**AI as an Epistemic Technology**

**Abstract:** In this paper I argue that Artificial intelligence and methods associated with it, such as machine learning, are first and foremost epistemic technologies—technologies which are exclusively and intrinsically designed to enhance our capacities as knowers. If this is so, then our relationship towards them cannot be immediately drawn from our analysis of our relationship to just any other technologies and non-trivial philosophical efforts will be in need to truly grasp the nature of such relationship. As we will see, understanding AI as an epistemic technology will also have significant implications for important debates in the epistemology and ethics of artificial intelligence. Particularly those debates pertaining to explainability, opacity and even epistemic harms related to data science methods.

**Keywords:** AI, technology, epistemic

**AI as an Epistemic Technology (Unrevised Preprint)**

1. **Introduction**

Recently, a series of arguments have emerged about the ethical and epistemic implications of artificial intelligence technologies that either explicitly defend their use and our trust in them (London, 2019) or that find no good reason to not use them or trust them (Mazurowski, 2019). Very often these argumentative strategies rely on analogies to other ‘similar’ technology to make their case (Alvarado, forthcoming). Alex London (2019), for example, argued that skepticism towards the use of medical AI is undermotivated. He finds it mysterious that concerns about opacity seem to plague discourse in AI (Burrell, 2016; Alvarado, 2022) but are absent concerning our acceptance of other equally opaque devices such as some pharmaceuticals whose detailed inner workings we don’t have access to. In his view, if epistemic opacity is not a deal-breaker issue in the context of pharmaceutical interventions of the kind provided by an aspirin, then it should not be a deal-braker concerning the use of AI technologies in clinical contexts either. He applies a similar reasoning to the way we trust medical practitioner and medical practice as a whole. If we trust medical practice and practitioners despite them being opaque, atheoretical and often relying on associationist reasoning, why would we not trust medical AI which shares these same features? He asks. This is particularly the case, he argues, when someone’s well-being is at stake. This view has been echoed and expanded on by others such as Durán and Jongsma (2021), Ratti and Graves, (2022). Outside of the field of medical AI, similar approaches have emerged concerning our trust and reliance in artificial intelligence in a myriad of socially consequential contexts (Rossi, 2018; Knowles and Richards, 2021; Jöhnk et al., 2021).

What these views share in common is an argumentative strategy that compares artificial intelligence to other technologies. 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 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 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.[[1]](#footnote-1) In and of itself, this comparison is 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, and, of course, AI is indeed a kind of technology. However, not all technologies are the same and depending on what kind of technology they are, different normative concepts may apply to them that do not apply to others (Lankton et al., 2015).

What all these views neglect is the fact that artificial intelligence is unlike most other technologies. While we trust autonomous vehicles to drive safely from one point to another and pharmaceuticals to chemically intervene in our bodies, what we trust AI to do is completely distinct. When we trust AI we trust it mainly as a source of knowledge: we trust them to *manipulate epistemic content* such as visual input and propositional structures; we trust them to *carry out epistemic operations* on this content such as analysis, inferences and predictions; and finally, we trust it as an aid *in epistemic contexts* such as scientific inquiry, business decisions, etc. In contrast we do not trust aspirin to provide information, to predict statistical trends or to make decisions for us. The same applies to cars. Although people can sometimes be trusted in these respect, not all interpersonal trust is allocated in virtue of the epistemic capacities of the recipient, even if some epistemic capacities are assumed.

Hence, in this paper I argue that Artificial intelligence and methods associated with it, such as machine learning, are first and foremost epistemic technologies.[[2]](#footnote-2) If this is so, then our relationship towards them cannot be immediately drawn from our analysis of our relationship to just any other technologies. As we will see, understanding AI as an epistemic technology will also have significant implications for important debates in the epistemology and ethics of artificial intelligence.

1. **Epistemic enhancers and epistemic technologies**

Many important distinctions between epistemic technologies and other kinds of technologies need to be made for the arguments above to get traction. However, even before getting into those distinctions it is important to remind the reader of a few preliminary distinctions that will be of foundational importance to move ahead with the argumentative structure of this paper. The first one is that what something does and what something is used for is not always the same thing (Alvarado, 2021). This is important because at some point in this paper I will argue that AI is still distinct from other *epistemic* technologies— such as calculators— and I will do so on the basis of what they do, even if they are used for roughly the same purposes, namely epistemic purposes. Hence, I will deny that AI as a kind of technology can be defined merely by its uses and applications. For example, a machine learning method designed to analyze data can be used in many applications. What defines it as an epistemic technology however are its capacities for data analysis and not, say, its deployment in moving a robotic arm. While some people may want to say that AI moves robotic arms, this is not, strictly speaking, what AI does as much as what it is applied on once very sophisticated non-AI software and hardware interfaces are brought into the equation. [[3]](#footnote-3)

Relatedly then, the most important step in understanding the main distinguishing aspect of artificial intelligence vis-à-vis other technologies is to understand what it is that artificial intelligence does. However, before getting into those details, it would behoove us to delve into some preliminary conceptual distinctions that differentiate between epistemic technologies and other technologies. To start with, we must understand that artifacts are intentional, functionally identifiable objects (Kroes, 2002; 2009; Symons, 2010).[[4]](#footnote-4) Hence, knowing what an artifact is entails knowing what it does. Machine learning algorithms are obviously artifacts, they have intentional functions and they carry out specific tasks. As we will see below, in a very general sense, these tasks can be perfectly categorized as epistemic in nature. By itself, stating that AI methods such as machine learning help us in the acquisition and creation of knowledge may seem as a trivial assertion. After all, even a light switch in a laboratory can be said to help us in the acquisition and creation of knowledge. However, the importance of this assertion is best understood in context: not all technologies carry out epistemic tasks. According to philosophers of science like Paul Humphreys (2004), this already represents a crucial conceptual break between technologies such as computational methods and other technologies.

Humphreys (2004) distinguishes technologies that enhance our epistemic capacities like a microscope or telescope from other technologies. Technologies such as the microscope address deficiencies— or help us in a way— 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. The microscope and the telescope enhance vision and visual tasks. Visual tasks are already epistemic in a way that an orbital shaker in a chemistry lab is not. They engage with our cognitive and inferential capacities in a way that the shaker simply is not designed to do.

Epistemic enhancement, however, is not limited to our perceptual abilities. In fact, such perceptual enhancement may not be what is most interestingly captured by the concept of epistemic enhancement coined by Humphreys. 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.”[[5]](#footnote-5) (2004 p.6) A calculator would be a perfect example of a device that is meant to enhance something other than our physical or perceptual capacities. In a similar way, according to Humphreys, “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.” It is well known, for example, that computational devices can perform many more mathematical operations and faster than most humans can. Therefore, they transcend and help us transcend our natural epistemic limitations. 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, devices such as equations and models qualify as epistemic enhancers. However, as seen in the quote above, so do representational devices and translations of one representational mode to another. In other words, epistemic enhancers include not only formal equations but even the visualizations we derive from them (Morrison, 2015).[[6]](#footnote-6) [[7]](#footnote-7) Furthermore, Humphreys notes:

“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)

In this sense, many, if not all, computational devices employed in scientific inquiry qualify as epistemic enhancers. He suggests, for example, that computational technologies “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.” This would include, of course, similar devices we use in clinical and scientific medicine, which have an intrinsic relationship to the acquisition and deployment of medical knowledge.

The idea here is then that at the very least we can differentiate between kinds of artifacts: those that are designed, developed and deployed as epistemic aids and those that are meant to respond to other limitations of our species such as physical strength, patient and consistent endurance, etc. As we will see, when it comes to artificial intelligence the epistemic element is still more profoundly interconnected to the technology itself. This, in turn, serves to further distinguish AI as first and foremost an epistemic technology. While a calculator is removed from a bulldozer in that it carries tasks that could more appropriately be deemed as epistemic, AI is even further so.

Artificial intelligence can be said to do many things: play chess, successfully diagnose patients, fold proteins, perform image/object recognition, guide autonomous vehicles, solve video games, etc. It can also be applied in many domains: finance, robotics and medicine. 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 analytics (Mitchell, 2019). The rest of the aforementioned 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 and deep neural networks are involved.

Machine learning algorithms are a methodological subset of artificial intelligence and although they have recently gained significant notoriety due to their many successful applications, they have nevertheless been around for a 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 and Murphy, 2015). But they are also meant to be able to predict possible properties and patterns in other, previously unknown data sets. Similarly, they are trained to recognize instances of similar items across different data sets. And they are also designed to predict whether a given previously unknown data set is likely or not to include a given 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. While these applications may seem to be 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, 2022).

Similarly, deep neural networks function as multilayer information sifters that filter inputs across layers according to either predetermined or ‘learned’ threshold weights. Deep neural networks are a subset of machine learning 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 (Sarle, 1994; Samek et al, 2021). 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.

While it may be tempting to think of robotic applications in medical AI as a counter point to my position, in reality, in the medical field, almost the majority of AI applications are epistemic as well. Yan et al. (2019 p.586), for example, 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)

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., 2015; London, 2019; Jongsma and Durán, 2021). 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.

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 geared towards enhancing our epistemic abilities. Hence, following Humphreys, we can see these technologies as epistemic enhancers.

However, AI is not an epistemic enhancer *merely in virtue* of enabling the acquisition of knowledge alone.[[8]](#footnote-8) Rather, as can be drawn from the previous paragraphs, AI and other computational methods —including ML and DNNs—are epistemic enhancers *also* because they deal with and enhance our ability to 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, 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 *by* carrying out epistemic operations on it. In this sense we can think of AI manipulating epistemic content such as semantically laden information, but we also have to think of the manipulation itself as epistemic in nature given that the operations involve inferences, predictions, and even decisions.

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’. It is important to note at this point that such epistemic tasks are not, strictly speaking, the same as cognitive tasks. 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 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, therefore it must “know” what it is doing. [[9]](#footnote-9)

To see more clearly what I am getting at with this distinction and with the use of epistemic terms, consider the difference between a core memory which allows us to store and retrieve information and a decision-making AI application. While we can say that the core memory deals with epistemic content—namely information— it does not do so necessarily through an epistemic operation. For example, the automated retrieval system (ARS) in Santa Clara’s University Library helps people extract and store books. In a sense it enables people to gather knowledge and it deals with epistemic content such as books. However, we can see that in its capacity as a retrieval system it only deals with books and partakes in knowledge gathering through physical operations. 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 yield results in highly complex tasks (Bjerring and Busch, 2021). 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 positions are positions with something of an epistemic authority. A critic, a second opinion, an expert and a tutor are definitely sources of knowledge and sources that signal an asymmetry between interlocutors. These are not merely cognitive terms; they are not merely computational assignments and they are not merely information retrieval mechanisms. They are epistemic tasks.

Furthermore, 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)

While it is not good 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. Their epistemic status in this sense is epistemic across three dimensions: they enhance our ability to gain knowledge, they do it in virtue of manipulating epistemic content, and they manipulate this content by carrying out what can be deemed as epistemic operations.

Therefore, we can say that artificial intelligence, as used in specialized fields, is an epistemic enhancer in at least three important ways:

1. It is a technical artifact that is designed to *expand our epistemic capacities* in at least two ways: it is deployed in an epistemic context such as inquiry; and it is also unlike a bulldozer in that the capacity that it expands is the capacity to calculate or compute and not dig, mix, or other physical tasks.
2. AI *deals with epistemic content*: propositions, models, visual concepts.
3. AI is also the kind of technology that *carries out epistemic operations*, such as analysis, prediction, and inference.[[10]](#footnote-10)

In short, 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.

1. **Implications**

To see the main contribution of this paper, perhaps it would be useful to provide a contrasiting example. In particular, consider other tehcnologies which may fulfill one of the main three categories of epistemic labor. We can ask the following questions:

Are these technologies deployed in an epistemic context?

Do they deal with epistemic content?

Do the deal with that content through epistemic operations?[[11]](#footnote-11)

Although the aim of this paper is mainly to provide a first step towards a foundational framework to best understand the nature of artificial intelligence as an epistemic technology and hence to best understood the role it plays in inquiry, in this section I aim to offer a brief overview of some immediate implications of such an understanding. This overview is not meant to be exhaustive and it is not meant to provide full-fleshed solutions to the challenges it lists. Rather, it is meant as a brief preview of the possible disruptions this view may bring about in important contemporary debates surrounding AI. In particular, as Alvarado (forthcoming) notes, this disruption will become obvious in debates involving reliance and trust. But also, it will have a significant effect in debates about explainability, epistemic opacity and can even further explain epistemic harms particular to AI methods that Symons and Alvarado (2022) point to.

Let us now put our discussion back in context. The introduction of this paper began by stating that understanding AI as an epistemic technology was in sharp contrast to growing views in the literature that attempt to make the case for the acceptance and adoption of AI by comparing it with other technologies. It is in these analogies where the disruptive element of the view defended in this paper become more immediately obvious. Therefore, let us relate our discussion back to claims such as London’s (2019) concerning the analogy between accepting aspirin and accepting medical AI. 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). If this is so and we’ve deemed the deployment of opaque pharmaceuticals permissible, the argument goes, then we have no good reason to reject AI for being opaque, unless we also reject many instances of modern medical practice.

Although much could be said about this argument, [[12]](#footnote-12) for the purposes of our current conversation, let us focus on the fact that in order for the argument above to really work we must ignore that we are talking about two completely different kinds of things, used for different purposes, that carry out completely different sets of operations, and which are relied upon in distinct circumstances: a pharmaceutical and a computational artifact. Ignoring this while making an argument from analogy seems like a highly suspect argumentative strategy.

Consider 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. We do not deploy aspirin in or for inquiry, for example. We also do not deploy aspirin in decision-making or as a decision-maker. If we look at the tasks that are carried out and expected from both aspirin and AI technologies, we can easily see how far apart they are. While the expectation is that an aspirin will carry out some chemical intervention with biological and phenomenological outcomes, AI is not deployed in this way. AI takes data, analyzes it, and infers or predicts things about it. It also distinguishes, categorizes, and takes other actions on it: stores it, converts it, takes it into consideration for further operations, or simply conveys it. Aspirins, on the other hand, are designed and directly deployed to treat pain and other ailments. In other words, the tasks in which it is deployed are directly related to our health. The operations an aspirin carries out are biochemical and the material with which it interacts is organic and not propositional, attitudinal, symbolic or doxastic in any way, shape or form.[[13]](#footnote-13) The material with which artificial intelligence *is* directly and exclusively involved is of all these latter sorts. Hence, we can conclude that pharmaceutical artifacts are not involved in the same contexts, they do not interact with us in the same way and they do not carry out the same tasks as AI. Their use as analogical cases concerning the trust or reliance on AI technologies is simply is at best undermotivated and at worst simply misguided.

One could object here that some pharmaceutical interventions are indeed deployed with epistemic aims. 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. (Gerer, 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 as epistemic enhancers. If this is so, then we can appeal to them to make an analogy on how and why we could or should trust and accept AI in the same way. 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 interacts is interacting is not epistemic content.[[14]](#footnote-14) Rather, Adderall as a technology interacts with the chemistry in our biological system. 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 in a library. 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.

While the discussion above restricts itself to the comparison between AI technologies and pharmaceutical technologies, similar outcomes can be expected when comparing AI to other technologies such as autonomous vehicles, IT systems in general, etc. The same method of analysis can be deployed to test if any technological analogy is warranted: what is the technology designed, developed and deployed for? Does this technology share design intentions, context of deployment or kinds of operational processes with AI? If so, to what extent? Are they also epistemic technologies: deployed in epistemic contexts, manipulating epistemic content and doing so through epistemic operations? If not, then we should pause and reflect on our comparison. [[15]](#footnote-15)

Understanding AI as an epistemic technology, also has important implications for other central problems in the philosophy, epistemology and ethics of AI. Consider, for example, the interrelated debates regarding the challenges of explainability and opacity that AI faces. While we can accept the use of a hammer without knowing in detail how or why it works, 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 suggest (Humphreys, 2004; 2009; Alvarado, 2020; 2021a). 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 face challenging epistemic 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. 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: how reliable are their epistemic processes? How consistent and robust their operations and results? How exactly do we know how reliable they can be? 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 and Formanek, 2018). According to Alvarado (forthcoming), we should only seek to trust epistemic technologies in their capacities as information providers. That is, we ought to only allocate epistemic trust to epistemic technologies. Furthermore, according to Simion (2018; 2019), [[16]](#footnote-16) the proper deployment of epistemic concepts such as epistemic trust, ought to be guided *only* by epistemic considerations such as epistemic norms. If this is so, as I think it is, then sound epistemic means are only properly deemed so, justified, by appropriate epistemic reasons, concepts and warrants (Dretske, 2000; Symons and 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). But in order to do this, we need to understand these processes. 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.

The same applies to claims that explicitly or implicitly entail any sort of epistemic authority (Simion, 20018). 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— they do not speak to the truth, falsity or legitimacy of the claim to epistemic authority. 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 form a position of authority (as an expert or decider) will require appropriate justificatory efforts, some of which include the ability to explain itself. If someone in a position of authority is not able to explain themselves after a decision affecting others, they may be occupying that position illegitimately (Lazar, forthcoming; Symons and Alvarado, forthcoming).[[17]](#footnote-17) If AI technologies are essentially and representationally opaque as most literature suggests (Burrell, 2016; London, 2019; Alvarado, 2021a; Durán and Jongsma, 2021; Alvarado 2022), we cannot have epistemic access to its relevant features. This is very important when one is trying to know straightforward and reasonable things about a model, such as whether or not or when it extrapolates beyond the possibilities initially provided by the data it is trained on (. 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 suggest, then we simply cannot trust AI.

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, will not work. While it *may* be acceptable to adopt many other technologies despite their epistemic opacity, justifying the acceptance of epistemic technologies that are epistemically opaque may prove to be significantly harder. Appealing to prudential or practical considerations for their sanctioning will simply not suffice.

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 the technology. Namely, understanding AI as an epistemic technology could explain how and why epistemic harms (Symons and Alvarado, 2022) are so important in the ethical assessment of AI. 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 the kinds of harms that they are more likely to cause with these epistemic functions are epistemic harms, 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.

Hence, the understanding of 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.

1. **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.

**References**

Alvarado (forthcoming). What kind of Trust Does AI Deserve, if Any? Accepted for Publication in *AI and Ethics*.

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

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

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

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

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. (1983). An overview of machine learning. *Machine learning*, 3-23.

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.

Danks, D. (2019, January). The value of trustworthy AI. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 521-522).

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

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 *machine learning in radiation oncology* (pp. 3-11). Springer, Cham.

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.

Ferrario, A., & Loi, M. (2021). The meaning of “Explainability fosters trust in AI”. *Available at SSRN 3916396*.

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, *1*(6), 261-262.

Girer, N., Sasu, N., Ayoola, P., & Fagan, J. M. (2011). Adderall Usage Among College Students.

Hakkarainen, K., Engeström, R., Paavola, S., Pohjola, P., & Honkela, T. (2009, September). Knowledge Practices, Epistemic Technologies, and Pragmatic Web. In *I-SEMANTICS* (pp. 683-694).

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.

Hintikka, K. J. J. (1962). Knowledge and belief: An introduction to the logic of the two notions.

Horsburgh, H. J. N. (1960). The ethics of trust. *The Philosophical Quarterly (1950-)*, *10*(41), 343-354.

Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford University Press.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, *169*(3), 615-626.

Hurlburt, G. (2017). How much to trust artificial intelligence?. *It Professional*, *19*(4), 7-11.

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.

Kroes, P., & Meijers, A. (2002). The Dual Nature of Technical Artifacts-presentation of a new research programme.

Kroes, P. (2010). Engineering and the dual nature of technical artefacts. *Cambridge journal of economics*, *34*(1), 51-62.

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>

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.

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.

Meyer, J. J. C. (2003). Modal epistemic and doxastic logic. In *Handbook of philosophical logic* (pp. 1-38). Springer, Dordrecht.

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*. Penguin UK.

Ratti, E., & Graves, M. (2022). Explainable machine learning practices: opening another black box for reliable medical AI. *AI and Ethics*, 1-14.

Ratto, M. (2012). CSE as epistemic technologies: Computer modeling and disciplinary difference in the humanities. In *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.

Rossi, F. (2018). Building trust in artificial intelligence. *Journal of international affairs*, *72*(1), 127-134.

Sarle, W. S. (1994). Neural Networks and Statistical Models. In *Proceedings of the Nineteenth Annual SAS Users Group International Conference.*

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

Simion, M. (2018). The ‘should’in conceptual engineering. *Inquiry*, *61*(8), 914-928.

Simion, M. (2019). Conceptual engineering for epistemic norms. *Inquiry*.

Simpson, T. W. (2013). Evaluating Google as an epistemic tool. *Philosophical Engineering: Toward a Philosophy of the Web*, 97-115.

Stolz, S. (2012). Adderall abuse: Regulating the academic steroid. *JL & Educ.*, *41*, 585.

Symons, J. (2010). The individuality of artifacts and organisms. *History and philosophy of the life sciences*, 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., & Alvarado, R. (Unpublished manuscript). When Does Explainability Matter? Presented at the 2022 IACAP conference. Santa Clara, California, July, 2022

Taraborelli, D. (2008). How the Web is changing the way we trust. *Current issues in computing and philosophy*, 194-204.

*Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & Van Moorsel, A. (2020, January). The relationship between trust in AI and trustworthy machine learning technologies. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 272-283).*

Varga, M. D. (2012). Adderall abuse on college campuses: a comprehensive literature review. *Journal of evidence-based social work*, *9*(3), 293-313.

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.

1. See Ferrario et al., (2020) Ferrario and Loi, (2021a) Ferrario et al.,(2021b) for similar argumentative strategies. [↑](#footnote-ref-1)
2. This term is to be differentiated from its use in media, media studies literature where ‘epistemic technologies’ are refer to contemporary pedagogical methods and devices which range from powerpoint presentations to search engines (Ratto, 2012; Miller and Record, 2013; 2017; Record and Miller, 2018) Some new media epistemology literature introduced the term more than a decade ago. (Hakkarainen et al, 2009) [↑](#footnote-ref-2)
3. I want to thank an anonymous reviewer whose comments made clear this distinction needed to be explicitly stated. [↑](#footnote-ref-3)
4. Symons (2010) argues that intentionality is necessary to the identification of an artifact to differentiate them from other functionally identifiable objects such as organs. Nevertheless, the relationship between intention 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).

   [↑](#footnote-ref-4)
5. Humphreys admits of three kinds of epistemic enhancement: extrapolation, conversion and augmentation. For a detail analysis of what these are see Humphreys (2004 p.3-6) [↑](#footnote-ref-5)
6. The value of representational tools that are not strictly speaking factive is the subject of a rich debate in philosophy of science. For a thorough overview of the intricacies related to idealizations in scientific representation see (Pincock, 2015). For a similar overview related more closely to computational methods such as computer simulations see Morrison (2015). [↑](#footnote-ref-6)
7. While Humphreys claim can be easily dismissed or accepted as not having too much at stake, it is important to note that whether or not simplified representational devices do in fact furnish knowledge is a question with a long history in the philosophy of science (See footnote 14). 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; 2015). For a thorough refutation of this conventional view, see Sullivan and (2021). [↑](#footnote-ref-7)
8. This is important to note because, as noted above, some may want to make the case that in this respect a computational method is no different from a microscope, or perhaps a light switch at a laboratory. Both of which in one way or another enable and increase the carrying out of epistemic tasks such as inquiry and experiment. [↑](#footnote-ref-8)
9. 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. [↑](#footnote-ref-9)
10. As noted above, it is important to note that such epistemic task are not the same as cognitive tasks. It can be said that while epistemic tasks require cognitive processes, cognitive processes are not necessarily epistemic tasks. [↑](#footnote-ref-10)
11. It is here that the main distinction between the context, the use, and the functions of an artifact come to the fore. What something does, what it is used for and what it is deployed for may be distinct. (Alvarado, 2021 Simulations as Scientific instruments footnote no. X) [↑](#footnote-ref-11)
12. 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, as we will see below, 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 in a 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. [↑](#footnote-ref-12)
13. By doxastic here I do not mean to imply that the AI system, as it manipulates information, epistemic content, partakes in anything closely related to belief or other mental states with an associated epistemic connotation. Rather I mean to signal that it manipulates the kind of content that could formally be part of an epistemic logic. (See Hintikka (1962) for an overview of doxastic/epistemic logics as well as Meyer, (2003) for a discussion that briefly touches on their role in the AI literature).

    [↑](#footnote-ref-13)
14. As Alvarado (2021) 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. [↑](#footnote-ref-14)
15. We should 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) argue in detail, accepting a novel artifact into formal inquiry is simply not like accepting the input of other experts , it is not like accepting the people that build such technologies, and it is not like accepting the methods by which such technologies are built either (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 to) 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. (See Symons and Alvarado, 2019 to see a thorough account of the distinct epistemic warrants at play in each of those cases).

    [↑](#footnote-ref-15)
16. According to Simion, an epistemic norm is one that is closely related to an epistemic value (Simion 2019). [↑](#footnote-ref-16)
17. Presented at the meeting for the International Association of Computing and Philosophy, 2022 in Santa Clara, California. [↑](#footnote-ref-17)