Journal of Data Science 20 (3), 381–399
July 2022
DOI: 10.6339/22-JDS1053
Data Science in Action

# Do Americans Think the Digital Economy is Fair? Using Supervised Learning to Explore Evaluations of Predictive Automation

Emilio Lehoucq[1,*]

[1]Department of Sociology, Northwestern University, 1810 Chicago Avenue, Evanston, IL 60208, USA

**Abstract**

Predictive automation is a pervasive and archetypical example of the digital economy. Studying how Americans evaluate predictive automation is important because it affects corporate and state governance. However, we have relevant questions unanswered. We lack comparisons across use cases using a nationally representative sample. We also have yet to determine what are the key predictors of evaluations of predictive automation. This article uses the American Trends Panel's 2018 wave ($n = 4,594$) to study whether American adults think predictive automation is fair across four use cases: helping credit decisions, assisting parole decisions, filtering job applicants based on interview videos, and assessing job candidates based on resumes. Results from lasso regressions trained with 112 predictors reveal that people's evaluations of predictive automation align with their views about social media, technology, and politics.

**Keywords** *algorithmic fairness; artificial intelligence; machine learning; public understanding of science and technology*

## 1 Introduction

We live in the "tech age" (Torpey, 2020). The technology industry has grown since the 1990s, far outperforming the economy at large (Oppenheimer, 2020, p. 190). The digitization of information and the rise of the Internet have opened new possibilities across sectors (Carlsson, 2004). The economy is now digital, meaning that all sectors are transformed by the computer-enabled digitization of information (Brynjolfsson and Kahin, 2000, p. 2). Predictive automation is a pervasive and archetypical example of the digital economy (Leavitt et al., 2021). Predictive automation refers to computer programs that mine data and make predictions to help decision-making (Zuboff, 2019). For example, companies use data about people's behavior and personal characteristics to help decide whether they are worthy of credit. Other organizations use the same kind of data to assess whether someone should be released in parole or not. Yet other corporations use resumes of job applicants to decide whether they deserve further review for hiring or use recordings of interviews with job candidates to assess whether they would be a good hire.

The norms that sustain the digital economy are changing. Since the mid 2010s, social media has been under increasing scrutiny. Facebook was involved in scandal after experimenting with users' emotions (News, 2014). The social media company was further involved in scandal for its relation with Cambridge Analytica during the 2016 United States election (González et al., 2019).

* Email: emilio-l@u.northwestern.edu.


Received December 2, 2021; Accepted May 26, 2022

Similarly, famous technology entrepreneur and venture capitalist Peter Thiel—Facebook's first outside investor and board member—jumped to mainstream attention by supporting Donald Trump (Chafkin, 2021). Trump itself was notable for his use of Twitter, and for the social media platform's decision to ban him (Clayton, 2022). In the meantime, activists have been busy mobilizing to protect privacy, leading to the General Data Protection Regulation (Lehoucq and Tarrow, 2020). Overall, the technology industry is increasingly politicized and faces regulatory risk. In this context, it is important to study how every day citizens evaluate technology. These evaluations are a key part of understanding governance. Corporations consider consumer demand and employee's preferences when making decisions about what technologies to use and how to do it. Similarly, government agencies assess public opinion when using and regulating technologies. In these two ways, everyday citizens' evaluations of technology matter.

The relation between people's evaluations of technology and governance is mediated and non-linear. However, if we want to understand the changing landscape of the digital economy, we need to understand how every day citizens evaluate technology. This article is a study in service of that end, focusing on evaluations of predictive automation, because this technology is pervasive and archetypical of the digital economy.

Despite the relevance of Americans' evaluations of predictive automation, we have important questions unanswered. First, there is mounting evidence that evaluations of predictive automation vary across use cases (Baleis et al., 2019; Kiviat, 2021). Few studies compare how Americans evaluate different uses of predictive automation (Araujo et al., 2020). Second, we are still learning the key predictors of evaluations of predictive automation, including demographics (Araujo et al., 2020; Gran et al., 2020; Grgić-Hlača et al., 2020; Wang et al., 2020) and broader views about technology (Araujo et al., 2020; Hidalgo et al., 2021; Lee, 2018; Lee et al., 2019; Newman et al., 2020; Wang et al., 2020; Woodruff et al., 2018). Finally, most scholars rely on experiments based on convenience samples (Gran et al., 2020; Grgic-Hlaca et al., 2018; Lee, 2018; Wang et al., 2020). Few studies use nationally representative data to explore Americans' evaluations of predictive automation (Kiviat, 2021).

This article uses a sample of 4,594 respondents representative of American adults to study how they evaluate the fairness of predictive automation across four use cases. The data comes from wave 35 of the Pew Research Center's American Trends Panel, which was fielded in 2018 and is publicly available. Results come from lasso regressions predicting whether Americans think the following use cases of predictive automation are fair: (1) a finance score to help decisions about credit, (2) a criminal risk score to assist decisions about parole, (3) a score to support decisions about whether a job applicant deserves further review based on interview videos, and (4) a score to support decisions about whether a job candidate would be a good hire based on resumes. Lasso regressions help select the most important predictors among 112 variables measuring demographics such as gender, age, and education; Americans' perceptions of technology and social media; and political views such as ideology and beliefs about the power of different groups.

The results reveal that Americans' evaluations of predictive automation align with their views about social media, technology, and politics. Americans who think that predictive automation is fair tend to be those that accept controversial social media practices, have a positive view of technology corporations, and lean conservative. On the contrary, Americans who think that predictive automation is unfair are usually those who disapprove of controversial social media practices, are skeptical of technology corporations, and lean liberal. Overall, these findings suggest that Americans are polarized when it comes to predictive automation.

## 2 Evaluations about the Fairness of Predictive Automation

There are different kinds of evaluations, such as fairness and trust. Scholars of judgment and decision making have mostly focused on evaluations of trust (Dawes, 1979; Dietvorst et al., 2015; Logg et al., 2019; Meehl, 1954). Trust refers to accepting vulnerability based on positive expectations about another actor (Rousseau et al., 1998). In contrast, this article focuses on evaluations of fairness since fairness is often the main moral claim in discussions about economic practices, criminal justice, and hiring (Kiviat, 2021, p. 42). Trust and fairness are certainly related: we are less likely to trust those who violate our sense of fairness. But fairness goes beyond trust. Fairness is a moral evaluation about who owes what to whom (Haidt and Kesebir, 2010; Wilson, 1997).

Perhaps most striking is the fact that Americans overwhelmingly think predictive automation is unfair. They tend to reject citizen scoring, particularly when it touches on credit (Hidalgo et al., 2021) and criminal justice (Grgic-Hlaca et al., 2018). Americans tend to negatively judge the use of automation in job interviews (Langer et al., 2019). The predictive technologies used in hiring decisions are perceived as unfair because they are reductionistic and leave important qualitative information aside (Newman et al., 2020). Given this perception, Americans judge predictive technologies more harshly than humans when they make mistakes in hiring (Hidalgo et al., 2021).

To understand the changing foundations of the digital economy, we need to further disaggregate evaluations of fairness. It is not enough to know that most Americans think predictive automation is unfair. Do Americans think that only based on technical features of predictive automation? Are there broader views about the production of technology that matter? Do usage and perceptions about social media—one of the most salient and controversial businesses within technology—matter? Are there differences across demographics such as race and gender? Do people's political views also factor in? These are all questions we need to answer to understand how evaluations of fairness can impact the future of the digital economy.

Technical capabilities matter for how people evaluate the fairness of predictive automation. People think predictive technologies are fair in the case of mechanical tasks because they are efficient and objective (Lee, 2018). Does this generalize beyond mechanical tasks? This article considers whether people are more likely to think predictive automation is fair when they think it is effective. But effectiveness is only part of the issue.

There is also evidence that computer literacy increases the perception that predictive technologies are fair (Wang et al., 2020). It is not yet clear whether people care about if they understand how predictive technologies work (Binns et al., 2018; Dodge et al., 2019; Kizilcec, 2016). They do seem to care about whether they can control the decision suggested by predictive technologies (Lee et al., 2019). One particularly important outcome is unequal impact. The use of predictive automation could in fact reinforce race and gender stratification (Angwinn et al., 2016; Benjamin, 2019; Harcourt, 2008; O'Neil, 2017; Eubanks, 2018; Ferguson, 2017; Noble, 2018). Do Americans care about this when evaluating the fairness of predictive automation?

Biases in predictive technologies hint at a broader issue. The technical capabilities of technologies are often as important as the way in which those technology are produced. Do Americans care about the production of technology? The answer seems to be yes, but we lack a deeper understanding. People who have concerns about privacy are more prone to think that predictive technologies are unfair (Araujo et al., 2020). But do Americans care beyond privacy?

Qualitative studies have found that once people learn about the fairness of predictive automation, they tend to have negative feelings and think differently about technology corporations

(Woodruff et al., 2018). This article extends this line of inquiry by exploring the relation between evaluations about the fairness of predictive automation and whether people think technology corporations 1) can be trusted, 2) are ethical, 3) should be more regulated, 4) have a good or bad impact, 5) fail to anticipate their impact, 6) benefit the privileged, 7) are biased toward men, and 8) support the views of liberals.

Of course, not all technology companies are the same. As previously explained, social media companies—notably Facebook—have been at the core of the scandal surrounding predictive technologies since the mid 2010s. Indeed, social media platforms use predictive technologies to create user experiences, something people have strong feelings about (Bucher, 2017). For example, while many people are unaware of how predictive technologies curate Facebook's News Feed, they are surprised and angry when they find about it (Eslami et al., 2015). Is controversy around social media driving Americans' evaluations about the fairness of predictive automation? This article explores the relation between people's engagement with—and perceptions of—social media and their evaluations about the fairness of predictive automation.

So far, I have considered technical features of predictive technologies and whether people care about their production more broadly. This gets us one step closer to understanding the changing foundations of the digital economy. However, given the pervasiveness of technology, people's evaluations about the fairness of predictive automation could run deeper. How people evaluate the fairness of predictive automation could be entangled with the core of who they are.

Yet, we do not even know who thinks predictive automation is (un)fair. Existing studies have started to explore the relation between demographics and evaluations about the fairness of predictive automation. But findings are not conclusive. There are contradictory studies about the relation between perceptions about the fairness of predictive automation and gender (Gran et al., 2020; Grgić-Hlača et al., 2020; Wang et al., 2020), age (Araujo et al., 2020; Gran et al., 2020; Grgić-Hlača et al., 2020), and education (Gran et al., 2020; Wang et al., 2020). While gender, age, and education are important, we need to explore other demographics as well. What aspects of people's lives are related to how they evaluate the fairness of predictive automation? This article considers race, marital status, religious affiliation and attendance, family income, civic participation, and census region.

Finally, there is some evidence that people's political views factor into how they evaluate predictive automation. Conservatives seem to be more likely than liberals to think that predictive technologies used in criminal justice are fair (Grgić-Hlača et al., 2020). This is an important finding because political polarization could create partisanship around predictive automation and affect the governance of the digital economy. We thus need to explore how politics factors into people's evaluations about the fairness of predictive automation. Are conservatives more likely to think that predictive automation is fair beyond criminal justice? Is it political ideology or partisan affiliation that matters? This article gets at these questions by considering partisan identity, ideology, and what people think about the power of technology corporations, finance, unions, and other organized groups.

In sum, we care about how Americans evaluate the fairness of predictive automation because this is a key part of governance. Given the relevance of the issue, this article expands previous research. It starts by considering how technical capabilities of predictive automation matter for how people think about its fairness. This study further broadens the view, investigating how the production of technology is related to evaluations about the fairness of predictive automation. Lastly, this article goes one step further in assessing not only technology and its production, but people's core lives by considering demographic characteristics and political views. Overall, this exploration helps us tackle a key piece of the changing landscape of the digital economy.

# 3 Data and Methods

## 3.1 Survey

This article draws on a survey of 4,594 respondents representative of adults living in households in the United States. GfK Custom Research fielded the survey between May 29th and June 11th, 2018, on behalf of the Pew Research Center as part of the American Trends Panel. The Pew Research Center developed the questionnaire in consultation with GfK and publicly released the survey at the end of 2019. Respondents were recruited from three large, national, overlapping, dual-frame landline and cellphone random-digital-dial (RDD) surveys conducted for the Pew Research Center in 2014, 2015, and 2017. At the end of each RDD survey, respondents were invited to join a panel. Respondents with a known residential address received a postcard with the survey. Respondents with an email address received an email invitation and up to four email reminders. All respondents that consented to SMS messages received an SMS invitation and up to four SMS reminders.

## 3.2 Responses

Each respondent saw two use cases of predictive automation, presented in random order. The four use cases were (1) a finance score for credit decisions, (2) a criminal risk score for parole decisions, (3) a score based on interview videos to filter job applicants, and (4) a score to assess job candidates based on resumes. These scenarios captured use cases of predictive automation that are either common or are the subject of current public policy debate.

In each scenario, respondents read a vignette explaining the use case, the data mined, and how a computer program is used for automation. One of the items, about finance score, captured respondents' answers to the following prompt:

> Companies have developed automated programs to calculate a new type of personal finance score, similar to a credit score. These programs collect information from many different sources about people's behavior and personal characteristics – such as their online habits or the products and services they use. They then assign people an automated score that helps businesses decide whether to offer them loans, special offers or other services. // How FAIR do you think this type of program would be to consumers?

The second item, about criminal risk score, measured responses to this scenario:

> Companies have developed automated programs to calculate a new type of criminal risk score for people in prison who may qualify for parole. These programs collect information from many sources about a person's past behavior and personal characteristics. They then compare this data to others who have been convicted of crimes, and assign a score that helps decide whether someone should be released on parole or not. // How FAIR do you think this type of program would be to people in parole hearings?

The third item captured respondents' answers to this prompt:

> In an effort to improve the hiring process, some companies are now recording interviews with job candidates. These videos are analyzed by a computer, which matches

Table 1: Distribution of the responses before re-coding.

| | Very fair | Somewhat fair | Not very fair | Not fair at all | Missing | Total |
|---|---|---|---|---|---|---|
| Finance score | 4% | 25% | 37% | 34% | 1% | 2279 |
| Criminal risk score | 7% | 40% | 35% | 17% | 1% | 2315 |
| Video mining | 5% | 29% | 39% | 27% | 1% | 2320 |
| Resume mining | 6% | 36% | 37% | 21% | 0% | 2274 |

> the characteristics and behavior of candidates with traits shared by successful employees. Candidates are then given an automated score that helps the firm decide whether or not they might be a good hire. // How FAIR do you think this type of program would be to people applying for jobs?

Finally, the item about mining resumes for hiring captured answers to this prompt:

> In an effort to improve the hiring process, some companies are now using computers to screen resumes. The computer assigns each candidate an automated score based on the content of their resume, and how it compares with resumes of employees who have been successful. Only resumes that meet a certain score are sent to a hiring manager for further review. // How FAIR do you think this type of program would be to people applying for jobs?

Overall, the vignettes designed by the Pew Research Center in consultation with GfK reflected a range of practices in predictive automation.

The Pew Research Center and GfK randomly assigned respondents two vignettes. Approximately half of the respondents were assigned each set of vignettes. Respondents who were assigned the vignettes about finance and criminal risk were also assigned the vignette about hiring. This implies that responses about finance and criminal risk can serve as predictors for those about hiring and vice versa. However, responses about criminal risk cannot serve as predictors for those about finance and responses about mining videos cannot serve as predictors to those about mining resumes. I divided the full dataset into four datasets (one for each response) and kept the observed responses for analysis.

After reading each vignette, respondents were asked whether the use case of predictive automation was "very fair," "somewhat fair," "not very fair," and "not fair at all." Respondents evaluated one use case at a time, presented in random order. The survey captures quick-response evaluations rather than more deliberative moral reasoning (Lizardo et al., 2016; Vaisey, 2009). Table 1 displays the distribution of the responses. Very few people answered "very fair," and relatively few answered "not fair at all." These responses cannot be treated as continuous since the shift from one category to another is not a one-unit change. Further, there is not enough data to train classification models with 112 predictors and four response categories. After exploratory data analysis, I recoded all responses as dummies (0 "fair," 1 "unfair"). While recoding the responses created some loss of information, it enabled the investigation of the patterns of associations between predictors and responses.

### 3.3 Predictors

I trained all models using most variables in the survey (112 predictors). This is an unusually large number of predictors for the social sciences. Studies typically focus on a handful of predictors that researchers select based on previous knowledge. I too use predictors for which there is a plausible substantive relation with the response. Section 2 above explains why the capabilities of predictive automation, broader views about technology and social media, demographics, and political views are relevant. However, I use a different method for variable selection. The common approach in the social sciences requires that researchers have enough a priori knowledge to select variables and specify a correct functional form. This is a strong assumption in exploratory research. Since Americans' evaluations about the fairness of predictive automation is still an exploratory domain, I use a more inductive approach. As I explain in more detail below, I use models that require less a priori knowledge and help in variable selection.

I explored every variable in the survey. The supplementary material explains how each variable is measured and processed. All variables in the survey are categorical. During exploratory data analysis, I collapsed some categories to avoid sparsity. I dropped seven predictors that created sparsity (as noted in the supplementary material). I also dropped six open questions and a measure about whether Americans think it is acceptable to use predictive automation. Views about fairness should correspond to perceptions about whether it is acceptable to use predictive automation (Wilson, 1997; Woodruff et al., 2018). These variables are indeed highly associated. Before recoding, the correlations are 0.63 for finance score, 0.67 for risk score, 0.68 for mining videos, and 0.67 for mining resumes. Most importantly, these associations are not informative of what people think about the fairness of predictive automation.

While there are missing values, they do not rise to a level of concern. No variable has more than three percent of non-random missing values. The survey randomized the assignment of 12 questions to only some respondents. An additional 22 questions were only asked based on respondents' answer to a previous question. Both survey design features create missing values. To avoid losing observations, I recoded missing observations for the predictors as "No response." This recoding adds an additional category to every predictor with missing values. If missingness is random, "No response" should not be an important predictor.

### 3.4 Modeling

Following common practices in supervised learning (Hastie et al., 2009; James et al., 2013), I trained six models to predict each of the responses and selected the models that best fit the data. Supervised learning is useful to explore a broad range of predictors and possible functional forms (Molina and Garip, 2019; Mullainathan and Spiess, 2017; Varian, 2014). Results should not be interpreted as inferring the correct relation between predictors and responses or as causal (Breiman, 2001). I trained models that could (1) perform well with a high number of predictors, (2) model potential non-linearities, and (3) be interpreted in terms of which predictors are more important and how they relate to the response.

I fitted logistic regressions to predict each of the (binary) responses. However, the coefficients of logistic regressions are likely to suffer from excessive variance because of the high number of correlated predictors (James et al., 2013, p. 218). To account for multicollinearity, I estimated ridge and lasso regressions. Ridge and lasso regressions shrink coefficients, which introduces bias but reduces variance and hence potentially increases the accuracy of the model (Hastie et al., 2009, pp. 61–93). Lasso regression is also useful to select the most important predictors since it can shrink coefficients to zero. However, ridge and lasso regressions cannot flexibly model potential non-linearities. I therefore used tree-based methods—specifically, random forests, boosted trees, and Bayesian additive classification trees—to model potential non-linearities. Tree-based methods also deal efficiently with categorical predictors (James et al., 2013, pp. 303–323).

With binary outcomes, both ridge and lasso regressions are fitted by minimizing the deviance and a term shrinking the size of the coefficients. Ridge regression adds the term $\lambda \sum_{j=1}^{p} \beta_j^2$, while Lasso adds the term $\lambda \sum_{j=1}^{p} |\beta_j|$, where $\lambda$ is a shrinkage parameter selected through cross validation to minimize the deviance and $p$ is the number of predictors. Ridge and lasso differ in how they shrink the size of the coefficients, $\beta$. Lasso shrinks some coefficients to zero and has the advantage of performing variable selection, while Ridge only makes them converge toward (but not reach) zero.

Classification trees recursively partition the predictor space and assign the most commonly occurring class of observations in a terminal node, thus taking the form $f(X) = \sum_{m=1}^{M} c_m x 1_{(x \in R_m)}$, where $R_1, \ldots, R_M$ represent the partitions of the predictor space and $c_m$ the most commonly occurring class. Trees have the advantage of performing variable selection, since only the predictors that most reduce the Gini index are used in growing the tree. Random forests average across the predictions of trees grown on bootstrap samples of the dataset. Also, any single tree is constrained to be grown on a subset of predictors, which decorrelates the resulting trees and improves the accuracy of the model. On the contrary, boosted trees are grown sequentially on the same data—each tree is grown using the residuals from the previous tree, which improves the fit. Finally, Bayesian additive classification trees consist of a set of priors for the structure of the trees and the leaf parameters and a likelihood for the data in the terminal nodes (Chipman et al., 2010). For binary outcomes, Bayesian additive classification trees assume a probit model, in which the model estimates the conditional probit given the predictors.

I used `R` to train all models. Ridge and lasso regressions are fitted with the function `cv.glmnet()` from the packages *glmnet* and *glmnetUtils*. Random forests and boosted trees are estimated with the function `train()` from the package caret (which uses the packages *randomForest* and *gbm*, respectively). Bayesian additive classification trees are fitted with the function `bartMachineCV()` from the package *bartMachine*. I tuned the hyperparameters that most impact each of the models. Specifically, in ridge and lasso regressions, the shrinkage parameter; in random forests, the number of predictors to split at each node; and in boosted trees, the number of trees and the learning rate.

I used cross validation to tune and compare the models. Cross validation is a family of resampling methods to estimate the error rate of models and avoid overfitting (James et al., 2013, pp. 176–186). I split each dataset into a training set for model tuning and a test set for model comparison (80% and 20% of the data, respectively; $n_{financescoretuning} = 1812$, $n_{riskscoretuning} = 1830$, $n_{videohiringtuning} = 1845$, $n_{cvhiringtuning} = 1813$, $n_{financescoretesting} = 453$, $n_{riskscoretesting} = 458$, $n_{videohiringtesting} = 461$, and $n_{cvhiringtesting} = 453$). I tuned all models with 5-fold cross validation on the training set and compared all models at the end on the test set. I used five folds since each training set is rather modest and there is a high number of categorical predictors, which could create sparsity with more folds. Some of the 112 predictors have more than two categories. After hot encoding there are 370 binary predictors.

Figure A1 in the online appendix reports the test misclassification rates. All models improve over a naïve classifier. I selected the lasso regressions. Logistic and ridge regressions perform worse across responses. Lasso regression, random forests, boosted trees, and Bayesian additive classification trees perform similarly. K-fold cross validation relies on random resampling. This implies different iterations will produce somewhat different results. Emphasizing small differences is not a good practice, since they may be caused by randomness. I selected the lasso regressions because they perform similarly than tree-based methods but are more interpretable.

# 4 Do Americans Think the Digital Economy is Fair?

## 4.1 Interpreting Models and Sensitivity Analyses

Before diving deep into the results, it is important to clarify how models should be interpreted and their robustness. One of the advantages of lasso regressions is that they can be trained on many correlated predictors. However, there is still multicollinearity: many predictors are a linear combination of other predictors in the model (James et al., 2013, p. 243). This creates uncertainty about which variables are truly predictive of the response. Models trained on an independent dataset could be different. The model obtained is valuable, but the results should not be overstated. Results come from one among several possible models.

Further, there is a special consideration for this study. Americans' views about the effectiveness of predictive automation are by far the most important predictor of perceptions of fairness. Before recoding, the correlations between effectiveness and evaluations about the fairness of predictive automation are 0.48 for finance score, 0.72 for risk score, 0.71 for video mining, and 0.71 for resume mining. These strong associations could be soaking most of the variation, which may hide other patterns in the data. I thus fitted all models excluding effectiveness as a predictor. Figures A2–A6 in the online appendix report the results. Models trained without effectiveness have a worse predictive performance. Yet, the conclusions are robust, and I thus interpret the models using effectiveness as a predictor.

I use two strategies to interpret the models. First, I only interpret the predictors selected both by the initial models and the sensitivity analyses. These predictors are more likely to be truly predictive of the response. Second, I focus on the predictors with larger coefficients (in absolute value). Controlling for other factors, these predictors have a stronger association with—and a higher likelihood of being truly predictive of—the response.

Overall, these strategies mean I do not interpret all 112 predictors. I used models that help me inductively select the most important predictors. I should not consider all predictors, but only those that help increase the accuracy of models. In addition, given the issue of multicollinearity, I focus on predictors that are selected by both the initial models and the sensitivity analyses, as well as those that have larger coefficients. This helps me center on the key predictors—those that are likely to matter for future studies using independent datasets.

## 4.2 Differences Across Cases: Fragmented Views of the Digital Economy?

As mentioned in the introduction, few studies compare how Americans evaluate uses of predictive automation (Araujo et al., 2020). This article explores the use of predictive automation across four use cases: assisting with credit decisions, assisting with parole decisions, filtering job applicants based on interview videos, and assessing job candidates based on resumes. This comparison is important for the digital economy, which is composed of different sectors and spans a wide variety of use cases. If we are to explore how people view the changing landscape of the digital economy and what effects those views have broadly, we must study the degree to which there could be fragmentation. With this objective in mind, this section explains the major differences across use cases.

At the core, there are differences across use cases in how (un)fair Americans believe predictive automation to be. Figure 1 reports 95% confidence intervals for the responses. In line with previous literature, most Americans think predictive automation is unfair. While this is not surprising, it is relevant. Predictive automation is at the core of and archetypical of the digital economy. If most people believe it is unfair, it is likely to have consequences for the long-term

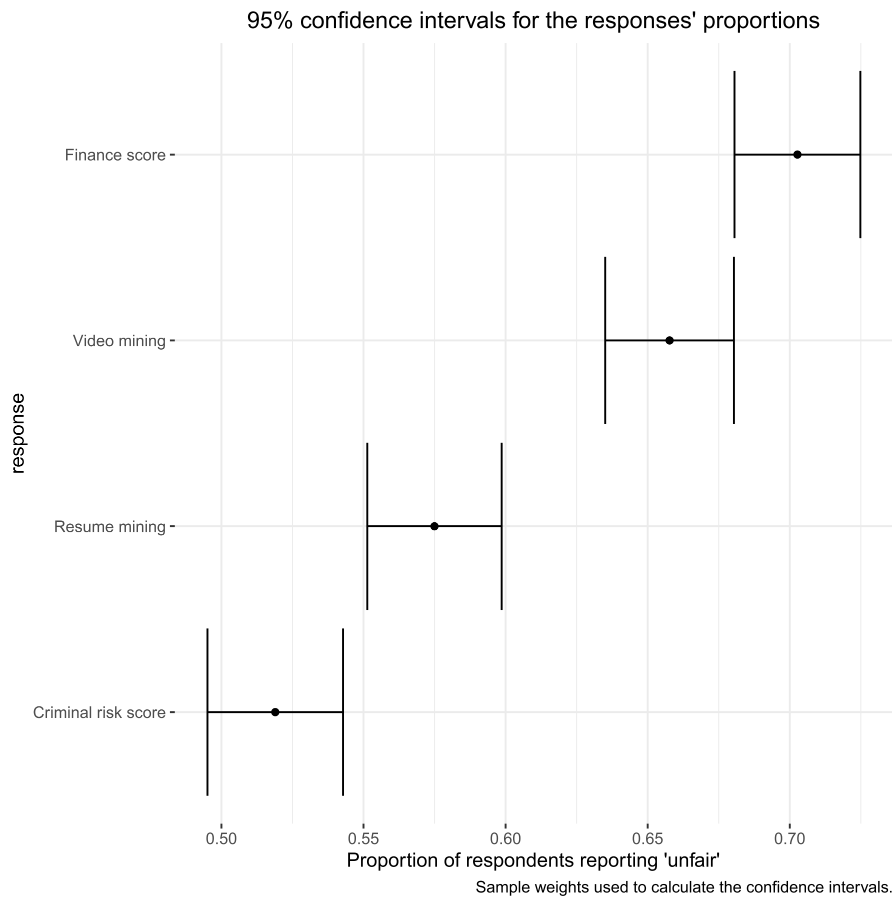


Figure 1: Confidence intervals for the responses' proportions.

viability of the digital economy. The possibility of a new automated finance score raises the more concerns. On the contrary, a criminal risk score for parole decisions seems to worry Americans the least. Finally, Americans are more skeptical about filtering job applicants with a score based on mining videos than they are about assessing job candidates based on mining resumes.

These key differences extend further. The most important predictors of Americans' evaluations about the fairness of predictive automation vary across use cases. Figures 2–5 present the results from lasso regressions. The first interesting difference concerns the technical capabilities of predictive automation. A key debate concerning predictive technologies is how biases creep into them, potentially leading to unequal outcomes (Angwinn et al., 2016; Benjamin, 2019; Eubanks, 2018; Ferguson, 2017; Harcourt, 2008; Noble, 2018; O'Neil, 2017). Based on the academic and mediatic interest, one would expect Americans to be majorly concerned with possible biases in predictive technologies and their consequences. Yet, this does not seem to be the case.

This study uses a measure of whether people think computer programs can avoid human biases. However, this measure only seems to be relevant in the case of a finance score to make decisions about credit. Lasso regressions did not select it in the case of risk scores assisting parole decisions, filtering job applicants based on video interviews, or assessing job candidates based on resumes. This is important for what it could mean for the ethics and regulation of predictive technologies (Dubber et al., 2020). If Americans are only concerned about biases in

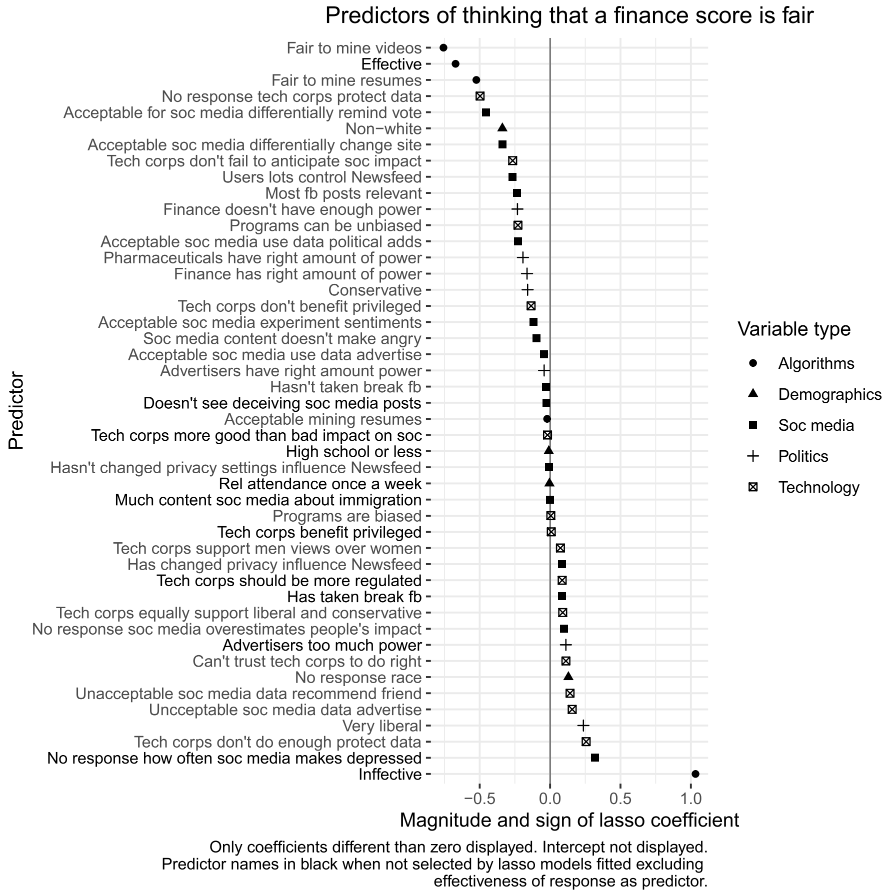


Figure 2: Results of lasso regression predicting whether people think a finance score to help credit decisions is fair (0 coded as “fair,” 1 as “unfair”).

certain cases, we could see the development of a fragmented regulatory and ethical landscape.

Differences across use cases run even deeper. Americans vary in how they evaluate the fairness of predictive automation beyond its technical capabilities. There are some political differences. Again, the use of predictive automation to create finance scores for credit decisions is the most interesting scenario. This article includes measures of people’s beliefs about the power of organized groups such as financial institutions, labor unions, advertisers, and small businesses. These measures are important to predict Americans’ beliefs about the fairness of predictive automation only in two use cases. In the first case, people who believe financial institutions do not have enough power and influence are more likely to think that using predictive automation to create finance scores is fair. On the other hand, those who think that small businesses have too much power are more likely to believe that risk scores for parole decisions are fair. This relation is unexpected and requires more research.

Finally, Americans’ evaluations about the fairness of predictive automation also vary across use cases in terms of demographics. Most notably, race does not seem to be a key predictor. Lasso regressions did not select race as an important predictor for risk scores to help decisions about parole and for mining videos to filter job applicants. Also, there is no evidence of a racial divide. Counterintuitively, and despite historical differential access to financial services between

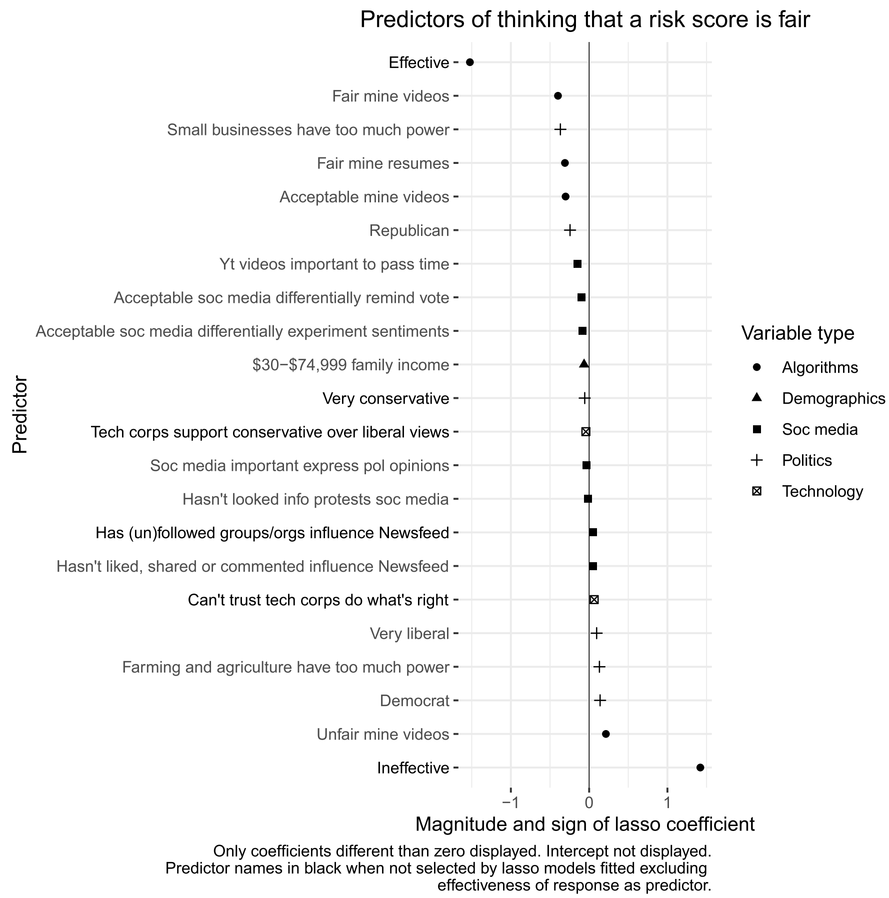


Figure 3: Results of lasso regression predicting whether people think a risk score to help decisions about parole is fair (0 coded as "fair," 1 as "unfair").

white and nonwhite people, being nonwhite is associated with thinking that finance scores for credit decisions are fair. The association of fairness is reversed in the case of mining resumes to assess job candidates. Nonwhite people are more likely to believe that mining resumes to assess job candidates is unfair. This is a more intuitive finding, but the difference between finance scores and mining resumes is puzzling and deserving of further research. Overall, race does not seem to be a predictor of a positive or negative view of predictive automation,

Finally, this article casts doubts about the importance of some of the demographic characteristics suggested by previous literature. Past studies suggest gender (Gran et al., 2020; Grgić-Hlača et al., 2020; Wang et al., 2020), age (Araujo et al., 2020; Gran et al., 2020; Grgić-Hlača et al., 2020), and education (Gran et al., 2020; Wang et al., 2020) are important predictors, although they are contradictory in terms of the direction of the relation. On the contrary, lasso regressions do not select age or gender as key predictors. Further, education was only selected in the case of mining videos to filter job applicants. People who think mining videos to filter job applicants is unfair are more likely to either have been only to high school or have less education. This is interesting in terms of the broader digital economy, which has differential effects across educational groups. Yet, these differences do not seem to be a key factor related to how people think about the fairness of predictive automation across use cases.

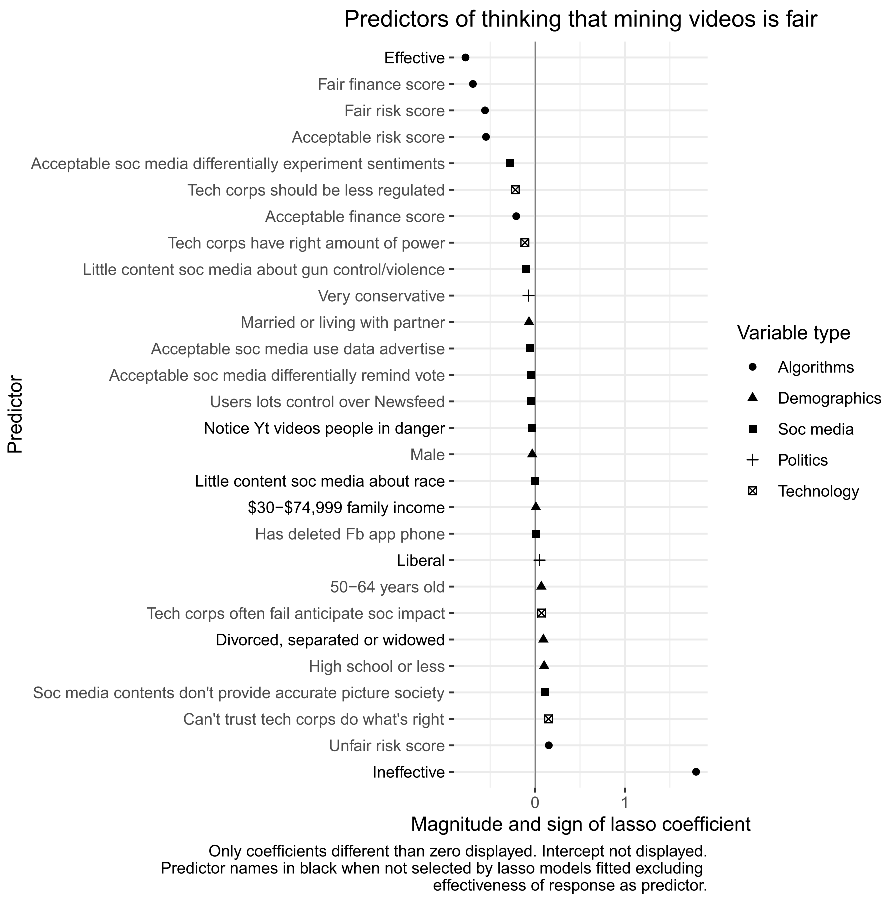


Figure 4: Results of lasso regression predicting whether people think mining videos to filter job applicants is fair (0 coded as "fair," 1 as "unfair").

Summing up, this section compared Americans' evaluations about the fairness of predictive automation across four use cases. While there are differences, there does not seem to be major fragmentation. People differ in terms of the importance attributed to technical capabilities of predictive automation. Similarly, political views matter somewhat differently across use cases. Finally, while demographics do not seem to be the most important predictors, some are relevant. This variation matters for the changing landscape of the digital economy, but the next section shows that there is converge across use cases.

### 4.3 Similarities Across Use Cases: Division in the Digital Economy?

The previous section focused on the major differences in how Americans evaluate predictive automation across use cases. But there are also similarities. In fact, evaluations across different use cases are related to each other. This suggests Americans think in a somewhat unified fashion about the digital economy rather than brining a unique repertoire of beliefs to each use case. For the most part, people who think finance and risk scores are (un)fair are also more likely to believe that mining resumes and videos is (un)fair.

The similarities in how Americans evaluate predictive automation are meaningful in what they tell us about the digital economy. Results suggest a divide among Americans. People who


Figure 5: Results of lasso regression predicting whether people think mining resumes to assess job candidates is fair (0 coded as "fair," 1 as "unfair").

think predictive automation is fair are more likely to be conservative, accept controversial social media practices, and hold technology corporations in a positive light. On the contrary, Americans who perceive predictive automation as unfair tend to be liberals who are skeptical of controversial social media practices, and have a more negative view of technology corporations. The rest of this section unpacks this division.

Americans who are more favorable towards social media tend to think that predictive automation is fair. Those who believe it is acceptable for social media to treat some users and not others differentially and exploit users' data to send targeted advertisements are more likely to perceive finance scores and mining resumes to assess job candidates as fair. Users who trust predicted finance scores are more likely to see Facebook as giving them high control over the Newsfeed and assess most posts they encounter as relevant. Similarly, Americans who consider it appropriate for social media to experiment with users' sentiments also tend to think that mining resumes and videos is fair. Additionally, believing that mining resumes to assess job candidates is fair is correlated with thinking that technology corporations protect users' data.

These associations suggest that satisfaction with the production of digital technologies goes hand in hand with evaluating their fairness. Process matters. If people believe that technology corporations protect users' data, anticipate their social impact, and are well regulated, they are more likely to judge favorably predictive automation.

People are harsher in judging digital technologies when they do not agree with how they are produced. For example, Americans who think a finance score to help decisions about credit is fair also believe that technology corporations do a good job at anticipating their social impact. Similarly, those who perceive mining videos to filter job applicants as fair are more prone to judge that technology corporations should be more regulated. Americans who think a finance score to help decisions about credit is unfair tend to believe that technology corporations do not adequately protect users' data. Along the same lines, people who perceive mining videos to filter job applicants as unfair tend to distrust technology companies. Finally, those who think mining resumes to assess job candidates is unfair tend to believe that technology corporations have a negative social impact.

These findings about social media are important for what they tell us about the digital economy. It is not obvious that specific beliefs about social media relate to how people think about predictive automation generally. The fact that they do is interesting. It suggests that Americans may be thinking of social media as an archetypical case of predictive automation. This would be impactful given how controversial social media has become.

Finally, Americans are politically divided in how they evaluate the fairness of predictive automation. Popular politicians have argued that technology corporations are biased against conservatives. As an example, consider the controversy surrounding Trump's ban from Twitter (Clayton, 2022). This could suggest that conservatives are more likely to think predictive automation is unfair, but this does not seem to be the case. While there is political polarization, it seems to go the other way around. Americans who think a finance score to help decisions about credit is unfair tend to be very liberal. Those who believe that risk scores are unfair area also very liberal or democrat. And people who think mining videos to filter job candidates is unfair are often liberal. On the contrary, conservatives are likely to believe that using finance scores to make credit decisions is fair and to perceive mining videos to filter job candidates as fair.

In a nutshell, section 4.2 showed that there is some variation in how Americans think about the fairness of predictive automation across use cases. These are differences in the importance of the technical capabilities, in political views, and the importance of demographics. These differences are relevant for what they could imply in terms of a fragmented regulatory and ethical landscape, and thus deserve more attention by future research. However, this section showed major similarities. These similarities are important for what they tell us about the digital economy. It seems that Americans are divided about how they evaluate the fairness of predictive automation. This division extends to people's views about the production of technology, social media, and politics. The next section discusses why we should care about this.

## 5 Discussion

While American liberals and conservatives have historically both been optimistic about technology and both been aligned with Silicon Valley (O'Mara, 2019), bipartisan agreement may be disappearing. In line with previous studies (Grgić-Hlača et al., 2020), I find that liberals seem to view predictive automation more negatively than conservatives. Generally, conservative Americans accept controversial social media practices, have a positive perception of technology companies, and tend to view predictive automation as fair. At the other pole, liberal Americans, who reject controversial social media practices and are skeptical of technology companies, are more likely to think that predictive automaton is unfair.

In academic fields, the conversation around this practice continues to grow, in fields such as the ethics of artificial intelligence (Dubber et al., 2020). In public discourse, the debate around predictive automation often relates to social media, which has been especially controversial for its data extraction and data selling practicing (Zuboff, 2019). For Americans, social media seems to be an archetypical case to think about predictive automation. Documentaries (like *The Social Dilemma* and *The Great Hack*) and stylized dramas (like *The Social Network*, *Catfish*, and episodes of BBC's *Black Mirror*) further bring social media companies and predictive technologies such as artificial intelligence to the public debate.

Previous studies have found that people with concerns about privacy are more skeptical of predictive technologies (Araujo et al., 2020). This article similarly finds that Americans who believe technology corporations protect users' data are more likely to think that predictive automation is fair. This study also finds that Americans who believe technology corporations do a good job at anticipating their social impact and should be less regulated generally view predictive automation as fair.

My research suggests people who think predictive automation is unfair are particularly concerned about social media companies differentially changing sites for some users and not others and reminding some users and not others to vote. This aligns with previous studies that have found that social media users react negatively to the predictive technologies that are used to curate platforms (Bucher, 2017; Eslami et al., 2015). These people do not seem to be comfortable with social media companies experimenting with users' emotions and exploiting their data to send advertisements.

Unlike previous studies (Araujo et al., 2020; Gran et al., 2020; Grgić-Hlača et al., 2020; Wang et al., 2020), this article did not find that gender or age are among the most important predictors of evaluations of predictive automation. Similarly, the literature is ambiguous about how education relates to views about predictive automation (Gran et al., 2020; Wang et al., 2020). This study only found that people with a high school education or less are more likely to believe that mining videos to filter job applicants is unfair. Race seemed to matter in only some cases, and its relation was inconsistent across cases.

In general, the variation in Americans' evaluations about the fairness of predictive automation will have substantial outcomes for governance and business. Previous literature suggests that negative feelings about predictive automation could affect how people judge technology corporations (Woodruff et al., 2018). As politicians and courtrooms are increasingly tasked with regulating predictive automation and reckoning with their aftermath, it may be their duty to sort through the complications. In the future, corporations, governments, and other entities interested in using predictive automation may consider consumer demand and employee's preferences. Just as concerns and demands for privacy prompted technology companies to respond (Rahnama and Pentland, 2022), predictive automation will too become a salient (and perhaps contentious) issue for companies and consumers.

## Supplementary Material

This article includes a replication file with an R project, unprocessed and processed data, and a table listing all the predictors used in the models, how they are measured, and their preprocessing. The online appendix referred to in the text is also available as a supplement online at the journal's website.

## Acknowledgement

I am very thankful to Laura Acosta, Anna Berg, Charles Camic, Berkeley Dietvorst, Steve Epstein, Wendy Espeland, Jeremy Freese, Alex Hanna, Alex Kindel, Arend Kuyper, Erik Lovell, Andrés Montealegre, Jim Murphy, Hatim Rahman, Jake Schauer, Qi Song, Nicolás Torres, Lincoln Quillian, and Vincent Yung, as well as the participants in the Culture and Society Workshop and the Applied Quantitative Methods Workshop at Northwestern University for their comments on previous drafts of this article.

## References

Angwinn J, Mattu S, Kirchner L (2016). Machine Bias. *ProPublica*.

Araujo T, Helberger N, Kruikemeier S, de Vreese CH (2020). AI we trust? Perceptions about Automated Decision-Making by Artificial Intelligence. *AI & Society*, 35: 611–623.

Baleis J, Keller B, Starke C, Marcinkowski F (2019). *Cognitive and Emotional Response to Fairness in AI – A Systematic Review*.

Benjamin R (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, Medford, MA.

Binns R, Van Kleek M, Veale M, Lyngs U, Zhao J, Shadbolt N (2018). 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*. ACM, New York, NY, USA. Event-place: Montreal QC, Canada.

Breiman L (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3): 199–231.

Brynjolfsson E, Kahin B (2000). *Understanding the Digital Economy: Data, Tools, and Research*. MIT Press, Cambridge.

Bucher T (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, 20(1): 30–44. Taylor & Francis.

Carlsson B (2004). The digital economy: What is new and what is not? *Structural Change and Economic Dynamics*, 15(3): 245–264.

Chafkin M (2021). *The Contrarian: Peter Thiel and Silicon Valley's Pursuit of Power*. Penguin Press, New York.

Chipman HA, George EI, McCulloch RE (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1): 266–298. Institute of Mathematical Statistics.

Clayton J (2022). A Year On, has Trump Benefited from a Twitter Ban? *BBC News*.

Dawes RM (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7): 571–582. American Psychological Association.

Dietvorst BJ, Simmons JP, Massey C (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them Err. *Journal of Experimental Psychology: General*, 114–126(1): 114.

Dodge J, Liao QV, Zhang Y, Bellamy RK, Dugan C (2019). Explaining models: an empirical study of how explanations impact fairness judgment. In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 275–285.

Dubber MD, Pasquale F, Das S (2020). *The Oxford Handbook of Ethics of AI*. Oxford University Press, Oxford.

Eslami M, Rickman A, Vaccaro K, Aleyasen A, Vuong A, Karahalios K, et al. (2015). "I always assumed that I wasn't really that close to [her]" Reasoning about Invisible Algorithms in News Feeds. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 153–162.

Eubanks V (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St.* Martin's Press, New York. Google-Books-ID: pn4pDwAAQBAJ.

Ferguson AG (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement.* NYU Press, New York.

González F, Yu Y, Figueroa A, López C, Aragon C (2019). Global reactions to the Cambridge analytica scandal: A cross-language social media study. In: *Companion Proceedings of The 2019 World Wide Web Conference, WWW '19*, 799–806. Association for Computing Machinery, New York, NY, USA.

Gran AB, Booth P, Bucher T (2020). To be or not to be algorithm aware: A question of a new digital divide? *Information, Communication & Society*, 24(12): 1779–1796. Taylor & Francis.

Grgic-Hlaca N, Redmiles EM, Gummadi KP, Weller A (2018). Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In: *Proceedings of the 2018 World Wide Web Conference on World Wide Web – WWW '18*, 903–912. ACM Press, Lyon, France.

Grgić-Hlača N, Weller A, Redmiles EM (2020). Dimensions of Diversity in Human Perceptions of Algorithmic Fairness. arXiv preprint: https://arxiv.org/abs/2005.00808.

Haidt J, Kesebir S (2010). Morality. In: *Handbook of Social Psychology*, 797–832. John Wiley & Sons, Inc., Hoboken, NJ.

Harcourt BE (2008). *Against prediction: Profiling, Policing, and Punishing in an Actuarial Age.* University of Chicago Press, Chicago.

Hastie T, Tibshirani R, Friedman J (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, New York.

Hidalgo C, Orghiain D, JA dAF, Martin N C (2021). *How Humans Judge Machines.* MIT Press, Cambridge, MA.

James G, Witten D, Hastie T, Tibshirani R (2013). *An Introduction to Statistical Learning with Applications in R*, volume 112. Springer.

Kiviat B (2021). Which data fairly differentiate? American views on the use of personal data in two market settings. *Sociological Science*, 8: 26–47.

Kizilcec RF (2016). How much information? Effects of transparency on trust in an algorithmic interface. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2390–2395.

Langer M, König CJ, Papathanasiou M (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, 27(3): 217–234. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ijsa.12246.

Leavitt K, Schabram K, Hariharan P, Barnes CM (2021). Ghost in the machine: On organizational theory in the age of machine learning. *Academy of Management Review*, 46(4): 750–777. Academy of Management.

Lee MK (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1): 1–12. SAGE Publications Ltd.

Lee MK, Jain A, Cha HJ, Ojha S, Kusbit D (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of* *the ACM on Human-Computer Interaction*, 3(CSCW): 1–26. ACM, New York, NY, USA.

Lehoucq E, Tarrow S (2020). The rise of a transnational movement to protect privacy. *Mobilization*, 25(2): 161–184. Mobilization: An International Quarterly.

Lizardo O, Mowry R, Sepulvado B, Stoltz DS, Taylor MA, Van Ness J, et al. (2016). What are dual process models? Implications for cultural analysis in sociology. *Sociological Theory*, 34(4): 287–310. SAGE Publications Sage CA, Los Angeles, CA.

Logg JM, Minson JA, Moore DA (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151: 90–103.

Meehl PE (1954). *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. University of Minnesota Press, Minneapolis.

Molina M, Garip F (2019). Machine learning for sociology. *Annu. Rev. Sociol.*, 45: 1–26.

Mullainathan S, Spiess J (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2): 87–106.

Newman DT, Fast NJ, Harmon DJ (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, 160: 149–167.

News B (2014). Facebook Emotion Experiment Sparks Criticism. *BBC News*.

Noble SU (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, New York.

O'Mara M (2019). *The Code: Silicon Valley and the Remaking of America*. Penguin Press, New York.

O'Neil C (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Random House, Danvers, MA.

Oppenheimer PC (2020). *The Long Good Buy: Analysing Cycles in Markets*. John Wiley & Sons, Chichester, UL.

Rahnama H, Pentland AS (2022). *The New Rules of Data Privacy. Harvard Business Review*. Section: Data management.

Rousseau DM, Sitkin SB, Burt RS, Camerer C (1998). Introduction to special topic forum: Not so different after all: A cross-discipline view of trust. *The Academy of Management Review*. 23(3): 393–404. Academy of Management.

Torpey J (2020). A sociological agenda for the tech age. *Theory and Society*, 49(5): 749–769. Springer.

Vaisey S (2009). Motivation and justification: A dual-process model of culture in action. *American Journal of Sociology*, 114(6): 1675–1715. The University of Chicago Press.

Varian HR (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2): 3–28.

Wang R, Harper FM, Zhu H (2020). Factors influencing perceived fairness in algorithmic decision-making algorithm outcomes, development procedures, and individual differences. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14.

Wilson JQ (1997). *The Moral Sense*. Simon and Schuster, New York.

Woodruff A, Fox SE, Rousso-Schindler S, Warshaw J (2018). A qualitative exploration of perceptions of algorithmic fairness. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems – CHI '18*, 1–14. ACM Press, Montreal QC, Canada.

Zuboff S (2019). *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. PublicAffairs, New York.