Perspective

On the *sui generis* value capture of new digital technologies: The case of AI

Emre Kazim,[1,2,10,*] Enzo Fenoglio,[1,3] Airlie Hilliard,[2] Adriano Koshiyama,[1,2] Catherine Mulligan,[7] Markus Trengove,[2] Abigail Gilbert,[5] Arthur Gwagwa,[6] Denise Almeida,[8] Phil Godsiff,[9] and Kaska Porayska-Pomsta[4]

[1]Department of Computer Science, University College London, Gower St, London WC1E 6EA, UK
[2]Holistic AI, 18 Soho Square, London W1D 3QL, UK
[3]Cisco Systems France, 11 Rue Camille Desmoulins, 92782 Issy-les-Moulineaux, France
[4]School of Public Policy, University College London, 29 Tavistock Square, London WC1H 9QU, UK
[5]Institute for the Future of Work, London, UK
[6]Ethics Institute, Department of Philosophy & Religious Studies, Utrecht University, Janskerkhof 13, 3512 BL Utrecht, The Netherlands
[7]Central, Department of Engineering and Informatics, Instituto Superior Técnico, Avenida Rovisco Pais 1, 1049-001 Lisbon, Portugal
[8]Department of Information Studies, University College London, Gower St, London WC1E 6EA, UK
[9]INDEX, Business School, University of Exeter, Exeter EX4 4PY, UK
[10]UCL Institute of Education, Knowledge Lab, 23–29 Emerald Street, London WC1N 3QS, UK
*Correspondence: e.kazim@ucl.ac.uk
https://doi.org/10.1016/j.patter.2022.100526

THE BIGGER PICTURE Digital technologies are emerging at a fast rate, with applications ranging from farming to recruitment. Much of the research on these technologies has concerned optimization and applications, with less focus on the regulation and governance of these systems and how they might bring about foundational and theoretical shifts. Indeed, much of the literature is concerned with forwarding technical approaches, and the potential opportunities and harms, without offering theoretical or philosophical perspectives; few have asked what a digital thing is, what the ontological nature and state of phenomena produced and expressed by digital things are, and if there are distinctions between the conceptions of digital and non-digital technologies. We forward this discussion by investigating the question of what value is being expressed by an algorithm, which we conceptualize in terms of a digital asset, which we define as a valued digital thing that is derived from a particular digital technology.

Concept: Basic principles of a new data science output observed and reported

---

SUMMARY

Much of the academic interest surrounding the emergence of new digital technologies has focused on forwarding the engineering literature, concentrating on the potential opportunities (economic, innovation, etc.) and harms (ethics, climate, etc.), with less focus on the foundational and theoretical shifts brought about by these technologies (e.g., what are "digital things"? What is the ontological nature and state of phenomena produced by and expressed in terms of digital products? Are there distinctions between the traditional conceptions of digital and non-digital technologies?. We investigate the question of what value is being expressed by an algorithm, which we conceptualize in terms of a digital asset, defining a digital asset as a valued digital thing that is derived from a particular digital technology (in this case, an algorithmic system). Our main takeaway is to invite the reader to consider artificial intelligence as a representation of the capture of value *sui generis* and that this may be a step change in the capture of value vis à vis the emergence of digital technologies.

She arrived -
Exorbitant smile
An exquisite avowal
Of her spirit and gile

Crowned by a throne
Of fire - screaming
Hello, I am here

— Trafalgar Tavern

INTRODUCTION

With the emergence of new digital technologies—such as, but not limited to, the Web, blockchain, artificial intelligence (AI), and Internet of Things—we have witnessed significant change in the production and structuring of products and services. While much of the interest from commentators has been focused on the potential opportunities (economic, innovation, research, etc.), the potential harms (economic, ethics, climate, etc.),[1] and forwarding the engineering techniques grounding such technologies, in concrete terms less has been explored in the conceptual realm regarding foundational and theoretical shifts brought about by these technologies. Indeed, many questions remain unanswered, such as what exactly are "digital things"? What is the ontological nature and state of phenomena produced by and expressed in terms of digital products? And is there anything distinct about digital technologies in comparison with how technologies have been conceived of traditionally?

In this article, we investigate the question of what value is being expressed by an algorithm, which we conceptualize in terms of a digital asset, where we take the notion of a digital asset to be a valued digital thing that is the product/derived from a particular digital technology. In this paper, we focus on algorithmic systems, which we read as a digital asset that is valued because it realizes a particular aim, such as a decision or recommendation for a set task. Our main contribution is an invitation to the reader to consider novel digital technologies, such as AI, as a representation of capture of value *sui generis* (compared with traditional technologies); i.e., digital technologies (such as AI) potentially have unique value capture compared with traditional technologies, and this may be a step change in the capture of value vis à vis the emergence of digital technologies (such as AI). Stated alternatively, we invite the reader to reflect on how value, as substance or relation, is captured by algorithmic systems, and suggest that this does represent a step change in the history of technology.

We forward our argument through division of this article in the following manner:

(1) Things: in preparation for our discussion of what a digital thing is, we first discuss what we mean by a thing. Here, we briefly sketch, conceptually, how something digital can be thought of and use this in subsequent sections to inform our discussion of a digital asset.

(2) Value: we follow the above by discussing value in the context of a digital asset (we frame AI in terms of a digital asset). Here we do so by borrowing from debates centered on value theory from classical economics.

(3) AI as case study: in order to illustrate our thesis, we then move from the conceptual mapping to provide an account of AI in terms of the capture of value associated with intellectual labor.

(4) Conclusion: we note that the value capture discussed relates to the activity side of intentionality (practical reason) and that the questions remain regarding attitude and ultimate value determination (rooted in the will, desire, substantive moral difference, etc.).

The broader importance of our investigation is that the transition in value-capture—from physical labor value to intellectual la-

bor value—constitutes a categorical step change, rather than an escalation along the same continuum. Accordingly, it is inadequate to extrapolate from existing conceptual frameworks, if we are fully to comprehend the changes and challenges occasioned by algorithmic systems. The full implications of this conceptual shift warrant further investigation, and so we suggest further research priorities in our conclusion.

Before turning to our arguments, we note the context in which this article was written: the authors represent an interdisciplinary group—from fields such as engineering, computer science, economics, law, information studies, and philosophy—interested in both responding to the advent of this new digital age (within particular focus on the algorithmic systems) and simultaneously moving beyond such *reacting* by venturing with a movement toward the realm of grounding and laying the conceptual foundations of this new period. Indeed, we seek to provide an interdisciplinary perspective of this new technological era by laying conceptual foundations regarding the relationship between value and information technologies. As such, the article should be read as an articulated thought experiment, as an invitation and call to action for others to join in this space.

THINGS

Ontology

As a point of departure, it is first necessary to outline, in broad terms, what we understand to be a thing. Before doing so, it is important to note that, within the philosophical literature, there is a considerable legacy and history of debate concerning what a thing is. We can state the debate in terms of its extremes. At one extreme is the view that things are a product of human framing and, as such, it does not make much sense to speak of things as discrete objects outside of this framing;[2,3] on the other extreme is the view that things can be thought of as discrete objects in themselves.[4] While we recognize the richness of this literature, in this paper we will not explore the debate but rather frame the extremes in terms of intrinsic and extrinsic ontology, where the former encapsulates the idea of things as discrete entities and the latter as entities which are framed (in terms of relationships between objects and within value systems). As such, we must provide a caveat that the below introduces concepts that we will use in this paper to flesh out the ontological status of digital objects, digital assets, and, ultimately, algorithmic systems.

Accordingly, a thing can be defined in a number of ways; at the most abstract, we can think of a thing as being thought of with respect to questions that are intrinsic to the thing, and questions that are extrinsic to the thing. Intrinsic and extrinsic approaches to things and value can be broadly mapped onto the following:

(1) Intrinsic (ontological categorization): an intrinsic definition of a thing or object relates only to the properties contained by the thing itself, independently of other objects. Producing an intrinsic definition of a thing requires investigating its essential and accidental properties, its wholeness and parthood, particulars and universals, being and becoming, and *substance*. Here a thing is categorized only according to its inherent status and properties (hence

why we refer to the intrinsic defining of thinghood as ontological categorization).

(2) Extrinsic (relational ontology): an extrinsic definition of a thing conceives of it only in terms of how it relates to other things or objects. Here, what is key to understanding the essential properties of an entity is to understand its position in a complex of relationships. For instance, we can think of the relationship between people (where the existence of persons as discrete entities is posited), the relationship between persons and inanimate objects (where objects are posited), and the relations between inanimate objects. Here an understanding of a thing is inextricably related to its relation to another thing, hence why we refer to extrinsic defining of thinghood as relational ontology.

Note that these two approaches to discussing a thing depend very much on the purpose with which the investigation is taking place. For example, if we are interested in metaphysical truth regarding such things as a logical possibility, we can readily and appropriately appeal to ontological categorization. Contrastingly, it may be entirely appropriate to engage in the relational ontology where the metaphysical status (in terms of necessity and truth) of the thing is not in question but instead how that thing exists concerning other things. The relationship between persons and inanimate bodies in the world may be fully investigated without exhaustive investigations into the ontological categorization of either concept (hence why we state that the objects are *posited*). Indeed, we can go further and state that the same one thing can be thought simultaneously in ontological and relational terms when the relationship between the thing and itself and/or another object requires this form of analysis. An example of this would be scholastic debates concerning the nature of the Trinity (see Augustine in De Trinitate[5]). As such, although the two, namely ontological categorization and relational ontology, are conceptually distinct, they are nonetheless not competitive.

Digital things

The relevance of our broad sketching of ways of thinking about thinghood to questions of the digital comes from the need to elucidate what a thing is in the digital realm. To give substance to this concept, we will first consider the relationship between the digital and the physical, and then pose the problem of categorization (intrinsic) and relational (extrinsic) ontology.

Things and physicality

Common notions of what a thing is typically appeal to physical objects (physicality/materiality) in the world, such as mountains or tables, but there are numerous ways in which we can think of things that do not directly appeal to physical objects. A paradigmatic example of this is mathematics; according to some metaphysical frameworks, mathematics is a property of the universe and exists in and of itself (lending itself to ontological categorization), while, according to others, mathematics is an abstract expression of ways in which humans express/communicate propositional claims about various things such as the physical world, and axioms (lending itself to relational ontology). This example is instructive because it also shows that, even in cases where physicality/materiality is posited (mathematics as

grounded in physical objects), abstract thinghood (in the form of mathematical propositional claims) is possible (albeit as a derivative).

In other words

(1) An abstract thing, that is derived from physicality/materiality, is also possible.

(2) To be a thing, something need not appeal to physicality/materiality at all.

For example, there are competing narratives regarding how humans have grasped the concept of number; one account has it that numbers are grounded in human experience, where we have evolved to count, through the grouping of objects and then abstracting concepts such as single, double, one, two, addition, and subtraction. Here there is an abstraction that is derived from physicality. Another account is that numbers are pure concepts that arrive from purely rational cognition; here our ability to count real objects in the world comes from our applying of this rational concept to real-world objects. There is no abstraction from the physical but instead an imposition to the world. In either case, irrespective of the origin of the concept (abstraction through experience, a derivative of physicality; or pure conceptualism, with no material grounding), the concept of a number is still abstract and certainly something with ontological status.

Digital irreducibility

Given the above, we may assert that digital things are abstractions; that, irrespective of whether they are derived from physicality/materiality, they can be thought of as discrete and distinct, with the ontological status that can be thought of in both categorical and relational terms. Indeed, these abstract (digital) things can be thought of both in terms of ontological categorization (investigation into the intrinsic of the digital) and relational ontology (investigation into relationships between digital things, persons, and digital things, etc.), and, as a corollary, we can assert that in cases where an account of the physicality/materiality—that the digital is derived from—is required, the novel digital technologies (listed in the introduction as AI) provide this account. They are the grounding of these derivatives (i.e., digital things). However, this does not exhaust the ontological debate regarding what digital things are. Although we are to think of digital things as derivatives of these novel digital technologies, digital things are non-reducible to the materiality of those technologies (in other words, they are epiphenomenal).

From the above, we can conclude that digital things require ontological investigation in terms that do not reduce the digital to the physicality of the technologies from which the digital is a derivative.

In the next section, we approach this problem from the standpoint of value. However, before doing so, it is worth noting that Bergen and Verbeek offer a relevant insight when commenting upon the dictum "To the things themselves!", which is inspired by Husserl's proclamation; Peter-Paul Verbeek introduced his theory of technological mediation as part of a "thingly turn," a "philosophy of artefacts."[6] As part of the post-phenomenological tradition, mediation theory aims to take concrete, socially situated technological artefacts seriously, recognizing the

constitutive role they play in how we experience the world, how we act in it, and even the way we as subjects are constituted and can constitute ourselves.[6,7] This literature approaches the problem we are interested in from the side of how the systems (that already exist) become things in the world, rather than how this paper approaches it, namely, by thinking about what human form of cognition these systems (digital assets) mimic. Indeed, on this final point it is worth noting that our position in this paper assumes a model of cognition which is mentalist, computational, and representationalist. Recently, challenges to this model are being offered (e.g., a 4E version of cognition[8]); although we will not explore them here, we recognize that a reformulation of our argument would have to be presented in light of these alternative models of cognition.

VALUE

The reader may ask why we have not appealed to information theory with its vast resource of ontological investigation concerning digital information. Indeed, one may read ontological categorization in terms of the storage and quantification of information as data (digital things), and relational ontology in terms of communication theory (which is necessarily relational: the relationship between digital things, and persons and digital things, etc.). The reason we have chosen not to do this is that we are interested in a high-order investigation into digital things and value. Indeed, our investigation is axiological (criteria of value and value judgments). Given that we are conjugating the question of how value and values are captured by digital things, we will now refer to digital assets, where an asset is something that is valued, and thus a digital asset is a category that is grounded on a digital thing within a value system.

We are motivated to do this because, later in the article, we will explore our interest in digital and the question of value from the perspective of digital assets that are derived from a particular technology (we do this by case studying AI). What we are referring to here is the notion that not all digital derivatives (digital things/assets that derive from digital technologies) are ontologically (and epistemically) the same. Technologies such as the Web, AI, and blockchain beget information expressing differing values; i.e., these digital assets grounded in these technologies are not the same sorts of information-value datum. Indeed, the nature of the respective digital asset will reflect the uniqueness of the particular technology.

As such there is a matrix of concepts that will need to be explored; these are namely questions of:

- Value: what exactly is the value being spoken about when discussing the value of a digital asset?
- Novel value capture: how can we explain how a digital asset can be thought of as an expression of the value that goes beyond digital as a simple datum?
- Digital technology and derivative value: can we explain the products of new digital technologies by going beyond information in its most basic form (information *qua* data), to more sophisticated forms such as information as immutability and information as intelligence? (This question is mostly explicated in section "AI" via a discussion of AI.)

What value?

The relationship between a digital thing and value—a digital asset—can be thought of in numerous ways, each of which appeals to how value is understood itself. To forward our investigation, below we provide two ways of thinking about value and digital assets that do not appeal to the capture of new value (but instead, at best, increases value in terms of utility). Note that this is not a comprehensive survey but instead serves to cast some alternative notions of value in terms of digital things.

Neutral: In value terms, there is nothing particular about the value of digital assets

One may argue that it is necessary to pose the question of value as preceding questions of digital thinghood, by asserting that a concept of value is required to assess the value of a digital thing (the digital asset). This prioritizes the question of value more generally over and distinct from that of a digital asset. In this account, when determining the value of a digital asset, one would be engaging in the same kind of activity as one would be when valuing any other thing. Here, digital assets would be a class of things like anything else that is evaluated within an independent value system. One way of cashing this out would be to think of the notion of a digital asset as value neutral in itself. For example, Aristotle's words are what is valuable, irrespective of whether we are reading them on papyrus, paper, or a screen.

Utility: Digital assets are valuable due to their utility

Here, via the use of digital technologies, there is a value that is to be ascribed to digital assets under the step change they represent in terms of utility. The usefulness of digital assets is such that the value in utility when in digital form is of a different category to the non-digital. Taking basic information (data) as an example, digital storage, sorting, sourcing, and searching render the data in digital form—as a digital asset—a value that did not exist before the advent of such digital technologies. In this respect, the value of digital assets is *new* and created by new digital technologies themselves. This can be thought of in terms of efficiency also. For example, one may argue that the act of research is an entirely different activity in light of the age of digital assets (comparatively with pre-digital ages and technologies). From a value standpoint, this is assuming that utility is a value that is worth maximizing: digital things are assets when this value is assumed. Indeed, we can note a distinction between the utility and the value theory of labor's relational notions of value above, and the ability of the latter to recognize that utility is not served to universal benefit. As such, as value structures labor to design some forms of digital asset (and, for instance, their regulation) in Europe, this directly impedes the interests of those in the Global South (see debates about EU and digital colonialism etc.). In the Global South, maximizing the utility offered by digital assets can present unique challenges and influence how valuable these technologies are to historically marginalized communities who may experience further hardship if the technologies are not implemented in appropriate ways, further emphasizing that the value of digital assets is dependent on their utility (with conditions to determine which assets are made, in the service of which ends, structured by wider structural factors).[9–11] Moreover, we can expand the understanding of utility to one thought of in terms of a scientific instrument. For example, a microscope represents an increase in utility insofar as it allows for a greater degree of magnification/resolution when viewing objects.

Algorithmic systems can also be thought of in terms of utility increase for scientific discovery; for instance, a pattern-matching algorithm that enables the users to pose new questions, rather than find new answers. Indeed, in the field of recruitment, the use of machine-learning-based algorithms that recognize patterns in data is enabling the advancement of algorithmic recruitment tools (e.g., Hilliard et al.[12]), which is broadening the opportunities for research in this field.

Although we recognize these two ways of thinking as important notions within the discussions of digital assets (indeed, we recognize these as values), we will leave these to now turn instead to questions regarding whether digital assets can be thought of as expressing/capturing new forms of value.

Novel value capture

In the above points regarding value neutrality and the utility value of digital assets, we have implicitly appealed to the notion that a digital asset is a kind of store of information. Indeed, respectively, in the accessing of that information and in rendering that information more useful (through its digitization), the value of the digital asset is understood. As such, we can express a digital asset in terms of the digital store of information; therefore, digital information is an abstraction of new digital technologies. In the cases above, notwithstanding the value of utility, the nature of the information itself remains the same (its inherent value). Certainly, new information can be derived through the bringing together of an analysis of digital assets; however, this would not posit a novel category of information expression. Another way of stating this is that it would not posit a novel category of value itself; the reason for this is that the information is valuable and hence, in expressing the information, the digital asset is storing that value.

In this section, we want to explore the notion that novel digital technologies may capture and represent information (as digital assets) *sui generis*. To do this, we must first pass comment on what is understood by information and then, second, explore the capture of value. In the following section, we will elaborate our discussion via the case study of AI.

On information

We take data to be a unit of information and we take information to be propositional, where a proposition can be thought of as an "about statement"; these are statements that have content that relates to concepts. For example, expression of [1.23] is informational (propositional) if and only if it is qualified by a concept. The units attached to 1.23 m, 1.23 s, 1.23 m^3 give meaning to the numbers, and in doing so, 1.23 becomes information (propositional). Indeed, at a more abstract and basic level, the symbols (referred to in parentheses) [1][.][2][3] are themselves informational insofar as they are qualified by concepts such as number, natural number, integer, decimal, sequence, and directionality (sinistrodextral). In the digital realm, data are an expression of meaning; 1.23 s as a data point capturing time and 1.23 m as a data point capturing length. Meaningfulness comes from what the symbols express and within which conceptual framework (referred to above in terms of the qualifier) the symbols are expressed. Hence why we can speak of data as the capture of meaning.

Stating data as a capture of meaning is contentious, given that information is defined (by some) as a capture of meaning and, as

such, in the above data and information are being synonymized (hence the contention). This may be articulated as a contention that can be stated in terms of meaning emerging in the context of metadata. Of course, one cannot discuss digital assets' association with information and data without considering wider archival and information studies literature.[13,14] In some ways, the contention around the definition of information—and its association with the concept of meaning—could be viewed in parallel as the contentions around the definition of a record, and its association with the concept of evidence (for further discussion about the definition, ethics, and philosophy of information see Floridi[15,16]).

According to Shannon's mathematical theory of communication (MTC), "frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem."[17] Note that this aligns with the digital irreducibility thesis above, insofar as the phenomena (information and semantic [meaning]) of ontological concern is independent of the genesis (materially) of the phenomena. MTC is a theory of information without meaning not in the sense of being meaningless but in the sense of not yet being meaningful,[18] whereas we are appealing to a conception of information as propositional and hence inherently about something and thus inherently meaningful.

The relationship between meaningfulness and value is 2-fold.

(1) Intrinsic value: in one dimension, data can be thought of as expressing/capturing information along a spectrum that spans the extent to which meaning is being captured. This can be thought of in terms of the richness of the data. At one extreme, atomistic elements of data can be conceived as expressing basic concepts such as number or space, and, at the other extreme, complex forms of data denoting an ensemble of relations and properties can be considered. Here we can think of value as proportionate to how rich the data is (indeed this is valuable intrinsically). This value can also be thought of as appealing to the assumption of the value of utility.

(2) Extrinsically: in another dimension, data can be thought of with respect to the role they play when interacting with a context. The context is provided by other data that are not intrinsic to the data themselves. Consider data [Wimbledon tennis ball], expressing a cluster of meaning related to a game, a particular tournament, a weight, a color. Now consider [Jamal's Wimbledon tennis ball], where the introduction of Jamal changes the expression of meaning to something that is distinct from [Wimbledon tennis ball]. Let us define two ways in which the meaning is changed:

(i) [Jamal's Wimbledon tennis ball] taken in terms of a meaning that relates to ownership. It is a description of fact relating to the ball; i.e., that particular tennis ball is the property of Jamal. Here the expression can be thought of as a meaning that remains intrinsic; it is a richer description and hence has more meaning and thus greater value in utility terms.

(ii) [Jamal's Wimbledon tennis ball] taken in terms of Jamal's relationship to a particular tennis ball. It is an

expression of how Jamal consciously relates to the particular ball; i.e., that particular tennis ball has meaning for Jamal. Here the expression is extrinsic, it is inextricably an expression that is derived from the relationship between Jamal's subjective relationship to the particular tennis ball and the tennis ball itself. Here, with respect to the extrinsic, the value is not in the richness of the data but rather in that it captures meanings of relations in terms of value. Clusters of properties and relations, for example, can enhance the richness (intrinsic value) of expressed meaning in data (indeed, clusters of more primitive data into complex datapoints is precisely what this is: such data express a richer array of information). However, by contrast, we can think of the capture of meaning, in extrinsic terms; that is, inextricably *subjective*. Jamal's intention, his subjective consciousness, is being expressed/captured. The capturing of subjective consciousness, in the form of Jamal's intentionality, is the value captured in such data.

Capture of value: Digital assets uniquely capture and express forms of (intellectual) labor

It is our contention that new digital technologies have the capacity to capture meaning *sui generis* because they represent (express), for the first time, a capture of *aspects* of intentionality.

Intentionality is a rich concept, which we will not cover exhaustively here. It suffices for the purposes of our discussion and investigation to think of intentionality in terms of an attitude toward something, and, in its practical form, the actioning of that intention (where actioning is the mechanism or process by which the attitude (intention) is realized (brought about)). Indeed, in this investigation we will always be referring to intentionality as attitude plus action, which is the subjective state plus the practical process of bringing this about.

In order to explain and expand upon how intentionality may be captured by new digital technologies, we can make use of concepts that emerged as a result of debates centered on the labor theory of value in classical economics. We do this not to appeal to the strength of the arguments for a labor theory of value but instead because we utilize some of the conceptual thinking regarding the way in which the theory discusses how value is in the world with respect to intentionality (where labor is understood as acting on the means needed to bring about an end). Scholars of classical and political economy will readily contend that the reconstruction we provide does not do justice to the core texts and continued criticisms. This would be correct; indeed, it is not our intention to frame our discussion within this literature. There are a number of reasons for this. First, any reconstruction would be lengthy and contentious, as well as challenged by other readings of the tradition. Second, we are not interested in this kind of scholastic contribution; instead, we seek only to *borrow* the language for the purpose of teasing out the value question with respect to the digital asset of AI. Finally, we believe that the conceptual repertoire, in classical value theory debates, reflects the axiological tension between labor as utility value and labor as subjective value of human development and intellectual exercise. In fact, we are skirting both a utility value theory and a value theory that appeals to a notion of value as it pertains to the activity of thinking (although we know of no specific term for this, we could cash this out as human flourishing, self-realization, self-actualization, and the exercising of our humanity, dignity, and respect for personhood; perhaps the closest notion is eudaemonia). In our conclusion, we return to this point.

We reconstruct the labor theory of value as a theory that seeks to address the question of what determines the value of something (a good or service). Importantly, this is not to be confused with the question of what determines the price of something. As economics takes place in the real world, there are often events or extremities that determine a price; for example, a famine may exponentially increase the price of grain and a heatwave may massively increase the price of air conditioning. We are not interested in these kinds of circumstances and will refrain from thinking about the question of price and value. Rather, we will discuss things in terms of normal circumstances (establishing equilibrium over indefinite periods) and we will conceptually appeal to the abstract stream that relates value to labor (rather than engaging in scholastic debates regarding the veracity of such classical economic theories). We are assuming a social and communal context, and will not explore subjective value that may be ascribed to a particular person in a particular circumstance (Jamal may be willing to value, at a non-equilibrium value, a tennis ball because he ascribes a nostalgic subject value to that particular thing).

In highly simplistic terms, the labor theory of value is a claim that the value of something is proportionate to the socially necessitated labor required to produce it. Of course, the thing being produced has to be necessary/required in some sense. For example, it will take a tremendous amount of labor to crush large rocks into smaller ones; however, without the context of those smaller rocks being used, say in building materials that need to be of a particular (small) size, the relationship of labor to value would not exist. Indeed, stated in terms of what is socially necessitated, labor can be thought of as the bringing about of a good (where what is socially necessary is good and therefore valuable). In other words, the value of the labor is contingent on the bringing about of that good. For example, food is a socially necessary good that can be thought of as brought about via farming, where farming represents the labor according to which the good of food is valued.[19]

Our purpose in discussing this is to use it as an example of intentionality playing itself out in terms that are related to the capture of value. Here, that food is valued is an attitude, and that farming is required to bring this about is an action. In other words, labor is a capture of value in the form of the food produced (food can be thought of as an expression of this value; its concretized form). Given that the attitude that determines the value is assumed to be socially necessary (perhaps we can appeal to basic needs such as nourishment), let us focus on investigating the action component. We asserted that action is the practical process of bringing about an end (where the end is the value that seeks to be realized). In the action dimension of labor, there are two parts:

● Physical labor: this is the work required to bring about the end that is sought with respect to the material world. Water needs to be physically directed to the plant, the soil
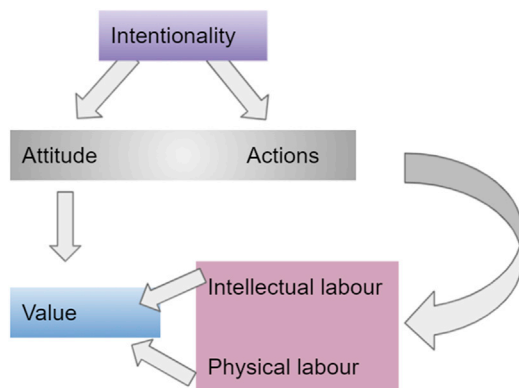
Figure 1. Outline simplifying the process of actioning an intention

Figure 2. Identification of where mechanization automates physical labor

needs to be physically prepared, the harvest needs to be physically collected. This is something that simply needs to be done in the physical world through an exercise of force.
● Intellectual labor (know-how): this is the work required to solve the problem of how to bring about the end that is sought. A plant needs to be watered at particular times of the day, it is to be cultivated in particular soil, and it is to be harvested at particular times of the year. This all requires intellectual labor. This is usually something that is learned through experience or is taught, but the knowledge (in terms of know-how) would have required individual and/or collective effort (intellectual activity) to have come about as a form of know-how.

The partition of labor into physical and intellectual is quite similar to the partition of knowledge into explicit and implicit that is done for the industrial context, where the implicit knowledge is totally unexpressed and represents the know-how acquired by experienced workers (also known as tacit knowledge). Indeed, the Nonaka-Takeuchi model[20] is used to postulate how knowledge in an organization is created through continuous social interactions of tacit and explicit knowledge. This tacit knowledge is quantifiable only indirectly unless it is transformed into explicit knowledge and captured into organizational knowledge in the enterprise. Thus a farming machine is animated through explicit intellect/knowledge (what we can learn from manuals) and tacit/knowledge (what we have learned by ourselves through experience, socialization, and externalization with other workers), but only the explicit knowledge can be quantified. Actually, the value of goods produced is made of tacit and explicit labor/knowledge combining learning and explicit knowledge. See Fenoglio et al.[21]

Through this division we can point to derivative forms of value. Farming (physical labor) is a capture/expression of the value that is also tethered to the ultimate value (i.e., the good), which again is food. The know-how of how to farm (intellectual labor) is a capture/expression of value that is tethered to the ultimate value (i.e., the good), which is food (Figure 1).

In this scheme, the ultimate value determines the value of actions that are required to bring it about. The actions that bring about the ultimate value are therefore valued (derivative value) and those that do not are not valuable.
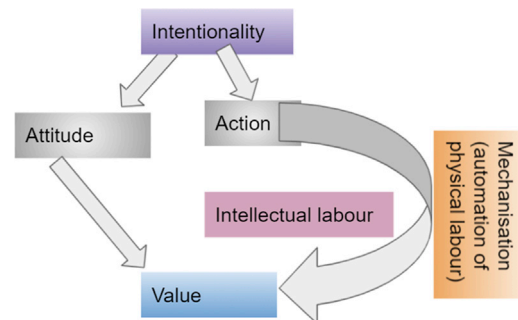
Continuing with the example of farming, we can think about the advent of mechanization; i.e., with industrialization, farming machines were invented that replaced the physical labor dimension of the active component of bringing about a value. In this sense, the farming machines were expressions/captures of the value of physical labor (the machines are concretized captures of the value associated with the intellectual labor required to manifest them). In this case, they replace the role of physical labor with humans and occupy the derivative value that the human labor had captured. In other words, the action dimension of labor has been changed with respect to physical labor through mechanization; however, note that the know-how is still very much a derivative value that remains within the purview of humans. In fact, the invention of the machine is an expression of such intellectual labor: it is an expression of a superior (in terms of efficiency and scale with which to produce the desired good) form of know-how. Note also that action can now be thought of as collapsing solely into intellectual labor (Figure 2).

By discussing mechanization as an expression of intellectual labor, rather than in terms of the loss of physical labor, we seek to point to the fact that the value being captured in the good is still inextricably dependent upon the derivative value of human intellectual labor. The farming machines are brought about, designed, and used (i.e., animated) through the human intellect.

With respect to digital technologies, the question is to be posed that investigates whether they are simply analogous to the mechanization of physical labor. We will argue that, at a base level (concerning primitive data), this is the case and that, with respect to more sophisticated forms (the digital assets derived from novel digital technologies), this is not the case.

Basic forms of digital assets

Let us begin with respect to cases of data as a digital asset that do appear analogous with mechanization of physical labor. In order to do this, allow us to appeal to the view that technologies are extenders, the implications of which are discussed by Hernández-Orallo and Vold.[22] According to this view, technologies augment and/or extend the physical limitations of our bodies. In the case of the farming machines, the capacity of the machine with respect to its speed and the force it can exert is certainly an extension of what physical labor can do *to a lesser extent*. Indeed, it is particularly instructive to speak of extension here
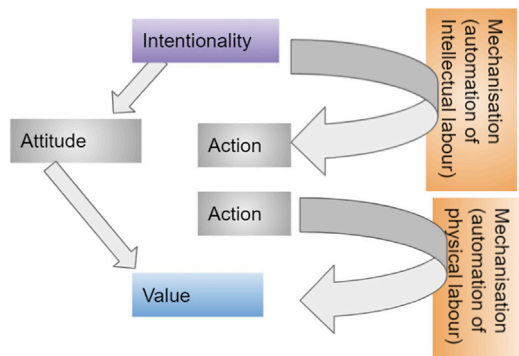
Figure 3. Identification of mechanization of intellectual labor

given that the acts are the same in essence that a human would do physically (i.e., prepare the land, harvest), but amplified and extended.

At one level, some digital technologies can indeed be described as extenders and what they replace in physical labor (above stated in terms that align with the limitations of human physical capacity) is the limitation of certain cognitive functions. For example, we can think of the human capacity to recall and record information: compared with digital data and infrastructure, humans have limited capacities in this regard and digitization can indeed be thought of in terms of mechanization and extenders of human limited capacity; although these are cognitive functions, the assumption is that these cognitive functions are in some sense limited by the embodied mind (i.e., physicality). At this most basic, primitive level, digitization is certainly a continuation of mechanization and derivative of human intellectual labor through the design and creation of digital infrastructure for data (which, in this context, we can think of as a primitive digital asset). It is akin to the mechanization of physical labor (technological extenders). Indeed, the understanding of the digital as primitive data aligns with the notion of intrinsic value, namely richness of data (in our discussion of meaning and value), and ontological categorization; this is because the value of the data is understood both in terms of what is being expressed (its richness; more technically, its propositional content) and in terms of its use (value in terms of utility).

Digital assets of novel digital technologies

Moving beyond the mechanization of physical labor, we believe that more sophisticated forms (expressions) of data that are derived from novel digital technologies present a mechanization of intellectual labor. To reflect this, we believe that the generation and use of these data could be conceptualized as a step change (a discontinuity in the story of value capture) in the relationship between humans and technology. As a point of departure, we should note that we take intellectual labor to fall within relational ontology, insofar as the action of intellectual labor (as we have specified in practical terms) is necessarily directed to something other than the intellectual laborer (indeed this is trivially true). This relates to thingness/ontology. Furthermore, the relation between the intellectual laborer and the task (which is to action; i.e., determine the process by which the intention is realized in the end, which is the generation of the ultimate value), is a subjective relationship. This relates to, and indeed it is necessarily relational

and falls into, the extrinsic category of the relationship between expressions of meaning and value.

In the past, this kind of relationship was impossible to express or capture through mechanization primarily because the technologies were limited to the technologies of the age (think of the development of tools, processes, and machines with respect to physical labor). Indeed, relating to human intellectual activity (what it is to think as a human; i.e., human cognitive capacities and functions, inference, memory), the technologies that have existed could only be utilized to a limited extent. Think of writing, of recording on parchment and paper, as indeed captures of intellectual labor that are limited in very tangible and physical ways, hence why we have thus far discussed more basic forms of digital assets as an extender to these. It is worth asserting our view here that we read human innovation as moving toward a high level of abstraction of labor. We call "technologies of the age" all those technologies that were only capable of extending physical labor and providing the lowest level of abstraction with the transformation of physical objects. Transport networks, fossil fuel extraction infrastructures, are the next level of abstraction with the transformation of physical objects and ontological structures. AI systems represent a higher level of abstraction with an initial transformation of knowledge into digital assets. We may speculate that the next level may be represented by distributed ledger technology (DLT) for a blockchain of knowledge. A blockchain-enabled token economy will efficiently and transparently incentivize and coordinate an integrative approach to fuel collective intelligence for contributing work, infrastructure, management, governance, arbitration, and exploitation of projects for creating value.

Distinct from what we referred to as technology developments that relate more directly to physical processes is the kind of intellectual activity that is higher order; the use of rules, decision making, and evaluation, for example. This has remained exclusively the purview of human intelligence and uncaptured through extenders. In the primitive data forms, for example, recalling and recording information, the value capture is reducible to the value in terms of utility (at best) and aligned with the mechanization of physical labor (Figure 3). The realization of an end requires concepts of problem solving and means and ends (higher-order intellectual labor), rather than simple basic functions such as recall and record, which may certainly be utilized in the problem-solving process (Figure 4). To capture high-order intellectual labor would indeed be *sui generis*.

Our argument is an attempt to show that as yet there is no mechanism to capture intellectual labor. In the following section, to make our argument that new digital technologies represent, for the first time, the capacity to capture higher-order forms of intellectual labor, we will case study AI and argue that the digital assets that are derived therein are expressions of values that capture functions traditionally the exclusive purview of high-order human intelligence. In other words, the derived digital assets associated with these respective technologies are to be understood as expressions/captures of value, namely higher-order forms of intellectual labor.

AI

AI is a broad term that encompasses a host of technologies. These technologies span from more simplistic techniques such
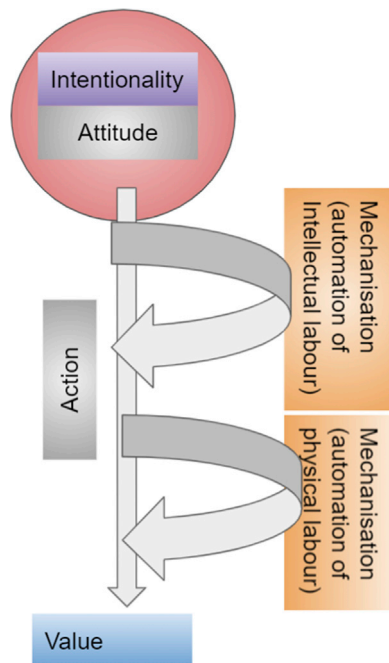
Figure 4. Mapping of what an intuiting machine would look like

as linear regression to more sophisticated techniques such as deep learning. It is outside the scope of this paper to cover all of AI concerning philosophical AI or the philosophy of AI, or the techniques used in AI.[23] For the purposes of our discussion, we can use the term AI as a synecdoche for neural networks and deep learning (DL). That is about capturing a phenomenon understood in terms of replicating processes that were traditionally the purview of human intelligence for solving a problem (or, stated in more basic terms, generating what a human agent can read as an insight). More explicitly, AI is a term used to describe the intelligence demonstrated by machines to solve intelligent problems. AI programs may mimic or simulate cognitive behaviors or traits associated with human intelligence, such as logical reasoning, problem solving, object detection, and learning. However, it does not subsume human intelligence or replicate processes of human intelligence. A neural network has nothing to do with biological neurons.[24] A truly human-like AI would be as useless as a truly pigeon-like aircraft. We note this because the term is highly anthropomorphic and emotional, which, at least imaginatively, brings us relatively too close to thinking of AI in terms of replacing/directly replicating genuine human cognition and consciousness (and to the apocalyptic exaggerations of Bostrom);[17] this is an entirely different investigation. Because we are interested in the activities that AI can represent and that those activities are cognitive (e.g., knowledge representation, inferential reasoning), we use the most apt analogy or description, namely the form of human reasoning it resembles but without limiting the scope of AI to the study of how to mimic human behavior.

By way of illustration, consider an example of a more specific DL system. We can think of a doctor that has become a neurologist specialized in stroke detection. Through time, the neurologist improves at determining which particular kind of stroke (what caused it) is at play. The key concern is what is referred to as improvement and how such an improvement occurs. As a human thinker, the neurologist will be drawing on a host of abilities (such as episodic memory, inference, and deduction) in a complex of cognition; this is a high level of sophistication. If we were to make an analogy with DL, we could say that the neurologist is optimizing for stroke cause detection and that the accuracy increase, with which the doctor makes the correct judgment, represents a decrease in the error rate of the neurologists' judgments (i.e., how often the neurologist gets it wrong). In this analogy, DL can be thought of as replicating this process through experience. In DL terms, this may mean more data analysis, larger datasets, greater computational power, better models, more training, which is the analogy to sophisticated human learning referred to in terms of doctors improving their accuracy—the accuracy increases for what it is optimized for.

However, this analogy only works on the surface. We have described both the doctor as improving and the DL as improving in terms of accuracy. We note that the accuracy of a prediction is the ratio between all correct predictions and all possible cases. It is appropriate in a large sense for describing how the doctors improve their skills, which has been described in broad terms (as sophisticated cognitive phenomena) and is very much an open question; i.e., we cannot fully explain this phenomenon taking into account only quantitative observations and measurements. For DL—more specifically, for supervised learning—it is preferable to talk of improving the model level of uncertainty to reduce the impact of uncertainties during model optimization and decision making. Uncertainty is a quantitative measure that affects human trust, which can be modelled probabilistically, unlike accuracy, which is a distance or closeness to the true value. The doctor may infer from seeing similar symptoms present themselves with particular strokes and generate a rule that those symptoms indicate a particular stroke, or that the accuracy of the stroke cause detection increases by asking particular questions in a particular order and therefore the insight is in the questions to be asked in the particular order that they are asked. Indeed, it does not matter how this occurs so long as the accuracy increases, given the end of accurate stroke cause detection is sought. In the case of AI, there are explicit processes and literature facilitated by engineers and high-performance computing systems that we can concretely point to in terms of reasons for improvement. Indeed, this is how the scientific/engineering literature progresses in terms of science and methodology. How the AI does what it does (in terms of decreases in error rate) is of primary concern precisely because the AI in its practical incarnation represented by DL and supervised learning is informational (meaningful, therefore knowledge) but without real understanding. Contrastingly, the improvement of the neurologist is closer to causal reasoning. A neurologist is a cognitive agent who can express informed judgments. An AI system (in this case, a DL supervised learning system) is a black box for solving narrow problems making it hard to understand how a decision has been made. The paradox here is that AI went from getting computers to do tasks for which humans have explicit knowledge, to getting computers to do tasks for which we (humans) only have tacit knowledge. Learning tacit knowledge from data, without any explicit knowledge taken from humans,
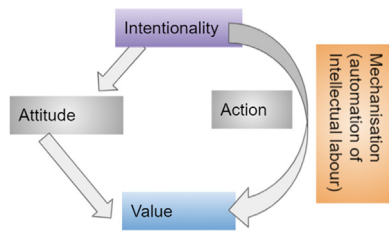
Figure 5. Identification of AI as mechanization of intellectual labor

is a source of interpretability, bias, and robustness concerns.[25] It is even not completely true that the more data, the better the accuracy will be, since the aleatoric uncertainty—the irreducible uncertainty in data—will not be affected. That said, suppose that the performances of a DL model for classification can be related to its accuracy. Classification accuracy alone is not enough information to make an informed decision. Sometimes it may be even desirable to select a model with a lower accuracy because it has a greater predictive power on the problem. This is typically true for class imbalance cases that can be solved using other measures such as precision. It is also well known that increasing privacy protection requires decreased accuracy. Economists recognize this as a resource allocation problem that can be solved in an economic framework where the production of statistics that are sufficiently accurate is balanced against privacy loss.[26] Back to our example, we can say that, in the same way doctor's skills improve and stroke detection accuracy increases through sophisticated cognitive phenomena, the engineers improve the accuracy of the DL model. Thus, the DL model improves, creating value through intellectual labor.

DL architectures have excellent abilities, but, despite the successes in image classification, sequences prediction, and language translation, they should be used with extreme diligence to avoid known vulnerabilities during operation, such as adversarial attacks,[27] or during learning, such as catastrophic forgetting.[28] In both cases, the ultimate decision stays with the human operator, who decides among different options like, for example, model retraining or more advanced learning methods (e.g., self-supervised learning or active learning).[29] Curing by retraining is a cost that implies a more complex infrastructure for better ground-truth datasets and computational resources. Using more sophisticated learning methods does not ensure better results. The unfortunate reality is that AI works only because a plethora of human beings is taking care, controlling, and intervening to mitigate issues and critical problems, to reduce external costs and side effects. At a fundamental level, AI is technical and social practices, institutions and infrastructures, politics, and culture.[30]

If we conceive the doctor's activity as problem solving, by analogy AI can be thought of as a digital asset expressing this problem-solving information, capturing the value that is problem solving. This is not to say that it is operationally the same as human learning and/or improving. Note that problem solving is a perspectival or relational attitude; its meaning being extrinsically generated and imposed upon the data. When designers decide what they want to be optimized, the designers are in effect investing meaning (in the form of the desired outcome) in the data; the process of optimizing (c.f., the doctor getting better at stroke detection) occupies the domain of higher-order intel-

lectual labor that humans have traditionally done. This is the capture of intellectual labor *sui generis*. Importantly, what is optimized and how, in a particular instance, the optimization is realized, is not of concern; the digital technology of AI is an expression of the process of problem solving. The digital asset (e.g., the optimized stroke cause detector) is the ultimate good (rooted in the value of stroke detection, medicine) and is the value captured; the problem-solving process that the algorithm undertakes is a derivative value tethered to this ultimate value, and this is precisely the capture of intellectual labor (Figure 5). Intellectual labor is being automated to determine the means to the end. There is nothing autonomous that follows the mental processes of a human being in terms of setting the value (in this case, stroke determination): the current AI is still about patterns matching and classifying observations.[31,32]

This is not akin to a technological extender that captured the automation of physical labor. Human intelligence is required to animate the machines. More simplistically, as soon as the hand is removed from the hammer, the hammer becomes functionally irrelevant; human intelligence is required to animate/direct the hammer. AI is the invention of a hammer capable of determining which nail needs to be struck (here, the concept of relevant nail is incorporated into the hammer). This mimics a capture of the aspect of intentionality inextricably related to means-end human intelligence (that the automation of physical labor does not reflect). In terms of the digital asset derived from AI, we can state it as a store of this kind of value. Indeed, one can use the same hammer used by Michelangelo, but one will not be able to sculpt the David. An AI-driven hammer will be able to sculpt a replica after proper training with massive datasets. An AI system is made of different subsystems or tools (e.g., face recognition tools, language translation tools, object detection tools) each with a very specific functionality. It is indeed possible to compare hammer tool with face recognition tool from a functional perspective. The difference is that the hammer is a physical extender of a hand to strike a nail, while a face recognition tool is animated by the intellectual labor used for learning the model from data.

In this respect, the current form of AI is still a technological extender for transforming huge masses of data into actionable items that can be consumed by humans. In this sense, an AI system is not so different from an automatic controller that produces the same set of outputs given a certain set of inputs using a transfer function computational algorithm for mapping inputs to outputs. The difference is that the AI system learns the model directly from data under certain contexts during the training phase and relies on certain assumptions, leading to misoperation when the training assumptions are not met. This is what we refer to as *sui generis* value capture. For example, a dataset used for training a cybersecurity system for malware detection is only valid at a given time and in a given possible world, after which it is too unsafe to be useful and requires retraining with trustworthy data. As such, the value captured is very much time dependent (which we may also state in terms of human know-how being time dependent).

Finally, it is of critical importance to state that, when the assertion that AI is a capture of value is made, it is to be understood within the framework of relational ontology; it is not the AI as the expression of meaningful information in and of itself.

Rather, it is in the metadata and in all the meta-learning, tacit knowledge, and ontologies that force the AI system to both be meaningful and, in practical terms, behave properly. In other words, an AI system does not express value (create and capture value) in isolation of human value ontologies but still operates and interacts with a vast society of cognitive agents (c.f. Minsky's society of mind)[33] for capturing value so that machine epistemology and human cognition can form an organic interface, creating a new system of human-machine cognition in harmony.[34]

CONCLUSION

In this article we have investigated the question of what value is being expressed by an algorithmic system, which we have read as a species of a digital asset. In doing so, keen readers will note that we have assumed that technologies are artefacts that are value laden, which, according to our reading, are both defined by human value ontologies and embody (which we have referred to synonymously as express) value that is ascribed to aspects of human cognition. More precisely we have drawn attention to the aspect of rationality in relation to volition, both in terms of practical reason and human labor. Our main takeaway is to invite the reader to consider the broader notion of novel digital technologies such as AII or CDLT, as a representation of the capture of value *sui generis* (compared with traditional technologies), and that this may be is a step change in the capture of value vis à vis the emergence of novel digital technologies. We hope that this piece is received with a view to stimulating a wider, more foundational debate on the nature of artefacts produced by novel digital technologies.

Before closing, we will do two more things. First we will note some limitations of our study, and second, we will offer some more speculative points concerning the implications of our excursion.

Limitations

We recognize that this is an ambitious project and that the treatment provided above serves as a point of departure to other explorations. Indeed, we recognize that the study will need to be expanded and situated within other hermeneutic traditions. Core limitations relate to the extent to which our argument regarding *sui generis* value capture stands and how well the conceptual claim can be applied to the digital technologies mentioned other than AI; i.e., how well does this argument stand for DLT and what exactly is the value expressed in a new type of economy that emerges, as stated by Mance Harmon, co-founder of Hedera Hashgraph: "A marketplace of things commerce made up of people, computers, and things that it is entirely distributed."[35]

Socio-political speculations

Our points center on socio-political implications, labor and value change, the relations of production, and the relationship between labor and meaning. The underlying point in these interventions is that the step change from physical labor value to intellectual labor value will constitute a discontinuity in our material relations, potentially forming a *sui generis*. Herewith, we also propose further research priorities to better understand this step change.

- First, the evaluative conception of algorithms we propose here, in our view, has interesting potential implications for our social and political relationship to AI. The first is that, insofar as the value of an algorithm is relational, its value depends on its position in a complex of relationships. This claim might seem trivial, but it is important nonetheless, because it invites further investigation into the relational conditions that give an algorithm its value. In other words, this claim suggests that an algorithm cannot be conceived as valuable in its own right, and so we must investigate the relative ends and objects that imbue it with relational value. To return to the Michelangelo example, it is only in relation to our existing aesthetic values that an intelligent hammer would have value, and so key to our understanding of the hammer's value is understanding its relationship to these extant sources of value.

- Second, consider our conception of an algorithm as a store of intellectual value. The step change in focus from physical labor (in automation systems) to intellectual labor (in algorithmic systems) means that we cannot straightforwardly extrapolate from previous technological transitions if we are to speculate about the social and political changes that algorithmic systems will produce. Consider two senses in which this might be true. First, intellectual labor is generative: it does not simply create value by solving existing problems (as physical labor typically does), but creates new fields of value. By analogy, laboring in the field might create value by meeting existing needs (i.e., creating food to satiate hunger), but the intellectual labor of inventing new recipes generates new experiential values the value of which does not depend on solving existing problems or meeting prior needs. Next, consider how this form of value creation corresponds to the meaning that individuals gain from work. To use Hannah Arendt's exposition of human activity, automation hitherto has only relieved our need for labor, in the limited sense that denotes activity undertaken for subsistence.[36] However, the expansion of automation into the realm of intellectual value begins to overlap with the human activities of work, through which we shape the material world, and action, through which we shape the social world. It is these latter activities through which Arendt thinks we best understand our source of value and meaning, and so the relationship between our sources of value and the incursion of automation into these domains should be considered carefully.

- Third, how we conceive of the value generated by algorithmic systems should also inform our conception of the relationships between capital, algorithmic systems, governments, workers, and consumers. The automation of intellectual labor value opens new domains of surplus value for capital to accrue, domains of value that, given the generative quality of intellectual labor, are potentially much more expansive than the domain of surplus labor value. Workers and governments will have to respond to

the encroachment of automation into the domain of intellectual labor, as this step change presents new and unique challenges to the security and value of human labor.

AUTHOR CONTRIBUTIONS

E.K. led the writing and conceptualization. E.F. wrote and contributed to sections on information theory and led the section on AI. M.T. wrote the conclusion. C.M., P.G., A. Gwagwa, K.P.-P., and A.K. contributed through workshopping and conceptualization. A.H., M.T., and A. Gilbert edited and workshopped. D.A. contributed to conceptualization and via information theory.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Kazim, E., and Koshiyama, A. (2020). A high-level overview of AI ethics. SSRN Electron. J. *2*, 100314. https://doi.org/10.2139/ssrn.3609292.

2. Harman, G. (2010). Technology, objects and things in Heidegger. Camb. J. Econ. *34*, 17–25. https://doi.org/10.1093/cje/bep021.

3. Gutland, C. (2018). Husserlian phenomenology as a kind of introspection. Front. Psychol. *9*, 896. https://doi.org/10.3389/fpsyg.2018.00896.

4. Bergmann, G. (1960). Ineffability, ontology, and method. Philos. Rev. *69*, 18. https://doi.org/10.2307/2182265.

5. Parker, S. (2020). Summary of Augustine's De Trinitate. Park. Pensées. https://trendsettercase.wordpress.com/2020/01/13/summary-of-augustines-de-trinitate/.

6. Verbeek, P.-P. (2005). What Things Do: Philosophical Reflections on Technology, Agency, and Design (Penn State University Press).

7. Verbeek, P.-P. (2011). What Things Do: Philosophical Reflections on Technology, Agency, and Design (Penn State University Press).

8. A. Newen, L. De Bruin, and S. Gallagher, eds. (2018). The Oxford Handbook of 4E Cognition, 1st ed. (Oxford University Press).

9. Metz, T. (2021). African reasons why Artificial Intelligence should not maximize utility. In African values, ethics, and technology: Questions, issues, and approaches (Palgrave Macmillan), pp. 55–72.

10. Arthur, G., Kazim, E., and Hilliard, A. (2022). The role of the African value of ubuntu in global AI inclusion discourse: a normative ethics perspective. Patterns *3*, 100462.

11. Gwagwa, A., Kazim, E., Kachidza, P., Hilliard, A., Siminyu, K., Smith, M., and Shawe-Taylor, J. (2021). Road map for research on responsible artificial intelligence for development (AI4D) in African countries: the case study of agriculture. Patterns *2*, 100381. https://doi.org/10.1016/j.patter.2021.100381.

12. Hilliard, A., Kazim, E., Bitsakis, T., and Leutner, F. (2022). Measuring personality through images: validating a forced-choice image-based assessment of the big five personality traits. J. Intell. *10*, 12. https://doi.org/10.3390/jintelligence10010012.

13. McKemmish, S. (2005). Traces: document, record, archive, archives. In Archives: Recordkeeping in Society (Elsevier), pp. 1–20.

14. Yeo, G. (2007). Concepts of record (1): evidence, information, and persistent representations. Am. Arch. *70*, 315–343. https://doi.org/10.17723/aarc.70.2.u327764v1036756q.

15. Floridi, L. (2013). The Philosophy of Information (OUP).

16. Floridi, L. (2018). Soft ethics, the governance of the digital and the general data protection regulation. Philos. Trans. R. Soc. Math. Phys. Eng. Sci. *376*, 20180081. https://doi.org/10.1098/rsta.2018.0081.

17. Shannon, C.E. (2001). A mathematical theory of communication. ACM SIGMOBILE Mob. Comput. Commun. Rev. *5*, 3–55. https://doi.org/10.1145/584091.584093.

18. Lombardi, O., Holik, F., and Vanni, L. (2016). What is Shannon information? Synthese *193*, 1983–2012. https://doi.org/10.1007/s11229-015-0824-z.

19. Zuniga, G.L. (1998). An Ontology of Economic objects:.: an Application of carl manager's Ideas. Am. J. Econ. Sociol. *58*, 299–312. https://doi.org/10.1111/j.1536-7150.1998.tb03474.x.

20. Nonaka, I., and Takeuchi, H. (1995). The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation (Oxford University Press).

21. Fenoglio, E., Kazim, E., Latapie, H., and Koshiyama, A. (2022). Tacit Knowledge Elicitation Process for Industry 4.0. Discover Artificial Intelligence 2, 6.

22. Hernández-Orallo, J., and 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 (ACM)), pp. 507–513.

23. Bringsjord, S., and Govindarajulu, N.S. (2018). Artificial intelligence. In The Stanford Encyclopedia of Philosoph.

24. Jasanoff, A. (2018). The Biological Mind: How Brain, Body, and Environment Collaborate to Make Us Who We Are (Basic Books).

25. Kambhampati, S. (2021). Polanyi's revenge and AI's new romance with tacit knowledge. Commun. ACM *64*, 31–32. https://doi.org/10.1145/3446369.

26. Abowd, J.M., and Schmutte, I.M. (2019). An economic analysis of privacy protection and statistical accuracy as social choices. Am. Econ. Rev. *109*, 171–202. https://doi.org/10.1257/aer.20170627.

27. Huang, L., Joseph, A.D., Nelson, B., Rubinstein, B.I.P., and Tygar, J.D. (2011). Adversarial machine learning. In Proceedings of the 4th ACM workshop on Security and artificial intelligence - AISec '11 (ACM Press), p. 43.

28. Coop, R., and Arel, I. (2013). Mitigation of catastrophic forgetting in recurrent neural networks using a Fixed Expansion Layer. In The 2013 International Joint Conference on Neural Networks (IJCNN).

29. Bengar, J.Z., van de Weijer, J., Twardowski, B., and Raducanu, B. (2021). Reducing label effort: self-supervised meets active learning. Preprint at arXiv. https://doi.org/10.48550/arXiv.2108.11458.

30. Crawford, K. (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence (Yale University Press).

31. Bengio, Y. (2020). Deep Learning for System 2 Processing.

32. Marcus, G. (2020). The next decade in AI: four steps towards robust artificial intelligence. Preprint at arXiv. https://doi.org/10.48550/arXiv.2002.06177.

33. Minsky, M. (1988). The Society of Mind 6. Pb-Pr (Simon and Schuster).

34. Bai, H. (2022). The epistemology of machine learning. Filos. Sociol. *33*, 40–48. https://doi.org/10.6001/fil-soc.v33i1.4668.

35. Radio, C. (2018). Founder Interview: Mance Harmon of Hashgraph. http://cryptoradio.io/mance-harmon-hashgraph/.

36. Arendt, H. (2013). The Human Condition (The University of Chicago Press).

About the author

Emre Kazim is the co-founder and COO Holistic AI, a start-up focused on software for auditing and risk management of AI systems. He is also a research fellow in the Computer Science Department at University College London and is a thought leader in AI ethics. He is particularly interested in AI governance and regulation and works closely with engineers to translate governance into practice.