Penultimate draft. Forthcoming in *Philosophy of Science for Machine Learning: Core Issues and New Perspectives*, edited by Juan Duran and Giorgia Pozzi, *Synthese Library,* Springer

MACHINE LEARNING AND THE ETHICS OF INDUCTION

Emanuele Ratti[1]

University of Bristol

Abstract. This chapter analyzes the inferential structure of machine learning (ML) systems, and shows how these can be value-laden in unexpected ways. ML systems follow an inductive inferential strategy, which is based on two components. First, there is the basic assumption that we are entitled to predict future events on the basis of past occurrences because the world will not drastically change. This assumption is called 'uniformity of nature' (UoN). Second, 'canons of inductive inference' (CIIs) are required to narrow down the set of possible hypotheses that one can generate from UoN. Debates on the ethics of ML have focused on CIIs. Here I show that UoN plays an important ethical role, in particular in eroding human agency.

Keywords: machine learning; induction; ethics of AI; AI ethics

1 INTRODUCTION

In the past few years, there has been a growing concern about the problem of 'algorithmic bias' in machine learning (ML) tools. In a general sense, 'bias' is understood as a systematic error due to a deviation from a norm or standard, where norms or standards can be statistical, moral, epistemic, etc (Danks and London 2017). Research has shown that, because of the way they are designed, algorithmic systems incorporate plenty of biases – in particular social and moral biases – that cause harmful impacts, especially to minorities. Stemming from these works, several proposals on how to identify biases and design algorithmic systems that are less biased have been developed (Fazelpour and Danks 2021). All in all, this rich literature is focused on how choices behind the design of algorithmic systems have repercussions on their outputs and impact individuals differently.

In this chapter, I want to show how investigating ML tools from the point of view of their inferential structure can be fruitful. In particular, I will show that (supervised) ML tools follow an inductive inferential structure, and that considering different components at play in induction allows to make different types of ethical arguments about the nature of ML tools.

The structure of the chapter is as follows. In Section 2, I will make the case that (supervised) ML tools follow an inductive strategy. In particular, this strategy is based on two components. First (2.1), there is the uniformity of nature (UoN) component, according to which we are allowed to infer future observations on the basis of similar past observations because of assumed uniformity between the past and the future. This principle is insufficient to ground any reliable inference, as the past and the future may look alike in an indefinite number of ways. UoN is supplemented by non-evidential background assumptions (that is, the second component), typically called 'canons of inductive inference' (CII). CIIs make sure that the space of possible ways in which the past and future are similar is narrow enough to allow inference. I then show (2.2) how ML follows this particular strategy, where CIIs take the form of different choices that shape the final ML model, and elucidate (2.3) how debates in ethics

[1] mnl.ratti@gmail.com

of ML are mostly about CIIs [see Chapter 14 in this volume]. In Section 3, I turn to the analysis of UoN, first by describing the 'harmless' role that it plays in the natural sciences (3.1), and then (3.2) to show how, in the context where human agency is salient, UoN can actually erode human agency itself.

2 INDUCTION AND MACHINE LEARNING

Let me start by describing the main features of induction and inductive inferences, and in which sense ML systems are inductive.

2.1 Varieties of Induction and Machine Learning

Here I refer to 'inductive inference' in a somewhat broad way - "whenever we note that evidence lends support to an hypothesis (…) while not establishing it with deductive inference" (Norton 2005, p 1). Induction is said to be *ampliative* because, unlike deduction, the conclusion is not contained in the premises, or the conclusion is more than just a reformulation of the available evidence. According to these broad indications, there are several types of inferences that (at least at first) can count as 'inductive'. These include enumerative induction; abduction; eliminative induction (Earman 1992; Norton 1995); variational induction; material induction etc. Let me briefly illustrate them.

Enumerative induction – also known as induction by simple enumeration - is the most common form of induction, an archetype of inductive generalization (Norton 2005). What enumerative induction is supposed to establish is a general dependency between two types of events. The idea is that if there is a sufficient number of co-occurrences of these two types of events, then we can infer that there is a general dependency between them (Canali & Ratti 2024). It is obviously ampliative: if the co-occurring instances are seen as premises, and establishing a general dependency is the conclusion, then the conclusion is not contained in the premises.

Eliminative induction is also known as 'induction by means of deduction' (Hawthorne 1993), 'strong inference' (Platt 1964), or 'eliminative inference' (Forber 2011). It works in the following way. First, a set of premises generates a finite set of competing hypotheses. These premises are of difficult characterization, but the idea is that they represent a prior state of individual practice (Kitcher 1993; Ratti 2015) which drives the selection of competing hypotheses in the first place. Next, other premises and new evidence stimulates a process where hypotheses are discarded until the 'true' hypothesis remains. Famously, this scheme captures Sherlock Holmes' strategy that can be summarized with the slogan 'when you have eliminated the impossible, whatever remains, however improbable, must be the truth'. Eliminative induction has a renowned tradition in science and philosophy, from Mill to Dretske. But whether it is really a kind of induction is an open question, as the means are in fact deductive, and its ampliative nature is dubious, given that the hypothesis ending up to be true is one of premises that go through the process of elimination (so, in a sense, the conclusion is contained in the premises).

Variational induction is not as common as the previous two. A comprehensive characterization is provided by Pietsch (2021; 2022), who borrows the term from Russo (2009). Pietsch conceptualizes it as a method to infer causal relevance between a phenomenon and a set of circumstances, "by relying on variational evidence, i.e., on evidence that tracks changes in a phenomenon resulting from systematic variations of circumstances" (Pietsch 2021, p 29). A precursor of variational induction is, according to Pietsch, Mill's methods, in particular the method of difference and the method of agreement. In my understanding, variational induction is ampliative because the conclusion (i.e., establishing causal relevance) is not contained in the premises (i.e., variational evidence).

Abduction/inference to the best explanation is another form of inductive inference. According to Norton (2005) there is a family of inductive inferences known as 'hypothetical induction', whose archetype is the so-called 'saving the phenomena' strategy: "the ability of an hypothesis to entail deductively the evidence is a mark of its truth" (p 5). Abduction is an extension of this strategy, where an additional constraint is put on the hypothesis. In addition to entail the evidence, the hypothesis must also explain it. This is considered to be ampliative, because the conclusion does not follow from the premises.

Now, it is usually said that ML works by adopting a form of inductive inference. Given that there are different types of inductive inference, which one does characterize induction? Because ML is a rather broad family of techniques and algorithms, we might expect that more than one type of inductive inference will fit it.

We can exclude two cases from the start. First, ML does not fit a type of 'inference to the best explanation' type of induction. This is because ML algorithms work by fitting a real function, and so in a sense they aim at 'saving the phenomena', but they do not pose explanatory constraints on the models themselves. To be more precise, what is learned by an algorithm is called a ML model (Facchini and Termine 2021). This is a learned function, consisting in a mathematical representation of the statistical patterns that the algorithm has learnt from a data set. Sometimes it is said that the function or the model is selected, among the ones possible. But other than fitting the data in the right way, the model need not meet explanatory requirements or desiderata to be 'selected' by the algorithm[2]. Of course, scientists using ML models may opt for the model that better explains the data from a variety of different angles, but the algorithm *per se* is neutral with respect to this particular explanatory constraint. We can also exclude eliminative induction as well. Even if eliminative inferences characterizes disciplines making ample use of ML techniques like genomics, the eliminative inductive strategy is usually ascribed to the whole system of inquiry, rather than the ML algorithm (Ratti 2015). In other words, scientists in genomics may use ML techniques in their own eliminative inductive investigations, but the algorithm *per se* does not proceed by eliminative inferences. This is because the algorithm does not start with any predefined hypothesis (Pietsch 2022).

Let me now turn to the positive cases. According to Pietsch (2022), variational induction seems to characterize a significant number of strategies used in Big Data analytics. He notices that in many cases (e.g. decision trees, naïve Bayes, neural networks), ML algorithms are in fact trained by relying on variational evidence, where confirmation or performance increases with the variety of evidence. For instance, decision trees are constructed by taking into account both positive and negative instances, and the relevance of predictor variables is established in terms of the difference that these variables make for the accuracy of a classification – these can be seen as characteristics of variational induction. The same can be said, according to Pietsch, for neural networks. He also recognizes that simple enumerative induction characterizes some methods, including association rules. But it is possible to show that variational induction and enumerative induction are based on similar ideas, and that in a way variational induction can be translated in terms analogous to enumerative induction. One important aspect that the two have in common is that both enumerative and variational induction assumes that more data is generally better (Canali and Ratti 2024). As mentioned above, in enumerative induction it is generally assumed that we should provide a sufficient number of co-occurrences between two events to make the case for their dependency, and in absence of specific guidelines on how much is enough, then we should assume that the higher

[2] But, as we will see, they come with several other constraints

the amount, the better. But simple enumerative induction is agnostic with respect to the type of evidence of co-occurrence that one should use. Variational induction is more precise on this; we need to increase the number of features considered (or 'circumstances') in order to observe "as many different situations in terms of changing circumstances as possible" (Pietsch 2021, p 30). Under this lens, variational induction is a more precise version of enumerative induction when it comes to the kind of evidence required for inference. Sure, one can say (as Pietsch does) that variational induction establishes causal relations, while enumerative induction does not, but whether variational induction is legitimate for causal inference – especially when understood in the context of ML, where causal inference is rare - is contentious (Canali and Ratti 2024). In other words, I take that it is often possible to reconstruct cases of variational induction-based ML as specific cases of enumerative inductions. In fact, I agree with Harman and Kulkarni (2007) when they point out that enumerative induction "applies to many examples of machine learning, including perceptron and feed-forward neural net learning" (p 25)

2.2 Enumerative Induction in Machine Learning

Now it is time to show more precisely the inductive structure of ML. Let me go back to enumerative induction.

In (2023), Johnson explores the relationship between methods like ML and the value-free ideal in science. She does this by tracing back the problematic nature of this relation to Hume's problem of induction. Typically, Hume's problem is understood as being about the justification of induction. In deduction, true premises will always guarantee true conclusions, which is not what happens in induction. Consider for instance a case where you have a set of premises arguing that many apples have been found to have seeds and no apples have been found not to have seeds. From these premises, the hypothesis/conclusion/prediction that the next apple will have seeds can turn out to be false, even if the premises are true. One may say that the more apples with seeds, the more likely that the next apple will have seeds. Psychologically, this is what we would expect, but whether we are justified in having this expectation is a different matter. In fact, it does not matter how much support a hypothesis has; there is nothing that can justify why known instances would provide any degree of support to predict unknown instances: there is "nothing logically at odds with the world suddenly becoming drastically different" (Johnson 2023, p 6). This means that we cannot justify (without further constraints) the practice of using patterns identified in evidence to predict novel cases. However, this practice is common in the sciences, and in ML in particular.

This problem is typically overcome by making one or more non-evidential assumptions. The most basic assumption that one can do – and that is used by Hume himself – is assuming that nature is uniform. Given that the world will not be drastically different tomorrow, we can use the past to reliably predict what there will be in the future on the basis of similarities between the world as it is now, and the world as it will be tomorrow. Let us call this UoN (i.e., Uniformity of Nature). The problem with UoN is that, as noted by Johnson and many others, there are many ways in which the world of today can be similar to the world of tomorrow. Today's world may have apples with seeds and blue skies, but tomorrow's world – while still having blue skies – may have apples without seeds. This implies that there is not much guidance in restricting the set of possible 'similarities' between past and future worlds (which then translate into possible hypotheses/predictions about novel instances) to a tractable size.

Therefore, UoN can be a ground norm used for inductive inferences in the sciences, but we need also other assumptions, which will provide additional "non-evidential way[s] of limiting the hypothesis space to a tractable size" (Antony 2016, p 161). Johnson calls these assumptions ‘biases’. In the history of science, philosophy of science, and epistemology, assumptions vary, and they are known under many names, such as cognitive heuristics (Bechtel and Richardson 2010), theoretical/epistemic virtues (Kuhn 1977; McMullin 1983), non-cognitive values (Douglas 2009), canons of inductive inference (Levi 1960), etc. I prefer to use ‘canons of inductive inference’ (CII), because it leaves open the question of whether these assumptions are epistemic or not – CII avoids entering this mine field. Moreover, unlike the use of ‘biases’, using CII is not normatively charged (even though Johnson means ‘bias’ without normative connotations). So the idea is that we are entitled to infer from known instances the prediction of novel instances because of UoN *plus* a host of other assumptions that would constrain the way patterns should look like. Certain kinds of CIIs (most notably, epistemic and non-epistemic values) are also invoked to solve old problems in philosophy of science, such as underdetermination (Longino 1990) – an example being the so-called ‘gap feminist empiricism’ (Solomon 2012), where the impasse of theory choice is solved by values.

By following Harman and Kulkarni (2007), it is possible to show that typical problems in statistical learning theory – whose applications are embedded in many ML systems – have an inductive structure of the kind just showed. In particular, consider classificatory problems. At a basic level, this is a problem about reaching conclusions about the next X on the basis of observing prior Xs, where the number of correct conclusions should be as high as possible. We want to find a method for doing this – in this context, the method is sometimes called ‘a rule’, or ‘model’. The rule is a classification rule because its goal is to help us to classify something as X, on the basis of observed characteristics of previous Xs.

This way of framing the problem seems misleading though, since it reflects old problems. Consider the question ‘what kind of classification rule do we want?’ in the context of, for instance, distinguishing between French Bulldogs and Boston Terriers (Castro 2022). The rule we are looking for will be in the form of “if you run into a dog that is either a French Bulldog or a Boston Terrier, take measurements X and Y (e.g. height and weight); if x is below z, and y is above z, then it is a French Bulldog; otherwise it is a Boston Terrier”. But how do we establish if this is the best rule? One may say that the best classification rule is the one with least errors. But, as Harman and Kulkarni note, this is not a precise way of identifying the best rule. If we think about the rules to select as forming a set, then an unconstrained classification problem like this one will have a set that includes all possible rules, and as a consequence there will be also many rules with similar error rates (as well as different types of errors), which will classify new cases differently. To avoid this problem, there should be restrictions on the rules that can be selected. This is exactly the same issue motivating the use of CII as a way to constrain UoN. Harman and Kulkarni (2007) call these restrictions ‘inductive biases’. They say that the framing of a classificatory problem,

> “must prefer some rules over others. It must be biased in favor of some rules and against others. If the method is the sort of enumerative induction which selects a rule from C [i.e. rules for classifying] with the least error on the data, there has to be a restriction on what rules are in C. Otherwise, we will not be able to use data in that particular way to restrict the classifications of new cases” (p 23)

This reflects typical problems in underdetermination and induction. There can be different hypotheses that make sense of the same data (in our case, there can be different rules with the least errors based on the same observations). Because of this, we need to restrict the range of possible hypotheses: there can be different ways in which the present and future worlds are similar, and these will grant different predictions. In the case discussed by Harman and Kulkarni, restricting the possible rules is necessary, and this is done by adopting an inductive bias. In the context illustrated by Johnson, in order for theory/hypothesis choice to overcome impasse of the vagueness of UoN or the underdetermination problem, we need CIIs (e.g. theoretical/epistemic virtues; non-epistemic values; biases; cognitive heuristics; etc). But while the contexts are slightly different, they are pointing to the same thing – in fact, the inductive bias Harman and Kulkarni talk about is just one instance of CIIs. It is also important to notice that CIIs do not do all the work; in the case of statistical learning theory and classificatory problems, UoN plays a substantial role too: it is only under the assumption that the world will not drastically change that the claim 'given these past Xs, here is a novel instance of X' makes sense – without UoN, there would be no starting point.

This framing of classificatory problems in statistical learning theory represents exactly the kind of problem solved by most supervised ML systems. Rather than using the word 'rule', several uses the word 'model'(i.e. it is the model of the data that is selected (James et al 2013)) or simply function[3]. The Xs that are used to select the best model are typically training data sets (i.e., data used to train the algorithm) and test data sets (i.e. data used to test the accuracy of the model selected as a result of first training). These data sets are 'labelled', in the sense that there is a label attached to different data points indicating what the data is about (i.e. the target variable such as a cancer sample; a dog; etc). After a model is 'selected', it is applied to unlabeled data sets to automatically classify new cases. In this process, CIIs play substantial roles. I take canons here to mean something broad, as anything that constrains or shapes the set of possible models that will be selected: *CIIs in ML are any choice made in building these algorithmic systems that constraint the 'selection' of the final model*. For instance, in another article (Ratti and Graves 2021) I decompose the ML construction process into various phases, and show how technical and micro choices in how to prepare data even before the actual training can have important consequences for the final model selected. Johnson focuses on similar things. In fact, it is a truism that predictive models can be built in different ways, and these will reflect different 'patterns' found in the data. But norms in the ML community will restrict the range of acceptable 'models' or 'methods' – from very mundane choices like parameters, to more important things such as best practices for when to use certain algorithms rather than others on the basis of the problem at hand, to different measures of performance (e.g. accuracy, precision, sensitivity, specificity, etc).

To sum up, many supervised ML systems proceed by induction. What they do is solving classification problems typical of statistical learning theory. These are conceptualized as problems of identifying the next X on the basis of observation of former Xs. The basis for the prediction is a rule R. Such a rule is selected by choosing among alternatives within a set of rules. This set is limited and constrained by what are called inductive biases. An additional and upstream assumption is that it makes sense to identify new Xs on the basis of past observed Xs only under the assumption of UoN. ML systems follow this structure, with 'models' usually standing for rules, and 'inductive biases' as being any constraint that is posed on the functioning of algorithms that impacts model selection. This structure is a typical inductive inference strategy, where UoN makes sure that the world of past observations will not look drastically different from the world of new instances, and CIIs constrain the set of competing

[3] 'Function', 'model', 'rule' (in case one refers to a classificatory problem), are used interchangeably in this literature. This can be easily checked in various introductory texts. For instance, James et al (2013) set the goal of statistical learning as the estimate of a function *f* mapping input variables into output variables. In discussing the trade-off between prediction accuracy and interpretability of the function *f'* which estimates the real function *f*, they refer to 'model interpretability', and 'model predictive accuracy', making clear that they are referring to the function *f'*. Burkov (2019) identifies the goal of ML as the construction of statistical models, and then defines the model explicitly as a function (p 5). Jiang (2021) also goes back and forth using 'model' and 'function' interchangeably. Harman and Kulkarni (2007) talk about 'rules of classification' as rules "for using observed features of items to classify them" (p 29). Rules are found by using data, where data are thought to be a random sample coming out of a background probability distribution, about which nothing is assumed. The best classification rule is the one that makes the least errors – in other words, the one that estimates as much as possible the background probability distribution. But the background probability distribution is a function, so the rule will be *f'* as meant by James et al and, ultimately, a model.

hypotheses that can be inferred from the past observations.

2.3 AI Ethics is about canons of inductive inference

I have shown that ML systems are typically designed to solve problems formulated within the remit of statistical learning theory, which in turn can be characterized as going through inductive procedures. In induction, two components stand out in making inferences possible. First there is UoN, namely the assumption that the world will not drastically change, thereby making it easier for us to predict the novel on the basis of past observations. The second component are CIIs, which are constraints posed on UoN, which facilitate the restriction of the space of hypotheses/predictions/rules to consider.

Debates on the ethical, societal, and political dimensions of ML tools are mostly about CIIs. The large literature on algorithmic bias (Fazelpour and Danks 2021) focuses on the different aspects of ML systems that may be salient or relevant for ethical, societal, and political issues. It has been shown how different choices in designing ML systems can be motivated by values whose nature is non-epistemic (Ratti and Graves 2022; Johnson 2023; Biddle 2020) [see Chapter 15 in this volume]. For instance, Biddle (2020) reconstructs the systematic presence of value-laden decisions in designing algorithmic systems in the justice system. In the typical way scholars of the inductive/epistemic risk literature proceeds, Biddle shows how at each step of the ML pipeline more than one decision is possible, and each choice comes with related risks. In deciding which risks are acceptable and which are not, ML practitioners use value-judgement, which often involves endorsing certain ethical, social, and political values. All these choices motivated and/or justified with values (Ward 2021) can be conceptualized as choices pertaining to CIIs. For instance, Biddle argues that in collecting data for building a tool to predict recidivism, questions about the baseline population or about the features that correlate with crimes come up – and choosing certain features rather than others is based on values and has consequences for values. Or another typical problem is about the evaluation of the performance of ML systems. In choosing the relevant metric (be it accuracy, precision, sensitivity, specificity or something else), the importance of increasing or decreasing false positives and false negatives is fundamental, and choosing one over the other reflects certain constraints that we want ML system to act within – and these are, ultimately, CIIs. These values may be ‘technical’, social, moral, political, etc, and they all constitute forms of CIIs that shape choices behind ML systems. One can also formulate the problem not by relying on ‘risks’, but rather on ‘contingency’. This is the ‘science and value’ approach developed by Brown (2020). Scientific activities like building a ML system requires practitioners to execute certain actions. These actions are contingent, in the sense that alternative actions are always possible. This is the case also in ML (Ratti 2020). *Practical reasons* resolve the impasse of alternative actions. Pragmatic choices based on practical reasons means taking into account various CIIs in order to orient the design of the ML tool in a direction rather than another. Problems of bias, fairness, privacy, safety, etc, can all be formulated as problems concerning the risks of (or contingent actions behind) constraining the design of ML systems via specific CIIs rather than others.

These observations are relevant to make a specific point: the ethics of ML is focused on a specific component of ML inferential structure, which is CII. However, focusing only on CII overlooks an important ethical dimension of ML tools, which is related to UoN. In the next section, I will show what this means.

3 MACHINE LEARNING AND UNIFORMITY OF NATURE

In the previous section, I have shown in which sense common supervised ML tools follow an inductive inferential structure, by drawing from well-known literature in statistical learning theory. In this section I will delineate some consequences of this. In particular, I will show how in cases where ML is applied to the natural sciences, the inductive structure does not pose any significant ethical problems. But in 3.2 I will show how, in the context where human agency is involved, that same inductive structure is problematic from an epistemic point of view and, most important, ethically controversial. But this is not just because of the typical ethical issues investigated by the ML ethics literature concerning CIIs; rather, the ethical problems are about UoN..

3.1 ML, induction, and the natural sciences

Let us see this inferential structure in action in the natural sciences [See Chapter 14 in this volume]. Let's consider, for instance, how supervised ML tools are used in biology, in particular in genomics. In this field, biologists have been using these tools for more than 20 years. Recently, genomics has benefited from the use of deep learning systems (Eraslan et al 2019), and the first steps have been made in adapting large language models to its context (Consens et al 2023). But these are complicated cases, and we do not need to consider them to show the kind of inferential structure in place in genomics when problems are approached from a ML perspective. For this reason, I will refer to simpler examples of the 'pre-deep learning' era.

Let me start with a basic example of a classificatory task in genomics, such as identifying transcription start sites (TSS) in genomics data (Libbrecht and Nobel 2015; Ratti 2020). Using ML to classify TSSs is a representative example of what people in this context are after. In genomics, the goal is to identify the genetic basis (that is, at the DNA level) of biological phenomena. It is common to gather data about genomes, and then look for biological entities that are biochemically active at the DNA level, including genes, DNA mutations, TSSs, transcription factors, etc. To get a preliminary understanding of TSS, consider the first phases of the mechanism of protein synthesis. Very briefly put, in eukaryotic cells the mechanism starts in the nucleus, where a given portion of DNA molecules (i.e. a DNA sequence known as 'gene') is copied (i.e. transcribed) into a RNA nucleotide sequence, which is known as pre-messenger RNA (pre-mRNA). Pre-mRNA as template-molecule is typically subjected to a process of chemical modification known as *splicing*, where the template is modified by removing some bases (*introns*) and the remaining (*exons*) are assembled together. The new molecule/template is known as messenger RNA (mRNA), and it is moved into the cytoplasm, where other processes (beyond the scope of this brief exposition) will lead to the synthesis of the primary structure of proteins. In this context, TSSs are DNA sequences where the transcription (i.e. the copying mechanism) of a gene approximately starts. We know a great deal about TSSs. For instance, they are usually associated with the so-called TATA boxes, namely sequences of Thymine-Adenine-Thymine-Adenine where transcription factors bind. This is taken to indicate where transcription should approximately start, around 25 nucleotides upstream TSSs themselves. Anytime there is a TATA box, then it is likely to have a TSS nearby. Similarly, TSSs are usually located downstream of CpG islands, which are regions where cytosine and guanine are separated by only one phosphate group. The interesting thing is that, in building a ML system to automatically predict the presence of TSSs, you do not need to use any of this theoretical information about TSSs[4].

Typically, the ML model is built in this way. An algorithm is trained on data sets to find 'rules' or 'models' or 'functions' that can be used to classify TSS sequences in new data. The data collected for the task are 'labeled', in the sense that it is specified whether given sequences contain a TSS. The data set is usually divided into training and test data sets. The

[4] Even though you may need to use your knowledge about TSSs to 'externally' validate the final ML model (see Ratti 2020).

training data set is provided as input to the algorithm, in order to produce a model that can be used to classify TSSs. This model is then ‘tested’ against the test data set. The training set with labeled TSS is the set of known instances, while the test data set is the set of unknown instances. But the latter are ‘unknown instances’ only for the sake of training; in fact this data set is labeled, but it is used only for testing the initial model, in the sense that the model will predict the labels for the testing data set, and then these are compared to the actual labels of the data. The model/rule/function is what we use to infer unknown instances, also beyond testing data sets. The way the model is built reflects various ‘canons’, in the sense that it is possible to have an indefinite number of models/rules/functions fitting the initial data sets, and so we need to restrict the space of possible models/rules/functions. Typical ‘canons’ are decisions taken at each step of the ML pipeline. When deciding which data to collect to use as input for training, several choices must be made. First, one would have to choose the data themselves. In a pioneering case of building a classifier for TSSs, Down and Hubbard (2002) extracted mammalian promoters from the EPD database, but they decided to discard those on human chromosome 22, and those with less than 500 bases of upstream sequences available. While they do not explain why they do this, this is certainly a case where a CII (that is, a choice in building the system) influences how the final model will look like. Another common CII is the so-called ‘feature selection’, which is how one reduces the dimensions of feature vectors by selecting only those features that are seen as relevant for the problem at hand. For instance, in building a support vector-machine (SVM) to classify polymorphisms (that is, DNA and amino acids mutations) for cancer genomics, Capriotti and Altman (2011) end up choosing 51 different features for building the input vector, where these features include components about local sequence environment, sequence profile, gene ontology (GO) terms associated with the relevant DNA region, etc. Another important choice or ‘canon’ is the choice of the algorithm itself (Jiang 2021). Whether one will choose a discriminative algorithm (e.g. nonlinear kernels, decision trees, convolutional neural networks, etc) or a generative algorithm (e.g. transformers, variational autoencoders, etc) will depend on many factors, one of which is the data modality (e.g. image, text, etc). And each algorithm will also come ‘pre-packaged’ with a number of assumptions about parameters, the kind of function family characterizing the model, etc. Moreover, even evaluating the performance of the model relies on an important choice as to which metrics to use (e.g. sensitivity, accuracy, specificity, etc). Finally, it is also important to note that in many cases there is also an ‘domain knowledge consistency check’, namely that the model is not evaluated only via quantitative measures, but also on the basis of whether what it has learnt is consistent with domain knowledge. In the case of TSSs, Down and Hubbard (2002) notice *en passant* that what the model has learnt was consistent with what we know about TSS.. In the case of polymorphisms/mutations, Capriotti and Altman (2011) even construct a new quantitative metric called GO log-odds (LGO) score. This measures the correlation between the effect of a mutation and a related gene ontology (GO) term. As shown in the literature (Leonelli 2016), GO terms are supposed to refer to specific biological entities and activities, By establishing the connection between a mutation and a GO term, one can get a grasp on the strength of the association between a mutation and typical biological processes. In this way, Capriotti and Altman can check whether their model is consistent with the known underlying biology.

The use of (supervised) ML in this context reflects the kind of inductive strategy described in Section 2. Problems are typical classificatory problems of statistical learning theory. These are conceptualized as problems of identifying the next X on the basis of observation of former Xs. In the genomics case we have just reviewed, the problem is to identify a TSS on the basis of data sets about TSS, or to identify certain mutations on the basis of data sets of mutations. The basis of the identification is a rule, which in the case of TSS or somatic mutation is a model. Such a rule/model is selected from a (metaphorical) set of alternatives rules/models, where the set is limited and constrained by what are called inductive biases in statistical learning theory, or in general by ‘canons of induction’ (again, understanding this term very broadly). These ‘canons’ are much wider than the typical list of CIIs in philosophy, and in the case of genomics and ML include all choices that have an impact on the final model. But in these procedures to build genomics ML models, an important role is played by UoN. In fact, it makes sense to identify new TSSs on the basis of past observed TSSs, or new mutations on the basis of past somatic mutations, only under the assumption of UoN. Therefore, ML systems follow an inductive strategy by grounding their inferences on UoN, plus a number of CIIs.

The role that UoN plays should not be overlooked. As an assumption, it plays an important normative role. I take ‘normative’ in the way discussed by Hans Radder in his work about the normativity of artifacts (2019). According to Radder, an artifact is normative when its realization implies that in the context in which an artifact used, people ought to follow certain (implicit and/or explicit) norms (where a norm is a directive about what to say, do, or be), where these norms will facilitate and not disturb the functioning of the artifact itself. The nature of norms vary, as they can be social, political, moral, epistemic, technical, cognitive, etc. Background assumptions in science can play an analogous ‘normative’ role. For instance, theoretical virtues are background assumptions on how to choose theories. Committing to the theoretical virtue of simplicity means that, in theory choice, you *ought to* choose the simplest theory among the alternatives, because simplicity is seen as a virtue/value, namely something desirable for scientific theories. UoN is a basic background assumption, and it can be interpreted as an epistemic norm. In fact, *via* UoN, you are entitled to use the past to infer about the future, on the basis of similarities between the past and the future. But you are not just entitled: in a sense it creates an epistemic obligation to do so. This can be put as a conditional: if one wants to predict what will happen, then the past *ought to* be used as a basis for inference. In a context like genomics, UoN is also a *desirable* epistemic norm (which, in a sense, is already implicit in the fact that it is a norm). Through UoN, we have a basis for controlling natural phenomena. From the fact that we can predict the future status of a biological phenomenon on the basis of its past observations, we can also vary the conditions that allow us to predict future occurrences, and we can control the phenomenon itself. This work of ‘varying conditions’ is done by selecting the right CIIs. But the very possibility of having a conversation about CIIs – that is, about how to restrict the space of possible similarities between the past and the future - is made possible by UoN, as it is only through UoN that it makes sense to take the past as a basis for predicting the future.

3.2 ML, induction, and human agency

UoN is considered mostly harmless, and in fact there is not much debate about it. It is of course important in the natural sciences, as it sets the starting point for many investigations, and it is certainly not considered controversial from other points of view. In particular, it is not considered ethically problematic, and as such it should not be the focus of scholars working in e.g., the ethics of ML. However, here I claim that, in cases where ML is used to take decisions concerning human agency, UoN is not just an epistemic norm – it also has an ethical dimension. This has some interesting consequences. In discussions on ethics of ML, ethical issues are usually identified as problems with CIIs. The idea is that a particular combination of CIIs generate unfair outcomes, e.g. biased data sets, or the wrong performance metrics, or a problematic feature selection process, etc. Under this lens, there will always be, at least in principle, a way to shape CIIs to make the algorithmic system more ‘ethical’. However, sometimes we might just want to argue that we *should not use* ML in a certain context – and this is what our ethical analysis of UoN can provide, unlike the literature focusing on CIIs. Let me unpack this.

In the context of natural sciences such as genomics, UoN is an assumption that is epistemically salient. The conditional statement put forth by UoN in this case is: if you want to investigate biological phenomena and be able to predict and control them, then you *ought to* use the past as a basis for your inferences. What UoN is saying, is that there are ways in which TSSs or somatic mutations of the past are indeed very similar to TSSs or somatic mutations of the present, and even of the future. But in the context where human agency is involved, acting under the assumption that there are strong reasons to take the future as similar to the past is not just a harmless epistemic norm.

First, it is epistemically problematic, unlike in the natural sciences. One may just notice that predicting human behavior is notoriously difficult. For instance, in political science (Larsen 2018) there have been discussions about this very issue for a long time. On the one hand, the side of predictivists claims that we can indeed discover patterns and regularities in human behavior, and these will form the basis of reliable predictions. On the other hand, anti-predictivists are quick in pointing out that the track-record of predictivism is quite poor, and includes some spectacular failures like the inability to see Trump's election coming; Obama's election; the 2008 economic crisis; etc. This is, in part, an epistemic problem, in the sense that "every future state [about human behavior] comes about as a unique result of a unique interplay of many factors – a one-of-a-kind situation for which there is no normal distribution". Moreover, the unpredictability of human behavior is exacerbated by the fact that humans will respond unpredictably to policies, which may then backfire "or produce new problems elsewhere in the policy space" (Larsen 2018, p 318). This does not mean that nothing at all can be predicted; in fact, one can take Hayek's distinction between point and pattern predictions[5], and claim that maybe pattern predictions are possible. I am not taking any side here, but I just want to show how controversial the idea of predicting is when it comes to human agency.

Second, in the context of human agency, assuming that we have strong reasons to believe that the future will be similar to the past is ethically controversial. Here, I define 'ethical' as pertaining to views on how one ought to live their own life. This a typical virtue-theory or Neo-Aristotelian way of defining ethics, which has found a recent revival in the philosophy and ethics of technology (Vallor 2016; Ratti and Stapleford 2021; Ferdman and Ratti 2024). According to this view, something is ethically salient when it shapes not just human lives concretely (as many CIIs do by shaping the final ML model), but also conceptions or views that individuals hold on how they ought to live their lives, as well as their perceptions on what they can possibly achieve. For instance, assaulting individuals in a neighborhood is not just ethically controversial because you are actually harming other individuals, but also because you are shaping individuals' perception of how they ought to live their lives (e.g. living in fear; not going out when it is dark; etc). Norms that impinge on these aspects are normative from an ethical point of view. Given this way of understanding 'ethical', the ethically controversial aspects of UoN in ML is two-fold:

- the past is usually subjected to a constant ethical evaluation by individuals, e.g. how they wish they could change or not; how it is seen as valuable or desirable or not; how the future can be shaped on the basis of the desirability of past experiences; etc. This can include very mundane things like wanting to change route to go to work to avoid rush hour or traffic, or more important things like wishing to have done a certain degree instead of another, to much more important things like. But saying that the future has

[5] Point predictions are about a specific element of a given phenomenon, while pattern predictions are about "some of the general attributes of the structures that will form themselves, but not cotaining specific statements about individual elements" (Hayek 1974)

to be similar to the past makes any reflective exercise on the past quite useless, as there is little point to reflect on the past if you are just doomed to repeat it, other than just gaining a better understanding of one's predicament. Therefore, we might say that something like UoN bypasses a moral evaluation of the past.

- the future is indeed the realm of agency and, as such, an important focus of ethics. Again, ethics is about how one ought to live their own life, and one's conception of living well. Whatever 'good life' one sees fit for him/herself, it will consist in certain plans and goals, and it will be realized by instantiating certain patterns of actions or behaviors. But by saying that the future will be similar to the past, we are bypassing those fundamental ethical and forward-looking considerations that any human being will gauge in order to exercise substantial freedoms. In other words, UoN can potentially erode human agent's substantial freedom to an open future by establishing how the future ought to look like

This commitment to erode the freedom to an open future is relevant to the ML context. Consider the well-discussed case of the use of ML tools to predict recidivism in the justice system (Biddle 2020; Pruss 2021). In the specific case of COMPAS, the debate ProPublica vs Northpoint/Equivant was all about CIIs. One crux of the controversy was that, according to ProPublica, COMPAS was racially biased because it did score poorly on a measure called 'predictive equality' (Castro 2022), and as a consequence blacks were more likely to be falsely classified as 'high risk', while whites were more likely to be falsely classified as 'low risk'. Northpointe/Equivant argued that that the tool was not racially biased at all, because it scored well on another measure called 'calibration', where risk group defendants were found to reoffend at similar rates, regardless of race. This debate, and the analysis under the lens of frameworks such as epistemic risk (Biddle 2020), sees the ethical issue as a problem of how we align the science to the right values. On this analysis, it is always possible (at least in principle) to adopt a different combination of CIIs, and make sure that a tool like COMPAS scores well in predictive equality, or any other measure[6]. But look at the role that UoN is playing. What a tool like COMPAS does because of UoN, is to estimate what an individual's conduct in the future will be on the basis of what other individuals (similar to the individual in question) have done in the past. In a sense, UoN is establishing the very fact that we should expect individuals' agency to unfold exactly as observed in the past (where the specific range is later established by CIIs). But one may use other considerations as basis for the inference. For instance, one may just base the estimation of future conduct on whether a defendant does indeed regret past conduct, or on the basis of how sincere the defendant's plans for the their future life are, or on the basis of the conditions that led the defendant to commit a certain crime (e.g. poverty). Therefore, one might make a strong claim, and say that our future agency should not be evaluated on the basis of how similar agents have acted in the past. This has been expressed in the literature as saying that we should not use 'bare statistical evidence' in court proceedings (Schmidt et al 2023). This, I think, may potentially apply even to 'individualized evidence' if this is the *only* criterion considered. One may argue, for instance, that agents should have the opportunity to reflect on their own past, and be evaluated on the basis of a combination of considerations about their own past actions, their intentions and plans about the future. In other words, one can argue that you ought to evaluate future agency on the basis of something different than just the past track-records of individuals or past track-records of individuals with a similar agency.

Debating the specific similarities between past and present agents (e.g. race, educational background, etc) as the basis through which the evaluation takes place is already

[6] It should be noted that ML tools cannot be tweaked to achieve a high score for all relevant measures, as various impossibility theorems have shown

framing the debate in terms of CIIs (most notably, in terms of feature selection). By addressing UoN, one can claim that setting up the evaluation of individual agents from the point of view of past observations as a ground criterion to predict the future is akin to denying agency to individuals in the first place, as their future actions are not determined by their own plans and goals, but *solely* by what other similar agents (or themselves) have done in the past. One may say that it is not the case that these decisions are taken on the basis of *only* considerations of past track-record. However, this does not consider that the use of ML – based on UoN – tends to promote the consideration of factors pertaining to past track-records as the 'objective' factors to follow.

These considerations about UoN in the ML context have interesting consequences. By shifting the focus from CIIs to UoN, one can argue that we ought not to use ML tools in specific contexts where human agency is salient. The view can be expressed as a conditional: if one wants to promote human agent's substantial freedom to an open future, then one ought to promote opportunities for agents to plan their lives on the basis of their own goals, and views; but ML tools, because of the focus of UoN, will consider agents' future action as a predictive problem to be solved on the basis of past instances, rather than as a problem concerning what agents plan to do on the basis of their own goals. This means that, from a ML perspective, the agent's future is not open; rather, it is similar to what has already happened. Therefore, ML tools, because of UoN, tend to distort human agency, and ought not to be used by those who want to promote agents' substantial freedom to an open future. To put more bluntly, focusing on UoN is one way to motivate the claim that we should not use ML tools in certain contexts[7]. By claiming that we do not want to frame the problem as a predictive problem where the past works as a basis to infer the future when human agency is involved, we immediately exclude the use of ML tools, which employ an inductive inferential structure. By focusing only on CIIs, we do not have this opportunity: we run the risk of giving the impression that it is possible, at least in principle, to come up with a combination of CIIs that will make the ML tool shape human agency in an uncontroversial ethical way. Another interesting consequence is that UoN, in a context like COMPAS, will necessarily frame the problem to solve as a predictive problem of crime-control, rather than as a problem of what a defendant might deserve on the basis of culpability and other factors (including defendant's willingness to pursue a different, good life). This has been seen as a 'domain distortion' by Pruss[8] (2021), as well as a value-promoting[9] aspect of ML tools (Ratti and Russo 2024).

4. CONCLUSION

In this chapter, I have shown how analyzing the inferential structure of ML can be useful to distinguish different classes of ethical problems. ML tools (at least supervised tools) follow an inductive strategy based on UoN and various CIIs. The literature on ethics of ML has focused especially on CIIs. Here I have shown how UoN raises issues about human agency – in fact, ML tools used in the context where human agency is involved are based on the assumption that the future must be similar to the past, thereby undermining the importance of human agency and open future. I have ended with some consequences concerning this analysis of UoN, especially in the kind of ethical arguments that focusing on UoN vs focusing on CIIs can

[7] As noted by a Reviewer, legal constraints may deliver this, even though it would be strictly a legal reason and not an ethical motivation

[8] By 'domain distortion' Pruss (2021) means how ML can influence the concepts and assumptions of a given context of application

[9] In a co-authored work with Federica Russo, we refer to 'value-promoting' science to the phenomenon when either scientific methodologies or concepts promote certain values at the expense of others because of their own identifiable characteristics, independently of the value endorsed by those using the methods or the concepts

provide.

REFERENCES

Antony, L. (2016). Bias: Friend or Foe? In Brownstein, M. and Saul, J., editors, Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology, pages 157–190. Oxford University Press.

Bechtel, W., & Richardson, R. (2010). *Discovering Complexity - Decomposition and Localization as Strategies in Scientific Research*. The MIT Press.

Biddle, J. B. (2020). On Predicting Recidivism: Epistemic Risk, Tradeoffs, and Values in Machine Learning. *Canadian Journal of Philosophy*, 1–21. https://doi.org/10.1017/can.2020.27

Brown, M. (2020). *Science and Moral Imagination*. University of Pittsburgh Press.

Burkov, A. (2019). The Hundred-Page Machine Learning Book.

Canali, S., & Ratti, E. (2024). Between quantity and quality: competing views on the role of Big Data for causal inference. In P. Illary & F. Russo (Eds.), *The Routledge Handbook of Causation and Causal Methods*. Routledge.

Capriotti, E., & Altman, R. B. (2011). A new disease-specific machine learning approach for the prediction of cancer-causing missense variants. *Genomics*, *98*(4), 310–317. https://doi.org/10.1016/j.ygeno.2011.06.010

Castro, C. 2022. Just Machine, *Public Affairs Quarterly*, 36(2)

Consens, M. E., Dufault, C., Wainberg, M., Forster, D., Karimzadeh, M., Goodarzi, H., Theis, F. J., Moses, A., & Wang, B. (2023). To Transformers and Beyond: Large Language Models for the Genome. http://arxiv.org/abs/2311.07621

Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. *IJCAI International Joint Conference on Artificial Intelligence*, *Ijcai*, 4691–4697. https://doi.org/10.24963/ijcai.2017/654

Douglas, H. (2009). Science, Policy, and the Value-Free Ideal. University of Pittsburgh Press.

Down, T. A., Hubbard, T. J. P., (2002). Computational Detection and Location of Transcription Start Sites in Mammalian Genomic DNA Computational Detection and Location of Transcription Start Sites in Mammalian Genomic DNA. *Genome Research* 458–461. https://doi.org/10.1101/gr.216102

Earman, J. (1992). Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory. The MIT Press.

Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Deep learning: new computational modelling techniques for genomics. In *Nature Reviews Genetics* (Vol. 20, Issue 7, pp. 389–403). Nature Publishing Group. https://doi.org/10.1038/s41576-019-0122-6

Facchini, A, & Termine, A. (2021). Towards a taxonomy for the opacity of AI systems, in *Conference on Philosophy and Theory of Artificial* Intelligence (pp 73-89). Springer.

Fazelpour, S., & Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, *16*(8). https://doi.org/10.1111/phc3.12760

Ferdman, A., & Ratti, E. (2024). What Do We Teach to Engineering Students: Embedded Ethics, Morality, and Politics. *Science and Engineering Ethics*, 30(1), 7. https://doi.org/10.1007/s11948-024-00469-1

Forber, P. (2011). Reconceiving Eliminative Inference. *Philosophy of Science*, *78*(2), 185–208.

Harman, G., & Kulkarni, S. (2007). *Reliable Reasoning - Induction and Statistical Learning Theory*. The MIT Press.

Hawthorne, J. (1993). Bayesian induction is eliminative induction. *Philosophical Topics*, *21*(1).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical

Learning. Springer. https://doi.org/10.1016/j.peva.2007.06.006

Jiang, H. (2021). Machine Learning Fundamentals - A Concise Introduction. Cambridge University Press.

Johnson, G. M. (2023). Are Algorithms Value-Free? Feminist Theoretical Virtues in Machine Learning. *Journal of Moral Philosophy*.

Kitcher, P. (1993). *The Advancement of Science*. Oxford University Press.

Kuhn, T. (1977). Objectivity, Value Judgement and Theory Choice. In *The Essential Tension: Selected Studies in the Scientific Tradition and Change* (pp. 356–367). University of Chicago Press.

Larsen, P. (2018). When Human Behavior Enters the Equation. Critical Review, 30(3–4), 316–324. https://doi.org/10.1080/08913811.2018.1565729

Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, *16*(6), 321–332. https://doi.org/10.1038/nrg3920

Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.

McMullin, E. (1983). Values in Science. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *2*, 686–709.

Norton, J. (1995). Eliminative induction as a method of discovery: How Einstein discovered general relativity. In J. Leplin (Ed.), The Creation of Ideas in Physics. Kluwer Academic Publishers.

Norton, J. (2005). A little survey of induction. In P. Achinstein (Ed.), *Scientific Evidence - Philosophy Theories and Applications*. The Johns Hopkins University Press.

Pietsch, W. (2021). *Big Data*. Cambridge University Press. https://doi.org/10.1017/9781108588676

Pietsch, W. (2022). On the Epistemology of Data Science - Conceptual Tools for a New Inductivism. Springer. https://link.springer.com/bookseries/6459

Platt, J. R. (1964). Strong Inference. *Science*, *146*(3642).

Pruss, D. (2021). Mechanical Jurisprudence and Domain Distortion: How Predictive Algorithms Warp the Law. *Philosophy of Science*, *88*(5), 1101–1112.

Radder, H. (2019). *From Commodification to the Common Good: Reconstructing Science, Technology, and Society*. The University of Pittsburgh Press.

Ratti, E. (2015). Big Data Biology : Between Eliminative Inferences and Exploratory Experiments. *Philosophy of Science*, *82*(2), 198–218.

Ratti, E. (2020). Phronesis and Automated Science: The Case of Machine Learning and Biology. In M. Bertolaso & F. Sterpetti (Eds.), *A Critical Reflection on Automated Science - Will Science Remain Human?* Springer.

Ratti, E. (2020). What kind of novelties can machine learning possibly generate? The case of genomics. *Studies in History and Philosophy of Science Part A*, *83*, 86–96. https://doi.org/10.1016/j.shpsa.2020.04.001

Ratti, E., & Stapleford, T. A. (Eds.). (2021). *Science, technology, and virtues: Contemporary perspectives*. k: Oxford University Press.

Ratti, E., & Graves, M. (2021). Cultivating Moral Attention: a Virtue-Oriented Approach to Responsible Data Science in Healthcare. *Philosophy and Technology*, *34*(4), 1819–1846. https://doi.org/10.1007/s13347-021-00490-3

Ratti, E., & Graves, M. (2022). Explainable machine learning practices: opening another black box for reliable medical AI. *AI and Ethics*. https://doi.org/10.1007/s43681-022-00141-z

Ratti, E., & Russo, F. (2024). Science and values: a two-way direction. *European Journal for Philosophy of Science*, 14(1). https://doi.org/10.1007/s13194-024-00567-8

Russo, F. (2009). *Causality and Causal Modelling in the Social Sciences. Measuring Variations*, New York: Springer.

Schmidt, E., Sesing-Wagenpfeil, A., & Köhl, M. A. (2023). Bare statistical evidence and the legitimacy of software-based judicial decisions. *Synthese*, 201(4). https://doi.org/10.1007/s11229-023-04141-2

Solomon, M. (2012). The web of valief: an assessment of feminist radical empiricism. In S. Crasnow & A. Superson (Eds.), *Out from the Shadows: Analytical Feminist Contributions to Traditional Philosophy* (pp. 435–450). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199855469.001.0001

Vallor, S. (2016). Technology and the Virtues - A Philosophical Guide to a Future Worth Wanting. Oxford University Press.

Ward, Z. B. (2021). On value-laden science. Studies in History and Philosophy of Science, 85, 54–62. https://doi.org/10.1016/j.shpsa.2020.09.006

Wheeler, G. (2016). Machine epistemology and big data. In L. McIntyre & A. Rosenberg, eds., *The Routledge Companion to Philosophy of Social Science*. London: Routledge.

Emanuele Ratti Lecturer in the Department of Philosophy University of Bristol, UK. His area of research and teaching is Ethics and Philosophy of Science and Technology (with a focus on the life sciences and data science). He is particularly interested in the aspects of the natural sciences and data science that stand at the intersection of ethical and epistemic questions.