Science and Engineering Ethics (2023) 29:32
https://doi.org/10.1007/s11948-023-00451-3

ORIGINAL RESEARCH/SCHOLARSHIP

# AI as an Epistemic Technology

**Ramón Alvarado**[1]

Received: 3 November 2022 / Accepted: 12 July 2023


## Abstract
In this paper I argue that Artificial Intelligence and the many data science methods associated with it, such as machine learning and large language models, are first and foremost epistemic technologies. In order to establish this claim, I first argue that epistemic technologies can be conceptually and practically distinguished from other technologies in virtue of what they are designed for, what they do and how they do it. I then proceed to show that unlike other kinds of technology (*including* other epistemic technologies) AI can be uniquely positioned as an epistemic technology in that it is primarily designed, developed and deployed to be used in *epistemic contexts* such as inquiry, it is specifically designed, developed and deployed to manipulate *epistemic content* such as data, *and* it is designed, developed and deployed to do so particularly through *epistemic operations* such as prediction and analysis. As has been shown in recent work in the philosophy and ethics of AI (Alvarado, AI and Ethics, 2022a), understanding AI as an epistemic technology will also have significant implications for important debates regarding our relationship to AI technologies. This paper includes a brief overview of such implications, particularly those pertaining to explainability, opacity, trust and even epistemic harms related to AI technologies.

**Keywords** AI · Technology · Epistemic · Knowledge tools

✉ Ramón Alvarado
ralvarad@uoregon.edu

1 Philosophy Department, University of Oregon, Eugene, OR, USA

Published online: 21 August 2023

## Introduction

In this paper I argue that Artificial Intelligence (AI) and the many data science methods associated with it are first and foremost epistemic technologies.[1] In order to establish this claim, I first argue that epistemic technologies can be conceptually and practically distinguished from other technologies in virtue of what they are designed for, what they do, and how they do it. I then proceed to show that unlike other kinds of technology (*including* other epistemic technologies) AI can be uniquely positioned as an epistemic technology in that it is primarily designed, developed *and* deployed to be used in epistemic contexts such as inquiry, it is explicitly deployed in such contexts to manipulate epistemic content such as data, *and* it manipulates such content specifically through epistemic operations such as inferences, predictions or analysis.

In order to best emphasize the relevance of the main claim(s) in this paper, I begin by providing an overview of the conceptual landscape, from philosophy and other disciplines, in which these claims are to be made. I make the case that an appeal to the uses of a technology—which most conventional approaches to this issue tend to rely on—is not adequate to capture their nature as epistemic technologies. In this first section, I also provide the conceptual groundwork upon which the concept of epistemic technologies is to be reengineered (Simion, 2018). In particular, I examine and expand on Paul Humphreys' (2004) notion of epistemic enhancers. Through this examination, both concepts—epistemic enhancers and epistemic technologies—are distinguished form other similar terms such as cognitive extenders and cognitive artifacts. Importantly, however, they are also distinguished from each other.

Once these distinctions are in place, the next section sets out to provide the positive argument of how and why AI does not just qualify as an epistemic technology according to the considerations in the previous section, but also qualifies as a paradigmatic case of an epistemic technology, which further differentiates it even from other epistemic technologies. As I will argue, while many technologies do take part in epistemic contexts, very few do so while *also* manipulating epistemic content *and also* manipulating such content through epistemic operations. As I will show, those technologies that partake in knowledge acquisition endeavors across these three dimensions are distinct from those that do not in a non-trivial manner and it is this distinct character that makes them epistemic technologies first and foremost. Capturing this distinct epistemic character, I will argue, requires that we take design intentions seriously as constitutive of the nature of technological artifacts. However, merely doing so will not suffice. We must also distinguish what something is designed, developed and deployed for from what something does and how it does it. Taking these distinctions seriously as constitutive dimensions of technical artifacts and seeing that at each of these dimensions there is a distinctly particular epistemic

[1] The term AI is here purposedly left as maximally inclusive to refer to many distinct computational technologies that characterize its development in the past decades, from longstanding machine learning (ML) methodology, including deep neural networks (DNNs), to more recent development in analytic tools like transformers, large language models (LLMs) and multimodal generative models that rely on techniques such as gradient descent.

element at play will elucidate in a clearer manner how AI and its related technologies fit into this exclusive category of epistemic technologies.

Finally, a section briefly outlining a non-exhaustive set of philosophical, ethical and epistemological implications of the view defended here is offered. This last section explores the repercussions that understanding AI as an epistemic technology may have on important contemporary debates regarding the significance of explainability, the challenge of epistemic opacity in computational methods, the issues related to the allocation of trust in AI, as well as in debates concerning the hidden dimension of harms that may be particular to AI technologies, as Symons and Alvarado (2022) first pointed out recently.

In short, I argue that epistemic technologies can be conceptually and practically distinguished from other technologies. This is in part because in order to understand artifacts as functionally identifiable objects, we must take into account that their *dual nature* (Kroes, 2010) is in a non-trivial way constituted by an intentional design (Symons, 2010). This intentional design tells us what the artifact is made for, which will in turn inform why it has the properties is has, which will in turn inform the way it is build and why it can do what it does in the way that it does.[2] Furthermore, I will argue that unlike other kinds of technology (*including* other epistemic technologies), AI is particularly and uniquely an epistemic technology in that it is primarily designed, developed and deployed to be an epistemic technology in *all three* of the above-mentioned senses. That is, they are primarily:

1. Designed, developed and deployed for use in epistemic contexts
2. Designed, developed and deployed to manipulate epistemic content *and*
3. Designed, developed and deployed to carry out epistemic operations on such content.

As I mentioned, I will show that the distinctive character of AI as an epistemic technology will prove important for a variety of philosophical debates. Yet the ramifications of their proper identification will also serve as a framework to understand the adequacy or inadequacy of our relationship to them in the contexts of use, trust and hence policy and regulation.

## Setting Up the Conceptual Space: Epistemic Technologies, What they are and What they are Not.

Many different artifacts that happen to partake in diverse knowledge acquisition, knowledge gathering and knowledge disseminating practices have been conventionally identified as epistemic technologies across multiple disciplinary frameworks.

[2] As we will see, this is so even if the artifact can be or is used for something other than what it was originally intended for. A brick can be used as a brick, as a door stop, as a step, or as a weapon, but it can only be used for other things than as a brick in virtue of the fact that it was built as a brick and not a gelatin dessert in the first place.

In media studies literature, for example, the term ‘epistemic technologies’ has been used to refer to a broad assortment of old and contemporary pedagogical methods, including things like search engines and slide presentations (Hakkarainen et al., 2009). In the philosophy of technology one can find the term in reference to iPhones and GPS navigation systems (Miller & Record, 2013, 2017; Ratto, 2012; Record & Miller, 2018). Similarly, in the history of science and technology, tools of knowledge—as they are commonly referred to therein—can include anything from Galileo’s demonstrative instruments to objects like recipe books (Becker et al., 2001;Daston, 2012; Friedrich, 2018).[3]

What these accounts have in common is that the designation of ‘epistemic technology’ or ‘tool of knowledge’ is ascribed to an artifact mainly in virtue of its deployment *within* an epistemic context or *for* a broadly-construed epistemic purpose such as inquiry.[4] Anthony (2018), for example, defines epistemic technologies as “instruments used to engage in investigation and construction of knowledge” (2018, p. 663). In their own words:

> Epistemic technologies are distinct from production technologies, which primarily enable the speed and efficiency of production processes. Although some epistemic technologies may also shorten time in the work of producing knowledge, the primary role of an epistemic technology is to enable the ongoing generation of knowledge. (p. 663)

As examples, Anthony uses the telescope, computer aided design and also slide presentations in a classroom or conference setting. Seemingly, under this view, if we are looking for answers and we deploy a tool, any tool, to helps us, that tool can be categorized as an epistemic technology.

As stated, however, and without any additional caveats or specifications, these types of conventional and widely used accounts simply fail at identifying the seemingly unique characteristics of epistemic technologies. This is because, while this criterion makes it so that it is uncontroversial that an oscillating shaker used in a chemistry laboratory as part of a scientific experiment fits the label, so does a light switch, a sitting stool, a glove, or a stirring stick in a similar setting. And while these latter items are *not* to be simply dismissed without reason from belonging to such a

[3] As we will see in the sections below, there are important developments in the philosophy of science where the epistemic component of computational methods is in fact recognized, albeit with some similar limitations to the ones discussed here. For now, however, let us focus on the accounts briefly mentioned here.

[4] A further common thread, that is orthogonal to our discussion, is a broad and liberal understanding of knowledge-acquisition practices that precludes a distinction between scientific undertakings and ordinary epistemic practices. Under this view, science is simply a continuation of ordinary empirical endeavors such as tasting, touching, or looking at things to gather information. In short, crossing the street and colliding subatomic particles at CERN are seen as both being and belonging to the same kind of epistemic enterprise. These views are in part the product of naturalized epistemology projects that understand a baby repeatedly throwing their milk bottle on the ground as a low budget experiment in gravitational physics. While the deflationary and reductionist assumptions inherent in these views are problematic, it is important to acknowledge that there is a rich an important debate pertaining to what constitutes scientific inquiry and what does not. This issue in the philosophy of science is called the “demarcation problem.”

category, their categorization as epistemic technologies does seem less immediately obvious than the aforementioned oscillating shaker. While there is a sense in which a sociologist of science may want to include them in a taxonomy of key tools in the pursuit of knowledge, there are non-trivial differences between kinds of technical artifacts that ought not to be neglected.

In order to more clearly see why this is the case, consider that there seems to be a marked distinction, however elusive, between the kind of work that a shelf or a wall does in a laboratory—and hence a marked distinction in the kind of artifact that a wall is—and the kind of work done by a spectrophotometer in a laboratory. While both could be construed as *enabling* the ongoing generation of knowledge or *engaging* in an investigation or construction of knowledge, what they do, how they do it, and what properties allow them to do such things in such ways, seem markedly distinct.

Hence, the first problem with this broad use of the term is obvious from a conceptual perspective. The criterion to qualify a tool as an epistemic technology simply on the bases of its deployment in epistemic contexts or its participation in a broadly-construed epistemic task seems inadequate. We want our concepts to capture something meaningful and to provide epistemically useful information when we use them (Simion, 2018). In other words, we want concepts to help us do things and identify things in the world (Baier, 1985). However, if we take into consideration only the reasons stated above—that an artifact be used in inquiry— as the sole criterion related to the identification of epistemic technologies, the distinguishing aspect of epistemic technologies as such, to which Anthony alludes in the quote above, ceases to be of much use. Suddenly all things can be epistemic tools or all things are already epistemic tools: cabinets, telescopes, paper towels, software-intensive devices, and coffeemakers alike.[5]

Thus, this criterion proves to be at best uninformative and at worst misguiding. It is uninformative in that it tells us nothing useful about the world or the objects we are interested in. It is misguiding because it leads us to believe that there are no meaningful distinctions between kinds of artifacts, their functions and therefore no meaningful distinctions in how we ought to relate to them, e.g., it makes it so that both an abacus and a hammer have no non-trivial differences as tools: perhaps they

[5] Notice that while this last item sounds as an exaggeration, the broadness of the criteria so far described allows so that any artifact deployed that "enables" the continuing acquisition of knowledge qualifies. If one accepts that culinary objects such as food and drink are artifacts and that they are used—often explicitly, as in the case of caffeinated drinks—to enable or enhance a practitioner's stamina in the laboratory, then we would have to accept the coffee maker's central role in inquiry. Furthermore, while one may want to debate the relevant distinctions between a culinary object such as coffee as an artifact and other mechanical objects such as technical artifacts, this distinction would do little to exclude the complex devices designed, developed and deployed to make the coffee itself from being considered.

are the same kind of tool or can be used for the same things.[6] It simply trivializes the concept in a way that it fails to capture anything meaningful.

From a practical perspective, this approach to understanding what Anthony calls the "primary role of epistemic technologies" also risks ignoring the fact that while one may find seemingly non-epistemic artifacts in a laboratory, such as glass containers, etc., such artifacts are not run-of-the-mill artifacts such as the ones found in our kitchens or other informal spaces. Rather, artifacts deployed as aids of formal inquiry are produced by a complex industry that specializes in precision standards. This has been so for at least a couple of centuries (Golinski, 1994). These artifacts are the product of detailed specifications that have gradually emerged from the consideration of complex theoretical frameworks, principled metrics and sophisticated epistemic considerations (Alvarado, 2021a). These specifications and considerations in turn sanction the use of such artifacts in such specialized settings (Helden & Hankins, 1994; Van Helden, 1994; Baird, 2004; Alvarado, 2022a, 2022b).

In other words, what qualifies these artifacts to be used in such settings is not simply that they are used for such purposes, but rather the fact that they were *primarily* designed and *made for* such uses. As already alluded to above, other kinds of technical artifacts, such as pharmaceuticals or bulldozers, for example, are not primarily made for these kinds of purposes, tasks, or contexts. Even if these artifacts partake in the enabling of knowledge acquisition and creation, they do so indirectly. They are made to do other kind of tasks such as digging, manipulate other kinds of content such as physical matter, and are deployed in other kinds of contexts such as construction or destruction projects.[7] That is, they are not in any *significant* manner designed, developed or deployed as aids in knowledge-acquisition, knowledge-creation, or knowledge management endeavors. As such, they are simply not epistemic technologies.

As we will see, design by itself, though a central aspect of epistemic technologies, may only provide a starting point on our quest to understand the distinctive characteristics of epistemic technologies. Nevertheless, as detailed in the following sections, when coupled with other features, such as the kind of content the technology manipulates and the kind of operations that the technology executes, we may have a better grasp about the relative uniqueness of epistemic technologies in the world of tools.

[6] Some may reply here that it is simply true that you can build an abacus from a hammer and a hammer from an abacus or abacuses from hammers and vice versa. However, notice that you still have to *build an abacus* from *hammers* and vice versa, i.e., the hammer in and of itself as is and in virtue of what it is does not do the job of an abacus and cannot do the job of an abacus unless appropriately arranged as such. This implies, at the very least, that there is *something* that differentiates them at a fundamental level. That something, as we will see, may be a product of function, engagement, materiality, target phenomena, etc., but it is not merely use.

[7] As we will see, things like pharmaceuticals are designed to intervene chemically with organisms, they work through biochemical interactions and are often deployed and designed for contexts of care and not contexts of inquiry.

## Epistemic enhancers

So far, our discussion suggests that deeming a technical artifact an epistemic technology simply on the basis of its use fails to capture the unique distinctive features that differentiate these tools from others. Simply stated, use in an epistemic context is not a sufficient condition for an artifact to be deemed an epistemic technology, at least not in any informative, distinctive, non-trivial sense. Anthony (2018) understood as much. Citing Dougherty and Dune (2012), she considers that what makes an epistemic technology *epistemic* is that when scientist use it, they are operating "with a technology of representation (not of treatments and intervention); relying on signs about the object of study rather than on material objects." (p. 663) With this brief quote, Anthony hints at, though does not expand on, a more nuanced view: what makes a technology epistemic is that the tool itself has a distinct function (a representational function) and that it manipulates a distinct material (symbols). As we will see, this is a step in the right direction to more properly identify and articulate the distinctive character of epistemic technologies. However, before we can fully grasp the implications of these added features it is important for us to understand in a more precise manner what we mean with the label 'epistemic' as a qualifier of kinds of artifacts.[8] Hence, in this section I expand on Paul Humphreys' notion of an *epistemic enhancers* (2004) in order to best understand what it is that makes epistemic technologies epistemic. While, as we shall see, Humphreys' concept also has some limitations, it nevertheless gets us closer to identifying the distinctive character of epistemic technologies alluded to above, particularly as the term relates to computational artifacts. Therefore, this section provides a clearer view of the epistemic element in all epistemic technologies while also providing a direct segue to the consideration of computational technologies specifically.

So, what is an epistemic enhancer? Humphreys concept seeks to capture a distinct set of technical artifacts in virtue of the human capacities that they enhance: while some tools enhance our physical capacities, others enhance our capacities to acquire knowledge. Technologies such as the microscope, for example, address deficiencies, or enhance existing capabilities, that other machinery, say a bulldozer, does not. In particular, these technologies can be said to enhance abilities that are not strictly speaking physical but that are rather cognitive in nature.[9] For example, the microscope and the telescope enhance vision and visual tasks. Visual tasks are already epistemic in a way that shaking a substance, the way an orbital shaker does in a chemistry lab, is not.

Epistemic enhancement, however, is not limited to our perceptual abilities. In fact, such perceptual enhancement may not be what is most interestingly captured by Humphreys' concept.

[8] Here I want to thank the anonymous reviewer that invited clarification on this key term.

[9] A detailed distinction between the terms 'epistemic' and 'cognitive' will be made at a later point in sections below, for now however, we can recognize the cognitive as distinct from the physical and at least as more closely related to knowledge-acquisition capacities than the latter, and this suffices for the claims so far.

Humphreys also notes that some abstract objects such as mathematical proofs, even in their pen-and-paper form, “are instruments that extend out native inferential abilities in ways which, at least, supplement our memory capacity.” (2004, p. 6) Notice here that the focus of the extension (the thing the enhancer enhances) are the inferential capacities, which are in turn connected to our memory capacities. While memory capacities may be intrinsically tied to physical processes, like vision or other perceptual capacities, it is not immediately obvious that inferential capacities are perceptual or physical in the way vision is, as echoed by Anthony’s quote at the start of this section.

According to Humphreys, there are three ways an epistemic enhancer can extend the reach of our understanding: extrapolation, conversion and augmentation. *Extrapolation,* expands “the domain of our existing abilities” (2004, p. 4). Humphreys uses the perceptual enhancement of optical instruments to exemplify extrapolation. As we saw above, artifacts such as telescopes and microscopes, for example, expand the domain of the visible for us.[10] When it comes to computational methods in general, they enhance our *already existing* ability of analysis. We have the ability to manipulate and entertain the relationship between values. We can also extract patterns from the relationships that the symbols that represent these values have and we can infer things from their transformations. Computational methods, from computer simulations to AI, indeed expand on this existing modality.

*Conversion,* on the other hand, occurs when “phenomena that are accessible to one sensory modality […] are converted into a form accessible to another” sensor modality (2004, p. 4). Sound patterns that are visualized on a screen are a good example of this kind of enhancement. Here, the conversion occurs twice, the sound intensity is translated into electrical signals, these signals are interpreted as numerical values and then these numerical values are displayed as a dynamic line or as pixels on a longitudinal bar graph. Many computational and modeling processes can be said to convert information in this sense.

Finally, *augmentation* occurs when—often in tandem with one or two of the other kinds of enhancement—an instrument gives us “access to features of the world that we are not naturally equipped to detect in their original form.” As an example of these compound cases of enhancement, Humphreys offers the following:

> Observational and computational enhancements complement one another when computationally assisted instruments are used with physical devices and mathematics working together to give us increased access to the natural world. An example of this is computerized axial tomography (CAT scans), within which physical detectors pick up differential rates of radiation and the output of those detectors transformed by mathematical algorithms into two-or three-dimensional images of the object under investigation. (2004, p. 5)

[10] Radio-aided telescopes, too, expand the range of the spectrum of electromagnetic radiation available to us. *Software-intensive* (Symons & Horner, 2014) instruments like more modern telescopes pose an epistemologically interesting question in the context of Humphreys’ enhancement taxology. Unfortunately, this question is beyond the space and aims of this current paper.

Accordingly, Humphreys suggests that “our natural mathematical talents have been supplemented by computational devices that enable us, in certain areas, to move beyond what our psychological capacities can achieve.” Computational devices, he wrote, “can greatly increase the speed with which we can perform certain mathematical operations, thus altering the time scale in ways analogous to the way that optical telescopes alter spatial scales.” (Ibid.).

Therefore, they transcend and help us transcend our natural epistemic limitations in several of the aforementioned ways. These technologies, he adds,

> can expand the domain of problem complexity, allowing us to move from studying a restricted range of simple topological structures to structures of far greater complexity or of different types. I can convert numerical results into graphical form, as is often done in statistical analysis, allowing easier access to massive data sets by shifting from one representational mode to another. (2004, p. 6)

Hence, equations and models qualify as epistemic enhancers, but so do other related representational devices as well as translations of one representational mode to another. In other words, epistemic enhancers include formal equations but also the visualizations we derive from them, as is the case with computer simulations (Morrison, 2015; Alvarado, 2021b).[11]

In this sense, many, if not all, computational devices employed in scientific inquiry qualify as epistemic enhancers. This includes AI and all related computational methods: machine learning algorithms, large language models, neural networks, etc. Furthermore, as Humphreys example of the CAT scan in the quote above shows, the types of enhancements that computational methods can provide are sometimes hybrid (Alvarado, 2021a, 2021b), e.g., they can enhance by conversion and/or by extrapolation at the same time.

Although it is not immediately relevant for the argument here, it is worth noting that it is not clear, however, whether or not computational methods allow us to *augment* in the strict sense specified by Humphreys. In his view, computational methods have *not yet* proven to have given us access to mathematical features of the world that, as he states, “we are not naturally equipped to detect in their original form.” Computational methods enhance our epistemic capacities by being faster than us at some epistemic tasks but not by providing us with a *novel* way of doing mathematics or by providing access to previously inaccessible mathematical realms (kinds of mathematical facts, mathematical relationships, etc.). Therefore, if they are

[11] As we will see later, note, that while in this instance Humphreys may be referring to the successful cases of enhancement, the term need not be a success term. In this sense, both a flawed model and a false model count as epistemic enhancers in principle, even if in practice they are not conducive to truths. For a thorough overview of the intricacies related to idealizations in scientific representation see (Pincock, 2011). For a similar overview related more closely to computational methods such as computer simulations see Morrison (2015). Recently, a more focused debate emerged regarding the role of non-factive content in scientific explanations, with an emerging consensus admitting that non-factive content could indeed be a significant part of a scientific explanation (Paez, 2009, 2019). For a thorough refutation of this conventional view, see Sullivan and Khalifa (2019).

epistemic enhancers, they are so only in the sense that they can calculate faster than humans or in that they can render the audible visible or compress intractable problems to be reasonably intelligible, but not in *augmenting* the kind of mathematical world we live in and have access to.[12] It is worth noting also, that this limitation is nevertheless relevant here because, to Humphreys, all computational methods are mathematical methods: they solve mathematical problems through mathematical means for mathematical purposes. The rest of their applications are only derivative of this primary task. This point will be important as an analogy when we analyze AI's epistemic nature as a technology that is first and foremost an epistemic technology in that it deals with epistemic content in epistemic ways within epistemic contexts.[13]

My use of the term 'epistemic' here is, of course, not gratuitous. Each of these references requires that qualifier for reasons that will become obvious in the next section of this paper: as we will see, I am not talking about just any type of content, but epistemic content when I refer to what AI manipulates; I am not taking about just any kind of operation but epistemic operations when I talk about the tasks that AI carries out, so on and so forth. Similarly, Humphreys is not talking about the enhancement of just any kind of human capacity but of the enhancement of epistemic capacities, specifically. Yet, because of this central argumentative use of the term, perhaps a more fundamental question arises: what exactly are these technologies enhancing and how is their enhancement epistemic? Let us address this question before we move on.[14]

As Humphreys' quotes above suggest, these technologies are sometimes enhancing our perceptual abilities, other times they are enhancing our mathematical abilities. Both of these are strongly tied to our psychological and inferential abilities, which Humphreys also mentions as being enhanced. Of course, many of these examples are directly tied to cognitive abilities and these terms—'epistemic' and 'cognitive'—are sometimes used jointly or even interchangeably in fields like epistemology of science (Weisberg & Muldoon, 2009). Because of their relationship, for Humphreys and his purposes in defining epistemic enhancers, these two dimensions of knowledge acquisition—the epistemic and the cognitive—can be grouped

[12] Alvarado (2021a, 2021b) suggests this claim is somewhat misguided. As he mentions in a footnote, novel representational devices such as notations (calculus) or aggregational insights (averages), have indeed enhanced our access to areas of knowledge previously unavailable. Perhaps, hidden in the representational opacity (Alvarado & Humphreys, 2017; Alvarado, 2021a; Burrell, 2016) of neural networks and other similar computational methods in machine learning and statistical analysis we may find a similar enhancement. Furthermore, at least in the mathematical community, genuine questions seem to have arisen about the mathematical novelty of computer-assisted proofs (Hartnett, 2015).

[13] It is worth noting here that there is an interesting question regarding Humphreys' views on the possibility of automated methods as full, albeit artificial, epistemic agents (Humphreys, 2004, p. 6, 2009a, b). If as specified both in the opening chapter of his book *Extending Ourselves* (2004) and in his paper *Network Epistemology* (2009), we can think of certain artifacts as fully autonomous epistemic agents, then perhaps these artificial agents themselves do access a kind of augmentation that is simply inaccessible to humans.

[14] I want to thank the anonymous reviewers that kindly invited clarification on the repetitive use of the term 'epistemic', on what was meant by epistemic and on the distinction between epistemic and cognitive, which now follows in detail.

together. Hence, any tool which enhances either our cognitive functioning or our abilities to conduct complex inquiry qualifies as an epistemic enhancer.

Here, I differ from Humphreys in that my focus is distinctly the latter, i.e., broadly defined epistemic features of an artifact or an agent such as inferential, explanatory and inquisitive abilities and not so much on cognitive processes. As I will argue, there is a sense in which, though related to one another, cognitive factors need not entail epistemic ones. In this sense they can be considered apart from each other. There is a rich and complex literature detailing the philosophical implications of both their possibly distinct nature and/or the relationship of one with the other. However, for our purposes here, it suffices that we distinguish the domain of the cognitive from the domain of the epistemic in a simple manner: by appeal to their object of study. Simply stated, as epistemologist Alvin Goldman observes (1986), the object of study of cognitive science relates more to the architecture and biochemical interactions of the human brain. In this sense, cognitive science can, of course, make important contributions to the broader—more philosophical oriented—field of epistemology (Goldman, 2018), which is roughly construed as the study of the nature and the possible acquisition of knowledge. However, it is widely recognized both in philosophy and in the cognitive sciences that these two fields of study are not the same and that it is not immediately obvious that the former implies the latter or that the latter can be reduced to the former.

More to the point of our discussion in this section, however, the abilities often emphasized by Humphreys are not simply cognitive in the way a neuron's functioning is cognitive. Inferential abilities or mathematical prowess, for example, are simply not identical with the cognitive processes that underlie them (Kim, 1982). Moreover, his focus is on broader knowledge-apt endeavors such as inquiry (2004, p. 6) and this is what makes the enhancement he refers to an epistemic one and not merely cognitive while still acknowledging that oftentimes enhancing the latter also enhances the former.

Taking this into consideration we can also note a distinction between what Humphreys had in mind when he used the term "epistemic enhancers" and other similar terms, such as "cognitive extenders" (Reiner & Nagel, 2017), used by technologists, or other technologies acknowledged as "cognitive artifacts" (Norman, 1991), which are often defined in ways that overlap with parts of the account described above (Viola, 2021).[15] For example, Hernández-Orallo and Vold (2019) refer to what they call *cognitive tools* or *cognitive extenders*. They define cognitive tools as something that can "carry the potential to offer humans entirely new cognitive abilities" (p. 507). While it is unclear what exactly the authors mean by a *new* cognitive ability, they suggest things like expanded memory processes, visual and auditory processing, boosted brain powers, etc. sometimes enabled directly by biological

[15] See Record and Miller (2018) for an example of how in discussions of certain technologies both the concept of 'epistemic technology' and the concept of 'mind extenders' are used in an overlapping manner.

implants powered by AI.[16] Drawing heavily from the theory of extended cognition made popular by Clark and Chalmers (1998), these views acknowledge that cognitive extenders "can not only replace or substitute for functions of the brain, they can also move beyond the brain to enhance our biological functions." Similarly, affective or cognitive artifacts are defined as "those artificial devices that maintain, display, or operate upon information in order to serve a representational function and that affect human cognitive performance." (Pireda, 2020) Hence, in this sense, and prima facie, these terms overlap.

Cognitive extenders however, are not necessarily epistemic technologies. I can take drugs, for example, that heighten neural connections in such a way that I can better emote when stimulated by the bass sounds of electronic music, and thus extend my cognitive functionality, without any aspect of the drug, its biochemical reaction or my use of it having any epistemic dimension to them (Schifano, 2020).

Furthermore, as we can see from Humphreys' examples above—which include empirical practices such as science, complex computational methods such as computer simulations and things like mathematical proofs—the concept of "epistemic enhancers", though related to cognitive extensions is simply a richer term as it refers to a more externalist account of knowledge acquisition as well as to a broader set of practices and instruments associated with it. It is, for example, harder to imagine the Large Hadron Collider (LHC) at CERN as a cognitive extender in Hernandez-Orallo and Vold's view than it is to understand it as an epistemic enhancer according to Humphreys' account.

Similarly, my use of the term 'epistemic' here is meant to refer to the broader category of practices that relate in a direct manner to the acquisition, retention, use and creation of *knowledge*. An epistemic technology that is *designed* as an epistemic enhancer *is* an epistemic technology in this sense. AI is an epistemic technology, as I will argue in the next section, in large part *because it is an epistemic enhancer*. However, the terms are not synonymous and they have non-trivial conceptual and practical differences between them. The role of epistemic technologies in knowledge acquisition, for example, is not as straightforward as either the term 'epistemic enhancer' would suggest or as the concept of a cognitive extender would imply. While the concept of a cognitive extender implies the extension of a cognitive capacity and the term 'epistemic enhancer' may be understood by some as a success term, an epistemic technology—as a technical artifact often instantiated through material implementation—is a term that accepts fallibility. In other words, epistemic technologies may fail at generating knowledge without ceasing to be epistemic technologies. Relatedly, because of this last point, the concept of epistemic technology that I will rely on in the next section is distinct from the concept of epistemic enhancers in that it can account for the fact that there are *epistemically opaque* (Humphreys, 2004, 2009a, 2009b) epistemic technologies. That is, the term allows that some 'blackbox' technologies are capable of producing/furnishing knowledge

[16] While Hernández-Orallo and Vold want to argue that the tools they are referring to are not "merely cognitive prosthetic" (p.508) but that they can give humans novel cognitive capacities, their examples all seem to refer to already existing human cognitive capacities, only enhanced.

about a phenomenon (Duran & Jongsma, 2021) while not necessarily enhancing *our* epistemic abilities (Alvarado, 2021a, 2021b). In fact, some of the intrinsically opaque AI systems that we rely on daily (Alvarado, 2021b; Alvarado & Humphreys, 2017; Burrell, 2016) get in the way of our conventional knowledge-acquisition practices such as explanations (Alvarado, 2022a, 2022b). Yet, as we will see, they may be, in some non-trivial sense, doing some of the epistemic work themselves (Baird, 2004, p. 67; Humphreys, 2004, p. 6; 2009b; Russo, 2022, p. 185).

As we will see below, when it comes to artificial intelligence, the epistemic element is still more profoundly interconnected to the technology itself since it is *not only* the context of deployment, its design intention, or the capacities it is meant to enhance that makes it an epistemic technology, but also the kind of content it manipulates as well as the kinds of operations it performs on such content. This, in turn, serves to further distinguish AI as first and foremost an epistemic technology. While a calculator can be distinguished as a kind of technology from a bulldozer in that it carries out tasks that can be best characterized as epistemic, AI, as I will show, proves to be even *more distinct* in this regard.

## AI as an Epistemic Technology

We can now begin to address the main aim and claim in this paper: understanding AI as first and foremost an epistemic technology. In doing so, two important philosophical distinctions regarding the nature of technical artifacts can help us establish some fundamental considerations and provide guidance as we seek to better capture the unique characteristics of epistemic technologies and as we move forward to understand AI as such.

In addition to material properties, accounting for the functions of objects is often necessary for a robust ontological account of what some things are (Polger, 2013). This is particularly the case when it comes to functionally identifiable objects such as artifacts. Hence, the first important philosophical distinction for our discussion is that artifacts are intentional, functionally identifiable objects—objects in the world which we identify and distinguish primarily in virtue of the functions they perform, e.g., hearts, printers, kidneys, corkscrews, etc. Because, as the list implies, artifacts are not the only functionally identifiable objects in the universe, identifying an artifact solely on the criterion of function is insufficient to tell them apart from other functionally identifiable objects. Therefore, the *intentional* aspect of artifacts is equally important to their proper identification. Symons (2010), for example, emphasizes the intentional nature of artifacts to appropriately distinguish them from things like biological organs, which are functionally identifiable objects as well.[17]

[17] The relationship between intention (intended design, more precisely) and the ontological status of an artifact is contentious. One could easily imagine an artifact whose intended function no longer figures in a future use. However, Symons' distinction holds. All he needs from this claim is that an intention was present in the original design of the artifact to differentiate its artifactual nature from either organs or other *pseudo-artifacts* (Kroes, 2002)—i.e., ready-made natural objects whose properties happen to coincide with an existing human interest or need, e.g., a rock formation in the form of a bowl.

Consequently, a full account of what an artifact is must include details of both what it does (its function) and what it is *intended* to do (its overall intentionally ascribed purpose) (Kroes & Meijers, 2002; Kroes, 2010; Miller, 2021).[18]

Importantly, as I briefly mentioned above, the term epistemic technology—as a term that deals with technical artifacts—is not a success term. Epistemic technologies do not need to provide us with reliable knowledge to be epistemic technologies. When we say that an epistemic technology's primary role is in enabling the creation of knowledge, we should be careful not to give too much epistemic credence to novel epistemic technologies ahead of time, or to expect too much from a technology's involvement in knowledge practices. As artifacts—which are always functionally identifiable and for the most part materially implemented devices (whether this implementation be digital or mechanical)[19]—their identity is able to survive failing at their primary function.

Furthermore, there may be instances of epistemic technologies that qualify as such even when they are *not* deployed in epistemic contexts or in the aid of an epistemic task. For example, a microscope does not cease to be a tool of knowledge when it is used as a doorstep. This is because—as with many other tools and artifacts—microscopes continue to be functionally identifiable even when they fail to carry out the function for which they are identified (Kroes, 2003; Symons, 2010; Alvarado, 2021a, 2021b).[20] As stated in the previous section, the intended design of a technical artifact is as important to what the artifact is and does, even when what it is intended to do ceases or changes. Therefore, accepting AI as an epistemic technology does not mean that we must also accept that it is an epistemically *trustworthy* epistemic technology. In fact, as also briefly mentioned above, we may have to come to terms with the fact that AI is an epistemic technology despite the fact that it can even *hinder* our epistemic abilities, as the issues with epistemic opacity of AI methods discussed below will suggest. In this sense, the term is distinct from both the narrower 'cognitive extender' category and from the richer concept of epistemic enhancers explored above.

Yet another relevant and directly related philosophical distinction is the nuanced point that what something does and what something is used for are not always the exact same thing (Alvarado, 2021b).[21] Use and function are closely related, of course. For example, if an artifact is intended to be used to cut through certain

[18] The intended function of an artifact—even if it does not endure the existence of the object, e.g., the object is no longer used for what it was intended and/or the intention is all but forgotten—also explains the continued functional identification of an artifact beyond fallibility. For example, we can refer to a broken corkscrew as a corkscrew, even when it does not fulfill its function. We can also continue to refer to a permanently-grounded airplane at a museum as an airplane.

[19] Consider, as a contrast to a broken corkscrew, a blueprint or a non-implementable design.

[20] Capturing the nature of error in software-based epistemic technologies may prove to be a non-trivial philosophical issue. For example, Floridi et al. (2015) posit that software as an artifact is the kind of object that cannot malfunction. Unlike a broken hammer, which may still be a hammer despite its inability to perform the function through which it is identified (as a functionally identifiable object), a word processor that fails to process words is not a word processor.

[21] I want to thank an anonymous reviewer whose comments made clear this distinction needed to be explicitly stated.

substances, the artifact cannot be constructed with just any substance, say gelatin. In this sense the intended use of a designed artifact determines what an artifact does (to cut) and what kinds of material implementations will actually allow it to be used for such purposes (solid and sharp materials). Yet, as we can see, each of those considerations do not necessarily refer to the same things (properties, contexts, etc.). I can use a knife to eat, to stab or to open a package. In all those uses what I need the knife to *actually do* is to cut. In fact, the reason I can use it for such purposes is directly related to that more fundamental capacity.[22]

Finally, another important distinction to make is that while what AI does and what it is used for are both important to what AI is, they are nevertheless two different aspects of what AI is. As we will see in this section, what defines AI as an epistemic technology is not just what it is used for—though this will play an important part—or which of our capacities it enhances, but also *its own internal capacities.* Just as we can use a knife for distinct purposes primarily because of its capacities to cut, the capacities of AI are, fundamentally, capacities for data analysis and not, say, moving a robotic arm. While some may want to argue that AI *does* move robotic arms, this is not, strictly speaking, what AI *does* as much as what it is applied to do once very sophisticated non-AI software and hardware interfaces are brought into the equation.

Together, these distinctions will help us buttress the distinction drawn in the previous section between AI as an epistemic technology and other kinds of technologies, but they will also help us further distinguish AI from other epistemic technologies as a *paradigmatic* member of this category of tools. As we will see, other epistemic technologies are not epistemic to the extent that AI is. This is because while they may deal with epistemic content, they do not do so in an epistemic manner; while they may be deployed in epistemic settings, they may not perform epistemic operations, so on and so forth, while AI participates across all three epistemic dimensions mentioned.

## What AI Does

Artificial intelligence can be said to do many things: play chess, successfully diagnose patients, simulate the folding of proteins, perform image/object recognition, guide autonomous vehicles, solve video games, etc. It can also be applied in many domains: finance, robotics, medicine amongst many others. However, although artificial intelligence can be examined through its many distinct applications, strictly speaking, the practical computational methodology referred to as artificial intelligence is first and foremost designed, developed and deployed as a branch of data science, such as the many statistical analytic methods used in the classification, prediction and generation of data (Mitchell, 2019). The rest of the aforementioned

[22] As Alvarado (2021a, 2021b) notes, what an artifact is meant to do, the functions it carries out to achieve this purpose and the teleological context in which it is deployed are all distinct. His example is that of a carburetor: its purpose is to mix fuel and oxygen, it does so through the manipulation of valves, and it enables a combustion engine to run. All three are distinct.

applications are simply that, *applications* of AI and not AI itself (Hengstler et al., 2016 p. 107). This is particularly the case when technologies such as machine learning, deep neural networks, and large language models are involved.

Consider machine learning algorithms, which are at the core of most contemporary AI developments, including the transformer models in generative AI such as large language models. Machine learning algorithms are a methodological subset of artificial intelligence. Although they have recently gained significant notoriety due to their many successful applications, they have nevertheless been around for quite a while (Carbonel et al., 1983). Functionally speaking, they are algorithms designed to analyze massive amounts of data in search for statistically significant patterns. They are also designed to draw inferences about existing relationships between items in a data set, inferences about possible trajectories of organizational trends in the data set, or inferences about structural similarities between data sets or items in distinct data sets. These algorithms are trained, through the analysis of the properties in one data set, to be able to discern properties in other data sets (El Naqa & Murphy, 2015), or to discern similarities between the many items in a data set and a new item outside of the data set. They are also designed to predict whether a previously unknown data set is likely or not to include an item found in the training set. For example, in image recognition tasks, the algorithm may be tasked with quantifying the probability that a given picture contains the image of a dog or not, or to predict whether a given data set is likely to include the image of a dog at all, or to tell whether a given picture is that of a dog or a wolf (Bergstrom & West, 2021). While these applications may seem trivial, foundational work has emerged from them towards image recognition of more complex patterns that usually require humans extensive training to identify, such as those in medical images (Alvarado, 2022a, 2022b).

Similarly, deep neural networks (DNN) function as multilayer information sifters that filter inputs across layers according to either predetermined or 'learned' threshold weights that activate their nodes. Deep neural networks are a subset of machine learning (ML) methodologies. Unlike other more conventional machine learning algorithms which work through the application of linear regression analysis, deep neural networks consist of several layers of non-linear regressions (Samek et al, 2021; Sarle, 1994). Regardless of these technical differences, the tasks that both conventional machine learning and deep neural networks carry out are of the same kind: pattern recognition, statistical analysis and predictive inferences. The point is that, regardless of their many applications, AI methodologies are, at their core, technologies of analysis. Although they may use distinct AI methodology (e.g., deep neural networks, or transformers), this applies to the new generation of generative AI starting from the original '*Netflix challenge*' of matrix completion (Barocas et al., 2017) to more recent developments in large language models (LLMs) that generate text from text prompts or multimodal models that can generate, amongst other things, images or video from text-prompts (Wolfram, 2023).

It may be tempting to think of the many robotic applications of AI, as well as in its applications in graphic design and illustration as a counter point to my position. We use and make AI for other things besides epistemic endeavors, so the argument goes: to make art, to optimize the speed of an engine, etc. One may say, for example, that deep neural networks are deployed in recommender systems solely in the service of entertainment and in a manner far removed from any epistemic enhancement. However, even then, in the realm of applications, the majority of AI applications are simply epistemic. Take their applications in the medical field. Yan et al., (2019, p. 586) offer the following summary concerning the value and applications of AI in the medical field:

> Firstly, AI can help clinicians diagnose disease and optimize treatment processes. After being applied to traditional medical procedures, AI can reduce the rate of misdiagnosis and improve diagnostic efficiency. Then, with the advent of deep learning, AI has the ability to recognize medical images and provide clinicians with more reliable imaging diagnostic information. Thirdly, by using big data analysis (AI can analyze extremely large data sets that are difficult to analyze by traditional data processing methods), AI algorithms can often provide more accurate results for patient prediction. AI can also help support drug research and improve the efficiency of new drug development. (p. 586)

As can be noted, it is only at the end of this list that Yan et al., mention non-epistemic tasks such as “providing patients with high quality medical services”. However, even then, this is only listed as the product of a combination of the analytic prowess of AI with advancements in its applications in robotics and other bureaucratic procedures. Similarly, in a study of 152 deployments of AI, Davenport and Ronanki (2018) found that the categories that best described the uses of AI technology were the following:

- Cognitive Process Automation: Automation of back office administrative and financial activities.
- Cognitive Insights: Detecting data patterns and interpreting them through statistically-based machine learning algorithms.
- Cognitive Engagement: Engaging employees or customers using natural language processing and machine learning. (As cited by Duan et al. 2019)

Other contexts of deployment are similarly telling. For example, Duan et al. (2019) position AI expert systems as historically deployed in the following six tasks: assistant, critic, second opinion, expert, consultant, tutor, and automaton.

As we can see, these kinds of roles are roles with something of an epistemic dimension. A critic, a second opinion, an expert and a tutor are definitely sources of knowledge and sources that signal an epistemic asymmetry between interlocutors. One may even say that these applications entail somewhat of an epistemic authority. What these systems do are not merely cognitive processes; they are not merely computational assignments and they are not merely information retrieval mechanisms. They are epistemic tasks.

While it is generally speaking not good philosophical practice to rely on definitions and nomenclature provided by practitioners to make philosophical claims, at the very least, the descriptions of these uses can give us an intuitive sense of how different the tasks in AI are from mere computation, calculation and memory retrieval.

Besides making use of memory, these technologies also carry out epistemic operations such as complex analysis, pattern recognition, inferential processes, etc. Similarly, debates concerning trust in artificial intelligence and its uses in medicine are almost exclusively focused on the deployment of analytic and inferential tools of this sort (Hengstler et al., 2016; London, 2019; Jongsma & Durán, 2021).

Whichever one of these tasks or specific technology one is referring to when using the term AI, it is clear that what we are tasking the technology to do is centrally related to both its analytic and inferential capacities and prowess. Furthermore, as seen above, the best understanding of artificial intelligence is as a branch of data science (Mitchell, 2019). And data science is a particular set of methodologies and devices that are *specifically* geared towards the enhancement of epistemic abilities: extract patterns, distinguish items and groups of items based on their properties, etc. Hence, following Humphreys, we can see these technologies are evidently epistemic *enhancers*. Yet, they are also something more.[23]

Consider the following: AI is not an epistemic enhancer *merely in virtue* of enabling the acquisition of knowledge alone or because they are *deployed to* carry out epistemic tasks. Rather, as can be drawn from the previous paragraphs, AI and other computational methods —including ML, LLMs and DNNs—are epistemic enhancers *also* because they deal with *epistemic content* such as propositions, symbols, etc. In this respect, they are different not only from an ordinary hammer or a precision grade hammer at the laboratory, but also from perceptual tools such as microscopes, telescopes and other laboratory instruments. While these other instruments are also deployed in an epistemic context, namely inquiry, they do not deal with epistemic content and they do not perform epistemic operations on that content. Rather, some of them deal with tissue, and manipulate it through chemical or biological means. Others deal with molecules and deal with them by performing physical operations on them such as shaking them. On the other hand, AI technologies deal *with* epistemic content such as models, propositions, and 'images'[24] *by* carrying out epistemic operations on it. In this sense we can think of AI manipulating epistemic content such as semantically laden information (data), but we also have to think of the manipulation itself as epistemic in nature given that the operations involve inferences, predictions, and even deliberative decisions.[25]

[23] They may be epistemic agents in their own right as they carry out unequivocally epistemic tasks for us (Baird, 2004; Humphreys, 2004, 2009a, 2009b; Russo, 2022).

[24] It is important to emphasize that strictly speaking these computational artifacts no not deal with images, precisely. Rather they manipulate or engage with a complex web of statistical patterns related to numerical values in pixel information. Hence why some philosophers and technologists deem some of these technologies in image recognition to be epistemically opaque in a specific manner, namely representationally opaque (Burrel, 2016; Alvarado & Humphreys, 2017; Alvarado, 2020, 2022a, 2022b).

[25] Although calling the operations of machine learning methods 'epistemic' in this sense may sound to some as a bit of a stretch, there is a sense in which there is a straightforward distinction between low level computational operations such as the ones carried by a compiler and the higher-level operations carried out by machine learning algorithms. Hence the use of the term 'epistemic' and not 'cognitive'. While fleshing out the relationship and the independence between these two concepts goes beyond the scope and aim of this paper, it can be said, without much controversy, that while epistemic tasks require cognitive processes, cognitive processes are not always necessarily epistemic tasks. The processing of light by plants in photosynthesis may, for example, meet the definition of a cognitive task, according

Therefore, more explicitly stated, we can say that artificial intelligence, as used in specialized fields, is an epistemic technology in several important ways. AI is designed, developed and deployed to:

1. Be *used in an epistemic context* and *for an epistemic purpose,* such as inquiry
2. *Deal with epistemic content*: propositions, models, data. And,
3. *Carry out epistemic operations*, such as statistical analysis, pattern recognition, predictions, and inferences on such epistemic content.

While we have seen several examples of how the first of these conditions are met by AI technologies, it may be useful to offer more detail on the last two points.

Consider the automated retrieval system (ARS) in Santa Clara's University library. This machine helps people extract and store books. It is like a vending machine for library books. In a sense, it enables people to gather knowledge. It can also be said to deal with epistemic content since it handles books, which in turn have words and propositions within them, etc. Hence, it may seem that this artifact meets several of the conditions specified above as an epistemic technology: it is deployed in an epistemic context (a library) and deals with epistemic content, namely books. However, after closer inspection we can see that in its capacity as a retrieval system it only deals with books and partakes in knowledge practices through *physical* operations: storage and retrieval through mechanical means. The ARS is manipulating only the physical aspect of the books and not their content. This is evidently different from what Paul Humphreys had in mind when talking about epistemic enhancers. In this sense, the ARS is more akin to a bulldozer than a calculator. But even more importantly, it is markedly different from what machine learning and other AI technologies do. AI in decision-making systems carries out complex operations on complex content that yields results in highly complex tasks (Bjerring & Busch, 2021).

When it comes to the handling of epistemic content,[26] consider that strictly speaking there is no such thing as raw data (Boyd & Crawford, 2012; Davies & Frank, 2013). Rather, data are often meaningful and well-defined abstractions interpreted as discrete values. Before data are part of a data base, a well-defined inquiry, method, and selected sample must be in place. What counts as a data point, the way that a data point is defined and included in a data set, the weights given to some features

Footnote 25 (continued)
to some philosophers (see Calvo, 2016), but it would be a non-trivial inferential leap, for example, to postulate that because a plant processes information in a dynamic way as a response to its environment, that such processes directly correlate to the generation of knowledge in the plant or by the plant. The epistemic status of plants is an issue far beyond the scope and aims of this paper. The point made here is simply to signal a distinction between the concept 'cognitive' and the concept 'epistemic' with the assumption that cognitive concepts do not necessarily imply any mental states or dispositions such as beliefs, propositions, etc. Although, it is true that 'epistemic' as used here, is more related to knowledge creation and acquisition practices and less related to conventionally understood mental states, it is also the case that I am here using an understanding of 'epistemic' that implies fallibility. That is, I am referring to things as epistemic in nature even when they fail to achieve the function for which they were designed, e.g., the creation, retention, expansion, of knowledge itself.

[26] Thanks to the anonymous reviewer for inviting clarification of this point.

of a data point over others and how that data point is ultimately interpreted within a cohesive system of references and implications, is what makes data useful to start with (Barocas et al., 2017). Furthermore, while AI is constituted by basic computational processes, what makes AI methods further valuable is that they are able to draw form a multiplicity of semantic specifications that go beyond the numeric symbols and values (Bhatt et al., 2020; Studer et al., 2006). They deal with symbols, yes. But they also deal with compounded signifiers such as words, and word couples and sentences and thematic references therein (Wolfram, 2023).

As we can see, we can further differentiate AI as an epistemic technology even from other epistemic technologies. Whatever *epistemic content* may be, data as a strongly curated—often theoretically informed and semantically rich—type of content qualifies as such. That is, unlike a blade of grass processing some chemical substance, data is epistemic content.

Similarly, whatever *epistemic operations* may be, inferences, predictions and analysis qualify as such in virtue of the fact that they are a distinct—richer—kind of operations, more closely aligned with knowledge acquisition and creation practices than the physical operations performed by robotic arms, or the optical operations performed by conventional telescopes. Whatever an argument is, the inferential principles involved in its processing and assessment as either valid or coherent are more akin to knowledge-related practices than the pneumatic processes of a jackhammer. Furthermore, the operations that AI technologies carry out are also distinct even from more basic computational operations such as the binary processes of compilers (Wolfram, 2023) or the kind of cognitive processes that can be ascribed to plants (Calvo, 2016). Hence, processing epistemic content through epistemic operations cannot be simply equated/reduced to carrying out simpler cognitive or even computational tasks.

This is a distinguishing feature that can also apply when comparing AI to cognitive extenders, tools or artifacts mentioned in the previous section. Consider, for example that if we look at the context in which an aspirin is deployed, one can clearly see that it is not, strictly speaking, an epistemic context.[27] One could

[27] London makes the following analogy concerning the acceptance of both modern pharmaceuticals and opaque AI technologies: "modern clinicians prescribed aspirin as an analgesic for nearly a century without understanding the mechanism through which it works. Lithium has been used as a mood stabilizer for half a century, yet why it works remains uncertain." (London, 2019 p. 17). This argumentative strategy works at a rhetorical level. As a reductio at absurdum, this argument pushes us into a corner because in ordinary settings most of us would not want to condemn widespread medical practices as undesirable. However, the soundness of the premises in the argument—mainly that we accept opaque and associationist methodology from medical practitioners without any significant reservation—depends on who are the interlocutors and what is the given situation. When a medical practitioner is talking to a peer as an epistemic source, say as when a doctor is consulting with a radiologist, opaque and merely associationist reasoning will not be as acceptable as when a doctor recommends a treatment to a patient. In fact, given the modern rejection of historical medical paternalism (Millar, 2015), many informed patients would not accept opacity or arbitrary associationism from their care providers unless in the context of a medical crisis. Although views that minimize the importance of epistemic opacity often operate under an assumption of crisis (e.g., by stating that these technologies could save lives) to ethically motivate and practically justify our reliance on opaque technologies, it is not immediately obvious that such a strategy is genuinely informative in the long term or as a default position concerning the world or the human condition. Most humans, for most of the time, in most contexts will not be in such a crisis mode when we design, develop, deploy or interact with these technologies.

object here that *some* pharmaceutical interventions are indeed deployed with epistemic aims and in epistemic contexts. One example of this is the use of Adderall for academic performance. It has been shown that students at various stages of their education use Adderall in order to improve their focus, retain and retrieve information more easily, and succeed academically in tests and other endeavors: homework, writing assignments, etc. (Girer et al., 2011; Stoltz, 2012; Varga, 2012; Kiernan et al., 2016). These could all be understood as epistemic tasks and indicate a broader epistemic context: the optimization of the acquisition of knowledge. Hence, we can clearly say that at least some pharmaceuticals are designed and deployed in ways and for purposes that could make them qualify surely as cognitive extenders, perhaps even as epistemic enhancers. However, it is important to note that even in cases such as Adderall, the operations expected of the artifact are not epistemic operations. Similarly, the content with which the artifact (the chemical compound of Adderall) interacts *is not* epistemic content. Even if we can say that it is used for epistemic purposes, its epistemic involvement and the epistemic context in which it is deployed are even less clearly epistemic than the involvement of the Automated Retrieval System mentioned above. While one could say that the ARS is deployed in an epistemic context (retrieving books which will help in the acquisition of knowledge) *and* deals with epistemic content (books), the latter cannot be said of Adderall. It does not deal with epistemic content at all. Its interactions are biochemical. Similar outcomes can be expected when comparing AI to other technologies such as autonomous vehicles, IT systems in general, etc.

In short, capturing the true nature of an artifact is essential to understanding our relationship with it. Essentially an artifact is what it does or is designed to do. AI and its associated methods are exclusively and particularly designed for and deployed in epistemic tasks. Hence AI technologies are epistemic technologies. Furthermore, they carry out these epistemic tasks by manipulating epistemic content and by doing so through epistemic operations. Therefore, they are epistemic technologies not just in ways that other technologies are not, but also in ways that even other epistemic enhancers are not. They are, paradigmatically and exclusively an epistemic technology.

## Implications

Understanding AI as an epistemic technology has important implications for central problems in the philosophy, epistemology and ethics of AI. In particular, it has direct implications for debates concerning explainable AI and its relationship to epistemic authority, epistemic opacity (Alvarado, 2020; Boge, 2022), epistemic trust (Alvarado, 2022a, b), and epistemic injustice (Symons & Alvarado, 2022). In this section I will provide a brief overview of some of the key implications of our lengthy discussion in these important debates.

Let us first consider the interrelated debates regarding the challenges of explainability, transparency, epistemic opacity and trust that AI faces. While we can accept the use of a hammer without knowing in detail how or why it works because we do not need to assess its epistemic soundness, it will be increasingly difficult to justify the epistemic trust we allocate to an epistemic technology if we do not have access to the epistemically relevant elements of its functioning, as the problem of epistemic opacity suggests (Alvarado, 2020, 2021a; Humphreys, 2004, 2009a, 2009b). When it comes to epistemic technologies particularly—technologies that are exclusively designed and deployed for the manipulation of epistemic content through epistemic operations in the service of an epistemic context—such as AI—the need to involve challenging epistemic and methodological scrutiny is even more immediately pressing. This is because they are intrinsically designed and deployed for the creation, exploration, evaluating *and* transmission of epistemic claims, through epistemic means in epistemic contexts. This is the case even when we consider recent generative AI methods such as large language models which build their outputs on predictive processes. Assessing the likelihood or the accuracy of such outputs will depend on how well we think the model analyzed the training data. Hence, epistemic reliability of the analytical processes will be relevant, albeit not determinant,[28] to the trust we can allocate to AI generated text, search, etc.

Some of the main ways in which we assess the value of an epistemic claim or an epistemic source is through an analysis of the reasons that justify them as epistemic claims and as epistemic sources (consistency, soundness, accuracy, verisimilitude, etc.). The undertaking of such an analysis by an interlocutor requires the possibility of reason-giving endeavors such as access to the epistemically relevant elements of the process, explanations about the processes, justifications or argumentation from a counterpart. In other words, if these technologies are going to function as information providers, they have to be particularly responsive to questions regarding their epistemic status. If we are to assess and hence to adopt these technologies as epistemically reliable for epistemic tasks, for example, we ought to do it through sound epistemic means (Durán & Formanek, 2018).

Following McKnight et al.'s (2011) work on the distinct kinds of trust apt for distinct kinds of technologies, Alvarado (2022a, 2022b) suggests that we should only seek to trust epistemic technologies in their capacities related to knowledge acquisition. That is, we ought to only allocate *epistemic trust*—trust allocated solely in virtue of the capacities (or potential) of the recipient to be a conveyer of knowledge (Wilholt, 2013) or trust allocated solely for epistemic reasons (McCraw, 2015)—to epistemic technologies.[29] Furthermore, according to Simion (2018, 2019), the proper deployment of epistemic concepts such as epistemic trust, ought to be guided *only* by epistemic considerations such as epistemic norms.[30] If this is so, then sound epistemic means are only properly justified as such by appropriate epistemic reasons,

[28] As Symons and Alvarado (2019) point out, the transferring of epistemic warrants from one process to another is a non-trivial issue in the epistemology of instruments deployed in inquiry. Having good reason to trust an underlying process or method (or person), does not automatically grant reason to trust a novel technology derived from it.

[29] Note that what Alvarado is saying here is not that we *should* trust epistemic technologies to provide reliable epistemic content. Rather, what Alvarado is suggesting is that *if* we are to trust epistemic technologies, we should only try to allocate epistemic trust and no other kind of trust (e.g., interpersonal trust or trust in capacities that are not epistemic). Whether or not one can epistemically trust AI is still an open question according to him (2022a).

[30] According to Simion, an epistemic norm is one that is closely related to an epistemic value (Simion, 2019).

concepts and warrants (Dretske, 2000; Symons & Alvarado, 2019). These epistemic warrants will only emerge through the ability to assess whether the trust we allocate to such technologies is well-grounded (appropriately allocated). If so, then explainability—the ability of a process to elucidate the ways in which it operates—will play a crucial role in sanctioning such technologies. Understanding that AI is primarily an epistemic technology allows us to understand the epistemic requirements it needs to fulfil and hence the serious challenges its opaque nature represents to our ability to sanction it in this same capacity.

In other words, if we accept that AI is an epistemic technology—and we accept that epistemic technologies deployed in epistemic contexts for epistemic purposes have more rigorous epistemic expectations (Symons & Alvarado, 2019; Alvarado, 2020, 2021b, 2022a; Simon, 2010)—then the epistemic challenges its opacity represents becomes the more salient and serious. The same applies to claims that explicitly or implicitly entail any sort of epistemic authority on the part of AI technologies (Simion, 2018). Their validity as such is also exclusively determined via epistemic norms and concepts. For example, if someone claims to be an authority on particle physics, we would not take a fact about their height or their moral standing as appropriate reasons to justify their claim, as these are not epistemic reasons (Fricker, 2007, 2017)—they do not speak to the truth, falsity or legitimacy of the claim to epistemic authority (Dretske, 2000). Similarly, accepting an AI technology as a decision-maker or as a consultant in an expert domain, both of which are claims to authority and claims made from a position of authority (as an expert or decider) will require appropriate justificatory efforts, some of which include the ability for AI to explain itself. If someone in a position of authority is not able to explain themselves, provide reasons, after a decision affecting others, they may be occupying that position illegitimately (Lazar, forthcoming). Note that we would think this of the particle physics expert as well: if they cannot provide reasons that directly justify their claim to authority as a particle physics expert, or to a particular procedure with which they may have arrived at a domain-relevant conclusion, we would not confer them a position in a highly sensitive laboratory, for example. Similarly, if AI technologies are essentially and representationally opaque, as most relevant literature suggests (Burrell, 2016; Alvarado & Humphreys, 2017; London, 2019; Alvarado, 2020, 2021a, 2022b; Durán & Jongsma, 2021), we cannot have epistemic access to its *epistemically relevant* features. In other words, explainability is simply not possible. If this is so, we cannot allocate epistemic trust. And if epistemic trust is the only trust we ought to allocate to epistemic technologies, as Alvarado (2022a, 2022b) suggests, then we simply cannot yet trust AI. In fact, we may be dealing with a fundamentally challenging technology, one that simply cannot be epistemically trusted.

Furthermore, considering the point in the previous paragraphs which suggests that explainability and the availability of appropriate epistemic warrants are essential to the adoption of epistemic technologies, it will mean that appeals such as London's (2019), Ratti and Graves (2022), or Duran and Jongsma's (2021) to circumvent the challenge of opacity to accept and ratify AI technologies in practical domains such as medicine, will not work.[31] As we saw with the example of the hammer, while it *may* be acceptable to adopt many other technologies despite their epistemic opacity, epistemically justifying the acceptance of epistemic technologies that are epistemically opaque may prove to be significantly harder (Symons & Alvarado, 2019; Alvarado, 2020, 2021a, 2021b, 2022). Appealing to prudential or practical considerations for their sanctioning will simply not suffice. Similarly, appealing to superficial similarities in the context of use of other kinds of technologies will prove to be misguided.

Consider the following: some recent arguments invite us to accept and incorporate AI into practical domains on the basis of their similarities to other methods or practices. Alex London (2019), for example, finds it problematic that concerns about opacity are prioritized in discourse about medical AI but are absent concerning our acceptance of other equally opaque things in equally sensitive contexts, e.g., pharmaceutical interventions of the kind provided by an aspirin or medical practitioners. We trust medical practitioners, London argues, *despite* the fact that their thought processes are often opaque (perhaps even to themselves), despite the fact that their inferential processes for diagnosis and treatment are often atheoretical, and despite the fact that they also often rely on purely associationist reasoning. We trust aspirin, London suggests, in a similar way despite similar epistemic obstacles. If this is so, why would we mistrust medical AI on the basis of sharing these same features? He asks. In a similar vein, other arguments point out that there is simply *no good reason* to not to trust opaque AI technologies (Mazurowski, 2019). This view has been echoed and expanded on by others such as Durán and Jongsma (2021), Ratti and Graves (2022) and others. Outside of the considerations related to medical AI, similar approaches have emerged concerning our trust and reliance in artificial intelligence in a myriad of socially consequential contexts (Rossi, 2018; Sethumadhavan, 2019; Ryan, 2020; Knowles & Richards, 2021; Jöhnk et al., 2021). Similar argumentative strategies exist drawing not on shared shortcomings but on comparable or superior accomplishment metrics vis-à-vis other humans or other technologies (Hinton, 2016; Jha & Topol, 2016; Chockley & Emanuel, 2016; Lombardo et al., 2020; Mazurowski, 2019;).[32]

In short, so these arguments go, we should trust/rely on/use/replace humans with AI because we trust other things with similar success records or similar shortcomings and challenges in similar settings. What these views share in common is an argumentative strategy that compares artificial intelligence to other kinds of technologies or other kinds of agents (e.g., cars, human experts, etc.). Hengstler et al. (2016), for example, argue for the adoption of AI from the perspective of considerations related to the development of autonomous vehicles. Similarly, Danks, (2019) makes reference to

[31] While some of these authors use different examples (and Duran and Jongsma's work is a broader take on the issues of reliability and opacity), these papers all refer to medical settings and all suggest that pragmatic considerations such as appeals to success records and accurate outputs may suffice to circumvent worries about opacity (see Duran & Jongsma, 2021, p. 332). In this sense and in reference to the problem of opacity they seem to imply a similar argumentative strategy and a similar solution to the problem of opacity.

[32] See Alvarado, 2022a, 2022b for a critical review of these arguments.

the trust he allocates to his car every morning when trying to understand trust in AI; Cho et al. (2019) allude to the trust we allocate to IT and basic computer-based systems in general and include AI in their analysis. As we saw in the previous paragraph, London and the other authors mentioned above allude to pharmaceuticals such as aspirin or lithium in order to make their case for AI in medicine and other decision-making contexts.[33] In and of themselves, these comparisons are not unreasonable. After all, the philosophy of technology is rich with epistemological and ethical debates that can provide much insight to our current technological condition. However, as we saw in our lengthy discussion in previous sections of this paper, not all technologies are the same and depending on what kind of technology we are talking about, different normative concepts may apply to them that do not apply to others (Lankton et al., 2015). If so, then perhaps it is worth asking whether we may be dealing with false or misguided analogies. While we trust autonomous vehicles to drive safely from one point to another and pharmaceuticals to chemically intervene in our bodies, what we seek to trust AI technologies with does not seem to be similar in any relevant way. We seek to trust them to *manipulate epistemic content* such as visual input and propositional structures; we seek to trust them to *carry out epistemic operations* on this content such as analysis, inferences and predictions; and finally, we seek to trust AI technologies as aids *in epistemic contexts* such as scientific inquiry, business decisions, etc. In contrast we do not trust aspirin or cars to provide information, to predict statistical trends or to make decisions for us. Furthermore, when the comparison shifts to agents such as other humans, it is clear that although people can sometimes be trusted in this respect, not all interpersonal trust is allocated in virtue of the epistemic capacities of the recipient, even if some epistemic capacities are assumed.

Hence, acknowledging the status of AI as an epistemic technology radically modifies the landscape of appropriate analogies that can be deployed in arguments towards its inclusion and our trust in its processes. This acknowledgement also forces us to also consider that the advent of AI represents the introduction of a novel artifact into the practice of formal inquiry. As Symons and Alvarado (2019) as well as Alvarado (2021b) argue in detail, accepting a novel artifact into formal inquiry is simply not like accepting the input of other experts, it is not like sanctioning the people that build such technologies, and it is not like accepting the methods by which such technologies are built either.[34] Hence, accepting AI in virtue of the reasons or the ways that we accept other epistemic technologies, other experts or other people is simply not warranted.[35]

Finally, another implication in which the view of AI as an epistemic technology can have explanatory value is in understanding the kinds of harms that are the most immediately related to this technology as also being epistemic in nature.

[33] See Ferrario et al. (2020), Ferrario and Loi (2021a), and Ferrario et al. (2021b) for similar argumentative strategies.

[34] By the term ‘accepting’, Symons and Alvarado seem to mean something akin to justifying the reliance/trust on, believing the result of, permitting it to count as, admissible. For the sake of the argument, here, I do so too.

[35] See Symons and Alvarado, 2019 to see a thorough account of the distinct epistemic warrants at play in each of those cases.

For example, understanding AI as an epistemic technology could explain how and why epistemic harms are so important in the ethical assessment of AI, as Symons and Alvarado recently suggested (Symons & Alvarado, 2022). While concerns about the pernicious effects of artificial intelligence are usually fleshed out in terms of the social and financial harms that accompany the deployment of such technologies, after our discussion, we can see that these are more appropriately attributed to the applications of AI and not to AI itself. When it comes to AI and its epistemic functions, it becomes clearer that perhaps other kinds of harms that we are likely to cause with these epistemic functions include epistemic harms, such as the unjust diminution in the epistemic standing of others or in the unjust diminution of the epistemic capacities and entitlements of others (Fricker, 2007, 2017). Importantly, such discriminatory harms, though often related, are not necessarily distributive harms.

Understanding AI technologies as epistemic technologies can provide both disruptive and enriching insights into important debates as we struggle to understand our relationship to such technologies. It can inform the reasons we trust these technologies, the tasks we trust them with, and even the role we allow them to play in our politics. Hence, such an understanding has implications in important debates at the philosophical, personal and institutional level. However, although our discussion elucidated the sparks of the imminent fire that this novel understanding of AI implies, this is a topic whose depth would be best explored at length in a book.

## Conclusion

In this paper I have shown that artificial intelligence and associated methods such as machine learning are first and foremost epistemic technologies. This is not only because they are deployed in epistemic contexts such as inquiry, but also because the content they manipulate is epistemic in nature and because the manipulations they carry out are of an epistemic kind. Hence, in the realm of epistemic enhancers, AI technologies are epistemic across at least three different dimensions. This fact distinguished them from many other technologies and makes it so that arguments by analogy to other technologies that seek to sanction their uncritical adoption are at best misguided and at worst disingenuous. Furthermore, understanding AI as an epistemic technology will have significant repercussions in other important debates concerning their epistemic and ethical status. Issues to do with explainability, epistemic trust and epistemic injustice can be best explained through the framework provided in this paper. The aim of this paper was to show that artificial intelligence is first and foremost an epistemic technology, understanding it as such is the first step towards understanding our relationship to it, how we want to involve it in our lives and how we can draft policy surrounding it.

**Acknowledgements** Arzu Formánek, Nico Formánek, John Symons, Gabriela Arriagada Bruneau, the Ethics of Technology Early-career Group (ETEG)

**Author contributions** RA is the sole author of this manuscript and carried out all the associated tasks in producing it.

## References


Alvarado, R. (2021a). Explaining epistemic opacity. (Preprint here: http://philsci-archive.pitt.edu/19384/)

Alvarado, R. (2020). Opacity, big data, Artificial Intelligence and machine learning in democratic processes. In K. Macnish (Ed.), *Big data and democracy.* Edinburgh University Press.

Alvarado, R. (2021). Computer simulations as scientific instruments. *Foundations of Science, 27*, 1–23.

Alvarado, R. (2022a). What kind of trust does AI deserve, if any? *AI and Ethics*. https://doi.org/10.1007/s43681-022-00224-x

Alvarado, R. (2022b). Should we replace radiologists with deep learning? Pigeons, error and trust in medical AI. *Bioethics, 36*(2), 121–133.

Alvarado, R., & Humphreys, P. (2017). Big data, thick mediation, and representational opacity. *New Literary History, 48*(4), 729–749.

Anthony, C. (2018). To question or accept? How status differences influence responses to new epistemic technologies in knowledge work. *Academy of Management Review, 43*(4), 661–679.

Barocas, S., Hardt, M., & Narayanan, A. (2017). Fairness in machine learning. *Nips Tutorial, 1*, 2017.

Baier, A. C. (1985). What do women want in a moral theory? *Noûs, 19*(1).

Baird, D. (2004). *Thing knowledge: A philosophy of scientific instruments*. University of California Press.

Becker, P., & Clark, W. (Eds.) (2001). *Little tools of knowledge: Historical essays on academic and bureaucratic practices*. University of Michigan Press.

Bergstrom, C. T., & West, J. D. (2021). *Calling bullshit: The art of skepticism in a data-driven world.* Random House Trade Paperbacks.

Bhatt, S., Sheth, A., Shalin, V., & Zhao, J. (2020). Knowledge graph semantic enhancement of input data for improving AI. *IEEE Internet Computing, 24*(2), 66–72.

Bjerring, J. C., & Busch, J. (2021). Artificial Intelligence and patient-centered decision-making. *Philosophy & Technology, 34*(2), 349–371.

Boge, F. J. (2022). Two dimensions of opacity and the deep learning predicament. *Minds and Machines, 32*(1), 43–75.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662–679.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 2053951715622512.

Calvo, P. (2016). The philosophy of plant neurobiology: A manifesto. *Synthese, 193*(5), 1323–1343.

Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (Eds.) (1983). An overview of machine learning. In *Machine learning* (pp. 3–23). Springer.

Cho, J. H., Xu, S., Hurley, P. M., Mackay, M., Benjamin, T., & Beaumont, M. (2019). Stram: Measuring the trustworthiness of computer-based systems. *ACM Computing Surveys (CSUR), 51*(6), 1–47.

Chockley, K., & Emanuel, E. (2016). The end of radiology? Three threats to the future practice of radiology. *Journal of the American College of Radiology, 13*(12), 1415–1420.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.

Danks, D. (2019). The value of trustworthy AI. In *Proceedings of the 2019 AAAI/ACM conference on AI, ethics, and society* (pp. 521–522).

Daston, L. (2012). The sciences of the archive. *Osiris, 27*(1), 156–187.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, *96*(1), 108–116.

Davies, T., & Frank, M. (2013). 'There's no such thing as raw data' exploring the socio-technical life of a government dataset. In *Proceedings of the 5th annual ACM web science conference* (pp. 75–78).

Dougherty, D., & Dunne, D. D. (2012). Digital science and knowledge boundaries in complex innovation. *Organization Science, 23*(5), 1467–1484.

Dretske, F. (2000). Entitlement: Epistemic rights without epistemic duties? *Philosophy and Phenomenological Research, 60*(3), 591–606.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial Intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71.
Durán, J. M., & Formanek, N. (2018). Grounds for trust: Essential epistemic opacity and computational reliabilism. *Minds and Machines, 28*, 645–666.
Durán, J. M., & Jongsma, K. R. (2021). Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *Journal of Medical Ethics, 47*(5), 329–335.
El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In I. El Naqa, R. Li. & M. J. Murphy (Eds.), *Machine learning in radiation oncology* (pp. 3–11). Springer.
Ferrario, A., & Loi, M. (2021). The meaning of "Explainability fosters trust in AI". *Available at SSRN 3916396*.
Ferrario, A., Loi, M., & Viganò, E. (2020). In AI we trust incrementally: A multi-layer model of trust to analyze human-artificial intelligence interactions. *Philosophy & Technology, 33*(3), 523–539.
Ferrario, A., Loi, M., & Viganò, E. (2021). Trust does not need to be human: It is possible to trust medical AI. *Journal of Medical Ethics, 47*(6), 437–438.
Floridi, L., Fresco, N., & Primiero, G. (2015). On malfunctioning software. *Synthese, 192*, 1199–1220.
Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
Fricker, M. (2017). Evolving concepts of epistemic injustice. In *The Routledge handbook of epistemic injustice* (pp. 53–60). Routledge.
Friedrich, M. (2018). *The birth of the archive: A history of knowledge*. University of Michigan Press.
Girer, N., Sasu, N., Ayoola, P., & Fagan, J. M. (2011). *Adderall usage among college students*.
Goldman, A. I. (1986). *Epistemology and cognition*. Harvard University Press.
Goldman, A. I. (2018). *Philosophical applications of cognitive science*. Routledge.
Golinski, J. (1994). Precision instruments and the demonstrative order of proof in Lavoisier's chemistry. *Osiris, 9*, 30–47.
Hakkarainen, K., Engeström, R., Paavola, S., Pohjola, P., & Honkela, T. (2009). Knowledge practices, epistemic technologies, and pragmatic web. In *I-Semantics* (pp. 683–694).
Hartnett, K. (2015). Will computers redefine the roots of math? *Quanta Magazine*, *19*.
Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change, 105*, 105–120.
Hernández-Orallo, J., & Vold, K. (2019). AI extenders: The ethical and societal implications of humans cognitively extended by AI. In *Proceedings of the 2019 AAAI/ACM conference on AI, ethics, and society* (pp. 507–513).
Hinton, G. (2016). Machine learning and the market for intelligence. In *Proceedings of the machine learning and marketing intelligence conference.*
Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford University Press.
Humphreys, P. (2009a). The philosophical novelty of computer simulation methods. *Synthese, 169*(3), 615–626.
Humphreys, P. (2009b). Network epistemology. *Episteme, 6*(2), 221–229.
Jha, S., & Topol, E. J. (2016). Adapting to Artificial Intelligence: Radiologists and pathologists as information specialists. *JAMA, 316*(22), 2353–2354.
Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes—an interview study of organizational AI readiness factors. *Business & Information Systems Engineering, 63*(1), 5–20.
Kiernan, J., Reid, C., & Zavos, P. (2016). Pulling an all-nighter: Current trends of college students' use of adderall. *MOJ Womens Health, 3*(1), 167–170.
Kim, J. (1982). Psychophysical supervenience. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 41*(1), 51–70.
Knowles, B., & Richards, J. T. (2021). The sanction of authority: Promoting public trust in AI. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 262–271).
Kroes, P. A. (2003). Physics, experiments, and the concept of nature. In *The philosophy of scientific experimentation* (pp. 68–86). University of Pittsburgh Press.
Kroes, P. (2010). Engineering and the dual nature of technical artefacts. *Cambridge Journal of Economics, 34*(1), 51–62.
Kroes, P., & Meijers, A. (2002). *The dual nature of technical artifacts-presentation of a new research programme*. University Library.

Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems, 16*(10), 1.
Lazar, S. (forthcoming) Legitimacy, authority, and the political value of explanations. *To be presented as Keynote for Oxford Studies in Political Philosophy*. https://philpapers.org/archive/LAZLAA-2.pdf
Lombardo, P., Boehm, I., & Nairz, K. (2020). RadioComics–Santa Claus and the future of radiology. *European Journal of Radiology, 122*.
London, A. J. (2019). Artificial Intelligence and black-box medical decisions: Accuracy versus explainability. *Hastings Center Report, 49*(1), 15–21.
Mazurowski, M. A. (2019). Artificial Intelligence may cause a significant disruption to the radiology workforce. *Journal of the American College of Radiology, 16*(8), 1077–1082.
McCraw, B. W. (2015). The nature of epistemic trust. *Social Epistemology, 29*(4), 413–430.
Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS), 2*(2), 1–25.
Millar, J. (2015). Technology as moral proxy: Autonomy and paternalism by design. *IEEE Technology and Society Magazine, 34*(2), 47–55.
Miller, B. (2021). Is technology value-neutral? *Science, Technology, & Human Values, 46*(1), 53–80.
Miller, B., & Record, I. (2013). Justified belief in a digital age: On the epistemic implications of secret Internet technologies. *Episteme, 10*(2), 117–134.
Miller, B., & Record, I. (2017). Responsible epistemic technologies: A social-epistemological analysis of autocompleted web search. *New Media & Society, 19*(12), 1945–1963.
Mitchell, M. (2019). *Artificial Intelligence: A guide for thinking humans*. Farrar.
Morrison, M. (2015). *Reconstructing reality: Models, mathematics, and simulations*. Oxford University Press.
Norman, D. A. (1991). Cognitive artifacts. *Designing Interaction: Psychology at the Human-Computer Interface, 1*(1), 17–38.
Páez, A. (2009). Artificial explanations: The epistemological interpretation of explanation in AI. *Synthese, 170*(1), 131–146.
Páez, A. (2019). The pragmatic turn in explainable artificial intelligence (XAI). *Minds and Machines, 29*(3), 441–459.
Pincock, C. (2011). *Mathematics and scientific representation*. Oxford University Press.
Piredda, G. (2020). What is an affective artifact? A further development in situated affectivity. *Phenomenology and the Cognitive Sciences, 19*, 549–567.
Polger, T. W. (2013). Physicalism and Moorean supervenience. *Analytic Philosophy*, *54*(1), 72–92.
Ratti, E., & Graves, M. (2022). Explainable machine learning practices: Opening another black box for reliable medical AI. *AI and Ethics, 2*(4), 1–14.
Ratto, M. (2012). CSE as epistemic technologies: Computer modeling and disciplinary difference in the humanities. In Wes Sharrock & J. Leng (Eds.), *Handbook of research on computational science and engineering theory and practice* (pp. 567–586). IGI Global.
Record, I., & Miller, B. (2018). Taking iPhone seriously: Epistemic technologies and the extended mind. In Duncan Pritchard (Ed.), *Extended epistemology.* Oxford University Press.
Reiner, P. B., & Nagel, S. K. (2017). Technologies of the extended mind defining the issues. In Judy Illes (Ed.), *Neuroethics: Anticipating the future* (pp. 108–122). Oxford University Press.
Rossi, F. (2018). Building trust in Artificial Intelligence. *Journal of International Affairs, 72*(1), 127–134.
Russo, F. (2022). *Techno-scientific practices: An informational approach*. Rowman & Littlefield.
Ryan, M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability. *Science and Engineering Ethics, 26*(5), 2749–2767.
Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE, 109*(3), 247.
Sarle, W. S. (1994). Neural networks and statistical models. In *Proceedings of the nineteenth annual SAS users group international conference.*
Schifano, F. (2020). Coming off prescribed psychotropic medications: Insights from their use as recreational drugs. *Psychotherapy and Psychosomatics, 89*(5), 274–282.
Sethumadhavan, A. (2019). Trust in Artificial Intelligence. *Ergonomics in Design, 27*(2), 34–34.
Simion, M. (2018). The ‘should’ in conceptual engineering. *Inquiry, 61*(8), 914–928.
Simion, M. (2019). Conceptual engineering for epistemic norms. *Inquiry*. https://doi.org/10.1080/0020174X.2018.1562373

Simon, J. (2010). The entanglement of trust and knowledge on the Web. *Ethics and Information Technology*, *12*, 343–355.
Stolz, S. (2012). Adderall abuse: Regulating the academic steroid. *J.L. & Educ., 41*, 585.
Studer, R., Ankolekar, A., Hitzler, P., & Sure, Y. (2006). A semantic future for AI. *IEEE Intelligent Systems, 21*(4), 8–9.
Sullivan, E., & Khalifa, K. (2019). Idealizations and understanding: Much ado about nothing? *Australasian Journal of Philosophy., 97*(4), 673–689.
Symons, J. (2010). The individuality of artifacts and organisms. *History and Philosophy of the Life Sciences, 32*, 233–246.
Symons, J., & Alvarado, R. (2019). Epistemic entitlements and the practice of computer simulation. *Minds and Machines, 29*(1), 37–60.
Symons, J., & Alvarado, R. (2022). Epistemic injustice and data science technologies. *Synthese, 200*(2), 1–26.
Symons, J., & Horner, J. (2014). Software intensive science. *Philosophy & Technology*, *27*, 461–477.
Van Helden, A. (1994). Telescopes and authority from Galileo to Cassini. *Osiris, 9*, 8–29.
Van Helden, A., & Hankins, T. L. (1994). Introduction: Instruments in the history of science. *Osiris, 9*, 1–6.
Varga, M. D. (2012). Adderall abuse on college campuses: A comprehensive literature review. *Journal of Evidence-Based Social Work, 9*(3), 293–313.
Viola, M. (2021). Three varieties of affective artifacts: Feeling, evaluative and motivational artifacts. *Phenomenology and Mind, 20*, 228–241.
Weisberg, M., & Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science, 76*(2), 225–252.
Wilholt, T. (2013). Epistemic trust in science. *The British Journal for the Philosophy of Science, 64*(2), 233–253.
Wolfram, S. (2023). What is ChatGPT doing… and why does it work. *Stephen Wolfram: Writings*.
Yan, Y., Zhang, J. W., Zang, G. Y., & Pu, J. (2019). The primary use of Artificial Intelligence in cardiovascular diseases: What kind of potential role does Artificial Intelligence play in future medicine? *Journal of Geriatric Cardiology: JGC, 16*(8), 585.