Philosophy & Technology (2024) 37:21
https://doi.org/10.1007/s13347-024-00705-3

RESEARCH ARTICLE

Algorithms Don't Have A Future: On the Relation of Judgement and Calculation

Daniel Stader[1]

Received: 21 August 2023 / Accepted: 12 January 2024 / Published online: 15 February 2024

Abstract
This paper is about the opposite of judgement and calculation. This opposition has been a traditional anchor of critiques concerned with the rise of AI decision making over human judgement. Contrary to these approaches, it is argued that human judgement is not and cannot be replaced by calculation, but that it is human judgement that contextualises computational structures and gives them meaning and purpose. The article focuses on the epistemic structure of algorithms and artificial neural networks to find that they always depend on human judgement to be related to real life objects or purposes. By introducing the philosophical concept of judgement, it becomes clear that the property of judgement to provide meaning and purposiveness is based on the temporality of human life and the ambiguity of language, which quantitative processes lack. A juxtaposition shows that calculations and clustering can be used and referred to in more or less prejudiced and reflecting as well as opaque and transparent ways, but thereby always depend on human judgement. The paper clearly asserts that the transparency of AI is necessary for their autonomous use. This transparency requires the explicitness of the judgements that constitute these computational structures, thereby creating an awareness of the conditionality of such epistemic entities.

Keywords judgement · calculation · algorithms · AI · hermeneutics · prejudice · artificial neural networks

1 Introduction: From the Pentagon Papers to AI

In 1971, the US public was shaken by the leak of the Pentagon Papers. A report containing secret documents proved the government had systematically lied to the public about the intentions, the origins and the course of the Vietnam War. That same year, Hannah Arendt published her essay "Lying in Politics: Reflections on

✉ Daniel Stader

[1] Echo – Center for the Study of Rhetoric between Old and New Media, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany

the Pentagon Papers", asking how such a failure of politics was possible. First of all, Arendt criticises the fact that foreign policy has been driven by the purpose of image-building. But she finds it even more problematic that the decision-makers in charge "prided themselves on being 'rational'", moreover, they

> were eager to find formulas, preferably expressed in a pseudo-mathematical language, that would unify the most disparate phenomena with which reality presented them; that is, they were eager to discover *laws* by which to explain and predict political and historical facts as though they were as necessary, and thus as reliable, as the physicists once believed natural phenomena to be. (Arendt, 1972a, p. 11)

Arendt further clarifies that human affairs are contingent and cannot be theorised like natural laws and thus become objects of calculation. By specifying the difference, Arendt introduces the antithesis of judgement and calculation:

> Reading the memos, the options, the scenarios, the way percentages are ascribed to the potential risks and returns – 'too many risks with too little return' – of contemplated actions, one sometimes has the impression that a computer, rather than 'decision-makers,' had been let loose in Southeast Asia. The problems-solvers did not *judge*; they calculated. Their self-confidence did not even need self-deception to be sustained in the midst of so many misjudgments, for it relied on the evidence of mathematical, purely rational truth. (Ibid., p. 37)

Joseph Weizenbaum takes up Arendt's contrast in his book *Computer Power and Human Reason: From Judgment to Calculation*, thus inaugurating a tradition of critique that repeatedly refers to the difference of judgement and calculation. Weizenbaum suggests that this problematic development dates back to "Francis Bacon's misreading of the genuine promise of science", aiming at "an age of rationality […], but with his vision of rationality tragically twisted so as to equate it with logicality" (Weizenbaum, 1976, p. 13). This twist is grounded in the intention to eliminate reality's "disconcerting *contingency*" (Arendt, 1972a, p. 12), which Arendt identifies as the central characteristic of the sphere of human affairs. This tradition of critique assumes that the datafication and hence calculability of human affairs leads to the replacement of judgement by calculation, even if this calculation is only apparently objective.

Criticism of the increasing replacement of judgement by calculation, or the narrowing of human environment and society by instrumental reason, has taken on unprecedented urgency in a world where "algorithms are inevitable" (Schumann & Taddicken, 2021, p. 8). In 1976, Weizenbaum worries that it is even a considered question of "whether or not every aspect of human thought is reducible to a logical formalism, or, to put it into the modern idiom, whether or not human thought is entirely computable" (Weizenbaum, 1976, p. 12). Nearly half a century later, it still seems widespread "to assume that judgment can be understood as essentially reducible to a calculative or quantitative process" (Malpas, 2020, p. 1073). Critics deny that judgement can be reduced to calculative

or quantitative processes. But at the same time they deplore the replacement of judgement by calculation and assert the uniqueness of the human relation to the world that gives judgement an indispensable legitimacy (cf. Cooley, 2007; Malpas, 2020; Raveendhran & Fast, 2021). But if judgement and calculation are so different in nature, how can one replace the other?

This paper argues that far from being replaced by calculation, judgement increasingly relies on calculative patterns such as algorithms. Agreeing with Arendt on its uniqueness, I state that the specific character of human judgement is incommensurable with quantitative processes like algorithmic calculation. It is only judgement – or prejudice – that attributes meaning or purpose to calculation. In short, contrary to Arendt's and Weizenbaum's claims, I reject the idea that judgement is replaced by calculation, but that computational infrastructures alter the range, the transparency and the possibilities of judgement. This adjustment of the critical measure is important to show that AI does not simply do what humans have done before, but that it is embedded by human judgement, thereby changing it and with it the way humans relate to themselves and others. Since the contexts of critique depend on the concrete structure they target – algorithms, specific machine learning processes, AI in general – I will contrast judgement with the epistemic entity of the algorithm, taking it as a primal element of the computational infrastructure, from the Turing machine to AI.

I begin by exploring the concepts of algorithm and judgement in order to highlight the general epistemological differences (section 2). In 2.1, I introduce the most common definitional understandings of what an algorithm is to set the framework of discussion. This section shows that the understanding of what a classical algorithm is already depends on human judgement. Nowadays, when there is talk of AI, in all likelihood it is referred to artificial neural networks (ANNs). That is why in 2.2, I extend this account to connectionist AI by explaining the basics of ANNs and its functionalities. Even though in a different way, ANNs depend on human judgement as well to be provided with meaning and purposiveness. In 2.3 I draw on three philosophical classics to emphasise the unique character of judgement and its contrast to the process of algorithms. I will show that human judgement does not simply rely on generalities such as rules or laws, but requires some purposefulness in order to interpret these generalities. This temporal purposefulness is the basis of human orientation and community, and is given its plasticity by the ambiguity of language, which makes it suitable for the contingency of human affairs. Human judgement, I clarify, has a different quality than algorithmic processes. So in 2.4, I use these insights to postulate three theses on algorithms to show that they depend on human judgement to play a role in the course of life, by showing that algorithms themselves have neither purpose nor meaning unless it is attributed to them externally. This already indicates that the epistemic difference between judgements and algorithms makes them incapable of substitution. Section 3 then shows how judgements can nevertheless depend on structures like algorithms, which can only be used in a prejudiced way. I argue that judgement is not replaced by algorithmic calculation, but that the prejudiced use of algorithms can be reflected or unreflected, depending on the opacity or transparency of the algorithm's conditionality. Therefore, in 3.1 I contrast judgement and algorithmic calculation to highlight the difference in terms of the

inherent temporality of judgement that calculation lacks. This demonstration serves to show that the integration of an iterative and inherently complete structure as an algorithmic process into the human lifeworld is a performance of human judgement, because it requires the attribution of qualities that it does not possess by itself. In 3.2, I distinguish between different types of the use of prejudices to show how algorithms can and cannot be used. I concentrate on the requirement of transparency, i.e. the disclosure of the basic assumptions as well as the selection and the extent of data that form the algorithm, to raise awareness of the conditionality of AI. This requirement is very briefly demonstrated using the example of the ethical AI tool Delphi.

2 Digging into Concepts: On Algorithms, Artificial Neural Networks, and Judgement

The issue of AI often revolves around how it replaces human judgement. But is this really the case? Before considering the relationship between judgement and calculation, it is necessary to clarify the nature of the most discussed computational structures and of judgement itself.

2.1 Definitions of Classical Algorithms

Although the principle of the algorithm is thousands of years old (cf. Knuth, 1972), the term has gained popularity in the context of discussions about machine decision making. Weizenbaum understands an algorithm as "an *effective procedure*", defined by "a set of rules which tells a player precisely how to behave from one moment to the next" (Weizenbaum, 1976, p. 46). These rules ought to be complete and consistent, i.e. they should provide a specific instruction for any given case, assuming that neither the rule nor the case leaves room for ambiguity. They work with symbolic representations that have to be well defined. Cormen et al. describe an algorithm as "any well-defined computational procedure that takes some value, or set of values, as *input* and produces some value, or set of values, as *output*. An algorithm is thus a sequence of computational steps that transform the input into the output" (Cormen et al., 2009, p. 5). Harari also notes that an algorithm is not a calculation itself, but "a methodical set of steps that can be used to make calculations, resolve problems and reach decisions. An algorithm isn't a particular calculation, but the method followed when making the calculation" (Harari, 2016, p. 97). It is important to keep in mind this distinction between an algorithm, its implementation (Moschovakis, 2001, pp. 934f.) and a calculation based on the algorithm, the algorithmic process. While Knuth states that an algorithm is "a finite set of rules that gives a sequence of operations for solving a specific type of problem", which "has five important features" (Knuth, 1997, p. 4), namely finiteness, definiteness, input, output and effectiveness, for others the input and output components do not belong to the algorithm itself: "For an algorithm to be considered valid it must have three characteristics: (1) it should be finite; (2) it should have well-defined instructions; and (3) it should be effective" (Schuilenburg & Peeters, 2021, p. 3). Hill gives a more elaborate

definition, claiming that "[a]n *algorithm* is a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given conditions" (Hill, 2016, p. 47).[1] She explicitly links the effectiveness of algorithms to the fact that they require "no judgement […], learning, insight, or understanding", that they are "devoid of open questions" and "not subject to interpretation" (ibid., p. 45), suggesting rather definiteness than effectiveness. Furthermore, in contrast to the previous definitions, she surprisingly adds purposefulness – although an algorithm may not 'have' a purpose in itself, we cannot understand it "without grasping its purpose" (ibid., p. 46). Algorithms in themselves must be understood by something that lies outside of them, that is, a context that provides purposefulness. Even if someone could perfectly describe the algorithmic processes of a traffic light, this description is not sufficient to understand what a traffic light *is* or what its calculation actually *does*.

2.2 Artificial Neural Networks: Supervised, Reinforced, and Unsupervised Learning

The history of AI is a history of two competing models, symbolic and connectionist AI.[2] The first model describes implementations of the classical algorithm as mentioned in the last section, which is a series of well-defined steps for processing symbolic representations. Connectionist AI, in the form of artificial neural networks (ANNs), on the other hand, consists of self-organising layers of cells that are able to process probabilistically, i.e. beyond discrete definitions of the input data. Although the idea of connectionist AI is about as old as the classical AI, it has only become practically useful with recent advances in computing power and the availability of large amounts of data (cf. Anderson & Rosenfeld, 1988; Bechtel, 1996, pp. 271f.). Working in a very different way, the structure of their process is often also called algorithm (cf. Alpaydin, 2016, p. 108; Goodfellow, Bengio & Courville, 2016, p. 96; Bechman & Bowker, 2019), although it lacks the definition and discreteness of the original use of the term.

An ANN that is capable of advanced learning processes consists of multiple layers of so-called neurons, processing cells, i.e. computational units, which are connected to each other and in various cases also to themselves (cf. Haykin, 2009, pp. 22–24). These connections of neurons to themselves or to their own layer in general also provide short-term memory, because the state of a neuron at a given time is influenced by its own state before and affects its state in the future (cf. Alpaydin, 2016, pp. 92f.). Moreover, instead of the simple approach of excitatory and inhibitory connections that forward or stop an impulse, each of the multiple input connections of a neuron can be weighted differently (cf. Aggarwal, 2018, pp. 1f.). This

[1] 'Imperatively given': "Until we issue a command, or order an action, we have not conveyed an algorithm. Although the articulation is not the algorithm itself, the algorithm itself being the procedure is the imperative verb" (Hill, 2016, p. 50).

[2] There are several attempts to combine the advantages of the two (cf. d'Avila Gacez & Lamb, 2023), but I will leave those aside here.

allows the input data to be processed in a highly abstract way, as these weights can be continuously adjusted, depending on their previous influence and accuracy. This is also where the learning of neural networks takes place, even more with each layer of neurons, processing input data at increasingly abstract levels (cf. Alpaydin, 2016, p. 103). This quality of ANNs to deconstruct and reassemble input data in an opaque, non-conceptual process leads to the problem of their interpretability (see 3.2).

ANNs can be trained supervised, by reinforcement learning, or unsupervised. Supervised learning often uses backpropagation. If the result (e.g., the recognition of objects in images) is not considered accurate, the network is traced back to the most influential neurons to change their state in the process (cf. Rumelhart, Hinton & Williams, 1986; Garcia Rosa, 2019, p. 30). Supervised learning always relies on training sets where examples have already been labelled "by an instructor or teacher who shows the machine learning system what to do" (Goodfellow, Bengio & Courville, 2016, p. 102), and the machine is trained to categorise them correctly, so it learns to do so on its own with cases that have not yet been labelled. But there are forms of learning that do not require such supervision, either by a person or an algorithm. An ANN can learn by reinforcement in the interaction with an environment, by trial and error in solving an explicit problem or completing an explicit task, and be rewarded for achieving accurate results. Recurrent successes will strengthen connections between neurons that happen to be activated repeatedly at the same time – such strengthening of correlated connections as a 'reward system' is the core of reinforcement learning (cf. Sutton & Barto, 2018). In contrast, so-called unsupervised learning does not require external feedback such as backpropagation or reward. Its advantage is that it can deal with large amounts of unlabelled data and detect patterns or regularities (e.g. through clustering, cf. Alpaydin, 2016, pp. 112f.) that are unknown to programmers or users. It "can be thought of as finding patterns in the data above and beyond what would be considered pure unstructured noise" (Ghahramani, 2004, p. 73). This means

> to find the 'best' representation of the data. By 'best' we can mean different things, but generally speaking we are looking for a representation that preserves as much information about x [the input] as possible while obeying some penalty or constraint aimed at keeping the representation *simpler* or more accessible than x itself. (Goodfellow, Bengio & Courville, 2016, p. 142)

Such representations "can be used for decision making, predicting future inputs, efficiently communicating the inputs to another machine, etc." (Ghahramani, 2004, p. 73). For example, insurance companies can find correlations between occupations and diseases, or online shops can target specific clusters of customers with specific advertisements, because they match a specific consumer profile.

When considering all types of ANN processes, it becomes clear that they also depend on external contexts of purpose. Supervised and reinforcement learning work with direct feedback that is grounded in human evaluation of desired or undesired outcomes (i.e.: human purposiveness). And even though the mere process of unsupervised machine learning works without direct human intervention or involvement, it becomes clear that firstly, data itself is always operationalised in purposeful

contexts, and secondly, the resulting representation has to be reintegrated into meaningful contexts by humans (see 3.). In addition to the specific assumptions of such unsupervised learning ANNs, it is the possibility of "classification", which is "another word for generalization" (Bechmann & Bowker, 2019, p. 4) that is assumed and addressed here (which is a first indication of the merely prejudiced use of their output).

2.3 What is Judgement? – Aiming Commonly at a Problematic Whole

When it comes to the claim that human judgement is being replaced by algorithmic calculation, there is a link to the origins of the concept of judgement more than two and a half millennia ago that is more than an anecdote, but shows a continuity at the heart of human affairs. Looking at recent debates about the (problematic) use of algorithms and their far-reaching responsibilities that can be infiltrated by bias, the two areas most discussed are healthcare (only recently Lix et al., 2022; Oxholm, Christensen & Nielsen, 2022; Rangareddy & Kurpad Nagaraj, 2022; Owens & Walker, 2020; Wiens, Price & Sjoding, 2020) and jurisdiction (Binns, 2022; Johns & Compton, 2022; Sommerer, 2022; Ugwudike, 2022; Slobogin, 2021). The etymological origins of the concept of judgement (though not of the term, which comes from the Latin *iudicium*) can be traced back to the ancient Greek verb *krínein* (*kríno*), the substantive *krísis* and the adjective *kritikós,* all of which share the same root, and all of which refer to a turning point in decision and distinction. *Krínein* has the meanings of to "*separate*", to "*distinguish*", to "*pick out, choose*", to "*decide* disputes" or "a contest", to "*judge*", to "*bring to a crisis*" (medically) as well as to "*bring to trial, accuse"* and to "*condemn*" (Liddell & Scott, 1996, p. 996). In ancient Greek, the substantive *krísis* is also particularly associated with medicine (meaning the critical turning point – or decision – of a disease or injury) and jurisdiction (meaning the judgement of a court as the resolution of a case) (cf.ibid., p. 997). As such, judgement also has a history in rhetoric, which has its roots in Greek forensic speech and continues its tradition from Latin rhetoric to the Enlightenment.[3] Judgement, but also critique and crisis, go back to these conceptual origins, which gather around the act of decision by distinction and/or choice. In order to judge an individual case, to make a decision, it is necessary, especially in medicine and jurisdiction, to refer to a general principle or law and to assess whether the case falls within this generality or not. The same applies to criticism, which always refers to an implicit or explicit (aesthetic, moral, etc.) standard against which a concrete object, person, group or institution is measured.

I will draw on a philosophical tradition of conceptualizing judgement that begins (broadly speaking) with Immanuel Kant's *Critique of the Power of Judgement* (*CPJ*). In the 1950s, two prominent theorists, Hannah Arendt and Hans-Georg Gadamer, interpreted Kant's *CPJ* with antagonistic intentions. Arendt understands Kant's *Critique* as a political legacy that actually deals with judgement as the

[3] I can barely touch the complex history and meaning the term judgement had during its course, cf. Van Zantwijk, Gabriel & Ogorek, 2001.

backbone of community-building, while Gadamer criticises Kant's apparent reduction of judgement to taste and seeks to rehabilitate prejudice as a common tradition of understanding. Despite their conflicting intentions, I aim to provide an integrative account of judgement that combines the strengths of these thinkers. Apart from the fact that Arendt and Gadamer both refer to Kant, they are compatible in terms of my purpose for two reasons: (1) All three of them assume the open-endedness of judgement; Kant clarifies that judgement is that which cannot be reduced to a rule; Arendt understands judgement as the central ability to deal with the unpredictable contingency of human affairs; and Gadamer asserts that judgement is an act that realises the traditional interpretability of the world, which is provided by language, "whose virtuality opens for us the infinity of discourse" (Gadamer, 2004, p. 553). (2) Although Gadamer and Arendt interpret Kant with different agendas, they share a background as Heidegger scholars.[4] On this issue, it is in particular their respective concepts of prejudice that are inspired by Heidegger. In outlining prejudice, Gadamer explicitly refers to Heidegger's concept of understanding (cf. Gadamer, 2004, p. 268; Heidegger, 2010, § 32), while Arendt implicitly refers to the Heideggerian 'They' (cf. Arendt, 2005, p. 100; Heidegger, 2010, § 27). This common reference leads Gadamer and Arendt to the idea that judgement needs to be explored not only in social, but also in temporal terms. Gadamer focuses on the anchoring of judgement in the past, Arendt emphasises its directedness towards the future; and both look – with Kant – at the present act of judgement, which is always performed by the member of a community. Bringing together important aspects of these theorists provides a full account of the temporal rootedness of judgement – and prejudice.

So what is judgement? Kant basically distinguishes between determining and reflecting judgement: Determining judgement is the derivation of something particular from a given general premise or law, while reflecting judgement means that something particular is given that is judged with respect to a generality that is not (yet) given. For example, determining judgement is required of a referee, who has to decide whether something is a punishable handball or just knee-jerk self-protection and he or she does so by exegesis of the valid football rules. The rule book is given as a totality of normative claims, but it contains conditions and characteristics to be determined – and the referee has to decide which concrete actions or situations fall within the scope of the rule book and *are* therefore handballs, fouls, goals or offside. Like the 'dash of salt' in a recipe, actual football matches show that even the most precise set of rules is based on concepts that require interpretation, e.g. whether a movement is 'intentional' or an 'unnatural increase in body surface area'.[5] Reflecting

[4] Heidegger himself is only lightly touched upon here, mainly because he rejects the concept of judgement, associating it with propositional logic, which claims some kind of objective truth (cf. Heidegger, 2010, p. 30–32). This way, judgement would be something "which is valid 'timelessly' in itself" (ibid., p. 151). This obviously does not apply to the tradition of judgement to which I refer.

[5] According to Weizenbaum, the rules of a game that leaves room for interpretation are simply not 'complete': "Another way to state the condition that the rules of a game must be complete and consistent in the sense here intended, is to say that no two referees faced with the same game situation would fail to agree in their judgment. Indeed, 'judgment' is not the proper word, for decision would be reached by the application of logic only. It would, in effect, be nothing more than a determined calculation, a logical process which could have only one outcome" (Weizenbaum, 1976, p. 44). The point is that most games

judgement, on the other hand, begins with something particular and seeks its generality, assuming that there is one. For Kant, this kind of judgement is guided by the principle of purposiveness. It is the search for an unknown purpose of a concrete phenomenon, which is particularly necessary in the life sciences, but it is also the basis for aesthetic judgement (since there is no explicit general definition of beauty on which to base a determining judgement). Reflecting judgement also has an important function in self- and social relations. It allows to reflect on one's emotions as well as one's thinking – one can be happy that the professor broke her leg, because she is not very nice or the exam gets postponed. But one can also reflect on this reaction and condemn this happiness as contrary to the principles of one's values, self-image or social norms (and it is purposeful to reflect on them). For Kant, the entanglement of judgement with social relations lies in the way it is based on common sense. Reflecting judgements, such as aesthetic ones, refer to an implicit standard as if it were given, a standard whose acceptance is also demanded of other people. This is why people discuss and try to persuade each other in matters of taste, although it is clearly subjective. Implicit common standards of judgement are common sense, and to demand the agreement of others requires that one's individual judgement be somehow linked to common sense and thus made communicable (cf. Kant, 2000; Arendt, 1989).

Hannah Arendt also denies that judgement is a mere "logical operation" (Arendt, 1978, p. 215). She draws on the idea of judgement as an essential element of sociability in general. For her, reflecting judgement means reflecting on oneself through the eyes of others, which is what makes communication and togetherness possible in the first place (cf. Arendt, 1961c, p. 220). Judgement is only possible because we are able to reflect on something through the eyes of someone else, be it a person or a community, which is used as its virtual measure. Therefore "judgment is endowed with a certain specific validity but is never universally valid. Its claims to validity can never extend further than the others in whose place the judging person has put himself for his considerations" (ibid., p. 221). As with Kant, Arendt's understanding of reflecting judgement aims at a totality or a general principle that in most cases is not or cannot be given. For her, it is not only the lack of an explicit concept of beauty, but also of 'right' and 'wrong' in the contingent course of living and, more generally, of most categories of human social life. Moreover, the incompleteness of judgement is inevitable with regard to the constitution of human life as temporal. The judgement of human affairs is always provisional because of the openness of the future. Here she raises the opposite concept of calculation: Especially in political matters, human judgement is "confronted with situations for which there are at most precedents, but no general rules" (Arendt, 2022, p. 569), because there is no way of foreseeing the influence of decisions or of public thinking on others and on the public sphere itself (cf. Arendt, 1998, p. 244).

Footnote 5 (continued)
faced with countless possible situations lack this virtual unanimity of referees, unless the moves themselves are principally finite and clearly marked out as in chess.

These approaches of Kant and Arendt need to be complemented by the philosophical tradition of hermeneutics, which emphasises the necessity of prejudice (or prejudgement) in the act of judgement. Gadamer states that it is "*the prejudices of the individual, far more than his judgments,* [that] *constitute the historical reality of his being"* (Gadamer, 2004, p. 278). For him, judgements are always embedded in the frame of reference of tradition, and it is only tradition that gives meaning to human life and is therefore the necessary background for any human act of meaningful orientation. Because tradition cannot be grasped or be made explicit in its entirety, it becomes the non-given totality to which every act of judgement refers. Drawing on Heidegger, Gadamer considers this to be true for every act of human understanding, which "has the structure of [understanding] *something as something*" (Heidegger, 2010, p. 144) and therefore has to rely on meaningful pre-structures in order to make sense of something – e.g. language (cf. Gadamer, 2004, pp. 446, 547f.). These pre-structures as the frame of reference of judgement are addressed, but can not be subjected in this very act, which is why Gadamer claims that prejudice plays a major role compared to judgement. More than that, as judgements happen with prejudices in their background, they are a means to carry these prejudices through time. Heidegger and Gadamer both would not come up with the idea to overcome prejudice in this meaning, because it first and foremost enables understanding of the world.

Even though Arendt has a different concept of prejudice, she agrees with its necessity,[6] especially for the sake of the possibility of community, a common understanding of the world. In a complex world, people are simply incapable of judging every aspect of their lives. Moreover, prejudices with their experienceless generality offer ways of socialising with others by referring to such kinds of commonplaces. But she denies that this is inescapable, given the uniqueness of individual experience, the confrontation with the alterity of other people's experiences through communication, and the openness of the future (cf. e.g. Arendt, 2005). Making one's individual, idiosyncratic experiences communicable, reflecting one's own standpoint through the eyes of others, or starting something new are ways of escaping the givenness of tradition (cf. Arendt, 1989; Arendt, 1998). Kant also provides an antidote to the strong emphasis of predetermination by tradition made by Gadamer: For him, a mature power of judgement is capable of making preliminary judgements, that is, the ability to judge with the awareness that one's judgement is based on and/or aimed at uncertain principles or facts, "the awareness that my judgement was problematic" (Kant, 1966, p. 545).

My position on prejudice for the argument on judgement and calculation in relation to AI lies somewhat between Gadamer and Arendt. Gadamer says that prejudice provides the unreflected light of reflection, Arendt argues that it is possible to get hold of prejudices, that they can be spelled out and outlined in their entirety. I argue that the pre-structures of understanding can indeed be understood as prejudices, which cannot simply be made conscious and transparent by mere reflection, because

[6] Arendt also thinks that understanding is "closely related to and inter-related" (Arendt, 1994, p. 313) with judging.

they usually appear in cohesive clusters, whose connections cannot be made visible in their entirety. Prejudice is a substitute for the active act of drawing a connection between something particular and something general; because the act of judgement cannot draw these connections in every direction (i.e. subsume an individual case under every possible generality; or search every kind of generality for the individual case), but must be specific according to its purpose; therefore, interpretive frameworks must be taken for granted. Purposes determine, narrow the focus of judgement. Nonetheless, the functioning of and the capacity for judgement depends on the occasional reflection and the constant permeability to resonance in the use of these pre-structures. Although I agree with Gadamer that common and traditional interpretive frameworks (as clusters of prejudgements) allow concrete situations and experiences to be subordinated to generally accepted patterns of meaning, I assume that these frameworks are best used provisionally. They are not inescapable, they can be confirmed, rejected, shifted, etc. in their use, as they are applied in a world with changing conditions of application. They can be questioned reflectively as a consequence of experience, of the overdue of success of their application, of discourse (i.e. the arguments and experiences of others). Questioning individual elements in interpretive frameworks can cause chain reactions simply because they consist of interconnected prejudgements/prejudices. I come back to the topic of prejudices in 3.2.

Judgement as an act means making a specific decision about a particular issue, referring to a kind of generality that is assigned with a purpose. Doctor, judge and politician are professions that require judgement, albeit in different ways: The typical medical purpose is health (i.e. promoting the physical quality of life), the judge's purpose justice, while the political purpose differs according to systems and contexts (social justice, prosperity, plurality, national unity, etc.). These purposes, as ungraspable totalities with long-standing social traditions, are the background to the interpretation of explicit generalities such as therapies, laws or constitutions. The judgements of the referee require the interpretation of ambiguous rules according to the purpose of fairness or consistency in the application of the rules (i.e. deciding in accordance with the virtual whole of all previous valid decisions on the same matter).[7] But the life of every individual, as well as the life of every community, requires judgements in order to orient and to shape, to adapt to and to implement itself in the factual world. However, the conditions and foundations of one's own judgement cannot be completely and explicitly transparent, due to the entanglement of human beings with the history of their communities as well as their own history (broadly speaking: tradition and biography). This is not a lack of judgement, but goes hand in hand with its plasticity, which offers judgements the possibilities of adaptation, revocation, expansion, restriction, etc., and its general orientation towards a future that is considered open (cf. Klinger, 2011, pp. 20f., 48). The meaning and the authority of abstract purposes are grounded in a community tradition, and they can be adapted, preserved, shifted or thwarted by acts of judgement. Their application in judgement refers to and relies on their shared understanding, or, with Arendt: "We

[7] Or, as Gadamer would put it, with the idea of the rules (cf. Gadamer, 2004, p. 324).

are contemporaries only so far as our understanding reaches" (Arendt, 1994, p. 323). Judging, then, is an activity that is embedded in a community and is always in some way related to it. This activity, I presume, is not reducible to its result, but it creates, preserves or attacks the net of an implicit social consensus. In mutual understanding, in discussion, even in arguing with each other, there is implicit common agreement of how to judge properly, even though specific cases, asserted generalities or drawn conclusions may be rejected. My integrative account of judgement means this social activity that is dealing with common, but fluent generalities and purposefulness, applying and testing it in concrete experience. Judging is at the same time such a present social activity, a recourse to traditional interpretive frameworks and directed at an open future; it is defined by the perspective of a temporal being that itself is embedded in those structures, and itself is ever-changing.

2.4 Three Theses on Algorithms

In what follows, I will look more closely at the difficulties of defining what an algorithm is. Using the account of judgement, I will show that both the development and the understanding of algorithms depend on judgement. Algorithms will be addressed here in the broadest sense, in the classical idea of AI as deductive processing, and connectionist AI, which focuses on inductive processing. In terms of the basic act of judgement, which involves precisely these ways of grasping something, AI and judgement seem to have the same object and the same function. According to symptoms, the doctor judges the risk or probability of a patient having a particular disease on the basis of her knowledge and experience, just as the algorithm seems to do on the basis of a database with disease taxonomies and/or previous cases of former patients with similar symptoms. Are these two acts not the same?

Moschovakis noted a quarter of a century ago that "there is no absolute notion of *algorithm,* independent of any given" (Moschovakis, 1998, p. 95). The irony of attempting to give a precise, complete and definitive concept of algorithm is, firstly, that it requires contingent judgements about the crucial characteristics of an algorithm, making the concept of algorithm itself ambiguous. Moreover, the definientia themselves are ambiguous and open to debate. For example, the understanding of algorithms as effective procedures becomes problematic because of the ambiguity of the term 'effective' – even Church and Gödel could not find a way to define effective calculability (Blass & Gurevich, 2003, pp. 4f.), let alone the broader notion of a procedure. The category of effectiveness raises the questions of in what way for what purpose something should be (most) effective, and thus already depends on judgement.[8] The value of judgement as a pragmatic act is revealed in the operation of a common and mostly consistent understanding of what an algorithm is, which is not the intensional definition of the concept, but the extensional attribution to concrete

[8] Even Hills notion of effectiveness is ambiguous, because some aspects are pointing at the result and mean definiteness ("devoid of open questions"), others refer to the way of processing ("'mechanical'") (Hill, 2016, p. 45).

definite procedures with finite input and output, provided with purposes[9] and effective in some way. Agreeing with Hill on her definition of algorithms, my first thesis on algorithms is:

Algorithms are always embedded in purposeful contexts and cannot be defined or understood without external references that provide a basis for extensional judgements about whether something is an algorithm, what an algorithm is or what its output actually means.

Since algorithms are abstract procedures that are applied to machines but do not at all originate in their use or design, extensional definitions are often tested against everyday manual instructions to sharpen them. A classic example of the search for rule-following algorithms are recipes: Recipes are considered to be effective procedures, finite sets of rules with an input and an output. For Weizenbaum, a recipe could be an algorithm only if

> two conditions were fulfilled: first, that there exists a language in which precise and unambiguous cooking rules could be stated; and, second, that all people are identical in every respect having anything to do with cooking. These conditions are not independent of one another, for one way in which everyone would have to be like everyone else is that they would all have to interpret the cooking language identically. (Weizenbaum, 1976, pp. 47f.)

Knuth takes up the ambiguity of language and individuality and rejects to identify a recipe with an algorithm because "it notoriously lacks definiteness" (Knuth, 1997, p. 6). Hill does not count recipes, like other instructions for everyday procedures, as algorithms, because they "are not effective except in perversely rigorous cases", for they "involve unstated assumptions". For her, "[t]he learning and enactment of those procedures admits of mistakes, training, and improvement, while the mental grasping of an algorithm is discontinuous. If a set of instructions can be said to be followed *well* or to be followed *badly*, it is not an algorithm" (Hill, 2016, p. 49). The unambiguity that is required by representational algorithms excludes any form of plurality or space for interpretation. Human algorithmic activity seems to be reduced to calculations guided by prescribed formulas or calculation paths, but it may be possible to count some rigorous instructions for traditional or religious rites among algorithms. In most cases, a shared lifeworld and similar experiences captured in language enable people to communicate in more ambiguous ways, making room for individual capacities, choices, equipment, contexts and/or preferences without losing the ability to give sufficiently definite instructions.[10] When it comes to connectionist algorithms, it is even less clear what they are. Although it is clear that they are complex structures

[9] Moschovakis also calls algorithms "purposeful interpretations" (Moschovakis, 1998, p. 75) of specific computing equations.

[10] Though, the ambiguity of human language is much more effective than definiteness because it takes shortcuts by human abilities of contextualisation and understanding that make it possible to connect ambiguous expressions to concrete situations.

that deal with a lot of data to calculate probabilities, and their technical medium is known, the process after the input and maybe the first layer is opaque – "not much is known about what happens in later layers" (Alpaydin, 2016, p. 100). In fact, the process is discussed under the term of interpretability (see 3.2), which already indicates the responsibility of judgement to determine what it actually is or means.

The definiteness of calculation of the classical algorithm or the reliance of connectionist algorithms on large data sources can be camouflaging features that make their processes appear objective. But algorithms depend on definite inputs. For this reason, machine-implemented algorithms in particular are in need of a highly formalised elaboration of their input and process. Although it may appear that the algorithmic calculation implemented in a machine is completing a task or solving a problem, it is performing calculation with operationalized, quantified entities: "The problem solved by the Turing machine is not 'our' problem (print a paper, booking a hotel, read the newspaper, etc.) but a series of mathematical functions" (Possati, 2020, p. 18). Of course, the implementation of algorithms in our lifeworld depends on the mediation of the in- and outputs so that they can be properly used for external purposes (Gill, 2017, p. 318). This means that they are highly dependent on judgement, both in their design and in their use. So my second thesis is:

Algorithms, as purpose-embedded entities, emerge from clusters of judgements.

In order to translate a non-formal problem into a formal input, it is necessary to turn ambiguous contexts into unambiguous, definite values. This requires judgement, firstly because "facts always require interpretation" (Malpas, 2020, p. 1074) and every simple step of programming is in need of "countless decisions and trade-offs" (Elish & Boyd, 2018, p. 69). To take the example of a recipe for the classical algorithm again, a typical instruction would be to 'add a dash of salt', which is quite ambiguous: "A 'dash' is defined to be 'less than 1/8 teaspoon,' and salt is perhaps well enough defined; but where should the salt be added? – on top? on the side?" (Knuth, 1997, p. 6) For humans, measuring a dash of salt for a given amount of cake dough is a learning process that is facilitated by interpersonal guidance and/or the resonance of the result – the cake. It is never necessary to know the exact amount of salt that has been added. But if you were to design a fully automated cake baking machine that had to add the right amount of all the ingredients by dropping them from containers above the mixing bowl, the representational algorithmic instruction would have to be guided by an input that gave an exact guideline for how much salt to add. This could be done by setting a certain time for the salt container to be opened, or by scaling the amount of salt added in the mixing bowl. In any case, the programmer had to decide on a certain measured amount of salt, a certain time to add it, a certain spot of adding, and so on. While the pure algorithm is unrelated to ingredients, dough, cake or even the concept of baking, the programmer relates the algorithmic code, the operating instruments and the ingredients, dough, and cake to each other by judging their workflow in order to achieve the desired result. These judgements are incorporated into the algorithmic set-up as individual definite decisions.

The different approach of ANNs, however, goes beyond this definite decisions that are preserved in the course of representational algorithms deducting from them. Still, even unsupervised ANNs depend on operationalised inputs, which again depend on the judgements of programmers and the selection or availability of data. Possati has already interpreted the process of programming with reference to Kant's concept of imagination, which is closely linked to judgement in Kant's theory: "Therefore, the programmer has to interpret the problem and create a new representation of this problem (the program) that can mediate between the problem itself and the Turing machine. This is an act of imagination" (Possati, 2020, p. 20). It is crucial that the programming of an algorithm or at least the operationalisation of input data is based on a concrete *interpretation* of a problem to be solved or a task to be coped with, and that this interpretation itself is based on and initiates judgements. These judgements (and, of course, the prejudices they contain) are incorporated into the algorithm and its calculations, even though ANNs deal with "hidden unit values [that] are not 0 or 1 but continuous", which "allows a finer and graded representation of similar inputs" (Alpaydin, 2016, p. 104). The operationalisation of input data, which can be understood as an equivalent of defining concepts, gives an axiomatic base to processing algorithms and cannot be denied or reflected. To treat judgements as axioms, as a continuous basis of (rule-following or probabilistic) calculation, inscribes them into the result of the process. In other words, these judgements become the unreflected conditionality for the validity of the result. Hence, my third thesis is:

The outputs of algorithms can only be used in a prejudiced way.

This proposition aims at the de-facto results of symbolistic and connectionist algorithms.[11] Their re-integration in purposeful contexts of the human lifeworld is of course dependent on human judgement or prejudice due to the nature of algorithms (see first thesis).[12] But even though the use may be reflected, the result forces a prejudiced use in the sense that the algorithmic process is formed by its axiomatic basis and intertwined with it in a non-resonant way. The process may deconstruct the input data as in ANNs, but it can do so only under the conditions of the input data and the implemented idea of order. Big Data and AI are "*socio*-technical concepts", meaning that "the logics, techniques, and uses of these technologies can never be separated from their specific social perceptions and contexts of development and use" (Elish & Boyd, 2018, p. 58; cf. ibid., pp. 71f.). Even in unsupervised machine learning, there is a demand of hyperparameters, e.g. for the way in which input data is fed into feedback loops (cf. Watson, 2023). The programmer's assumptions as

[11] I simplify the steps of the use. Of course, the use of algorithms that work with operationalised inputs requires a back-translation of the output, which can also be done by automated algorithms that duplicate the phenomena described.

[12] As Bechtel put it, the activities of neural networks "begin and end with symbolic representations constructed by humans, and hence any intentionality they seem to exhibit in response to their symbolic representations is all mediated by humans for whom these symbolic representations do have content" (Bechtel, 1996, p. 274).

well as the decisions are made according to purposes inscribed into the structure of algorithms and therefore in their results. In order to get a useful result, "to be meaningful to us" even unsupervised clustering is in need of human "setting number of topics, cleaning data in a particular way with an a priori understanding of 'meaningful' clusters and interpreting clusters with parent classes manually" (Bechmann & Bowker, 2019, p. 6f.). Chatbots and other language processing tools depend on models of statistical patterns of lingual symbols, on presumed trustworthiness of sources and websites etc. The process is not supposed to and cannot question its basis, though this is the key feature of judgement, reflecting according to alternatives provided by a human lifeworld. Judgement deals with the ambiguity of language and re-evaluates concepts in a given situation, whereas the algorithmic output sticks to the conceptual framework given in the input data. The consciousness of the conditionality of algorithmic outputs cannot change the prejudiced use, because of the axiomatic basis and specific data sets that are meant to provide unambiguous "ground truths" (Jaton, 2017) about human beings and their lifeworld. Nonetheless, one can reflect on the use of algorithmic outputs and their effects themselves. The third thesis is by no means meant in a strictly pejorative way, according to the approach of the necessity of prejudice assumed in this paper. But an appropriate use of prejudiced structures has to fulfil certain requirements to keep all stakeholders of that use in a sovereign position, and the most important is the transparency of the axiomatic basis and the data sources.

3 Future(s) of Judgement and Calculation

The examination of the concept of algorithm has shown that it cannot be understood without being embedded in a purposeful context. This context depends on the ambiguity of language and traditional frameworks of interpretation. Thus, my thesis on the relationship between judgement and algorithmic calculation is that algorithms do not have inherent purposes, but they are *used* for purposes in the course of judgement and/or prejudice. This means that they cannot in any way *replace* human judgement, not even in a problematic way, but their proliferation in human life *changes* the use of judgement and prejudice.

3.1 On the Temporality of Judgement and Calculation

I can only touch briefly on the question of temporality here, but the cornerstones have already been mentioned. Gadamer and Arendt are both trained by Heidegger and are therefore well aware that acts of understanding are constituted by a temporal dimension. Heidegger claims the temporal rootedness of understanding.[13]

[13] Besides the fact that Gadamer's idea of prejudice and judgement relies on Heidegger's conceptualisation of understanding, there is another reason of compatibility. As mentioned in 2.3, for Heidegger understanding has the structure of "*something as something*" (Heidegger, 2010, p. 144). The concept of judgement presented here can be subsumed under this notion, because it is broadly speaking the understanding of something (general) as something (particular) or reversed.

For him, "Dasein always has understood itself and will understand itself in terms of possibilities", and he grasps this quality of understanding as its "project character", which does not mean that it is concerned with concrete possibilities but that "in projecting, project throws possibilities before itself as possibility, and as such lets it *be*", "understanding is the mode of being of Dasein in which it *is* its possibilities as possibilities" (Heidegger, 2010, p. 141). To be able to relate to concrete possibilities, human understanding must inherently be dimensioned in the form of possibilities, and thereby to an openness inherent in itself and the world. Human beings only understand themselves and the world in terms of possibilities, and, moreover, in understanding they become their possibilities. Human understanding, though, always happens against the background of a non-given whole (cf. ibid., p. 146) that is intrinsically temporal and always incorporates the openness of the future. A being that is capable of understanding is projecting itself, it is essentially a "*being toward possibilities*"; only such a being can also develop its understanding in form of "*interpretation*" (ibid., p. 144); and only such a being can have meaning, defined as "*the upon which of the project in terms of which something becomes intelligible as something*" (ibid., p. 147). It is a special quality of human understanding that the virtual whole it presumes and projects gives the possibility to allocate places for something by interpretation and thereby provide it with meaning in the course of life as the being towards possibilities.

Gadamer then takes up these concepts and concentrates on historicity and the fact that the human lifeworld, based on the Heideggerian idea of Dasein and its understanding, cannot not be reduced to mere facts. The "*life-world*" as "the antithesis of all objectivism" means "the whole in which we live as historical creatures" and it implies "infiniteness of the past, and above all the openness of the historical future" (Gadamer, 2004, p. 239). He further clarifies that "we are always situated within traditions" (ibid., p. 283), that we always "understand ourselves in a self-evident way in the family, society, and state in which we live" (ibid., p. 278). Judgement is, according to Gadamer, always dependent on historical prejudices, because the act refers to meanings and forms of reasoning that are provided by tradition. Gadamer claims that reasons are in fact only possible because of "what we call tradition: the ground of their validity" (ibid., p. 282). Further, he sees understanding as "a process of transmission in which past and present are constantly mediated" (ibid., p. 291), and this way of being historical "*means that knowledge of oneself can never be complete*" (ibid., p. 301). This indicates a specific relationship of the present to the past:

> In fact the horizon of the present is continually in the process of being formed because we are continually having to test all our prejudices. An important part of this testing occurs in encountering the past and in understanding the tradition from which we come. Hence the horizon of the present cannot be formed without the past. (Ibid., 305)

What Gadamer describes here is the reformation and reevaluation of the past in different contexts of the present, and while tradition is considered as a non-given whole, the present is constituted by a specific situation and a specific act of understanding. In other words: We deal with judgement as a mediator between something general and something particular, and this mediator always depends on the past as a

provider of validity, of a virtual whole, as an interpretive framework. Judgement can only rely on meaning because it is situated in a historical lifeworld and refers to tradition that shapes, but does not limit the possible ways of relating a generality with a particularity. For my provisional concept of judgement, I agree with Gadamer's idea of a referential rootedness of judgement in a past tradition that cannot be given as an explicit whole because it belongs to an overarching lifeworld.

Judgement is dependent on a past tradition, but it is not at its mercy. As I mentioned in 2.3, there are ways of questioning and changing these grounds of validity. Moreover, as Heidegger's concept of the project of understanding already implies, there is an inherent orientation of judgement towards the openness of the future. And while judgement may be based on a purposefulness shaped by tradition, it is not directed towards the past. This is particularly important regarding Arendt's concept of prejudice. Prejudice is dangerous because "it is always anchored in the past" (Arendt, 2005, 101), which also means that the traditional interpretive framework is bound to a different lifeworld, to different conditions of application. Judgement, then, can serve as empowerment against "the pseudo-divinity named History of the modern age, without denying history's importance but denying its right to being the ultimate judge" (Arendt, 1978, p. 216). This act of judgement is, of course, not itself ultimate, it is an act that is bound to its present. Arendt's idea of judging is to deal with particularities without any given generalities in the realm of human affairs, with the need to project into the future. Nonetheless, there is no way to calculate it properly because "the world is daily renewed through birth and is constantly dragged into what is unpredictably new by the spontaneity of each new arrival" (Arendt, 2005, p. 127). Although tyranny tries to force its people into conformity and predictability, total control is not possible in the face of the uniqueness of each individual, who is capable of starting chains of causality anew, which Arendt describes with the concept of natality (cf. Arendt, 1998, p. 9). This human capacity forces judgement to include the unpredictability, albeit the openness, of the future. For Arendt, the only way to hedge against this common uncertainty lies in "the faculty to make and keep promises", which only makes community possible as a temporal bond, because "binding oneself through promises, serves to set up in the ocean of uncertainty, which the future is by definition, islands of security without which not even continuity, let alone durability of any kind, would be possible in the relationships between men" (ibid., p. 237). The possibility of politics lies in this ability to promise, in the ability to decide and adhere to common norms in unforeseeable circumstances (cf. Arendt, 1961a, p. 164; Arendt 1972b, pp. 92f.).

To sum up these insights into my integrative account: Judgement is a purposeful act and, as such, an act that has an intrinsic temporal directedness. Indeed, the temporality of judgement is not simply constituted by the fact that it is purposeful and therefore aims at achieving a certain purpose in the future. Meaningful judgement is inherently interwoven with a past tradition and an uncertain future. Further, it is directed towards abstract purposes that are not achieved once, but are orientation values that require continuous interpretation and implementation. To ascribe to such a purpose, whether it be justice, prosperity or the good life in general, is to ascribe to a tradition of understanding such a purpose, to an interpretative framework that is grounded in history but never given in its explicit totality. This vacancy, the

unavailability of a whole even if it is addressed, is at the core of the human faculty of judgement. The incompleteness provides flexibility for the actual application in concrete situations, leaves room for plural discourse and gives plasticity for change and new historical adaptations over time. Judgement is tied to a past tradition and, as a present act, projects into the future.[14] The non-given whole as the background of judgement cannot be given because of the openness of time. Judgement transfers abstract meanings through time by means of concretisation, and it can only do this in ambiguous language – the temporal character of judgement implies that judgements are never definite, but themselves become ambiguous subjects of judgement in the future.

Both classical and connectionist algorithms, I argue, are not temporal in the same way, their process is fundamentally different from human understanding and therefore cannot replace it in any way. I will mark the crucial difference in three steps that build onto each other: The difference in processing language, the difference in processing reference, and finally the difference in processing temporality.

The difference in processing language is a result of operationalisation. Ambiguous human language is able to adapt in the process of its application, to generalise or concretise, to change or reaffirm its traditional meaning and use due to the conditions it is confronted with. Consequently, human judgement does not simply draw subsuming connections between generalities and particularities, but inevitably deals with its own framework. However, operationalisation requires a static framework; this is obviously true for classical, representational algorithms with predetermined functional paths, but also for unsupervised ANNs that adhere to the operationalised framework of their input data. Quantified representations have a definiteness that is necessary for calculative or probabilistic processes,[15] but at the same time they do not allow these quantified conceptual representations to be reflected in the course of that process. The algorithmic process necessarily assumes that the input data is both accurate and sufficient for the required output; judgement challenges its basis by exposing it to unpredictable applications, and therefore requires ambiguous language.

The difference in processing language correlates with the different ways in which the processes relate to their subject. Judgement is an act that takes place against the background of a non-given whole, whereas the output of algorithms is always based on a given, definite amount of data and a given framework of processing. Algorithms provide a distinct output at a distinct point in time, providing completeness of data and processing steps, which are in principle explicit. On the contrary, it has been shown that judgement has an intrinsic incompleteness, an incompleteness that is constituted by a vacancy of time. The temporality of judgement makes meaning

[14] This is true even for the most ambitious judgements claiming universality. They are projecting their continuous validity in all possible future scenarios, even though they cannot designate them.

[15] In what follows, I limit my remarks to the temporal aspect and cannot pursue what expressions of quantity actually are and how they root in human experience; Krämer (2014) suggests that the symbolism of computation, however abstract, is a derivative of human spatial sensuality.

and purposefulness possible, because both rely on an ever-changing lifeworld that cannot be made explicit in its entirety, but can be handled with ambiguous language.

Completeness and incompleteness, mediated through different languages, imply different forms of temporality, a different relation to time and a different connectedness in time. The rootedness of judgement in the past and its directedness towards the future contrasts with algorithmic processes that only work in iteration and/or iteration loops. Because this process of symbolic or probabilistic calculation is holistic in the literal sense, it is not linked to previous iterations as alterities, but only as different starting points for processing. Machine learning is a refinement that is achieved through large data sets and multiple iterations, leading to different starting conditions in the next iteration.[16] Algorithms cannot work with openness and therefore not with any concept of future but only with iteration under different conditions and available data, hence, not a changing world, but a different data set, not a different time, just another present calculation. In many ways, neural networks don't have a future, but only a reiterating present. Social media algorithms accumulate more and more of the same history and present, projecting their own process of iteration on human users, which equals the assumption that users continue to want to see more of the same (cf. Bechmann & Bowker, 2019, p. 4). In short, AI is supposed to learn "ground truths" about human beings (cf. Jaton 2017) by calculating with the assumption that humans are not learning at all. In internal temporal regard, every iteration is complete; whatever is ascribed to a machine as a learning success or a refinement, is, in fact, a human ascription based on an external purposiveness.

Algorithmic calculation is deprived of purposefulness and thus of any kind of temporal directedness. Even if algorithms are indeed *used* in time and their calculations *take* time, algorithms do not *have* time in the sense of a temporal directedness that separates past, present and future and somehow relates perception and activity to them. Operationalisation of data leads to a process that is not concerned with a lifeworld, but with isolated representations that to not have a meaning within that process. This reduction to timeless quantities may, however, give the output an appearance of timeliness in the process of retranslation into ambiguous language.[17] But it is human judgement or prejudice that applies the calculation to a temporal experience of reality, semanticising the output with its own temporal directedness.[18] Pointing at the limits of algorithmic expression is not a depreciation, but

[16] Even the internal memories of ANNs only work as present factors: "If we define the state of a network as the collection of the values of all the neurons at a certain time, recurrent connections allow the current state to depend not only on the current input but also on the network state in the previous time steps calculated from the previous inputs" (Alpaydin, 2016, p. 93).

[17] This is part of what makes algorithms suitable for machinewashing, cf. Seele & Schultz, 2022.

[18] This is not unique to algorithms, but applies to all kinds of formulae, even the laws of natural science. The difference is that the universality and continuity of formulae applied to nature are descriptive models that are constantly exposed to the possibility of being falsified. And the identification of real-life-objects with quantities in formulae is a matter of judgement and therefore a temporalisation, the attribution of certain temporal characteristics that transfer the formal relations into the temporal world: The timeless figures are transferred to what is judged a *continuity* of natural law, i.e. the continuity of the causes and effects in nature. There is also an approach to link the specific human temporality according to Gadamer's hermeneutics with computer-brain-interfaces, cf. Lindia, 2022.

an advocacy for the use of algorithms as instruments of aware preliminary human judgement. The danger of the unreflective use of algorithms lies primarily in the support of prejudices and the growing loss of bonds to experience and grounding in the actual world.

3.2 The Prejudice(s) of Judgement and Calculation

The investigation has repeatedly referred to the constitutive role of prejudice. Hence, prejudices are not to be understood in a pejorative sense, but as unavoidable conditions of the conduct of life. For Arendt, a human prejudice is a former judgement "which originally had its own appropriate and legitimate experiential basis, and which evolved into a prejudice only because it was dragged through time without its ever being reexamined or revised" (Arendt, 2005, p. 101). As mentioned above, I see prejudice as an interpretive framework provided by tradition, and thus have a broader concept than Arendt, who is concerned with specific ways of relating generalities to particularities. In what follows, I focus on her concept of prejudice, not to rethink my hermeneutic approach, but to draw attention to a kind of human relationship to this interpretive framework, i.e. prejudiced behaviour. This is an appropriate interpretation because Arendt, too, is primarily concerned with how people *use* prejudice, not with what it means as an epistemic entity.

Arendt ascribes two important functions to prejudice: First, it enables life in a highly complex environment, because no one is able "to form an original judgment about everything on which he is asked to pass judgment in the course of his life" (ibid., p. 99); second, it is a means of communitisation. Both of these functions are provided by the central characteristic of prejudice, the fact that it is "not tied to personal experience" (ibid., p. 100; cf. Arendt, 1971, p. 418). This allows people to act intuitively on prejudicial patterns rather than having to judge every concrete situation or case in everyday life, and also to connect with each other through prejudice because it is simply not tied to individual experience. Commonplaces and generalisations are typical examples of the socialising potential of experienceless prejudices.[19] The condition for proper prejudices is that they actually work without too much friction, that they simplify everyday life and/or are approved by a community. Times of crisis are the parting of the ways for prejudices. Either, the old prejudices are replaced by actual judgements that have a direct link to the experience of reality:

> The disappearance of prejudices simply means that we have lost the answers on which we ordinarily rely without even realizing that they were originally answers to questions. A crisis forces us back to the questions themselves and requires from us either new or old answers, but in any case direct judgements. (Arendt, 1961b, p. 174)

Or, the prejudices turn

[19] They can be identified with abstract concepts forming Gadamers interpretative frameworks that communities share in their common sense – if those concepts were forced to explicitness and claimed general truths.

> into something that by nature they most definitely are not – that is, into pseudotheories, which, as closed worldviews or ideologies with an explanation for everything, pretend to understand all historical and political reality. If it is the function of prejudice to spare the judging individual from having to open himself to, and thoughtfully confront, every facet of reality he encounters, then worldviews and ideologies are so good that they somehow shield us from all experience by making ostensible provision for all reality. (Arendt, 2005, p. 103)

Arendt's criterion for the legitimate use of prejudices is clearly that they prove themselves in practice. There is no need to reestablish links with experience where certain procedures or problems can be successfully handled, where certain purposes can be achieved, where certain things are understood without having to evaluate the approach. The failure of prejudice ends either in reestablishing a link to experience, or in sealing off from experience and raising prejudice from a condition of purpose to a purpose in itself (and thus extending it to a 'closed worldview'). Theories, even scientific theories, are often used and referred to as prejudices, and it is important to mark them as grounded, but always conditional models of reality.

So, what is the prejudiced use of algorithms? Two main characteristics of algorithms have been determined as the operationalised basis and the data-driven completeness. The algorithmic output, then, is inseparably interwoven with these characteristics, it is bound to the operationalised input and the axiom of completeness. The use of such an output,[20] however, requires its reintegration into the meaningful and temporal context of the human lifeworld, which lacks the unambiguity of data as well as a completed whole of reference. What the output actually means for human beings, is subject of human interpretation, and this interpretation therefore is to ascribe a worldly relationship to the algorithmic output it actually does not have. As the algorithmic process is completed under its own conditions, it is also not tied to personal experience, it springs from operationalisation that can be considered ossified judgement, it can, following Arendt's conception, only be used in a prejudiced way. This is not to say that the use of algorithms cannot be reasonable or justified, even an enriching technology for the organisation of social life. However, the appropriate use of algorithms – i.e. the reflected prejudiced use – depends on specific conditions. In the following, I will concentrate on the demand for transparency.

A common criticism of the opacity of algorithms concerns their black-box nature (Rai, 2020; von Eschenbach, 2021; Zednik, 2021), because "not much is known about what happens in later layers" (Alpaydin, 2016, p. 100) of ANNs. This problem is related to the question of interpretability, which has become urgent with the rise of machine learning and ANNs that collect large amounts of data autonomously. The question of interpretability – how to understand the epistemic operation of AI, what it means, what it does, or why it produces a certain output – has thus provoked many different answers, some of which undermine their own intentions. This is especially true of attempts that simply "define interpretation as a process taking

[20] I reduce my account here to the end user and leave out the programmer.

one explanation to another, more understandable explanation" (Erasmus, Brunet & Fisher, 2021, p. 858; cf. quite similarly Pirozelli, 2022). Such accounts imply a definiteness of understanding that allows explicit understanding by reduction and deduction, and thus equate human understanding with explicit explanation, assuming that it relies on available completeness. At least according to my methodological paradigm of hermeneutics, there is an irreconcilable difference between understanding and explanation, as has been emphasised by Wilhelm Dilthey, the founder of modern hermeneutics (cf. Dilthey, 2010, p. 147; Harrington, 2000). The explicitness of explanations is reserved to the mere relations of cause and effect applied in the natural sciences, while the sphere of human affairs requires other categories which cannot be made explicit in their entirety and which are not simply the subject of description but aim at orientation, provide meaning. Human judgement, when directed to abstract purposes, is at some point subject to choices that may be well founded but are never fully transparent and explicable. This is why there have been legitimate interventions against the double standard, because AI is demanded to be fully transparent, which human judgement is incapable of (cf. Zerilli et al., 2019). Furthermore, Krishnan (cf. Krishnan, 2020) calls for an evaluation whether the mere algorithmic results serve the purpose of the algorithm, rather than demanding transparency of its internal processes.

My claim is to demand full transparency not of the technicalities, but of the human judgements that form the axiomatic basis of an algorithm, especially in the context of machine learning. Only by reflecting the output of an algorithm on the conditionality of the judgements and the data on which it is based can I autonomously assess its value for my purposes. However, the use of algorithms is prejudiced in any case and must be reflected upon in a certain way. Reflection on conventional prejudices works by intervening in their function, in their formation, their concepts, in the connection with their basic experience; this is not possible with algorithms, because the algorithmic formation is not a process that originates from the conduct of human life. There are many different opacities of algorithms that are considered problematic (cf. Burrell, 2016), but I argue in particular for the transparency of the axiomatic basis that underlies the process and/or the operationalisation of the input data as well as the selection of data that is fed into the algorithm. Making this form of validity of the output available and visible helps to raise the awareness of its conditionality and its mere instrumental improvement in a lifeworld that has to provide its own meaning and purpose. Without this transparency, algorithmic outputs appear as unconditional parts of our lifeworld. Considering the assumed amount of data, they seem to be superior to our limited realm of experience, and with their mere eligibility for prejudiced use they hide the purposes (e.g. economic or power interests) and the assumptions (e.g. about the character of particular ethnicities, residential areas) that guided the construction of their architecture and the selection of their input data. Without reflection, the prejudiced use of algorithms is unintentionally dependent on these purposes and assumptions.

What would such transparency look like? I demonstrate this by means of ethical AI tool Delphi. Delphi is a trained ANN that purports to model "people's moral judgments on a variety of everyday situations", with the aim of helping "AI systems be more ethically-informed and equity-aware" (https://delphi.allenai.org/). The

website can only be accessed after confirming that you understand the limitations of such a model, and the input screen constantly reminds you that the results should not be used as actual advice for humans. You can then enter a situation or a question and Delphi will answer whether an action is right or wrong, normal or understandable etc. Compared to the most common applications of AI, Delphi has the advantage of being a research project that is interested in publicly discussing its architecture and limitations. But no matter what is explained in an external paper, the tool appears as an interlocutor evaluating my moral questions, giving normative answers. If I ask Delphi, "Can I kill a tyrant?", it tells me, "It's wrong", whereas it answers the question "Can I kill a Tyrant?" (with a capital T) "It's okay". Leaving these technicalities aside, I focus on the appropriate use of such tools, and what they mean for us.

In their paper on the development of Delphi, its creators reveal its axiomatic foundations as well as its data set. The data set consists of 1.7 million moral judgements provided by crowdworkers (cf. Jiang et al., 2022, p. 5; Metz, 2021). The assumption of the deep learning process is that "we present a large amount of examples to the learning algorithm and let it learn the implicit rules from those examples in a bottom-up manner, rather than humans prescribing rules in a top-down fashion for machine" (Jiang et al., 2022, p. 7). The idea of "implicit rules" as a normative basis contained in the examples is based on the assumption that there *are* common and underlying rules of human moral judgement that can be made operable (and thus in principle explicit) for a machine by training it with large amounts of data. At the same time, however, the creators reject relying on any normative claims, stating that the entire approach of Delphi is merely descriptive (cf. ibid., pp. 26f.). However, the database does serve as a normative basis for the evaluation of the tool: "When tested with unseen examples from COMMONSENSE NORM BANK, Delphi predicts the correct judgment 92,8% of the time, performing much better than state-of-the-art language models such as GPT-3, which only makes correct predictions 60,2% of the time" (ibid., p. 5). Furthermore, there is no information about the crowdworkers and whether they, or the situations they were judging, were selected in any representative way, or whether the people were just hired at random. The sheer volume of data seems to be expected to compensate for the method. The collection of such judgements does not represent people's actual moral actions, but only their judgement in hypothetical cases, and their judgement in hypothetical cases not as an act of inner self-assurance, but under the social pressure of surveillance. The creators expect the database to represent common sense "without specifically endorsing the correctness or appropriates of particular judgments" (ibid., p. 9, fn. 2).

This shows that the architecture of the tool is based on highly problematic, if not contradictory assumptions and that the data set is opaque in its actual composition and significance.[21] Despite these problems, the algorithmic process itself does not judge moral issues in any way, because it simply relates its data in a probabilistic way, without finding an 'implicit rule' the programmers are not aware of. Reflecting

[21] If I problematised the mere result, as Krishnan (cf. Krishnan, 2020) suggests, I could not raise these specific but general concerns about such an algorithm and link its basic assumptions to the orientation of my own judgements.

judgement, for example in the context of ethical theories, must take into account the ambiguities of the theory, its conditional nature, and the purpose of judgement itself. This awareness of its limitations, tentativeness and incompleteness is the strength of judgement, because it leaves room for the resonance of reality and experience through time. In a critique of Delphi, Seele defines the characteristics of ethical judgement:

> An ethical judgement is the reflection of a moral situation or dilemma through the lens of an ethical theory. It is the theory that makes ethics a concept driven by neutrality and distance from personal moral values and beliefs (that everyone has, including ethicists). Ethicist[s] – unlike preachers – deliver a reason-guided argument based on one of the existing theories, such as virtue ethics, deontology, consequentialism, contractualism, utilitarianism or discourse ethics. (Seele, 2022)

The structure of judgement, linking something general to something particular, allows it to have reasons, to justify itself. Delphi cannot give reasons, because it does not have reasons in the way a judgement has, it has data and statistical calculation. It does not refer to a constantly changing lifeworld, but to a present data set, which it calculates iteratively. Considering that the concept of ethics means the reasonable reflection of principles and theories, and the discipline of doing so, Delphi is not an ethical tool, but only a tool whose data deals with moral topics.

To use the tool in a reflected prejudiced way means to be aware of these conditions, limitations and problems, and seems to leave no other conclusions than that its answers have nothing to do with our meaningful lifeworld. Nevertheless, Delphi is quite useful for the understanding of AI tools, precisely because it is a research project that aims to debate itself and therefore provides transparency, by opening itself up to public discourse. This transparency is a prerequisite for proper criticism and improvement, whereas the opacity of more common AI tools prevents such a reflected prejudiced use. Given the role of common understanding and purposefulness, the problem with the opacity of algorithmic axioms (leading to mere unreflective prejudiced use) is that their way of relating to purposes and their constitutive judgements are withdrawn from common discourse and individual reflection. The result-oriented account of mere utility runs the risk of impoverishing the horizon of human judgement and gives way to instrumental reason, which simulates the objectivity and definiteness of purposes and their attainment.

4 Conclusion

Capturing Arendts thoughts on the failure of American policy in the Vietnam War, it becomes clear that the mistake was not (as she said) simply to substitute judgement for calculation. Rather, the 'experts' misjudged by judging that their calculations were adequate and sufficient representations to capture reality. Since calculation has a different epistemic object, judgement and calculation are not just two different ways of dealing with the same problem, one of which is better than the other. Structures of calculations, such as algorithms, emerge from a sum of judgements that are taken as axioms and thus become static prejudgements through operationalisation. They are also based on a finite amount of data, which is treated as a definite completeness. This process in itself

is not problematic, as prejudices are necessary in many aspects of complex human life. It is important, however, to treat necessary prejudices as preliminary judgements, i.e. to monitor their impact on reality, one's own experience and discourse, with a constant openness to their resonance. In order to use algorithms in this way, it is necessary to have a transparent insight into the judgements that were made during their programming and the selection of data, so that these conditionalities can be reflected upon during the adaptation of the output for one's own purposes. The current increase in the use of algorithms can be seen as problematic in that opaque algorithms are – and can only be – used as mere unreflected prejudices (or they have to be completely refused). This development threatens to impoverish the human capacity for reflecting judgement and to give oneself orientation, for which algorithms can never be more than mere tools. Community members who rely on the illusion that AI will replace their judgements with better conclusions or decisions will not improve their lives, but deprive themselves of the ability to connect abstract concepts with concrete experiences of themselves and others. It is therefore social poverty and the atrophy of meaningful lifeworlds, not technological superiority, that creates the dangers and power of algorithms.

The inescapability of judgement is more than a mere assertion of a normative humanist approach. On the contrary, it goes to the heart of a world-shaping species that cannot escape its own ambiguities. Ironically, the fact that there is no generally accepted definition of what an algorithm is already shows that algorithms only make sense in the context of a meaningful life.

Acknowledgments Parts of the article were developed during an open fellowship at the Weizenbaum Institute for the Networked Society in Berlin and made possible by the immensely helpful comments of the multidisciplinary scholars I met at the workshops there. Special thanks go to Christian Strippel, who provided me with numerous remarks on a draft of this paper. I would also like to thank the two anonymous reviewers for their extremely helpful comments, questions and encouragement to expand and deepen my topic.

Author Contributions The author (DS) assumed sole responsibility for all aspects of the work presented herein.

Funding Open Access funding enabled and organized by Projekt DEAL. The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability Not applicable.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication The author (DS) consents to the publication of this article.

Competing interests The author has no relevant financial or non-financial interests to disclose.

References

Aggarwal, C. C. (2018). *Neural Networks and Deep Learning: A Textbook*. Springer. https://doi.org/10.1007/978-3-319-94463-0

Alpaydin, E. (2016). *Machine Learning: The New AI*. The MIT Press.

Anderson, J. A., & Rosenfeld, E. (Eds.). (1988). *Neurocomputing: Foundations of Research*. The MIT Press. https://doi.org/10.7551/mitpress/4943.001.0001

Arendt, H. (1961a). What is Freedom? In H. Arendt, *Between Past and Future: Six Exercises in Political Thought* (pp. 143–171). The Viking Press.

Arendt, H. (1961b). The Crisis in Education. In H. Arendt, *Between Past and Future: Six Exercises in Political Thought* (pp. 173–196). The Viking Press.

Arendt, H. (1961c). The Crisis in Culture: Its Social and its Political Significance. In H. Arendt, *Between Past and Future: Six Exercises in Political Thought* (pp. 197–226). The Viking Press.

Arendt, H. (1971). Thinking and Moral Considerations: A Lecture. *Social Research, 38*(3), 417–446.

Arendt, H. (1972a). Lying in Politics: Reflections on the Pentagon Papers. In H. Arendt, *Crises of the Republic: Lying in Politics. Civil Disobedience. On Violence. Thoughts on Politics and Revolution* (pp. 1–47). Harcourt Brace & Company.

Arendt, H. (1972b). Civil Disobedience. In H. Arendt, *Crises of the Republic: Lying in Politics. Civil Disobedience. On Violence. Thoughts on Politics and Revolution* (pp. 49–102). Harcourt Brace & Company.

Arendt, H. 1978. *The Life of the Mind: One / Thinking – Two / Willing: One-volume Edition*. Hartcout, Inc.

Arendt, H. (1989). *Lectures on Kant's Political Philosophy*. R. Beiner (Ed.). Chicago University Press.

Arendt, H. (1994). Understanding and Politics. In H. Arendt, *Essays in Understanding. 1930–1954: Formation, Exile, and Totalitarianism* (pp. 307–327). J. Kohn (Ed.). Schocken Books.

Arendt, H. (1998). *The Human Condition* (2nd ed.). The University of Chicago Press.

Arendt, H. (2005). Introduction *into* Politics. In H. Arendt, *The Promise of Politics* (pp. 93–200). Ed. and intr. J. Kohn. Schocken Books.

Arendt, H. (2022). *Denktagebuch: 1950–1973* (2nd ed.). Ed. U. Ludz & I. Nordmann. Piper.

Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society, 6*(1). https://doi.org/10.1177/2053951718819569

Bechtel, W. (1996). Responsibility and decision making in the era of neural networks. *Social Philosophy & Policy, 13*(2), 267–295. https://doi.org/10.1017/S026505250000354X

Binns, R. (2022). Human judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance, 16*, 197–211. https://doi.org/10.1111/rego.12358

Blass, A., & Gurevich, Y. (2003). Algorithms: A quest for absolute definitions. *Bulletin of European Association for. Theoretical Computer Science, 81*.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12. https://doi.org/10.1177/2053951715622512

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction into algorithms* (3rd ed.). The MIT Press.

Cooley, M. (2007). From judgment to calculation. *AI & Society, 21*, 395–409. https://doi.org/10.1007/s00146-007-0106-5

d'Avila Garcez, A., & Lamb, L. C. (2023). Neurosymbolic AI: the 3rd wave. *Artificial Intelligence Review, 56*, 12387–12406. https://doi.org/10.1007/s10462-023-10448-w

Dilthey, W. (2010). Ideas for a descriptive and analytic psychology. In W. Dilthey, *Selected Works. Volume II: Understanding the Human World* (pp. 115–210). R. A. Makkreel & F. Rodi (Eds.). Princeton University Press.

Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs, 85*(1), 57–80. https://doi.org/10.1080/03637751.2017.1375130

Erasmus, A., Brunet, T. D. P., & Fisher, E. (2021). What is Interpretability? *Philosophy & Technology, 34*, 833–862. https://doi.org/10.1007/s13347-020-00435-2

Gadamer, H.-G. (2004). *Truth and Method*. Transl. J. Weihnsheimer & D. G. Marshall. Continuum.

Garcia Rosa, J.L. (2019). Biologically Plausible Artificial Neural Networks. In: Suzuki, K. (Ed.), *Artificial Neural Networks – Architectures and Applications* (pp. 25–52). Intech d.o.o. https://doi.org/10.5772/54177

Ghahramani, Z. (2004). Unsupervised Learning. In: O. Bousquet, U.v. Luxburg & D. Rätsch (Eds.), *Advanced Lectures on Machine Learning* (pp. 72–112). Springer.

Gill, K. S. (2017). Hermeneutic of performing data. *AI & Society, 32*, 309–320. https://doi.org/10.1007/s00146-017-0727-2

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press.

Harari, Y. N. (2016). *Homo Deus: A brief history of tomorrow*. Harvill Secker.

Harrington, A. (2000). In Defence of *Verstehen* and *Erklären*: Wilhelm Dilthey's *Ideas Concerning a Descriptive and Analytical Psychology*. *Theory & Psychology, 10*(4), 435–451.

Haykin, S. (2009). *Neural Networks and Learning Machines* (3rd ed.). Pearson Prentice Hall.

Heidegger, M. (2010). *Being and Time. A Translation of* Sein und Zeit. Transl. J. Stambaugh. Revised and with a Foreword by D.J. Schmidt. State University of New York Press.

Hill, R. K. (2016). What an Algorithm Is. *Philosophy & Technology, 29*, 35–59. https://doi.org/10.1007/s13347-014-0184-5

Jaton, F. (2017). We get the algorithms of our ground truths: Designing referential databases in digital image processing. *Social Studies of Science, 47*(6), 811–840. https://doi.org/10.1177/0306312717730428

Jiang, L., Bhagavatula, C., Liang, J. T., Dodge, J., Sakaguchi, K., Forbes, M., Borchardt, J., Gabriel, S., Tsvetkov, Y., Rini, R. A., & Choi, Y. (2022). *Can Machines Learn Morality? The Delphi Experiment*. https://doi.org/10.48550/arXiv.2110.07574

Johns, F., & Compton, C. (2022). Data jurisdictions and rival regimes of algorithmic regulation. *Regulation & Governance, 16*, 63–84. https://doi.org/10.1111/rego.v16.1

Kant, I. (1966). *Kant's gesammelte Schriften. Herausgegeben von der Königlichen Preußischen Akademie der Wissenschaften. Band XXIX. Kant's Vorlesungen. Band I: Vorlesungen über Logik. Zweite Hälfte.* Walter de Gruyter.

Kant, I. (2000). *Critique of the power of judgment.* Ed. P. Guyer, transl. P. Guyer & E. Matthews. Cambridge University Press.

Klinger, F. (2011). *Urteilen*. diaphanes.

Knuth, D. E. (1972). Ancient Babylonian Algorithms. *Communications of the ACM, 15*(7), 671–677.

Knuth, D. E. (1997). *The Art of Computer Programming. Volume 1: Fundamental Algorithms* (3rd ed.). Addison-Wesley.

Krämer, S. (2014). Mathematizing Power, Formalization, and the Diagrammatical Mind or: What Does 'Computation' Mean? *Philosophy & Technology, 27*, 345–357. https://doi.org/10.1007/s13347-012-0094-3

Krishnan, M. (2020). Against Interpretability: A critical examination of the interpretability problem in machine learning. *Philosophy & Technology, 33*, 487–502. https://doi.org/10.1007/s13347-019-00372-9

Liddell, H. G., & Scott, R. (1996). *A Greek-English Lexicon* (9th ed.). Clarendon Press.

Lindia, M. S. (2022). Gadamer in a wired brain: Philosophical hermeneutics and neuralink. *Philosophy & Technology, 25*, 27. https://doi.org/10.1007/s13347-022-00522-6

Lix, L., Vasylkiv, V., Ayilara, O., Dahl, L., Poppel, A. & Al-Azazi, S. (2022). A Synthesis of Algorithms for Multi-Jurisdiction Research in Canada. *International Journal of Population Data Science* 7(3): 137. https://doi.org/10.23889/ijpds.v7i3.1911

Malpas, J. (2020). The necessity of judgment. *AI & Society, 35*, 1073–1074. https://doi.org/10.1007/s00146-020-01021-y

Metz, C. (2021). Can a Machine Learn Morality? *The New York Times*, 11/19/2021. Retrieved December 5, 2023, from https://www.nytimes.com/2021/11/19/technology/can-a-machine-learn-morality.html

Moschovakis, Y. N. (1998). On founding the theory of algorithms. In H. G. Dales & G. Oliveri (Eds.), *Truth in Mathematics* (pp. 71–104). Clarendon Press.

Moschovakis, Y. N. (2001). What Is an Algorithm? In B. Engquist & W. Schmid (Eds.), *Mathematics Unlimited: 2001 and Beyond* (pp. 919–936). Springer.

Owens, K., & Walker, A. (2020). Those designing healthcare algorithms must become actively anti-racist. *Nature Medicine, 26*, 1327–1328. https://doi.org/10.1038/s41591-020-1020-3

Oxholm, C., Christensen, A.-M. S., & Nielsen, A. S. (2022). The ethics of algorithms in healthcare. *Cambridge Quarterly of Healthcare Ethics, 311*, 119–130. https://doi.org/10.1017/S0963180121000864
Pirozelli, P. (2022). Sources of understanding in supervised machine learning models. *Philosophy & Technology, 35*, 23. https://doi.org/10.1007/s13347-022-00524-4
Possati, L. M. (2020). Digital Reflective Judgement: A Kantian Perspective on Software. *Critical Hermeneutics, 4*, 1–33. https://doi.org/10.13125/CH/4317
Rai, A. (2020). Explainable AI: from black box to glass box. *Journal of Marketing Science, 48*, 137–141. https://doi.org/10.1007/s11747-019-00710-5
Rangareddy, H., & Kurpad Nagaraj, S. (2022). Artificial Intelligence and Healthcare. *Journal of Clinical and Diagnostic Research, 16*(11), YI01–YI03. https://doi.org/10.7860/JCDR/2022/56148.17020
Raveendhran, R., & Fast, N. J. (2021). Humans judge, algorithms nudge: The psychology of behavior tracking acceptance. *Organizational Behavior and Human Decision Processes, 164*, 11–26. https://doi.org/10.1016/j.obhdp.2021.01.001
Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536. https://doi.org/10.1038/323533a0
Schuhmann, C. & Taddicken, M. (2021). Algorithms as Research Objects for Communication Science. In C. Schuhmann & M. Taddicken (Eds.), *Algorithms and Communication* (pp. 7–23). SSOAR. https://doi.org/10.48541/dcr.v9.1
Schuilenburg, M., & Peeters, R. (2021). The algorithmic society: An introduction. In M. Schuilenburg & R. Peeters (Eds.), *The Algorithmic Society: Technology, Power, and Knowledge* (pp. 1–15). Taylor & Francis Group.
Seele, P. (2022). AI ethics inflation. Delphi and the restart of theory. *AI & Society*. https://doi.org/10.1007/s00146-022-01414-1
Seele, P., & Schultz, M. D. (2022). From Greenwashing to Machinewashing: A Model and Future Directions Derived from Reasoning by Analogy. *Journal of Business Ethics, 178*, 1063–1089. https://doi.org/10.1007/s10551-022-05054-9
Slobogin, C. (2021). *Just Algorithms: Using Science to Reduce Incarceration and Inform a Jurisprudence of Risk*. Cambridge University Press. https://doi.org/10.1017/9781108988025
Sommerer, L. M. (2022). *Self-imposed Algorithmic Thoughtlessness and the Automation of Crime Control: A study of person-based predictive policing and the algorithmic turn*. Nomos. https://doi.org/10.5771/9783748929550
Sutton, R.S. & Barto, A.G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). The MIT Press.
Ugwudike, P. (2022). Predictive Algorithms in Justice Systems and the Limits of Tech-Reformism. *International Journal for Crime, Justice and Social Democracy, 11*(1), 85–99. https://doi.org/10.5204/ijcjsd.2189
Van Zantwijk, T., Gabriel, G. & Ogorek, R. (2001). Urteil. *Historisches Wörterbuch der Philosophie online*. https://doi.org/10.24894/HWPh.5515
Von Eschenbach, W. J. (2021). Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philosophy & Technology, 34*, 1607–1622. https://doi.org/10.1007/s13347-021-00477-0
Watson, D. S. (2023). On the philosophy of unsupervised learning. *Philosophy & Technology, 36*, 28. https://doi.org/10.1007/s13347-023-00635-6
Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgement to Calculation*. W. H. Freeman and Company.
Wiens, J., Price, W. N., & Sjoding, M. W. (2020). Diagnosing bias in data-driven algorithms for healthcare. *Nature Medicine, 25*, 25–26. https://doi.org/10.1038/s41591-019-0726-6
Zednik, C. (2021). Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *Philosophy & Technology, 34*, 265–288. https://doi.org/10.1007/s13347-019-00382-7
Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? *Philosophy & Technology, 32*, 661–683. https://doi.org/10.1007/s13347-018-0330-6