The Incompleteness of Knowledge: An Epistemological Study of Machine Learning.

Devin Barui. Student no: s4640456. 8576 words.

A thesis presented for the degree of Bachelor of Politics, Philosophy and Economics (Honours).

I. Introduction

In the midst of the 'data revolution', *data science* and *machine learning* have become increasingly eminent disciplines. A Google Trends search using keywords 'data science' and 'machine learning' shows that in only five years, interest over time has almost doubled. Additionally, more and more search results show that this phenomenon is taking root in the business world with terms such as the 'data-driven' in 'data-driven decision making' becoming part of the common discourse. Quoting a consumer article on the topic: "One of the greatest advantages to come with data-driven decision-making is the ability to make decisions more confidently than ever before" (Grant, 2023).

However, this development raises as many concerns as it does opportunities. Where rationalists and empiricists have long debated the prospect of acquiring objective knowledge via reason or through the senses, other schools have questioned our ability to obtain this knowledge at all (i.e. Structuralism, Linguistics, etc). Therefore, this idea that machine learning and big data will spell "The end of theory" or the obsoletion of the scientific method (Anderson, 2008), raises questions as to what has changed.

There is a growing belief that if we take an infinitely large dataset, pass it through a perfectly optimized algorithm, and run it on an infinitely fast computer with no computational limits, we would be able to peer into the fabric of reality and supersede the problems of attaining objective knowledge. In other words, the only obstacle towards the right answer in most or at least the majority of situations is a technological problem rather than a philosophical one. However, based on current knowledge, do we have enough information to determine whether this claim is true?

This paper will investigate this claim by examining machine learning through a framework of philosophical relevance. In this, it argues that the knowledge acquired through machine learning results from a methodological separation of different types of knowledge into similar and dissimilar regions. This understanding is crucial for interpreting current challenges in machine learning development, including algorithmic bias, a phenomenon where machine learning algorithms yield results that unjustly favor or discriminate against particular groups or situations (Kordzadeh, 2022). Additionally, it addresses the black-box problem, stemming from our limited understanding of the inner workings of intricate deep learning algorithms (Rudin, 2019). Through this study of these issues, the necessity of establishing a connection between knowledge derived from machine learning and the nature of such knowledge is emphasized.

Related Works

In relation to this topic, numerous prior works have taken similar approaches to philosophically articulate different aspects of statistical learning theory.

Vallverdu (2015) and Vallverdu (2020) similarly categorize the historical development of statistical theory from frequentism to Bayesianism and discuss how these paradigms relate to our understanding of causal inference in light of deep learning algorithms. In his 2020 paper, he responds to the criticisms of eminent statistician Judea Pearl (2018) regarding deep learning where Pearl argues that deep learning models cannot create causal knowledge due to its model-blindness and lack of human inferential capabilities such as managing counterfactuals. Vallverdu investigates this criticism through reference to the ontological relation of machine learning models to human knowledge and how these connections facilitate the ability to create causal connections. The arguments of this paper work alongside Vallverdu (2020) but present a stronger articulation of the logical paradoxes of machine learning as they relate to inductive inference.

Anirudh and Bengio (2022) published a highly-influential paper which conducted a qualitative analysis on to what extent human and animal intelligence could be reduced to a set of inductive principles. The broad idea behind this research is that through understanding these implicit biases from a neurocognitive perspective, we would better be able to better understand optimal architectures for different AI problems. From a philosophical perspective this can be denoted as a wide-scale empirical investigation of summative uniformity principles for inductive inference. The authors conclude with mixed results where certain inductive biases were capable of simplifying and generalizing to a large set of problems whereas the implementation of other inductive biases was not as well understood. This information is useful in the context of this paper and will incorporate this knowledge as an implicit part of its argument.

Along these lines, Desalles (2019) takes a more philosophical approach to investigating the inductive inferences inherent in human neurocognition to improve machine learning models. Notably, he describes two concepts: complexity drop and contrast. As will be shown these elements will arise as a natural course of the research question for this paper and represents a point of agreement for this paper.

Research Method

Considering that machine learning is a dense field which requires a herculean amount of dedicated study to understand in its entirety, to keep this paper in line with the intended audience (those interested in the philosophy who have a basic understanding of statistics), this paper will incorporate as much technical detail as necessary without listing out every feature of the formalized proofs for each relevant concept.

Additionally, this paper will draw from a wide range of textbooks (Geron, 2022; Witten and James, 2015; Emmhert-Streib et al., 2023) and professional handbooks (Ta Liu, 2022; Fernandes De Mello and Antonelli Ponti, 2018) in order to sufficiently establish the basic concepts relevant to this paper. Beyond all of these textbooks being reputed and written by respected authors, the ethos for this decision is derived from the amount of technical experience and careful consideration required to write a textbook which simplifies but doesn't misrepresent information into a publishable standard. After adapting the basic forms and terminology from the textbooks on machine learning, these will be then compared to the handbooks used in statistical learning theory intended for professional use to provide a sufficient amount of rigor to each concept.

The research methodology for this paper will involve:

1) Creating analytical objects which summarize a set of machine learning concepts $z \in Z$ with a representation of these concepts T which are philosophically interrogable and provide a chain of explainability between these objects and their root concepts. As will be demonstrated, even where this methodology is applied, the end results brings us to the forefront of issues relating machine learning with philosophy in a way which allows us a loose vocabulary for discussing them.
2) Using these analytical objects will then be employed in the discussion section to provide commentary on issues such as the black-box problem and algorithmic bias

A notable weakness of this research methodology comes from the lack of nuance or the information loss between z_i and its corresponding component in representation T. While this paper has employed a means-end justification where the correspondence of the results with other papers is used to assess the success of representation T, this still represents a significant methodological drawback which limits the scope of this research method to generate new information. Ironically, a full description of why this is a weakness in the context of this paper will be presented throughout this paper.

Nevertheless, this function of synthesizing a heuristic for philosophically discussing machine learning seems to be a worthwhile contribution and can assist in interdisciplinary translation as it pertains to key issues.

II. Results

This section of the paper presents the findings of the research method. It is divided into three parts: 1) A summary of the foundations of statistical theory through analyzing the age-old debate between frequentism and bayesianism and deriving each their respective epistemological assumptions about the nature of statistics; 2) A philosophical synthesis of modern statistical learning theory as it was transformed by Wolpert and Macready (1997) through the no-free-lunch theorems; 3) A direct comparison between this synthesis of statistical learning theory with the philosophical problem of induction.

Foundational Epistemology of Statistical Theory.

Although statistics is a dominating force which is deeply entrenched in our methods for understanding the world, the relationship between statistics and reality is not implicitly obvious.

A relevant concept for this idea is the notion of platonism in mathematics. Despite ongoing investigation and debates which have spanned centuries, presently, there exists no widely-accepted formalism which bridges a fundamental knowledge implicit in reality consistently across all of the most basic mathematical operations. This struggle has led to the default status of platonism as the least presumptive school of philosophical thought in relation to how mathematics imprints on reality. In summary, while many people incorrectly assume that math and reality is directly associated, the knowledge that we obtain from mathematical reasoning exists more strongly in a conceptual, ideal world than it does in the material world. However, this former world could be seen as existing in a more 'truer' state than the latter (Horsten, 2007).

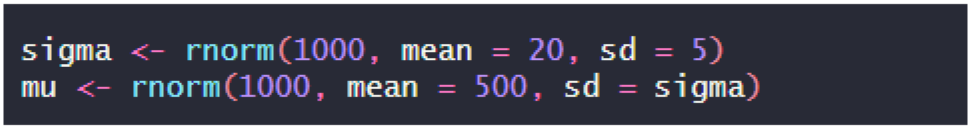
Accordingly, the idealist description of statistics can be found through measure theory. The measure-theoretic formalism of probability denotes a population set, Ω; the set of events as a σ-algebra, *E*; and a probability measure μ. Combining these elements, they depict a powerful formalized system which partitions an object Ω through *E* based on a normalized probability measure μ through a set-theoretic approach. In practice, statisticians attempt to discover an unknown population Ω, based on inferences on the observations *E* (Roussas, 2014). However, even through this system, we are still not informed on the epistemic status of these objects and how they relate to events in reality.

The most fundamental problem in statistics is to provide a methodological solution to mitigate the problem of induction. After all, just because something was observed in the past does not mean that it will happen in the future. This struggle to overcome this problem can be seen in the age-old debate between frequentism and bayesianism.

Under frequentism, the probability measure μ is defined as the limiting value of the frequency of an event when performed on a large number of observations. Through this definition, we gain a perception of the world which connects probability to reality as a collection of the physical manifestations of stochastic processes (natural phenomena which generate events based on reproducible mechanisms). This is the most common perception of statistics when applied to scientific investigation, as stochastic processes are presumed to exhibit law-like properties in the physical sciences (Vallverdú, 2015). Simply, flipping a coin with an equal tendency to either result and conceptualizing it as a fifty percent probability is the easiest way to envision this idea .

However, there is a certain element which is missing in this interpretation. Bayesianism provides an alternate view of μ and how it relates to observations in reality. Alongside statements of certain probability, Bayesianism also embeds a degree of belief in those statements as a way to systematize the subjective nature of observing *E* from Ω. This is not only a pragmatic considering that Ω cannot be directly observed, but also more philosophically agnostic as it does not prescribe a metaphysical status to Ω in the same way that frequentism does. Instead, Bayesianism asks the question, "given a prior belief X, what degree of belief would a rational agent assign to Y given X" (Vallverdú, 2015). Referring to the coin toss example, if an observer began with a belief that the chance towards either result was equal, the Bayesian method would output a posterior probability denoting a probability assessing whether the chance was actually fifty percent.

To further articulate this, suppose we have a set of measurements $s_i = (\mu_i, \sigma_i)$ which denote different sample means and standard deviations and attempt we are trying to describe the population mean μ.



```
sigma <- rnorm(1000, mean = 20, sd = 5)
mu <- rnorm(1000, mean = 500, sd = sigma)
```

Fig 1. Randomly generated samples with predesignated mean and random standard deviations in R.

<u>Frequentist approach</u>

$$P(s_i \mid \mu) = (2\pi\sigma_i^2)^{-1/2} exp(\frac{-(\mu_i - \mu)^2}{2\sigma_i^2})$$ → *Normal distribution for each i.*

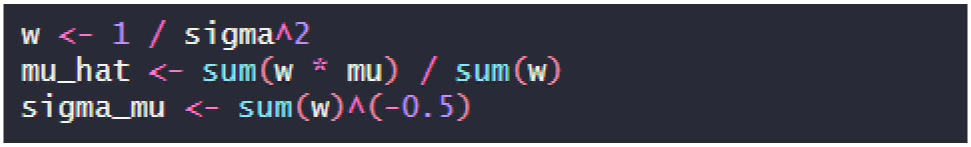
$$L(s \mid \mu) = \prod_{i=1}^{N} P(s_i \mid \mu)$$ → *Maximum likelihood estimator.*

Using the log-likelihood we can find a weighted point estimate for μ which assigns more weight for observations with less variance (sum of precision weights).

$$w = 1/\sigma^2$$

$$\hat{\mu} = \sum w\mu / \sum \mu$$ → *Point estimate.*

$$\sigma_\mu = (\sum w)^{-1/2}$$ → *Standard error.*



```
w <- 1 / sigma^2
mu_hat <- sum(w * mu) / sum(w)
sigma_mu <- sum(w)^(-0.5)
```

Fig 2. Calculating point estimate for $\mu \pm \sigma_\mu$.

$$\hat{\mu} = 500 \pm 1$$

<u>Bayesian approach</u>

$$P(\mu \mid s) \ = \ \frac{P(s|\mu)\, P(\mu)}{P(s)}$$

Bayes' theorem as presented above can be applied to this very same problem. Let's examine each of the terms in the expression:

- $P(\mu)$ represents the prior probability; our degree of belief in μ prior to the analysis.
- $P(s \mid \mu)$ represents the likelihood - this is near identical to the frequentist likelihood $L(s \mid \mu)$.
- $P(\mu \mid s)$ represents the posterior probability; our renewed degree of belief in light of new information (note that this does not appear in the frequentist approach).
- $P(s)$ represents the model evidence; a normalizing parameter which equals one in this instance since there's only one model under consideration.

Although this interpretation is vastly different from before, if we set the prior probability to be constant over μ (i.e. $P(\mu) = 1$), then:

$$P(\mu \mid s) \ = \ P(s \mid \mu)$$

This makes the Bayesian approach proportionally equivalent and maximize at the same value as the frequentist approach. This is due to the flat prior chosen for the problem which is uninformative and does not cause any deviations than what is expected in the frequentist approach. While useful, in practice it is not always feasible to introduce a flat prior into the model which will necessarily cause each approach to differ. In these situations, it could be argued that even though an informative prior assumption is intrinsically necessary for Bayesianism, frequentism naively avoids the problem and also makes implicit assumptions which are not acknowledged (VanderPlas, 2014).

In summary, frequentism and bayesianism ask two different questions. Frequentism asks a physical question: 'given a large series of events, how many of those events represent the fixed parameter x?'. Whereas Bayesianism asks an epistemic question: 'Given a prior probability, what degree of confidence ought to be had in x given a set of new information y'.

While the distinction between physical probability and epistemic probability is firmly established in the traditional understanding of statistical philosophy, an alternative perspective on the probabilities inherent in statistical methods is viewing them as expressions of epistemic attitudes. These probabilities can be seen as doxastic, representing the opinions of an idealized rational agent about data and hypotheses. Alternatively, they can be decision-theoretic, involving a more elaborate representation of the agent, which determines their inclinations towards decisions and actions regarding the data and hypotheses. Finally, they can be understood as logical, serving as a formal representation that establishes a normative ideal for uncertain reasoning, without necessarily suggesting that the numerical values correspond to anything of psychological significance (Romeijn, 2014).

In essence, statistics is a deeply philosophical endeavor which often gets mistaken for a purely numerical one. Not only is there the freedom to interpret information through a variety of different perspectives, but also the ability to question our ability to gain knowledge at all.

.

Fundamentals of Statistical Learning Theory.

Furthering this discussion, as the relationship between classifiers and their outputs has become increasingly intricate, the need to justify the outputs of learning algorithms within their capacity to consistently approach a learnable subspace has gained prominence.

Formal learning theory, residing at the intersection of mathematics and philosophy, addresses the question of how "an agent should use observations about their environment to arrive at correct and informative solutions" (Hansson, 2007). Accompanying this field is computational and statistical learning theory which are concerned with understanding the learning problem from the perspectives of computability and statistics, respectively (Fernandes de Mello, 2018). Since our research question assumes infinite computational power, computational learning theory is of less significance in this context.

This section will cover fundamental concepts which illustrate how machine learning models characterize their own inductive bias and how this bias necessarily arises in order to create a learning model which outperforms random guessing.

Classifying Joint-Probability distributions

Starting with two variables: an input variable X and an output variable Y, while each of these variables bear their own distributions, for solving learning problems we are interested in finding a joint probability distribution $P(X \times Y)$ which identifies the combined behavior for both of these variables. However, since we cannot directly observe $P(X \times Y)$, we must find a function f such that:

$$P(X \times Y) \simeq (X, f(X))\ ; \quad f\colon X \to Y$$

Here, we are attempting to algorithmically find a mapping function $\hat{f}_i \subset F$, where f_i represents all possible learning outcomes, which most closely approximates $P(X \times Y)$ and informs us about the joint relationship between the two variables. However, no assumption is made about the joint probability distribution, allowing it to take any form (Witten, 2013; Fernandes de Mello, 2018).

Risk and Generalization Error

Assuming we can compute all f_i in F (this is actually fairly straightforward in binary classification problems), in order to determine which f_i perform well for the problem, we need a metric which characterizes the loss of each learner function, or the amount that a learner guesses correctly or incorrectly. This is called a loss function and can take on various forms depending on what we are trying to find.

Where $l(.)$ denotes the specified loss function, we can quantify a risk which identifies the integral of divergences from the expected outcome $P(X \times Y)$ and $f(x)$:

$$R(f) \; = \; \int_{X \times Y} ||l(P(X \times Y) \; - \; (X, f(X)))||dX \times Y$$

Accordingly,

$$f_{best} = argmin_{f_i} R(f_i) \; \forall i$$

However, since $R(f)$ considers all possible data examples (Including those outside of the training sample), we cannot observe this parameter in practice except in simple examples where the data has a finite set of outcomes (such as a dice roll or a coin toss). Therefore, we must introduce an estimator for $R(f)$, the empirical risk - typically calculated as the sample average of the loss function (Fernandes de Mello, 2018; Witten, 2013).

$$R_{emp}(f) \; = \frac{1}{n} \sum_{i=1}^{n} l(x_i, y_i, f(x_i))$$

Hence,

$$\hat{f}_{best} = argmin_{f_i} R_{emp}(f_i) \; \forall i$$

These two concepts introduce a fundamental concept which will be relevant for the rest of this discussion, *generalization*. The generalization error arises as a way to express the convergence of an empirical risk to its expected risk. It can be defined as:

$$G \; = \; |R_{emp}(f) \; - \; R(f)|$$

A learning algorithm is said to generalize well when the difference between the expected risk and empirical risk is small. Typically, models which generalize well are said to discern well between the noise in a sample and the general 'gist' that we want to capture in the learning problem (Fernandes de Mello, 2018). However, considering that we cannot observe the generalization error directly in most cases, we must construct an upper bound ϵ on $R_{emp}(f)$ such that it eventually converges to $R(f)$across large datasets. To demonstrate this upper bound, Vapnik-Chervonenkis (VC) dimensions and uniform consistency principles are applied to show how the complexity of the hypothesis space transforms relative to the law of large numbers (Ta Liu, 2020). Since this justification is unnecessarily complex for the purposes of this paper, this bound can simply be shown as:

$$P(|R_{emp}(f)\ -\ R(f)|\ >\ \epsilon)\ \leq\ 1\ -\ \delta$$

Therefore, even without observing $R(f)$, we can choose values ϵ and δ such that it is bounded within this generalization error based on our intuition of what is discernible by f (Ta Liu, 2020).

No-Free-Lunch Theorems and learnability limits

Based on everything so far, we currently have X, Y and our mapping function F as well as the means to create an algorithm to find $f_i \subset F$ whose empirical risk generalizes to the true risk. At this point, we introduce the no-free-lunch theorems - one of the most fundamental insights of statistical learning theory which relates back to the problem of induction.

The no-free-lunch (NFL) theorems (Wolpert and Macready, 1997) comprise a collection of impossibility theorems that broadly assert the following: when considering all possible function mappings f for a given problem, no learning algorithm can outperform any other algorithm consistently. Put simply, this is because for every algorithm that correctly predicts a specific set of data points, there exists another set of data points that it predicts incorrectly. Averaged over the entire function space, the generalization error $R(f)$ for each learning algorithm converges to 0.5, indicating performance no better than random guessing (Fernandes de Mello, 2018).

However, this is somewhat contradictory, as our basic analysis showed that minimizing the empirical risk unilaterally improves a predictive outcome. To reconcile these two ideas, consider the following (Fernandes de Mello, 2018):

> Let M be a learning algorithm for a binary classification task whose loss function produces a 0 or 1 over a domain X. With sample size n denoting any number smaller than $X/2$ (implying that the training data is limited and may not fully represent the underlying distribution), a sample S and a distribution P such that for $P(|R_{emp}(f)\ -\ R(f)|\ >\ \epsilon)\ \leq\ 1\ -\ \delta$ there exists a concept $f\ \subset\ F$ in $P(X\ \times\ Y)$ which we want to learn, however:
>
> $$P(|R_{emp}(f)\ -\ R(f)|\ >\ \epsilon)\ \leq\ 1\ -\ \delta$$

$$P(|R_{emp}(f) - R(f)| > \epsilon) = 1/8 \text{ and } 1 - \delta = 1/7$$
$$1/8 \leq 1/7$$

In other words, while there exists a learnable concept in our joint probability distribution, even the most optimal algorithm we could choose is not confined within the upper bound for the generalization error. Therefore, even our best empirical risk minimization algorithm will fail to generalize due to the complexity of the hypothesis space outpacing the law of large numbers (Emmert-Streib, 2023).

What these theorems articulate is that there is no universally best algorithm that performs well without prior knowledge of the learning problem. In effect, this means that to reliably produce a model that generalizes to our desired learning concept f, we need to introduce a degree of bias that simplifies the hypothesis space. This can translate to anything from changing the function space F, changing the probability distribution P, changing the input space X, or changing the loss function L (Emmert-Streib, 2023).

When adjusting these elements, the objective is to modify the learning problem in a way that facilitates the *generalization* of the underlying concept represented by $R(f)$ or $R_{emp}(f)$. This problem typically involves navigating the trade-off between bias and variance, commonly known as the bias-variance trade-off, which deals with managing the model's complexity and its ability to capture underlying patterns effectively(Witten, 2013).

Underfitting occurs when the model inadequately capture the complexity of the data, resulting in a high generalization error, $R(f)$. This situation arises when the model's representation of the underlying concept is over simplistic and fails to grasp the nuance of the distribution $P(X \times Y)$. On the other hand, overfitting occurs when the model excessively fits the noise in the training data, leading to a large disparity between the empirical risk, $R_{emp}(f)$, and the true risk, $R(f)$. This phenomenon signifies that the model is overly complex and struggles to generalize well beyond the training data, impairing its predictive performance on unseen data points (Géron, 2022).

Learnability and Inductive Bias

Collectively, this summary of statistical learning theory provides us with a robust vocabulary for discussing learning problems. We can establish a learning problem as the measurable object $P(X, Y)$ as well as a predicate object $R(f)$ which captures specific aspects of the relationship between F, X, and P which we want to use as a general rule for solving the problem. Additionally, the object $R_{emp}(f)$ quantifies our ability to observe our predicate object $R(f)$ through its feasibility when applied to all possible dimensions of $P(X, Y)$.

When viewed through this angle, the connection between statistical learning problems and philosophy becomes a lot more obvious. Instead of blindly searching for what predicates $R(f)$ work and don't work for a particular learning problem, this scope to introduce bias allows us, as philosophers, to inquire into what dimensions of the original learning problem are relevant and which ones aren't. As an illustration, we can draw comparisons to the story of Plato and his assembly of philosophers in order to discern the 'true essence of man.' According to the tale, after concluding that the essence of man was that of a 'featherless biped,' Diogenes bursts into the room with a plucked chicken in hand, quipping, 'Behold, man!' (Fantuzzi, 2021). This historical anecdote portrays Plato's pursuit of a concise conceptualization $R(f)$ in an effort to elucidate the essence of 'man' within the broader context of existence. The emphasis here does not lie in Plato's definitive discovery of the essential characteristic defining 'man', but rather in his pursuit of a concept that is both comprehensive and yet empirically determinable, representing a nuanced balance between philosophical abstraction and measurability.

Additionally, where we are tasked with a rational solution to the problem of induction, the convergence to a predicate object $R(f)$ not only facilitates a response but also shapes its own inductive bias. This is in contrast to the dynamic between frequentism and bayesianism. While frequentism traditionally relies on the premise that probabilistic measures represent a reflection of physical frequencies, the implications of the NFL theorems reveal inherent biases within this foundational assumption. On the other hand, although Bayesianism acknowledges the inherent limitations in acquiring statistical knowledge without prior assumptions, it may lack the ability to explicitly articulate the biases inherent in the predicate object $R(f)$.

However, it is crucial to acknowledge that the presence of inductive bias in a statistical model does not automatically imply presumption. For instance, theory and research in the natural sciences often employ frequentist statistical approaches without presuming a predetermined outcome. While the term 'inductive bias' carries certain implications, it primarily denotes the incompleteness between the information required by a model and the mathematical structures employed for a certain problem (Schurz, 2017). This underlies the idea that the connection between mathematical concepts and reality is not inherently self-evident. As a result, encoding an inductive bias into a machine learning model can potentially render it more agnostic than biased, contingent upon the philosophical validation of which inductive assumptions hold greater agnosticism than others.

Rearticulating Inductive Bias.

Since the formalization of the no-free-lunch theorems, their influence has extended beyond the realm of computer science and permeates the domain of epistemological philosophy. Specifically, Schurz (2017) contends that these theorems present a 'radicalized version of Hume's induction skepticism,' while Lauc (2020) draws intriguing parallels between these theorems and Goodman's new riddle of induction. These philosophical challenges emphasize the constraints of relying on inductive reasoning for the establishment of knowledge, yet upon closer examination, they do not singularly impede the exploration of how machine learning contributes to our comprehension of the world.

The problem of induction is most famously associated with David Hume's critique of the scientific method. Hume's argument centers around the idea that any attempt to justify inductive inferences, particularly within the natural sciences, inevitably leads to circular reasoning. Starting with the premise that 'nature is uniform,' a fundamental assumption crucial for scientific investigation, the logic might follow that we should establish the validity of the scientific method based on this principle. However, as Hume points out, the only support for this premise comes from empirical and inductive evidence. Given the absence of a conclusive means to prove this postulate, Hume concludes that there is no solid logical foundation for placing unwavering trust in the scientific method or in any form of inductive reasoning (Sterkenburg, 2021).

In the context of machine learning, this understanding conveys the recognition that no purely data-driven or deductive approach exists to derive a flawless solution. However, as Sterkenburg (2021) points out, this perspective disregards a critical intuition that seemingly contradicts certain aspects of machine learning. Despite the lack of a universally deducible method for optimal data analysis, several general processes and methods intuitively demonstrate their effectiveness without controversy. For instance, although the NFL theorems stress the absence of a singular optimal learning algorithm across all scenarios, empirical risk minimization consistently outperforms empirical risk maximization (Sterkenburg, 2021). Similarly, despite the lack of a formal proof validating the scientific method, its effectiveness remains uncontested among rationalists and scientists in solving scientific problems compared to its non-application. In fact, what appears more controversial is the notion that the essence of inductive reasoning can be reduced to a universal uniformity principle.

Accordingly, several scholars have put forth compelling arguments countering the assertion that inductive reasoning can be reduced to a singular uniformity principle or precisely defined (Putnam 1981; van Fraassen 1989, 2000). Sober (1991) particularly highlights that the proposition of a universal uniformity principle for all inductive inferences results in a quantifier-shift fallacy. He suggests that instead of every inductive inference relying on a single assumption, each one necessitates a distinct and localized empirical assumption. In essence, each induction relies on specific and context-bound empirical factors, challenging the notion of a universal uniformity principle. However, even if we manage to circumvent Hume's argument about the circularity of any non-deductive justification for induction, we find ourselves confronted with an infinite regress, where each empirical assumption can only find validation through another induction process, each laden with its own set of empirical assumptions (Sterkenburg, 2021).

Despite this problem, this reformulation aligns with the scientific intuition that inductive inferences give insight into systemic views of certain processes and how they relate to the world. Additionally, through the use of a predicate object $R(f)$, we can articulate this bias in a way that specifies its relevance to certain problems. This approach offers a clearer understanding of why generalized methods, such as empirical risk minimization, consistently outperform other methods, like empirical risk maximization. In this sense, the challenge lies in finding the appropriate predicate objects $R(f)$ that facilitate law-like generalizations of certain processes (Lauc, 2020).

This introduces another 'problem of induction' which concerns how we can determine which predicates are able to describe the world in such a 'law-like' general manner. Goodman's new riddle of induction illustrates this problem:

> Suppose that all emeralds examined before a certain time t are green .... Our evidence statements assert that emerald a is green, that emerald b is green, and so on....Now let us introduce another predicate less familiar than "green". It is the predicate "grue" and it applies to all things examined before t just in case they are green but to other things just in case they are blue. Then at time t we have, for each evidence statement asserting that a given emerald is green, a parallel evidence statement asserting that emerald is grue. The question is whether we should conjecture that all emeralds are green rather than that all emeralds are grue when we obtain a sample of green emeralds examined before time t, and if so, why? (Goodman 1983).

Through this thought experiment, Goodman underscores the difficulty in determining which predicates can be relied upon as lawlike. Despite our natural intuition about emerald t being green, given that all emeralds before t were green, the thought experiment reveals the lack of a logical foundation favoring green over grue. Goodman argues that our familiarity with concepts like green and blue does not preclude the possibility of another syntactic organization using terms like grue and bleen. In this context, relying on an inherent quality of the predicate green or its syntactic ordering to determine which predicates can be generalized becomes an additional reliance on inductive or empirical elements that cannot be logically proven (Goodman, 1983).

While these assertions offer valuable insights into the nature of inductive inference, they do not unveil anything substantially different beyond what the NFL theorems already establish - the absence of a logically deducible basis for favoring one predicate over another without reliance on some form of inductive bias. However, as discussed earlier, this realization alone does not invalidate the ongoing discussion about the role of machine learning in enhancing our understanding of the world. Just as our dependence on inductive and empirical claims does not completely undermine our confidence in the scientific method's efficacy in expanding our knowledge, our reliance on these claims should not impede our ability to gain insights through machine learning.

In practice, the choice of predicate objects often rests on intuitive grounds, which may not always conform to strictly valid premises. Consequently, although this pragmatic understanding may limit the application of confirmation theory in formalizing machine learning, it fosters the exploration of alternative philosophical perspectives to address the inherent challenges and opportunities in machine learning problems (Hansson, 2007).

III. Discussion

This section will use the established representation of statistical learning theory through philosophical relevance T, to study current problems plaguing the development of machine learning: 1) Will explore the black-box problem and argue why the paradigm shift towards deep learning algorithms may qualitatively affect our understanding of statistical learning theory; 2) will then branch towards algorithmic bias and use the establish knowledge of this paper to argue a pragmatic philosophy of machine learning which divides metaphysical knowledge into similar and dissimilar spaces.

The Paradigm Shift of the Black-Box Problem

Following the successes of deep neural networks, this success has not been matched by the theoretical progress to explain their behaviour. The black-box problem in machine learning emerges as one of the most pressing challenges in the field today. Put simply, while certain algorithms, such as linear regression, offer higher interpretability and transparency, others are near completely uninterpretable and, therefore, opaque. In straight-forward, simple learning problems, researchers often have the flexibility to choose between transparent and opaque models, usually opting for simpler and more interpretable ones. However, in more complex problems with intricate datasets, a certain degree of opaqueness becomes a necessity in order to derive a well-performing predicate object $R(f)$ (Voulodimos, 2018). While the black-box problem exists with a large variety of machine learning algorithms, this problem is most prevalent in large-scale, multi-layered neural networks when attempting to solve complex problems.

A significant aspect of this challenge stems from the elusive nature of understanding why complex neural networks can achieve the level of generalization they exhibit. In the words of Zhang et al. (2021), “Conventional wisdom attributes small generalization error either to properties of the model family or to the regularization techniques used during training… these traditional approaches fail to account for the exceptional generalization capability of large neural networks in practical settings.” Ordinarily, regularization techniques devised to counteract overfitting have played a pivotal role in creating models that effectively generalize across both training and testing data. Typically, when a model displays a high training rate (indicating accurate predictions for the majority of observations in the training set), the natural expectation is for it to exhibit a low test rate (indicating incorrect predictions for the majority of observations in unseen data points). However, neural networks characterized by expansive architectures seem to defy this principle, consistently demonstrating near-perfect training rates alongside proportionately high testing rates (Zhang, 2021).

In their endeavor to illustrate this point, Zhang et al. (2021) execute a series of 'extensive systematic experiments.' Initially, they train a neural network on a dataset with accurately labeled samples and record the corresponding training and test errors. Subsequently, they introduce label alterations for each observation through nonparametric randomization and attempt to retrain the model using the same dataset. In theory, within the standard dataset, where observations are correctly labeled, there should exist sufficient information to accurately identify the predicate object $R(f)$. Consequently, one would anticipate low training and test errors. However, in the randomized dataset, where there is minimal information to discern any pattern, no identifiable target function or distinguishable predicate object $R(f)$ should be discernible within the noise. As a result, both the training and test rates should remain low.

Although the experiment revealed a dramatic decrease in the testing rate, as anticipated, the training rate remained extraordinarily high. This finding suggests that despite the absence of identifiable patterns in the model, it demonstrated a level of sophistication enabling it to identify a possibly arbitrary $R_{emp}(f)$ corresponding to a hypothetical $R(f)$ that was not replicated in the testing set (Zhang, 2021). This outcome potentially signifies an inductive inference that lies entirely beyond the scope of our conventional frame of reference for machine learning problems and challenges traditional frameworks within the field.

Informational-Bottleneck theory of generalization.

In light of this problem, one popular approach for explaining the gap in our theoretical understanding of deep learning models has been informational bottleneck theory. The core idea behind this theory is to find a representation, denoted as Z, between the data X and the output Y that yields a minimally sufficient statistic which minimizes the unnecessary information between X and Z while maximizing the information between Z and Y (Räz, 2022):

$$L(P(Z \mid X)) = argmin_{P(Z|X)} I(X;\ Z) \ - \ \beta \cdot I(Z;\ Y)$$

In this equation, the term $I(X; Z)$ represents the minimally sufficient information shared between X and its representation Z, while β represents the weighted coefficient of $I(Z; Y)$, the maximal information shared between Z and Y. Essentially, the conceptualization of the information bottleneck aims to identify a bottleneck point where the information unique to X is maximally compressed to give the most effective predictive output of Y through representation Z. This inherently supposes a generalizing process facilitated through stochastic gradient descent (a part of the fitting process) (Kawaguchi, 2023).

Comparatively, this theory successfully explains why deep learning models seem to generalize effectively even with indications of overfitting. This is because at the root of this algorithm, Z is already generalized and is used as a basis to determine exceptions in the data rather than interpolating it. Additionally, there has been substantial research which demonstrates the benefits of the informational bottleneck framework for understanding deep learning models, with recent studies making progress in directly linking this framework to the generalization error (Räz, 2022).

However, a challenge remains. The issue with this framework is that it limits the capacity to generalize a problem in such a way that is readily understandable. In the earlier formulation of the generalization error, the underlying premise was that we presuppose some inherent quality regarding an object that we wanted the algorithm to identify and generalize upon. While IB theory aims for a similar goal, this process of interacting with the model to nudge it towards identifying the same generalizable features that we can see, this is not inherently the same as Z, implying that our predicate objects Z and $R(f)$ are not functionally equivalent despite theoretically being derived from the same idea.

This puts us in a rather intriguing position. Just as Plato endeavored to define the essence of man through the notion of a 'featherless biped', although not the specific target variable we sought, we currently find ourselves reliant on a representation Z that might offer a more refined depiction of the desired object, yet a representation that remains beyond our direct perception. Due to the nature of machine learning's capacity to learn from datasets that surpass human comprehension, we often encounter challenges in comprehending the intricate representations that large-scale neural networks may generate. Consequently, even if the concept of a 'featherless biped' is not flawless, it still provides enough utility to discern the shortcomings of the model when faced with a plucked chicken. Conversely, when a compressed object Z fails, we lack the same ability to derive utility from its lack of success.

To demonstrate this issue, consider the following graphic from Goodfellow, Shlens and Szegedy (2014):

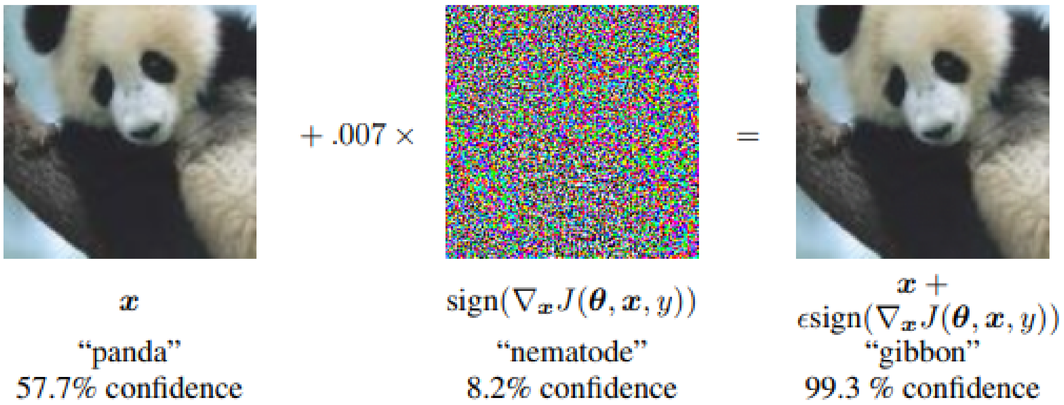

$+ .007 \times$
$=$
$\boldsymbol{x}$
"panda"
57.7% confidence
$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"nematode"
8.2% confidence
$\boldsymbol{x} + \epsilon \text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"gibbon"
99.3 % confidence

Fig 3. Adversarial training example used to intentionally fool GoogLeNet on an animal classification task.

This graphic representation showcases how machine learning models can be intentionally tricked using adversarial training examples which manipulate the gradient-descent algorithm to produce an unfounded result. While models are designed such that a small vector of noise should not greatly alter predictive outcomes, this image shows how a slight alteration to the image which is imperceptible from a human perspective can be used to fool widely-used deep learning architecture, GoogLeNet, on the ImageNet classification dataset. The problem here not only comes from the incorrect classification, but also from our inability to comprehend how this vector of noise translates into a misclassification (Goodfellow, 2014). While this model may accurately predict the majority of cases, in situations where it fails to make accurate predictions, the absence of interpretability as in a more perceivable object $R(f)$, we derive disproportionately more disutility from an adverse result than we otherwise would.

Algorithmic Bias through Methodological Similar and Dissimilar regions.

Arising from the black-box problem, a challenge for machine learning is the growing controversies surrounding the application of unfair and biased algorithms. Notable instances of this problem include the Apple Card's credit algorithm, which faced accusations of granting lower credit limits to women compared to their spouses, even when the wife had a higher credit score (Thorbecke, 2019). Another example is Amazon's discontinuation of algorithmic systems for recruitment decisions due to the discovery of gender bias (Hamilton, 2019). Additionally, Bolukbasi et al. (2016) demonstrate how gender bias can emerge in natural language processing algorithms. Even when trained on Google News articles, "exhibit female/male gender stereotypes to a disturbing extent".

Various models have been developed to address these concerns. In Bolukbasi et al. (2016), the authors leverage a large training dataset of gender specific words to mitigate gender bias in the model's output. However, this model necessitates a substantial amount of input. Raff et al. (2018) propose fair forest designs that integrate protected classes into the model to establish both 'group fairness' and 'individual fairness' metrics for comprehensive evaluations. Similarly, Zhang et al. (2018) explore an adversarial approach to model protected classes Z, ensuring that the learning algorithm optimizes predictions for Y while minimizing its capacity to predict Z.

While these models generate tenable solutions, they largely tend to circumvent the fundamental problem - a lack of theoretical understanding of methodologically unsound features of algorithms. As a result, these amount to issue-solving rather than problem-solving. When faced with challenges like racial/gender bias in hiring decisions, the interpretive challenge within machine learning becomes notably more intricate than Plato's contemplation over the plucked chicken. Describing to an algorithm, which lacks direct access to our perceptible world, the intricacies of how race factors into job applications without directly correlating to the problem at hand presents a paradoxical challenge that even we, as humans, struggle to articulate in full.

This is not to say that employing one of these methods will not give us a suitable predicate object $R(f)$ which will provide utility to some degree. However, in attempting to surmount Hume's problem of induction, machine learning has become 'data science' in that it has taken the same epistemic turn as the scientific method in deriving its own vocabulary of methods which purport to combine the ‘right’ inductive biases and empirical evidence to point us in the direction of truth. However, in this attempt to generate agnostic methodologies to solve problems, we ignore the nature of the knowledge we are attempting to understand in the first place (Lauc, 2020).

Consider the fact that in this example of a learning problem involving hiring applicants in an environment still bereft with gender and racial bias, our historical evidence implicitly points us in the direction which perpetuates these biases. After all, if we are considering a CEO position in a company and attempting to rank applicants for this position, we must acknowledge that in this scenario, the most agnostic statement we can make is that there have been less women and individuals of certain ethnicities who have occupied (let alone excelled) in that position. Attempting to identify some inherent trait which can distinguish between good applicants for a role, regardless of gender/racial bias, suggests that there is a minimally presumptive and yet maximally effective algorithm which doesn’t point out something implicitly obvious. This statement falls into the ‘magical machine learning’ trap and fails to recognize the implicitly obvious reality that the most generalizable and effective algorithm cannot help but be predisposed toward the lowest hanging fruit (i.e. systemic racial and gender bias) (Vallverdú, 2020).

What is not obvious from this scenario is that even when faced with these biases, as humans we can perceive a truth that exists beyond what is empirically discernible and in contradiction to what may be obvious to a learning algorithm - that race and gender should not determine learning outcomes. From a phenomenological perspective, the term 'bias', whether attributed to gender or race, suggests a distortion of reality, with the removal of these biases presumed to converge towards a 'true' state of the world. This assumption fosters the notion that machines can perceive this 'truth' akin to humans, and that with the appropriate architecture, the resulting algorithm would remain unbiased towards these distorted realities, unlike humans. However, this fundamentally misrecognizes that what is bizarre or even alien in this configuration, isn’t the bias itself, but the idea that this ‘truth’ that we perceive may not exist in a physicalist or epistemic sense at all.

Beyond denying that this ‘truth’ that we experience is even true at all, this opens the way to consider the knowledge that we gain through the ontological fact that we exist as beings in the world. This signifies that the knowledge we acquire may stem from a fundamental essence we possess through our existence. This idea potentially extends not only to more ‘objective’ knowledge (such as that obtained through the scientific method) but also our subjective understanding as it relates to our experience of reality. The challenge here lies in the disparity between our subjective experience, derived from the object itself $P(X \times Y)$, which perceives a set of general predicates $f_i \in F$, and our learning algorithm which produces and estimates its own set of predicates $g_i \in G$.

Where our previous problem of $R_{emp}(f)$ not converging to $R(f)$ was caused from an insufficient amount of bias, the current predicament involves $R_{emp}(f)$ converging to $G(f)$ rather than what we perceive as our target object $R(f)$. If we are to seriously entertain the notion that the problem we are attempting to solve lies beyond the realms of physical or epistemic understanding, the reason why our empirical risk aligns with the wrong predicate object can be attributed to the disparity between the 'truth' we conceive in our mind, existing in an entirely different realm of knowledge from that being modeled by the learning algorithm

Although aspects of our subjective experience may be understood in an empirical manner (such as in our ability to engage with science), when considering the fickle nature of subjectivity in philosophical discourse, it does not follow that all aspects of our subjective experience can be modeled under the same assumptions. This lends evidence to the idea that, beyond any ability to objectively model aspects of reality, certain problems cannot help but be a reflection of our own subjective interpretation of them. Ignoring this fact gives way to a naive methodology which willfully denies the problem.

To clarify, this is not to say that there indeed exists separate realms of knowledge, or a dualism composed of physical and non-physical substances. Instead, what this is to say is that this distinction between different realms of knowledge arises pragmatically as a part of machine learning methodology. That is, where the world could harbor some metaphysically ideal concept, machine learning methodologically divides this space into different regions which correspond in likeness to certain types of knowledge. This allows various metaphysical notions to coexist alongside the practical philosophy of machine learning whilst recognizing that the practice itself divides this metaphysical space into different subspaces which are each enabled through their own set of inferential postulates.

Practically speaking, this notion articulates the necessity for more dedicated research on what specifically makes problems more or less similar to each other, as well as what inferences correspond to them. In relation to this, this finding corroborates the arguments made by Sober (1991) while also being reminiscent of Hempel (1950) in his solution to the New Riddle of Induction which used similarity as a way to distinguish between lawlike and non-lawlike predicates. Additionally, this concept substantiates Vallverdu (2020) who argues that human involvement (i.e. human subjectivity) is intrinsically tied to the processes which create causal connections, albeit in a different context.

Upon this reflection, it becomes apparent why the greatest efforts in improving machine learning are derived from modeling aspects of how we, as ontological beings, perceive and interact with the world. This can be seen in Anirudh and Bengio (2022) and Desalles (2019) have made contributions to our neurocognitive understanding of how machine learning algorithms function, revealing insights that have enhanced existing algorithms. As has been revealed, this could be due to the fact that the reality that any algorithm shows us, is only a small part of the reality that there is.

Therefore, in answering our research question, we have arrived at the notion that there may exist fundamental incompatibilities between the current theoretical paradigm of machine learning and its application in different types of learning problems. While this has been studied in relation to problems with a distinct ethical dimension, such as the controversies relation to algorithmic bias, this insight does not preclude the idea that some or all of the knowledge that we gain through machine learning might be affected by this problem to some degree. Through the methodological partitioning of reality into similar and dissimilar regions, we do not have a complete framework for understanding the transitive relationships at the boundaries between one region and another. While somewhat discouraging for the AI takeover of the human race, this poses questions for further research to prove whether this similarity hypothesis (reminiscent of Hempel 1950) exists when a methodological separation of different areas of knowledge such as in machine learning is applied, as well as what these properties are.

IV. Conclusion

In conclusion, through exploring the epistemic components derived from the age-old debate between frequentism and bayesianism and instantiating these epistemic components through the relationship between the generalization error, the no-free-lunch theorem, and inductive bias, this paper synthesized a workable heuristic for understanding a predicate object classed through $R(f)$. This was analogized through the historical anecdote of Plato in his quest to find the 'essence of man' where Plato used the concept of a 'featherless biped' to interpret the object of 'man' as a feature of the world but through a concise concept. Additionally, upon rearticulating this notion through the problems of inductive inference as articulated by Hume and Goodman, it was demonstrated that there was likely no logical foundation for assigning a universal uniformity principle for inductive inference and that there was similarly no such foundation for assigning a generalizability or 'lawlike' tendency to any predicate object $R(f)$.

After constructing this framework, the black-box problem and algorithmic bias were examined through this perspective. It was argued that the shift of the discipline towards justifying deep-learning architectures has functionally changed the 'goal' of statistical learning theory as it related previously towards characterizing the predicates as a part of the problem-solving process and has taken a further step towards agnostically attempting to determine knowledge analogous to science. In this, it was argued that algorithmic bias arose from an incompatibility between the problem and the inductive inferences inherited through the 'agnostic' machine learning methodology. In this, the paper articulates the arguments of Sober (1991) and Hempel (1950) to construct the hypothesis that machine learning methodologically divides the world we metaphysically experience into similar and dissimilar regions, and concludes that further study is required to prove this concept and discover what these similar and dissimilar properties are.

Bibliography

1. Achille, A., & Soatto, S. (2018). Information dropout: Learning optimal representations through noisy computation. IEEE transactions on pattern analysis and machine intelligence, 40(12), 2897-2905.
2. Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Retrieved 26/10 from https://www.wired.com/2008/06/pb-theory/
3. Angius, N. a. P., Giuseppe and Turner, Raymond. (2021). The Philosophy of Computer Science. https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=computer-science&archive=win2017
4. Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Advances in neural information processing systems, 29.
5. Bringsjord, S. a. G., Naveen Sundar. (2022). Artificial Intelligence. The {Stanford} Encyclopedia of Philosophy. https://plato.stanford.edu/archives/fall2022/entries/artificial-intelligence/
6. Buckner, C. (2019). Deep learning: A philosophical introduction. Philosophy compass, 14(10), e12625.
7. Cheung, S., Darvariu, V., Ghica, D. R., Muroya, K., & Rowe, R. N. (2018). A functional perspective on machine learning via programmable induction and abduction. Functional and Logic Programming: 14th International Symposium, FLOPS 2018, Nagoya, Japan, May 9–11, 2018, Proceedings 14,
8. Cohnitz, D., & Rossberg, M. (2014). Nelson goodman. Routledge.
9. Dessalles, J.-L. (2019). From Reflex to Reflection: Two Tricks AI Could Learn from Us. Philosophies, 4(2), 27. https://www.mdpi.com/2409-9287/4/2/27
10. Emmert-Streib, F., Moutari, S., & Dehmer, M. (2023). Foundations of Learning from Data. In (pp. 489-520). Springer International Publishing. https://doi.org/10.1007/978-3-031-13339-8_17
11. Fantuzzi, M., Morales, H., & Whitmarsh, T. (2021). Reception in the Greco-Roman World: Literary Studies in Theory and Practice. Cambridge University Press.
12. Favaretto, M., De Clercq, E., & Elger, B. S. (2019). Big Data and discrimination: perils, promises and solutions. A systematic review. Journal of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0177-4
13. Fernandes de Mello, R., Antonelli Ponti, M., Fernandes de Mello, R., & Antonelli Ponti, M. (2018). Statistical learning theory. Machine Learning: A Practical Approach on the Statistical Learning Theory, 75-128.
14. Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.".

15. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.
16. Goodman, N. (1983). Fact, fiction, and forecast. Harvard University Press.
17. Goyal, A., & Bengio, Y. (2022). Inductive biases for deep learning of higher-level cognition. Proceedings of the Royal Society A, 478(2266), 20210068.
18. Grant, D. (2023). What is Data-Driven Decision Making? (And Why It's So Important). Retrieved 26/10/2023 from https://www.driveresearch.com/market-research-company-blog/data-driven-decision-making-ddm/
19. Hamilton, M. (2019). The sexist algorithm. Behavioral sciences & the law, 37(2), 145-157.
20. Hansson, S. O., & Zalta, E. N. (2007). Formal Learning Theory. In: Stanford Encyclopaedia of Philosophy.
21. Hempel, C. G. (1950). Problems and changes in the empiricist criterion of meaning. Revue internationale de philosophie, 41-63.
22. Horsten, L. (2007). Philosophy of mathematics. https://plato.stanford.edu/entries/philosophy-mathematics/
23. Kawaguchi, K., Deng, Z., Ji, X., & Huang, J. (2023). How Does Information Bottleneck Help Deep Learning? arXiv preprint arXiv:2305.18887.
24. Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. European Journal of Information Systems, 31(3), 388-409.
25. Lauc, D. (2020). Machine Learning and the Philosophical Problems of Induction. In (pp. 93-106). Springer International Publishing. https://doi.org/10.1007/978-3-030-37591-1_9
26. Minati, G. (2019). On theoretical incomprehensibility. Philosophies, 4(3), 49.
27. Pearl, J. (2018). Theoretical impediments to machine learning with seven sparks from the causal revolution. arXiv preprint arXiv:1801.04016.
28. Raff, E., Sylvester, J., & Mills, S. (2018). Fair Forests.
29. Räz, T. (2022). Understanding Deep Learning with Statistical Relevance. Philosophy of Science, 89(1), 20-41. https://doi.org/10.1017/psa.2021.12
30. Romeijn, J.-W. (2014). Philosophy of statistics. https://plato.stanford.edu/entries/statistics/#BasClaSta
31. Roussas, G. G. (2014). An introduction to measure-theoretic probability. Academic Press.
32. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x
33. Saltz, J. S., & Dewar, N. (2019). Data science ethical considerations: a systematic literature review and proposed project framework. Ethics and Information Technology, 21(3), 197-208. https://doi.org/10.1007/s10676-019-09502-5
34. Schurz, G. (2017). No free lunch theorem, inductive skepticism, and the optimality of meta-induction. Philosophy of Science, 84(5), 825-839.
35. Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied soft computing, 90, 106181.

36. Sober, E. (1991). Reconstructing the past: Parsimony, evolution, and inference. MIT press.
37. Sterkenburg, T. F., & Grünwald, P. D. (2021). The no-free-lunch theorems of supervised learning. Synthese, 199(3-4), 9979-10015. https://doi.org/10.1007/s11229-021-03233-1
38. Ta Liu, S. (2020). Machine Learning: A Practical Approach on the Statistical Learning: by Rodrigo Fernandes de Mello and Moacir Antonelli Pontintor. Springer International Publishing AG, part of Springer Nature, 2018, xv+ 362 pp., ISBN: 978-3-319-94988-8. In: Taylor & Francis.
39. Thorbecke, C. (2019). New York probing Apple Card for alleged gender discrimination after viral tweet. ABC News. Retrieved february/22/2020 from https://abcnews. go. com/US/york-probing-apple-card-alleged-genderdiscrimination-viral/story.
40. Vallverdú, J. (2015). Bayesians versus frequentists: a philosophical debate on statistical reasoning. Springer.
41. Vallverdú, J. (2020). Approximate and Situated Causality in Deep Learning. Philosophies, 5(1), 2. https://doi.org/10.3390/philosophies5010002
42. Van Fraassen, B. C. (1989). Laws and symmetry. Clarendon Press.
43. Van Fraassen, B. C. (2000). The false hopes of traditional epistemology. Philosophical and Phenomenological Research, 253-280.
44. VanderPlas, J. (2014). Frequentism and bayesianism: a python-driven primer. arXiv preprint arXiv:1411.5018.
45. Vidal, R., Bruna, J., Giryes, R., & Soatto, S. (2017). Mathematics of Deep Learning. arXiv pre-print server. https://doi.org/None
46. arxiv:1712.04741
47. Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. Computational Intelligence and Neuroscience, 2018, 1-13. https://doi.org/10.1155/2018/7068349
48. Witten, D., & James, G. (2013). An introduction to statistical learning with applications in R. springer publication.
49. Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. IEEE transactions on evolutionary computation, 1(1), 67-82.
50. Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating Unwanted Biases with Adversarial Learning.
51. Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. Communications of the ACM, 64(3), 107-115. https://doi.org/10.1145/3446776