March 2022

Working Paper

Ethical AI development: Evidence from AI startups

James Bessen, Stephen Michael Impink, and Robert Seamans

This working paper is available online at: https://www.brookings.edu/center/center-on-regulation-and-markets/

Center on Regulation and Markets at BROOKINGS

The Center on Regulation and Markets at Brookings creates and promotes rigorous economic scholarship to inform regulatory policymaking, the regulatory process, and the efficient and equitable functioning of economic markets. The Center provides independent, non-partisan research on regulatory policy, applied broadly across microeconomic fields.

Ethical AI Development: Evidence from AI Startups

March 17, 2022

James Bessen, Technology & Policy Research Initiative, Boston University

Stephen Michael Impink, Stern School of Business, New York University

Robert Seamans, Stern School of Business, New York University

Abstract: Artificial Intelligence startups use training data as direct inputs in product development. These firms must balance numerous trade-offs between ethical issues and data access without substantive guidance from regulators or existing judicial precedence. We survey these startups to determine what actions they have taken to address these ethical issues and the consequences of those actions. We find that 58% of these startups have established a set of AI principles. Startups with data-sharing relationships with high-technology firms; that were impacted by privacy regulations; or with prior (non-seed) funding from institutional investors are more likely to establish ethical AI principles. Lastly, startups with data-sharing relationships with high-technology firms and prior regulatory experience with General Data Protection Regulation are more likely to take costly steps, like dropping training data or turning down business, to adhere to their ethical AI policies.

JEL codes: O33, J21, L10

Keywords: artificial intelligence, ethics, data, startup

1. Introduction

Digitalization, including the use of data to train algorithms in artificial intelligence (AI) products, is important for future macroeconomic growth (Brynjolfsson et al. 2017, Furman & Seamans 2018, Farboodi & Veldkamp 2019). Many future AI products will replace some aspects of human labor, augment human capabilities, and revolutionize data analysis (Bessen 2016, Bessen et al. 2018, Acemoglu & Restrepo 2018, Felton et al. 2019); yet, these gains are not without potential risks to consumer welfare. The mass collection and use of data, particularly personally identifiable information (PII), raises concerns about firms' ethical obligations to protect individuals and limit the likelihood of discriminatory outcomes (boyd & Crawford 2012, Barocas & Nissenbaum 2014, Whittaker et al. 2018, Barocas & Levy 2020). Moreover, scholars highlight many fairness issues stemming from the development and use of AI, such as unrepresentative training data, biased programmers, and an overemphasis on prediction accuracy, which may negatively impact decision-making or harm a demographic subgroup (Friedman & Nissenbaum 1996, Barocas & boyd 2017, Cowgill & Tucker 2019).

Over the past several years, AI's ethical use has been a popular topic among academics and journalists alike, with much of the research focusing on issues related to privacy and outcome fairness. Recent research extends this conversation to explain how managers use AI products within organizations, solidifying the importance of increased education around potential bias and lack of causal identification in many machine learning (ML) models (Martin et al. 2019, Cowgill 2019, Cowgill et al. 2020a, Kleinberg et al. 2018, Tarafdar et al. 2020). Despite the increased usage of AI in organizations, awareness of the potential pitfalls of using AI lags. A recent study of Google search trends shows a significant increase in searches related to AI-based recruitment and algorithmic bias since 2016; however, there has been no increase in similar searches focused on ethical recruitment and hiring (Mujtaba & Mahapatra 2019). Though there have been numerous calls to action, urging governments to regulate AI and provide substantive guidelines for ethical

AI development and usage (Whittaker et al. 2018), no widescale policies have emerged. The General Data Protection Regulation (GDPR) and California Consumer Protection Act (CCPA) are the most wide-scale ethical regulations; however, these regulations only cover privacy, one aspect of broader ethical concerns. Similar to privacy regulation before GDPR, few policies tackle ethical dilemmas—such as the impact of bias and outcome fairness holistically— and those that do exist are not stringently enforced. Additionally, most of the existing ethical regulations developed by sovereign governments[1] are from countries concentrated in a single geographic area: the European Union.

Given the lack of guidance from governments, guidelines focused on AI design, deployment, and usage are usually created, mandated, and adhered to at the firm level, leaving ample room for ambiguity and limited guidance for startups. As such, ethical AI principles may take on very different meanings across firms and industries. Despite these apparent differences in exact definitions, the firm-level policies signify the importance of ethics to these firms and their managers, investors, and stakeholders. Many large high-technology firms have established norms around AI development and share their guidance on potential negative externalities from AI usage (Guo et al. 2019, Smith & Browne 2019). As a result, private firms in the U.S., U.K., and E.U. published a quarter of the available AI guideline documents (Jobin et al. 2019)[2]. Intel[3], IBM[4], Microsoft[5], SAP[6], Sony[7], and Google[8] have shared ethical guidelines that provide insight into

[1] 18 ethical AI policies from governments, Jobin et al. 2019
[2] Jobin et al. 2019 details 84 ethical guidelines produced worldwide in the last several years.
[3] Intel's AI Privacy Policy White Paper: Protecting Individuals' Privacy and Data In The Artificial Intelligence World (Intel, 2018).
[4] Everyday Ethics for Artificial Intelligence (IBM, 2018). Transparency and trust in the cognitive era. IBM https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/ (2017).
[5] Microsoft AI principles. Microsoft https://www.microsoft.com/en-us/ai/our-approach-to-ai (2017).
[6] SAP's guiding principles for artificial intelligence (AI). SAP https://www.sap.com/products/leonardo/machine-learning/ai-ethics.html#guiding-principles (2018).
[7] Sony Group AI Ethics Guidelines (Sony, 2018). https://www.sony.net/SonyInfo/csr_report/humanrights /AI_Engagement_within_Sony_Group.pdf
[8] DeepMind, acquired by Google in 2014, issues guidelines in addition to Google. DeepMind ethics and society principles. DeepMind https://deepmind.com/applied/deepmind-ethics-society/principles/ (2017). Artificial intelligence at Google: our principles. Google AI https://ai.google/principles/ (2019).

algorithmic transparency, bias, and fairness, describing many implications for justice, equity, and privacy.

Larger high-technology firms have crafted ethical AI policies that align with their business models, reduce liabilities, and limit the chances of a public relations fiasco. There are, however, many examples of these larger firms turning down business that conflicts with their principles[9]. We are not yet fully aware of all the ethical implications of AI usage, even as AI actively reshapes many aspects of conducting business. Moreover, we have limited information about how creating and adhering to ethical AI policies is connected with larger firms' performance and even less information about how they correlate with AI startup performance. This paper explores the relationship between prior resources and AI startups' actions to address numerous data-related ethical issues. More specifically, our research question is whether and how prior resources, such as funding, data sharing relationships, and prior experience with GDPR impact the adoption and use of ethical AI principles.

Even though the exact nature of ethical AI principles can vary to some degree across firms, the focus of these policies often includes guidance on understanding potential biases in data and algorithms and how these biases could impact the fairness of their AI product's recommended outcomes. These policies may require firms not to use certain training data or to implement checks to ensure a level of ethical compliance, creating real costs that may negatively impact AI product development. For example, these policies may require that data containing PII or information on legally protected statuses like ethnicity, gender, or age are not used in AI production, limiting access to needed training data resources and possibly inhibiting algorithm performance. Certain AI startups have prior resources, experience, and relationships that better

[9] Washington Post: Microsoft won't sell police its facial-recognition technology, following similar moves by Amazon and IBM, https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/ and Big tech companies back away from selling facial recognition to police, https://www.vox.com/recode/2020/6/10/ 21287194/amazon-microsoft-ibm-facial-recognition-moratorium-police

prepare them to make these trade-offs. Furthermore, some startups may even turn down deals that conflict with their ethical guidelines, such as refusing to work with the military or certain authoritarian police forces, directly reducing their revenues.

On the other hand, having an ethics policy may signal that a startup is more willing to adapt to the industry's broader norms. High technology firms, which often share their data sources with startups, may be more comfortable sharing these resources or partnering with AI startups that take ethics seriously. Moreover, these firms may influence partnering startups to adapt their ethical norms through their corporate accelerator and incentive-based marketing programs. Certain investors may be more willing to invest in AI startups with ethics policies and in leaders who prioritize developing ethical AI norms. For example, investors could worry about PR-related issues or reduced exit opportunities for startups with ethical issues or products that facilitate discriminatory behavior. Furthermore, as potential algorithmic bias and resulting discrimination are more broadly acknowledged as a risk, there could be increased liability related to product usage once judicial rulings are established.

We assess these issues by collecting and analyzing novel survey data from 225 AI startups. We find that more than half of responding firms have ethical AI principles. However, many of those firms have never invoked their ethical AI principles in a costly way, such as firing an employee, dropping training data, or turning down a sale. We contribute to the developing literature on ethics in AI development by highlighting the case of AI startups, where these ethical issues are often related to the acquisition and use of resources needed to survive. For small, young firms like AI startups, presence or absence of ethical AI principles could impact product development in many tangible ways. Having to drop certain data and fire programmers could delay product releases; yet, firm survival often depends on quickly getting your product to market. We also use this study to contribute further research on how data privacy, often related to ethical issues around the use of personal data, creates an important trade-off between adhering to regulation and accessing needed training data.

This paper proceeds as follows. First, we introduce the academic literature on ethical development and usage of AI, highlighting gaps in the research on how ethics interplays with data access issues. Next, we discuss the data collected from our survey of AI startups ending in March 2021, including several data limitations, such as issues of non-response (response rate: 6%) and the cross-sectional nature of our data. We use cross-sectional regression models, relying on Heckman's selection to help overcome some issues of non-response and Coarsened Exact Matching (CEM) to support that our treatment group—those firms with AI ethics principles—are demographically similar to startups with these principles. Our analysis highlights correlations between a firm's prior resources and ethics outcomes, focusing on prior funding, data-sharing collaborations with high technology firms, including large cloud services providers (i.e., Amazon, Microsoft, and Google), and prior experience with GDPR. Lastly, we discuss possible antecedents to AI startups adapting ethical AI principles and conclude.

2. Prior Research on Ethical AI

First, we review research on ethical issues stemming from the mass collection of personal data, monitoring individuals, and related privacy implications. This literature highlights the trade-off between using certain types of data and possible ethical issues that could arise. Next, we focus on more recent issues of algorithmic bias stemming from training data and programmers. In addition to the choice of training data, how programmers code the algorithm could introduce bias, impacting the fairness of outcomes. Lastly, there is nascent literature on the impact of managers within organizations using potentially biased AI outcomes in a way that accounts for these biases, stressing the importance of education.

Privacy and Monitoring. Firms must choose what types of data they use to train the algorithms underlying their AI products. Often, firms use data about unique individuals, such as information on past purchases or preferences, which in conjunction with other demographic information is personally identifiable to that individual. The ability to identify someone from their data, passively collected through normal business activities, leads to numerous concerns about possible negative externalities (boyd & Crawford 2012). In most cases, consumers willingly share this information to enjoy services or social media platforms they highly value (Acquisti et al. 2016, FTC 2020). In response to growing privacy concerns, governments created substantive legislation (e.g., GDPR, CCPA) covering a broad range of ethical issues, including the right to access, delete, and prevent the sale of one's personal data held by a firm and to know what personal data is being collected and stored by one's employer.

Ethical concerns also arise from using AI to monitor individuals, capturing personal information, such as motions and gestures, to identify anomalies or unwanted behavior patterns. AI products focused on monitoring have a wide range of applications with socially positive outcomes, such as increasing fairness by ensuring students do not cheat on exams (Bellamy et al. 2018) or by reducing road hazards by buzzing truck drivers that start to doze. Alternately, there are applications with ethically ambiguous or even actively negative social outcomes, such as search crowds for individuals or using the technology to discriminate against a particular group. These "big-brother" style products collect lots of personal information that could be stored and reused to train other AI products. Even for socially beneficial outcomes, monitoring technologies raise concerns about the organization's role in forcing members to be subjected to these more intrusive, albeit passive, technologies.

The literature on the impact of data privacy regulation also gives us insight into how increased adherence to ethical norms impacts smaller, newer firms. For example, increased government regulation, though providing guidance for handling ethical issues that could harm consumers, may also reduce the amount of training data collected and used in AI development. This trade-off between increased regulation and data availability could asymmetrically have a negative impact on smaller firms that need data to develop their products and grow, creating competitive barriers for some startups (Jia et al. 2018, McSweeny & O'Dea 2018, Johnson & Shriver 2020; Bessen et al. 2020).

Algorithmic Bias. Programmers are often unaware that their product's results are biased. Even if they do know that their results are biased, they often cannot determine that bias's exact source (Selbst & Barocas 2018). AI products are often referred to as a "black box," meaning programmers have limited ability to explain causal relationships (Athey 2017, Donnelly et al. 2018). Data regulation has required that certain types of AI products and their outputs, such as training data and algorithms used in recruitment and hiring, be more "explainable." However, even additional transparency may not entirely reduce bias, especially when the source of the bias and relationships among the model's inputs are causally unknown.

Issues with an algorithm's training data can introduce bias into outcomes, impacting the fairness of outcomes (Barocas et al. 2018, Mitchell et al. 2021). An unrepresentative sample of data collected from a single race or gender could lead to biased outcomes. The most common example of this is the use of photos of human faces to train AI products focused on recognizing an individual. If you only use photos of one demography's faces, the AI product will be better at identifying individuals similar to those in the sample. Models may need to control for sensitive demographic information instead of ignoring it; otherwise, the endogeneity of related variables may confound results, introducing unaccounted-for biases (Cowgill & Stevenson 2020).

Even if the sample is representative, bias could emerge from biased programmers who build the algorithms in a certain way. There has been a recent push to better educate programmers against possible sources of algorithmic bias, including unconscious biases. In addition to education, many large-scale efforts have focused on increasing diversity in STEM programs or hiring more qualified minority programmers. A recent field study has shown that no one demographic subgroup is more biased per se; still, there is a benefit (i.e., reduced prediction errors) to having more demographic subgroup diversity (Cowgill et al. 2020b).

The coding of outcomes could introduce bias into an algorithmic model (Cowgill & Tucker 2019). For example, including all employees who left a firm in a turnover algorithm may capture individuals who left the firm for poor performance and, additionally, individuals who left because they were from a background that did not fit with the firm's culture. In many cases, outcomes are endogenous with other aspects of demography not accounted for in an AI product's underlying algorithms. Moreover, feedback loops, where outcomes are used in future iterations of the model, can further exacerbate these initial biases (Cowgill & Tucker 2019). Lastly, algorithmic bias can also occur when programmers are overly focused on prediction accuracy instead of weighing the benefits of accuracy with other pro-social outcomes (Cowgill & Stevenson 2020).

<u>Impact of Managers</u>. Most algorithms have one goal, prediction accuracy, and in most cases, the algorithm type and the associated training data are chosen as the first step in reaching this goal. In cases where programmers prioritize algorithmic accuracy, it may be best for humans to make the final decision weighing the potential source of biases and the organization's pro-social goals. Even if an AI product is biased, managers can use the output from AI combined with other information to produce a less biased outcome. Just because the AI is biased does not mean that the resulting decisions are biased; managers can use AI as a tool to complement their constrained decision-making processes (Kahneman & Tversky 2013).

Often bias from big data and algorithms emerge in similar processes within an organization. Managers benefit from education about recognizing sources of algorithmic bias and other forms of unconscious bias that could impact their decision-making (Martin et al. 2019, Rambachan & Roth 2019, Tarafdar et al. 2020). Even though managers may benefit from a metered approach, using AI to complement other information in their decision-making process, they often shy away from using AI when made aware of potential biases (Cowgill 2019, Cowgill et al. 2020a). Though recent frameworks have been introduced to help identify ethical issues (Silva & Martin 2019) and raise awareness of the trade-offs between aspects of ethics and the use of big data in machine learning processes (Morse et al. 2020), theoretical development needs to reflect the speed at which organizations are adapting to the use of AI Some managers are unaware of the possibility that their AI products are biased, producing less fair outcomes; however, given the amount of evidence provided in the current stream of research, managers should be made aware of ethical issues and held accountable to ameliorating potential issues (Martin 2019).

3. Survey Data & Measures

We use data from a survey of AI startups, including questions regarding the impact of ethical AI principles on product development and sales. We list these questions in Appendix A. We pre-tested the survey with several academics and practitioners associated with startups and then administered the survey from January 2021 to March 2021 through Qualtrics. We received 225 responses from AI startups in our sample; these firms confirmed that they develop AI products in the first survey question. Respondents to our survey came from several sampling frames; however, our sample's largest frame came from Crunchbase. We reached out to 3,790 AI startups worldwide. From Crunchbase, we identified firms associated with the keyword "artificial intelligence" that have received funding, are in operation, and have not yet experienced an IPO. In addition to Crunchbase, we received a contact list of AI startups from the Creative Destruction Lab, a startup incubator based in Toronto, and another contact list from Philipp Hartmann and Joachim Henkel (Hartmann & Henkel 2018). We addressed the survey to founders, chief technical officers, or other executives who know their firm's business model and technologies.

Firms in our survey are about four years old and employ, on average, 36 employees. However, almost half of firms have less than eleven employees. Even though the survey was administered worldwide, most of our responses are from more developed countries, with almost 80% of responses from the United States, Canada, and Europe. This lower response rate (6%) can bias our analysis if certain types of firms are more or less likely to respond to the survey. We use Heckman's selection, described in the Methods section below, in the first stage of our analysis to help correct this potential non-response bias based on demographics observables. However, we are still limited by potentially correlated unobservable variations of these startups not captured by our data. We report this summary in Table 1.A.

To build our measures, we paired the survey data with firm-level data from both Crunchbase and Pitchbook. Of the 225 survey respondents, 151 responding firms had available demographic and funding information in Crunchbase or Pitchbook.[10] From the survey, we created dummy variables for a) if the firm had established an ethical AI policy (58% SD 0.5); b) if the firm acted upon those principles by turning down business (23% SD 0.43), dismissing an employee (7% SD 0.25), or dropping training data (21% SD 0.41); and c) if the firm had taken the following ethics-related actions, including considering diversity when selecting training data (46% SD 0.50), hiring a minority or female programmer (69% SD 0.47), offering bias training (27% SD 0.45), or seeking expert advice (35% SD 0.48).

We create measures from survey responses on whether a firm has a data-sharing collaboration with a high-technology firm (45% SD 0.50) or has prior experience with GDPR (63% SD 0.49). To measure prior experience with GDPR, we create an indicator variable for if the respondent created a new position or reallocated resources to manage GDPR compliance or

[10] Some startups in our sample are listed in Crunchbase and/or Pitchbook with nothing more than a description. These startups are likely small, nascent ventures that will have additional data paired in the future as they grow. We drop any firms that are public (IPO), have acquired another firm, or are more than 10 years old.

deleted data due to GDPR compliance. We also create measures for prior resources from our paired firm-level data, including if the firm had received funding before the end of 2019 (26% SD 0.44) or if the firm had received funding from an institutional investor (Series A or later) (24% SD 0.43). Additionally, we use firm demographics to create the dummy variable used in the Heckman selection equation (described below) and as controls in the main equation. The majority of firms are less than three years old (64% SD 0.48), and a third of firms have less than 11 employees (32% SD 0.47). About 18% (SD 0.38) of firms are located in cities with a higher concentration of venture capital firms (San Francisco, London, New York, Boston, and Hong Kong)

We report these measures in the Summary in Table 1.B and related correlations in Appendix B.

4. Methods

We use regression models to explore the relationship between proprietary training data and venture capitalist (V.C.) funding. We use Heckman's selection approach (Heckman 1976, 1979) and Coarsened Exact Matching (CEM, Iacus et al. 2019) to help address selection and endogeneity issues. First, given our lower survey response rate and reliance on cross-sectional data, we analyze if our survey respondents are similar to the broader population of startups in Crunchbase and Pitchbook. From initial t-tests, we find that responses from the United States are overrepresented, but California, where many startups are based, is underrepresented. Moreover, younger firms are less likely to respond to our survey. To confirm this, we use a probit regression model to estimate the likelihood of response (Table 2, selection equation).[11]

[11] We look at response to the survey against the total population of AI startups matching our criteria, currently available in Crunchbase and Pitchbook (4,956 startups)

Based on this, we use Heckman's two-step procedure to account for selection issues from possible respondent missingness to support the argument that our sample of respondents does not bias our main OLS model estimates. We include dummy variables for young startups and startups with H.Q. locations in the United States or California in the first step, below, to obtain estimates of γ.

$$(1)\ response_i = w_i\gamma + \mu \qquad \text{[selection equation]}$$

where,

$response$ takes the value of 1 if a firm responds to the survey, otherwise 0.

w_i is a vector of firm demographic dummy variables (e.g., U.S., California, Young (<3 years old)) that are plausibly correlated with sample response.

Now that we have obtained the estimates of γ from the selection equation, we compute the inverse Mills ratios of each observation.

$$(2)\ \lambda = \frac{\phi(w_i\gamma)}{\Phi(w_i\gamma)} \qquad \text{[inverse Mill's ratio]}$$

where,

$\phi(w_i\gamma)$ is the probability density function

$\Phi(w_i\gamma)$ is the complementary cumulative distribution function

Next, we use CEM to ensure that the firms with an ethical AI policy are observationally similar to those without an ethical AI policy. We include H.Q. location (U.S. dummy), age (young), and employment size (small) as parameters in the CEM model and drop two firms (151 to 149 firms) that lack survey, funding, and demographic data. Moreover, we use CEM weighing in our main regression to support further that these two groups are observationally similar. The match reduces the difference in standardized means across these observable demographic variables between the respondents who have and do not have ethical AI principles. We provide a table comparing these demographic variables of firms before and after the match and weighting in Appendix E.

We use the following regression specification for our analysis:

$$\textit{(3)}\ ethics_i = \beta_0 + \beta_1 prior\_resource_i + \boldsymbol{X} + \rho + \mu \quad \text{[main equation]}$$

where,

$ethics_i$ is a dummy variable created from the survey responses for if the firm has a set of ethical AI principles (yes, 1; otherwise, 0). We also analyze other measures of ethics-policy related outcomes: a) turned down business due to their ethical AI principles, b) dismissed an employee due to their ethical AI principles, or c) stopped using certain training data due to their ethical AI principles; and ethics-related actions: d) considered diversity in training data selection, e) hired a minority or female programmer, f) offered bias training, or g) sought expert advice.

$prior\_resouce_i$ is a dummy variable (a) funding raised on or before Dec 31, 2019, a year before the survey, (b) received funding from an institutional investor during this period (Series A or later), c) collaborated with a high-technology firm to access data, or d) have prior experience with GDPR.

X are controls for small size (<11 employees) and H.Q. location in a top V.C. city.

λ is the inverse of the Mills ratio, included to control for the representativeness of our sample compared with the population of AI startups that we sourced and contacted.

μ is the error term; we use robust standard errors in all regressions.

We control for firm size and if a firm is located in a city with a high concentration of venture capital firms, which may be related to outcomes. For example, smaller firms have fewer employees, so they may be less likely to fire an employee as they have fewer employees to fire. Additionally, proximity to the largest V.C. firms may capture the impact of these interactions between larger institutional investors and similar firms in their same location. These learnings may be informally shared, impacting our results.

We pay close attention that these control variables are not highly correlated with firm age and H.Q. location in the U.S. or California since we use those variables in the first stage of the Heckman selection model. We provide these correlations in Appendix B.3.

5. Results

Most firms responding to our new questions have established codified ethical AI principles. However, initial analysis unveils that larger startups are more likely to have these established principles. About 75% of firms with more than 50 employees have ethical AI policies, whereas 60% of firms with fewer than 50 employees have similar policies. Also, there is variation by region, with Asia headquartered firms being more likely to have policies (71%) and Middle-Eastern and African headquartered firms being least likely to have policies (50%). There is, however, little difference between policy adoption of firms in North America (61%) and Europe (58%), where more than three-quarters of responding firms are located.

Similar to ethical AI policy adoption, we find that business outcomes resulting from startups adhering to their ethical AI principles vary by firm size and H.Q. location. Larger startups are much more likely to dismiss an employee and drop data (Figure 1.A), and firms in Europe are more likely to turn down business and drop data (Figure 1.B). Larger startups with more than ten employees are more likely to hire a female or minority programmer (Figure 2.A). Additionally, firms in the U.S. are more likely to acquire additional, more diverse training data and conduct unconscious bias training than firms in Europe and other parts of the world (Figure 2.B). These initial tabulations provide insight into which control variables are most important to include in our regression models: size and location are important determinants in AI adoption.

Figure 1.A. Ethical AI Policy and Business Outcomes by Firm Size

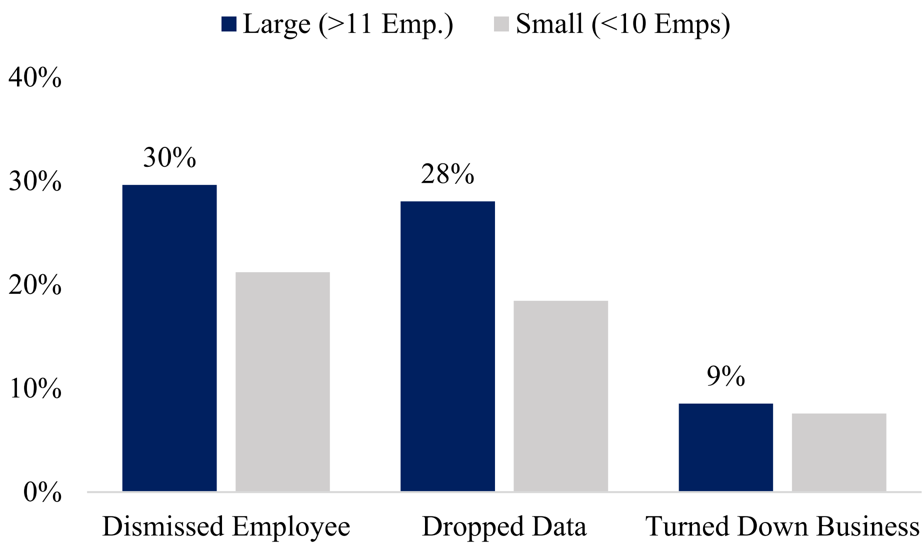


Figure 1.B. Ethical AI Policy and Business Outcomes by Region

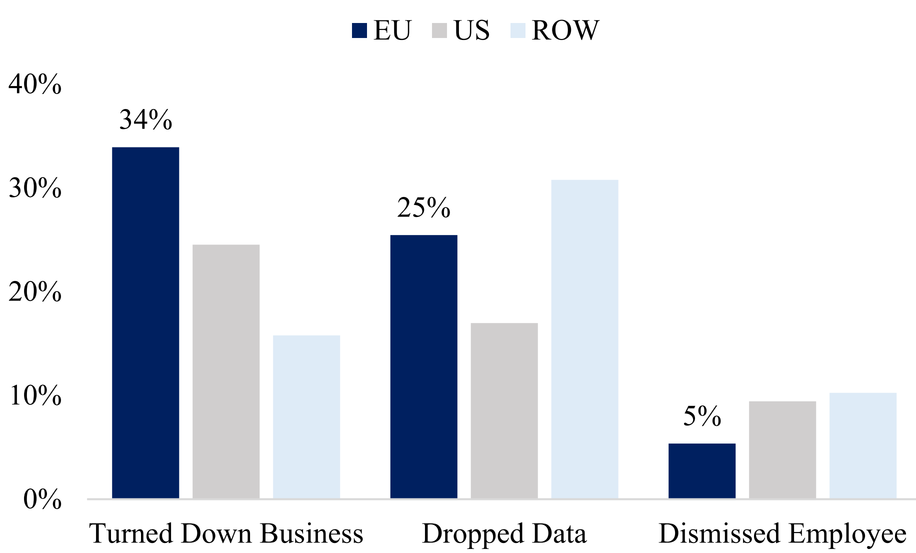


Figure 2.A. Ethics-related Actions by Firm Size

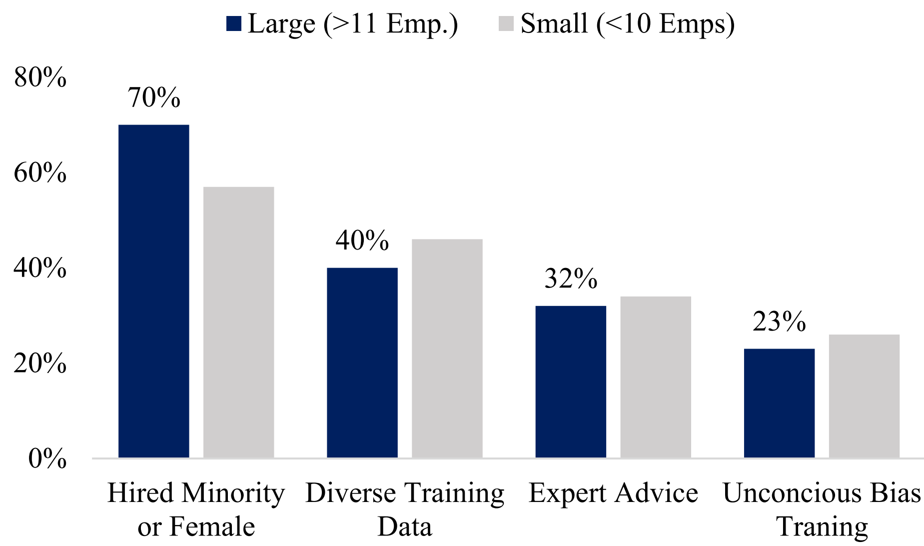
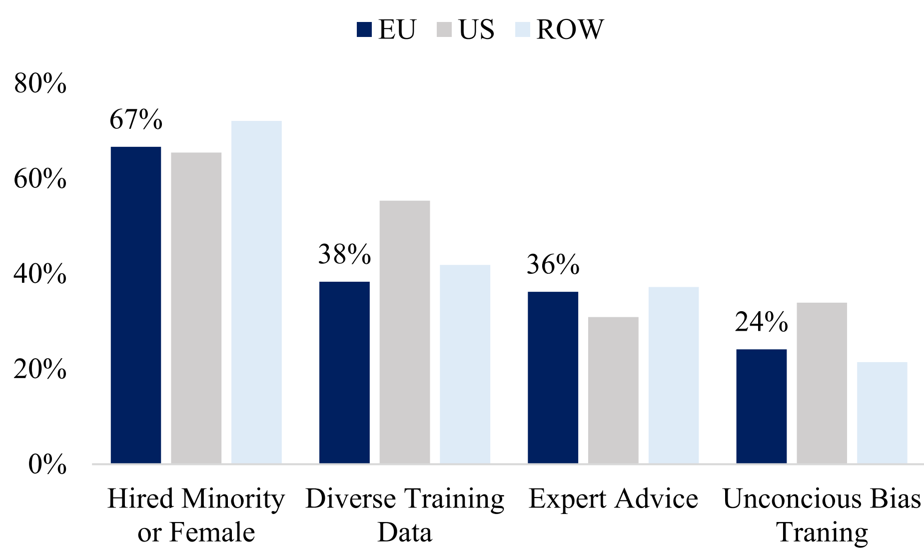


Figure 2.B. Ethics-related Actions by Region



Ethical AI Policy Adoption. Through a series of cross-sectional OLS regressions, we explore the relationship between prior resources and policy adoption. We find that prior funding is not related to adopting ethical AI principles. Similarly, we find that prior early- and late-stage (primarily Series A and B) institutional investor funding is negatively related to policy adoption. We report these findings in Table 3.A, additively building up the results: (1) base model, (2) including Heckman selection controls (the inverse Mill's ratio), (3) including CEM and regression weighting, and (4) including controls for size and H.Q. location in a city with a high concentration of V.C. firms. There is no relationship between these two funding measures and policy adoption, individually. However, there is a crossover interaction where firms with prior funding before the end of 2019 and prior funding from an institutional investor have relatively higher AI principle adoption than other firms (Table 3.C, model (4): +0.37 SD 0.19).

When using a matched sample, controlling for respondent selection, size, and location, there is a significant positive relationship between firms that have a data collaboration with a high-technology firm and adoption of AI principles (Table 3.B, model (4): +0.23 SD 0.09). While we make no causal assumptions, possibly these firms are more likely to share data with AI startups with ethical AI principles to reduce their own liability or limit the risk of data misuse. At the same time, it could be that these technology firms provide information through their corporate accelerators, marketing programs, or cloud-service relationships about the importance of developing AI more ethically.

Furthermore, using this same approach, there is a significant positive relationship between prior experience with GDPR and the increased adoption of AI principles (Table 3.B, model (8): +0.24 SD 0.08). These findings support a positive relationship between prior resources, such as certain types of funding, data relationships, or prior regulatory experience, and establishing AI principles.

As a robustness check, we run these models using probit and report results in Appendix C, Table C.1. This additional analysis finds similar results for high-technology firm data-sharing relationships and prior experience with GDPR on ethical AI principle adoption. Furthermore, we run our main results without Heckman's selection (IMR), but including matching and weighting (CEM) based on the treatment effect of adopting ethical AI principles. We report these in Appendix C in Tables C.2 and C.3. Next, we use a linear regression model with two types of lasso procedures (double selection and partialing out) to select control variables. This model is also run without Heckman's selection in the first stage, as we must ensure there is limited correlation between selection variables in the first-stage regression and controls in the second-stage regression. These results remain similar to our main results (Appendix C, Table C.4).

<u>Business Outcomes due to Ethical AI Principles.</u> A set of ethical AI principles in and of itself is not important unless firms adhere to those principles. From the survey, we asked firms with AI policies to provide additional information on how adherence to these policies impacts their business outcomes to determine if ethics policies are followed rather than simply being signals to investors. More than half of the firms with AI principles experienced at least one costly business outcome because they adhered to their ethical AI principles. Though this is not an exhaustive list of outcomes, these fields provide insight into how being more ethical impacts AI startups' operations, costs, and revenues in data-centric production. Though this question asks respondents if they have incurred a negative business outcome due to their AI principles, there is a possibility that these outcomes could have occurred anyways or were misattributed to the adopting AI principles.

Startups with a data-sharing collaboration with a high technology firm are substantially more likely to dismiss an employee (Table 4, model (2): +0.1 SD 0.04), drop training data (Table 4, model (5): +0.15 SD 0.06), and turndown business due to their AI principles (Table 4, model (8): +0.23 SD 0.06). Moreover, startups with prior experience with GDPR are more likely to drop training data when adhering to their principles (Table 4, model (5): +0.18 SD 0.06).[12] Despite these findings, we find limited support for the impact of prior funding on these business outcomes.

Ethics-related Actions. Furthermore, many firms have taken ethics-related actions not necessarily connected to adopting ethical AI principles, such as hiring minorities, using more diverse training data, seeking expert advice, and conducting bias training. We examine the relationship between prior resources and these ethics-related actions. There is also a positive relationship between having a data-sharing collaboration with a technology firm and an increased focus on acquiring more diverse training data (Table 5, Model (2): +0.19 SD 0.08). Lastly, there is a positive relationship between the prior experience with GDPR and requiring bias training for employees (Table 5, model (6): 0.14 SD 0.07). We find no other significant relationships between prior resources and business-related actions.

Lastly, in Appendix D, we report additional findings from the survey focused on variation in AI policy adoption and types of algorithms, data protection, customer industry, and other measures capturing the importance of training data to these startups.

6. Conclusion

This study provides insight into how AI startups—firms on the front lines of AI product development—address ethical issues. Combined with our prior surveys, this data is a step towards better understanding entrepreneurship around AI and the issues impacting startups relying on big datasets in production. Given the value of AI to the economy more broadly, these firms' innovations are anticipated to be important for labor productivity and future macroeconomic growth. Similar to data privacy, there is a trade-off between many ethical issues and the ability to access and use certain data.

[12] There is no significant difference between how GDPR effects startups in the US versus Europe. Even though it is a European regulation, it is widely followed in most developed countries.

There are a few main findings from our analysis. First, AI startups likely follow norms developed in their customer's industries or codified by technology firms. In almost all scenarios explored, the relationship with a technology firm was significantly related to more ethical startup behavior. Next, larger technology firms have created norms that are often cited and followed by startups since there are no extant government policies other than privacy regulation addressing many of these ethical issues. Lastly, prior resources, relationships, and regulatory experience enable startups to navigate complex issues and make certain ethical choices that increase costs, such as training or hunting for more diverse data, or reduce revenues, such as dropping data or refusing to sell to less ethical customers.

Our results are derived from cross-sectional survey data, which has its limitations. We have attempted to address these issues as much as possible by using Heckman selection correction and Coarsened Exact Matching approaches. Though we control for size and having a headquarters location in a city with a large concentration of V.C. firms, we cannot entirely rule out that firms with ethical AI policies are not more likely to partner with high technology firms. Moreover, there are other unobservable aspects of prior resources that may be correlated with policy adoption. For instance, having a supplier relationship with a specific cloud services supplier or other aspects of relationships with these high technology firms may impact outcomes.

The conversation around the ethical use of AI is becoming even more important as startups use more sophisticated algorithms that require even more data. For example, when more firms use neural networks, they may be less willing to drop data as that data may be needed to create a functional AI product. From our results, it is apparent that many AI startups are aware of possible ethical issues, and more than half have taken steps by providing codified AI principles to guide their firm. However, firms with prior resources, such as data sharing relationships with larger high technology firms and prior regulatory experience with GDPR, are more likely to act on these principles in a material way.

Tables & Figures

Table 1.A – Firm Summary

Measure	Mean	SD	Min	Max	Firms	Source
Age	3.85	1.86	0.1	9		Crunchbase, Pitchbook
Young (less than 3 years)	0.64	0.48	0	1	102	
Employee (Avg. Count)	36	39.54	10	250		Crunchbase, Pitchbook, Q2
Small (less than 11 employees)	0.32	0.47	0	1	51	
Headquarters Location						Crunchbase, Pitchbook, Q4
Asia	10%				16	
Middle East & Africa	5%				8	
United States	41%				66	
California	*7%*				11	
New York	*6%*				10	
Massachusetts	*2%*				3	
United Kingdom	11%				18	
Canada	5%				8	
Germany	3%				5	
France	2%				3	
Other Countries	33%				53	

Table 1.B – Measure Summary

Measure	Mean	SD	Min	Max	Firms	Source
Dependent Variables						
Do you have Ethical AI Principles?	0.58	0.50	0	1	93	Q51
Due to these principles, has your firm:						
Dismissed Employee	0.07	0.25	0	1	10	Q52
Dropped Data	0.21	0.41	0	1	33	Q52
Turned Down Business	0.23	0.43	0	1	37	Q52
Has your firm done the following:						
Considered Diversity in Data Selection	0.46	0.50	0	1	74	Q50
Hired Minority/Female Programmer	0.69	0.47	0	1	110	Q50
Offered Unconscious Bias Training	0.27	0.45	0	1	43	Q50
Sought Expert Advice	0.35	0.48	0	1	56	Q50
Independent Variables						
Log (Funding before 2019)	3.16	5.45	0	15.66		Crunchbase, Pitchbook
Funding before 2019, dummy	0.26	0.44	0	1		Crunchbase, Pitchbook
Seed Funding	0.24	0.43	0	1		Crunchbase, Pitchbook
Institutional Investors (Series A or later)	0.11	0.31	0	1		Crunchbase, Pitchbook
Tech Firm Data Collaboration	0.45	0.50	0	1		Q30
GDPR (Prior Capability)	0.63	0.49	0	1		Q37, Q38, Q39

Notes: 160 Observations/Startups.

Table 2 – Response Probit

DV, Dummy:	(1) **Response**
Young (<3 Years)	0.177*
	(0.096)
California (dummy)	-0.828***
	(0.128)
Firms	4296

Notes: * p<0.1, ** p<0.05, *** p<0.01. 225 firms respond to the survey, 4956 firms are the current population of AI startups in our third-party data. Coefficients are estimated using Probit regression, which supports the variables used in the first stage of the Heckman selection procedure.

Table 3.A – AI Ethics Principles (Funding)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV, Dummy:	**Does your firm have Ethical AI Principles?**							
Sample	All	All	Matched	Matched	All	All	Matched	Matched
Funding before 2019 (dummy)	-0.051	-0.048	-0.049	-0.056				
	(0.083)	(0.085)	(0.085)	(0.086)				
Inst. Investors (Series A or later, dummy)					-0.243***	-0.250***	-0.251***	-0.251***
					(0.090)	(0.094)	(0.094)	(0.094)
IMR		0.076	0.080	-0.187		0.259	0.255	0.004
		(0.311)	(0.317)	(0.357)		(0.286)	(0.288)	(0.339)
Employees (<11)			0.008	0.001			-0.012	-0.020
			(0.085)	(0.086)			(0.082)	(0.083)
VC Location				0.222				0.207
				(0.137)				(0.136)
CEM Weighting:	No	No	Yes	Yes	No	No	Yes	Yes
Firms	151	149	149	149	151	149	149	149
Adj R2	-0.001	-0.012	-0.019	-0.011	-0.004	-0.013	-0.020	-0.012

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS regression and include robust standard errors, in parentheses below the coefficient. Models (3), (4), (7), and (8) include matching (CEM), based on firm age (young.), employment size (employment small), and region (US), dropping two firms. Additionally, these models use CEM weighing. All but the base models (1) and (5) use Heckman's selection procedure, controlling with IMR.

Table 3.B – AI Ethics Principles (Prior Resources)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV, Dummy:				**Does your firm have Ethical AI Principles?**				
Sample	All	All	Matched	Matched	All	All	Matched	Matched
Tech Firm Data Collab.	0.229***	0.216***	0.225***	0.229***				
	(0.076)	(0.079)	(0.081)	(0.081)				
GDPR (Prior Capability)					0.243***	0.229***	0.235***	0.238***
					(0.080)	(0.082)	(0.083)	(0.083)
IMR		-0.044	-0.029	-0.317		0.111	0.128	-0.152
		(0.286)	(0.298)	(0.338)		(0.301)	(0.308)	(0.349)
Employees (<11)			0.048	0.040			0.039	0.032
			(0.086)	(0.087)			(0.086)	(0.086)
VC Location				0.235*				0.231*
				(0.138)				(0.132)
CEM Weighting:	No	No	Yes	Yes	No	No	Yes	Yes
Firms	160	160	160	160	160	160	160	160
Adj R2	0.0467	0.0346	0.0305	0.0318	0.0506	0.0377	0.0330	0.0340

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS regression and include robust standard errors, in parentheses below the coefficient. All models include matching (CEM), based on firm age (young.), employment size (employment small), and region (US), dropping two firms. Additionally, all models use CEM weighing and Heckman's selection procedure, controlling with IMR.

Table 3.C – AI Ethics Principles (Funding Interaction)

	(1)	(2)	(3)	(4)
DV, Dummy:	**Does your firm have Ethical AI Principles?**			
Sample	All Matched			
Funding before 2019 (dummy)	-0.475***	-0.471***	-0.472***	-0.470***
	(0.120)	(0.125)	(0.125)	(0.127)
Inst. Investors (Series A or later, dummy)	-0.055	-0.050	-0.052	-0.062
	(0.106)	(0.109)	(0.112)	(0.113)
Interact.: Funding x Inst. Investors	0.370**	0.357*	0.360*	0.366*
	(0.180)	(0.188)	(0.192)	(0.193)
IMR		0.201	0.203	-0.051
		(0.281)	(0.285)	(0.334)
Employees (<11)			0.007	0.001
			(0.086)	(0.086)
VC Location				0.208
				(0.135)
CEM Weighting:	All CEM Weighted			
Firms	160	160	160	160
Adj R2	0.0493	0.0399	0.0337	0.0334

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS regression and include robust standard errors, in parentheses below the coefficient. Models (3), (4), (7), and (8) include matching (CEM), based on firm age (young.), employment size (employment small), and region (US), dropping two firms. Additionally, these models use CEM weighing. All but the base models (1) and (5) use Heckman's selection procedure, controlling with IMR.

Table 4 – Ethical Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
DV, Dummy:	**Dismissed Employee**			**Dropped Training Data**			**Turned Down Business**		
Sample	All Matched								
Funding before 2019 (dummy)	0.011			-0.065			0.125*		
	(0.043)			(0.066)			(0.074)		
Tech Firm Data Collab.		0.097**			0.150**			0.233***	
		(0.044)			(0.062)			(0.064)	
GDPR (Prior Capability)			0.050			0.184***			0.080
			(0.038)			(0.059)			(0.064)
IMR	-0.094	-0.144	-0.083	-0.519*	-0.607**	-0.495**	-0.459	-0.567*	-0.428
	(0.219)	(0.213)	(0.215)	(0.276)	(0.269)	(0.246)	(0.357)	(0.316)	(0.337)
Employees (<11)	0.013	0.033	0.022	-0.099	-0.076	-0.077	-0.015	0.045	0.011
	(0.043)	(0.046)	(0.046)	(0.068)	(0.065)	(0.066)	(0.071)	(0.070)	(0.071)
VC Location	0.071	0.085	0.078	0.306	0.308*	0.308*	0.437**	0.494***	0.474***
	(0.178)	(0.164)	(0.170)	(0.191)	(0.186)	(0.168)	(0.196)	(0.177)	(0.179)
CEM Weighting:	All CEM Weighted								
Firms	160	160	160	160	160	160	160	160	160
Adj R2	-0.0212	0.0166	-0.0118	0.0122	0.0394	0.0542	0.0408	0.0978	0.0286

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS regression and include robust standard errors, in parentheses below the coefficient. All models include matching (CEM), based on firm age (young.), employment size (employment small), and region (US), dropping two firms. Additionally, all models use CEM weighing and Heckman's selection procedure, controlling with IMR.

Table 5 – Ethical Firm Actions

	(1)	(2)	(3)	(4)	(5)	(6)
DV, Dummy:	**Training Data Diversity**			**Bias Training**		
Sample	All Matched					
Funding before 2019 (dummy)	0.047			0.016		
	(0.085)			(0.078)		
Tech Firm Data Collab.		0.193**			0.054	
		(0.079)			(0.072)	
GDPR (Prior Capability)			0.124			0.138*
			(0.082)			(0.072)
IMR	-0.755**	-0.852***	-0.727**	0.101	0.074	0.128
	(0.303)	(0.283)	(0.302)	(0.385)	(0.380)	(0.387)
Employees (<11)	0.039	0.083	0.063	0.021	0.033	0.044
	(0.085)	(0.082)	(0.085)	(0.078)	(0.077)	(0.077)
VC Location	0.554***	0.587***	0.577***	0.097	0.107	0.114
	(0.092)	(0.090)	(0.089)	(0.212)	(0.212)	(0.217)
CEM Weighting:	All CEM Weighted					
Firms	160	160	160	160	160	160
Adj R2	0.0237	0.0579	0.0359	-0.0186	-0.0154	0.00363

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS regression and include robust standard errors, in parentheses below the coefficient. All models include matching (CEM), based on firm age (young.), employment size (employment small), and region (US), dropping two firms. Additionally, all models use CEM weighing and Heckman's selection procedure, controlling with IMR.

References

Acemoglu, D., & Restrepo P. Artificial Intelligence, Automation and Work. No. w24196. National Bureau of Economic Research. 2018.

Acquisti, A., Taylor, C., & Wagman, L. The Economics of Privacy. Journal of Economic Literature, 54(2), 442-92. 2016.

Athey, S. Beyond Prediction: Using big data for policy problems. Science 355, no. 6324, 483-485. 2017.

Barocas, S., & Boyd, D. Engaging the ethics of data science in practice. Communications of the ACM, 60(11), 23-25. 2017.

Barocas, S., & Nissenbaum, H. Big data's end run around anonymity and consent. Privacy, big data, and the public good: Frameworks for engagement, 1, 44-75. 2014

Barocas, S., & Selbst, A. D. Big data's disparate impact. Calif. L. Rev., 104, 671. 2016.

Barocas, S., Hardt, M., & Narayanan, A. Fairness and Machine Learning. URL: http://www. fairmlbook.org. 2018.

Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., ... & Zhang, Y. AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. arXiv preprint arXiv:1810.01943. 2018.

Bessen, J. E. How Computer Automation affects Occupations: Technology, Jobs, and Skills. Boston Univ. School of Law, Law and Economics Research Paper, 15-49. 2016.

Bessen, J. E., Impink, S. M., Reichensperger, L., & Seamans, R. GDPR and the Importance of Data to AI Startups. Available at SSRN 3576714. 2020.

Bessen, J. E., Impink, S. M., Reichensperger, L., & Seamans, R. The Business of AI Startups. Boston Univ. School of Law, Law and Economics Research Paper, 18-28. 2016.

boyd, D., & Crawford, K. Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society 15, no. 5: 662-679. 2012.

Brynjolfsson, E., Rock, D., & Syverson, C. Artificial Intelligence and the Modern Productivity Paradox: A clash of expectations and statistics (No. w24001). National Bureau of Economic Research. 2017.

Cowgill, B., & Tucker, C. E. Economics, Fairness and Algorithmic Bias. Journal of Economic Perspectives. 2019.

Cowgill, B., and Stevenson, M. Algorithmic Social Engineering. AEA Papers and Proceedings, vol. 110, pp. 96-100. 2020.

Cowgill, B., Dell'Acqua, F., & Matz, S. The Managerial Effects of Algorithmic Fairness Activism. AEA Papers and Proceedings, vol. 110, pp. 85-90. 2020a.

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. Biased Programmers? or biased data? A field experiment in operationalizing ai ethics. In Proceedings of the 21st ACM Conference on Economics and Computation, pp. 679-681. 2020b.

Donnelly, R., Ruiz, F. R., Blei, D., & Athey, S. Counterfactual Inference for Consumer Choice Across Many Product Categories. arXiv preprint arXiv:1906.02635. 2019.

Farboodi, M., & Veldkamp, L. A Growth Model of the Data Economy. Working Paper, Columbia Business School, New York, June. 2019.

Felten, E. W., M. Raj, and R. C. Seamans. The Variable Impact of Artificial Intelligence on Labor: The role of complementary skills and technologies. Available at SSRN. 2019.

Friedman, B., & Nissenbaum, H. Bias In Computer Systems. ACM Transactions on Information Systems (TOIS) 14, no. 3, 30-347. 1996.

Goldberg, S., Johnson, G., & Shriver, S. Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes. Available at SSRN 3421731. 2019.

Guo, A., Kamar, E., Vaughan, J.W., Wallach, H., and Morris, M. R. "Toward fairness in AI for people with disabilities: A research roadmap." arXiv preprint arXiv:1907.02227. 2019.

Heckman, J. J. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement, v. 5, n. 4 (pp. 475-492). 1976.

Heckman, J. J. Sample selection bias as a specification error. Econometrica: Journal of the Econometric Society, 153-161. 1979.

Iacus, S. M., King, G., & Porro, G. Causal inference without balance checking: Coarsened exact matching. Political Analysis, 1-24. 2012.

Jia, J., Jin, G. Z., & Wagman, L. The Short-run Effects of GDPR on Technology Venture Investment (No. w25248). National Bureau of Economic Research. 2018.

Jobin, A., Ienca, M., & Vayena, E. The Global Landscape of AI Ethics Guidelines. Nature Machine Intelligence 1, no. 9: 389-399. 2019.

Johnson, G. A., Shriver, S. K., & Du, S. Consumer privacy choice in online advertising: Who opts out and at what cost to industry?. Marketing Science. 2020.

Kahneman, D., and Tversky, A. "Choices, values, and frames." In Handbook of the fundamentals of financial decision making: Part I, pp. 269-278. 2013.

Kleinberg, J., Ludwig, J., Mullainathan, S. and Rambachan, A. Algorithmic fairness. In AEA Papers and Proceedings (Vol. 108, pp. 22-27). 2018.

Martin, K. E. Designing ethical algorithms. MIS Quarterly Executive June. 2019.

Martin, K., Shilton, K., & Smith, J. Business and the ethical implications of technology: Introduction to the symposium. 2019.

McSweeny, T., O'Dea, B. Data, Innovation, and Potential Competition in Digital Markets. CPI Antitrust Chronicle, February. 2018.

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. Prediction-based decisions and fairness: A catalogue of choices, assumptions, and definitions. arXiv preprint arXiv:1811.07867. 2018

Mujtaba, D. F., & Mahapatra, N. R. Ethical considerations in AI-Based recruitment. IEEE International Symposium on Technology and Society (ISTAS), pp. 1-7. IEEE, 2019.

Nanda, R. Financing high-potential entrepreneurship. IZA World of Labor.05530. 2016.

Rambachan, A. and Roth, J., 2019. Bias in, bias out? Evaluating the folk wisdom. arXiv preprint arXiv:1909.08518.

Silva, S., & Kenney, M. Algorithms, platforms, and ethnic bias. Communications of the ACM, 62(11), 37-39. 2019.

Smith, B., and Browne, C.A. Tools and weapons: The promise and the peril of the digital age. Penguin. 2019.

Tarafdar, M., Teodorescu, M., Tanriverdi, H., Robert, L. P., & Morse, L. Seeking Ethical use of AI Algorithms: Challenges and Mitigations. 2020.

Teodorescu, M. H., Morse, L., Awwad, Y., & Kane, J. A Framework for Fairer Machine Learning in Organizations. In Academy of Management Proceedings (Vol. 2020, No. 1, p. 16889). Briarcliff Manor, NY 10510: Academy of Management. 2020.

Whittaker, M., Crawford, K., & Dobbe, R. Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Mysers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. AI now report 2018. New York: AI Now Institute at New York University. 2018.

Appendix

Appendix A.1 – Survey Questions

Ethics Questions:

Q50. Has your firm taken any of the following actions?

	Yes	No	I don't know or N/A
Offered unconscious bias training	◯	◯	◯
Hired an under-represented minority or female programmer	◯	◯	◯
Considered gender or racial diversity as criteria for selecting training data	◯	◯	◯
Sought expert advice on navigating ethical issues	◯	◯	◯

51 Does your firm have a set of ethical AI principles?

◯ Yes

◯ No

◯ I don't know

52 If your firm has ethical AI principles, has your firm ever:

	Yes	No	I don't know or N/A
Turned down business due to a conflict with these ethical AI principles	◯	◯	◯
Dismissed an employee that did not follow these ethical AI principles	◯	◯	◯
Stopped using certain training data that did not align with these ethical AI principles	◯	◯	◯

Data Sharing Question:

30 Do you collaborate with other technology firms to access data?

- ◯ Yes
- ◯ No
- ◯ I don't know (99)

GDPR Questions:

37 Have you created a new position to handle compliance associated with the General Data Protection Regulation ("GDPR")?

- ◯ Yes
- ◯ No
- ◯ I don't know or N/A

38 Have you reallocated resources in your firm to handle GDPR compliance?

- ◯ Yes
- ◯ No
- ◯ I don't know or N/A

39 Have you deleted some data in order to comply with GDPR?

- ◯ Yes
- ◯ No
- ◯ I don't know or N/A

Appendix B – Correlations

Table B.1 – AI Ethics Principles and Ethical Outcomes Correlations

	Ethical AI Principles?	Dismissed Employee	Dropped Data	Turned Down Business	Funding before 2019	Tech Firm Data
Dismissed Employee	0.2263* 0.0051					
Dropped Data	0.4317* 0	0.3267* 0				
Turned Down Business	0.4578* 0	0.4307* 0	0.4336* 0			
Funding before end of 2019	-0.0756 0.356	-0.0355 0.6656	-0.1126 0.1686	0.0441 0.5905		
Tech Firm Data Collaboration	0.2313* 0.0041	0.1882* 0.0202	0.2013* 0.0129	0.2791* 0.0005	0.1349 0.0985	
GDPR (Prior Capability)	0.2477* 0.0021	0.0959 0.2398	0.2234* 0.0057	0.0897 0.2719	-0.0087 0.9157	0.0683 0.4029

Table B.2 – Firm Ethical Action Correlations

	Diverse Training Data	Sought Expert	Hired Minority/ Female	Bias Training	Funding before 2019	Tech Firm Data
Sought Expert	0.2103* 0.0093					
Hired Minority/ Female Prog.	0.2137* 0.0082	0.1502 0.0647				
Bias Training	0.3604* 0	0.2396* 0.0029	0.1337 0.1004			
Funding before 2019 (Dummy)	0.0886 0.2793	0.0733 0.3713	-0.0758 0.3552	-0.0541 0.5096		
Tech Firm Data Collaboration	0.1774* 0.0287	0.1191 0.1439	0.1676* 0.039	0.0494 0.5454	0.1349 0.0985	
GDPR (Prior Capability)	0.1159 0.1552	0.139 0.0876	0.1635* 0.0441	0.1646* 0.0428	-0.0087 0.9157	0.0683 0.4029

Table B.3 – Firm Demographics Correlations

	US	VC Location	Less than 11 Emps.
VC Location	0.2103* 0.0093		
Less than 11 Employees	0.2137* 0.0082	0.1502 0.0647	
Less than 3 years old	0.1159 0.1552	0.139 0.0876	0.1635* 0.0441

Appendix C – Robustness

Table C.1 – Probit AI Ethics Principles

	(1)	(2)	(3)	(4)
DV, Dummy:	**Does your firm have Ethical AI Principles?**			
Sample:	All CEM Matched			
Funding before 2019 (dummy)	-0.145			
	(0.212)			
Inst. Investors (Series A or later, dummy)		-0.706***		
		(0.243)		
Tech Firm Data Collab.			0.590***	
			(0.209)	
GDPR (Prior Capability)				0.638***
				(0.212)
Employees (<11)	-0.116	-0.195	-0.012	-0.028
	(0.209)	(0.213)	(0.214)	(0.213)
VC Location	0.407	0.614	0.357	0.479
	(0.435)	(0.463)	(0.437)	(0.440)
Firms	160	160	160	160

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using Probit regression, which supports the variables used in the first stage of the Heckman selection procedure.

Table C.2 – AI Ethics Principles (Funding)

	(1)	(2)	(3)	(4)	(5)	(6)
DV, Dummy:	**Does your firm have Ethical AI Principles?**					
Sample	All	Matched	Matched	All	Matched	Matched
Funding before 2019 (dummy)	-0.051	-0.046	-0.058			
	(0.083)	(0.085)	(0.086)			
Inst. Investors (Series A or later, dummy)				-0.243***	-0.231**	-0.250***
				(0.090)	(0.092)	(0.093)
Employees (<11)		0.005	0.006		-0.019	-0.020
		(0.084)	(0.084)		(0.081)	(0.082)
VC Location			0.154			0.208
			(0.155)			(0.130)
CEM Weighting:	No	Yes	Yes	No	Yes	Yes
Firms	160	160	160	160	160	160
Adj R2	-0.001	-0.019	-0.011	-0.004	-0.020	-0.012

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS regression.

Table C.3 – AI Ethics Principles (Prior Resources)

	(1)	(3)	(4)	(5)	(7)	(8)
DV, Dummy:	**Does your firm have Ethical AI Principles?**					
Sample	All	Matched	Matched	All	Matched	Matched
Tech Firm Data Collab.	0.229***	0.224***	0.222***			
	(0.076)	(0.080)	(0.081)			
GDPR (Prior Capability)				0.243***	0.232***	0.240***
				(0.080)	(0.083)	(0.082)
Employees (<11)		0.049	0.048		0.035	0.036
		(0.085)	(0.086)		(0.085)	(0.085)
VC Location			0.120			0.176
			(0.160)			(0.148)
CEM Weighting:	No	Yes	Yes	No	Yes	Yes
Firms	160	160	160	160	160	160
Adj R2	0.0467	0.0366	0.0339	0.0506	0.0379	0.0393

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS regression.

Table C.4 – AI Ethics Principles (Lasso Model, Control Selection)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV, Dummy:	**Does your firm have Ethical AI Principles?**							
Sample	All Matched (CEM)							
Funding before 2019 (dummy)	-0.063	-0.074						
	(0.083)	(0.082)						
Inst. Investors (Series A or later, dummy)			-0.235**	-0.246***				
			(0.091)	(0.086)				
Tech Firm Data Collab.					0.247***	0.246***		
					(0.077)	(0.078)		
GDPR (Prior Capability)							0.258***	0.228***
							(0.080)	(0.077)
Lasso Procedure	DS	XPO	DS	XPO	DS	XPO	DS	XPO
Firms	160	160	160	160	160	160	160	160

Notes: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are estimated using OLS regression.

Appendix D – Other Survey Results

We report the survey results across various grouping created from other survey questions: customer industry, data protections, algorithms, and the importance of data in their markets.

Respondents with ethical AI principles are more likely to sell their products into specific industries (Figure 1). For instance, more than 40% of respondents with ethical AI principles sell to customers in the financial services and healthcare industry. On the other hand, less than 20% of those firms sell to customers in the government, agriculture, education, and law enforcement industries. This variation of ethics policies usage by industry may point to possible differences in the underlying business models and requisite training data. For instance, AI startups that sell into the healthcare industry may use more sensitive data (i.e., data about human patients) than startups that see into the agriculture industry (i.e., data about the soil quality and rainfall). Moreover, variation by customer industry points to different norms based on who uses the product and product-related regulations in that industry.

There is also variation in data protections used by AI startups with and without Ethical AI principles (Figure 2). Firms that have ethical AI principles are generally more likely to use data protection. This is particularly noticeable for more technical data protections, like encryption and logged access. Additionally, firms with ethical principles are more likely to deidentify data, safeguarding consumer privacy by removing personally identifiable information. More common data protection methods, such as legal contracts or passwords, are used with similar frequency across both groups.

Firms with ethical AI policies are more likely to use all types of algorithms, except neural networks (Figure 3). This finding points to the possibility that firms using neural networks, an algorithm frequently discussed as needing colossal training data sets, may be more focused on data access than ethical concerns. Forth-five of the 58 firms in our study without ethical AI policies use neural networks in product development (78%). However, many firms (48%) are developing AI using neural networks, and the majority of those firms have ethical AI principles

A larger share of firms without ethical AI policies respond that data is the most important resource in their market, providing additional support for the possibility that training data's importance is related to the ethical AI policy adoption (Figure 4). Firms with ethical AI policies are more likely to respond that data provides a major advantage in their market, though it is not necessarily the most important resource. Moreover, we do not find a significant relationship between firms that rank human expertise as their most important resource establishing ethical AI principles.

Figure D.1 - Ethical AI Policy by Customer Industry

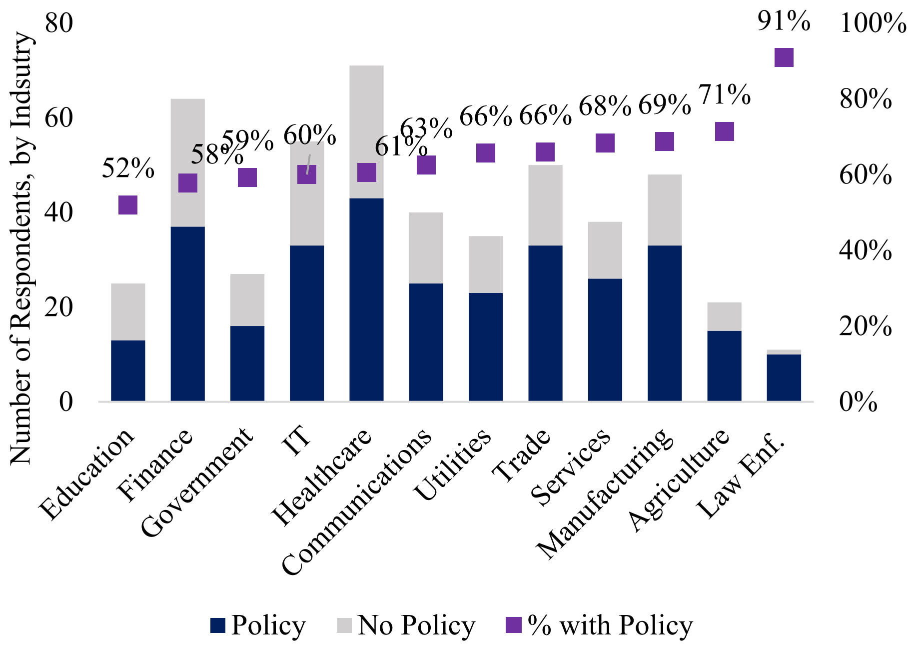


Figure D.2 – Ethical AI Policy by Data Protection

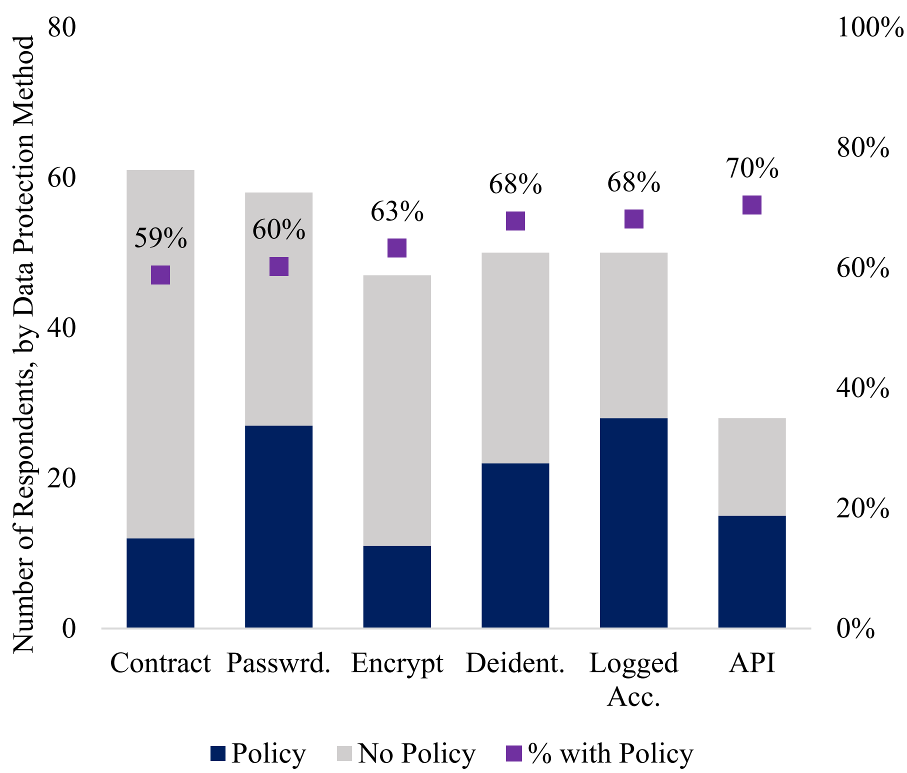


Figure D.3 – Ethical AI Policy by Algorithm Type Used

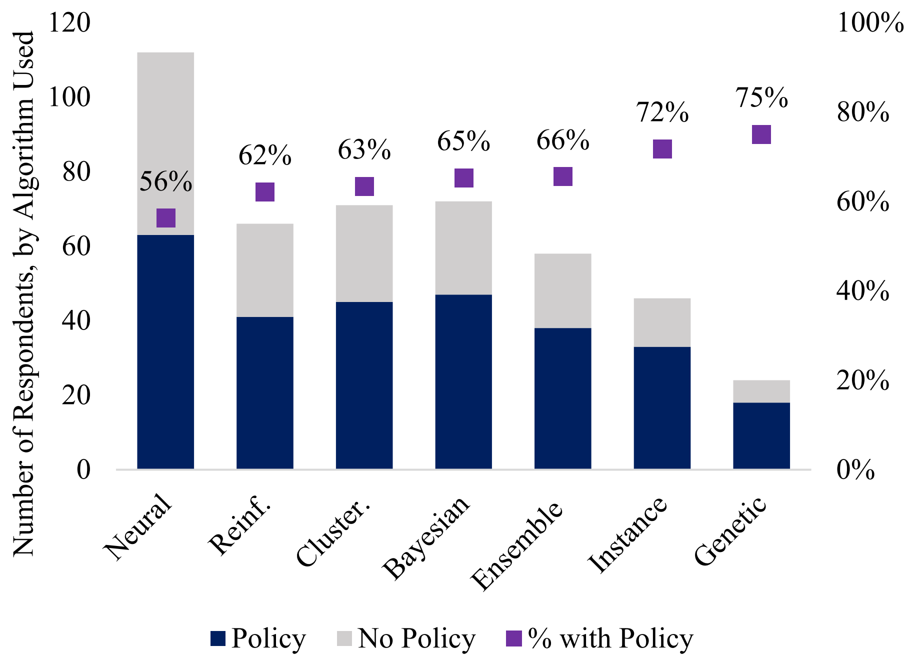
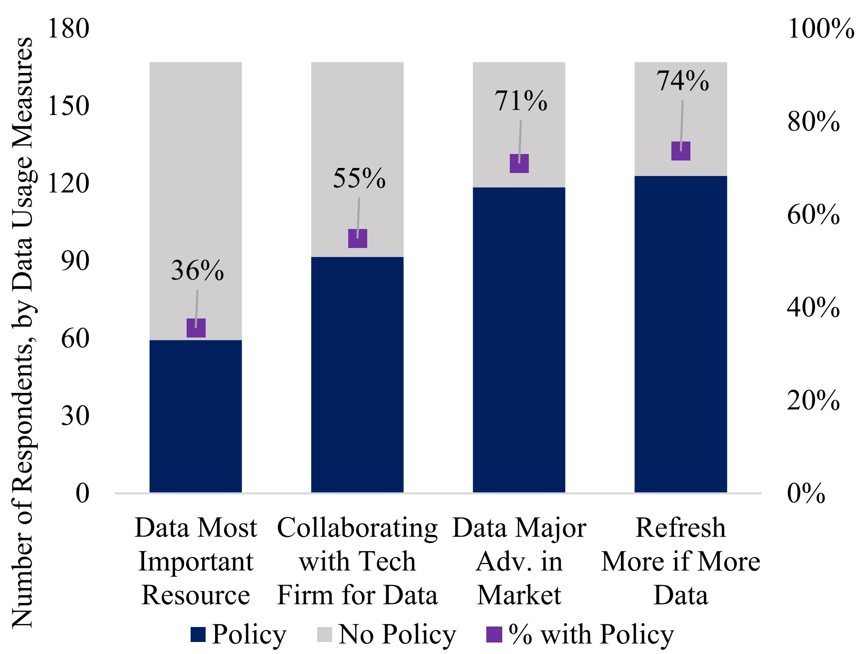


Figure D.4 - Ethical AI Policy by Data Usage Measures



Appendix E – CEM Matching

Table E.1 - CEM

	Treatment: Established an ethical AI policy?					
	Pre/All		Matched		Matched & Weighted	
	No	Yes	No	Yes	No	Yes
Young	0.75	0.56	0.77	0.56	0.56	0.56
US HQ	0.41	0.41	0.42	0.41	0.41	0.41
Small	0.37	0.28	0.35	0.28	0.28	0.28
Firms	160	160	160	160	160	160

Center on
Regulation and Markets
at BROOKINGS

The Center on Regulation and Markets at Brookings provides independent, non-partisan research on regulatory policy, applied broadly across microeconomic fields. It creates and promotes independent economic scholarship to inform regulatory policymaking, the regulatory process, and the efficient and equitable functioning of economic markets.

Questions about the research? Email communications@brookings.edu.
Be sure to include the title of this paper in your inquiry.