Are the AI models used in science really *models*? Lessons from models in economics

Word count: 4485 words (excluding abstract and references)

Abstract

We often refer to Deep Neural Nets (DNNs) as models. This paper utilizes the use of DNNs in economics as a case study to examine whether they really function as models in science. I argue for two main claims. First, DNNs function differently from the main types of models used in science. Second, DNNs should be understood as measuring instruments that are not models. Thus, we should not refer to the DNNs used in science as models. This argument has important implications for the philosophical questions that we can legitimately ask about the DNNs used in science.

1 Introduction

Philosophers and AI researchers often refer to deep neural nets (DNNs) as “models”. But simply saying that DNNs are models does not say much. After all, philosophers have identified an (overwhelming) number of different models (Frigg and Hartmann 2024). If DNNs are models, what type of models are they (Boge 2022; Sullivan 2024)?

My starting point for answering this question is Emily Sullivan's proposal that DNNs function as toy models in science (2022, 2024). One distinctive feature of Sullivan's proposal is that she characterizes models based on their functions (Morrison and Morgan, 1999, §2.2; Frigg, 2022, §16.6). If DNNs have the same functions as models of a given type, they function as models of that type. I evaluate Sullivan's proposal by looking at how DNNs are used in economics. Economics is an ideal testing-ground for Sullivan's proposal because economic models have a variety of different functions (Morgan and Knuuttila 2012). If DNNs and toy models really have the same functions, we should be able to find evidence for this in economics.

Based on my case study, I argue for two main claims about the role of DNNs in science. First, DNNs do not function as toy models (pace Sullivan), nor as other common types of models used in science. Second, the DNNs used in most sciences should be understood as measuring instruments rather than models.[1]

But why does it matter that DNNs are measuring instruments rather than models? The answer has important implications for the philosophical questions that we can legitimately ask about the use of DNNs in science. Consider the question of what type of understanding of a target system DNNs deliver (Boge 2022; Sullivan 2022). I show that since DNNs are measuring instruments, we should expect DNNs to provide scientists with pragmatic understanding and not explanatory understanding (c.f. Kieval and Westerblad 2024).

The paper proceeds as follows. I begin by outlining Sullivan's proposal (Section 2). I then discuss how DNNs function in economics (Section 3). Next, I argue why DNNs do not function like toy models and other common model types (Section 4) and why they should be understood as measuring instruments (Section 5) before concluding (Section 6).

1. I do not deny that DNNs might be models in certain sciences (e.g., neuroscience), but only deny that they are models in *most* sciences.

2 The functions of models

To determine the model type of DNNs, we need to find a way of distinguishing different model types. One common way of distinguishing model types is based on their functions (Morrison and Morgan, 1999, §2.2; Frigg, 2022, §16.6). We can use such a functional characterization to determine if a given model functions like the models of a particular type:

> **Functional characterization of models.** An object O functions as models of type T if object O is used for the same objectives as models of type T.

The functional characterization does not tell us that object O shares the same characteristics as type T models, but only that it has the same functions as type T models. This information could still tell us something useful about the object. For example, if I learn that a given stool functions like a chair, the functional characterization suggests that this stool belong to the type of chairs. This information is useful because it tells me that I can sit on a stool. Similarly, if I learn that a model functions like a model of type T, I also learn what I can use that model for, which could be helpful.

I adopt the functional characterization for two reasons. First, it is a relatively weak condition relative to other characterizations (e.g., see Frigg 2022, §16). Second, one of the main objectives of this paper is to evaluate Sullivan's proposal (2024) and she adopts this characterization. In particular, she argues that DNNs function as toy models. Toy models are characterized by three main features (Reutlinger, Hangleiter, and Hartmann, 2018; Nguyen, 2020). They represent a target system, are strongly idealized, and are extremely simple. DNNs *prima facie* look nothing like toy models. For example, we usually consider DNNs to be extremely complex rather than extremely simple. However, this characterization of toy models is not functional: it focuses on the core features of toy models. When Sullivan (2024) says that the DNNs "used in science *function* as highly idealized toy models" (my emphasis), she has in mind a functional characterization of models. DNNs "function representationally and epistemically in a similar way as highly idealized toy models do in science" (Sullivan, 2024, p.1446).

Sullivan then explains why toy models and DNNs have the same functions in science. For her, the main function of toy models is to provide how-possibly explanations (Sullivan 2022, §2). A how-possibly explanation of why a phenomenon occurs does not offer the actual explanation of why it occurs, but merely a *possible* reason why it occurs (Sullivan, 2022, p.112). Since DNNs also provide how-possibly explanations, they have the same function as toy models (Sullivan, 2024, §3.3; 2022, §5.1). For example, consider deep patient models in medicine (Miotto et al. 2016; Sullivan 2022). These models use electronic health records to predict the probability of having a certain disease. Like toy models, deep patient models provide a how-possibly explanation. They show that it is possible to perform accurate disease diagnosis solely based on past medical records. Like toy models, they do not provide an actual explanation. They do not tell us, for instance, how a particular patient actually developed a medical problem. This is why Sullivan thinks that DNNs function like toy models in science.

While Sullivan's proposal was motivated by deep patient models, she takes it to apply globally across science. She talks about DNNs "used in science" and not just in parts of science (Sullivan 2024, p.1445). So, her proposal entails that DNNs should provide how-possibly explanations across all sciences.

3 DNNs in economics

To test Sullivan's proposal, this section describes how economists use DNNs.[2] Economics is an ideal domain for identifying their functions because models are used for a great variety of tasks there (Morgan and Knuuttila 2012). If DNNs function like toy models, we should be able to find evidence in economics. Moreover, several economists have recently reviewed the functions of DNNs in their subdomains (Korinek 2023; Fernández-Villaverde, Nuño, and Perla 2024; Dell 2025). This recent literature helps us to identify the two main functions of DNNs in economics:[3]

1. **Predicting**: Economists use DNNs to predict accurately outcomes of interest.

2. **Measuring**: Economists use DNNs to measure economic concepts that they do not directly observe.

We will now examine how economists have utilized DNNs for predicting and measuring in more depth.

3.1 Predicting

DNNs allow us to model complex nonlinear relations between inputs and outputs. In predicting problems, economists are not interested in whether the inputs explain the outputs, but how well the model predicts the outputs. Empirical asset pricing is one of the major predicting problems in economics (Gu, Kelly,

2. I refer to DNNs as models in this section for two reasons: i) this is how economists refer to them and ii) for now, we have not yet questioned this classification.

3. The literature discusses two other functions of DNNs. First, economists use DNNs as tools to increase their productivity (Korinek 2023). Second, DNNs can be used to find approximate solutions for complex dynamic economic models (Fernández-Villaverde, Nuño, and Perla 2024). I excluded these functions because they treat DNNs as instruments rather than models. So, they do not provide a good test for Sullivan's proposal. However, these functions support the general point in this paper: DNNs are instruments, not models.

and Xiu 2020; Kelly, Xiu, et al. 2023). In fact, economists explicitly recognize it as such: empirical asset pricing is "fundamentally a problem of prediction" (Gu, Kelly, and Xiu 2020).

The goal of empirical asset pricing is to accurately predict the future return on a financial asset such as a company's stock using past financial data. To find the optimal model for this task, economists compare how well different models predict asset returns given a fixed set of inputs. The inputs include various financial indicators, considered as reliable predictors of asset returns (e.g., the most recent company's profits). The early literature utilized linear models, but Gu, Kelly, and Xiu (2020) pioneered the use of machine learning (including DNNs). In their study, Gu, Kelly, and Xiu (2020) use a dataset that contains asset returns for almost 30000 stocks for the period 1957-2016 and over 900 predictors (§2.1).

Their analysis compares six models: linear models, linear models estimated with penalized least squares, linear models after dimensionality reduction, generalized linear models, tree-based models, and DNNs. They find that DNNs and tree-based models significantly outperform traditional models in terms of out-of-sample predictions (Tables 1 and 5). These findings sparked a large literature, in which DNNs have become arguably the most popular model for empirical asset pricing (Kelly, Xiu, et al. 2023). More recently, economists have also started utilizing DNNs to process alternative data sources that they could not use before. For example, they have used text data instead of standard financial indicators as predictors in empirical asset pricing (Lopez-Lira and Tang 2024). These examples show how economists have successfully utilized DNNs for the goal of predicting.

Unfortunately, predicting problems are relatively uncommon in economics. Many economists view this as the main obstacle to the broader use of DNNs

(Mullainathan and Spiess 2017). After recognizing DNN's great ability to predict, Gentzkow, Kelly, and Taddy (2019) say: "In many social science studies, however, the goal is to go further". What they mean is that economists usually do not want just to predict a variable, but to estimate causal effects or design policy interventions. The fact that DNNs excel at predicting does not help with more conventional problems in economic research.

3.2 Measuring

But if pure predicting problems are rare in economics, why have economists recently got so excited about DNNs (Dell 2025; Korinek 2023)? The reason is that they have found new ways of utilizing DNNs' predictive capabilities to solve more conventional problems. Specifically, economists have started thinking of DNNs as methods for *measuring* rather than just predicting (see also Tal 2017a; Mussgnug 2022). Measuring is more demanding than predicting because it not only requires that a DNN predicts an output well but also that the predicted output is connected to a measurement of an economic concept.

Thinking of DNNs as excelling at measuring rather than predicting has allowed economists to solve more conventional problems in their domain. An economist can first use a DNN to measure an economic concept and then use this newly measured concept to resolve such a problem. Dell (2025) expressed this insight clearly:

> [DNNs] impute low-dimensional structured data from unstructured texts or images ... This structured data is then used for causal or descriptive analyses

We can identify two steps in the workflow of economists who use DNNs for measuring. First, they use a DNN to measure an economic concept using unstructured data. We can call this step the *measuring step*. Second, they utilize the economic concepts in a more standard "causal or descriptive analysis" such as estimating a treatment effect. We can call this step the *analysis step*. The two-step procedure is crucial for understanding why economists have started using DNNs for measuring economic concepts.

3.2.1 Case study: measuring crop residue burning

We will now examine in more depth a case study of how two-step procedure works: Jack et al. (2025). The authors study how to reduce crop residue burning in India. After harvesting their field, Indian farmers often burn their crop residue to reduce the waiting period between harvests. The trouble is that this practice is a major source of air pollution, which has various negative health consequences (Ghude et al. 2016). This is why Jack et al. (2025) conducted a randomized controlled trial that evaluated two policies aimed at reducing crop residue burning in Indian villages. The policies provided cash transfers to farmers on the condition that they refrained from crop residue burning. A farmer would receive a transfer only after the economists have verified that she had not burnt her fields.

The economists randomly assigned the participating farmers to three groups. First, the control group received no transfer. Second, the first treatment group received the whole transfer only after the economists had successfully examined the farmers' fields following the harvest and verified that no burning has occurred. Third, the second treatment group received the first part of the transfer unconditionally *before* the harvest and the second part conditionally *after* the harvest, following verification.

The main data collection challenge was verifying if a farmer had burned their fields after the harvest (Jack et al, 2025, pp.44-46). It was not practically feasible for the local researchers to visit each farmer and confirm if burning had occurred. The reason is that the researchers did not know when a farmer would decide to harvest. If the researchers visited a farmer's fields too early, the harvest would not have yet been collected, so they would not be able to check for crop residue burning. If they visited the fields too late, the evidence for burning might already have disappeared. This was because any evidence from crop residue burning was quick to disappear. Figure 1 illustrates this fact. It shows satellite images of a field that was burnt on November 4. The evidence from crop burning had almost disappeared within a few days. There was a very short window in which crop residue burning could be verified.

Figure 1: Satellite images of a field burnt to remove crop residue on November 4. Source: Walker et al. (2022), i.e., the data companion paper to Jack et al. (2025).

To tackle the difficulties with verifying crop residue burning, Jack et al. (2025) used satellite images. These images allowed them to monitor a farmer's field with high frequency and avoid missing crop residue burning. They built a dataset with satellite images that tracked the state of the fields belonging to the participating farmers. They then trained a model that predicted if a field showed burn scars.[4] While they *predicted* burn scars in their models, the burn scars were then interpreted as *measurements* of crop residue burning. This inference assumed that burn scars were due to crop residue burning done by the farmer and not for any other reason.

4. In their final paper, Jack et al. (2025) did not end up using a DNN for that task, but a random forest model. However, as they acknowledge in their companion paper (Walker et al. 2022), they could have used a DNN but found the random forest more convenient in this particular case. Several other papers in economics that utilize satellite images employ DNNs instead of random forests for measuring (Jean et al. 2016; Huang, Hsiang, and Gonzalez-Navarro 2021). The fact that Jack et al. (2025) utilized random forests rather than DNNs does not invalidate the argument about the functions of DNNs.

After measuring crop residue burning, Jack et al. (2025) estimated the causal effect of the two treatments. Their results showed that giving the full cash transfer after the harvest (the first treatment) did not decrease crop residue burning. In contrast, giving part of the cash transfer before the harvest (the second treatment) significantly decreased it.

The study by Jack et al. (2025) maps seamlessly into the two-step procedure for the workflow of economists who use DNNs for measuring. First, the *measuring step* involved using satellite images to predict burn scars and then measure crop residue burning. Second, the *analysis step* involved using the measurements of crop residue burning to estimate the causal effect of different interventions. While in Jack et al. (2025) the measurements from the first step were used for causal inference in the second step, they could also be used for other purposes (Balboni, Burgess, and Olken 2025). These studies show how economists have successfully integrated DNNs into their workflow by using them for measuring.

4 Do DNNs function as models in science?

In this section, I use the case study to argue that DNNs do not function as toy models (Section 4.1), but as data processing methods (Section 4.2).

4.1 DNNs do not function as toy models

The case study establishes that DNNs have two main functions in economics: predicting and measuring. In contrast, Sullivan's proposal entails that DNNs should be providing how-possibly explanations across all of sciences. The case study puts pressure on her proposal. DNNs do not share the objectives of toy models in economics and thus do not seem to function as toy models in science.

We might worry that my argument focuses too much on economics. From the fact that DNNs do not function as toy models in economics, it infers that toy models do not function as toy models in all of science. Perhaps we can qualify Sullivan's proposal by saying that DNNs function as toy models in *most* sciences. Maybe economics is just an exception, and Sullivan's proposal still applies to most sciences. Then, my case study will no longer pose a threat to her proposal.

DNNs undoubtedly function as models in some sciences. For example, early convolutional neural networks were biologically inspired, and their predecessors aimed at representing the vision system of living organisms (Lindsay 2021). However, this fact does not entail that DNNs function as *toy* models in these sciences. On the functional characterization, whether they function as toy models depends on how they are used. The problem is that DNNs seemed to be used for different functions than toy models. For example, Buchholz and Grote (2023) and Mussgnug (2022) show that they are used for predicting and measuring rather than how-possibly explanations in social sciences outside of economics. At best, Sullivan's proposal holds outside social sciences, but it definitely does not hold across all sciences.[5]

4.2 DNNs do not function as data models

If DNNs do not function as toy models, is there another model type that can accommodate their functions in economics? One plausible candidate is data models (Suppes 1962; Harris 2003; Leonelli 2019; Bokulich and Parker 2021). Data models are "arrangements of data that are evaluated, manipulated and modified with the explicit goal of representing a phenomenon" (Leonelli 2019, p.2). In other words, data models are transformations of data whose function

5. DNNs seem to be used for functions unrelated to how-possibly explanations in natural sciences as well. This seems to be the case for the life sciences (Lawrence et al. 2026) and physics (Stoll 2025).

is to represent a phenomenon. Kepler's laws are an example of a data model (Harris 2003). Kepler used data on the position of the planets over time, to which he applied curve-fitting methods. The data models were the final elliptic curves that represent planetary orbits. However, data models need not be statistical (Leonelli 2019). They could be images that are processed with the aim of representing the traits of plants in biology or galaxies in astronomy.

Predicting and measuring also seem to be among the functions of data models. When we predict an output, we manipulate data to represent a feature of our target system. In empirical asset pricing, the objective is to represent future asset returns, which are features of the target system (the financial markets). When we measure an economic concept, we are using data to represent an unobserved feature of our system. Jack et al. (2025), for example, aim to determine whether a field has undergone crop residue burning. For them, crop residue burning is the represented feature of the target system. Since data models can be used for predicting and measuring, DNNs in economics seem to function as data models.

However, notice that DNNs are not arrangements of the data, but the method that arranges the data. The DNNs are a data processing method rather than a data model. Data processing methods are the methods used to transform a data input into another data output.[6] To see the difference between data models and data processing methods, suppose we have a sample of heights of people drawn from a given population. We can use the method of taking the average to calculate the average height. Here the average height is an arrangement of our data and could be taken as a representation of a feature of the target system, i.e., the expected height in the population. Average height is a data model

6. Leonelli (2019, pp.22-23) understands data processing *procedures* differently. For her, such procedures aim to constrain the set of knowledge claims that we can infer from a dataset. In contrast, I understand data processing *methods* as the specific methods used to analyze data. They are what connect the data input and data output.

whereas taking the average is a data processing method.

Data models are arrangements of data whose function is to represent. They come after we have applied data processing methods to our data. On the other hand, the function of data processing methods is to arrange the data so that the output can be a data model. The crucial difference between data processing methods and data models is that the former is not necessarily representational. The method of taking the average on its own does not represent any physical objects, but is a mechanical procedure applied to the data. This is in contrast to data models, whose function is to represent.

This idea suggests that DNNs function as data processing methods rather than data models in economics. When we use financial data to represent future asset returns, the data model is the predicted future asset returns. Similarly, when we use satellite images to detect burn scars, the data model is a binary variable indicating if crop residue burning has occurred. In both cases, the DNN is the method used for processing the satellite images that outputs the data models.

5 DNNs as measuring instruments

The previous section suggests that DNNs do not function as data models or toy models, but as data processing methods. As used in economics, DNNs seem more akin to investigative instruments such as thermometers and telescopes than to standard model types.[7] This raises the question of whether DNNs are models or instruments. The answer to this question would require us to abandon the functional characterization of models. While that characterization helped us to identify the *functions* of DNNs in economics, it cannot tell us if they are models

7. Mitchell (2020) offers a complementary argument that reaches a similar conclusion using a case study from biology.

or instruments. This is why this section leaves aside the functional characterization and directly examines the more fundamental question of whether DNNs are models or instruments.

One difficulty with addressing this question is that certain models *do* function as measuring instruments. For example, a model of a real pendulum can function as a measuring instrument: it can be used to measure gravitational acceleration (Morrison 1999). However, not all measuring instruments are models: a thermometer is a measuring instrument that is not a model (Morrison and Morgan 1999, p.25). This insight suggests how we can rephrase the original question so that it becomes more manageable. Are DNNs i) models that function as measuring instruments or ii) measuring instruments that are not models?

We can answer this new question by examining a domain that clearly distinguishes between measuring instruments and models. The philosophy of measurement provides such a domain (Tal 2017a, 2017b, 2020; Boumans 2012; Mussgnug 2022). If DNNs are more akin to what philosophers of measurement call models, then they are models. If they are not, they are measuring instruments.[8]

To understand how philosophers of measurement distinguish measuring instruments and models, we need to introduce Erin Tal's distinction between *instrument indications* and *measurement outcomes* (2017a, p.235; 2017b, pp.34-35). Instrument indications are the properties of measuring instruments that we directly observe at the end of a measurement process. They could be the level of mercury in a thermometer or the digits displayed on a digital clock.

8. Mussgnug (2022) was the first to draw the connection between philosophy of measurement and DNNs used in social science. While I agree with his general framework, his discussion is not always clear about whether DNNs are what philosophers of measurement call models or instruments. At first, he understands DNNs as measurement models that function as measuring instruments and provide instrument indications (p.13). Later, however, he considers them as models that deliver measurements (p.15). Consequently, we cannot use his argument to answer the main question in this section.

In these cases, the thermometer and the clock are the measuring instruments whose properties are the instrument indications. On the other hand, measurement outcomes are the quantities whose values we actually want to measure. They could be the exact boiling point of a liquid or the time it took to reach that boiling point. The measurement outcomes are unknown properties of the object that we are trying to measure.

It is not always straightforward to relate an instrument indication to a measurement outcome. Instrument indications are properties of the measuring instrument, whereas measurement outcomes are properties of the object we study. Suppose we have instrument indications of the same quantity from five different measuring instruments. How can we extract the true measurement outcome from them? This is exactly where models enter philosophy of measurement (Tal, 2020, §7; Boumans, 2012, pp.414-415). They map instrument indicators into measurement outcomes. They tell us how to combine the five instrument indications to derive a single measurement outcome under certain assumptions, such as the reliability of the different instruments. If five thermometers deliver different instrument indications of the boiling point of a liquid, we can use a model to determine the true boiling point, i.e., the measurement outcome.

While measuring instruments deliver instrument indications, models map these instrumental indications into measurement outcomes. Given this distinction, we can see why DNNs, as used in economics, are more akin to measuring instruments. In the case of measuring crop residue burning, the objective was to determine if crop residue burning had occurred in a field. Jack et al. (2025) took burn scars as evidence for crop residue burning. The burn scars predictions from the DNNs were the instrument indications, and residue crop burning was the measurement outcome. Since the DNN was used to predict burn scars given images of the fields, it was a measuring instrument rather than a model.

The model here tells us how to use the images with predicted burn scars on a field to determine whether crop residue burning has occurred. The same verdict holds for empirical asset pricing. There, the DNN provided a number indicating asset return. To come up with the best possible forecast for the future asset returns (i.e., the measurement outcome), financial economists would usually combine the instrument indication with other information, such as predictions from other sources, or the DNN's reliability. It thus seems that we can only accommodate economists' use of DNNs if we understand DNNs as measuring instruments. This insight supports the idea that DNNs are measuring instruments rather than models.

6 Conclusion

The scientific practice of economists suggests that DNNs do not function as the common model types. Instead, they seem to be measuring instruments rather than models. This conclusion has important implications for the legitimate philosophical questions that we can ask about the use of DNNs in science. Consider the question of whether DNNs provide scientists with an understanding of the target system. What is the right account of understanding to use (Boge 2022; Sullivan 2022; Kieval and Westerblad 2024)?

Philosophers of science distinguish between explanatory understanding and pragmatic understanding (De Regt and Dieks 2005; Stuart 2025). According to explanatory understanding, a scientist understands a phenomenon if she possesses an explanation of why it occurs. If a scientist can explain why the planets revolve around the sun, she has explanatory understanding of this phenomenon. On the other hand, pragmatic understanding states that a scientist understands a particular system if she has the ability to analyze or intervene on this system to achieve her objectives. If a scientist knows how to predict the future positions of the planets in the solar system, she has a pragmatic understanding of the system.

The fact that DNNs are measuring instruments suggests that scientists using DNNs gain a pragmatic understanding of a system.[9] The reason is that measuring instruments (other than models) do not generally provide explanations, but still allow scientists to analyze a system to achieve their objectives. So, DNNs provide pragmatic understanding rather than explanatory understanding. We can thus see how our characterization of DNNs as measuring instruments sheds light on an important philosophical question about DNNs.

9. This conclusion supports the positions of Kieval and Westerblad (2024) and Stuart (2025) and objects to Sullivan (2022).

References

Balboni, Clare, Robin Burgess, and Benjamin A Olken. 2025. "The origins and control of forest fires in the tropics." *Review of Economic Studies,* rdaf088.

Boge, Florian J. 2022. "Two dimensions of opacity and the deep learning predicament." *Minds and Machines* 32 (1): 43–75.

Bokulich, Alisa, and Wendy Parker. 2021. "Data models, representation and adequacy-for-purpose." *European Journal for Philosophy of Science* 11 (1): 31.

Boumans, Marcel. 2012. "Measurement in Economics." In *Philosophy of economics,* edited by Uskali Mäki, Dov M. Gabbay, Paul Thagard, and John Woods, 395–423. North Holland.

Buchholz, Oliver, and Thomas Grote. 2023. "Predicting and explaining with machine learning models: Social science as a touchstone." *Studies in History and Philosophy of Science* 102:60–69.

De Regt, Henk W, and Dennis Dieks. 2005. "A contextual approach to scientific understanding." *Synthese* 144 (1): 137–170.

Dell, Melissa. 2025. "Deep learning for economists." *Journal of Economic Literature* 63 (1): 5–58.

Fernández-Villaverde, Jesús, Galo Nuño, and Jesse Perla. 2024. *Taming the curse of dimensionality: quantitative economics with deep learning.* Technical report. National Bureau of Economic Research.

Frigg, Roman. 2022. *Models and theories: A philosophical inquiry.* Routledge.

Frigg, Roman, and Stephan Hartmann. 2024. "Models in Science." In *The Stanford Encyclopedia of Philosophy,* Fall 2024, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as data." *Journal of Economic Literature* 57 (3): 535–574.

Ghude, Sachin D, DM Chate, C Jena, G Beig, R Kumar, MC Barth, GG Pfister, S Fadnavis, and Prakash Pithani. 2016. "Premature mortality in India due to PM2. 5 and ozone exposure." *Geophysical Research Letters* 43 (9): 4650–4658.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. "Empirical asset pricing via machine learning." *The Review of Financial Studies* 33 (5): 2223–2273.

Harris, Todd. 2003. "Data Models and the Acquisition and Manipulation of Data." *Philosophy of Science* 70 (5): 1508–1517. https://doi.org/10.1086/377426.

Huang, Luna Yue, Solomon M Hsiang, and Marco Gonzalez-Navarro. 2021. "Using Satellite Imagery and Deep Learning to Evaluate the Impact of Anti-Poverty Programs." *NBER Working Paper.*

Jack, B Kelsey, Seema Jayachandran, Namrata Kala, and Rohini Pande. 2025. "Money (not) to burn: payments for ecosystem services to reduce crop residue burning." *American Economic Review: Insights* 7 (1): 39–55.

Jean, Neal, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. "Combining satellite imagery and machine learning to predict poverty." *Science* 353 (6301): 790–794.

Kelly, Bryan, Dacheng Xiu, et al. 2023. "Financial machine learning." *Foundations and Trends® in Finance* 13 (3-4): 205–363.

Kieval, Phillip Hintikka, and Oscar Westerblad. 2024. "Deep Learning as Method-Learning: Pragmatic Understanding, Epistemic Strategies and Design-Rules." PhilSci Archive 23489. https://philsci-archive.pitt.edu/23489/.

Korinek, Anton. 2023. "Generative AI for economic research: Use cases and implications for economists." *Journal of Economic Literature* 61 (4): 1281–1317.

Lawrence, Elsa, Adham El-Shazly, Srijit Seal, Chaitanya K Joshi, Pietro Liò, Andreas Bender, Shantanu Singh, Pietro Sormanni, Ola Spjuth, and Matthew Greenig. 2026. "Understanding Biology with Machine Learning: Compression, Intelligibility, and Dependency." *Artificial Intelligence in the Life Sciences,* 100161.

Leonelli, Sabina. 2019. "What Distinguishes Data From Models?" *European Journal for Philosophy of Science* 9 (2): 22.

Lindsay, Grace W. 2021. "Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future." *Journal of Cognitive Neuroscience* 33 (10): 2017–2031.

Lopez-Lira, Alejandro, and Yuehua Tang. 2024. *Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models.* arXiv: 2304.07619 `[q-fin.ST]`. https://arxiv.org/abs/2304.07619.

Miotto, Riccardo, Li Li, Brian A Kidd, and Joel T Dudley. 2016. "Deep patient: an unsupervised representation to predict the future of patients from the electronic health records." *Scientific reports* 6 (1): 1–10.

Mitchell, Sandra D. 2020. "Instrumental perspectivism: Is AI machine learning technology like NMR spectroscopy?" In *A Critical Reflection on Automated Science: Will Science Remain Human?,* edited by Marta Bertolaso and Fabio Sterpetti, 27–42. Springer.

Morgan, Mary S, and Tarja Knuuttila. 2012. "Models and modelling in economics." In *Philosophy of economics,* edited by Uskali Mäki, 13:49–87. Amsterdam: Elsevier.

Morrison, Margaret. 1999. "Models as autonomous agents." In *Models as Mediators: Perspectives on Natural and Social Science,* edited by Mary S. Morgan and Margaret Morrison, 38–65. Cambridge University Press.

Morrison, Margaret, and Mary S. Morgan. 1999. "Models as mediating instruments." In *Models as Mediators: Perspectives on Natural and Social Science,* edited by Mary S. Morgan and Margaret Morrison, 10–37. Cambridge University Press.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine learning: an applied econometric approach." *Journal of Economic Perspectives* 31 (2): 87–106.

Mussgnug, Alexander Martin. 2022. "The Predictive Reframing of Machine Learning Applications: Good Predictions and Bad Measurements." *European Journal for Philosophy of Science* 12 (3): 1–21.

Nguyen, James. 2020. "It's not a game: Accurate representation with toy models." *The British Journal for the Philosophy of Science.*

Reutlinger, Alexander, Dominik Hangleiter, and Stephan Hartmann. 2018. "Understanding (with) Toy Models." *British Journal for the Philosophy of Science* 69 (4): 1069–1099.

Stoll, Frauke. 2025. "Understanding (with) Deep Neural Networks in Particle Physics." Talk presented at the European Philosophy of Science Association Conference (EPSA 2025), Groningen, Netherlands, August 29, 2025.

Stuart, Michael T. 2025. "A New Account of Pragmatic Understanding, Applied to the Case of AI-Assisted Science: MT Stuart." *Philosophical Studies,* 1–25.

Sullivan, Emily. 2022. "Understanding from machine learning models." *The British Journal for the Philosophy of Science.*

———. 2024. "Do Machine Learning Models Represent Their Targets?" *Philosophy of Science* 91 (5): 1445–1455.

Suppes, Patrick. 1962. "Models of Data." In *Logic, methodology and philosophy of science,* edited by Ernest Nagel, Patrick Suppes, and Alfred Tarski. Stanford UniversityPress.

Tal, Eran. 2017a. "A Model-Based Epistemology of Measurement." In *Reasoning in Measurement,* edited by Nicola Mößner and Alfred Nordmann, 233–253. Taylor & Francis.

———. 2017b. "Calibration: Modelling the measurement process." *Studies in History and Philosophy of Science Part A* 65:33–45.

———. 2020. "Measurement in Science." In *The Stanford Encyclopedia of Philosophy,* Fall 2020, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Walker, Kendra, Ben Moscona, Kelsey Jack, Seema Jayachandran, Namrata Kala, Rohini Pande, Jiani Xue, and Marshall Burke. 2022. *Detecting Crop Burning in India using Satellite Data.* arXiv: 2209.10148 [cs.CV]. https://arxiv.org/abs/2209.10148.