Predatory Predictions and The Ethics of Predictive Analytics

Kirsten Martin, PhD.
William P. and Hazel B. White Professor of Technology Ethics
University of Notre Dame
IT, Analytics, and Operations
Kmarti33@nd.edu

In this paper I critically examine ethical issues introduced by predictive analytics. I argue firms can have a market incentive to construct deceptively inflated true-positive outcomes: individuals are over-categorized as requiring a penalizing treatment and the treatment leads to mistakenly thinking this label was correct. I show that differences in power between firms developing and using predictive analytics compared to subjects can lead to firms reaping the benefits of predatory predictions while subjects can bear the brunt of the costs. While profitable, the use of predatory predictions can deceive stakeholders by inflating the measurement of accuracy, diminish the individuality of subjects, and exert arbitrary power. I then argue that firms have a responsibility to distinguish between the treatment effect and predictive power of the predictive analytics program, better internalize the costs of categorizing someone as needing a penalizing treatment, and justify the predictions of subjects and general use of predictive analytics. Subjecting individuals to predatory predictions only for a firms' efficiency and benefit is unethical and an arbitrary exertion of power. Firms developing and deploying a predictive analytics program can benefit from constructing predatory predictions while the cost is borne by the less powerful subjects of the program.

Predatory Predictions and The Ethics of Predictive Analytics.

Within the larger scope of AI and data analytics, the use of predictive analytics has proven appetizing for organizations. Organizations use predictive analytics to infer and predict human behavior and actions, using data from the past to create a model that predicts events in the future (Birhane, 2021; Hälterlein, 2021). Organizations attempt to predict who will buy their products, who will cheat, who will be a good employee, which students will need help, who will commit more crime (Susser, 2020), who is likely to become more sick.

Rather than test causal explanations or describe a data sample, a predictive model, as the name suggests, predicts the value of an outcome variable from current data sources (Hälterlein, 2021).[1] For example, an AI program may categorize who currently has cancer in an image whereas a predictive analytics program may attempt to predict who will get cancer in the future. A predictive model learns from historical data centered on similar situations with the goal of forecasting what will happen in the future and, importantly, allowing the organization to intervene in some fashion (Waldman, 2019b).

This need to use the prediction for differential treatment – e.g., personalized medicine (Carnevale et al., 2021), personalized learning (Regan & Jesse, 2019; Williamson, 2017), personalized content (Mittelstadt, 2016), personalized insurance (Barry & Charpentier, 2020; Cevolini & Esposito, 2020), or personalized pricing (Seele et al., 2021) – introduces two distinctive ethical concerns where I will focus. First, the need for some subjects to receive a treatment (a greater sentence, harsher mortgage terms, etc), while others do not, leads to the creation of cutoffs to label subjects from data that is ambiguous at best. While the model may predict with a certain probability that an event may occur in the future, the application of the model requires a positive/negative label in order to apply a treatment (or not). This is the ethical issue of choosing the outcome variable.

A second distinctive feature of predictive analytics is that the measure of whether an individual was correctly predicted is in the future. In other words, whether or not the prediction that a patient actually would get cancer can only be known once the patient is allowed to continue on for a given amount of time. In a more well-known example, whether or not an individual predicted to commit future crimes will actually commit those crimes can only be

known after a defined amount of time (Larson et al., 2016). This is the ethical issue of measuring accuracy.

The goal of this paper is to critically examine ethical issues introduced by predictive analytics including categorizing subjects with a prediction and measuring the accuracy of the prediction. The article differentiates between (a) the developer, who designs and sells the predictive program, (b) the adopting organization, who applies the program within their context, and (c) the individual, who is the subject of the predictive program. Each actor has a different role in being impacted by or impacting the construction of the outcome variable and the measurement of accuracy of the predictive analytics program. And each actor has different voice or power in the market to have their interests and harms acknowledged in that design. As I show below, the difference in power between developer, organization, and subject can lead to market actors (developer and organizations) reaping the benefits of predictive analytics while subjects bear the brunt of the costs.

This paper proceeds as follows. First, I justify a critical approach as a useful lens to illuminate the ethical issues of using predictive analytics. In doing so, I join many others seeking to question not only the presumed objectivity and neutrality of analytics (Johnson, In Press) but also the power dynamics at play in building the algorithm, collecting and using the data, and selling AI and analytics (Beer, 2017; Benjamin, 2019; Diakopoulos, 2015; D'Ignazio & Klein, 2020; Floegel & Costello, 2022; Gibson & Martin III, 2019; Leavy et al., 2020; Levy, 2015; Paris et al., 2022; Poole et al., 2020; Waldman, 2019b; Zuboff, 2019). I extend critical approaches by explicitly examining the power of market forces on firms in their design of predictive analytics.

Second, I critically examine the categorization of the individual (positive versus negative) and the measurement of the accuracy of the prediction (true versus false) as constructed in design. I propose, for a given decision context, clearly identifying the moral implications of each type of result (true-positive, true-negative, false-positive, and false-negative), identifying who benefits and who is marginalized in each quadrant, and, importantly, understanding who is in power to construct and measure each of the quadrants. I show firms are vulnerable to the construction of deceptively inflated true-positive outcomes: where the construction of the outcome variable and the creation of accuracy can lead to more individuals being categorized as requiring treatment (the aperture effect) and the treatment actually leading to mistakenly thinking

this label was correct (the treatment effect). Finally, I ground why such predatory predictions are morally wrong. While profitable, the use of predatory predictions can unnecessarily harm individuals in being labeled as requiring a penalizing treatment, deceive stakeholders in inflating the measurement of accuracy, diminish the individuality of subjects, and exert arbitrary power in the design of predictive analytics.

I focus on the role of organizations, who have power over developers in purchasing the predictive analytics model and an obligation to subjects as the actor inflicting harm and arbitrary power. First, organizations should account for the costs of over-labeling subjects more directly. Organizations currently adopt predictive programs without the corresponding infrastructure to adjudicate claims of wrongfully labeled subjects or accounting for the costs of the treatment. Second, organizations should clearly measure the treatment effect versus predictive power of the model to more accurately explain to stakeholders whether the predictive model adds value or provides a fog of ambiguity for the organization making decisions. Third, organizations should justify the design of the outcome variable, and the cutoff as to when subjects receive a penalizing treatement, in order to avoid wielding unjust and arbitrary power over individuals. The arguments here suggest that decisions where the subjects have little power or voice are not only particularly vulnerable to predatory predictions but also where predatory predictions are particularly unethical in taking advantage of those who already are disadvantaged.

The use of predictive analytics has been disproportionately and arbitrarily deployed on subjects without power or voice thus allowing developers and organizations to over-categorize subjects without bearing the standard 'costs' of the treatment. This paper has implications to the examination and assessment of predictive analytics and data analytics more broadly. The questionable claims of AI's efficiency and accuracy runs at a fever pitch (Birhane et al., 2021). This article illustrates that such fixation on efficiency and accuracy may be because more the powerful actors construct the measurement of accuracy and benefit from the current measurements.

Critical Examination of Predictive Analytics

Critical theoretical approaches maintain a healthy skepticism towards any assumptions of neutrality or objectivity and contextualize situations in a way that accounts for the power and influence of different actors. Importantly, critical theoretical approaches seek to identify and critique systemic power relations with an intention to contribute to structural change and even emancipation (Poole et al., 2020; Stahl, 2021). A critical approach to examine predictive analytics is justified to overcome (a) the presumed objectivity of the model as well as (b) the power dynamics underlying the design, development, and deployment of predictive analytics in the market.

First, a dominant approach to data analytics, including AI, machine learning (ML), and other types of analytics, is to falsely presume objectivity and neutrality of the decision (Johnson, In Press; Martin, 2022). Such models satisfy our Dewian quest for certainty and predictability (Dewey, see also Birhane, 2021), where predictions "are accepted as valid, interpreted as the product of intelligent and objective technical assessments" (Gill, 2020). While predictive analytics is framed as objective and neutral, the data, models, and outcomes are the culmination of value-laden human decisions. Unfortunately, such programs "often uncritically inherit, accept and incorporate dominant cultural and belief systems, which are then normalized" (Gill, 2020).

Second, predictive analytics are increasingly implemented within systems of control and power – particularly in the market. As Ari Waldman correctly states, "[u]sing algorithms to make commercial and social decisions is really a story about power, the people who have it, and how it affects the rest of us" (Waldman, 2019b, p. 615). While all "data are a form of power" (Iliadis & Russo, 2016), predictive analytics are used to "impose order, equilibrium, and stability to the active, fluid, messy, and unpredictable nature of human behaviour and the social world at large" (Birhane, 2021). And our current use of predictive analytics illustrates the danger of data-driven decisions being in the control of powerful single actors (Carnevale et al., 2021). The current use of inferences to predict attributes of people "magnifies the power of organizations that collect and process data, while disempowering the people who provide data and who are affected by data-driven decisions" (Solow-Niederman, Forthcoming, p. 1).

The marketplace, within which predictive analytics programs are designed, sold, and used, is not a neutral site. The market is political and built on social and structural relations that are connected to inequalities (Henderson & Williams, 2013; Poole et al., 2020). Within the critical examination of Big Tech as a market, previous research has focused on the damaging influence of corporations on the direction of AI ethics research (Abdalla & Abdalla, 2021), the power of the corporation over data and privacy (Waldman, 2021), and powerful corporations prioritizing efficiency and freedom for a subset of society (Cohen, 2019; Waldman, 2019b).

The critical examination of predicative analytics herein continues a rich line of scholarship seeking to understand who gains power and who is disenfranchised by the design decisions in data and analytics.[2] A critical examination would be to ask not only about alternatives but also question the power dynamics within the market and seek "to dismantle entrenched hierarchical marketplace dynamics" (Poole et al., 2020). This explicit lens of power – who has it and who benefits and is harmed from the decisions made – would be turned to the design decisions of predictive analytics. The power dynamics of markets is an important addition as many ethical examinations of fairness of AI are considered outside the pressures of markets. Markets are not perfect and, as we see here, firms can create harms to external stakeholders who do not have power to negotiate their interests through market transactions. A critical approach to markets embraces this power differential rather than ignore it.

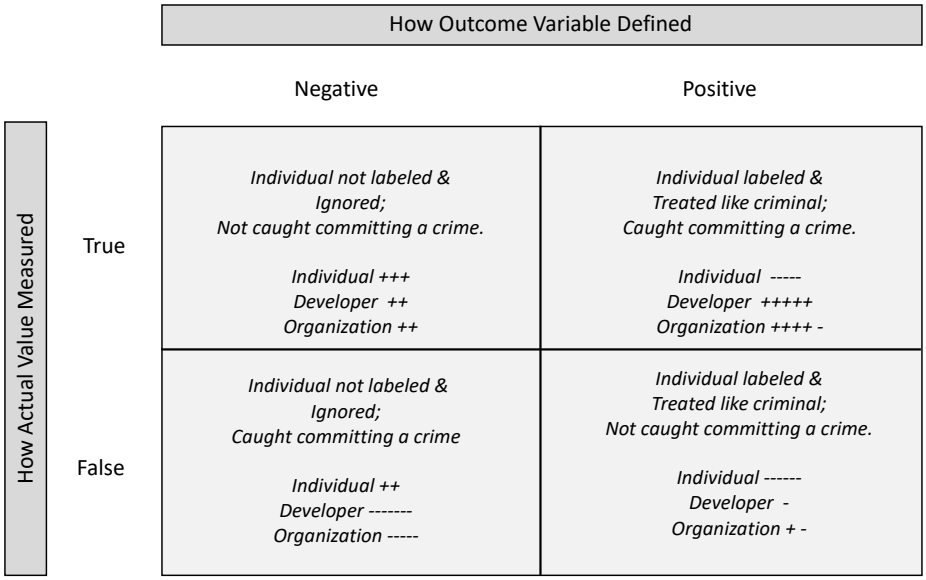
Critical Approach to Predictive Analytics

While predictive analytics shares many of the concerns and moral implications of data analytics, AI, and ML generally,[3] I focus on two facets that differentiate predictive analytics in morally important ways. First, the outcome variable is more likely to be dichotomous, due to the need to treat people differently as a result of the prediction. While the prediction algorithm may still produce an outcome weight that is on a continuous scale, the need to apply a treatment to a subset of the subjects requires the outcome variable to be cut into categories or labels. In the examples used here, the outcome variables are binary (positive and receive treatment versus negative and do not receive treatment). [4]

How this outcome variable is defined impacts the number and type of people categorized as 'positive' and requiring the assigned treatment. For example, for a predictive analytics program

attempting to identify and treat individuals as possible criminals, how the outcome variable is defined will impact the top axis of Figure 1.

Figure 1: Result Matrix for Predictive Analytics – Predicting Criminals



In addition, determining whether the predictive label is 'true' or 'false' is conducted after time passes. In terms of the Figure 1 confusion matrix, I am examining how the matrix is carved in to positive/negative x-axis and true/false y-axis. For example, organizations predict criminality using analytics in two different realms. First, courts have used predictive analytics, COMPAS most famously, to predict whether a defendant will commit another crime in the future (Dieterich et al., 2016). The prediction is used to determine parole and sentences. In addition, police and sheriff departments have used predictive analytics to predict whether students as well as individuals will commit a crime in the future (Bedi & McGrory, 2020; McGrory & Bedi, 2020). This prediction is then used for interventions by the police.

The three primary actors would be considered party to the design, development and use of a prediction model in that these actors are influenced by and influence the decisions. First, the developer of the prediction analytics program makes value-laden design decisions such as the type of data sets used in training the model, the specific features of the individual that are included in the training data, the assumptions made about the data and the outcome weights, as well as defining the outcome variable (Martin, 2022). Second, the organization purchasing and adopting the predictive analytics program – the school, police department, bank – uses the model

for a particular context. Finally, individuals are subject of the prediction analytics program and are impacted by the prediction and subsequent treatment. Each actor can benefit from the design of the predictive analytics program such as appearing accurate, avoiding a penalizing treatment, or fulfilling their mission. These actors can also be harmed in the design decisions in having their rights taken away, having to implement a costly treatment, suffering financial harms, being treated unfairly, having additional burdens placed on them, having their reputation harmed, etc.

Importantly, a critical approach to understanding the market influence on the design decisions of the predictive analytics program does not take for granted that all market actors (individuals, organizations, and developers) have equal standing in the market with the requisite power to have their concerns and preferences incorporated in their transactions. For example, the subject of a predictive program may wish to contest a decision or may be upset at a bad outcome. While a perfect market, if one could exist, would have any costs felt by individual incorporated into their interaction with the organization using the analytics program (Coase, 1960), the current use of predictive analytics tends to be focused on individuals without voice or power in the market. This lack of power has implications as to whether subjects are able to 'negotiate' their interests in the design, development, and use of predictive analytics and whether the costs of the program are fairly distributed as explored below.

Figure 1 includes an example analysis of the distribution of who benefits and who is harmed based on the type of result for a specific predictive analytics program – predicting likelihood of a person to commit a crime.

- For true-positives, the organization treats individuals (defendants/students) like a criminal and the individual is later recorded as committing a crime. The organization implementing the program benefits from claiming an accurate prediction and fulfilling their mission of identifying people thought to be criminals. The organization could have costs associated with labeling someone as requiring a treatment: e.g., the treatment and the possibility of individuals contesting the categorization. However, the examples used here have fixed costs for the treatment and minimal ability for individuals to appeal. For this decision context, the police assigned to the school are fixed and judges requiring additional prison time do not 'pay' for the prison.[5] I discuss relaxing this assumption as an important approach to changing the market dynamics in the last section. Another benefit for the organization is demonstrating that the predictive analytics program works

and adds value by predicting behavior that is then shown to be 'true.' The individual is harmed in that they are treated like a criminal before committing any crimes – including diminishment of rights, harassment, and additional burdens place on them. The developer benefits by claiming a higher accuracy rate (true positive), which is the primary metric of success for selling predictive analytics programs. There are no costs for the developer for the predictive analytics program categorizing someone as requiring a penalizing treatment.

- For true-negatives, the organization ignores individuals who are never caught committing crimes. The individual benefits as they are not treated like a criminal and are never caught committing a crime. The organization benefits by demonstrating their program is accurate and avoiding any minimal treatment costs identified above. The developer benefits by demonstrating their predictive model is accurate. However, too many true-negatives would suggest the program is not needed which would be a cost for the organization and even more so for the developer. For example, if all subjects are categorized as not being a future criminal (and not needing the treatment), the utility of the predictive program is minimal.
- For false-negatives, the organization ignores defendants/students who then are recorded as committing a crime in the future. The organization could be harmed both reputationally for a bad prediction and in not achieving their mission, if the number of false negatives is revealed and publicized. This could be the worst-case scenario for the organization, who adopted the predictive analytics program to minimize a negative outcome only to have the program be wrong and still have the negative outcome (crime as recorded). Similarly, the developer would be harmed reputationally for an incorrect prediction and in missing the opportunity of the organization applying a penalizing treatment. However, the individual benefits by not being treated like a criminal before being recorded as having committed a crime.
- For false-positives, the organization treats an defendants/students as a criminal who are later never caught committing any crime. The individual is harmed by being treated as a criminal even though they never commit a crime. The organization benefits slightly in pursing their mission but then appear inaccurate. *If the instances of false positives are recorded, believed, and publicized*, the organization runs a risk of their reputation being

tarnished. However, in this case of predictive analytics on prisoners and students, the voices of those incorrectly labeled positive, who are then treated like future criminals, do not have the position in the market to force the organization or developers to bear any of their costs. As a counter example, if predictive analytics was used to predict executives of committing fraud, false-positives would be considered quite expensive due to lawsuits, employment grievances, complaints, as well as reputational harm to the organization and developer. Similarly, the developer could be harmed reputationally by their accuracy rating being diminished and producing a mistake if this metric is tracked, believed, and publicized.

Thus far I have argued that who benefits and who is harmed for a given use of predictive analytics differs across the results matrix and is not evenly distributed. Further, that developers, organizations, and individuals who are subject to the prediction do not have the same power or voice in the market to have their interests addressed. I turn now to critically examine the construction of the outcome variable (x-axis) and accuracy (y-axis) which impacts the number of individuals who fall in each quadrant in Figure 1.

Ethics of Choosing Outcomes

Similar to AI models in general, the outcome variable chosen has implications as to what the organization thinks is important and whose interests are prioritized in the design of the data analytics program. While the outcome variable is a proxy for 'what you really care about' (Thomas & Uminsky, 2020), the outcome variable can be a sloppy proxy for the phenomenon of interest. In fact, we often cannot measure the phenomena that matter the most (Thomas & Uminsky, 2020). In addition, most outcome variables incorporate an underlying model of behavior or theory of how things work. For example, an outcome variable for recidivism makes assumptions as to whether family, societal, environmental, behavioral factors are important to the prediction (Hälterlein, 2021). These factors can also be discriminatory or unfair (Barocas & Selbst, 2016; Martin, 2019).

Predictive analytics carries two additional burdens that make the outcome variable more problematic. First, with predictive analytics the outcome is a probability about the future rather than the present. The ambiguity of attempting to categorize a current situation is exacerbated by adding the probability that situation will occur in the future. So, identifying cancer in a patient is hard, predicting who will get cancer is even harder (e.g., Huang et al., 2020). Second, the need to treat people differently based on the value of the outcome variable pushes us to create a dichotomous label in a particularly ambiguous situation. For example, for a predictive analytics program to predict an individual's' likelihood to commit a crime, those predicted positive (likely to commit a crime) receive the assigned treatment (additional scrutiny, searches, more parole requirements, etc).

When predictive analytics are forced into dichotomous choices, organizations must choose a seemingly artificial cut off point to transform a continuous outcome to a binary categorization. Hidlebrand (2008) differentiates between distributive and non-distributive data profiles (see also Vedder, 1999), where a distributive profile can be applied to all members of a group, thus making dichotomous cutoffs easier. Hildebrand uses the example of bachelors, who all share 'not being married' to the same degree, and "the profile will apply without qualification to all members" (Hildebrandt, 2008, p. 21), thus making dichotomous cutoffs easier since one is a bachelor or not.

However, most groups and labels are not distributive and are not true by definition. Hildebrandt uses the example of a checklist for psychopaths where each facet is valued 0-2 and

the total possible ranges from 0-40. The psychopath test was used to predict if a prisoner posed a danger to the community. The cutoff for being labeled a psychopath in one particular program was 30; those with a score above 30 were deemed a psychopath and not released from prison. The meaning of the differential treatment – to be released from prison versus extending their sentence – is clear and large, yet the associated difference between an individual scoring a 29 versus a 31 is not clear or meaningful. At a minimum, the outcome variable cutoff is not obviously neutral, clear, or unambiguous.

Take the example of the use of proctoring software to track remote students to predict if they are cheating or not (Harwell, 2019).

> One system, Proctorio, uses gaze-detection, face-detection and computer-monitoring software to flag students for any "abnormal" head movement, mouse movement, eye wandering, computer window resizing, tab opening, scrolling, clicking, typing, and copies and pastes. A student can be flagged for finishing the test too quickly, or too slowly, clicking too much, or not enough (Harwell, 2019).

The threshold to trigger the alarm for Proctorio is sensitive with many people being flagged for merely leaning back in their chair or having someone walk in the room. Students whose eyes wonder were flagged as well. In 2019, the company reported 6% of all exams had been flagged as "confirmed breaches of integrity" (Harwell 2019). However, what the cutoff is for being categorized as 'positive' and a predicted cheater is not discussed. In this way, the categorizing of individuals in predictive analytics carries the veneer of precision but is actually imprecise.

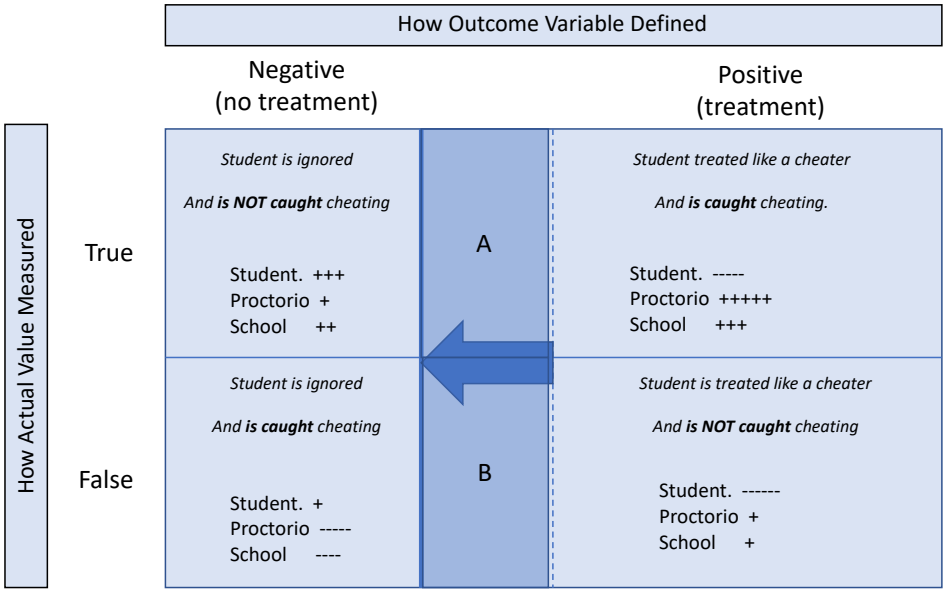
Returning to the result matrix, how the outcome variable is transformed into dichotomous categories impacts the top axis. When the positive label is broadened to include more individuals, the line moves to the left and more individuals receive the treatment. In Hildebrandt's psychopath test, this would be if the cutoff was shifted from a 30 to 25 so that more individuals qualify as a psychopath and receive treatment (staying in prison). Proctorio can broaden who is considered a possible cheater and increase the number of individuals being labeled 'positive,' by including more triggers and/or shifting the threshold so that more students are flagged as possible cheaters.

The gap, shaded in a Figure 2 for cheating prediction software, exemplifies those newly labeled positives based on shifting the criteria for the cutoff of the outcome variable. In effect, the developer is opening the aperture to let in more individuals to fall to the right of the line,

being categorized as positive, and requiring a treatment. A wider aperture on the model's lens, the more people are categorized as positive.

Determining the optimal point to draw the line is an optimization problem and dependent on the positive and negative costs identified within the results matrix. This is true whether the cutoff is determined manually or the outcome of an optimization program. For individuals being labeled possible future criminals, widening the criteria for who is categorized 'positive' means more individuals being labeled future criminals, more individuals being tracked and surveilled by the police, more students with a file sent to the police, more knocks on their houses to check in, and a shift in the presumption of innocence for both students (Bedi & McGrory, 2020) and citizens (McGory, 2021). This is true for both those that end up being caught (A) and those that never are caught committing a crime (B).

Figure 2: Result Matrix for Predictive Analytics Outcome Shifted for Proctorio



Importantly, shifting the line to the left in Figure 2 and opening the aperture to categorize more subjects as positive can be attractive if the harm of both quadrants is not felt by the developers or the organization adopting the program but only by the individual who is subject to the treatment.

For cheating software, if only Proctorio and the school are considered, then the benefits of widening the aperture appear overwhelming. Proctorio has the hope of identifying more true positives which makes the developer look better and the school can report a greater number of students caught cheating. Proctorio can also avoid the dreaded false negative (missing a student

who cheats) which would damage their reputation. Importantly, the treatment cost is minimal since students are notified with an automatic email or testing is suspended automatically. And the initial design did not leave students with an ability to appeal; cheating detection programs were used for years before Dartmouth medical students who were over-categorized as having cheated complained and had their story covered by national newspapers. The cost of over-predicting to the organization was minimal and almost nothing for Proctorio.

When a firm sells a cheating detection program to school administrators, the students feel the harm of being aggressively surveilled and categorized as a possible cheater. Since those students are not paying for the software, their concerns can be minimized because the 'cost' is not felt by the firm or the school. This is the market pressure developers feel to maximize true-positives by not incorporating the costs borne by others. This problem of creating third party externalities in markets is common because a firm has no immediate incentive to consider the costs not a part of the transaction but borne by others, particularly when the costs are not identified, acknowledged, or publicized for some subjects.[6]

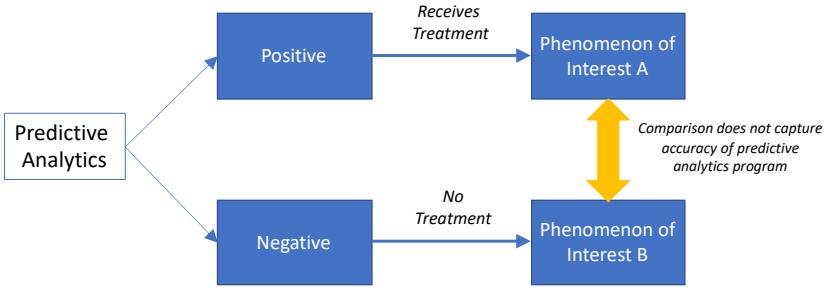
Ethics of Creating Accuracy.

I have argued that the construction of the outcome variable is a value-laden design decision and that how the outcome variable is defined can be optimized for the developer or the organization adopting the program while undermining the interests of the individuals or subjects. With predictive analytics, the optimal cutoff to require a treatment may be inflated, when the cost of widening the model to identify more individuals as needing a treatment is not felt by the developer or organization. I turn now to the ethics of measuring and creating accuracy.

In general, predictive analytics shares the existing issues with measuring efficacy and accuracy as other data analytics, AI, and ML programs. The use of historical data is inherently conservative by reproducing and reinforcing norms, practices, and traditions of the past (Birhane, 2021). And when historical data reflects the unjust decisions of individuals, the model inherits the discriminatory recording of history only to reproduce that discrimination in the future (Barocas et al., 2018; Benjamin, 2019; Birhane, 2021; O'Neil, 2016). In addition, the data from predictive outcomes and recommendations are fed back into the system, thereby reproducing and confirming biased correlations (Gill, 2020). This can create a feedback loop where the categorization or prediction becomes evidence to include in the training data (O'Neil, 2016). These are known issues.

However, predictive analytics faces an additional challenge in *creating* accuracy**,** where individuals with a particular outcome variable are treated differently than those with an alternative predictive score. In other words, when organizations measure and compare the end state of those categorized positive and negative, organizations do not compare similarly treated conditions. Figure 3 illustrates the problem.

Figure 3 Measuring accuracy and treatment effects.



Predictive analytics runs into the possible problem of creating accuracy in categorizing someone with an outcome variable (not promotable, not hirable, untrustworthy, high likelihood of recidivism, etc), which pushes the individual into a course of treatment that then creates the outcome predicted by the program. For example, predicting students as possible criminals brings closer scrutiny to their lives, making it more likely for police to find evidence of bad behaviors even if the incidence of bad behaviors is equal across those labeled positive (and scrutinized) and negative (and ignored).

Perdomo et al. refer to those types of predictions as 'performative,' when the prediction influences the target or when predictive models trigger actions that influence the outcome they aim to predict (Perdomo et al., 2020; see also O'Neill, 2016). This issue of a prediction influencing the resultant 'accuracy' is not new to predictive analytics. As noted by Perdomo et al, in regards to accessing the trustworthiness in loan applicants, "In a self-fulfilling prophecy, the high interest rate further increases the customer's default risk" (Perdomo et al., 2020). Perdomo et al treat the problem as an issue of corrupting the data: when ignored, performativity can surface as a form of distribution shift in the data.[7] The authors focus on the technical details of retraining a model on new data (Perdomo et al., 2020). Here I focus on the ethical implications of measuring the performance of predictive analytics and the power dynamics influencing how accuracy is measured. The problem of creating accuracy compounds the ethical implications of over categorizing individuals as positive.

The case of predictive policing is perhaps the quintessential example of creating accuracy. Identifying a particular neighborhood as possibly being more likely to have petty crimes (the prediction) leads to more officers sent to look for crimes in that neighborhood (the treatment). A larger number of officers finds more crimes and arrests more people.[8] This is the standard argument as to how predictive analytics can feed into creating a new reality and create the perception of being 'accurate:'

The case of Chicago's predictive policing program is an even more stark example of creating accuracy. Robert McDaniel, a citizen in Chicago, was visited by Chicago police and told,

> an algorithm built by the Chicago Police Department predicted — based on his proximity to and relationships with known shooters and shooting casualties — that McDaniel would be involved in a shooting. That he would be a "party to violence," but it wasn't clear what side of the barrel he might be on. He could be the shooter, he might get shot. They didn't know. (Stroud, 2021).

McDaniel was also told that the police would be watching him. And while McDaniel had no violent history, he was suddenly under constant surveillance by the police. The predictive analytics program made a prediction, the police department was the immediate treatment. However, this increased attention and visits by the police looked suspicious to those in his neighborhood who thought he was working *with* the police. McDaniel was then shot – twice – by those in his neighborhood who believed he was a snitch given the amount of police attention he was receiving (Stroud, 2021). In this case, the program appears quite accurate since the prediction was that McDaniel would be involved in a shooting, and he was shot – twice.

Figure 4: Result Matrix with Treatment Effect Included for Predicting Shootings

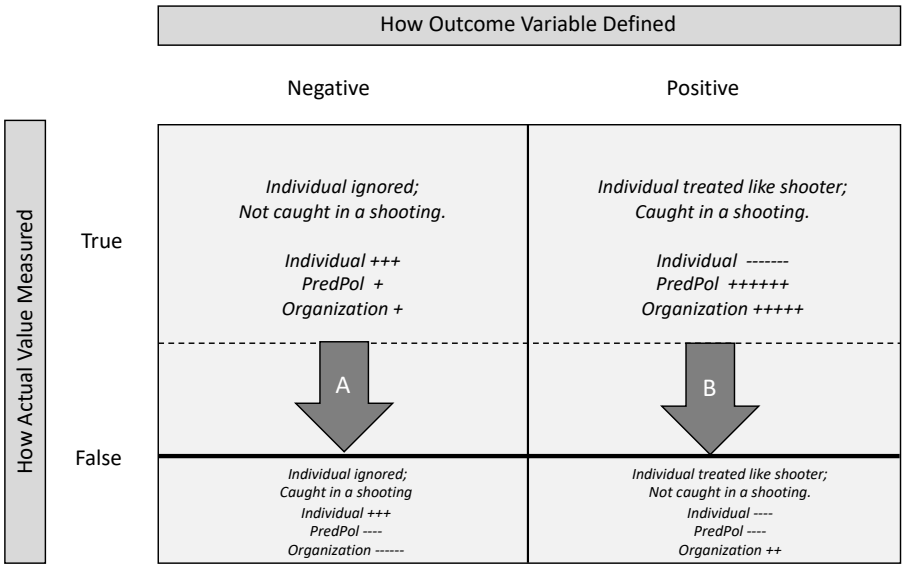


Figure 4 illustrates how the treatment could work to make more individuals classified positive (criminal, involved in shooting, cheater, bad credit risk, etc) to be later found to have been correctly classified (Arrow B).

This problem of creating accuracy occurs frequently in predictive data analytics. At a parole hearing, courts may classify a prisoner with a higher recidivism score predicting they are likely to commit more crimes (e.g., COMPAS). The courts then place additional requirements on those paroled prisoners (where they can live, drug tests, how frequently they need to check in, who they can live with, etc) which make them more likely to commit a parole violation. Those classified as likely to recidivate have additional rules places on them to make the more likely to break those rules. The program or treatment contributes to the creation of accuracy.

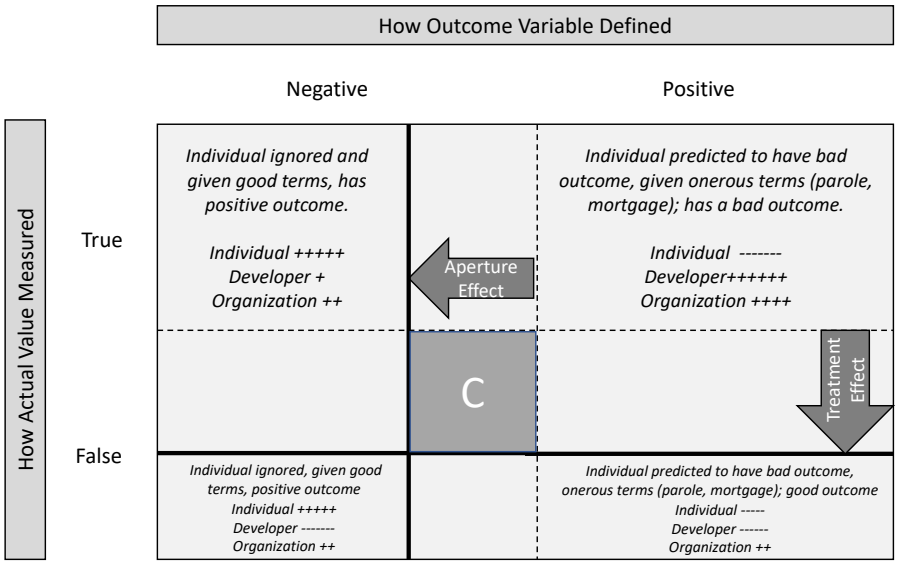
In addition, the treatment effect in our examples also creates arrow A in Figure 4, where the treatment works to make more individuals classified negative (not a criminal, not involved in a

shooting, not a cheater, good credit risk) and not requiring a treatment to be later found to have been correctly labeled. In fact, the goal of the treatment in some situations is to make the categorization more likely to be found 'true.' For example, when employees are labeled as future leaders and are given leadership development, those labeled should be more likely to be future leaders.

The issue is similar to measuring the treatment effect in medical research, only in medical research the organization developing the treatment has an interest in the effectiveness of the treatment and a market incentive to measure its effect. For example, firms selling a drug intervention measure the treatment effect of the given drug on, for example, cholesterol. And, within econometrics we have robust methods to measure the causal effects of the treatment on a response variable (Lee, 2005). Firms use these methods to prove treatments are effective in order to sell those treatments.

Here, for firms developing a predictive analytics program, measuring the treatment effect – the size of the Arrows A and B in Figure 5 – means acknowledging that the data analytics program is, perhaps, not as accurate as once thought. Firms selling predictive analytics programs may not welcome research measuring treatment effects to isolate whether their prediction analytics program actually 'works,' because the more impact a treatment has – e.g., sending more police to a given area or to a given person – the less the predictive analytics program actually works in adding value as a prediction.

Figure 5: Predatory Predictions



Taken together in Figure 5, the optimization problem of creating the outcome variable and measuring accuracy will depend on whose interests are acknowledged and influence the decisions. First, by opening the aperture to allow more individuals to receive treatment (predicted cheater, future criminal), the organization hopes to catch more true positives without much (if any) costs. This is a classic optimization problem where the benefits of over producing or, here, over categorizing individuals are larger than the associated costs if and only if the costs to the individuals or society are not included.[9] The cost for drawing the line with a larger number of true-positives 'caught' is borne by the subject of the predictive analytics program: the student, defendant, citizen.

Second, the developer and organization benefits from then conflating the value of the treatment with the prediction. While the individual is harmed by the large treatment effect (high interest rates, onerous parole terms), the developer is able to show a larger true-positives (and fewer false-positives). In other words, the focus on true positives by firms developing predictive analytics is no accident.

This highlights the harms of the individual particularly in Box C in Figure 5, who were arbitrarily categorized as positive and who, because they were so labeled, saw that label become true. Consider someone labeled as untrustworthy, given onerous loan terms, and then struggles to pay back the loan. Or a defendant is arbitrarily predicted to commit another crime, given harsher parole terms, which are then broken. Unlike those who have a bad credit history, and unlike those who would have struggled no matter the terms, individuals in Box C are subject to predatory predictions: where individuals do not deserve the positive categorization and the positive label then negatively impacts their outcomes. This paper has used market forces to explain why firms are drawn to predatory predictions.

Normatively Grounding Corporate Responsibility and Predictive Analytics

Thus far, I have argued that when constructing the outcome variable and measuring accuracy, firms developing and adopting predictive analytics can have a market incentive to create more true positives. Firms open the aperture of who deserves a treatment and that treatment then makes the categorization more likely to be true. I turn to explore why such predatory predictions are morally wrong and the associated responsibility of developers and organizations.

First, the conflation of the treatment effect with the measurement of the predictive power of the program inflates the efficacy of the predictive analytics program and is a form of deception. If a developer inflates the accuracy measurements of a predictive analytics program by including the treatment effect, the organization is under the mistaken impression as to the predictive power of the program. The subject or potential subjects, as well as society and regulators, are deceived into believing that the predictive program is performing a service that otherwise could not have been completed by humans.

Organizations would have an obligation to not deceive by conflating the treatment effect with the predictive power of the program. Organizations would need to clearly measure and communicate the treatment effect and require developers to report the difference in their accuracy claims. Clearly measuring the treatment effect versus the predictive power of the model would more accurately explain to stakeholders whether the predictive model adds value or provides a fog of ambiguity for the organization to make hard decisions with less accountability.

Second, for predatory predictions, the construction of the outcome variable can be optimized for the developer or client organization but without the interests of the subjects taken into account. The individuals would be over-categorized as requiring a penalizing treatment. When the interests of the subjects are not considered, the organization developing and adopting the predictive analytics program would be considered morally wrong for treating the subjects as a mere means (Kant, 1785). In some circumstances, e.g., in Box C in Figure 5, opening the aperture of who requires treatment and having a strong treatment effect would pull individuals into a penalizing treatment and the treatment would cause a negative outcome for that individual only because the organization benefits from the labeling.

Organizations should ameliorate the incentive to open the aperture and categorizing additional subjects as requiring a treatment by better internalizing the costs of the categorization.

Organizations should incorporate the treatment costs in their decisions to categorize subjects to internalize the costs of labeling subjects as requiring a penalizing treatment. For example, decision makers using predictive analytics should more directly bear the cost of the treatment (jail, police) as well as the cost of making a mistake in the categorization. This could be accomplished by how decision-makers are evaluated (the cost associated with the treatment in addition to the accuracy of the program) or how budgets are allocated. Further, organizations currently adopt predictive programs without the corresponding infrastructure to adjudicate claims of incorrectly labeling subjects. Organizations should bear an internal cost to alleviate some of the cost to subjects to contest a decision or voice harms and concerns.

The case of the Dartmouth medical student is illustrative in that the aggressive over-categorization of students labeled as cheating. Subjecting individuals to predatory predictions only for a firm's efficiency and benefit is unethical and was known as an issue for years, but it took 17 medical students at an elite university before the voices of students had credence. Further, part of the issue identified was the students' inability to challenge or appeal the program's decision (Singer, 2021). Currently, the use of predictive analytics focuses on subjects who do not have power to have their concerns acknowledged by the organization. However, false positives and harms to the subject should be tracked and publicized to force developers and organizations to bear a 'cost' for opening the aperture to categorize more subjects as needing a treatment.

Third, the developers and adopting organizations can be seen as wielding arbitrary power over individuals in the cases explored here. The traditional examination of the arbitrary use of power focuses on humans as those who use power arbitrarily (Pettit, 1996, 1997) and has been extended to corporations (Hseih 2004).[10] An act is arbitrary when it is performed with a lack of rules, reason, or controls or at the whim of the actor in power (Pettit 1997; Hseih 2004). Such arbitrary interference in the lives of others treats someone as though their interests and judgments do not matter (Hseih, 2004, p. 653). For justice scholars, arbitrary decisions are by definition unjust in that there is no legitimate rationale or systematic rules being applied.

As explored here, firms using predictive analytics can exert power over subjects through the design of predictive analytics programs[11] and arbitrarily interfere with an individual's choices or undermine an individual's right to not be dominated by another (Hsieh, 2004, 2005; Mink, 2020). Subjecting individuals to predatory predictions only for a firms' efficiency and benefit is

unethical and an arbitrary exertion of power. Decisions would be arbitrary when the prediction is based merely on the benefits that accrue to the developer or organization and without regard to the subject (consider Box C of Figure 5).[12]

In order to avoid arbitrarily using power in the use of predictive analytics, organizations should justify the cutoff for the outcome variable. For some industries and organizations, current stakeholders and governing bodies – boards, top management team, etc – would ask managers adopting predictive analytics to justify the design of the outcome variable. Managers have obligations to stakeholders that do not disappear when managers adopt AI decision systems including predictive analytics (Martin and Parmar, 2022). Just as those managers must justify their decisions to stakeholders and governing bodies when their team makes a decision, so too must those managers justify their design decisions as to who needs a penalizing treatment when using predictive analytics. For industries with a more robust regulatory oversight, e.g., public schools, legal system, etc, those regulatory bodies would require justification for how the organization decided to determine who needs treatment using the predictive analytics program.

Finally, there are situations where de-individualization – treating individuals differently based only on their shared attributes with others – is unethical. Someone is treated as an individual when decisions are informed by relevant information, including the individual's unique and idiosyncratic features (Birhane, 2021; Lippert-Rasmussen, 2011). The deindividualization of the person is the tendency to judge and treat people on the basis of group characteristics instead of on their own individual characteristics and merits (Vedder, 1999).

One does not need to be the victim of discrimination or subject to a mistake to encounter the harm of de-individualization. The idea is that we, in some circumstances, have an obligation "to recognize certain features of other persons qua persons, such as the intrinsic value of their well-being or the character of their individual autonomy" (Eidelson, 2015). The harm, therefore, is not about being wrong with a predictive analytics program but that the organization makes decisions about someone based on what 'people like you' do. For example, Binn notes that individual justice is when each case is assessed on its own merits without comparison to a reference set of cases (Binns, 2019). Individual justice can only be achieved through human judgement because the mere idea of judging someone without regard to a reference set of individuals and cases is antithetical to algorithmic decision making in general and predictive analyics in our case. In such a situation, additional human oversight may be required.

Therefore, there may be sets of decisions that require Binn's individual justice, other decisions that warrant deferring to the individualization of the subjects, and still others that are normatively appropriate to be augmented with predictive analytics, which relies on the subject's shared attributes with others to make a prediction. For example, Virginia Eubanks rightly critiques the flattening the individual when using predictive analytics for the distribution of social goods, such as TANF, SNAP, Medicaid in Indiana, homeless services in LA, or child welfare in Allegheny county. And Eubanks highlights the extent to which those already marginalized are being subject to predictive models that "tag them as risky investments and problematic parents" (Eubanks, 2018). For us, Eubanks correct critique suggests two related concerns. The power and voice of the subject of predictive analytics and the context of the decision (here, social services) should impact whether individualized decisions and justice are required. In addition, the threat of de-individualization may not be fairly distributed in society. In terms of the arguments of this article, predatory prediction programs can be disproportionally being used on marginalized communities (Benjamin, 2019; D'Ignazio & Klein, 2020; Eubanks, 2018; Noble, 2018; Paullada, 2020), since their concerns and interests are more easily dismissed in the market. Organizations would have an obligation to ensure their predictive analytics program is not being used in decision contexts where individual treatment is expected or is regularly afforded to more powerful individuals.

The arguments here suggest that decisions where the subjects have little power or voice are not only particularly vulnerable to predatory predictions but also where predatory predictions are particularly unethical in taking advantage of those who already are disadvantaged. In other words, the lack of a strong market correction through the voice of the subject means both the developer and adopting organization are more likely to ignore the cost borne by the subjects, thereby overcategorizing subjects as needing a penalizing treatment. In doing so, the predatory predictions take advantage of the subjects' vulnerability in the market to their own advantage. This suggests that the degree of power or voice of the subjects is one way to distinguish between different degree of moral wrongness of predatory predictions. And an organization adopting predictive analytics should first ensure subjects' concerns and preferences are heard in both design and implementation.

TABLE 1: Obligations of Firms with Predictive Analytics

Problem	**Obligation of Firms…**	**Recommendation**
Treatment Effect	…to not deceive	*Measure Treatment Effect*. clearly measure the treatment effect versus predictive power of the model to more accurately explain to stakeholders whether the predictive model adds value
Aperture Effect	…to not treat subjects as a mere means	*Decrease Incentive to Open Aperture*. Internalize the costs of labeling subjects as requiring a penalizing treatment including the cost of the treatment (jail, police) as well as the cost to give subjects voice to identify harms and hear appeals.
Aperture & Treatment Effect	…to not exert arbitrary power	*Justify Outcome Variable.* Justify the cutoff for the outcome variable in order to avoid wielding unjust and arbitrary power over individuals. Any justification that the organization benefits from current optimization would be considered unjust.
Use of Predictive Analytics	…to minimize deindividualization	*Justify Use of Predictive Analytics.* Be judicious when predictive analytics is used, particularly when the program is used in decision contexts that are pivotal (loans, educations, judicial system) or the subjects lack power (or both).

Conclusion

I have argued that firms developing predictive analytics can have a market incentive to categorize a larger number of subjects as requiring a treatment by increasing the possibility of true positives without the bearing the costs of the accompanying false labels. In addition, the measurement of the accuracy of a predictive analytics program can conflate the predictive value and a possible treatment effect, where those being treated are more likely to achieve the predicted result. Firms developing predictive analytics, therefore, can have a market incentive to increase the measurement of true-positives by opening the aperture to allow in more positives and by inflating the number of 'true' predictions through predatory predictions. However, the costs are primarily borne by the subjects.

Based on the arguments here, firms could design a predictive analytics program assuming the subjects are powerful actors with access to resources and relationships to have their concerns and preferences acknowledged in the market – no matter what the market power of the subjects actually is. This would forestall firms adopting or developing predictive analytics from taking advantage of vulnerable subjects without power to enforce their own rights or preferences. For firms adopting predictive analytics programs, the organization should ensure that the subjects have adequate voice to enforce their rights and preferences in the market.

Firms have an associated responsibility to incorporate the interests of subjects because the inflated count of true positives is a form of deception, because their acts can be a form of arbitrary, unjust exertion of power, and because the design of the program may undermine the individualization of the subject. The market forces that reward such *predatory predictions* explain why we continue to see unfairness in predicting trustworthiness of consumers for loans, recidivism for prisoners, child neglect, mental disorders, future victims of violence, employability, leadership, home prices, cheaters.

Endnotes

[1] Specifically, predictive analytics are differentiated from other types of analytics in that predictive analytics predicts the future rather than categorizing the present or explaining the past: "It predicts the future by analyzing current and historical data. The future events and behavior of variables can be predicted using the models of predictive analytics. A score is given by mostly predictive analytics models. A higher score indicates the higher likelihood of occurrence of an event and a lower score indicates the lower likelihood of occurrence of the event" (Kumar & Garg, 2018).

[2] Taking a critical approach to the examination of predictive analytics extends existing critical research within data analytics more generally such as the examination of information marginalization of vulnerable individuals (Tang et al., 2021), feminist technoscience in information systems (Floegel & Costello, 2022), and structural and power vulnerabilities in higher ed use of online platforms (Paris et al., 2022). For example, the examination of whether technology is helping only those with power and advantage (Mohammad, 2021), who benefits from making predictions with AI (Kerr & Earle, 2013), if due process rights of individuals are undermined (Citron, 2007), or if AI is used to further disenfranchise people in poverty (Eubanks, 2018), reinforce systemic racism (Benjamin, 2019) and misogyny (D'Ignazio & Klein, 2020) and disproportionately impact LGBTQ+ (Waldman, 2019a). Even more generally, we see this critical lens being used to highlight when marginalized groups are disparately harmed by privacy violations (Skinner-Thompson, 2020) or are victims of nonconsensual pornography (Citron & Franks, 2014; Keats Citron, 2018).

[3] Here I focus on the ethics of predictive analytics above and beyond the concerns outlined within AI programs generally: such as the reproduction of biases (Gill, 2020), transparency (Levy & Johns, 2016), problematic proxies (Barocas & Selbst, 2016; O'Neil, 2016; Williamson, 2017), fairness (Barocas et al., 2018; Belitz et al., 2022; Hoffmann et al., 2018), data (Whitman, 2020), virtues (Vallor, 2016), and principles (Floridi & Cowls, 2019; Mittelstadt, 2019; Mittelstadt et al., 2016). Predictive analytics, like all other AI and data analytics programs, inherit these same ethical issues.

[4] Treatments could be graded and the outcome variable could be split into three options (no treatment, light treatment, harsh treatment). This would complicate the calculation of the treatment effect I develop below and the confusion matrix in each of the figures. The critical and ethical analysis, however, would remain the same. Whether or not how outcome variable is cut into X categories is ethical or not would be based on analyzing the results matrix with the costs and benefits of each actor and whether the construction treats subjects as a mere means, deceives stakeholders into thinking the predictive analytics program is more accurate than it is, or exerts arbitrary power over the subjects.

[5] Courts and judges sentencing defendants do not pay for prisons. And for-profit prisons have been shown to create an incentive for putting more people in prison https://www.sentencingproject.org/publications/capitalizing-on-mass-incarceration-u-s-growth-in-private-prisons/ The cost of imprisonment is seen to be 'free' to COMPAS and the courts.

[6] For businesses and organizations, this phenomenon – where the transaction between two parties creates a harm to a third, less powerful party – is not uncommon. When a company sells steel to a car manufacturer, the community feels the harm of the pollution. And for decades, steel companies would not incorporate this environmental cost into the manufacturing design.

[7] "As the decision-maker acts according to a predictive model, the distribution over data points appears to change over time. In practice, the response to such distribution shifts is to frequently retrain the predictive model as more data becomes available. Retraining is often considered an undesired—yet necessary—cat and mouse game of chasing a moving target" (Perdomo et al., 2020).

[8]. e.g., while Black and White Americans sell and use drugs at similar rates, Black Americans were 6.5 times as likely as White Americans to be incarcerated for drug-related offenses. https://www.hamiltonproject.org/charts/rates_of_drug_use_and_sales_by_race_rates_of_drug_related_criminal_justice

[9] A classic forecasting optimization case for MBAs would be how to forecast the number of CDs to produce with the opportunity to sell a CD as a large benefit ($12) and the cost to throw away an unsold CD as minimal ($0.50). Businesses had an incentive to produce 'extra' CDs in the possible hopes of a future sale since the costs are minimal. In 1997, the case would be taught as if the cost to throw away the CD was minimal because we did not consider the community or environment in the calculation.

[10] For Pettit, interference is arbitrary if it is done on the whim of the agent or, here, firm. Pettit says: "what makes an act of interference arbitrary, then – arbitrary in the sense of being perpetrated on an arbitrary basis? An act is

perpetrated on an arbitrary basis, we can say, if it is subject just to the arbitrium, the decision or judgment, of the agent; the agent was in a position to choose it or not choose it, at their pleasure" (Pettit, 1997, p. 55).

[11] I am thankful to XYZ for pointing out this connection to Pettit. See also Mink's chapter on oppression and AI systems (Mink, 2020).

[12] This domination turns to oppression when the measurements and optimization are defined in ways that correlate with socio-economic groups that are already systematically disadvantaged

References

Abdalla, M., & Abdalla, M. (2021). *The Grey Hoodie Project: Big tobacco, big tech, and the threat on academic integrity*. 287–297.

Barocas, S., Hardt, M., & Narayanan, A. (2018). *Fairness and machine learning: Limitations and Opportunities*.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*.

Barry, L., & Charpentier, A. (2020). Personalization as a promise: Can Big Data change the practice of insurance? *Big Data & Society*, *7*(1), 2053951720935143.

Bedi, N., & McGrory, K. (2020, November 19). Pasco's sheriff uses grades and abuse histories to label schoolchildren potential criminals. The kids and their parents don't know. *Tampa Bay Times*. https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/school-data/

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, *20*(1).

Belitz, C., Ocumpaugh, J., Ritter, S., Baker, R. S., Fancsali, S. E., & Bosch, N. (2022). Constructing categories: Moving beyond protected classes in algorithmic fairness. *Journal of the Association for Information Science and Technology*.

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons.

Birhane, A. (2021). The Impossibility of Automating Ambiguity. *Artificial Life*, *27*(1), 44–61.

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2021). The Values Encoded in Machine Learning Research. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 173–184. https://arxiv.org/abs/2106.15590

Carnevale, A., Tangari, E. A., Iannone, A., & Sartini, E. (2021). Will Big Data and personalized medicine do the gender dimension justice? *Ai & Society*, 1–13.

Cevolini, A., & Esposito, E. (2020). From pool to profile: Social consequences of algorithmic prediction in insurance. *Big Data & Society*, *7*(2), 2053951720939228.

Citron, D. K. (2007). Technological due process. *Washington University Law Review*, *85*, 1249.

Citron, D. K., & Franks, M. A. (2014). Criminalizing revenge porn. *Wake Forest L. Rev.*, *49*, 345–391.

Coase, R. H. (1960). Problem of social cost, the. *Journal of Law and Economics*, *3*, 1.

Cohen, J. E. (2019). *Between truth and power*. Oxford University Press.

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415.

Dieterich, W., Mendoza, C., & Brennan, T. (2016). *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*. Northpointe Inc Research Department. http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. Mit Press.

Eidelson, B. (2015). *Discrimination and disrespect*. Oxford University Press.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Floegel, D., & Costello, K. L. (2022). Methods for a feminist technoscience of information practice: Design justice and speculative futurities. *Journal of the Association for Information Science and Technology*, *73*(4), 625–634.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Issue 1.1, Summer 2019*, *1*(1).

Gibson, A. N., & Martin III, J. D. (2019). Re-situating information poverty: Information marginalization and parents of individuals with disabilities. *Journal of the Association for Information Science and Technology*, *70*(5), 476–487.

Gill, K. S. (2020). *Prediction paradigm: The human price of instrumentalism*.

Hälterlein, J. (2021). Epistemologies of predictive policing: Mathematical social science, social physics and machine learning. *Big Data & Society*, *8*(1), 20539517211003120.

Henderson, G. R., & Williams, J. D. (2013). From exclusion to inclusion: An introduction to the special issue on marketplace diversity and inclusion. *Journal of Public Policy & Marketing*, *32*(1_suppl), 1–5.

Hildebrandt, M. (2008). Defining profiling: A new type of knowledge? In *Profiling the European citizen* (pp. 17–45). Springer.

Hoffmann, A. L., Roberts, S. T., Wolf, C. T., & Wood, S. (2018). Beyond fairness, accountability, and transparency in the ethics of algorithms: Contributions and perspectives from LIS. *Proceedings of the Association for Information Science and Technology*, *55*(1), 694–696.

Hsieh, N. (2004). The obligations of transnational corporations: Rawlsian justice and the duty of assistance. *Business Ethics Quarterly*, *14*(4), 643–661.

Hsieh, N. (2005). Rawlsian justice and workplace republicanism. *Social Theory and Practice*, *31*(1), 115–142.

Huang, S., Yang, J., Fong, S., & Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Letters*, *471*, 61–71.

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, *3*(2), 2053951716674238.

Johnson, G. (In Press). Are algorithms value-free? Feminist theoretical virtues in machine learning. *Journal Moral Philosophy*. https://philpapers.org/rec/JOHAAV

Keats Citron, D. (2018). Sexual privacy. *Yale LJ*, *128*, 1870.

Kerr, I., & Earle, J. (2013). Prediction, preemption, presumption: How big data threatens big picture privacy. *Stan. L. Rev. Online*, *66*, 65.

Kumar, V., & Garg, M. (2018). Predictive analytics: A review of trends and techniques. *International Journal of Computer Applications*, *182*(1), 31–37.

Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016, May 23). How We Analyzed the COMPAS Recidivism Algorithm. *ProPublica*. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Leavy, S., O'Sullivan, B., & Siapera, E. (2020). Data, Power and Bias in Artificial Intelligence. *ArXiv Preprint ArXiv:2008.07341*.

Lee, M.-J. (2005). *Micro-econometrics for policy, program and treatment effects*. OUP Oxford.

Levy, K. E. (2015). The contexts of control: Information, power, and truck-driving work. *The Information Society*, *31*(2), 160–174.

Levy, K. E., & Johns, D. M. (2016). When open data is a Trojan Horse: The weaponization of transparency in science and governance. *Big Data & Society*, *3*(1), 2053951715621568.

Lippert-Rasmussen, K. (2011). "We are all Different": Statistical Discrimination and the Right to be Treated as an Individual. *The Journal of Ethics*, *15*(1–2), 47–59.

Martin, K. (2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, *160*(4), 835–850.

Martin, K. (2022). Algorithmic Bias and Corporate Responsibility: How companies hide behind the false veil of the technological imperative. In K. Martin (Ed.), *Ethics of Data and Analytics*. Taylor & Francis.

McGory, K. (2021, July 24). Pasco Sheriff's Office letter targets residents for 'increased accountability.' *Tampa Bay Times*. https://www.tampabay.com/investigations/2021/07/24/pasco-sheriffs-office-letter-targets-residents-for-increased-accountability/

McGrory, K., & Bedi, N. (2020, September 3). Pasco's sheriff uses data to guess who will commit crime. Then deputies 'hunt down' and harass them. *Tampa Bay Times*. https://www.tampabay.com/news/pasco/2020/09/03/pascos-sheriff-uses-data-to-guess-who-will-commit-crime-then-deputies-hunt-down-and-harass-them/

Mink, K. (2020). *The Disciplinary Power of Algorithms: Domination, Agency and Resistance*.

Mittelstadt, B. (2016). Automation, Algorithms, and Politics| Auditing for Transparency in Content Personalization Systems. *International Journal of Communication*, *10*, 12.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*, *3*(2), 1–21.

Mohammad, S. M. (2021). Ethics Sheets for AI Tasks. *ArXiv Preprint ArXiv:2107.01183*.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. nyu Press.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.

Paris, B., Reynolds, R., & McGowan, C. (2022). Sins of omission: Critical informatics perspectives on privacy in e-learning systems in higher education. *Journal of the Association for Information Science and Technology*, *73*(5), 708–725.

Paullada, A. (2020). Machine Translation Shifts Power. *The Gradient*. https://thegradient.pub/machine-translation-shifts-power/

Perdomo, J., Zrnic, T., Mendler-Dünner, C., & Hardt, M. (2020). *Performative prediction*. 7599–7609.

Pettit, P. (1996). Freedom as antipower. *Ethics*, *106*(3), 576–604.

Pettit, P. (1997). *Republicanism: A theory of freedom and government*. Oxford University Press.

Poole, S., Grier, S., Thomas, F., Sobande, F., Ekpo, A., Torres, L., Addington, L., Henderson, G., & Weekes-Laidlow, M. (2020). Operationalizing critical race theory (CRT) in the marketplace. *Journal of Public Policy and Marketing*.

Regan, P. M., & Jesse, J. (2019). Ethical challenges of edtech, big data and personalized learning: Twenty-first century student sorting and tracking. *Ethics and Information Technology*, *21*(3), 167–179.

Seele, P., Dierksmeier, C., Hofstetter, R., & Schultz, M. D. (2021). Mapping the ethicality of algorithmic pricing: A review of dynamic and personalized pricing. *Journal of Business Ethics*, *170*(4), 697–719.

Skinner-Thompson, S. (2020). *Privacy at the Margins*. Cambridge University Press.

Solow-Niederman, A. (Forthcoming). Information Privacy and the Inference Economy. *Northwestern Law Review*, *117*(2), 1–68.

Stahl, B. C. (2021). Concepts of Ethics and Their Application to AI. *Artificial Intelligence for a Better Future*, 19.

Stroud, M. (2021). *Heat Listed*. https://www.theverge.com/22444020/chicago-pd-predictive-policing-heat-list

Susser, D. (2020). *Predictive policing and the ethics of preemption*.

Tang, R., Mehra, B., Du, J. T., & Zhao, Y. (2021). Framing a discussion on paradigm shift (s) in the field of information. *Journal of the Association for Information Science and Technology*, *72*(2), 253–258.

Thomas, R., & Uminsky, D. (2020). The problem with metrics is a fundamental problem for ai. *ArXiv Preprint ArXiv:2002.08512*.

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

Vedder, A. (1999). KDD: The challenge to individualism. *Ethics and Information Technology*, *1*(4), 275–281.

Waldman, A. E. (2019a). Law, privacy, and online dating:“Revenge porn” in gay online communities. *Law & Social Inquiry*, *44*(4), 987–1018.

Waldman, A. E. (2019b). Power, process, and automated decision-making. *Fordham L. Rev.*, *88*, 613.

Waldman, A. E. (2021). *Industry Unbound: The Inside Story of Privacy, Data, and Corporate Power*. Cambridge University Press.

Whitman, M. (2020). “We called that a behavior”: The making of institutional data. *Big Data & Society*, *7*(1), 2053951720932200.

Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. Sage.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power, 2019. *New York: Hachette Book Group*.