AI & SOCIETY
https://doi.org/10.1007/s00146-023-01648-7

OPEN FORUM

A phenomenological perspective on AI ethical failures: The case of facial recognition technology

Yuni Wen[1] · **Matthias Holweg**[1]

Received: 22 May 2022 / Accepted: 13 March 2023

Abstract
As more and more companies adopt artificial intelligence to increase the efficiency and effectiveness of their products and services, they expose themselves to ethical crises and potentially damaging public controversy associated with its use. Despite the prevalence of AI ethical problems, most companies are strategically unprepared to respond effectively to the public. This paper aims to advance our empirical understanding of company responses to AI ethical crises by focusing on the rise and fall of facial recognition technology. Specifically, through a comparative case study of how four big technology companies responded to public outcry over their facial recognition programs, we not only demonstrated the unfolding and consequences of public controversies over this new technology, but also identified and described four major types of company responses—Deflection, Improvement, Validation, and Pre-emption. These findings pave the way for future research on the management of controversial technology and the ethics of AI.

Keywords AI ethics · Public controversy · Company responses · Facial recognition · Case study

1 Introduction

Artificial intelligence (AI) has been widely adopted in many areas in business and society, such as improving efficiency in manufacturing and scientific research, automating processes in the financial service sector, and augmenting human decision-making in medical and legal contexts. Yet this powerful technology is also embroiled in ethical failures, and several high-profile cases have been widely reported: Facebook, for example, was being widely scorned for sharing users' data with Cambridge Analytica, a consultancy company that conducted AI-based political campaigns; Google's TensorFlow has been attacked by the public for helping the U.S. Defense Department analyze drone footage with AI (Holweg et al. 2022). When an AI project openly violates social norms and values, it would provoke public controversy (Ouchchy et al. 2020); the ensuing controversy has not only the power to cause lasting damage to corporate reputation, but also potentially to spark regulatory reform that puts an end to such technology altogether (Hekkila 2021).

Many recent attempts have been made to develop theoretical underpinnings on ethical use of AI technology and consensus has been achieved on several ethical principles of AI, such as fairness, transparency, accountability and autonomy (e.g., Floridi and Cowls 2019; Choung et al. 2022). Recent initiatives have begun to translate these high-level principles into practice, such as the "human-in-the-lop" methods and ethics-based auditing protocol (e.g., Floridi et al. 2022; Morley et al. 2022). However, scholars so far have paid relatively little attention to how companies respond to address the ethical issues of AI. Stahl et al. (2022) highlighted the importance of empirical studies of company responses in this field. Through case studies, they identified company responses *within* the organizations—organizational awareness, technical approaches, human oversight, ethical training, and balancing competing goods—to mitigate the ethical risks of AI. Complementary to Stahl et al. (2022)'s study, this paper focus on company responses to the public *outside* the respective organization. Specifically, we examine how companies respond to AI ethical failures, which we define as the AI applications that have been deployed and caused public criticisms by violating social norms and values.

✉ Yuni Wen
yuni.wen@sbs.ox.ac.uk

Matthias Holweg
matthias.holweg@sbs.ox.ac.uk

[1] Said Business School, University of Oxford, Oxford, UK

Published online: 01 April 2023

Empirically, we focus on the case of facial recognition technology, which has seen a rapid rise in the last decade following major developments in deep neural networks applied to this problem (LeCun et al. 1989; LeCun and Bengio 1995). Facial recognition technology uses biometric data from a still or moving image, or a real-time camera feed, to identify and verify individuals. As such, it has been widely applied in financial services, security, policing, and social media applications (Hamann and Smith 2019). The ethical failures related to facial recognition technology commonly relate to aspects of *privacy* with regards to image data used for training and prediction, *bias* with regards to discrimination against certain parts of the population, and *safety* with potential malicious use of the technology (Smith and Miller 2022). The ethical debate about facial recognition reached its peak in May 2020, when the "Black Lives Matter" movement led to widely voiced public concerns about biased treatment of the black population as the result of adopting facial recognition in law enforcement. Under this public pressure, most technology companies either stopped facial recognition projects entirely or issued promises not to sell facial recognition technology to law enforcement agencies (Heilweil 2020).

We study how leading companies respond to the ethical failure of facial recognition technology is the subject of our investigation. We conducted a comparative case study of four leading technology companies—Google, Microsoft, IBM and Amazon—that all faced public controversy over their respective use of facial recognition technology. We explore the nature of the ethical failures they were facing, as well as how they responded to the ensuing controversy. We found that these companies, though facing very similar criticisms at the same point in time, display very distinct reactions—ranging from an accommodative stance to a defensive one. We identified four types of organizational responses to AI ethical failure: pre-emption, validation, improvement and deflection. Based on these findings, we had attempted to elucidate the determinants of a company's response to ethical AI failure, and propose three antecedents for further research.

2 Literature review

2.1 Technology failure and public controversy

Failure is generally understood as the state of not achieving a desirable objective (Merriam-Webster 2022), yet what is count as desirable varies by different groups and value systems (Akaka and Parry 2019). Failure in the context of technologies is "a product of judgement" (Appadurai and Alexander 2019: 1) and featured with "interpretive flexibility" (Bijker and Pinch 1987: 40). It means that technology failure is a matter of social contestation in which different stakeholders try to make their interpretation. For instance, by engineers' criteria, failed technologies refer to those without the quality of workability, reliability and efficiency (Gooday 1998). From the point of view of shareholders, only the technology that obtains a big market share can be treated as success (Braun 1992). In the eye of customers, technology failure simply means misfit for their individual purposes or needs (Pye 1987). More recently, in the narratives of entrepreneurs in Silicon Valley, technology failure is not an obstacle but a steppingstone on way to success (Appadurai and Alexander 2019).

Relatedly, technologies should be treated as a sociotechnical system, which consists of not only artifacts but also social practices, social arrangements, and social relationships (Kline 1985; Tonkinwise 2016). Therefore, resistance to a new technology is affected by the wider society in which potential users are embedded (Slowlkowski and Jarratt 2007). Failure can be the result of organizations or institutions in denial (Easterling 2016). As Bruland (1995: 24) notes, the study of resistance to technology is about "interaction between the technology and its social context". For example, a technology can fail at the organizational level, because technology development can be halted by cultural susceptibility to innovation, interdepartmental competition and lack of R&D input (e.g., Calantone et al. 1993; Souder and Sherman 1993; Schilling 1998). A new technology could also fail in the adoption/diffusion process when potential users may not know about the technology, or decide to reject the technology after being informed (Rogers 1962).

Social resistance to a new technology often takes the form of public controversies, a situation in which the public express their disagreement (Venturini 2010). Technology has become an increasingly important part of the "list of contestable issues" (Feenberg 1999: vii). Recent examples include the controversies over social media technologies due to their intrusion on privacy, and the malicious use of drones (Zwickle et al. 2019). The central message in this stream of research is that public controversies "are not simply collisions between the public on one side and the scientists and/or industry on the other"; instead, both the "public" and "industry" are heterogenous (Cambrosio and LImoges 1991: 388). In line with this tradition, our study will not only seek to unpack the public controversies over the AI technology, but also understand the heterogeneity in the response from organizations affected.

2.2 The ethics of AI

The main thrust of the research into the ethics of technologies has technical in nature, first and foremost seeking to develop tools and techniques to assess and prevent ethical issues. Research into the question how companies make

ethical choices in developing and deploying these technologies, however, is nascent (Martin et al. 2019). The key insight of this emerging field is that technology is never value free, and it is the responsibility of technology firms to assess and address the ethical consequences of the technology in use. AI has not only increased the efficiency and effectiveness of many tasks, but also replaced parts (or all) of human decision-making (Dafoe 2018). Ethical issues are, thus, an inherent concern for any company developing and operating it. With the rapid development and ubiquitous use of AI systems, the focus has extended from AI functionality toward AI ethics (Raji et al. 2022), such as privacy violation (Mazurek and Małagocka 2019), biased prediction (Gebru 2020) and lack of explainability (Doran et al. 2017). As a result, it exposes organizations to ethical failure of AI, which we define as a situation where AI violates social norms and triggers public criticisms (Holweg et al. 2022). On this definition, we need to clarify two points. First, social norms of AI are phenomenologically constructed. In other words, ethical principles of AI are generated in response to a series of scandals in association with AI applications, such as Google's TensorFlow, Uber's autonomous vehicles and Amazon Rekognition. Second, to constitute an AI ethical failure, such norms of are violated in a way that it evokes wide criticisms in the public space, such as media, social media and policy debates. In other words, AI ethical failure is featured with certain level of salience and publicity.

In response, companies, research institutes, government bodies, and NGOs have largely focused on making high-level ethical principles on the use of AI, such as the "Ethics Guidelines for Trustworthy AI" issued by the European Union (2019) or Google's (2018) statement entitled "Artificial Intelligence at Google: Our Principles". While a plethora of ethical principles has been published, too numerous to review here, several common core principles for AI ethics are emerging: transparency, accountability, responsibility, fairness and autonomy (Floridi and Cowls 2019; Vakkuri et al. 2019; Khan et al. 2021).

Yet devising a list of high-level ethical principles per se is not immediately useful in practice: When it comes to specific cases, the principles are often coming into conflict with each other; different groups may interpret the principles differently; and organizations may find the principles too vague to operationalize (Whittlestone et al. 2019; Arvan 2018). It has also been pointed out that AI ethics cannot exist in a company without a broader culture of ethics (Lauer 2021) and companies' commitment to ethics conflicts with the industry commitments to meritocracy, technological solutionism, and market fundamentalism (Metcalf and Moss 2019). Therefore, there are many recent governance initiatives to translate the high-level principles into practices (Morley et al. 2022), such as the "human-in-the-lop" methods that introduce human operators to intervene to prevent harmful impacts (Lin et al. 2020), ethics-based auditing protocol (Floridi et al. 2022), and specific responsibility assignments to AI engineers (Rochel and Evéquoz 2021). Most of these governance initiatives are based on experimental, hypothetical ideas, however, and ethical considerations may be compromised when they clash with commercial incentives that value efficiency and profit (Mittelstadt 2019; Lauer 2021).

What is amiss still is empirical evidence that considers how organizations address AI issues in reality. Stahl et al. (2022) have taken an initiative in summarizing a set of mitigation strategies employed by AI companies from ten case studies. They have insightfully identified five company responses *within* the organizations to mitigate the ethical risks of AI, including organizational awareness, technical approaches, human oversight, ethical training, and balancing competing goods. However, they have not touched upon how companies respond to the public *outside* the organizations. This is a significant omission, given that there are an increasing number of public scandals with regards to the use of AI (Ouchchy et al. 2020) and such criticisms initiated by external stakeholders can cause long-lasting damage to corporate reputation (Holweg et al. 2022). It is this gap that we seek to address by focusing on AI ethical failures—a special circumstance where the AI application has caught public attention and criticisms for violating social norms.

2.3 The ethics of facial recognition technology

Facial recognition is an AI-based technology to identify people by capturing the features of a face in a video or image and comparing it with an existing database of human faces. What is common to most facial recognition technology applications is the use of deep convolutional neural networks, which results in a need for large and representative training datasets to achieve consistent, accurate, and unbiased predictions (Hamann and Smith 2019). Over the last decade, this technology has experienced a booming growth based on advances in deep learning, while being marred by public controversy at the same time. It was found to generate biased predictions and violate privacy, and its application in policing caused concerns on a potential threat to democracy and freedom. In 2020, facial recognition run into a great setback when the European Commission proposed to ban the technology in the public space (BBC 2020), and big tech companies announced to stop selling the technology to the police (Heilweil 2020). Public outcry has since prompted companies to respond to mitigate reputation loss (Elsbach 2003; Pfarrer et al. 2008).

3 Research design and methodology

We conducted a comparative case study of companies engaged in facial recognition technology to analyze company responses to controversial AI technology. We selected four big technology companies—Amazon, Google, IBM and Microsoft[1] that have been involved in developing, adopting and selling facial recognition technology as an important part of their respective AI cloud platforms: Amazon AWS, Google Cloud, IBM Watson and Microsoft Azure. Their facial recognition projects have been widely reported and discussed on media. They all faced criticisms related to privacy violations, biased predictions, and/or malicious use of facial recognition. Despite facing very similar public criticisms, the companies' reactions were very different. In our study we explore and analyze the differences in the companies' responses to public controversy over their facial recognition technologies.

As a first step, we examined the consequences of companies' responses. We compared the change in public sentiment and financial returns for each of the four companies before and after the major public outcry. To measure financial returns, we used abnormal stock returns measured during the time of public outcry to capture investors' reactions to and perceptions of company responses (Feldman et al. 2016; Zhang and Wiersema2009). We considered the weekly abnormal returns—the difference between actual returns and normal (or expected) returns on a given week—using an eight-week (−4 weeks to +4 weeks) event window, with the event (Day 0) being the earlies critical media report of the particular facial recognition program. The stock market price data was provided by Yahoo Finance. To measure public sentiment, we adopted the Thomson Reuters MarketPsych Indices. MarketPsych generates sentiment scores of news articles for all the major firms in the world based on textual analysis that identifies the valence of references to a firm. The negative numbers reflect negative sentiment relative to similar firms, positive numbers reflect positive relative sentiment. Similarly, using abnormal return calculation, we analyzed sentiment scores for a focal firm in the four weeks before and after the public outcry. However, we only could identify short-term backlash variations from the mean for both aspects, yet no significant long-term effects could be isolated. The graphs of abnormal stock returns and sentiment scores are presented in Appendix 1.

We then proceeded to codify the company responses. As response strategies generally capture what firms "say and do" after criticism (Coombs 2007: 170), we collected our data from media reports and company documents. First, we obtained relevant media reports from two media databases: *Google News*, which focuses on online news, and *Factiva*, which is a collection of newspapers. We searched the keyword "facial recognition" along with a company name in the two databases for the timeframe between 2010 and 2020. We further filtered those media reports by only including the ones that either covered public reactions or company responses to facial recognition and the ones from widely recognized media, such as Forbes, Medium, CNBC, and Reuters. Second, we obtained these companies' statements on facial recognition in their official websites, including company presentations, strategic announcements, public letters, and articles written by managers. To complement the data on the longitudinal aspects of company documents, we used an internet archive tool *WayBack Machine* to trace the deleted company statements on their websites. We collected 166 relevant media reports and 48 company documents. Our data analysis followed an iterative process, which is "the process by which a researcher moves between induction and deduction practicing the constant comparative method" (Suddaby 2006: 639), combining insights from process studies (Langley 1999) and cross-case comparison (Eisenhardt 1989). Our data analysis was specifically divided into two parts: first, we analyzed the activities and quotes from stakeholders, including academics, civil organizations, governments departments and media comments, from which we identified what kind of public controversies emerged; second, we looked at the activities and quotes from the four companies, from which we summarized company responses. We then built a timeline for each case mapping out stakeholder and company actions (see Appendix 2).

We used axial coding to classify the first-order codes into more abstract aggregate dimensions. It involved cross-case analysis, identifying differences and similarities among reactions from key stakeholders and responses from the four companies. In so doing, we combined a process-based logic within each case and a variance-based logic across the four cases. This variance in process analysis provides the basis for our findings. We presented our coding process in Fig. 1. As a conclusion of the process, we identified distinct types of company responses to controversial facial recognition technology deployment, as discussed in the next section.

4 Case findings

4.1 Amazon Rekognition

In November 2016, Rekognition was released as part of Amazon Web Service (AWS) to provide machine learning-based vision capability to cloud customers (Amazon 2016).

[1] Apple and Facebook were also accused of similar issues with regard to their FRT, yet we excluded them because they only developed and deployed facial recognition for use within their own products, instead of selling it as a standalone technology.

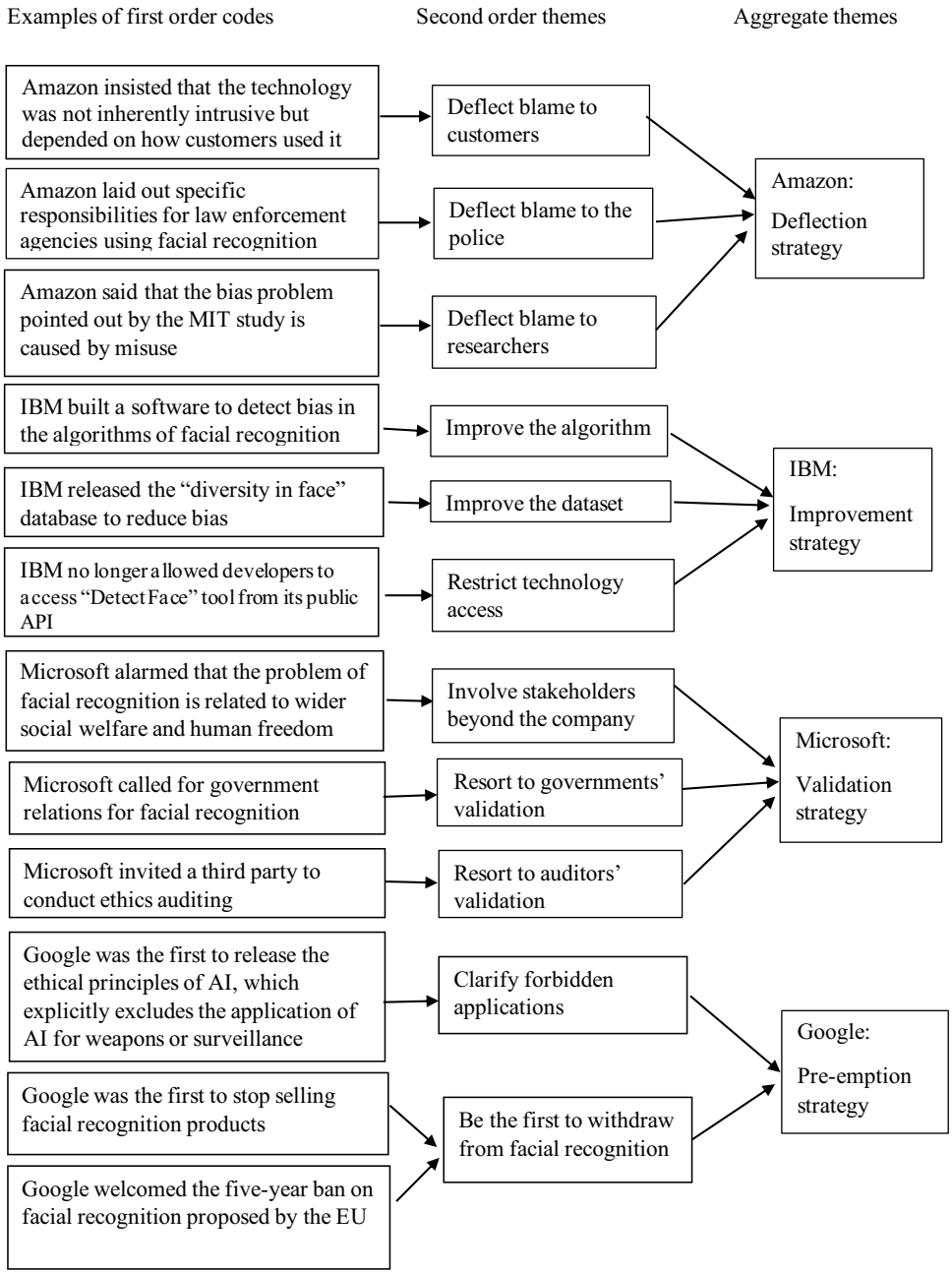


Fig. 1 The coding process

Rekognition was expected to help attract customers to different cloud platforms on AWS, increasing the both revenues and keeping customers engaged for longer periods. In November 2017, Amazon announced a series of updates to Rekognition, including real-time face searches in an existing database (Amazon 2017). The newly developed function was believed to target primarily law enforcement agency customers.

In May 2018, the American Civil Liberties Union (ACLU) was first to publicize a report that Amazon was marketing Rekognition to Washington County for suspect identification. The ACLU criticized that Rekognition was a surveillance system which posed a potential threat to freedom and democracy (Dwoskin 2018). It was soon followed by a public petition with 150,000 signatures and an employee protest within the company, demanding to stop the sale of facial recognition services to the U.S. government agencies (Vincent 2018).

In response, Amazon deflected blame to customers. The company insisted that the technology was not inherently intrusive but depended on how customers used it, stating:

"Our quality of life would be much worse today if we outlawed new technology because some people could

choose to abuse the technology. Imagine if customers couldn't buy a computer because it was possible to use that computer for illegal purposes? Like any of our AWS services, we require our customers to comply with the law and be responsible when using Amazon." (Lee 2018)

Like its competitors, Amazon Rekognition soon found itself involved in bias accusations. In July 2018, ACLU conducted a test of Rekognition to recognize members of the US Congress and it incorrectly identified 28 members of Congress as criminals. Nearly 40 percent of Rekognition's false matches in the test were of people of color, even though they make up only 20 percent of Congress (Snow 2018). In January 2019, the MIT study extended its investigation on commercial facial recognition programs to Amazon Rekognition. They found that Rekognition performed even worse in identifying the gender of female and darker-skinned faces than similar programs from IBM and Microsoft; in their study Rekognition was classifying the gender of dark-skinned women 31.4 percentage points less accurate than that of light-skinned men (Singer 2018).

Amazon immediately attributed the bias problem to the misuse of technology by researchers. In a statement by Matt Wood (2019), general manager of artificial intelligence of AWS, he pointed out that the studies did not properly use Rekognition service and that Amazon had found no differences in error rates by gender and race when running similar tests. In a blog post, Michael Punke, Vice President of Amazon's Global Public Policy, reiterated this stance:

"New technology should not be banned or condemned because of its potential misuse. Instead, there should be open, honest, and earnest dialogue among all parties involved to ensure that the technology is applied appropriately and is continuously enhanced." (Punke 2019)

Amazon emphasized the responsibility of law enforcement agencies for the responsible use of facial recognition technology. It proposed guidelines for any law enforcement agency using facial recognition, suggesting law enforcement agencies ought to arrange human review to ensure no violation of civil rights by the facial recognition technology, to be transparent about how they use the technology, and to give notice when video surveillance was used (Punke 2019).

The pressure reached a peak in May 2019, when a group of Amazon shareholders put forth a proposal to prohibit sales of facial recognition technology to governments and study how it might threaten civil rights and people of color. The board of directors was recommending shareholders vote against the proposal, hailing facial recognition as "a powerful tool for business purposes, but just as importantly, for law enforcement and government agencies to catch criminals, prevent crime, and find missing people" (PHYS 2019a). In the end, the proposal was voted down at the company's annual shareholders meeting (PHYS 2019a).

Despite the mounting pressure, Amazon Web Service CEO Andy Jassy publicly defended the value of facial recognition for government use and claimed that Amazon would continue to sell the facial recognition technology to government customers. He further called for federal laws to limit the misuse of facial recognition software, and issued a policy that customers who misuse the technology will be barred from using the platform. He said that "(..) whether it's private-sector companies or our police forces, you have to be accountable for your actions and you have to be held responsible if you misuse it" (quoted in PHYS 2019b). Our analysis finds that after the series of public announcements in 2019, Amazon's stock price remained stable, but its public sentiment suffered from a temporary decline.

However, the death of George Floyd in May 2020 and the consequent "Black Lives Matter" movement triggered a public discussion on police abuse of facial recognition technology. Unsurprisingly, Amazon declared its strong support for the movement. Amazon called for an end to "the inequitable and brutal treatment of black people" on Twitter and put a "Black Lives Matter" banner at the top of its home page. Jeff Bezos posted an email from a customer criticizing the BLM banner on Amazon's home page, and said the emailer is the kind of customer he's "happy to lose" (Paul 2020). But this statement was criticized as empty and hypocritical while the company was maintaining contracts with law enforcement agencies. At this time, Amazon announced a one-year moratorium on police use of Rekognition, and a year later, the company decided to extend the ban indefinitely (Allyn 2020).

4.2 IBM Watson's facial recognition algorithm

IBM was known for AI technological breakthrough in 2011, when IBM Watson defeated two human champions in the quiz game show "Jeopardy!". The company set out to use the promising technology to solve "society's big problems" beyond the capability of other computers, such as healthcare and poverty (Markoff 2011). However, in the following years, the technological capability shown off in the TV show had not been turned into the expected set of commercial applications (Hesseldahl 2014).

In January 2018, a group of MIT researchers tested facial recognition algorithms from IBM and Microsoft and found both skin color and gender bias. They suggested that the algorithms' error rates in determining the gender of light-skinned men were never worse than 0.8 percent. For darker-skinned women, however, the error rates of IBM ballooned to more than 34 percent, compared to 20 percent of Microsoft. This contrasts with an accuracy rate of more than 97

percent claimed by IBM about their facial recognition system (Hardesty 2018).

IBM issued a proactive response to the MIT study. It not only acknowledged the virtue of the study but also promised to make technological improvement on both datasets and algorithms. Ruchir Puri, the chief architect of IBM's Watson artificial-intelligence system, replied through media:

> *"This is an area where the data sets have a large influence on what happens to the model. We have a new model now that we brought out that is much more balanced in terms of accuracy across the benchmark that Joy (researcher) was looking at. It has a half a million images with balanced types, and we have a different underlying neural network that is much more robust. It takes time for us to do these things. We've been working on this roughly eight to nine months… She was bringing up some very important points, and we should look at how our new work stands up to them."* (Hardesty 2018)

Since, IBM made a series of technological improvements in addressing the accuracy and bias problem; it also reached out to the researchers at MIT to figure out how to fix its bias problems. Five months later, IBM said that they had achieved a nearly tenfold decrease in error rate for facial analysis. To help improve the training of facial recognition and reduce bias in algorithms, IBM announced a plan to make the largest facial attribute and identity training set in the world, with more than a million images, and the dataset would be equally distributed across skin tones, genders, and age (IBM 2018). In September, IBM launched a software to analyze how and why algorithms make decisions, as well as to detect bias and recommend changes. In the meantime, a technical workshop was held by IBM Research in collaboration with the University of Maryland to identify and reduce bias in facial analysis (Burt 2018). In January 2019, IBM fulfilled its promise to tackle the bias problem of facial recognition by releasing the "Diversity in Faces" dataset, which contained the promised one million images that were meant to sample a more diverse group of faces. In a press release, the company said it hoped the dataset would "advance the study of fairness and accuracy in facial recognition technology" (Smith 2019).

However, while IBM was devoted to pursuing the value of fairness in facial recognition, it ignored other equally important values—the very attempt to build a more diverse database turned out to be a privacy violation. The media reported that the "one million photos" of the new IBM dataset were scraped (identified and downloaded automatically by a search algorithm) from Flickr without the consent of users, and were shared with outside researchers. It put the company in a media storm on privacy violation. One of the most widely cited reports is from NBC, which used a provocative headline—"Facial Recognition's dirty little secret: Millions of online photos scraped without consent, (to power technology that could eventually be used to surveil them)" (NBC 2019).

In the responding statement IBM reiterated its commitment to privacy: "We take the privacy of individuals very seriously and have taken great care to comply with privacy principles". The company explained that the data was only open to verified researchers, and that Flickr users can choose to opt out of the database. The company also included a list of steps users can take if they want their photos removed from the dataset (Liao 2019). Then, in September 2019, IBM quietly removed the "Detect Face" tool from its public API, so that developers could no longer simply buy access to the company's facial recognition. It also started gradually rolling back facial recognition for existing clients (IBM 2019a). However, according to our analysis, these actions failed to increase the company's stock price, and its public sentiment experienced a minor decrease before going up again.

In June 2020, in the wake of widespread protests over the killing of George Floyd, IBM announced that it was no longer researching, developing, marketing, or selling facial recognition tools to any client, and is not using the technology itself. IBM CEO Arvind Krishna wrote a letter to the Congress about the Racial Justice Reform bill and declared the company's position:

> *"IBM firmly opposes and will not condone uses of any [facial recognition] technology, including facial recognition technology offered by other vendors, for mass surveillance, racial profiling, violations of basic human rights and freedoms, or any purpose which is not consistent with our values and Principles of Trust and Transparency."* (IBM 2019b)

4.3 Microsoft azure face API

Microsoft initially adopted facial recognition technology as a complementary function to its own products or entertainment programs. For example, in 2010, Microsoft launched it as an Xbox add-on to log users into their live accounts (Carmody 2010). In 2015, the company rolled out several entertainment facial recognition programs, such as an app to guess a person's age with just one photo, which created a trend on social media to share guessed ages (Newcomb 2015). In 2017, it launched Face API as a part of its cloud computing service Azure to enable customers to access facial recognition technology (Microsoft 2017).

Like IBM, Microsoft's facial recognition algorithm was accused of race and gender bias by the MIT study published in January 2018. It found that Microsoft had a 20.8 percent error rate gap for identifying the darker-skinned women and light-skinned men (Hardesty 2018). But unlike

IBM's immediate response, Microsoft kept silent for five months after publication of the study. In a company blog post, Microsoft announced that it had significantly reduced the error rate for darker-skinned population and women by diversifying the training data. The company statement went beyond a mere technical discussion of the problem at hand, raising a more nuanced management challenge how and when to interfere and mitigate AI systems that reflect and amplify decisions made in a biased society (Roach 2018).

In response, Microsoft involved stakeholders beyond the company by framing the issue as an infringement of human wellbeing. In July 2018, Microsoft publicly raised the alarm about a potential problem with facial recognition that "goes beyond bias itself, raising critical questions about human freedom". President Brad Smith acknowledged technology companies' ethical responsibility in this context, and said that Microsoft had already rejected some customers' requests to deploy the technology in situations involving "human rights risks". He further called on government to regulate facial recognition technology. He wrote:

> *"Facial recognition technology raises issues that go to the heart of fundamental human rights protections like privacy and freedom of expression. These issues heighten responsibility for tech companies that create these products. In our view, they also call for thoughtful government regulation and for the development of norms around acceptable uses*." (Smith 2018a)

In December of that year the company issued ethical principles for the use of its facial recognition technology: fairness, accountability, non-discrimination, notice and consent, transparency and lawful surveillance. It said it would bar the technology from being used to engage in unlawful discrimination, and would encourage customers to be transparent when deploying such services (Microsoft 2017). In the meantime, the company's president Brad Smith made a speech named "Facial Recognition: It's Time for Action", reiterating the urgency to have a government regulation (Smith 2018b).

In January 2019, a group of 90 advocacy groups sent a letter to big tech companies, including Microsoft, requesting that the companies pledge not to sell facial recognition technology to governments (Catro 2019). Three months later, Microsoft said it rejected a California law enforcement agency's request to install facial recognition technology in officers' cars and body cameras. The company was concerned that it would lead to innocent women and minorities being disproportionately held for questioning, because the system had been trained on mostly white and male pictures (Menn 2019). Microsoft also deleted its massive database of 10 million facial images[2] which was being used to train facial recognition systems (BBC 2019).

Microsoft also invited influential third parties to conduct ethics auditing. For example, in October 2019, when media reported that Microsoft had funded the Israeli facial recognition startup AnyVision, which secretly had surveilled Palestinians in the West Bank. Commentators soon questioned the alignment between what Microsoft claimed and what it actually did. In response, Microsoft hired former United States Attorney General Eric Holder to conduct an audit of AnyVision, to determine whether it complies with Microsoft's ethical principles on the use of facial recognition technology. Although the independent audit later found out that AnyVision had not engaged in a mass surveillance program in the West Bank, Microsoft decided to divest from AnyVision and stop all investments in third-party facial recognition companies (Solo 2019).

Although Microsoft adopted an increasingly cautious approach toward facial recognition technology, it never intended to withdraw from the game entirely. In January 2020, Microsoft publicly expressed its reservation over the European Commission's proposed moratorium on using facial recognition technology in public areas. Brad Smith cited many unreplaceable benefits of the technology, such as finding missing children. He said that it was important to first identify problems and then craft rules to ensure that the technology would not be used for mass surveillance: "There is only one way at the end of the day to make technology better and that is to use it" (quoted in Chee and Chalmers 2020). This stance is illustrated by the facial recognition law passed in the State of Washington, which was heavily lobbied by Microsoft. Smith hailed it as "a significant breakthrough—the first time a state or nation has passed a new law devoted exclusively to putting guardrails in place for the use of facial recognition technology" (Smith 2020). Our analysis suggests that there were only short-term turbulences in the company's stock price and sentiment at this point, but no dramatic change in the long run.

After the emergence of the "Black Lives Matter" movement, hundreds of workers at Microsoft were calling on the company to cut ties with law enforcement customers, and to do more to demonstrate its commitment to anti-racism efforts. Following its competitors, Microsoft stated that it was not currently providing the technology to police, and would not do so until there were federal laws in place that would regulate this technology. Thus, unlike IBM, it did not rule out the possibility of selling the technology to police forces, but calls for regulation first (Statt 2020).

[2] To put this into context, the primary image database used for the training of AI systems, ImageNet, contains a total of 14 million images of labeled objects of all kinds.

4.4 Google cloud vision API

Google, for the most part, is only engaged in face detection without disclosing the identity of the person. According to the Google Vision API documentation, it only supports face detection that detects multiple faces within an image, along with the associated key facial attributes like emotional state or wearing headwear (Google 2021). This did not shield the company from public controversy, however: when Google Photos was launched in 2015, a machine-learning technology to automatically group photos with similar content was embedded. One month after its launch, a user post a tweet saying that the software categorized the pictures of him and his friends as "gorillas" and this post trigged more than 15,000 re-tweets. Google intervened immediately but could not find a solution. A public apology was issued, acknowledging that "there is still clearly a lot of work to do with automatic image labeling" and promised to prevent the mistakes from happening in the future (quoted in BBC 2015).

In January 2018, Google released an entertainment selfie tool that matched users' faces with historical portraits in museums. Yet avid users soon found out that the app tended to match faces to euro-centric art featuring white faces. Google said that it would expand its partnership with more museums around the world to bring diverse cultures. When people raised privacy concerns, Google promised not to store nor use user photos for any other purposes (Paul 2018).

Another wave of public outcry hit the company two months later. It started from a small group of Google employees internally raising ethical concerns over Google's collaboration with the U.S. Department of Defense to use AI technology to detect objects from drone footage. The case was soon reported by major newspapers, including The Guardian, New York Times, BBC, and Fortune (e.g., Vanian 2018). Thousands of Google employees signed a petition calling the company to cancel the military contract. In the petition letter, they warned that working with the military would harm Google's brand and reputation (Shane and Wakabayashi2018). In response, the company explained that it was limited to helping the military with non-offensive tasks and said the project would help save lives (Gibbs 2018). In June 2018, the head of Google Cloud Service announced that the company would not renew their contract with the Department of Defense when it expires a year later (Wakabayashi and Shane 2018).

Since, Google has switched from adopting passive stances to proactively leading the industry in limiting the use of facial recognition technology. In June 2018, Google became the first company to lay out principles for responsible AI, which include ensuring social benefits and avoiding bias and privacy violation. It also explicitly excludes the application of AI for weapons or surveillance (Google 2018). In December of that year, it in announced a stop to selling facial recognition products until they could put in place policies to prevent abuse of the technology. Walker Kent (2018), VP of the company, wrote in a blog post:

> *"…like many technologies with multiple uses, facial recognition merits careful consideration to ensure its use is aligned with our principles and values, and avoids abuse and harmful outcomes. We continue to work with many organizations to identify and address these challenges, and unlike some other companies, Google Cloud has chosen not to offer general-purpose facial recognition APIs before working through important technology and policy questions."* (Kent 2018)

In 2019, Google was alleged to have collected face scans from people with darker skin tones in exchange of gift cards, mainly homeless people and college students. A Google spokesperson responded that they were taking these claims seriously and investigating them. He also explained that the purpose of data collection was to diversify data to improve the face unlock feature for the Pixel 4 phone. Subsequently, the company decided to suspend facial recognition research for Pixel 4 altogether (Hamilton 2019).

In January 2020, Google expressed welcome for the EU proposal to impose a five-year ban on facial recognition. In contrast to Microsoft's reservation, Google CEO Sundar Pichai voiced support by stating "I think it is important that governments and regulations tackle it sooner rather than later and gives a framework for it". He cited the possibility that the technology could be used for malicious purposes, and one area of concern is so-called "deepfakes"—video or audio clips that have been manipulated using AI. Pichai said that Google had released open datasets to help the research community build better tools to detect such fakes (cited in Chee and Chalmers 2020). Our analysis indicates that there was only a short-term decline in the company's stock price before it bounced back, and that its public sentiment remained steady after these actions.

Finally, in the wake of the "Black Lives Matter" movement in June 2020, Google announced to stop selling facial recognition technology to police and reiterated its pioneering role in restraining the use of the technology:

> *"We were the first major company to decide, years ago, to not make facial recognition commercially available and we have very clear AI Principles that prohibit its use or sale for surveillance." (Google spokesperson, quoted in* Dickey 2020)

4.5 Cross-case analysis

Following the abductive approach, we moved back and forth between data and theory, and categorized these first-order events into more general concepts relevant to our research

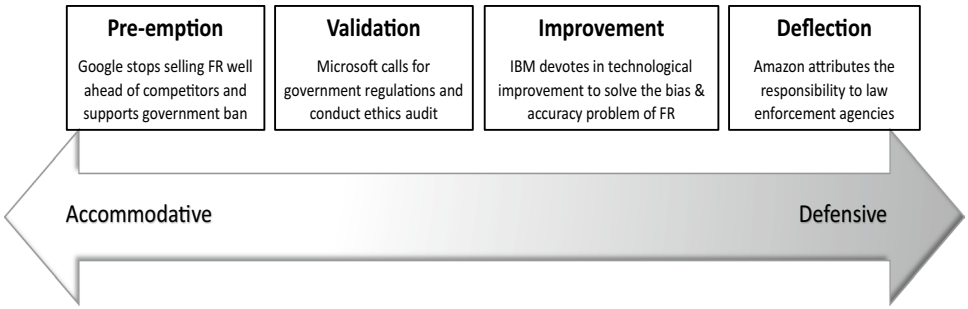


Fig. 2 Spectrum of company responses to public controversy over facial recognition technology

(Suddaby 2006). Our comparison of four technology companies engaging in facial recognition technology shows a spectrum of company responses. Past reputation research has categorized response strategies as ranging along a spectrum of accommodativeness, "(..) with endpoints of accepting responsibility/remediation and denial of a crisis." (Coombs 1998: 179–180). The more accommodative response refers to the firm proactively accepting responsibility (Elsbach 2003), such as apologies, expressions of regret, and promises of action. In contrast, more defensive strategies try to disassociate a firm with the misconduct, and may include excuses, justifications, and denials (Bundy et al. 2021). We identified four distinct types of company responses to controversial AI technology along this spectrum, as outlined in Fig. 2.

At the most defensive end sits the response by Amazon, which vigorously defended the benefits of facial recognition and attributed the reported wrongdoing as the consequences of misuse. The literature suggests that criticisms of wrongdoing are inherently ambiguous in that external actors find it hard to evaluate the situation (Faulkner 2011), and this it is especially the case for the new AI technology where there is no consensus on who should be responsible. Such ambiguity provides space for companies to strategically attribute the misconduct to others in order to reduce reputation loss (Boeker 1992). Amazon proactively deflected the responsibility to users—mainly businesses and law enforcement agencies—to ensure benevolent outcomes, such as the company guidelines on facial recognition that required law enforcement agencies to ensure transparency and no violation of civil rights by technology. We classified this response strategy as *deflection*.

Further along the spectrum, IBM's response moves slightly away from the defensive end in that it recognized the criticism and made constant efforts to improve the technology. According to Mishina et al. (2012), stakeholders make two primary types of reputational assessments of an organization. One type is *capability reputation*, where stakeholders are concerned about what the organization is capable of doing so that they judge it by its abilities and resources. Capability reputation damage entails a focus on technical fixes (Holweg et al. 2022). In our cases, IBM highlighted its technological progress in addressing the accuracy and bias problem of facial recognition, for example, by extensively working with academics, upgrading algorithms to reduce errors, and diversifying data, although the latter attempt backfired on it with the privacy issue. We labeled this approach as *improvement*.

Microsoft's response in comparison is situated even further towards the accommodative end. Microsoft openly acknowledged the risks of facial recognition and the responsibility of technology companies, but it opposed a complete ban. Unlike capability reputation, people would also pay attention to whether companies' intentions and goals are benevolent or malevolent, which is referred to as *character reputation* (Mishina et al. 2012). To avoid damage to character reputation, firms often seek to associate with a powerful or influential third-party actor, which can validate whether the firm is legitimate and credible (Prashantham et al. 2020). As such, Microsoft consistently called for interventions from the governments to regulate the facial recognition technology and determine what applications are valid; it also invited independent auditors to ensure their use of facial recognition is ethical. We categorized this response strategy as *validation*.

Finally, Google's approach is located at the most accommodative end of the spectrum, as it not only accepted the criticisms but also turned to support a ban on facial recognition technology. Although most literature considers how to repair reputation in the aftermath of substantial damage being observed, scholars generally agree that communication work can start before the actual crisis (Coombs 2010). According to the crisis life cycle framework developed by Wilcox and Cameron (2009), there are phases when all appear calm and when issues that are likely to emerge as crisis trigger points are occasionally identified. This is the time when firms can adopt pre-emptive measures, for example by differentiating themselves from others and seeking self-preservation, in order to prevent reputation

loss (Pang 2012). Google took pre-emptive moves before the peak of the public crisis: It announced to stop selling facial recognition well ahead of government bans and its competitors, and further pioneered in making ethical principles in the application of AI. Therefore, we labeled this response strategy as *pre-emption*. We summarize the four response strategies along the spectrum in Fig. 2 below.

5 Discussion

In this study, we have investigated the phenomenon of how four leading technologies firms that were developing, using or selling facial recognition have responded very differently to public criticisms. We have characterized their response strategies as: deflection, improvement, validation and pre-emption – ranging from the most accommodative approach to the most defensive approach. As the AI technology exposes organizations to new and potentially damaging controversy and public outrage, our findings can help organizations by laying out the strategic options at hand to mitigate potential reputation loss.

Although our case study analysis is exploratory and, thus, does not permit conclusively identifying the reasons for the difference in response, we would like to propose three possible antecedents for further research. First, one likely antecedent is the financial importance of the technology to the company, i.e., how much revenue has the technology brought, or is expected to contribute. A company would be reluctant to concede to public pressure when it conflicts financial interests. Amazon Rekognition, for instance, was considered financially important to the company. In 2018, less than two years after its launch, the sales of Rekognition had already accounted for around $3 million of Amazon's $25.7 billion in cloud revenue and was expected to deliver rapid revenue growth with public sector clients (Dastin and Bera 2020; AWS 2021). Driven by its financial prospect, we posit that Amazon was prone to defending the technology by attributing the responsibility to users.

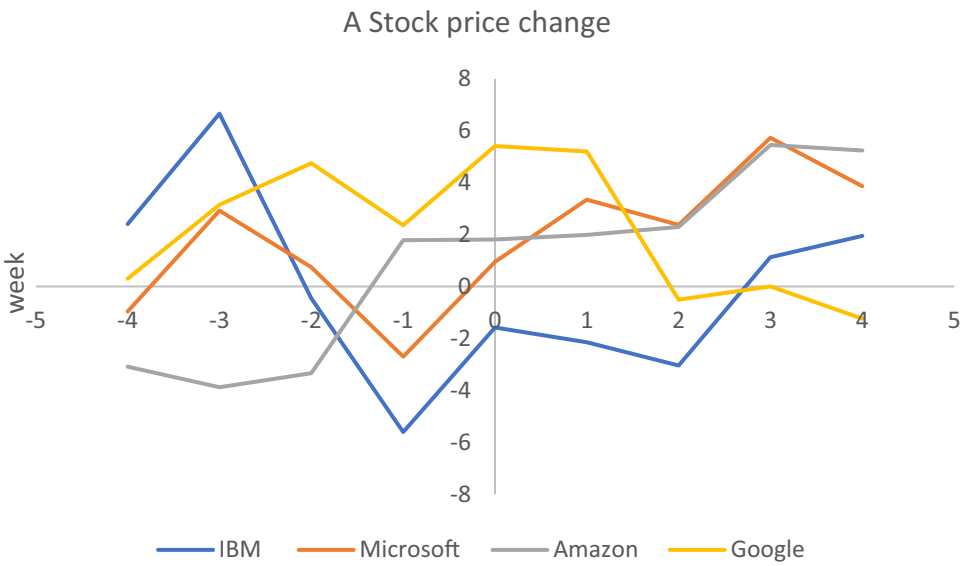
Second, we propose that company responses to controversial technology can also be influenced by the strategic importance of the technology—Is it peripheral or core to the product and service offering of the firm? Giving up a core product or service would challenge the very foundation of a company and therefore the company would be less receptive to public criticisms. IBM Watson, for example, has dominated the company's growth strategy since 2014 (Power 2014), and the facial recognition program herein (with the potential to bring government clients on board) was considered as an important part of Watson (Harries 2015). We posit that this position has incentivized IBM to "stay in the game" and keep improving the technology, in order to be able to retain it.

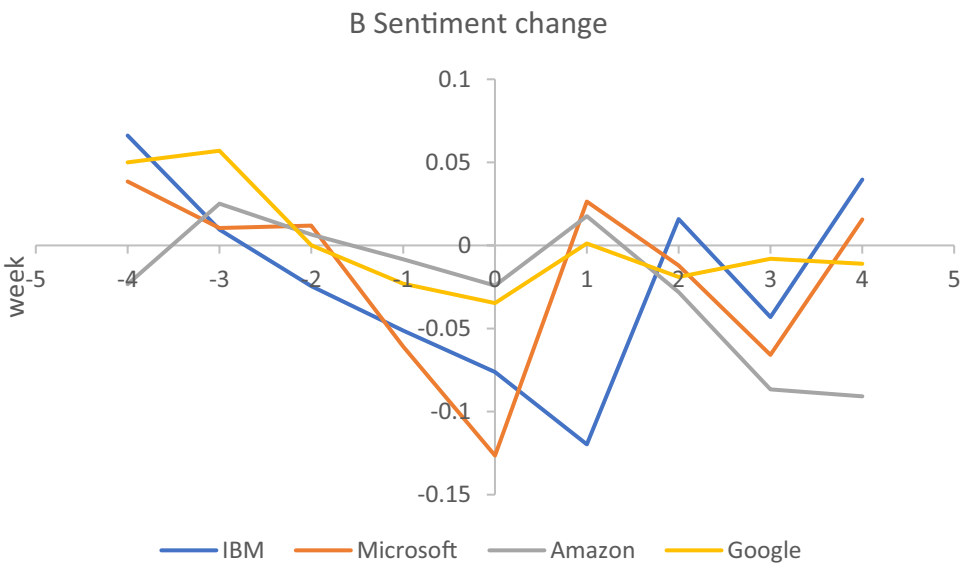
A third possible determinant for company responses is whether the controversial technology violates the company's stated public values. Value is a company's cultural cornerstone guiding all of its actions, including how it deals with a controversial technology. For instance, academics and employees cited Google's motto "Don't Be Evil" in the open letters against the company's partnership with Pentegan (Harnett 2018; ICRAC 2018). In response, Google proceeded to switching off the facial recognition project, and launched its own ethical principles for AI (Google 2018). This denotes a public set of corrective actions, aligning commercial activities with set principles.

Furthermore, as our analysis is limited to four big tech companies, further research on AI ethical failures should consider other types of organizations. For example, our model has relatively limited implications for small and medium-sized firms, as they attract less public attention to their AI programs and lack the resources to conduct some of the strategies suggested by the paper. Therefore, future research can explore strategies specific to small and medium-sized firms to address AI ethical failures. Similarly, the model may not apply to the public or non-profit sectors, which are driven by public interests and social welfare instead of financial returns. Future research can investigate solutions tailored to public and non-profit organizations in the face of public criticisms of their AI projects. Also, our model is unable to account for those that are not subject to public scrutiny and even intentionally use AI for bad ends, such as terrorist organizations and authoritarian regimes. Future research can look into other deterrence strategies for such intentional AI ethical failure.

We also suggest that future research can highlight the interplay between firm strategies and regulations on controversial AI technologies. In contrast to the ethical principles of AI that are converging towards a global consensus, regulations of AI vary by region—there now exist more than seven hundred policy initiatives across sixty countries. For instance, the Europe Union is proposing a ban on many high-risk applications of AI, such as facial recognition. China on the other hand shows much less caution in its governance of AI but is devoted to promoting the technology (OECD 2021). It seems most unlikely that there will be a global AI regulatory framework to rely on in the medium or even long term. Hence firms adopting AI technologies will need remain attentive to public criticism, and if accused of ethical failure, understand the nature of the criticism they are facing to develop the appropriate strategy on how to respond to safeguard their reputation.

Appendix 1 Abnormal return analysis of stock price and sentiment change





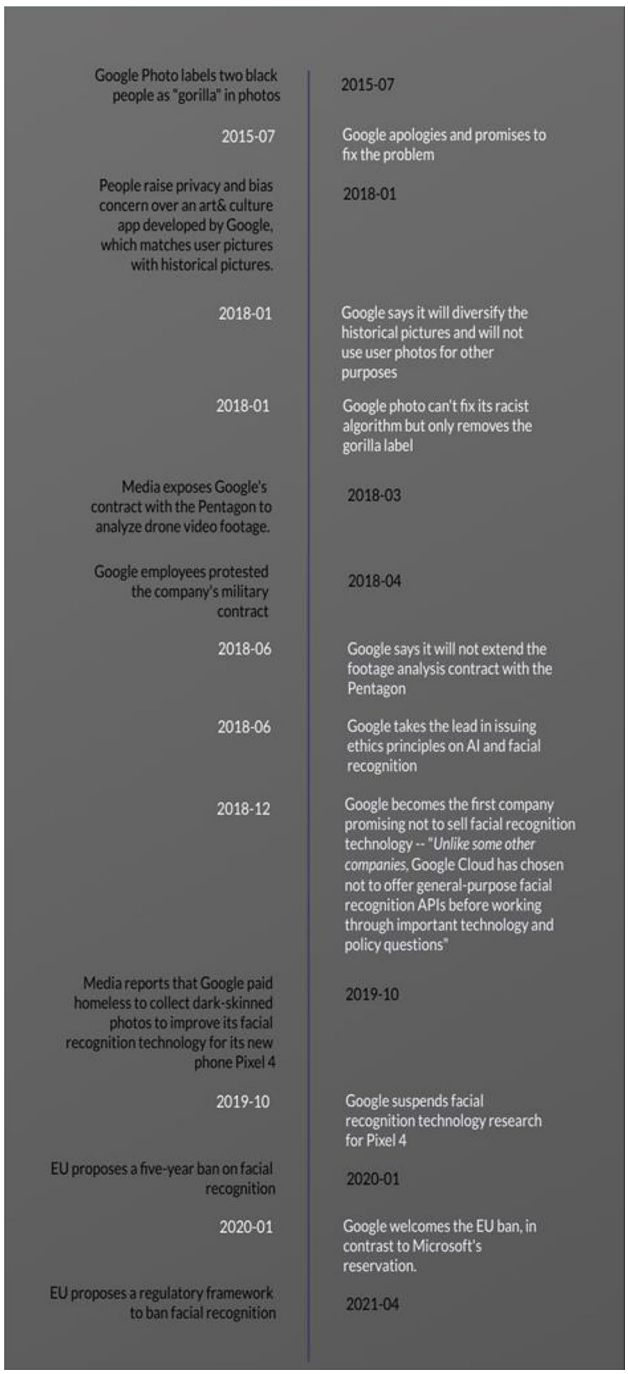
Appendix 2 Case timelines

Amazon Rekognition—Deflection

IBM Watson Facial Recognition—Improvement

Microsoft Azure Face API—Validation

Google Cloud Vision API—Pre-emption



Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Akaka MA, Parry G (2019) Value-in-context: an exploration of the context of value and the value of context. In Maglio, P. P., Kieliszewski, C. A., Spohrer, J. C., Lyons, K., Patrício, L., & Sawatani, Y. (Eds.). Handbook of service science, volume II. Springer, Cham, pp 457–477

Allyn B (2020) Amazon halts police use of its facial recognition technology. NPR. https://www.npr.org/2020/06/10/874418013/amazon-halts-police-use-of-its-facial-recognition-technology?t=1626010929990. Accessed 13 July 2021

Amazon (2016) Introducing Amazon Rekognition. Available at https://aws.amazon.com/about-aws/whats-new/2016/11/introducing-amazon-rekognition/. Accessed a11 July 2021

Amazon. (2017) Amazon Rekognition announces real-time face recognition, Text in Image recognition, and improved face detection. https://aws.amazon.com/about-aws/whats-new/2017/11/amazon-rekognition-announces-real-time-face-recognition-text-in-image-recognition-and-improved-face-detection/. Accessed 11 July 2021

Appadurai A, Alexander N (2019) Failure. Polity Press, Medford

Arvan M (2018) Mental time-travel, semantic flexibility, and AI ethics. AI & Soc. https://doi.org/10.1007/s00146-018-0848-2

AWS. (2021) Amazon Rekognition customers. https://aws.amazon.com/rekognition/customers/?nc=sn&loc=8. Accessed Dec 10 2021

BBC (2015) Google apologizes for Photos app's racist blunder Available at https://www.bbc.co.uk/news/technology-33347866. Accessed 2 Aug 2021

BBC (2019) Microsoft deletes massive facial recognition database. Available at https://www.bbc.co.uk/news/technology-48555149. Accessed 15 July 2021

BBC (2020) Facial recognition: EU considers ban of up to five years. https://www.bbc.co.uk/news/technology-51148501. Accessed 2 Sep 2020

Bijker WE, Pinch TJ (1987) The social construction of fact and artifacts. In Scharff, R. C., & Dusek, V. (Eds.). Philosophy of technology: the technological condition: an anthology, pp.107-139

Boeker W (1992) Power and managerial dismissal: Scapegoating at the top. Adm Sci Q 37:400–421

Braun HJ (1992) Introduction. Symposium on 'Failed Innovations.' Soc Stud Sci 22(2):213–230

Bruland K (1995) Patterns of resistance to new technologies in Scandinavia: an historical perspective. In: Bauer M (ed) Resistance to new technology. Cambridge University Press, Cambridge

Bundy J, Iqbal F, Pfarrer MD (2021) Reputations in flux: how a firm defends its multiple reputations in response to different violations. Strateg Manag J 42(6):1109–1138. https://doi.org/10.1002/smj.3276

Burt C (2018) IBM launches real-time algorithmic bias detection tools.BiometricUpdate.com. https://www.biometricupdate.com/201809/ibm-launches-real-time-algorithmic-bias-detection-tools. Accessed 2 Aug 2021

Calantone RJ, Benedetto CA, Divine R (1993) Organizational, technical andmarketing antecedents for successful new product development. R&D Manag 23(4):337–434

Cambrosio A, Limoges C (1991) Controversies as governing processes in technology assessment. Technol Anal Strateg Manag 3(4):377–396

Carmody T (2010) How facial recognition works in Xbox Kinect. https://www.wired.com/2010/11/how-facial-recognition-works-in-xbox-kinect/. Accessed 15 July 2021

Catro, A (2019) Google, Amazon, and Microsoft face new pressure over facial recognition contracts. https://www.theverge.com/2019/1/15/18183789/google-amazon-microsoft-pressure-facial-recognition-jedi-pentagon-defense-government. Accessed 1 Aug 2021

Chee JF, Chalmers J (2020) Alphabet CEO backs temporary ban on facial-recognition, Microsoft disagrees. Reuters. https://www.reuters.com/article/us-google-eu-idUSKBN1ZJ18O. Accessed 15 July 2021

Choung H, David P, Ross A (2022) Trust and ethics in AI. AI & Soc. https://doi.org/10.1007/s00146-022-01473-4

Coombs WT (1998) An analytic framework for crisis situations: Better responses from a better understanding of the situation. J Public Relat Res 10:177–191

Coombs WT (2007) Protecting organization reputations during a crisis: the development and application of situational crisis communication theory. Corp Reput Rev 10:163–176

Coombs WT (2010) Parameters for crisis communication. In: Coombs WT, Holladay SJ (eds) Handbook of crisis communication. Wiley-Blackwell, Malden, pp 17–54

Dafoe A (2018) AI governance: a research agenda. Governance of AI Program, Future of Humanity Institute, University of Oxford, Oxford, pp 1442–1443

Dastin J, Bera A (2020) Amazon pauses police use of its facial recognition tech for a year. Reuter. https://www.reuters.com/article/us-amazon-com-facial-recognition/amazon-pauses-police-use-of-its-facial-recognition-tech-for-a-year-idUSKBN23H3EO. Accessed 2 Aug 2021

Dickey MR (2020) Google employees demanded the company stop selling tech to police. Techcrunch. https://techcrunch.com/2020/06/22/google-employees-demand-company-stop-selling-tech-to-police/. Accessed 2 Aug 2021

Doran D, Schulz S, Besold TR (2017) What does explainable AI really mean? A new conceptualization of perspectives. arXiv preprint arXiv:1710.00794

Dwoskin E (2018) Amazon is selling facial recognition to law enforcement—for a fistful of dollars. The Washington Post. https://www.washingtonpost.com/news/the-switch/wp/2018/05/22/amazon-is-selling-facial-recognition-to-law-enforcement-for-a-fistful-of-dollars/. Accessed 11 July 2021

Easterling K (2016) Histories of things that don't happen and shouldn't always work. Soc Res 83(3):625–644

Eisenhardt KM (1989) Making fast strategic decisions in high-velocity environments. Acad Manag J 32(3):543–576
Elsbach K (2003) Organizational perception management. Res Org Behav 25:297–332
European Union (2019) Ethics Guidelines on Trustworthy AI. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html. Accessed 11 Nov 2021
Faulkner R (2011) Corporate wrongdoing and the art of the accusation. London: Anthem Press.
Feenberg A (1999) Questioning technology. Routlege, London
Feldman ER, Amit R, Villalonga B (2016) Corporate divestitures and family control. Strateg Manag J 37(3):429–446. https://doi.org/10.1002/smj.2329
Floridi L, Cowls J (2019) A unified framework of five principles for AI in society. Harvard Data Sci Rev 1(1):5–17. https://doi.org/10.1162/99608f92.8cd550d1
Floridi L, Holweg M, Taddeo M, Amaya Silva J, Mökander J, Wen Y (2022) capAI-A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act. https://doi.org/10.2139/ssrn.4064091
Gebru T (2020) Race and gender. In Dubber, M. D., Pasquale, F., & Das, S. (Eds.). The Oxford handbook of ethics of AI, pp 251–269
Gibbs S (2018) Google's AI is being used by US military drone programme. Guardian. https://www.theguardian.com/technology/2018/mar/07/google-ai-us-department-of-defense-military-drone-project-maven-tensorflow. Accessed 2 Aug 2021
Gooday G (1998) Re-writing the 'book of blots': Critical reflections on histories of technological 'failure.' Hist Technol Int J 14(4):265–291
Google (2018) Artificial Intelligence at Google: Our Principles. Available at https://ai.google/principles/. Accessed 11 Nov 2021
Google (2021) Detect faces. https://cloud.google.com/vision/docs/detecting-faces. Accessed 15 July 2021
Hamann K, Smith R (2019) Facial recognition technology. CRIM. JUST, 9. https://www.americanbar.org/groups/criminal_justice/publications/criminal-justice-magazine/2019/spring/facial-recognition-technology/. Accessed 3 Jan 2022
Hamilton A (2019) Google suspended facial recognition research for the Pixel 4 smartphone after reportedly targeting homeless black people. Business Insider. https://www.businessinsider.com/google-suspends-facial-recognition-research-after-daily-news-report-2019-10?r=US&IR=T. Accessed 2 Aug 2021
Hardesty L (2018) Study finds gender and skin-type bias in commercial artificial-intelligence systems. MIT News. https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212. Accessed 20 Aug 2021
Harnett S (2018) Google employees quit in protest over military artificial intelligence program. https://www.kqed.org/news/11668872/google-employees-quit-in-protest-over-military-artificial-intelligence-program. Accessed 10 Dec 2021
Harries D (2015) IBM acquires deep learning startup AlchemyAPI. Available at https://gigaom.com/2015/03/04/ibm-acquires-deep-learning-startup-alchemyapi/. Accessed 10 Dec 2021
Heilweil R (2020) Big tech companies back away from selling facial recognition to police. That's progress. Recode. https://www.vox.com/recode/2020/6/10/21287194/amazon-microsoft-ibm-facial-recognition-moratorium-police. Accessed 18 June 2020
Hekkila M (2021) European Parliament calls for ban on facial recognition. POLITICO. https://www.politico.eu/article/european-parliament-ban-facial-recognition-brussels/. Accessed 13 Dec 2021
Hesseldahl A (2014) IBM Doubles Down on Watson. Recode. https://www.vox.com/2014/1/9/11622154/ibm-doubles-down-on-watson. Accessed 20 Aug 2021
Holweg M, Younger R, Wen Y (2022) The Reputational Risks of AI. California Management Review Insights. https://cmr.berkeley.edu/2022/01/the-reputational-risks-of-ai/ Accessed 12 Feb 2022
IBM (2018) IBM to release world's largest annotation dataset for studying bias in facial analysis. https://www.ibm.com/blogs/research/2018/06/ai-facial-analytics/. Accessed 20 Aug 2021
IBM (2019a) Release notes. https://cloud.ibm.com/docs/visual-recognition?topic=visual-recognition-release-notes. Accessed 20 Aug 2021
IBM (2019b) IBM CEO's Letter to Congress on Racial Justice Reform. https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/. Accessed 20 Aug 2021
ICRAC (International Committee for Robot Arms Control) (2018) Researchers in Support of Google Employees: Google should withdraw from Project Maven and commit to not weaponizing its technology. https://www.icrac.net/open-letter-in-support-of-google-employees-and-tech-workers/. Accessed 10 Dec 2021
Kent W (2018) AI for social good in Asia Pacific. Google blog. https://www.blog.google/around-the-globe/google-asia/ai-social-good-asia-pacific/
Khan AA, Badshah S, Liang P, Waseem M, Khan B, Ahmad A, Akbar MA (2021 June) Ethics of AI: A systematic literature review of principles and challenges. In: Proceedings of the International Conference on Evaluation and Assessment in Software Engineering 2021, pp 383–392
Kline SJ (1985) What is technology?. Bull Sci Technol Soc 5(3):215–218
Langley A (1999) Strategies for theorizing from process data. Acad Manag Rev 24(4):691–710
Lauer D (2021) You cannot have AI ethics without ethics. AI and Ethics 1(1):21–25. https://doi.org/10.1007/s43681-020-00013-4
LeCunBengio YY (1995) Convolutional networks for images, speech, and time series. Handb Brain Theory Neural Netw 3361(10):1995
LeCun Y, Boser B, Denker J, Henderson D, Howard R, Hubbard W, Jackel L (1989) Handwritten digit recognition with a back-propagation network. Adv Neural Inf Process Syst 2. 396-404
Lee D (2018) Amazon defends providing police facial recognition tech. BBC. https://www.bbc.co.uk/news/technology-44220037. Accessed 13 July 2021
Liao S (2019) IBM didn't inform people when it used their Flickr photos for facial recognition training. Verge. https://www.theverge.com/2019/3/12/18262646/ibm-didnt-inform-people-when-it-used-their-flickr-photos-for-facial-recognition-training. Accessed 20 Aug 2021
Lin G, Li H, Ma H, Yao D, Lu R (2020) Human-in-the-loop consensus control for nonlinear multi-agent systems with actuator faults. IEEE/CAA J Autom Sin 9(1):111–122. https://doi.org/10.1109/JAS.2020.1003596
Markoff J (2011) Computer wins on 'Jeopardy!': Trivial, it's not. New York Times. https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html. Accessed 20 Aug 2021
Martin K, Shilton K, Smith J (2019) Business and the ethical implications of technology: introduction to the symposium. J Bus Ethics 160(2):307–317. https://doi.org/10.1007/s10551-019-04213-9
Mazurek G, Małagocka K (2019) Perception of privacy and data protection in the context of the development of artificial intelligence. J Manag Anal 6(4):344–364. https://doi.org/10.1080/23270012.2019.1671243
Menn J (2019) Microsoft turned down facial-recognition sales on human rights concerns. Reuters. https://www.reuters.com/article/us-microsoft-ai-idUSKCN1RS2FV?utm_campaign=trueAnthem:+Trending+Content&utm_content=5cb671670f1cfa0

001e6919f&utm_medium=trueAnthem&utm_source=twitter. Accessed 15 July 2021
Merriam-Webster (2022) Failure—the definition of failure. https://www.merriam-webster.com/dictionary/failure?utm_campaign=sd&utm_medium=serp&utm_source=jsonld. Accessed on 22 December 2022
Metcalf J, Moss E (2019) Owning ethics: Corporate logics, silicon valley, and the institutionalization of ethics. Soc Res Int Q 86(2):449–476
Microsoft (2017) C#: Face Detection and Recognition with Azure Face API. https://social.technet.microsoft.com/wiki/contents/articles/37893.c-face-detection-and-recognition-with-azure-face-api.aspx. Accessed 1 Aug 2021
Mishina Y, Block E, Mannor M (2012) the path dependence of organizational reputation: how social judgment influences assessments of capability and character'. Strateg Manag J 33:459–477. https://doi.org/10.1002/smj.958
Mittelstadt B (2019) Principles alone cannot guarantee ethical AI. Nat Mach Intell 1(11):501–507. https://doi.org/10.1038/s42256-019-0114-4
Morley J, Kinsey L, Elhalal A et al (2022) Operationalising AI ethics: barriers, enablers and next steps. AI & Soc. https://doi.org/10.1007/s00146-021-01308-8
NBC (2019) Facial recognition's 'dirty little secret': Millions of online photos scraped without consent. NBC. https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921. Accessed 20 Aug 2021
Newcomb A (2015) Microsoft wants to guess your age using face recognition app. https://abcnews.go.com/Technology/microsoft-guess-age-face-recognition-app/story?id=30707123. Accessed 15 July 2021
OECD (2021) National AI policies and strategies. https://oecd.ai/en/dashboards. Accessed 25 Oct 2021
Ouchchy L, Coin A, Dubljević V (2020) AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. AI Soc. https://doi.org/10.1007/s00146-020-00965-5
Pang A (2012) Towards a crisis pre-emptive image management model. Corp Commun Int J. 17(3): 358-378.
Paul K (2018) Why Google's selfie app works better for white people. MarketWatch. https://www.marketwatch.com/story/why-googles-selfie-app-works-better-for-white-people-2018-01-17?mod=article_inline. Accessed 2 Aug 2021
Paul K (2020) Amazon says 'Black Lives Matter'. But the company has deep ties to policing. Guardian. https://www.theguardian.com/technology/2020/jun/09/amazon-black-lives-matter-police-ring-jeff-bezos. Accessed 13 July 2021
Pfarrer MD, DeCelles KA, Smith KG, Taylor MS (2008) After the fall: reintegrating the corrupt organization. Acad Manag Rev 33:730–749. https://doi.org/10.5465/amr.2008.32465757
PHYS (2019a) Amazon shareholders support selling face recognition tech to police. https://phys.org/news/2019a-05-amazon-shareholders-recognition-tech-police.html. Accessed 11 July 2021
PHYS (2019b) Amazon speaks out in favor of regulating facial recognition. https://phys.org/news/2019b-06-amazon-favor-facial-recognition-technology.html?utm_source=TrendMD&utm_medium=cpc&utm_campaign=Phys.org_TrendMD_1. Accessed 13 July 2021
Power B (2014) How watson changed IBM. Harvard Business Review. https://hbr.org/2014/08/how-watson-changed-ibm. Accessed 10 Dec 2021
Prashantham S, Bhagavatula S, Kumar K (2020) Handle with care: Entrepreneurial reputation-borrowing in an emerging economy. J Bus Ventur Insights 13:e00156. https://doi.org/10.1016/j.jbvi.2020.e00156
Punke M (2019) Some thoughts on facial recognition legislation. https://aws.amazon.com/blogs/machine-learning/some-thoughts-on-facial-recognition-legislation/. Accessed 11 July 2021
Pye D (1978). The nature and aesthetics of design. New York.
Raji ID, Kumar IE, Horowitz A, Selbst A (2022) The fallacy of AI functionality. In: 2022 ACM Conference on fairness, accountability, and transparency, pp 959–972. https://doi.org/10.1145/3531146.3533158
Roach J (2018) Microsoft improves facial recognition technology to perform well across all skin tones, genders. Microsoft blog. https://blogs.microsoft.com/ai/gender-skin-tone-facial-recognition-improvement/. Accessed 15 July 2021
Rogers EM (1962) Diffusion of innovations. Simon and Schuster
Rochel J, Evéquoz F (2021). Getting into the engine room: a blueprint to investigate the shadowy steps of AI ethics. AI & SOCIETY, 36:609–622.
Schilling MA (1998) Technological lockout: An integrative model of the economic and strategic factors driving technology success and failure. Acad Manag Rev 23(2):267–284. https://doi.org/10.2307/259374
Shane S, Wakabayashi D (2018) The business of war': Google employees protest work for the Pentagon. The New York Times. https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html. Accessed 02 August 2021
Singer N (2018) Amazon is pushing facial technology that a study says could be biased. The New York Times 24
Slowlkowski S, Jarratt D (2007) The impact of culture on the adoption of high technology products. Mark Intell Plan 15(2):97–105
Smith B (2018a) Facial recognition technology: The need for public regulation and corporate responsibility. Microsoft blog. https://blogs.microsoft.com/on-the-issues/2018a/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/ Accessed 15 July 2021
Smith B (2018b) Facial recognition: It's time for actions. Microsoft blog. https://blogs.microsoft.com/on-the-issues/2018b/12/06/facial-recognition-its-time-for-action/. Accessed 15 July 2021
Smith J (2019) IBM Research Releases 'Diversity in Faces' Dataset to Advance Study of Fairness in Facial Recognition Systems. https://www.ibm.com/blogs/research/2019/01/diversity-in-faces/. Accessed 20 Aug 2021
Smith B (2020) Finally, progress on regulating facial recognition. Microsoft blog. https://blogs.microsoft.com/on-the-issues/2020/03/31/washington-facial-recognition-legislation/. Accessed 15 July 2021
Smith M, Miller S (2022) The ethical application of biometric facial recognition technology. AI & Soc 37:167–175. https://doi.org/10.1007/s00146-021-01199-9
Snow J (2018) Amazon's face recognition falsely matched 28 members of congress with mugshots. https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28. Accessed 11 July 2021
Solo O (2019) Microsoft hires Eric Holder to audit AnyVision over use of facial recognition on Palestinians. NBC. https://www.nbcnews.com/tech/security/microsoft-hires-eric-holder-audit-anyvision-over-use-facial-recognition-n1083911. Accessed 15 July 2021
Souder W, Sherman J (1993) Managing new technology development. McGraw-Hill, New York

Stahl BC, Antoniou J, Ryan M et al (2022) Organisational responses to the ethical issues of artificial intelligence. AI & Soc 37:23–37. https://doi.org/10.1007/s00146-021-01148-6

Statt N (2020) Microsoft won't sell facial recognition to police until Congress passes new privacy law. Verge. https://www.theverge.com/21288053/microsoft-facial-recognition-police-law-enforcement-pledge-regulation. Accessed 15 July 2021

Suddaby R (2006) From the editors: what grounded theory is not. Acad Manag J 49(4):633–642. https://doi.org/10.5465/amj.2006.22083020

Tonkinwise C (2016) Failing to sense the future: From design to the proactionary test drive. Soc Res 83(3):597–624

Vakkuri V, Kemell KK, Abrahamsson P (2019) AI ethics in industry: a research framework. arXiv preprint arXiv:1910.12695.

Vanian J (2018) Defense department is using Google's AI Tech to help with drone surveillance. Fortune. https://fortune.com/2018/03/06/google-department-defense-drone-ai/. Accessed 2 Aug 2021

Venturini T (2010) Diving in magma: how to explore controversies with actor-network theory. Public Underst Sci 19(3):258–273

Vincent J (2018) Amazon employees protest sale of facial recognition software to police. The Verge. https://www.theverge.com/2018/6/22/17492106/amazon-ice-facial-recognition-internal-letter-protest. Accessed 11 July 2021

Wakabayashi D, Shane S (2018) The business of war': Google employees protest work for the Pentagon. The New York Times. https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html. Accessed 2 Aug 2021

Whittlestone J, Nyrup R, Alexandrova A, Cave S (2019) The role and limits of principles in AI ethics: towards a focus on tensions. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp 195–200

Wilcox D, Cameron GT (2009) Public relations: strategies and tactics, 9th edn. Pearson Allyn & Bacon, Boston, MA

Wood M (2019) Thoughts on recent research paper and associated article on Amazon Rekognition. https://aws.amazon.com/blogs/machine-learning/thoughts-on-recent-research-paper-and-associated-article-on-amazon-rekognition/. Accessed 11 July 2021

Zhang Y, Wiersema MF (2009). Stock market reaction to CEO certification: The signaling role of CEO background. Strategic Management Journal, 30(7):693–710.

Zwickle A, Farber HB, Hamm JA (2019) Comparing public concern and support for drone regulation to the current legal framework. Behav Sci Law 37(1):109–124. https://doi.org/10.1002/bsl.2357