Testing the Black Box: Institutional Investors, Risk Disclosure, and Ethical AI

Trooper Sanders¹

Received: 25 February 2020 / Accepted: 11 June 2020 / Published online: 24 July 2020
© The Author(s) 2020

Abstract

The integration of artificial intelligence (AI) throughout the economy makes the ethical risks it poses a mainstream concern beyond technology circles. Building on their growing role bringing greater transparency to climate risk, institutional investors can play a constructive role in advancing the responsible evolution of AI by demanding more rigorous analysis and disclosure of ethical risks.

Keywords Artificial intelligence · Black box · Investors · AI ethics · Pension funds

In its most recent public filings to the United States Securities and Exchange Commission (SEC), Microsoft Corporation (2020) alerted investors to risks from its growing artificial intelligence business. The company's September 2019 10-Q filing warned: "AI algorithms may be flawed. Datasets may be insufficient or contain biased information. Inappropriate or controversial data practices by Microsoft or others could impair the acceptance of AI solutions.... Some AI scenarios present ethical issues. If we enable or offer AI solutions that are controversial because of their impact on human rights, privacy, employment, or other social issues, we may experience brand or reputational harm" (2019). These disclosures mark a meaningful step forward in bringing AI ethics from the academy and advocacy and into the mainstream of the marketplace. And while flagging risks to investors is not the same as the market rewarding companies for the ethical quality of their development, application, and commercialization of AI, it can help make emerging technologies and business practices powered by AI more accountable to investors and the public.

Since 2005, the SEC has required companies issuing shares to the public to disclose risks (Election Code of Federal Regulation 2020). Firms must "disclose material factors that may adversely affect the issuers business, operations, industry or financial position, or its future firm performance" (Filzen, McBrayer, and Shannon, 2016). For example, a pharmaceutical

Trooper Sanders is a former Rockefeller Foundation Fellow.

✉ Trooper Sanders
tsanders@wisewhisper.com

¹ Rockefeller Foundation, New York USA

company's filing might discuss growing competition from generic drug manufacturers while a financial services firm might discuss the impact of regulatory changes and associated costs on the business. Access to good information is an essential part of efficient markets. In economics, the ideal state is when consumers and producers have perfect knowledge about price, quality, and other factors affecting decision-making. While individual investors have access to considerable information, it is institutional investors, the investment funds, insurance companies, and pension funds with more than \$100 trillion under management globally, who have the means to track, analyze, and react to the vast quantity of data available today (Segal 2018). A study by finance professors Field and Lowry (2005) found that institutional investors make better use of publicly available information than individual investors (2006).

In theory, greater transparency about risks should improve investors' situational awareness and their ability to make sound decisions but in practice disclosures often fall short of the mark. An analysis by the Investor Responsibility Research Center Institute said that risk factor disclosures by large companies "do not provide clear, concise and insightful information...are not tailored to the specific company...[and] tend to represent a listing of generic risks, with little to help investors distinguish between the relative importance of each risk to the company" (2020). Indeed, one of Microsoft's leading competitors in AI summed up their risk factors in their quarterly filing crisply: "Our operations and financial results are subject to various risks and uncertainties...which could adversely affect our business, financial condition, results of operations, cash flows, and the trading price of our common and capital stock" (Alphabet Inc, 2019).

However imperfect, institutional investors can use their influence to bring greater transparency to AI in two ways: pushing regulators to demand more disclosure by public companies and assessing the ethical AI fitness of companies in their portfolio who have materially significant stakes as AI developers or consumers. The evolving role of institutional investors in climate change is instructive. First, climate risk is a core business concern for funds. "The prices of the assets we buy as an investor, and the degree to which these prices reflect climate risk, affect the fund's financial risk," noted Norway's Government Pension Fund, a climate risk hawk among large funds (Olsen and Grande 2019). In addition, strong corporate performance on climate change is often an indicator of shareholder-friendly efficiency and sound management and governance.

In 2007, American and European investors managing \$1.5 trillion in assets joined a coalition calling on the SEC to require companies to assess and publicly disclose their financial risk related to climate change. "Climate change can affect corporate performance in ways ranging from physical damage to facilities and increased costs of regulatory compliance, to opportunities in global markets for climate-friendly products or services that emit little or no global warming pollution," the coalition argued. "Those risks fall squarely into the category of material information that companies must disclose under existing law to give shareholders a full and fair picture of corporate performance and operations" (Environmental Defense Fund 2007).

A few years later, several of the world's largest funds began to formally factor environmental, social, and governance (ESG) matters in some investment decisions. While ESG investing has its limitations for both institutional investors (OECD 2017) and for tackling the relevant concerns (Rennison 2019), it has become an important vehicle to put capital behind business practices aligned with the public interest (Eccles 2019). In addition, loose principles in the early years of ESG have evolved into more

robust metrics and standards (Edgecliffe-Johnson, Nauman, and Tett 2020). For example, powerful investors, including billionaire Michael Bloomberg, and others recently kicked off an effort examining the physical, liability, and transition risks of climate change as part of establishing voluntary climate-related financial risk disclosure standards (Task Force on Climate-related Financial Disclosures 2020).

Institutional investors did not wait for climate law and regulation to settle and scale before seizing opportunities and asserting influence. A combination of hard law and regulation, non-legislative soft law, and climate ethics shaped by evolving consumer sentiment, political consensus, and social norms provided sufficient guidance and grounding. Similarly, the emerging consensus in AI ethics around transparency, justice and fairness, non-maleficence, responsibility, and privacy can provide a guidon to investors addressing AI concerns (Jobin, Inca, and Vayena 2019). Indeed, as scholars such as Gary Marchant have noted, the slow pace of legal and regulatory change in technology matters has created a void best filled by soft law tools such as professional guidelines, private standards, codes of conduct, and best practice (2019). Indeed, as key players in the economy, institutional investors could give the ethical AI field some essential oomph. As leading AI ethicist, Virginia Dignum notes: “Engineers are those that ultimately will implement AI to meet ethical principles and human values, but it is policy makers, regulators and society in general that can set and enforce the purpose” (2019).

Consider emotion recognition services that use algorithms to analyze facial features and make inferences about mood and behavior (Jee 2019). This growing segment of the AI market, worth more than \$20 billion and put to use in areas ranging from workplace hiring to law enforcement, poses several ethical challenges including:

- weak scientific foundations, with one recent review of more than 1000 scientific papers finding very little evidence that facial expressions alone can predict how someone is feeling (Chen 2019);
- concerns that racial and gender bias will exacerbate existing disparities (Rhue 2019); and
- it replaces human judgment and being used without appropriate human oversight (Qumodo Ltd. 2019).

For firms offering such services, material concerns that could fall under risk disclosure requirements include:

- biased data and shaky science undermining the quality of and confidence in products leading to declining sales and market share;
- controversial applications affecting public interest concerns such as employment discrimination and abusive policing leading to greater regulatory and public relations costs; and
- the confluence of business headwinds, public resistance, and technical vulnerabilities eroding market confidence and triggering a long winter or collapse of the sector.

Customers of these services face their own risks worthy of disclosure. They include:

- harmful AI infecting the quality and reputation of core products and services leading to increased litigation risk, declining sales and market share, and unexpected mitigation costs;
- damage to the corporate brand, including its brand valuation juices share price, and relationship with customers, stakeholders, and the public; and
- productivity loss from toxic AI polluting critical operations such as talent management or a negative experience in one application of AI slowing or stopping other AI efforts that offer material benefits.

Just as changes in climate change thinking and analytics moved influential market players to act, the evolving state of the art in AI ethics can help institutional investors probe beyond disclosures in public filings (Moss 2019). First, strong ethical AI performance can be an indicator for a strong and well-managed enterprise generally, and weak performance a warning sign for more fundamental challenges that could hurt shareholders. Furthermore, the well-established body of knowledge about algorithmic bias gives analysts a strong foundation to test the material ethical risks of companies buying and selling machine learning products and services in areas as diverse as human resources, health care, and consumer banking (Raghavan et al., 2019). Companies forthcoming about the limitations of training data, bias in services and products, and steps they are taking to mitigate harm are more likely to pose fewer risks while those denying data vulnerabilities or ethical soft spots should be viewed skeptically. Investors will be able to develop deeper layers of inquiry on risk, financial performance, and other priorities as the fairness, accountability, and transparency field expands beyond technical matters such as explainability and interpretability to include rigorous treatment of the real world use and the social and organizational impact of AI (Sendak et al., 2020). In addition, greater scrutiny given to the limits of AI in sensitive sectors such as health care can help investors avoid exposure to overblown claims that harm people, damage companies, and destroy shareholder value (Szabo 2019).

In sum, while institutional investors' involvement in AI ethics is no balm to the havoc rogue AI can cause, they can be constructive allies in the push to align the power of technology and the public interest. Whether putting money behind ethical performance yields returns that sustain their interest depends on pressure from and decisions by developers, regulators, and consumers who drive AI's course.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Alphabet Inc. (2019). Form 10-Q for the quarterly period ended September 30, 2019. https://abc.xyz/investor/static/pdf/20191028_alphabet_10Q.pdf?cache=376def7.

- Chen, A. (2019). Computers can't tell if you're happy when you smile. *MIT Technology Review*. <https://www.technologyreview.com/s/614015/emotion-recognition-technology-artificial-intelligence-inaccurate-psychology/>.
- Dignum, V. (2019). AI ethical principles are for us. *Medium*. . <https://medium.com/@viriniadignum/ai-ethical-principles-are-for-us-def54c64d9a8>.
- Eccles, R. (2019). Why it's time to finally worry about ESG. *Harvard Business Review*. <https://hbr.org/podcast/2019/05/why-its-time-to-finally-worry-about-esg>.
- Edgecliffe-Johnson, A., Nauman, B., and Tett, G. (2020). Davos 2020: companies sign up to environmental disclosure scheme. *Financial Times*. <https://hbr.org/podcast/2019/05/why-its-time-to-finally-worry-about-esg>.
- Electronic Code of Federal Regulations (2020), Title 17: commodity and securities exchanges §229.105 (Item 105) risk factors. Current as of January 16, 2020. https://www.ecfr.gov/cgi-bin/text-idx?amp;node=17:3.0.1.1.11&rgn=div5#_top.
- Environmental Defense Fund. (2007). *Major investors*. Environmental Groups Petition SEC to Require Full Corporate Climate Risk Disclosure: State Officials. <https://www.edf.org/news/major-investors-state-officials-environmental-groups-petition-sec-require-full-corporate-climat>.
- Field, L., and Lowry, M. (2005). Institutional versus individual investment in Ipos: the importance of firm fundamentals. AFA 2006 Boston Meeting Paper. https://papers.ssm.com/sol3/papers.cfm?abstract_id=613563.
- Filzen, J., McBrayer, G., and Shannon, K. (2016). Risk factor disclosures: do managers and markets speak the same language?. Accessed January 8, 2020, <https://www.sec.gov/comments/s7-06-16/s70616-369.pdf>.
- Investor Responsibility Research Center Institute (2020). "The corporate risk factor disclosure landscape," 21 and 3. Accessed January 8, 2020, <https://www.weinberg.udel.edu/IIRCResearchDocuments/2016/01/FINAL-EY-Risk-Disclosure-Study.pdf>.
- Jee, C. (2019). Emotion recognition technology should be banned, says an AI research institute. *MIT Technology Review*, <https://www.technologyreview.com/f/614932/emotion-recognition-technology-should-be-banned-says-ai-research-institute/>.
- Jobin, A., Jenca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399 <https://www.nature.com/articles/s42256-019-0088-2>.
- Marchant, G. (2019). "'Soft Law' governance of artificial intelligence," *AI Pulse*. . <https://aipulse.org/soft-law-governance-of-artificial-intelligence/>.
- Microsoft Corporation (2020). Form 10-Q for the quarter ended September 30, 2019. . https://c.s-microsoft.com/en-us/CMSFiles/MSFT_FY20Q1_10Q.docx?version=a8248fdc-67a9-45da-1db8-818e9e8abde9.
- Moss, E. (2019). Unpacking "Ethical AI". *Points Data & Society*. <https://points.datasociety.net/unpacking-ethical-ai-b770b964c236>.
- OECD (2017). Investment governance and the integration of environmental, social, and governance factors. <https://www.oecd.org/cgfi/Investment-Governance-Integration-ESG-Factors.pdf>.
- Olsen, O., and Grande, Trond G. (2019). Government pension fund global account of work on climate risk. **Letter sent to the Ministry of Finance**. . <https://www.nbim.no/en/publications/submissions-to-ministry/2019/government-pension-fund-global%2D%2Daccount-of-work-on-climate-risk/>.
- Qumodo (2019). Automatic facial recognition: why do we need a human in the loop? <https://medium.com/@1530019197930/automatic-facial-recognition-why-do-we-need-a-human-in-the-loop-de8366d10680>.
- Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. (2019) Mitigating bias in algorithmic hiring: evaluating claims and practices. Accessed January 12, 2020. arXiv:1906.09208v3, accessed January 10, 2020, <https://arxiv.org/pdf/1906.09208.pdf>.
- Rennison, J. (2019). ESG investing is a term that is too often misused. *Financial Times*. <https://www.ft.com/content/ac10773a-a975-1fe9-b6ee-3cdf3174eb89>.
- Rhue, L. (2019). Understanding the hidden bias in emotion-reading AIs. *MIT Technology Review*. . <https://www.technologyreview.com/s/614015/emotion-recognition-technology-artificial-intelligence-inaccurate-psychology/>.
- Segal, J. (2018). The asset management industry is getting more concentrated. *Institutional Investor*. . <https://www.institutionalinvestor.com/article/b1bk8n82qcc0kt/The-Asset-Management-Industry-Is-Getting-More-Concentrated>.
- Sendak, M., et al. (2020). The human body is a black box: supporting clinical decision-making with machine learning. arXiv:1911.08089. . <https://arxiv.org/abs/1911.08089>.
- Szabo, L. (2019). A reality check on artificial intelligence: are health care claims overblow?. *Kaiser Health News*. . <https://khn.org/news/a-reality-check-on-artificial-intelligence-are-health-care-claims-overblown/>.
- Task Force on Climate-related Financial Disclosures. Task force overview. Accessed January 6, 2020. <https://www.fsb-tcfd.org/about/>.