THE ETHICS OF BIASED ARTIFICIAL INTELLIGENCE: A STAKEHOLDER AND SHAREHOLDER THEORY INVESTIGATION

By

Ria Chinchankar

riac1@wharton.upenn.edu

An Undergraduate Thesis submitted in partial fulfillment of the requirements for the

WHARTON RESEARCH SCHOLARS

Faculty Advisors:

Prasanna Tambe

tambe@wharton.upenn.edu

Associate Professor, Operations, Information, and Decisions

Amy Sepinwall

sepin@wharton.upenn.edu

Associate Professor, Legal Studies & Business Ethics

THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA

MARCH 2022

ABSTRACT

This paper aims to investigate how shareholder and stakeholder theory apply to issues of biased artificial intelligence, specifically machine learning. It is done through an event study that measures whether a significant abnormal return in share price occurs on the day of a news leak about a biased AI event, which can be used to understand how market discipline and thus shareholder theory apply in this context. A sample of ten large publicly listed US-based technology companies was used. No consistent, statistically significant abnormal returns were found, which indicates no evidence that shareholder theory is an effective mechanism for preventing and responding to issues of biased artificial intelligence. This troubling finding prompts further exploration of stakeholder theory and government regulation as methods to ensure justice and fairness in algorithmic decision making.

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Research question

This paper aims to understand how shareholder and stakeholder theory apply to biased artificial intelligence, specifically, machine learning. Stories of biased artificial intelligence are increasingly in the news. Given that a significant proportion of the research in this field comes from corporations, it is worthwhile to investigate how corporations may respond to such issues. An event study was used to investigate whether market discipline and thereby shareholder theory is sufficient to correct and prevent instances of biased artificial intelligence.

1.2. Overview of the technology

Artificial intelligence is a term colloquially used to describe a host of technologies, but in this paper it will refer to decisions made by algorithms that are meant to mimic decisions typically made by humans. The most widely used of AI approaches— machine learning— takes input data, known as training data, that have a variety of variables. The goal is to predict a certain output variable using the other variables in the training data. This pattern is then applied to new data to predict an outcome. An example could be a machine determining what to serve for dessert—if it is hot outside, it could give ice cream, if it is cold outside, it could give a slice of pie. The machine could also learn that for those who are lactose intolerant, a popsicle should be served instead of ice cream. In this example, the input variables are weather and lactose sensitivity, the output variable is dessert choice. A sample data set could be the dessert choices of thousands of students at a dining hall. The decisions made or outcomes decided upon by an algorithm are a direct result of what patterns are in the training data. To put it simply, machine learning is coded, complex pattern recognition.

In real world instances, determining the impact each variable has on the outcome is nearly impossible; algorithms are notoriously opaque, sometimes referred to as a black box. Since the decisions made are based on patterns in the training data, any biases in the training data will be replicated in the algorithm. This leads to algorithms that produce outcomes which reflect the bias already present in society and in the data it was trained on. In “The New Jim Code,” Ruha Benjamin writes that “discriminatory practices are becoming more deeply embedded within the sociotechnical infrastructure of everyday life” (Benjamin 2020). Benjamin speaks to how this bias in algorithms mimics the bias in our society, but with minimal explanation, minimal accountability, and the authority of technology behind it. Importantly, humans design these systems. Safiya Noble writes that these technologies are often perceived as “benign, neutral, or objective” but they are not—the values of the people and organizations that develop these algorithms are reflected in the decisions they make (Noble 2018). This flawed mechanism is now being applied to make decisions in domains as wide reaching and consequential as education, hiring, health, and law enforcement. Bias within machine learning is the focus of this paper.

1.3. Ethics in artificial intelligence literature review

Examples of biased decisions are plenty, but a few especially egregious examples will be highlighted. An innocent person was incarcerated when blurry video camera footage mistook him for someone else committing a crime (Burke 2021). Someone with endometriosis was denied painkillers when an algorithm confused her pets’ prescriptions for hers (Szalavitz 2021). Families are denied home loans because of algorithms with racial bias built into them (Martinez 2021). This bias is unethical in that it denies people necessities due to error or discrimination on the basis of protected characteristics. Contextualizing these instances within broader literature relating to AI and ethics is important, as it helps establish where the field currently lies.

Despite much discussion, there is little consensus on what ethical AI means (World Health Organization, 2021). There are over 20 definitions for fairness, over 47 principles for the ethical use of AI, and dozens of frameworks developed by governing bodies across the planet (World Health Organization, 2021). There are competing methods about how to best minimize bias in AI, but no agreement on what to do.

These scattered approaches could lend themselves to a situation where developers choose the set of principles that best suit their needs under the guise of ethical AI, in effect paying lip service to the concept without rigorously ensuring equitable outcomes (Floridi 2019). It can also lead to organizations feigning ignorance on what ethical practices are, given that there are no definitively "right" processes, methods, and definitions. This heterogeneity is cause for concern.

At the moment, it is unclear whether corporations have developed systems to respond to such ethical dilemmas. Some companies hire staff as "ethics owners," people meant to translate broader principles into practice (Metcalf 2019), but the level of authority they have within the corporation is unclear. Google created an AI ethics council, but it was so inadequately resourced that Google canceled it (Knight 2019). Initial reviews of AI ethics tools and methods show confusion, a lack of consistency, and a need for a multidisciplinary approach (Morley 2020). Expert interviews of managers working in artificial intelligence have revealed that ethics is not a priority and that there are a variety of definitions of AI ethics (Baker-Brunnbauer 2021). Clearly, the existence of ethics principles is not enough to prevent ethical dilemmas, especially when there are no common aims or fiduciary duties, no professional norms, no proven methods to translate principles into practice, and no robust legal or professional accountability mechanisms (Mittelstadt 2019).

This concern becomes especially pressing when the pace of AI growth and the entities that house it are considered. Between 2011 and 2019, U.S. spending on artificial intelligence grew from under $300 million to roughly $16.5 billion (CB Insights 2020). Much of this investment lies within the hands of large corporations; a key example is Microsoft investing $1 billion in Open AI, an organization trying to develop artificial general intelligence in 2019 (Nellis 2019). In 2020, global total corporate investment in artificial intelligence reached almost $68 billion (Stanford University 2021). Furthermore, the non-universally accepted values of these technology companies often end up embedded in the decisions the algorithms produce (Noble 2018).

It is also worth asking whether the responsibility to create ethical outcomes can or should lie within corporations, or whether profitability should have any influence on decisions relating to ethics. This paper aims to understand how dominant theories of corporate social responsibility apply to issues of bias in machine learning.

1.4. Dominant corporate social responsibility theories: shareholder and stakeholder theory

There are two dominant normative perspectives regarding corporate social responsibility, or ideologies about how corporations ought to consider their impact on society. They are used to interpret the purpose of the corporation, including the identification of moral or philosophical guidelines for its operation and management (Donaldson 1995).

The first is shareholder theory, popularized by Milton Friedman in his 1970 essay "The Social Responsibility of Business Is to Increase Its Profits". The second is stakeholder theory, pioneered by R. Edward Freeman in his 1984 book *Strategic Management: A Stakeholder*

Approach. Both can be useful when thinking about artificial intelligence given that a large portion of this development and these applications lie within corporations, but impact entities outside of their direct control.

Friedman's shareholder theory dictates that the corporation's main duty is to maximize shareholder value, or profit. Any other activities that the corporation tries to pursue, namely, social initiatives, involve using money that belongs to shareholders for ends to which they did not consent. This is a problem in and of itself, but furthermore, the managers of a corporation are not well positioned to carry out social initiatives. Not only do they lack adequate knowledge of what society needs and how to provide it, but they are also free from checks and balances on their social policymaking. Whereas the government is built and trained to provide social services, a corporation is not, and it should avoid acting in domains outside of its profit maximizing activities. Friedman concludes that "while conforming to basic rules in society, rule of law, and rules of ethical custom," the corporation ought to focus wholly on increasing its profits (Friedman 1970).

Freeman's perspective is that managers have a duty to the corporation's stakeholders (Smith 2003). Stakeholders are defined as those groups who can affect or are affected by the corporation's acts (Freeman 1984). Primary stakeholders are those with direct and well-established legal claims on organizational resources, like shareholders or creditors (Jones 1999). Secondary stakeholders are those whose claims on organizational resources are less well established in law and/or are based on non-binding criteria such as community loyalty or ethical obligation, like dismissed workers or the natural environment (Jones 1999). Corporations have two responsibilities: to ensure that the ethical rights of no stakeholder are violated, and to balance the legitimate interests of the stakeholders when making decisions (Smith 2003). These interests must be considered as an end in themselves, even if they could reduce company profitability (Smith 2003).

To flesh out how these theories would counsel managers to act, consider a non-AI related example—namely, the question of whether to close a plant in a small town that employs a large proportion of the population in favor of moving manufacturing to another country with lower costs. Shareholder theory would dictate that the corporation should do the bare minimum that law mandates in the interest of workers, unless doing otherwise promises to raise share value. Stakeholder theory would involve true consideration of the dismissed workers, the impact on the environment the firm had, and how the town's government might continue to exist without taxes paid by the corporation.

1.5. Application of shareholder and stakeholder theory to biased AI

Both of these perspectives produce a different approach to AI development. A shareholder perspective would encourage optimization of the algorithm around profit, with controls in place only to prevent illegal activity. In this case, deciding what is illegal is tricky—who is ultimately responsible, and therefore liable, for the outcome of an opaque algorithm is a question with little legislation or precedent to answer it. With this perspective, little ought to be done to minimize bias or consider how the algorithm impacts the world it operates in. The only exception would be if it is profitable to have ethically minded algorithms. On the other hand, a stakeholder perspective would call for a careful consideration of how the algorithm would impact the various groups of society who would be subject to the decisions. It would mandate this broader focus because it is the ethical thing to do in and of itself, regardless of whether it is a precondition to profit.

The overlap between Friedman and Freeman's perspectives is an interest in shareholders and thereby profit. If a biased AI event threatens profits, shareholders will care about it, so therefore shareholder theory and market discipline could be useful as a doctrine underlying the ethics of AI in the corporations that currently develop it. If these biased events do not consistently threaten profits, it would indicate that shareholder theory and market discipline are insufficient mechanisms for minimizing bias in AI. This would point to further research needed in the realms of stakeholder theory or regulation to address the problem. Given the sheer volume of unethical incidents already reported and lack of prioritization of ethics mentioned in the literature, it appears that corporations leading AI development adopt the shareholder perspective. The hypothesis for the paper is

> **market discipline, and thereby shareholder theory, is not a sufficient mechanism for correcting and preventing instances of biased AI.**

1.6. Contribution statement

Understanding whether biased artificial intelligence results in abnormal market returns for large corporations is important because it helps us understand if the market imposes enough discipline to correct instances of biased AI. That determination could, in turn, provide grounds for an evaluation of what direction further research and policy should go in. Corporations are important to investigate since so much of the research in the field is funded by them. Examples are Microsoft's investment in Open AI, Google's Deep Mind, and Amazon's AWS.

Regulators, sociologists studying the impact of science on society, and tech ethicists will find this research useful. Determining how corporations respond to existing ethical pressures can depend on whether biased artificial intelligence threatens their market value and responsibility to shareholders. The outcomes of the event studies will help identify how high on the priority list biased AI is for corporations. If it is to mirror attitudes of managers in this field, the answer is that other tasks are higher priority (Baker-Brunnbauer 2021). If this is the case, it shows that accountability and a prioritization of ethics alone is unlikely to come from corporations, indicating the insufficiency of market discipline and shareholder theory. This should prompt a change of attitude in regulators; they would be more focused on disciplining firms than collaborating with them.

Additionally, understanding whether abnormal market returns result from biased artificial intelligence tells us the quantitative value of such errors, which enables the scale of the market perception of biased AI to be understood. If markets do not respond significantly to instances of biased artificial intelligence, it indicates a broader disregard from shareholders. This is especially true given that markets are supposed to factor in all information relating to a company and its impacts in the future. A non-result or only a small abnormal return could indicate that market-based solutions or those based on shareholder theory may not be the right path forward to artificial intelligence ethics, further indicating how this research could be used to shape regulation.

All in all, understanding if biased AI results in abnormal market returns can yield results that are useful for regulators, tech ethicists, and those interested in how technology will shape our society, especially given that so much of the fate of artificial intelligence lies in the hands of a select few powerful corporations.

2. METHODOLOGY

2.1. Suitability of an event study

An event study is a well-established analysis tool in accounting and finance that allows for the economic impact of an event to be measured using security prices over a fairly short period of time (MacKinlay 1997). Event studies have been used for a variety of purposes, from identifying the impact of a merger to how the value of a firm changes given a shift in the regulatory environment. It will be used to determine the quantitative market impact of biased AI news leaks; that is, whether biased AI impacts shareholders. A news leak in this context is a story in the news about instances of biased machine learning. The underlying assumption of an event study is that the event will impact the cash flow of a company. If the event does threaten cash flows, it negatively impacts shareholders, thus compromising a corporation's ability to maximize shareholder value. The domains in which biased AI can threaten a company's cash flow are primarily legal and reputational.

At this point in time, the primary legal threats to cash flows come from the expenses related to a lawsuit that a firm would incur. The lawsuits would be in the domains of privacy or anti-discrimination. An example of a privacy-related lawsuit is the ACLU suing Clearview AI, a company that trained its facial recognition algorithm on millions of images of people in Illinois who did not consent to their images being used for this purpose. This practice violated the Illinois Biometric Information Privacy Act, and the lawsuit has been ongoing since 2020 (ACLU 2020). An example of an anti-discrimination and AI related legal concern is the 2021 EEOC Initiative on Artificial Intelligence and Algorithmic Fairness (U.S. Equal Employment Opportunity Commission 2021). The Initiative aims to ensure that AI and similar technologies used in hiring comply with federal civil rights laws, and it comes as AI plays a growing role in hiring and labor markets (Friedman 2020). These lawsuits are expensive for corporations, with a 2010 US Chamber Institute for Legal Reform survey of Fortune 200 companies reporting that the average outside litigation cost per respondent was nearly $115 million in 2008 (Lawyers for Civil Justice 2010). Clearly, the cost of litigation would impact the cash flows of a corporation. However, there is a lack of major precedent and federal law governing AI, making it hard for markets to know how the courts might respond. The event study can shed further light on how markets interpret such events.

This is especially important given that the efficacy of fines has been questioned significantly. Fines are often seen as a method of punishing corporations for wrongdoing, but civil and criminal penalties "almost never result in a stock price decline on (or shortly after) the date of their announcement, even when the fine is so high… that it establishes a new record" (Coffee 2021). When using the same logic of the event study and believing that corporations align more closely with the shareholder approach, if a large fine does not impact share price, we may infer that the market declines to see the news issue as cause for concern.

Reputational costs of unethical practices can range from consumer boycotts to clients cutting ties. Examples include consumers boycotting Canada Goose jackets for using animal fur, the U.K. government stopping its contracts with Palantir after their poor privacy practices were exposed, and fewer people eating at Chipotle after E.coli outbreaks (BBC 2021, Shead 2021, Polansek 2015). In all of these cases, the companies lost some form of business over their questionable practices, indicating how a poor reputation can follow corporations or pressure them to change their policies. This loss of business means a loss of revenue, thereby impacting cash flows and future business prospects.

2.2. Event study overview

The output of an event study includes the mean abnormal return, a measure of the percentage by which the returns for the company are abnormal compared to what "normal" returns are. Security prices are an appropriate metric because according to the efficient market hypothesis, the market considers all current information and accounts for its future impact when pricing a share. Since this investigation focuses on artificial intelligence ethics news leaks, a daily events study would be most apt since the news leaks can be attributed to a single day. News leaks of a bias related event are the correct units of observation because event studies are centered around the day that information makes its way to the market.

Event study analysis consists of certain steps. The methodology requires that one identifies the event date, at which t = 0, the event window, the estimation period, and the gap between the estimation period and the event window (Liu 2020). The estimation period refers to a span of time used to determine what a "normal" return would look like for the companies in question. An abnormal return model is used, as shown in Equation 1.

Abnormal return:

$$AR_{i,t} = R_{i,t} - \alpha_i - \beta_i Rm_t$$

Equation 1: Market and Abnormal Return Model (Liu 2020)

Wharton Research Data Services provides access to a tool called Eventus that can perform event studies if provided with the right input data. Eventus will be used to conduct the event study. The input data consists of a sample of firms whose market responses on the date will be measured, the date of the news leak, the span of the event window in number of days, and the estimation period. Eventus also provides statistical significance data, which is useful in determining the confidence level of the output provided. A statistically significant abnormal return could imply that the news leak did have an impact on the returns of a company, indicating that there is some financial or market based incentive for firms to pay attention to and take action about some element of the biased AI. A no result could indicate that there is no evidence to believe that firms have a financial or market based incentive to pay attention to biased AI.

The sample used is a set of ten large tech companies. The criteria for being in the sample were to be publicly listed, in the technology industry, based in the US, and be of a comparable size to the companies whose events are in question (Apple, Amazon, and Google). The companies analyzed are Oracle, Microsoft, Facebook, Apple, Amazon, Google, Adobe, Nvidia, Salesforce, and Tesla. All of the estimation windows, the windows used to determine what a "normal" return is, are set to be before the onset of COVID-19.

Market model abnormal returns will be used for the results given that the firms in the sample will likely have similar return patterns to the market even on non-news days. These results use the market model to create a prediction of what normal returns would be, then subtract the expected return from the actual to get the abnormal figure. The alternative, market adjusted returns, is inadequate because the firms studied are all in the same industry and therefore do not represent the market. The uncorrected Patell Z score will be used to determine statistical significance of results given the way in which it distributes abnormal returns across the cumulated event window (Patell 1976).

2.3. The events analyzed

Three events have been chosen to be the focal point of three event studies. They all address issues relating to biased AI and relate to some of the most influential firms in the industry: Amazon, Apple, and Google.

In 2018, news outlets reported that Amazon had been using biased algorithms in their hiring processes from at least 2014 to 2015 (Dastin 2018). The algorithms marked resumes that used terms associated with women, like a women's organization or university, as weaker. If hiring decisions were made by directly following the algorithm, it would have violated the EEOC as well as ethical guidelines governing personal data ethics and fairness in employment (Kodiyan 2019). The hiring practice is consequential—determining who does and does not have access to employment on the basis of gender has large downstream implications in the futures of job candidates. This bias could impact cash flows by making future employees less willing to work for or with Amazon, a reputational cost. Given that the algorithm was not blindly followed, lawsuits relating to EEOC violations are unlikely. The event study aimed to determine whether these reputational consequences pose a threat to the market value of Amazon, or if given the sheer size of the company, it is seen as quantitatively inconsequential.

In 2019, users on Twitter began realizing that the Apple and Goldman Sachs credit card was offering men credit limits up to 20x higher than women with identical assets (Telford 2019). Married couples noted that wives were offered lower credit limits than their husbands, which is a biased outcome that significantly impacts peoples lives. Upon hearing this news, the New York State Department of Financial Services investigated the Apple/Goldman Sachs partnership in question, ultimately finding years later that there was no evidence of gender discrimination in the processes used to decide credit limits (Farrell and Nasiripour 2021). There could be reputational consequences in addition to the investigation; consumers might have been unwilling to sign up for the card or continue using it after hearing reports of the bias. Therefore the event study would reveal whether the investigation announcement relating to biased outcomes posed a threat to Apple's cash flow given the legal and reputational consequences that could hurt its market value.

Again, it would also reveal if these instances, while questionable, pose enough of a threat to such a large company for it to register as a difference in their market value.

In 2020, Google fired one of its top AI ethicists and diversity advocates, Timnit Gebru. She was “forced out” for her research highlighting dangers of certain natural language processing (NLP) models that Google was developing, some of which related to bias (Hao, 2020). Her research had the potential to threaten Google’s business model, given that NLP is a huge element of Google’s search engine. This can pose reputational consequences for Google, and an event study will determine if these consequences are of a high enough magnitude for the market to be concerned and respond accordingly.

All of these events were reported within the span of a few days; there was no build up or story developing over a large span of time. This limited time span of the news of biased AI makes these events appropriate for an event study, given that a suitably narrow event window can be constructed.

2.4. Method validation

Event studies are highly sensitive to the estimation period and event windows, as demonstrated by Hood in his event study examining the impact of Tiger Woods’ personal scandals on the market value of his sponsors (Hood 2012). In order to ensure the methodology and parameters of the event study were appropriate, two control cases were run: Bill Gates leaving Microsoft on March 13 2021 and the news leak for the Cambridge Analytica/Facebook scandal on March 17, 2018 (Haselton 2020, Cadwalladr 2018). These events were picked for the occasions they represent: a sudden leadership change and an ethics-related news leak. The parameters used mirrored those employed in the studies of the biased AI related events, in that for events happening during the pandemic, a pre-pandemic estimation period was used. Across all of the events examined, the only variables that changed were the date, to reflect the event, and the estimation period, to avoid any pandemic-induced irregularities in calculating what a "normal" return might consist of. Table 1 indicates the estimation periods used for the events.

Table 1: Control group events and estimation period

Event	Estimation period	Biased AI event with same estimation period used
Bill Gates leaving Microsoft	Ends 300 days before the event date; 255 days in length.	Google
Cambridge Analytica/Facebook	Ends 46 days before the event date; 255 days in length.	Amazon, Apple

When evaluating the significance of the results, the primary data point will be the two-day cumulative average abnormal stock return, mirroring the data used by Karpoff et al (1999). The results are displayed in Table 2.

Table 2: Control group date, mean cumulative abnormal return, and uncorrected Patell Z

Event	Date	Mean cumulative abnormal return for the (-1, 0) window	Uncorrected Patell Z
Bill Gates departure	March 13, 2021	8.38%	12.297***
Cambridge Analytica scandal	March 19, 2018 ^	-1.25%	-2.657**

^The news leaks happened on March 17, but given the stock market being closed, March 19 was used as Day 0

The symbols ** and *** denote statistical significance at the 0.01 and 0.001 levels, respectively, using a generic one-tail test.

Table 3: Control group mean abnormal returns and statistical significance levels for Day 0 and Day +1

Event	Day 0 mean abnormal return	Day 0 Significance level	Day +1 mean abnormal return	Day +1 Significance level
Bill Gates departure	1.25%	0.001	4.48%	0.001
Cambridge Analytica scandal	-0.82%	0.01	-0.59%	0.01

In these two events, there are significant results over the two-day window in Table 2, indicating that the event study parameters show results consistent with expectations for the events in question. In Table 3 there is a pattern of sustained significance in the same direction after the event, indicating consistency in how the market perceived these events. Consistency refers to the same trend visible in every data point presented, from the (-1,0) window to the Day 0 and Day +1 mean abnormal return. This control group validates the event study set up. The Microsoft example reflects strong leadership given how the transition was announced; the Cambridge Analytica/Facebook scandal represents the unethical nature of the event, its political implications, and the weaknesses in Facebook’s policies.

3. RESULTS

Three separate event studies were conducted: one for the 2018 Amazon biased algorithm, one for the 2019 biased Apple Card, and one for the 2020 Google termination of a high profile AI ethics researcher. The mean cumulative abnormal return for the (-1, 0) window, in keeping with Karpoff's results presentation, is in Table 4. Table 5 demonstrates the Day 0 and Day +1 mean abnormal return as well as significance level.

Table 4: Biased AI event date, mean cumulative abnormal return, and uncorrected Patell Z

Biased AI owner	Event Date	Mean cumulative abnormal return for the (-1,0) window	Uncorrected Patell Z
Amazon	10 October 2018	0.12	-0.556
Apple	19 November 2019	0.7	0.627
Google	4 December 2020	-1.46	-2.486**

Table 5: Biased AI event mean abnormal returns and statistical significance levels for Day 0 and Day +1

Biased AI owner	Day 0 mean abnormal return	Day 0 Significance level	Day +1 mean abnormal return	Day +1 Significance level
Amazon	-0.68	0.05	1.83	0.001
Apple	0.3	N/A	0.05	N/A
Google	-1.21	0.01	1.23	0.1

Amazon's (-1, 0) window was insignificant. The Day 0 return was significant and negative, but upon closer inspection, there were other events with news releases on the same day that impacted the sample size, such as the Postal Service proposing a 10% price increase that would impact Amazon's costs, increased regulatory scrutiny facing Google, and speculation about Tesla's new chairman (Factiva 2022). The Day +1 return was significant and positive, but again, other events that could impact the cash flows happened on the same day. Amazon Pay began a new partnership to make it easier for customers to purchase subscription offerings, Facebook took down hundreds of misinformation pages, news of fraud relating to Apple IDs was released, and Apple began bolstering in-house chip design capabilities (Factiva 2022). Given that a value-weighted index was used, the news of Apple developing in-house chip design capabilities could have been enough to result in a significant, positive, mean abnormal return due to Apple's large size. There was no news about Amazon's biased AI impacting other firms in the sample. The Factiva database was used to search for other potentially impactful events.

Apple's results were insignificant in all regards. During this time, it was reported that Apple's stock had a 71% return, that Facebook was going through an antitrust probe, that Tesla's stock went up, that Apple's stock hit a new high, that expectations for Apple were looking positive given the roll out of 5G, and that there were rumors of Apple receiving a China tariff exemption (Factiva 2022). In a value-weighted index with Apple being one of the largest firms and such overwhelmingly positive news about the stock, Twitter accounts of bias for a card that is backed by Goldman Sachs simply did not have enough power to make a significant dent in the share price, even with the possibility of an investigation present.

Google had a significant negative return for the (-1, 0) window and for Day 0. This could be due to further context behind the termination being released in the news. However, other major news developments were also reported during the relevant window. Apple signed a $600m deal with Dialog Semiconductor as part of its efforts to bring chip production in house, which impacts competing firms in the hardware industry, including Google (Factiva 2022). Day +1 showed a result significant to the 0.1 level, with the mean abnormal return being 1.23%. While not significant to the desired level, it indicates a roughly equal and opposite reaction, which could indicate the share price quickly recovering from any potential negative effect of the Gebru firing.

To summarize, some of the results indicated a negative, significant abnormal market return for the sample on the day of, in the aftermath of, or in the two-day window of a biased AI news leak. However, there was no news about the impact of one firm's actions on the other, there was no consistent negative or positive signal persisting to Day +1, and there were always other events happening on the same day that could pose a threat to the cash flow of corporations in the sample. The event studies have yielded no result.

4. ANALYSIS AND EXTENSIONS

4.1. Analysis

The event studies for biased AI news leaks for Amazon, Apple, and Google yielded no result. Any evidence of a significant abnormal return was not robust or consistent enough to conclude that biased AI resulted in a reduction in the share prices of ten large, publicly traded technology companies based in the US.

There is no evidence that market discipline and thereby shareholder theory are sufficient mechanisms for correcting and avoiding instances of biased AI, since the event studies yield no significant, consistent negative abnormal return.

Furthermore, there is no evidence to suggest that under shareholder theory, corporations have an incentive to rectify and prevent bias in AI. If the duty of a corporation is to maximize shareholder value and there is no evidence that biased AI has a significant, consistent impact on share price, then under shareholder theory it follows that there is no evidence for corporations to care about biased AI, so long as no laws are violated. Using profit as a motivator to care about biased AI will not work if biased AI does not in fact threaten profits, as the event studies suggest. Market discipline is not shown to be effective in this domain.

Since this problem is of such a large magnitude, other methods of addressing it need to be investigated and implemented. Potential solutions could lie within stakeholder theory or government regulation.

4.2. Exploration of stakeholder theory: avenues for further research

A central premise of stakeholder theory—that all stakeholders have equal importance—is useful when thinking through algorithm development. AI calls for an expansion in the definition of a stakeholder based on how an algorithm is used and who it impacts. A few concepts within stakeholder theory are especially important: incorporation and dependent stakeholders. Incorporation is the idea that stakeholders are not just consulted, but brought onto the board of directors (Jones 1999). A powerful idea is giving those most impacted by an algorithm decision making power. This is supported by Benjamin, who writes that "if the training data is produced by a racist society, it won't matter who is on the team, but the people who are affected should also be on the team"—advocating for a broader variety of stakeholders to be considered in the algorithm development process (Benjamin 2020). An example of this would be a hospital allowing a random sample of residents in zip codes they serve to vote on whether and how an algorithm is used. A dependent stakeholder is a stakeholder without power, but with legitimacy and urgency (Mitchell 1997). These dependent stakeholders could be those severely impacted by algorithms but who have little power. Increased attention should be brought to them.

However, such approaches seem unlikely in practice. Even in non-biased AI-related situations, profits are repeatedly the top concern. A prime example is the deaths of Amazon workers who were told to keep working at a fulfillment center in the middle of a tornado (Reuters 2021); similar instances are easy to find. Even when corporations do take steps to do good for society, such as by donating to charities, ethical standards related to profit-generating activities are different. Indeed, Herman (1981) writes that "a firm can have formal stakeholder management programs without generating substantive stakeholder management outcomes," and Jones elaborates that "once publicly owned, a firm becomes hostage to performance criteria established by the financial markets, which value optimum economic returns rather than social responsibility—unless social responsibility pays" (Jones 1999). If all stakeholders were truly considered, the events relating to biased AI would not have happened. Further, ethical scandals at these companies, even unrelated to biased AI, would have been avoided had there been a true adoption of stakeholder theory.

Accountability, responsibility, and justice, especially relating to biased AI, will not solely come from these large tech corporations who are at the forefront of research. Many sets of ethical principles attempting to identify what needs to be done to harness AI for good have the same conclusion: that there is a need for enforceable regulation. The event studies support this conclusion. Corporations cannot be called upon to self-regulate or prioritize dependent stakeholders. The urgency for comprehensive regulation is further emphasized. Such regulation would be industry-wide, providing the same structural pressures on all firms (Jones 1999).

4.3. Policy recommendations and prescriptions

Designing this regulation is no easy task. The Algorithmic Accountability Act is a step in the right direction, asking companies to assess the impacts of automating decisions, calling for more federal standards in assessment and reporting, and giving people more ability to know what decisions relating to them are automated and how (Chu, 2022). Yet, there is room for improvement. Giving companies the responsibility to report gives them significant ability to hide information. Further, even if consumers have information about the algorithms making decisions, algorithms are inherently opaque and understanding them requires a level of tech literacy that cannot be expected from the general public. In order for regulation to be effective, more authority over these large corporations needs to exist. If they exist as they are, reporting could become a PR task that the largest corporations have the resources to conduct. Rather than just reporting, legislatures should determine where accountability lies and hold decisions made to the same standard as those by wholly human processes.

Furthermore, questioning the sources of research and development in the field is essential. If the bulk of research lies within the corporation, the research produced will reinforce profit maximization, as opposed to building a technology designed to benefit society. By contrast, if public entities that are independent from large corporations became major players in research, that could change the industry standard of what is considered ethical.

Additionally, when building standards and laws, multidisciplinary teams must be employed. The training of a computer engineer cannot include all the historical context required to understand how bias plays out in society, and therefore cannot inform what areas to check for when evaluating the outcomes of the algorithms.

Moreover, training data, where much of the bias can originate, should have much stricter standards. There should be more rigorous limits on the type of data that can be used to predict certain outcomes, with humanities and social science experts as well as those directly impacted by algorithms playing a large role in dictating these limits.

5. CONCLUSION

Biased AI is a growing problem. Algorithms are being employed to make decisions in fields impacting every facet of life, from pain medication dosage to incarceration to home ownership. The biases in the algorithms reflect the biases in our world given that the training data for the algorithms represents real decisions. Problems with biased AI are even more concerning given the authority of technology, the opacity of an algorithm's inner workings, and the power of the large companies developing this technology.

A significant portion of this investment, research, and development lies within large corporations. When it comes to matters relating to corporate ethics, there are two main fields of thought: shareholder theory, pioneered by Milton Friedman in 1970, and stakeholder theory, fathered by R. Edward Freeman in 1984. The former dictates that corporations should care only about profit, while the latter states that corporations ought to consider all stakeholders equally. The overlap between these two is an interest in profit.

If a consistent link or signal between biased AI and profit was found, it would follow that there could be some mechanism under which corporations would take action against biased AI while adhering to shareholder theory. To investigate whether such a link exists, an event study was used, a quantitative finance analysis to see if there was a significant abnormal market return associated with a particular date for a particular sample size of companies.

The events investigated were instances of controversial practices at Amazon, Apple, and Google relating to biased AI. The event studies yielded no result; that is, there was no evidence of a consistent, sustained, directly attributable impact that biased AI has on profit. It follows that there is no evidence that shareholder theory is a sufficient ideology if the ethical use of AI in society is desired. It is also worth questioning whether profit should be a consideration in ethics related decisions at all.

Given this result, alternatives to shareholder theory must be considered. Stakeholder theory could provide concepts useful for thinking through the problems. However, the current ethical behaviors of large corporations in matters relating to AI and other issues indicate an unwillingness to alter profit-generating activities for the sake of ethics, despite significant efforts to donate to and support nonprofits.

This reality prompts further thought of what comes next. This technology is wide-reaching and must be regulated, but the manner in which this happens must be carefully considered. Regulation must not rely on the public attaining advanced technology literacy. Nor should it yield PR reports that large corporations can easily finance. Finally, regulation should be crafted with a multidisciplinary approach; it should recognize training data as one of many origination points of bias; and it should promote more research and development housed outside of non-publicly accountable corporations.

References

2022. Alphabet, Apple, Amazon, Adobe, Meta, Oracle, Microsoft, Amazon, Nvidia, Tesla. Retrieved 27 February 2022, from Factiva database

ACLU. 2020. ACLU v. Clearview AI. *ACLU.org* (May 27).

Baker-Brunnbauer, J. 2021. Management perspective of ethics in artificial intelligence. *AI and Ethics* (1).

BBC. 2021. Canada Goose to end the use of all fur on coats. *BBC* (28 June).

Benjamin, R. 2020. Race after Technology: Abolitionist tools for the New Jim Code. *Social Forces* 98: 1-3.

Burke, G., Mendoza, M., Linderman, J., and M. Tarm. 2021. How AI-powered tech landed man in jail with scant evidence. *AP News*

Cadwalladr, C. and E. Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian* (17 March).

CB Insights, & PwC. (April 13, 2020). Artificial intelligence (AI) funding investment in the United States from 2011 to 2019 (in million U.S. dollars) [Graph]. In *Statista*. Retrieved March 06, 2022

Chu, K. 2022. Wyden, Booker and Clarke Introduce Algorithmic Accountability Act of 2022 To Require New Transparency and Accountability for Automated Decision Systems. *Ron Wyden United States Senator for Oregon* (February 3).

Coffee, J.C. 2021. Crime and the Corporation: Making the Punishment fit the Corporation. Available at SSRN: https://ssrn.com/abstract=3914961

Dastin, J. 2018. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters* (October 10).

Donaldson, T. and L.E. Preston. 1995. The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications. *Academy of Management Review* 20(1).

U.S. Equal Employment Opportunity Commission. 2021. EEOC Launches Initiative on Artificial Intelligence and Algorithmic Fairness. *EEOC.gov* (October 28).

Farrell, G., and S. Nasiripour. 2021. Goldman Cleared of Bias in New York Review of Apple Card. *Bloomberg* (March 23).

Floridi, L., and J. Cowls. 2019. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*.

Freeman, R.E. 1984. Strategic Management: A Stakeholder Approach. Boston: Pitman.

Friedman, G.D., and T. McCarthy. 2020. Employment Law Red Flags in the Use of Artificial Intelligence in Hiring. *American Bar Association* (October 1).

Friedman, Milton. 1970. A Friedman Doctrine—The Social Responsibility of Business Is To Increase Its Profits. *New York Times* (September 13).

Hao, Karen. 2020. We read the paper that forced Timnit Gebru out of Google. Here's what it says. *MIT Technology Review* (December 4).

Haselton, T. and J. Novet. 2020. Bill Gates leaves Microsoft board. *CNBC* (March 13).

Herman, E. 1981. Corporate Control, Corporate Power. *Cambridge University Press.*

Hood, M. 2012. The Tiger Woods scandal: a cautionary tale for event studies. *Managerial Finance* 38:5.

Jones, M.T. 1999. The institutional determinants of social responsibility. *Journal of Business Ethics* (20).

Karpoff, J.M, Lee, D.S, and V.P. Vendryzk. 1999. Defense Procurement Fraud, Penalties, and Contractor Influence. *Journal of Political Economy* (107).

Knight, W. 2019. Google employees are lining up to trash Google's AI ethics council. *MIT Technology Review* (April 1).

Kodiyan, A.A., 2019. An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool. *Researchgate Preprint*.

Lawyers for Civil Justice, Civil Justice Reform Group, and the U.S. Chamber Institute for Legal Reform. 2010. Litigation Cost Survey of Major Companies. *Duke Law School.*

Liu, X. 2020. Introduction to Event Study. *Wharton Research Data Services*.

MacKinlay, A.C. 1997. Event Studies in Economics and Finance. *Journal of Economic Literature* 35 (1): 13-39.

Martinez, E. and L. Kirchner. 2021. The Secret Bias Hidden in Mortgage-Approval Algorithms. *The Markup* (August 25).

Metcalf J, Moss E, and D. Boyd. 2019. Owning ethics: Corporate logics, Silicon Valley, and the institutionalization of ethics. Soc Res. 82(2):449–76.

Mitchell, R.K., Agle, B.R., and D.J. Wood. 1997. Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts. Academy of Management Review. 22(4).

Morley, J., Floridi, L., Kinsey, L., and A. Elhalal. 2020. From what to how: an initial review of publicly available AI ethics tools, methods, and research to translate principles into practices. *Science and engineering ethics* 26: 2141-2168.

Nellis, S. 2019. Microsoft to invest $1 billion in OpenAI. *Reuters* (July 22).

Noble, S. 2018. Algorithms of Oppression: How Search Engines Reinforce Racism. *NYU Press.*

Patell, J.M. 1976. Corporate Forecasts of Earnings Per Share and Stock Price Behavior: Empirical Test. *Journal of Accounting Research* (14).

Polansek, T. 2015. Americans aware of Chipotle outbreak eat there less often: poll. *Reuters* (22 December).

Reuters. 2021. Sadness, anger for Amazon workers who died during tornado. *Reuters* (17 December).

Shead, S. 2021. UK government ends one of its data contracts with Palantir. *CNBC* (September 10).

Smith, J.H. 2003. The Shareholders vs Stakeholders Debate. *MIT Sloan Management Review* 44(4).

Stanford University. 2021. Global total corporate artificial intelligence (AI) investment from 2015 to 2020 (in billion U.S. dollars) [Graph]. In *Statista*. Retrieved March 06, 2022

Szalavitz, M. 2021. The Pain Was Unbearable so Why Did Doctors Turn Her Away? *Wired* (August 11).

Telford, T. 2019. Apple Card algorithm sparks gender bias allegations against Goldman Sachs. *The Washington Post* (November 11).

World Health Organization. 2021. Ethics and governance of artificial intelligence for health.