one's own subjective involvement, which leads design teams, even if they realise their high degree of subjectivity required, to take measures and precautions so that this does not result in bias in the analyses carried out. It is well known that the designer's experience, culture and values influence their choices and, consequently, the value of the resulting design (Innella & Petroni, 2019).

^{cl} The consideration and definition of a hierarchy of values incorporated into a project is fundamental and must be clearly defined. Thus, the values were considered, but among them there was a hierarchisation in relation to relevance and duration, to generate transversal values (Spiekermann, 2016).

^{cli} Which are the various techniques that can be used to identify and elicit the relevant values for a project: questionnaires,, interviews, co-creation workshops, analyses of market data etc.

In the case of the Signo Project, in addition to the thematic analyses applied to processes and content, a kind of interactive simulation was also developed in which participants assessed the relevance in terms of values in given hypothetical situations created and managed by the design teams.

^{clii} Which plays an exceptional role in communicating the value of design. Diagrams, sketches, prototypes and visual presentations translate abstract concepts into concrete ideas and, as a result, *stakeholders* come to realise something closer, due to the objectivity that emerges from the various intentions. (Mozota, 2023)

Ever since we quoted Nagel at the beginning of this dissertation, and with every sentence we write, the question "what does it all mean?" has never come close to being satisfied. And, we believe, it never will be, considering our insatiability for knowledge. AI, after all, is closely related to this insatiability for knowledge, if seen as a means, as we argued at the beginning and as we can see for now.

Considering AI as an epistemic source, what would it be? How does knowledge come about through an AI? Perception? Memory? Logical inference? Testimony? Instrument?

Of the many possible ways of achieving knowledge, we can consider that AI, even if considered as a mere instrument, does not invalidate it as an epistemic source, as well as the other options considered. The assumption of what "the" truth is, however, is another matter altogether and a complex topic in itself, which we have only touched on superficially, and which deserves more attention, especially when we aim to achieve it by extrapolating the objectivity of the data towards subjectivity. Therefore, we can realise that this proposed liminality will really be overcome when it is used as an unconditional means that leads to knowledge.

However, we can consider that reaching this level of instrumentality to knowledge through technological artefacts - such as the AI of the future - will be a complex process involving *designers* and users. There is a gap in analysing the acquisition of knowledge from technologies that interact with natural language (Freiman, 2023).

From an epistemological point of view, contrary to what AI theorists argue, an instrument is designed and built to be, in fact, opaque: it encapsulates the complexities of the instrument's inner workings, allowing the user to understand and accept the output without further ado (Neges, 2018). Therefore, combating the opacity of an AI is the second absurdity. The first is to pose it as a problem.

In the *design* phase, instruments are also conceived and constructed to produce strategically selected signals, which are specific indicators of a prior and very well-defined intentionality, and this is produced to be even inaccessible to the user (just like the intentionality that is also supposed to be unfathomable, according to Tunhas, 2021).

The theoretical perspective of the instruments, established exclusively by the *designers*, must be taken into account - hence the concern for the meta-ontological system to operate as a protocol, with constructive advantages when adopted, and not as a legal determination in itself. From the users' point of view, there needs to be an understanding and acceptance of the propositional content of the outputs - a minimum level of understanding will be necessary, which seems to be jeopardised by the decline^{cliii} that is being seen globally. There is also the need for the user to trust the instrument, which needs to be operated correctly. Instruments need to be calibrated and, in the case of AI, such calibration can be considered to generate fair results - without problems. Thus, we can consider that an AI, as an instrument, can create knowledge from the complex interaction it is able to realise between users in their various forms of relationships (Neges, 2018).

^{cliii} Find out more at <https://www.oecd.org/en/about/programmes/pisa.html> accessed on 11/Jan/2025.

Conclusion and Final Considerations

In conclusion, we believe that the following argumentative constructions, made throughout the dissertation, have been sufficiently consolidated to conclude that:

(1) AI is both a supply and a generator

We argue that AI, as a product, is aimed at meeting human needs and desires, acting as an offer through *big tech*. However, it also impacts users and even social forms, which in turn consume more and create new demands.

This dynamic generates a cycle in which AI, while meeting demands, also stimulates the creation of new ones through strategies in line with capital's own intentions.

Capital plays a fundamental role in this process, with its intentions directing the development and application of AI. Any analysis and proposed solution to the problem, therefore, in relation to AI must necessarily be subordinate and related to this dynamic of capitalist interests, in order to be feasible.

(2) Dialogue is the most valuable product

We argue that *big tech* recognises the value of dialogue as a way of meeting users' needs and desires in a personalised and efficient way. Conversational AI, in particular, is a powerful tool for automated commercial exploitation of dialogue on a large scale and potentially reaching the level of customisation of individuality.

We stress that dialogue in this context is not limited to the mere exchange of information, but encompasses AI's ability to understand and respond to the epistemological, emotional and even cognitive needs - beliefs, desires, wishes and mental states of users. Through dialogue endowed with the interpretation of such subjective dimensions, AI can influence user behaviour, generate demand for products and services, and even shape the perception of reality.

Here, ethical challenges related to the commercialisation of dialogue have arisen - in which the problems mapped out are justified as relevant. AI's ability to persuade and manipulate raises concerns about user autonomy and cognition. We advocate the need for responsible design and control mechanisms to ensure that AI is used ethically and

transparently, since AI presents immense potential for the market, but also requires in-depth reflection on its ethical and social impacts.

The search for a balance between commercial exploitation and the protection of users is the real challenge for legislators, who are less able to perceive what is really going on, given the capillarisation of the actions of a conversational AI, given at an individual level.

(3) Capital has an intentionality that cannot be overlooked

We argue that the intentionality of capital in this context is manifested in the way AI is designed, developed and applied. Capital acts as the investor in AI (to create products and services that meet the demands of the market, generating profit and expanding its influence); and it does so by understanding the needs and desires of users (because it seeks to understand its users, or consumers, as much as possible) and thereby generating demand (by exciting users and thereby creating new demands by influencing their behaviour).

Therefore, exploiting and amplifying this cycle through dialogue is very convenient for capital. The risk occurs when the intentionality of capital overrides the interests of users, ignoring ethical and social aspects, jeopardising the autonomy, privacy and well-being of users in the unbridled pursuit of profit and market expansion.

We therefore argue that the same methodological principles that will allow capital to analyse the discourse of users will also allow the creation of tools that will analyse the influential discourse of capital. We have evoked the concept of pharmakon several times, when the difference between medicine and poison is simply the dose.

(4) Ethical problems exist, but they are poorly characterised

We argue that the lack of a consistent ontological foundation affects and prevents an accurate analysis of ethical problems. There is also the argument that there is a teleological dissonance - when a conversational AI offers exactly what it was designed to accomplish.

We cite that the inadequate characterisation of ethical problems manifests itself in the following aspects: a focus on effects rather than causes (due to teleological dissonance);

a static view of AI (using a static - or analogue - theorisation for something that is eminently dynamic - or digital); a disregard for the intentionality of capital (ignoring how it influences the way AI is designed, developed and applied).

To overcome these challenges, we advocate the advancement of ontological investigations, leading to the development of a metaontological protocol that allows for the representation of existential complexity and the integration of the objective, intersubjective and subjective dimensions, in addition to the capital dimension itself - given as the fourth dimension that influences and manages the other three.

(5) The map of ethical problems is inconsistent

We explored and criticised the map presented by (Mittelstadt et al., 2016), concluding that, due to its static approach to a dynamic system and its failure to consider the current teleology of AI, it presents an inconsistency between times, leading us to a technological liminality in terms of AI. However, the value of the map is that it is a guide for constructing, based on the problems considered, how an AI of the future could exist, in order to overcome these problems, although others may occur.

(6) Any improvement will require an ontological adjustment between the current and the digital

We argue that the lack of ontological correspondence between the real ("actual") world and its digital representation is the source of the ethical problems mapped out. Investing in a meta-logical approach will take AI a long way forward, as will the ethical risks, of course, but with the difference that, unlike today, there will be ways of extracting the "medicine" itself from the same source as the "poisons", through tools that could be useful for analysing data in epistemological terms or even in terms of communication strategically produced to influence or curtail freedoms, moral constraints or unwanted levels of prejudice, but not only.

As a final consideration (or recommendation):

(7) There is an urgent need to invest in modelling a meta-ontological protocol for representing the existential modal that is capable of capturing and emulating

relational complexity and thus operating values to result in an ethical and responsible AI.

We therefore believe, based on the various arguments and conclusions, that the creation of an AI that is ethical, transparent and that benefits both its users and society as a whole requires a change not just in *design*, but in paradigm, which abandons the static vision - in data - for another dynamic one, based on the existential modal represented in its maximum complexity, based on its relationships.

Bibliographical references

- Ajana, B. (2020). Personal metrics: Users' experiences and perceptions of self-tracking practices and data. *Social Science Information*, 59(4), 654-678.
<https://doi.org/10.1177/0539018420959522>
- Akın, Ö. (2021). *Design Added Value: How Design Increases Value for Architects and Engineers* (1.^a ed.). Springer.
- Alves, P. M. S. (2022). Ontology and Epistemology in Husserl's Ideen-I. *Phainomenon*, 33(1), 3-24. <https://doi.org/10.2478/phainomenon-2022-0002>
- Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. In *Wired*. *Wired*. <https://doi.org/https://www.wired.com/2008/06/pb-theory/>
- Ben Mahria, B., Chaker, I., & Zahi, A. (2021). A novel approach for learning ontology from relational database: from the construction to the evaluation. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00412-2>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001, May). The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities. *Scientific American Magazine*.
<https://www.scientificamerican.com/article/the-semantic-web/>
- Bostrom, N. (2014). *Superintelligence: paths, dangers, strategies*. Oxford University Press.
- Brey, P., & Søraker, J. (2009). Philosophy of Computing and Information Technology. In A. Meijers, D. Gabbay, P. P. Thagard, & J. Woods (Eds.), *Philosophy of Technology and Engineering Sciences* (Vol. 14). Elsevier.
- Castoriadis, C. (2000). *The Imaginary Institution of Society* (5.^a ed.). Paz e Terra.
- Dasgupta, S., Papadimitriou, C. H., & Vazirani, U. V. (2006). *Algorithms*. McGraw-Hill Higher Education.

- de Boer, B., Te Molder, H., & Verbeek, P. P. (2018). The Perspective of the Instruments: Mediating Collectivity. *Foundations of Science*, 23(4), 739-755.
<https://doi.org/10.1007/s10699-018-9545-3>
- Digman, J. M. (1990). Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, 41, 417-457.
<https://doi.org/https://doi.org/10.1146/annurev.ps.41.020190.002221>
- Dreyfus, S. (2002). Richard Bellman on the Birth of Dynamic Programming. *Operations Research*, 50(1), 48-51.
<https://doi.org/http://dx.doi.org/10.1287/opre.50.1.48.17791>
- Ezell, B. C., & Crowther, K. G. (2007). Philosophical issues and their implications for the systems architect. *Foundations of Science*, 12(3), 269-276.
<https://doi.org/10.1007/s10699-007-9108-5>
- Floridi, L. (2012). Big data and their epistemological challenge. In *Philosophy and Technology* (Vol. 25, Issue 4, pp. 435-437). <https://doi.org/10.1007/s13347-012-0093-4>
- Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology 1. In *THE BEHAVIORAL AND BRAIN SCIENCES* (Vol. 3).
- Foucault, M. (1979). *Microfísica do Poder* (R. Machado, Ed. & Trad.; 1.ª ed.). Edições Graal.
- Foucault, M. (1987). *The Archaeology of Knowledge* (L. F. B. Neves, Trad.; 3.ª ed.). Forense-Universitária.
- Frankfurt, H. G. (2005). About talking shit. *Intrínseca*. www.intrinseca.com
- Freiman, O. (2023). Analysis of Beliefs Acquired from a Conversational AI: Instruments-based Beliefs, Testimony-based Beliefs, and Technology-based Beliefs. *Episteme*, 1-17. <https://doi.org/10.1017/epi.2023.12>
- Genep, A. Van. (2011). *The Rites of Passage* (M. Ferreira, Trad.). Vozes Editora.

- Goddard, C. (2018). *Ten Lectures on Natural Semantic Metalanguage - Exploring Language, Thought and Culture Using Simple, Translatable Words*. Brill.
https://doi.org/10.1163/9789004357723_001
- Goddard, C., & Wierzbicka, A. (2011). Semantics and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(2), 125-135. <https://doi.org/10.1002/wcs.101>
- Guizzardi, Giancarlo (2005). *Ontological foundations for structural conceptual models*. Centre for Telematics and Information Technology, Telematica Instituut.
- Habermas, J. (2012). *Theory of communicative action - vol. 2: On the critique of functionalist reason (1^a)*. WMF Martins Fontes.
- Harding, S. (2011). *The Postcolonial Science and Technology Studies Reader*.
<https://www.dukeupress.edu/the-postcolonial-science-and-technology-studies-reader>
- He, B., Patel, M., Zhang, Z., & Chang, K. C.-C. (2007, May). Accessing the deep web. *Association for Computing Machinery*, 50(5), 94-101.
<https://doi.org/https://doi.org/10.1145/1230819.1241670>
- Heidegger, M. (1999). *Introduction to metaphysics* (C. Leão, ed.; 4^a). Tempo Brasileiro.
- Heidegger, M. (2008). *Parmenides*. Editora Vozes / Editora Universitária São Francisco.
- Heidegger, M. (2012). *Being and Time* (UNICAMP Multilingual Philosophy Collection). Edition in German and Portuguese; translated and organised by Fausto Castilho (10th ed.) Editora Unicamp/Editora Vozes.
- Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26(2). <https://doi.org/10.1007/s10676-024-09775-5>
- Innella, G., & Petroni, M. (2019). *Going Real: The Value of Design in the Era of PostCapitalism*.
- Kant, I. (2008). *Prolegomena to All Future Metaphysics* (A. Morão, Trad.; 1.^a ed.). Edições 70.
- Kant, I. (2018). *Critique of Pure Reason* (9.^a ed.). Calouste Gulbenkian Foundation.

- Keul, H.-K. (2002). Subjectivity and Intersubjectivity: Remarks on the Concept of Freedom in Kant and Habermas. *SUBJECTIVITY AND INTERSUBJECTIVITY The Journal of Value Inquiry*, 36, 253-266.
- King, M., Valle, R. S., & Halling, S. (1989). *Existential-Phenomenological Perspectives in Psychology: Exploring the Breadth of Human Experience* (R. S. Valle & S. Halling, Eds.).
- Lange, B., & Lechterman, T. M. (2021). Combating disinformation with AI: Epistemic and ethical challenges. <https://philpapers.org/archive/LANCDW-2.pdf>
- Lear, J. (1988). *The Desire to Understand*. Cambridge University Press.
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, Issue 7553, pp. 436-444). Nature Publishing Group.
<https://doi.org/10.1038/nature14539>
- Martinez-Cruz, C., Blanco, I. J., & Vila, M. A. (2012). Ontologies versus relational databases: Are they so different? A comparison. In *Artificial Intelligence Review* (Vol. 38, Issue 4, pp. 271-290). <https://doi.org/10.1007/s10462-011-9251-9>
- May, R. (2007). *Love and Will*. W. W. Norton & Company, Inc.
- McKay, S., & Meyboom, A. (2022). *Design Capital - The Hidden Value of Design in Infrastructure*. Routledge.
- Meersman, R. (2001). Ontologies and Databases: More than a Fleeting Resemblance. In A. d'Atri & M. Missikoff (Eds.), *Proceedings of the OES/SEO 2001 Rome Workshop*. LUISS Publications. <http://www.starlab.vub.ac.be/publications/OES-SEO2001.pdf>
- Merleau-Ponty, M. (1999). *Phenomenology of Perception* (C. A. R. de Moura, Trad.; 2.^a ed.) Martins Fontes.
- Miguens, S. (2007). *Philosophy of Language - an introduction*. Faculty of Letters, University of Porto.

- Miller, C. (2014). Moral Virtues, Epistemic Virtues, and the Big Five. In O. Flanagan & A. Fairweather (Eds.), *Naturalising Virtue* (pp. 92-117). Cambridge University Press.
- Mises, L. von (1968). *El Socialismo* (L. Montes de Oca, Trad.; 2.^a ed.). Instituto Nacional de Publicaciones de Buenos Aires.
- Mises, L. von (2010). *Human Action* (D. Stewart Jr., Trad.). Ludwig von Mises Institute Brazil.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society*, 3(2).
<https://doi.org/10.1177/2053951716679679>
- Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake News" Is Not Simply False Information: A Concept Explanation and Taxonomy of Online Content. *American Behavioral Scientist*, 65(2), 180-212.
<https://doi.org/https://doi.org/10.1177/0002764219878224>
- Mozota, B. B. de. (2023). *Design Management: Using Design to Build Brand Value and Corporate Innovation*. Allworth.
- Neges, M. S. (2018). *DISSERTATION / DOCTORAL THESIS [PhD dissertation]*. University of Vienna.
- Nelson, J. L., & Taneja, H. (2018). The small, disloyal fake news audience: The role of audience availability in fake news consumption. *New Media and Society*, 20(10), 3720-3737. <https://doi.org/10.1177/1461444818758715>
- Newman, L. (2023). Descartes' Epistemology. In E. Zalta & . Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023).
<https://plato.stanford.edu/archives/win2023/entries/descartes-epistemology/>
- Nietzsche, F. W. (2008). *The will to power* (M. S. P. Fernandes & F. J. D. de Moraes, Trads.). Counterpoint.
- Oberst, M. (2015). KANT ON UNIVERSALS. In *Source: History of Philosophy Quarterly* (Vol. 32, Issue 4). <http://www.jstor.org/stable/44076598>

- Oomen, P. (2019). Language about God in Whitehead's Philosophy: An Analysis and Evaluation of Whitehead's God-Talk. *Process Studies*, 198-218.
<https://philpapers.org/archive/OOMLAG.pdf>
- Ortolan, L. (2022). *The cynical and savage guide to the games of life* (1.^a ed.). KDP.
- Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023, October 29). Generative Agents: Interactive Simulacra of Human Behaviour. *UIST 2023 - Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. <https://doi.org/10.1145/3586183.3606763>
- Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Willer, R., Liang, P., & Bernstein, M. S. (2024, November 15). Generative Agent Simulations of 1,000 People. <https://arxiv.org/abs/2411.10109>
- Partridge, C. (2015). Why (and how) to use a metaphysicalist foundational ontology
Title: Why (and how) to use a metaphysicalist foundational ontology.
<https://www.researchgate.net/publication/370510623>
- Pennock, R. T. (2019). *An Instinct for Truth: Curiosity and the Moral Character of Science*. MIT Press. <https://mitpress.mit.edu/9780262042581/an-instinct-for-truth/>
- Plato (2011). *Timaeus-Critias* (M. do C. Fialho, Ed.; R. Lopes Plato, Trad.; 1.^a ed.). Centre for Classical and Humanistic Studies.
- Portmess, L., & Tower, S. (2015). Data barns, ambient intelligence and cloud computing: the tacit epistemology and linguistic representation of Big Data. *Ethics and Information Technology*, 17(1), 1-9. <https://doi.org/10.1007/s10676-014-9357-2>
- René Descartes (2005). *Metaphysical Meditations*. Martins Fontes.
- Ricoeur, P. (1989). *The Conflict of Interpretations: Essays on Hermeneutics* (M. F. Sá Correia, Trad.). Rés Editora.

- Ricoeur, P. (2016). *Hermeneutics and the Human Sciences Essays on Language, Action and Interpretation* (J. B. Thompson, Ed.). Cambridge University Press.
<https://doi.org/10.1017/CBO9781316534984>
- Romele, A. (2019). *Digital Hermeneutics: Philosophical Investigations in New Media and Technologies*. www.routledge.com/
- Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson. <https://doi.org/10.1109/MSP.2017.2765202>
- Saramago, J. (2002). *The duplicitous man*. Caminhos.
- Sartre, J.-P. (2011). *O ser e o nada - Ensaio de ontologia fenomenológica* (P. Perdigão, Trad.; 20.^a ed.). Vozes.
- Sartre, J.-P. (2015). *What is subjectivity?* (E. dos S. Abreu, Trad.; 1.^a ed.). Nova Fronteira.
- Shields, C. (2007). *Aristotle*. Routledge. <https://philpapers.org/rec/SHIA-7>
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). *Fake News Detection on Social Media: A Data Mining Perspective*. <http://arxiv.org/abs/1708.01967>
- Spiekermann, S. (2016). *Ethical IT Innovation - A Value-Based System Design Approach*. CRC Press.
- Suits, B. (2017). *The Philosophical Cicada - Is life a game?* Gradiva.
- Szopa, R. (2023). Ethical problems in the use of algorithms in data management and in a free market economy. *AI and Society*, 38(6), 2487-2498.
<https://doi.org/10.1007/s00146-021-01319-5>
- Tolley, C. (2012). Kant on the Content of Cognition. *European Journal of Philosophy*, 20(4), 200-228. <https://doi.org/10.1111/j.1468-0378.2011.00483.x>
- Tunhas, P. (2021). Kant: the domain of nature. In *Metaphysical science: itineraries of a concept* (pp. 215-215).
https://sigarra.up.pt/flup/pt/pub_geral.pub_view?pi_pub_base_id=517865

- Tunhas, P. (2023). Philosophy (J. F. Meirinhos & D. P. Aurélio, Eds.; 1^a). Institute of Philosophy : U. Porto Press. <https://doi.org/https://doi.org/10.21747/978-989-746-365-5/filo>
- Turner, V. W. (1974). *The Ritual Process - Structure and Antistructure* (N. C. de Castro, Trad.). Editora Vozes.
- Van Otterlo, M. (2013). Privacy, Due Process and the Computational Turn. In *Privacy, Due Process and the Computational Turn The Philosophy of Law Meets the Philosophy of Technology* (1st ed., pp. 41-64). Routledge. <https://doi.org/10.4324/9780203427644>
- Vergílio (2019). *Georgicas* (G. A. F. Silva, Trad.). Cotovia Books.
- von Eckartsberg, R. (1998). *Phenomenological Inquiry in Psychology: Existential and Transpersonal Dimensions* (R. Valle, Ed.). Springer.
- Waelbers, K., & Briggie, A. (2010). Three Schools of Thought on Freedom in Liberal, Technological Societies. https://www.pdcnet.org/techne/content/techne_2010_0014_0003_0176_0193
- Whitehead, A. North., & Sherburne, D. W. . (1981). *A key to Whitehead's Process and reality*. University of Chicago Press.
- Withehead, A. N. (2010). *Process and Reality - Essays on Cosmology* (M. T. Teixeira, Ed. & Trad.; 1.^a ed.). Philosophy Centre of the University of Lisbon.
- Wittgenstein, L. (2015). *Logical-Philosophical Treatise Philosophical Investigations* (6^a). Calouste Gulbenkian Foundation.
- Wood, A. W. (1975). Kant's Dialectic. *Canadian Journal of Philosophy*, 5(4), 595-614. <https://doi.org/10.1080/00455091.1975.10716970>
- Zamborlini, V. C. (2011). Study of alternatives for mapping ontologies from the ONTOUML language to OWL: Approaches for representing temporal information. <https://doi.org/>