AI Operationalism: The case of international development

Alexander Martin Mussgnug

Abstract:

Machine learning is rapidly transforming how society and humans are quantified. Shared amongst some machine learning applications in the social and human sciences is the tendency to conflate concepts with their operationalization through particular tests or measurements. Existing scholarship reduces these equations of concept and operationalization to disciplinary naivety or negligence. This paper takes a close look at equations of concept and operationalization in machine learning predictions of poverty metrics. It develops two arguments. First, I demonstrate that conflations of concept and operationalization in machine learning poverty prediction cannot be reduced to naivety or negligence but can serve a strategic function. Second, I propose to understand this function in the context of philosophical and historical research on operationalism in the social sciences.

I Introduction

Machine learning is rapidly transforming how society and humans are quantified. Among others, machine learning models are employed in the measurement and classification of depression (Priya et al., 2020; Sau & Bhakta, 2017), student engagement (Alruwais & Zakariah, 2023; Bosch, 2016), poverty (Blumenstock et al., 2015) and political affiliation (Pennacchiotti & Popescu, 2011). Yet, some machine learning applications conflate the concept of interest with the specific operationalization used as their target variable. For instance, machine learning practitioners conflate depression with particular psychometric tests of depression (Zulfiker et al., 2021), learning with particular behavioral data (Knox et al., 2020), and poverty with particular poverty metrics (Blumenstock et al., 2015).[1]

Existing scholarship reduces these equations of concept and operationalization to disciplinary naivety or negligence and suggests the integration of richer methodological resources as a solution. However, despite long-standing efforts to promote more sophisticated conceptual and methodological practice, conflations of concept and operationalization persist in many areas of empirical machine learning. This paper takes a close look at equations of concept and operationalization in machine learning predictions of poverty metrics. It develops two arguments. First, I demonstrate that conflations of concept and operationalization in machine learning poverty prediction cannot be reduced to naivety or negligence but can serve a strategic function. Second, I propose to understand this function in the context of philosophical and historical research on operationalism in the social sciences.[2]

This paper proceeds in four parts. Section two offers an overview of current scholarship on conflations of concept and operationalization in empirical machine learning. Section three provides a detailed case study of conceptual and operational collapse in machine learning poverty predictions. I highlight how the

[1] Throughout the paper, I use the noun operationalization roughly to refer to the operations used to empirically assess an (abstract) concept, not their concrete execution. In the case of social science measurement, operationalization then refers to the measurement model which specifies the mathematical relationship between observable properties and measurement outcome (which seeks to inform us about the unobservable theoretical concept). Colloquially, one often conflates measurement and measurement model and, thus, measurement and operationalization.

[2] Operationalism goes beyond mere operationalization. Operationalism (for ontological or pragmatic reasons) reduces a concept to a particular operationalization.

relevance of machine learning predictions for their intended use in policymaking depends on them approximating metrics that capture aspects of poverty relevant for actual or potential policymaking scenarios in a given context. However, machine learning researchers' choice of a target variable can be driven by convenience and expedience in a manner that renders it disconnected from policy needs. Here, equations of concept and operationalization can signal the prima facie policy relevance of machine learning poverty predictions. By collapsing the distinction between concept and operationalization, machine learning applications are presented as predicting "poverty" and, thus, seemingly provide information relevant to the aims of international development policymaking.

Section four proposes to understand such equations of concept and operationalization in empirical machine learning as an instance of *AI operationalism*. AI operationalism captures how equations of concept and operationalization in empirical machine learning temporarily fix the referent of complex concepts in ways that serve a strategic yet pragmatic function. While attributing conceptual and operational conflations only to naivety and negligence leads scholars to immediately move to disciplinary reformation, characterizing these conceptual practices as an instance of operationalism stresses them as worthy of more substantial investigation in their own right. Moreover, contextualizing equations of operationalization and concept in empirical machine learning within historical scholarship of operationalism can help better understand both the problematic implications of AI operationalism and the potential value of its limited and responsible use. I conclude in section five.

II Conflations of concept and operationalization: Disciplinary naivety?

A common critique of empirical machine learning in the social and human sciences concerns the tendency of some machine learning practitioners to conflate concepts and their operationalizations. Existing scholarship attributes this to disciplinary naivety and negligence. Machine learning practitioners' failure to effectively distinguish between concepts and operationalizations, so a prominent line of reasoning goes, stems from lacking sophistication in the absence of the appropriate methodological resources — and, in particular, machine learning practitioners' insufficient familiarity with measurement modeling and construct validation.

In their influential research on measurement and sociotechnical machine learning, Abigail Jacobs and Hanna Wallach (2021, p. 10) remark that "although measurement modeling is fundamental to the quantitative social sciences, it has not traditionally played a role in computer science. As a result, researchers and practitioners are often inclined to collapse the distinctions between constructs and their operationalizations, either colloquially or epistemically. […] Moreover, because most computational systems are developed by computer scientists, the practice of collapsing these distinctions is widespread." This concern is echoed by Jason Radford and Kenneth Joseph (2020) and in recent work by Moritz Herrmann and coauthors (2024). In a similar vein, Momin Malik attributes conflations of concept and operationalization to negligence. He observes that "many times in practice, when machine learning makes claims, it *forgets* to recognize and communicate that the measurements serving as "ground truth" are a black box that hide problems of measurement, and that the ground truth is not construct itself" (2020, p. 11, emphasis added).

Scholars have stressed the decisive ethical and epistemic implications of this practice. On the side of ethics, work has focused on how unacknowledged mismatches between concept and operationalization can lead to fairness-related harms when models are employed in decision-making scenarios. A widely-discussed example is the failure to sufficiently acknowledge potential discrepancies between "care need" as a relevant concept for decision-making in health care allocation and "care costs" as a possible operationalization of it. If the distinction between the concept of care need and its operationalization as care costs is underappreciated, patients' insurance status and its effect on billing might distort allocation decisions leading to an unfair provision of care (e.g., Jacobs & Wallach, 2021, p. 4; Tal, 2023). On the side of epistemology, an insufficient distinction between concepts and operationalizations might play into much-discussed replication issues in empirical machine learning. For example, failing to acknowledge how different data generation processes result in different operationalizations of the same abstract concept can mislead comparative analyses (Herrmann et al., 2024).

Attributing the conflation of concepts and operationalizations primarily to disciplinary negligence and naivety has led research to focus on two proposed solutions. First, scholars stress the need to cultivate a greater awareness for the distinction between concept and operationalization in applied machine learning research. Second (and relatedly), they suggest bringing in methodological resources, such as measurement modeling and construct validation, from more established disciplines like psychology and physics (Jacobs & Wallach, 2021; Radford & Joseph, 2020; Tal, 2023).

Despite long-standing efforts to promote more sophisticated conceptual practice (e.g., Myrtveit et al., 2005; Wagstaff, 2012) and the undeniable maturing of empirical machine learning as a field of research, progress has arguably been insufficient. As Herrmann et al. (2024, p.4) note, "it is puzzling that validity and other quality criteria of empirical research have gained little attention in ML so far." While it is true that many within the machine learning community have begun to engage substantially with issues such as construct validation and the limitations of proxy variables (e.g., Guerdan et al., 2022; Kleinberg et al., 2018), these efforts have not yet translated into sophisticated conceptual practice in large areas of empirical machine learning where conflations of concept and operationalization still regularly occur.[3]

One such area is international development. Existing scholarship often engages only briefly with illustrative cases of operational and conceptual conflation before articulating a vision for disciplinary reformation (Herrmann et al., 2024; Jacobs & Wallach, 2021; Radford & Joseph, 2020). In contrast, the following section will examine in greater detail conceptual and operational conflations in machine learning poverty predictions. I argue that, in the case of machine learning poverty predictions, there is more to these practices than disciplinary naivety and negligence alone. Equations of concept and operationalization can serve a strategic function.

III Machine Learning Poverty Prediction

This section begins by illustrating the tendency to equate particular operationalizations of poverty with the concept itself in some instances of machine learning poverty prediction. I then highlight how the relevance of these predictions for their intended use in policymaking depends on them approximating metrics that capture aspects of poverty relevant for actual or potential policy scenarios in a given context. The choice of a target variable, however, can be driven by convenience and expedience in a manner that renders it disconnected from actual policy needs. Here, equations of concept and operationalization can strategically signal the prima facie relevance of machine learning poverty predictions absent meaningful conceptual and metrological engagement.

[3] Much research on proxy variables or construct validity is motivated by its still lacking prevalence in empirical machine learning. Guerdan et al., for instance, (2022) note: "Despite its importance in establishing measures of appropriate reliance, outcome measurement error remains largely unacknowledged in AI-assisted decision-making literature. Instead, evaluations assume that the outcome proxy predicted by the AI-based tool (*Y*) is equivalent to the true outcome of interest *(Y*).*"

III.1 Equations of concept and operationalization

International development requires reliable statistics to monitor progress and target interventions. Traditional poverty data stems from costly in-person surveys, frequently outdated censuses, and social registries relying on proxies of questionable quality (Brown et al., 2016). In an effort to provide timely and accurate information on livelihood outcomes in the Global South, many have begun leveraging machine learning models trained on alternative data sources to supplement geographically and temporally scarce poverty measurements. Machine learning practitioners generally rely on two kinds of data: mobile network records and, much more frequently, satellite imagery. The vast majority of research employs a supervised learning or (more rarely) a transfer learning methodology. In the former case, models are directly trained on features extracted from mobile network or remote sensing data and existing survey-based poverty metrics. In the latter case, a convolutional neural network (CNN) is trained on widely available data such as daytime imagery to predict more scarce data such as nighttime light intensity. The learned features of the CNN are then used to train a second model to predict existing poverty measurement data.

Seminal to research on machine learning poverty prediction is the work of Joshua Blumenstock and, in particular, his coauthored 2015 *Nature* article "Predicting poverty and wealth from mobile phone metadata." In it, Blumenstock et al. leverage mobile network data to predict a wealth index for individual subscribers and microregions in Rwanda. Using automated feature extraction, Blumenstock et al. reduce the call detail records of nearly all mobile phone subscribers in the country to a range of quantitative features. As the target variable, the developers rely on a proxy of the Demographic and Health Survey (or DHS) wealth index. The DHS is a prevalent survey in developing nations administered by the US Agency for International Development. The principal component of asset-related questions in the DHS can be used as an indicator of poverty. Since DHS survey data is anonymized and not linked to mobile network records, the developers conduct a phone survey of about 900 mobile phone subscribers asking a limited number of asset-related questions found in the survey. The principal component of these questions serves as a proxy of the original DHS wealth indicator. Finally, Blumenstock et al. train a statistical model on the features extracted from the mobile network data to estimate the previously constructed proxy of the DHS asset index. Model outcomes, aggregated at the district level, are validated relative to the DHS survey with a predictive accuracy of $r = 0.916$.

In addition to being one of the first applications of machine learning to poverty estimation, Blumenstock et al.'s pioneering publication exemplifies the tendency of some applications of machine learning poverty predictions to collapse the distinction between operationalization and concept. Not only within the title but throughout the article, Blumenstock and coauthors speak of "predicting poverty" rather than predicting a particular measurement of it. Regarding the purpose of their application, for instance, they note: "Here we examine the extent to which anonymized data from mobile phone networks can be used to predict the poverty and wealth of individual subscribers" (p. 1073). In doing so, the authors blur the line between poverty and poverty metric to the point of effectively equating operationalization and concept. This conflation of operationalization and concept persists apart from one brief mention of the construction of a "wealth index" without any further information on its composition. Moreover, there is no mention of the assumptions and limitations of the particular operationalization employed or discussion of its contextual adequacy in assessing levels of deprivation in Rwanda. An epistemically significant tendency to let the operationalization stand in for the concept itself is, thus, indicated rhetorically by the conflation of poverty with its measurements and methodologically by a lack of substantive conceptual and metrological engagement elucidating the relation between them.

Blumenstock et al.'s 2015 article contrasts with other scholarship in the field that clearly and substantially distinguishes concepts from operationalizations. Instead of collapsing the distinction between operationalization and concept, many machine learning practitioners strive to responsibly elucidate theoretical, conceptual, and metrological choices made and limitations occurred in moving from concept to operationalization (e.g., Al Kez et al., 2024; Marcinko et al., 2022; McBride et al., 2022; Spandagos et al., 2023). At the same time, the article is not an outlier in the field. Similar tendencies can be observed in more recent publications. Yet their degree and definitiveness can differ greatly. Many share the rhetorical and methodological features outlined (e.g., Kim, 2021; Kondmann & Zhu, 2020; Ni et al., 2021). Only rarely do practitioners explicitly fix the referent of poverty to be a particular measurement of it by stating, for instance, that "throughout the paper, 'poverty' refers to the Global MPI" (Pokhriyal & Jacques, 2017, p. 9784).[4]

One might argue that equations of concept and operationalization are owed to the rather strict word limits in many publication venues and that more substantial conceptual engagement might be found in the supplementary material of empirical machine learning articles. However, most articles discussed here, including the one by Blumenstock et al. (2015), show no substantial conceptual engagement in their extensive supplementary material. Even in cases where one might encounter a clearer distinction between concept and operationalization, limiting such elucidation to the supplementary material of an article constitutes a choice that speaks to the strategic role of operational and conceptual conflation argued for in what follows. Understanding this role, however, requires us to first examine the stated purpose of machine learning poverty predictions.

III.2 Policy-orientedness

The expressed goal of machine learning poverty predictions is the provision of efficient and timely poverty metrics that can aid in policymaking. Proof-of-concept studies, such as Blumenstock et al. (2015), anticipate the future application of machine learning predictions to "population monitoring in remote and inaccessible regions, real-time policy evaluation, and the targeting of resources to those with the greatest need" (p. 1076). Since its publication, machine learning poverty prediction has moved from exploratory research to its first real-world policy implementations. Fueled by the COVID-19 pandemic which rendered difficult on-the-ground data collection, a limited number of machine learning applications have been employed in policy efforts, such as the targeting of poverty relief programs (e.g., Aiken et al., 2022; Blumenstock et al., 2021; Mukherjee et al., 2023; Smythe & Blumenstock, 2022).

On a political and social level, the goal of targeted interventions and development policy generally is to reduce poverty. Thus, empirical machine learning applications are relevant to policymaking to the extent that they successfully approximate a target variable that captures aspects of poverty relevant to actual or potential policy scenarios. Poverty, however, is what Cartwright and coauthors (2022) have termed a Ballung Concept: "an agglomeration, dense, unruly, and context dependent: there is a lot packed into it; there is often no central core without which an item does not merit the label; different clusters of features from the congestion can matter for different uses" (p. 103; see also Bradburn et al., 2016; Cartwright & Runhardt, 2014). Simply put, there is no one universally correct notion of poverty. Rather we group under poverty a contextually sensitive set of often interrelated aspects — different features of poverty are relevant in different contexts for different purposes. For instance, being poor in Berlin is different than being poor in Sub-Saharan

[4] MPI stands for Multidimensional Poverty Index.

Africa and long-term policy planning requires a focus on different aspects of poverty than targeted short-term interventions (Alkire et al., 2015, pp. 188–192; Sen, 2009, Chapter 11).

Consequently, establishing the policy relevance of any given metric requires determining what conception of poverty is relevant in a particular context and which operationalization of poverty best captures salient aspects of it.[5] In international development practice, these conceptual and metrological issues are often addressed together in a manner that involves a considerable amount of ethical and epistemic normative engagement.[6]

International development experts design and employ a range of different poverty measurements for distinct purposes — from setting global long-term policy agendas to the targeting of short-term interventions focused on distinct geographic or demographic contexts.[7] Often, selecting any particular operationalization involves extensive stakeholder consultations with both political decision makers and affected populations in order to reflect in the choice of poverty metric the relevant societal understanding of poverty and policy priorities (e.g., Angulo, 2016; Angulo et al., 2016). Moreover, development economists often account for the tradeoffs and decisions made in measurement design and outline the properties and limitations of their data (e.g., Alkire et al., 2015). Sabina Alkire, the chief architect of the most popular multidimensional poverty indicator, stresses the evaluative nature of poverty metrics by noting that the purpose of measurement, the dimensions assessed, and whether poverty metrics are triangulated with qualitative information are all "normative questions that are inherent to the measurement of welfare concepts" (Alkire, 2012, p. 1).

Machine learning practitioners must not (and arguably should not) independently answer complex ethical and epistemic questions involved in determining the adequacy of a given measurement (and operationalization) for development policymaking. However, to the extent that machine learning practitioners seek to establish that their predictions provide value to international development efforts, they must nevertheless establish two claims. Developers must evidence that their models approximate a given target variable with sufficient accuracy. And, more centrally to this paper, that the target variable captures aspects of poverty that can be considered relevant to actual or potential policy scenarios. This does not require machine learning practitioners to *answer* complex normative questions. But it requires justifying the choice of a given target variable by establishing the link between any operationalization of poverty predicted and a conception of poverty relevant to the particular policy context.

III.3 Expedience and target variable choice

However, meaningful conceptual and metrological engagement with the target variable is generally absent in cases where machine learning practitioners conflate concept and operationalization. As outlined earlier, Blumenstock et al. (2015), for instance, predict a proxy of the DHS wealth index (DHS WI). Since its publication, the DHS WI has become the "workhorse" measurement of the field of machine learning poverty prediction (Hall, Ohlsson, & Rognvaldsson, 2022, p. 3). Yet its adoption is rarely justified in any substantial manner. Blumenstock and coauthors (2015), for instance, demonstrate no meaningful conceptual engagement

[5] This is arguably unless developers rely on an operationalization commonly employed in a given local and policy context (see next subsection).
[6] Recent work by Johanna Thoma (2024) provides an overview of some of these issues and raises further complications.
[7] For international comparisons, development experts often rely on the global poverty line or the Global MPI. Poverty metrics, such as national MPIs (*National MPI Directory*, 2024) or indices tailor made for vulnerable subgroups (e.g., Madrigal et al., 2023) can be used to monitor specific geographic or demographic contexts. Certain metrics such as asset wealth, for instance, are less useful when it comes to targeting in acute crises since they fail to reliably capture short-term changes in economic welfare.

with poverty and fail to outline the assumptions and limitations of the particular operationalization employed. While their publication was not tied to any immediate policy application, even a meaningful proof-of-concept requires some deliberation and justification establishing that what is predicted is at least of potential relevance to real-world development policy.

It might be argued that establishing the relevance of a given target variable does not require substantial conceptual and metrological engagement, when predicting a measurement widely employed in development policymaking. Critically, however, machine learning predictions of the DHS WI or a proxy of it are no such case. In contrast to the exceptional and continued prevalence of DHS WI as a target variable in empirical machine learning, it does not present a particularly common metric in development policymaking. Its comparatively rare use in international development might be in part due to the fact that the DHS WI lacks a number of properties desirable for policymaking and research. As a survey-specific measurement, the DHS WI lacks comparability across both time and countries (Rutstein & Johnson, 2004). Moreover, as the DHS WI is constructed through principal component analysis on a set of asset ownership variables, its precise composition is determined statistically rather than in the more principled and intentional manner of other poverty measurements (Alkire et al., 2015; Ravallion, 2016, Chapters 3–5).

In fact, the properties of the DHS WI even conflict with the value proposition of machine learning poverty predictions. Machine learning models trained on alternative data sources promise frequent poverty estimates that can supplement traditional survey statistics which commonly lie multiple years apart. At the same time, the DHS WI captures the possession of assets such as a particular type of housing, a television, and a car which are relatively unaffected by short-term changes in livelihood — asset wealth reflects primarily long-term socioeconomic status. As a result, the operationalization of poverty through the DHS WI is, to some extent, contraindicated for the purpose that machine learning poverty predictions are most strongly marketed for.[8]

Rather than by its prevalence in or relevance for policymaking, the DHS WI's popularity amongst machine learning applications to poverty estimation can best be explained by its convenience and expedience. Extracted from widespread Demographic and Health Surveys and based on variables more easily attainable during questioning than, for instance, consumption figures, the DHS WI is widely and openly available. Most critically however, the DHS WI is particularly well suited to be predicted by machine learning models trained on mobile phone and remote sensing data. In contrast to other indicators such as consumption or income, assets wealth is more highly correlated with features extracted from mobile communications or satellite imagery.[9] Empirical evidence supports that the DHS WI can be predicted with much higher accuracy and ease than other indicators of poverty (Jean et al., 2016; Steele et al., 2017). Granted, traditional poverty measurement efforts are similarly driven and constrained by data availability. What sets machine learning poverty prediction apart is not feasibility and convenience, but its determining force in the absence of (rather than in combination with) substantial conceptual and metrological reflection regarding its policy relevance.

[8] To my knowledge, a substantial discussion of this issue is still missing in the scholarship on machine learning poverty prediction.
[9] Uses of the DHS WI are not without their limitations. Even when the average predictive accuracy of models is high, research has shown that machine learning predictions of the DHS WI often systematically underestimate instances of extreme poverty most relevant to development aid (Ratledge et al., 2022) and might not reliably track short-term changes in livelihood outcomes if models are trained on satellite imagery (Kondmann & Zhu, 2020).

III.4 A strategic role for conflations of concept and operationalization

This results in a justificatory gap. For machine learning poverty predictions to serve their intended purpose, practitioners would have to establish how the estimated target variable such as the DHS WI captures aspects of poverty that can be considered relevant to actual or potential policy scenarios in the context covered. Unless machine learning models predict a metric already widely used in this policymaking context, this would involve a significant degree of conceptual and metrological engagement in order to elucidate the particular perspective on poverty the operationalization endorses, its underlying epistemic and ethical commitments, and potential limitations. The choice of target variable, however, can often be driven not by its policy prevalence or relevance but predominantly by technical affordances and convenience. Machine learning applications predicting convenient metrics such as the DHS WI can thus risk appearing disconnected from actual policy needs.

Here, conflations of concept and operationalization can serve a strategic function by helping portray the prima-facie policy relevance of machine learning applications driven primarily by expedience and technical affordances. By collapsing the distinction between concept and operationalization, machine learning applications are presented as predicting "poverty" and, thus, seemingly provide information relevant to the aims of international development policymaking. Policy efforts seek to reduce poverty and, in the context of the machine learning application, poverty refers simply to whatever the metric measures.

The case of emerging machine learning applications to development economics cautions us to consider the extent to which equations of concept and operationalization might be a feature rather than a bug.[10] To be clear, I am not suggesting that conflations of concept and operationalization in machine learning poverty predictions are the result of extensive strategic deliberation. Disciplinary naivety and carelessness might very well play into collapsing concept and operationalization. At the same time, I seek to challenge the idea that equations of poverty with a particular metric are entirely *reducible* to negligence and naivety. Instead, I argue that the strategic role described here can help us make sense of their prevalence and their persistence in areas of empirical machine learning such as in international development.

Two considerations support this claim. First, one must examine the relative occurrence of conceptual and operational conflation in empirical machine learning applications in international development. Equation of concept and operationalization in machine learning poverty prediction appear most commonly in expedient estimates of the DHS WI, arguably most at risk of appearing irrelevant to real-world policymaking. They are considerably less common in cases where practitioners predict established poverty metrics constructed in a more deliberate manner (e.g., Marcinko et al., 2022; Spandagos et al., 2023). This trend also appears in other domains such as learning analytics or engagement tracking applications, where conflations of concept and operationalization appear particularly dominant when models are trained to predict behavioral proxies with disputable relevance for policy and decision-making (cf. Knox et al., 2020a).[11]

Second, reducing equations of concept and operationalization entirely to a lack of disciplinary sophistication ascribes to machine learning practitioners a questionable degree of naivety and carelessness. Here it serves to be precise in what exactly is required to reduce conflations of concept and operationalization entirely to accident. It might well be correct that many machine learning practitioners publishing on poverty

[10] I thank a reviewer for suggesting this formulation.
[11] Granted, one might counter that the use of the DHS WI is itself caused by lacking conceptual sophistication. However, this then conflicts with the second consideration offered.

estimation lack extensive familiarity with the often rather nuanced debates surrounding the conception and measurement of poverty and alleviation strategies.[12]

Critically, however, a lack of conceptual and metrological sophistication alone is not sufficient to explain persistent equations of concept and operationalization in machine learning poverty prediction. This is because practitioners might equally refer to their models predicting the DHS WI (rather than predicting "poverty") whilst still avoiding complex conceptual and metrological debates by leaving open the question to what extent and in what ways their predictions are informative of poverty. Persistent conflations of concept and operationalization in certain cases of machine learning poverty prediction can, thus, only be explained in two ways. Either practitioners exhibit a degree of naivety and negligence that renders them ignorant, not of nuanced conceptual and metrological issues, but of the very distinction between concept and measurement. Or the persistence of conceptual and operational conflations must also, to some extent, be understood in light of its strategic role. It is the latter view that this paper seeks to highlight.

In this account, equations of concept and operationalization are not reducible to disciplinary naivety and not exclusively *caused by* lacking conceptual engagement. Given that machine learning practitioners seek to signal the intended relevance of their research for actual or potential policymaking, equations of poverty with operationalizations such as the DHS WI, *enable* practitioners to sidestep complex conceptual and metrological questions. Once practitioners collapse this distinction, questions regarding the conceptual validity or adequacy of the poverty measurement appear, in a sense, meaningless. Within the scope of the particular machine learning application, poverty refers to whatever the metric measures. Consequently, often all that is left to the evaluation and validation of machine learning poverty predictions is their predictive accuracy (c.f., Mussgnug, 2022; Wagstaff, 2012).

IV AI Operationalism

The previous section argued that conflations of concept and operationalization in machine learning applications to international development cannot be reduced to disciplinary naivety but instead serve a strategic function. This section proposes to understand such equations of concept and operationalization in empirical machine learning as a form of operationalism. Two interpretations of operationalism in the history and philosophy of social science are discussed: an orthodox view of operationalism as an ontological position and a more recent pragmatic account. I hold that equations of concept and operationalization in machine learning poverty prediction can be understood as a specific instance of the latter which I label *AI operationalism*. I do so primarily for two reasons. First, designating these conflations as a case of operationalism emphatically highlights their strategic role in parts of machine learning. While attributing conceptual and operational conflations only to disciplinary naivety and negligence leads scholars to immediately move to disciplinary reformation, the notion of AI operationalism stresses these conceptual practices as worthy of more substantial investigation in their own right. Second and relatedly, contextualizing these conflations in empirical machine learning within historical accounts of operationalism can help better understand both the systematic risks of AI operationalism, and the potential value of its limited and responsible use.

[12] The degree of metrological and conceptual involvement necessary for predictive machine learning applications is rather different than for designing or measuring poverty metrics. Practitioners merely choose a target metric and feature variables serve an often purely instrumental function. This position is comparable with international development experts choosing already available data for research or policy purposes. However, a greater degree of metrological and theoretical involvement might be called for in many instances of machine learning poverty prediction to better ascertain robustness and anticipate model drift.

IV.1 Candidate 1: Ontological operationalism

Operationalism is most closely associated with the work of the physicist Percy W. Bridgman, who coined the term in his 1927 book *The Logic of Modern Physics*. There he advocated that "in general, we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations" (1927, p. 5). Regarding the concept of length, for instance, Bridgman argued that, in principle, "the operations by which length is measured should be uniquely specified. If we have more than one set of operations, we have more than one concept, and strictly there should be a separate name to correspond to each different set of operations" (p. 10). According to Bridgman, it is only for convenience that we might keep the same concept for different operations. Such convenience, however, comes at the expense of conceptual clarity (p. 23).[13][14]

Bridgman's operational agenda emerged out of his work as an experimental physicist and primarily criticized conceptual complacency in physics rather than providing a solution to the problems of philosophy (see also Chang, 2004, Chapter 3, 2021). It was as the latter, however, that Bridgman's position was initially interpreted by philosophers of science, particularly the logical empiricists (Feigl, 1969). For logical empiricists, scientific theories consisted of logical, empirical, and theoretic parts. Some early logical empiricists argued that for theoretical terms to be meaningful, i.e., possess cognitive significance, they must be reducible to observational vocabulary through correspondence rules and, thereby, empirically verifiable. As Frederick Suppe (1974, p. 12) notes, "initially correspondence rules had to have the form of explicit definitions which provide necessary and sufficient observational conditions for the applicability of theoretical terms; theoretical terms were cognitively significant if and only if they were explicitly defined in terms of the observation vocabulary." Confronted with Bridgman's operationalism, logical empiricists viewed operational definitions as one possible formulation of the correspondence rules providing the necessary and sufficient conditions for the meaningfulness of theoretical terms.

Enthusiasm regarding the marriage of logical positivism and operationalism, however, was rather short lived. On the one hand, continued criticism revealed the initially strict verificationism of some logical empiricists to be untenable. If correspondence rules were interpreted to provide necessary and sufficient conditions for cognitive significance, large parts of scientific practice must be considered meaningless. As a consequence, logical empiricists soon "mellowed and matured" their initially radical views (Feigl, 1969, p. 631). On the other hand, Bridgman strongly opposed the logical empiricists' adoption of his thesis, considering their attempt at reconstructing scientific language to be farfetched and artificial (Feigl, 1969, p. 662).[15] Bridgman (1956, pp. 74–75) complained: "I have only a historical connection with this thing called 'operationalism.' In short, I feel that I have created a Frankenstein, which has certainly got away from me. I abhor the word operationalism or operationism, which seems to imply a dogma, or at least a thesis of some kind. The thing I have envisaged is too simple to be dignified by so pretentious a name."

[13] Bridgman wasn't the first to advocate that the concepts of physics are best uniquely identified by their operationalization (Moyer, 1991, p. 244). Beyond physics, Bridgman's operationalism echoed the views of empiricist and pragmatist philosophers such as C. S. Peirce (1878/1995) and social scientists, such as J.B. Watson's behaviorism (Watson, 1913; see also 'Operationalism', 1998, p. 347). Nonetheless, Bridgman's eminence as a leading physicist of his time both lent to the appeal of and shaped his operationalist thesis.

[14] Curiously, what Bridgman initially conceived as a cautionary measure against conceptual complacency, in the case of machine learning poverty prediction, helps facilitate the avoidance of critical conceptual and metrological debate.

[15] Bridgman's views were also challenged, for instance, because a strict application of operationalism would lead to an inefficient proliferation of concepts, leaving scientific practice hopelessly fragmented (Russell, 1928). In face of these criticisms, Bridgman (quoted in Chang, 2021) also matured in his position later maintaining that reducing the meaning of concepts to their operation was "obviously going too far when taken out of context."

Regardless, it is this "Frankenstein" which found its way not only into the philosophical canon but also into social scientific practice. Whereas Bridgman's primary concern was the field of physics, operationalism became most impactful in the then-emerging field of experimental psychology. How exactly and in what form operationalism came to be incorporated within psychology is a topic of considerable debate (Feest, 2005; Green, 1992; Leahey, 1980; Michell, 1999). Often, however, operationalist orthodoxy within psychology is presented as the full-fledged ontological position that all there is to the meaning of a concept is its operationalization (Feest, 2005; Leahey, 1983; Michell, 1999, pp. 170–175; Vessonen, 2019). Some even go as far as describing operationalism as the "prototypical anti-realist view." According to Lovett & Hood (2011, p. 208), "[…] the operationist denies ontological realism. Moreover, the operationist will deny that we can attain knowledge about disorders qua theoretical entities, thereby denying epistemic realism as well."(ibid, p. 209).

IV.2 Machine learning poverty prediction and ontological operationalism

This account of operationalism resonates with a small number of authors who have linked conflations of operationalization and concept in machine learning to the philosophical strand of positivism. The claimed strengths of this association vary. While some point toward a continuing legacy of positivism in AI development (Jones, 2018; Williams, 2024), others endorse the stronger claim that empirical machine learning research is dominated by a positivist philosophy. On the arguably most prominent discussion platform of data science professionals, practitioners claim that "the prevailing paradigm of data analysis is buried deep in the arid soil of a philosophical school called Logical Positivism" (Keith, 2021). In conference publications, scholars identify data science as an "extremely positivistic discipline" (Luczak-Rösch, 2013, p. 4). A similar point is made by John McCarthy, who is among the founding figures of artificial intelligence. McCarthy (2007, p. 2) argues that "the mistaken philosophy of taking the world (or particular phenomena) as a structure of sense data has been harmful in artificial intelligence and machine learning research, just as behaviorism and logical positivism harmed psychology."

Some positivist legacy might well contribute to equations of concept and operationalization. Moreover, these conflations are indeed compatible with or might themselves play into positivist tendencies.[16] At the same time, identifying conflations of concept and operationalization in the case of machine learning poverty predictions as an instance of operationalism characterized by a thoroughgoing positivist stance is misguided. Such would require denying that there is anything more to the concept of poverty than any particular operationalization. Not only does one find no evidence for such a stance in the literature on machine learning poverty prediction, but it would also commit machine learning practitioners to a questionable and radical commonsense antirealism.

Poverty has long featured in our commonsense conception of the world. We find treatments of poverty, for instance, in the works of Aristotle (ca. 330 BCE/1944, section 1320a) and Confucius (Analects, 16.1, as translated in Eno, 2015). Moreover, the concept of poverty appears regularly in our ordinary everyday reasoning, for instance, as a consideration in political deliberations. Denying the existence of poverty beyond empirical observation, thus, amounts to a rather radical commonsense antirealism. Hausman (1998, p. 197) echoes this point with respect to economic phenomena more broadly:

[16] I am also leaving open to what extent positivism (and related stances such as behaviorism) might serve a more substantial explanatory role in other areas of applied machine learning such as learning analytics (Knox et al., 2020).

"Anti-realists seek to draw a line between the relatively unproblematic claims of everyday life and the problematic theoretical posits of science. Physics postulates new unobservables, to whose existence commonsense realism does not commit us. Although economics refers to unobservables, it does not, in contrast to physics, postulate new ones. Its unobservables - beliefs, preferences, and the like - are venerable. They have been a part of commonsense understanding of the world for millennia."

Such, commonsense antirealism is a position hardly attributable even to economic theorists. This is illustrated by the fact that despite early efforts from Henry Schultz (1928) and Paul Samuelson (1938), such a form of operationalism never gained much traction in economics (Hands, 2004; 'Operationalism', 1998). Moreover, one will be hard-pressed to find development economists or machine learning practitioners questioning the reality of poverty. The lack of evidence for it and the exceptionally far-reaching commitments it would, cast serious doubt on attributing conflations of concept and operationalization to an ontologically motivated operationalism in machine learning poverty prediction.

IV.3 Candidate two: Methodological Operationalism

More recently, Uljana Feest (2005) has questioned this orthodox view of operationalism.[17] Feest puts forth an interpretation of operationalism as practiced by early operationalist psychologists Stanley S. Stevens and Edward C. Tolman, rather than one defined by its association with logical positivism. According to her (p. 143), references to logical positivism in Stevens' and Tolman's writing "were largely rhetorical, aimed at backing up their views by appealing to cutting-edge philosophy of science." Their investigative practices, however, reveal that operational definitions were neither taken as a priori knowledge nor as analytically true — operationalism practiced by early psychologists was never aimed at defining the meaning of concepts.

Feest outlines how Stevens and Tolman were motivated to ground psychological research on experimental methodology rather than introspection. In order to do so, they had to get an empirical handle on so far obscure psychological phenomena. Operational definitions served as criteria for the application of concepts, *temporarily fixing* the referent of terms such as depression by specifying "what kinds of operations were to count as empirical indicators for the referents of their concepts" (p. 131). Rather than ontological, operational definitions were pragmatic and methodological in nature, enabling scientists to do research on particular kinds of psychological phenomena and thereby serving a strategic role in bootstrapping empirically grounded concept development.

Feest's "methodological operationalism" provides a convincing account of operational definitions in early experimental psychology. At the same time, its focus on experimental practice risks leaving out of sight how operationalism might have also proven opportune to (and, to some extent, might have been motivated by) signaling the policy relevance of early research in experimental psychology. Consider for instance, operationalist tendencies in early intelligence research. Understanding the use of operational definitions in intelligence research as a purely epistemic choice within an isolated process of inquiry risks painting a narrow and incomplete picture. Intelligence research obtained its funding and relevance at least partly due to its rather immediate application in educational interventions and for administrative purposes in both mental health and military settings (Carson, 2007). I take it to be no coincidence that operationalist frontman Stanley Stevens situates Edwin Boring's famous dictum "Intelligence is what the tests test" (Boring, 1923, p. 35) precisely within the context of his wartime contributions to the rollout of a mental testing program in the US Armed Forces (Stevens, 1973, pp. 45–46). Such evidence does not undermine Feest's characterization of

[17] Feest uses the synonymous term "operationism" (more common in the theory of psychology) rather than "operationalism" (more prevalent in philosophy).

operational definitions in early experimental psychology as pragmatic and strategic tools. However, it calls attention not only to their function in bootstrapping empirically grounded concept development but also to the ways in which operationalism was opportune in signaling the policy relevance of this emerging research agenda.

IV.4 Empirical machine learning and methodological operationalism

Acknowledging this second function of operationalism, ties historical cases of operationalism in the social sciences to operational and conceptual conflations in empirical machine learning. Similar to the historical case of experimental psychology, the previous section has outlined how conflations of concept and operationalization can signal the policy relevance of empirical machine learning in international development. Thus, I propose to interpret operational and conceptual conflation in machine learning poverty prediction in ways broadly aligned with the pragmatic account of operationalism in the social sciences as an instance of *AI operationalism*. AI operationalism captures how equations of concept and operationalization in empirical machine learning temporarily fix the referent of concepts in ways that serve a strategic yet pragmatic function. This function might lie in portraying the prima facie relevance of empirical results for potential policy applications or in bootstrapping empirical concept development.

Identifying conflations of concept and operationalization in empirical machine learning applications, such as in international development, as a case of operationalism might cause some objection. Differences clearly exist between empirical machine learning and paradigmatic cases of operationalism discussed in the history and philosophy of science. Most critically, scholars have commonly been interested in instances where researchers self-identify as "operationalist" and explicitly and deliberatively employ operational definitions. Machine learning practitioners do not self-identify as operationalists and equations of concept and operationalization likely lack the same degree of intention and deliberation. My claim, however, is not that machine learning practitioners are operationalists but that equations of concept and operationalization in empirical machine learning can be understood as a somewhat distinctive instance of operationalism.[18]

Designating equations of concept and operationalization in machine learning as an instance of operationalism serves two purposes. First, it emphatically highlights their strategic role in parts of empirical machine learning. It shifts attention to the ways in which questionable conceptual practices might be considered a feature rather than a bug. Attributing conceptual and operational conflations only to disciplinary naivety and negligence leads scholars to immediately move to disciplinary reformation. In contrast labeling such practices as an instance of AI operationalism can help identify them as worthy of more substantial investigation in their own right. Such work also involves being open to the possibility that instances of conceptual and operational collapse might not fit the category of AI operationalism outlined here. However, precisely this kind of investigation into their nature and function might lead us to arrive at a more nuanced picture of their implications. And this, in turn, can help current efforts more successfully target interventions that aim to establish greater conceptual sophistication in empirical machine learning practice.

[18] I believe this is warranted especially when a pragmatic rather than ontological reading of operationalism is endorsed and emphasis is put on the effects of operationalism. Nonetheless, some might worry that my notion of AI operationalism might water down the category and wish to reserve the label of operationalism to denote an explicit and intentional methodological or ontological position. Therefore, it is important to stress that none of what has been argued for so far, including the strategic function of conceptual and operational conflations depends on this semantic disagreement. And much of the following attempt to link these practices in empirical machine learning to historical research on the effects of operationalism in the social sciences bears relevance independent of whether one would like to adopt AI operationalism as a label.

Second, it appropriately contextualizes relevant conflations of concept and operationalization in empirical machine learning and its effects within historical instances of operationalism in the social sciences. Existing research focuses almost exclusively on the somewhat idiosyncratic fairness-related harms associated with the failure to effectively distinguish between concept and operationalization. Situating prevalent equations of concept and operationalization within the long-standing history of operationalism in the social sciences, however, can help underscore their broader and systematic implications — both when it comes to the problematic consequences of AI operationalism and the potential value of its limited and responsible use.

Existing historical scholarship provides an extensive and illuminating perspective on the complex ethical and epistemic implications of operationalism in social scientific research. The most prominent accounts examine the questionable role that operationalist tendencies played in early intelligence research and policymaking (e.g., Gould, 1996; Malabou, 2019). However, historical research also exists on the role of operationalist framings and rhetoric in, for instance, international development (e.g., Macekura, 2020).

A particularly critical observation of this scholarship is of how prevalent operationalism and operational rhetoric can systematically reduce the space for normative epistemic and ethical engagement. In the case of intelligence research, this can be most easily illustrated using the infamous example of Edwin Boring. When Boring (1923, p. 35) claims that "intelligence is what the tests test," he not only collapses the distinction between concept and operationalization, but also the space for critical normative engagement. To some extent, he exempts his work from ethical and epistemic questions of how we should conceptualize intelligence (for a given purpose) and to what extent any given operationalization might capture this understanding of intelligence. Once concept and operationalization are equated, questions regarding the various value-laden conceptualizations of intelligence, and the validity or adequacy of the respective intelligence test appear, in a sense, meaningless.

International development illustrates this relationship between operational rhetoric and normative engagement forcefully. Development economists have become increasingly invested in and able to explore the normative dimensions of poverty and poverty measurement, precisely as scholars such as Amartya Sen (1979, 1999) broke down the common equation of poverty and economic progress with the monetary poverty line and the Gross Domestic Product (GDP) (Macekura, 2020). Challenging equations of poverty and development with monetary poverty and the GDP has opened up greater space for ethical and epistemic scrutiny. International development practitioners have begun to embrace to a much greater extent different conceptions of poverty and investigate how value judgments inherent in the design of any chosen measurement shape its inevitably partial perspective on the complex concept. Often, such work also involves an increased emphasis on participatory and qualitative research. This can serve to explore facets of economic deprivation that are largely left out of common operationalizations of poverty and involves a more diverse set of voices in poverty research and policymaking, rather than paternalistically imposing a certain understanding of deprivation on affected communities (e.g., Narayan et al., 2000).

Much like these historical cases, prevalent AI operationalism might risk systematically reducing the space for critical normative engagement in certain areas of inquiry, such as international development. It does so partly by obscuring the value judgments inherent in a given target variable. This relates to the earlier claim that equations of concept and operationalization are not only *caused by* lacking conceptual and metrological engagement but also *facilitate* it. This lack of meaningful elucidation of the conceptual and metrological dimensions of machine learning poverty prediction conceals normative choices involved in the measurement of poverty. The case of AI operationalism is arguably particularly problematic since the prediction of existing poverty measurements by machine learning models adds an additional layer of methodological complexity, i.e., the lack of transparency pertains to the value judgments that underlie the target variable that underlies the machine learning model's predictions.[19] Affected communities, researchers, and policymakers are not only commonly unable to scrutinize the often opaque workings of complex machine learning models (e.g., Boge, 2022) but claims of machine learning models predicting "poverty" obscure even the concrete operationalization that the model seeks to predict (and its underlying normative dimensions). This might further add to a false sense of objectivity and value-neutrality in empirical machine learning (Talat et al., 2021) that can undermine recent efforts to facilitate greater and broader consideration of the ethical and epistemic dimensions of poverty measurement and policymaking.

In some ways, AI operationalism not only obscures the normative dimensions of machine learning poverty prediction but also closes them off from critical scrutiny by development experts and affected populations.[20] Presenting a given measurement as one possible way of empirically assessing aspects of poverty, leaves space to question the appropriateness and fit of operationalization and concept. AI operationalism, however, reduces poverty, even if just locally and pragmatically, to a given operationalization and leaves no such space for critical engagement. One might reject the equation of poverty with a particular operationalization. But if poverty is, for a given application, simply equated with the operationalization predicted such does not bear on the application since the critic then simply speaks of a different concept to begin with (cf. the example of Boring discussed earlier). Appreciating this systemic risk of widespread AI operationalism in reducing the space for normative engagement by obscuring value judgments and closing them off from critical scrutiny, urges us to look beyond idiosyncratic fairness-related harms arising from the conflation of concept and operationalization. Instead, it reveals how equations of concept and operationalization bear on current work investigating the relationship between AI and epistemic justice (Miragoli, 2024; Pozzi, 2023; Symons & Alvarado, 2022), value capture (Nguyen, 2024), and power dynamics more generally (Mohamed et al., 2020).

Positioning equations of concept and operationalization in empirical machine learning within the historical context of operationalism also points toward the potentially value of their responsible, sophisticated, and limited use. Machine learning and alternative data sources give us the ability to empirically investigate novel kinds of phenomena and existing phenomena from different perspectives and levels of granularity. In a similar way as detailed by Feest (2005) in her historical account of early experimental psychology, operational definitions in applied machine learning might serve a valuable role in bootstrapping concept development. While this might be less applicable to empirical investigations of a comparatively established concept such as poverty, operational definitions might prove beneficial to empirical machine learning in more nascent areas of research.[21]

As Herrmann et al. (2024) argue, this would involve more explicitly embracing the exploratory dimensions of empirical machine learning, as well as moving focus away from the chasing of predictive accuracy through supervised machine learning and toward the use of other learning methodologies and explainable AI methods. These can help probe and map out potential relationships (or "nomological web") between various operationally defined concepts. In these cases, a deliberative and sophisticated AI

[19]Or, in some cases, even the prediction of a proxy of a poverty measurement (e.g., Blumenstock et al., 2015)

[20]In a more sophisticated operationalism, substantive normative engagement can be involved in justifying the initial equation of poverty with a particular operationalization. But such is not commonly involved in the cases of empirical machine learning discussed here, and AI operationalism precisely obscures this lacking normative engagement.

21 One concrete example where a responsible AI operationalism might be fruitful are in potential machine learning applications exploring the concept of purity in moral psychology (Gray et al., 2023).

operationalism might be welcomed, especially if it can avoid some of the problematic implications discussed by using operational definitions responsibly in limited, explicit, and preliminary fashion (cf. Vessonen, 2021).

V Conclusion

Existing scholarship reduces equations of concept and operationalization in empirical machine learning to disciplinary naivety or negligence and argues for the integration of richer methodological resources. In contrast, this paper has argued that, in the case of empirical machine learning in international development, conceptual and operational conflation serves a strategic function. Equating the concept of poverty with a particular operationalization can help signal the prima facie policy relevance of empirical machine learning driven primarily by convenience and technical affordances. Policy efforts seek to reduce poverty and, in the context of the machine learning application, poverty refers simply to whatever the metric measures.

In response, I have proposed to understand some equations of concept and operationalization in empirical machine learning as an instance of *AI operationalism*. Characterizing these conceptual practices as an instance of operationalism stresses them as worthy of more substantial investigation in their own right. Moreover, contextualizing conflations of operationalization and concept in empirical machine learning within historical accounts of operationalism can help better understand its complex systematic implications — both when it comes to the risks of AI operationalism and the potential value of its responsible use.

Connecting historical and philosophical scholarship on the epistemology of science to a contemporary debate in the methodology and ethics of AI, this paper sought to contribute to better understanding a distinctive conceptual practice in social and human science machine learning. The first steps taken here can open avenues for more research on the implications of AI operationalism and its applicability beyond the case of international development. In particular, further work is needed on areas where non-supervised learning methodologies and model interpretability efforts are more common than in machine learning poverty prediction (Hall, Ohlsson, & Rögnvaldsson, 2022). Investigations of these fields can reveal to what extent AI operationalism is relevant to researchers' conceptual practices not only with respect to a chosen target variable but also with respect to features emerging within the inner structure of deep learning models.

VI References

Aiken, E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, *603*(7903), Article 7903. https://doi.org/10.1038/s41586-022-04484-9

Al Kez, D., Foley, A., Abdul, Z. K., & Del Rio, D. F. (2024). Energy poverty prediction in the United Kingdom: A machine learning approach. *Energy Policy*, *184*, 113909. https://doi.org/10.1016/j.enpol.2023.113909

Alkire, S. (2012). *Value Judgements in Multidimensional Poverty Indices*. https://www.ophi.org.uk/wp-content/uploads/LiteratureReviewValueJudgements.pdf

Alkire, S., Roche, J. M., Ballon, P., Foster, J., Santos, M. E., & Seth, S. (2015). *Multidimensional Poverty Measurement and Analysis*. Oxford University Press.

Alruwais, N., & Zakariah, M. (2023). Student-Engagement Detection in Classroom Using Machine Learning Algorithm. *Electronics*, *12*(3), Article 3. https://doi.org/10.3390/electronics12030731

Angulo, R. (2016). From Multidimensional Poverty Measurement to Multisector Public Policy for Poverty Reduction: Lessons from the Colombian Case. *OPHI Working Papers*, 27.

Angulo, R., Díaz, Y., & Pardo, R. (2016). The Colombian Multidimensional Poverty Index: Measuring Poverty in a Public Policy Context. *Social Indicators Research*, *127*(1), 1–38. https://doi.org/10.1007/s11205-015-0964-z

Aristotle. (1944). *Politics, Book 6*. https://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0058:book=6:section=1320a&highlight=poverty

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, *350*(6264), Article 6264. https://doi.org/10.1126/science.aac4420

Blumenstock, J., Lain, J., Smythe, I., & Vishwanath, T. (2021). *Using Big Data and machine learning to locate the poor in Nigeria*. https://blogs.worldbank.org/opendata/using-big-data-and-machine-learning-locate-poor-nigeria

Boge, F. J. (2022). Two Dimensions of Opacity and the Deep Learning Predicament. *Minds and Machines*, *32*(1), 43–75. https://doi.org/10.1007/s11023-021-09569-4

Boring, E. G. (1923). Intelligence as Tests Test It. *The New Republic*, 35–37.

Bosch, N. (2016). Detecting Student Engagement: Human Versus Machine. *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, 317–320. https://doi.org/10.1145/2930238.2930371

Bradburn, N. M., Cartwright, N., & Fuller, J. (2016). *A Theory of Measurement*. https://www.durham.ac.uk/media/durham-university/research-/research-centres/humanities-engaging-sci-and-soc-centre-for/CHESSK4UWP_2016_07_BradburnCartwrightFuller.pdf

Bridgman, P. W. (1927). *The logic of modern physics*. The Macmillan Company.

Bridgman, P. W. (1956). The Present State of Operationalism. In P. G. Frank (Ed.), *The Validation of Scientific Theories*. Boston: Beacon Press.

Brown, C., Ravallion, M., & van de Walle, D. (2016). *A Poor Means Test?: Econometric Targeting in Africa* [Working Paper]. World Bank. https://doi.org/10.1596/1813-9450-7915

Carson, J. (2007). American Psychology and the Seductions of IQ. In *The Measure of Merit* (pp. 159–194). Princeton University Press; JSTOR. https://doi.org/10.2307/j.ctv346qgf.10

Cartwright, N., Hardie, J., Montuschi, E., Soleiman, M., Thresher, A. C., Cartwright, N., Hardie, J., Montuschi, E., Soleiman, M., & Thresher, A. C. (2022). *The Tangle of Science: Reliability Beyond Method, Rigour, and Objectivity*. Oxford University Press.

Cartwright, N., & Runhardt, R. (2014). Measurement. In N. Cartwright & E. Montuschi (Eds.), *Philosophy of Social Science: A New Introduction*. Oxford University Press.

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress: Measurement and Scientific Progress*. Oxford University Press.

Chang, H. (2021). Operationalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/operationalism/

Eno, R. (2015). *The Analects of Confucius: An Online Teaching Translation.* https://tenthousandrooms.yale.edu/project/robert-enos-translations-analects-and-mencius

Feest, U. (2005). Operationism in psychology: What the debate is about, what the debate should be about. *Journal of the History of the Behavioral Sciences*, *41*(2), 131–149. https://doi.org/10.1002/jhbs.20079

Feigl, H. (1969). The Wiener Kreis in America. In D. Fleming & B. Bailyn (Eds.), *The Intellectual Migration: Europe and America, 1930–1960*. Belknap Press.

Gould, S. J. (1996). *The mismeasure of man* (2nd ed.). Norton.

Gray, K., DiMaggio, N., Schein, C., & Kachanoff, F. (2023). The Problem of Purity in Moral Psychology. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *27*(3), 272–308. https://doi.org/10.1177/10888683221124741

Green, C. D. (1992). Of Immortal Mythological Beasts: Operationism in Psychology. *Theory & Psychology*, *2*(3), 291–320. https://doi.org/10.1177/0959354392023003

Guerdan, L., Holstein, K., Steven, Z., & Wu, S. (2022). *Under-reliance or misalignment? How proxy outcomes limit measurement of appropriate reliance in AI-assisted decision-making.* https://www.semanticscholar.org/paper/Under-reliance-or-misalignment-How-proxy-outcomes-Guerdan-Holstein/317a02a572a450af28352869ffae3d8df0104abd

Hall, O., Ohlsson, M., & Rögnvaldsson, T. (2022). A review of explainable AI in the satellite data, deep machine learning, and human poverty domain. *Patterns*, *3*(10), 100600. https://doi.org/10.1016/j.patter.2022.100600

Hall, O., Ohlsson, M., & Rognvaldsson, T. (2022). Satellite Image and Machine Learning based Knowledge Extraction in the Poverty and Welfare Domain. *Papers*, Article 2203.01068. https://ideas.repec.org//p/arx/papers/2203.01068.html

Hands, D. W. (2004). On Operationalisms and Economics. *Journal of Economic Issues*, *38*(4), 953–968.

Hausman, D. M. (1998). Problems with Realism in Economics. *Economics & Philosophy*, *14*(2), 185–213. https://doi.org/10.1017/S0266267100003837

Herrmann, M., Lange, F. J. D., Eggensperger, K., Casalicchio, G., Wever, M., Feurer, M., Rügamer, D., Hüllermeier, E., Boulesteix, A.-L., & Bischl, B. (2024). *Position: Why We Must Rethink Empirical Research in Machine Learning* (No. arXiv:2405.02200). arXiv. https://doi.org/10.48550/arXiv.2405.02200

Jacobs, A. Z., & Wallach, H. (2021). Measurement and Fairness. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 375–385. https://doi.org/10.1145/3442188.3445901

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), Article 6301. https://doi.org/10.1126/science.aaf7894

Jones, M. L. (2018). How We Became Instrumentalists (Again): Data Positivism since World War II. *Historical Studies in the Natural Sciences*, *48*(5), 673–684. https://doi.org/10.1525/hsns.2018.48.5.673

Keith, G. (2021). The Dark Side of Data Science. *Medium*. https://towardsdatascience.com/the-dark-side-of-data-science-30b554bcf946

Kim, J. Y. (2021). *Using Machine Learning to Predict Poverty Status in Costa Rican Households* (No. arXiv:2111.13319). arXiv. https://doi.org/10.48550/arXiv.2111.13319

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human Decisions and Machine Predictions*. *The Quarterly Journal of Economics*, *133*(1), 237–293. https://doi.org/10.1093/qje/qjx032

Knox, J., Williamson, B., & Bayne, S. (2020). Machine behaviourism: Future visions of 'learnification' and 'datafication' across humans and digital technologies. *Learning, Media and Technology*, *45*(1), 31–45. https://doi.org/10.1080/17439884.2019.1623251

Kondmann, L., & Zhu, X. (2020, January 1). *Measuring Changes in Poverty with Deep Learning and Satellite Imagery*.

Leahey, T. H. (1980). The Myth of Operationism. *The Journal of Mind and Behavior*, *1*(2), 127–143.

Leahey, T. H. (1983). Operationism and Ideology: Reply to Kendler. *The Journal of Mind and Behavior*, *4*(1), 81–90.

Lovett, B. J., & Hood, S. B. (2011). Realism and operationism in psychiatric diagnosis. *Philosophical Psychology*, *24*(2), 207–222. https://doi.org/10.1080/09515089.2011.558498

Luczak-Rösch, M. (2013). *Beyond Positivism in Computer Science*.

Macekura, S. J. (2020). *The Mismeasure of Progress: Economic Growth and Its Critics*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/M/bo60921587.html

Madrigal, M., Álvarez, M., de los Santos, D., & Pardo, R. (2023). *Multidimensional Poverty Index with a Focus on Women: A proposal for Latin America and the Caribbean: Development Futures Series No. 56*. https://www.un-ilibrary.org/content/papers/10.18356/30053307-56

Malabou, C. (2019). *Morphing intelligence: From IQ measurement to artificial brains* (C. Shread, Trans.). Columbia University Press.

Malik, M. M. (2020). *A Hierarchy of Limitations in Machine Learning* (No. arXiv:2002.05193). arXiv. https://doi.org/10.48550/arXiv.2002.05193

Marcinko, C. L. J., Samanta, S., Basu, O., Harfoot, A., Hornby, D. D., Hutton, C. W., Pal, S., & Watmough, G. R. (2022). Earth observation and geospatial data can predict the relative distribution of village level poverty in the Sundarban Biosphere Reserve, India. *Journal of Environmental Management*, *313*, 114950. https://doi.org/10.1016/j.jenvman.2022.114950

McBride, L., Barrett, C. B., Browne, C., Hu, L., Liu, Y., Matteson, D. S., Sun, Y., & Wen, J. (2022). Predicting poverty and malnutrition for targeting, mapping, monitoring, and early warning. *Applied Economic Perspectives and Policy*, *44*(2), 879–892. https://doi.org/10.1002/aepp.13175

McCarthy, J. (2007). Challenges to Machine Learning: Relations Between Reality and Appearance. In S. Muggleton, R. Otero, & A. Tamaddoni-Nezhad (Eds.), *Inductive Logic Programming* (pp. 2–9). Springer. https://doi.org/10.1007/978-3-540-73847-3_2

Michell, J. (1999). *Measurement in Psychology: A Critical History of a Methodological Concept*. Cambridge University Press.

Miragoli, M. (2024). Conformism, Ignorance & Injustice: Ai as a Tool of Epistemic Oppression. *Episteme: A Journal of Social Epistemology*, 1–19.

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, *33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Moyer, A. E. (1991). P.W. Bridgman's operational perspective on physics Part I: Origins and development. *Studies in History and Philosophy of Science Part A*, *22*(2), 237–258. https://doi.org/10.1016/0039-3681(91)90026-O

Mukherjee, A. N., Bermeo Rojas, L. X., Okamura, Y., Vulembera, J., & Bance, P. G. A. (2023). Une Approche de Priorité au Numérique pour les Transferts Monétaires d'Urgence:Programme STEP-KIN en République Démocratique du Congo. *Social Protection Discussion Papers and Notes*, Article 181798. https://ideas.repec.org//p/wbk/hdnspu/181798.html

Mussgnug, A. M. (2022). The predictive reframing of machine learning applications: Good predictions and bad measurements. *European Journal for Philosophy of Science*, *12*(3), 55. https://doi.org/10.1007/s13194-022-00484-8

Myrtveit, I., Stensrud, E., & Shepperd, M. (2005). Reliability and validity in comparative studies of software prediction models. *IEEE Transactions on Software Engineering*, *31*(5), 380–391. IEEE Transactions on Software Engineering. https://doi.org/10.1109/TSE.2005.58

Narayan, D., Chambers, R., Shah, M., & Petesch, P. (2000). *Voices of the Poor: Crying Out for Change* (Vol. 2).

Nguyen, C. (2024). Value Capture. *Journal of Ethics and Social Philosophy*, *27*(3). https://doi.org/10.26556/jesp.v27i3.3048

Ni, Y., Li, X., Ye, Y., Li, Y., Li, C., & Chu, D. (2021). An Investigation on Deep Learning Approaches to Combining Nighttime and Daytime Satellite Imagery for Poverty Prediction. *IEEE Geoscience and Remote Sensing Letters*, *18*(9), 1545–1549. IEEE Geoscience and Remote Sensing Letters. https://doi.org/10.1109/LGRS.2020.3006019

Operationalism. (1998). In *The Handbook of Economic Methodology*. Edward Elgar Publishing. https://www.elgaronline.com/display/9781852787950.00087.xml

Peirce, C. S. (1995). How to Make our Ideas Clear. In *Pragmatism*. Routledge. (Original work published 1878)

Pennacchiotti, M., & Popescu, A.-M. (2011). A Machine Learning Approach to Twitter User Classification. *Proceedings of the International AAAI Conference on Web and Social Media*, *5*(1), Article 1. https://doi.org/10.1609/icwsm.v5i1.14139

Pokhriyal, N., & Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences*, *114*(46), Article 46. https://doi.org/10.1073/pnas.1700319114

Pozzi, G. (2023). Automated opioid risk scores: A case for machine learning-induced epistemic injustice in healthcare. *Ethics and Information Technology*, *25*(1), 3. https://doi.org/10.1007/s10676-023-09676-z

Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. *Procedia Computer Science*, *167*, 1258–1267. https://doi.org/10.1016/j.procs.2020.03.442

Radford, J., & Joseph, K. (2020). *Theory In, Theory Out: The uses of social theory in machine learning for social science* (No. arXiv:2001.03203). arXiv. https://doi.org/10.48550/arXiv.2001.03203

Ravallion, M. (2016). *The Economics of Poverty: History, Measurement, and Policy*. Oxford University Press.

Russell, L. J. (1928). [Review of *Review of The Logic of Modern Physics.*, by P. W. Bridgman]. *Mind*, *37*(147), 355–361.

Rutstein, S., & Johnson, K. (2004). *The DHS Wealth Index*. https://doi.org/10.13140/2.1.2806.4809

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61–71.

Sau, A., & Bhakta, I. (2017). Predicting anxiety and depression in elderly patients using machine learning technology. *Healthcare Technology Letters*, *4*(6), 238–243. https://doi.org/10.1049/htl.2016.0096

Schultz, H. (1928). Rational Economics. *The American Economic Review*, *18*(4), 643–648.

Sen, A. (1979). Equality of What? In S. M. McMurrin (Ed.), *Tanner Lectures in Human Values* (pp. 197–220). Cambridge University Press. https://doi.org/10.1017/CBO9780511570742

Sen, A. (1999). *Development as Freedom*. Oxford University Press.

Sen, A. (2009). *The Idea of Justice*. Penguin.

Smythe, I. S., & Blumenstock, J. E. (2022). Geographic microtargeting of social assistance with high-resolution poverty maps. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(32), e2120025119. https://doi.org/10.1073/pnas.2120025119

Spandagos, C., Reaños, M. T., & Lynch, M. Á. (2023). *Energy poverty prediction and effective targeting for just transitions with machine learning* [Working Paper]. ESRI. https://esri.ie/publications/energy-poverty-prediction-and-effective-targeting-for-just-transitions-with-machine

Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.-A., Iqbal, A. M., Hadiuzzaman, K. N., Lu, X., Wetter, E., Tatem, A. J., & Bengtsson, L. (2017). Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society, Interface*, *14*(127), 20160690. https://doi.org/10.1098/rsif.2016.0690

Stevens, S. S. (1973). Edwin Garrigues Boring. *Biographical Memoirs. National Academy of Sciences (U.S.)*, *43*, 41–76.

Suppe, F. (1974). *The Structure of Scientific Theories*. Urbana, University of Illinois Press.

Symons, J., & Alvarado, R. (2022). Epistemic injustice and data science technologies. *Synthese*, *200*(2), 87. https://doi.org/10.1007/s11229-022-03631-z

Tal, E. (2023). Target specification bias, counterfactual prediction, and algorithmic fairness in healthcare. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 312–321. https://doi.org/10.1145/3600211.3604678

Talat, Z., Lulz, S., Bingel, J., & Augenstein, I. (2021). *Disembodied Machine Learning: On the Illusion of Objectivity in NLP* (No. arXiv:2101.11974). arXiv. https://doi.org/10.48550/arXiv.2101.11974

Thoma, J. (2024). Social Science, Policy and Democracy. *Philosophy & Public Affairs*, *52*(1), 5–41. https://doi.org/10.1111/papa.12250

Vessonen, E. (2019). Operationalism and Realism in Psychometrics. *Philosophy Compass*, *14*(10). https://doi.org/10.1111/phc3.12624

Vessonen, E. (2021). Respectful operationalism. *Theory & Psychology*, *31*(1), 84–105. https://doi.org/10.1177/0959354320945036

Wagstaff, K. (2012). Machine learning that matters. *arXiv Preprint arXiv:1206.4656*.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177. https://doi.org/10.1037/h0074428

Williams, T. (2024). *The Ghost in the Machine: Logical Positivism and the Verification Challenge in AI*. https://www.linkedin.com/pulse/ghost-machine-logical-positivism-verification-ai-tim-williams-phd-9cvfc

Zulfiker, M. S., Kabir, N., Biswas, A. A., Nazneen, T., & Uddin, M. S. (2021). An in-depth analysis of machine learning approaches to predict depression. *Current Research in Behavioral Sciences*, *2*, 100044.