TECH ACCOUNTABILITY
FINANCE-FOCUSED STRATEGIES

AI Investment Standards

Building consensus from emerging efforts

Key Takeaways

- Artificial intelligence systems are already being deployed both within and beyond the tech sector, exposing communities to potential bias and rights violations, while also introducing new legal, financial, and reputational risks for companies. The nascent development of ethical standards for building and implementing AI tools, however, is presently too piecemeal and or too generalized to keep pace.

- Many stakeholders and third-party actors are developing sets of ethical AI principles and certification regimes. These efforts hold promise, particularly as the use of complex AI tools spreads into industry sectors where companies may not have the technical expertise to assess the impact of those tools in-house. As demand for AI certification and impact assessments increases, however, so does the risk of bad actors entering the space to "rubber stamp" poor governance practices for profit.

- Investors have an opportunity to support the evolution of specific, operationalizable ethical AI standards by demanding transparency regarding these "black box" technologies and the processes companies use to develop, assess, and deploy them. In coordination with civil society organizations and academia, investors are also well positioned to inject pragmatic business concerns into the discourse surrounding AI.

TECH ACCOUNTABILITY
FINANCE-FOCUSED STRATEGIES

NetGain
Partnership

OpenMIC

Contents

In September 2021, the NetGain Partnership initiated a research process designed to explore finance-focused strategies that would hold leading internet platforms accountable and "create a healthier digital public sphere." The partnership said it was interested in supporting shareholder engagement while also developing stronger ESG(+D) screens on tech issues. The research would aim to be "broadly useful to philanthropy and the broader public interest community."

In April 2022, the partnership commissioned Open MIC and Whistle Stop Capital to produce a series of reports that addressed those issues. Since then, the research team has conducted interviews with more than 40 practitioners, analysts and observers of shareholder engagement and finance-focused strategies in the global technology sector. The team has also done substantial research exploring current tactics and strategies employed in the finance-sector globally to check the power and harmful behaviors of Big Tech companies.

Click here or use the QR code at the right to view the four reports prepared by Open MIC

Overview

Artificial intelligence ("AI") is an emerging technology that, despite its relative novelty, has already become a powerful force in virtually every industry – including essential sectors with vast implications for human and civil rights.

While the precise definition of artificial intelligence is nebulous and evolving, it can be understood as using computer processes to perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement.[1] Consequently, AI tools can include a broad spectrum of processes from simple algorithms to sophisticated machine learning systems.

In practice, AI tools are typically used to supplement or replace human decision making. Companies and public service providers have adopted AI systems at such a rapid rate that these tools can now be found in nearly every industry and public sector, including policing, military operations, corporate hiring, financial services, housing, education, healthcare, fraud detection, and environmental safety.

A Sample of Reported AI Harms

- In 2013, Chicago police used an algorithm to direct surveillance of those deemed most likely to be involved in gun violence, even though nearly half of those listed by the system had never been arrested for any gun-related crime.[2]

[1] "Artificial Intelligence," Computer Security Resource Center, National Institute of Standards and Technology, https://csrc.nist.gov/Topics/technologies/artificial-intelligence.
[2] Mike Dumke and Frank Main, *Chicago Sun-Times*, "A look inside the watch list Chicago police fought to keep secret" (May 18, 2017), https://chicago.suntimes.com/2017/5/18/18386116/a-look-inside-the-watch-list-chicago-police-fought-to-keep-secret.

- In Michigan, an artificial intelligence system with an error rate of 93 percent falsely accused more than 20,000 people of unemployment fraud over a two year period.[3]
- Job applicant filtering systems used by 99 percent of Fortune 500 companies routinely screen out qualified candidates based on technicalities, perpetuating an ongoing staffing crisis.[4]
- Remote exam proctoring software routinely uses racially-discriminatory facial recognition technology that fails to verify Black students' identities,[5] and has falsely accused students of cheating due to their neurodivergent fidgeting behaviors or for wearing a hijab.[6]
- On June 28, 2022, the Organ Procurement & Transplantation Network, after years of controversy over racial disparities in donor selection, finally removed a longstanding explicit race factor from its algorithm that systematically disadvantaged Black candidates seeking kidney transplants.[7]

---

[3] David Eggert, *AP News*, "State apologizes for fraud fiasco, wants to reduce penalties" (Jan. 28, 2017), https://apnews.com/article/c0e2346e85854a5b827ca42653c1fb40; Ed White, *Detroit Free Press*, "Unemployed Michiganders wrongly accused of fraud can seek cash from state" (July 27, 2022), https://www.freep.com/story/news/local/michigan/2022/07/26/unemployed-wrongly-accused-fraud-can-seek-cash-state/10157193002/.

[4] Kathryn Dill, *Wall Street Journal*, "Companies Need More Workers. Why Do They Reject Millions of Résumés?" (Sept. 4, 2021), https://www.wsj.com/articles/companies-need-more-workers-why-do-they-reject-millions-of-resumes-11630728008?mod=article_inline.

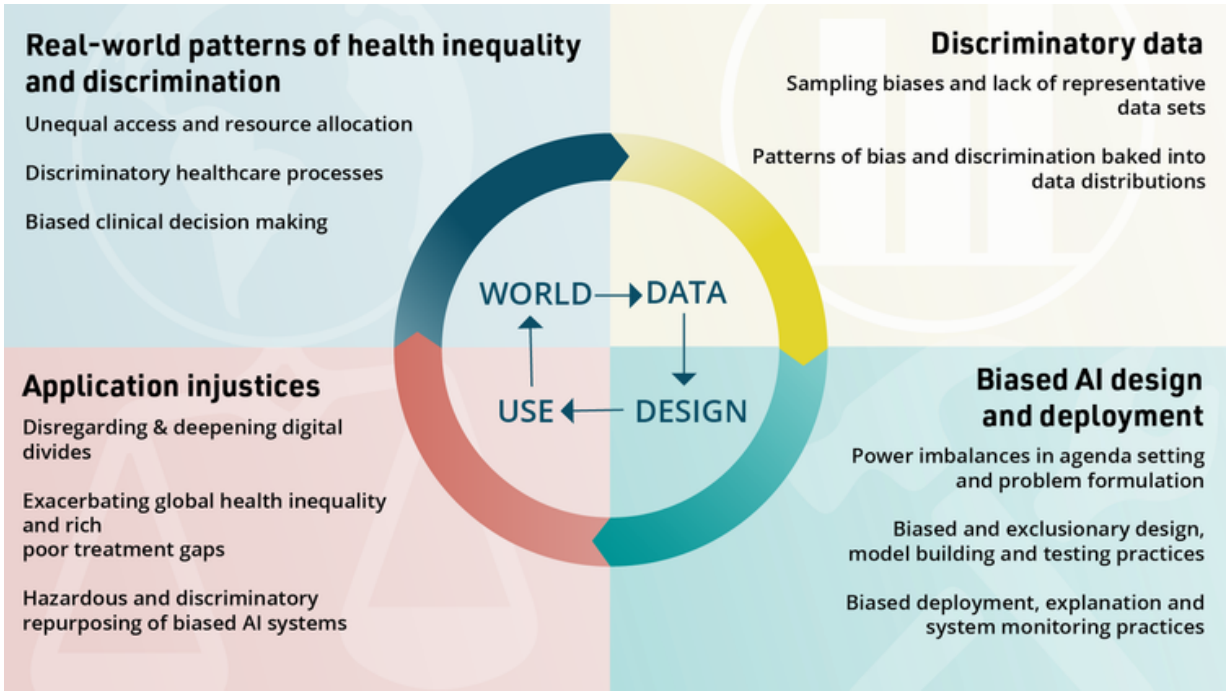
[5] Monica Chin, *The Verge*, "ExamSoft's proctoring software has a face-detection problem" (Jan. 5, 2021), https://www.theverge.com/2021/1/5/22215727/examsoft-online-exams-testing-facial-recognition-report.

[6] Shea Swauger, *NBC News*, "Remote testing monitored by AI is failing the students forced to undergo it" (Nov. 7, 2020), https://www.nbcnews.com/think/opinion/remote-testing-monitored-ai-failing-students-forced-undergo-it-ncna1246769.

[7] "OPTN Board approves elimination of race-based calculation for transplant candidate listing," Organ Procurement & Transplantation Network, U.S. Department of Health and Human Services (July 27, 2022), https://optn.transplant.hrsa.gov/news/optn-board-approves-elimination-of-race-based-calculation-for-transplant-candidate-listing/.

The opacity of AI systems and their implementation suggests that these manifold harms are in fact only the tip of the iceberg, since many systems fail to notify individuals when they are subject to algorithmic decision making. Without notification, it is tremendously difficult for individuals to seek remedy for any harms they experience.[8]

Across every sector, AI systems have been credibly accused of perpetuating discrimination, creating drastic systemic failures, and simply failing to deliver the functionality promised by vendors.



[9]

AI tools can cause harm to individuals and communities in a wide variety of ways. While often presented as solutions to human bias, an AI system can replicate discriminatory biases embedded within the data it analyzes and within the assumptions of those who designed the model. This is particularly evident in predictive policing tools, which often rely on datasets about crime collected by law enforcement agencies engaged in racially discriminatory policing. AI systems also routinely put privacy rights at risk by collecting and aggregating massive amounts of data in order to train and run models.

[8] Sasha Costanza-Chock, Inioluwa Deborah Raji and Joy Buolamwini, "Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem," Association for Computing Machinery ("ACM") Conference on Fairness, Accountability, and Transparency (June 2022), p.9.
[9] The BMJ, [graphic], https://www.bmj.com/content/bmj/372/bmj.n304/F1.medium.jpg

In other cases, AI tools may be insufficiently robust for the problem they are deployed to solve, or may be attempting to solve a conceptually impossible problem. For example, a Taiwanese beauty company recently launched an app that promises to assess users' personalities based on their facial structure.[10] In addition to the privacy concerns raised by collecting user images, and the bias concerns of directing advertisements to users based on those images, the essential premise of this AI tool is based on the debunked assumption rooted in phrenology that there is any link to be found between facial features and personality. Other AI systems may be set to perform a task that is practically impossible, such as predicting crime. As has been routinely documented, there can be no reliable data set of the locations, times, and people involved in crimes. All available data must be rendered an inappropriate proxy by virtue of law enforcement selection bias, crime reporting imbalances, and a myriad of other inequities.[11] Law enforcement agencies and AI vendors have not allowed this fundamental impossibility to halt or even slow the deployment of supposedly predictive AI crime tools.[12]

Civil society organizations and corporate actors are moving quickly to respond to rising concerns about the responsible development and operation of AI systems.

Numerous consortia of cross-industry representatives have formed to generate and promote best practices for AI design and procurement. Even apart from these broader efforts, most major technology companies have put forth their own sets of responsible AI principles, although these principles typically fail to include concrete plans for implementation. Think tanks and policy advocacy organizations have developed additional standards for ethical AI design and remediation, and worked to call attention to the existing and potential harms of such technologies. For example, nonprofit Mijente has expanded its longstanding work opposing discriminatory policing to include policing algorithms.

Despite this energy, the field of investor advocacy has been slow to respond to the burgeoning conversation around AI. Interviews with leading ESG firms indicate that

---

[10] Zara Stone, *The Information*, "The Secret Life of Selfies: How a Beauty Tech Startup Is Using AI to Match Faces with Products" (Sept. 9, 2022), https://www.theinformation.com/articles/the-secret-life-of-selfies-how-a-beauty-tech-startup-is-using-ai-to-match-faces-with-products.

[11] Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz and Andrew D. Selbst, "The Fallacy of AI Functionality," ACM Conference on Fairness, Accountability and Transparency (June 2022), pp.6-7.

[12] Pranshu Verma, *Washington Post*, "The never-ending quest to predict crime using AI" (July 15, 2022), https://www.washingtonpost.com/technology/2022/07/15/predictive-policing-algorithms-fail/.

while there is general awareness among investors of the increasingly widespread risks of AI, only a handful of projects to promote responsible AI currently exist.

Those projects have also tended to be fairly narrowly focused – for example, shareholder advocate As You Sow is partnering with nonprofit EqualAI to include a pledge against AI bias as part of its Racial Justice Scorecard, focusing on company commitments to auditing AI systems used for human resources and other internal processes. Ongoing shareholder campaigns calling for companies to perform racial equity audits or civil rights audits have also targeted algorithms and AI systems, particularly facial recognition, as technologies most in need of discriminatory impact assessment.

Only one investor indicated active engagements with both tech and non-tech companies regarding AI systems. The investor underscored the significant challenges in determining the most effective asks for shareholder advocacy, as well as concerns about the impact of campaigns to win proxy resolution votes at major tech companies with dual-class share structures and concentrated power.

Investors generally indicated a high degree of awareness regarding regulatory efforts around AI, including notable international shifts such as the European Union's proposed Artificial Intelligence Act.[13] Others expressed frustration at an emerging narrative positioning AI technology as a competitive race to innovate between the U.S. and China, which has tended to drown out other valuable conversations regarding the responsible governance of AI systems. Many investors expressed a general openness to exploring more AI-focused engagements.

This work represents a meaningful beginning, but as the scope of AI tools rapidly expands to all essential sectors, investor engagement must as well.

The AI landscape is such that vendors are virtually unregulated, or rather, self-regulated with widely divergent degrees of success. Responsible AI development is a popular topic, and most developers are eager to get out ahead of the issue, if only to preempt governmental regulation.

This has resulted in each AI developer crafting its own individual strategy and set of governance practices for designing responsible AI systems. While certain common themes are emerging, it's too early to assess which practices are most effective – not least because

[13] "The AI Act," European Union (June 1, 2021), https://artificialintelligenceact.eu/the-act/.

many of these internal standards are limited to principles, with little to no information regarding how the company actually plans to operationalize those governance standards.

There are some promising counter-examples. Microsoft has been building and publishing quite comprehensive responsible AI guidance, but it's unclear exactly how these standards will be implemented in practice, and if they will prove to be effective.[14]

Moreover, since AI vendors have a vested interest in profiting off the AI tools they develop, regardless of whether those tools are developed responsibly, there is ample reason to be skeptical of how much industry work may be productive and not mere window-dressing.

For non-technology companies interested in procuring AI systems rather than designing them in-house, the situation is perhaps even more troubling.

Lack of transparency regarding the complex technological functionality and governance design practices behind AI tools makes it extremely difficult for a company purchasing from an AI vendor to perform its own assessment of whether the tool will uphold the company's mission values. Further, an overabundance of company standards and third-party responsible AI certification organizations may offer these customers a potentially false sense of security in the absence of true transparency.

Consequently, both AI vendors and customers are failing to do basic due diligence in developing and deploying AI systems, routinely exposing their brands to serious financial and reputational risks they have not even begun to adequately assess.[15]

Investors seeking to engage companies on these issues must contend with these challenges as well. Both shareholders and company executives may have insufficient understanding of the technological workings of AI systems to anticipate or assess the full implications of the tools being designed or procured. Lack of disclosure renders the development process fatally opaque. In the absence of clear and credible information, the grandiose marketing claims of vendors are given more weight than they rightly deserve, and they can undermine the validity of third-party certifiers' judgments.

In this way, the development of shareholder advocacy regarding responsible AI systems can be considered a test case for a variety of emerging technologies.

---

[14] "Responsible AI Resources," Microsoft.com, last reviewed Sept. 22, 2022, https://www.microsoft.com/en-us/ai/responsible-ai-resources.

[15] Comments of Center for Democracy & Technology, *Request for Information on Financial Institutions' Use of Artificial Intelligence, including Machine Learning*, FR Doc. 2021-06577 (July 1, 2021), p.9.

The greatest challenges for advocacy on AI – technical complexity, lack of transparency, regulatory free-for-all, and breathless science-fiction marketing – are common hallmarks of a market eager for innovation. The opportunities outlined here represent not only a pathway towards successful shareholder advocacy for responsible AI, but also a blueprint for movement-building around future emerging technologies.

There are major opportunities for educational efforts targeted to investors, specifically addressing the insidious scope of AI projects and debunking the hyperbolic claims of their functionality. There is a clear need for increased assessment of the effectiveness of the plethora of emerging ethical AI standards, as well as of the burgeoning market for independent certifiers, which investors are uniquely positioned to meet.

Challenges

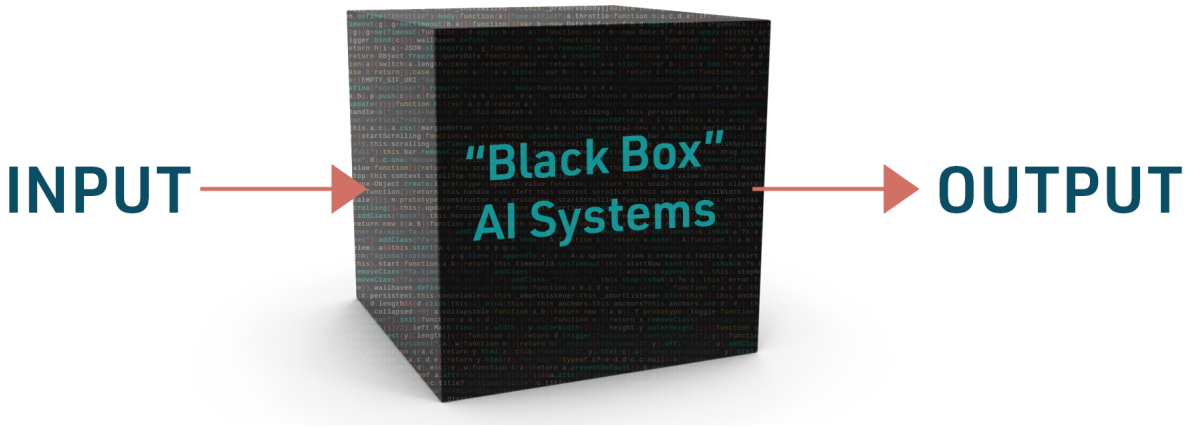
→ Technical complexity and lack of transparency

AI technologies are often extremely complex, "black box" systems in which the decision-making process is not immediately evident to human observers. Even in cases where a problem is identified with an AI tool, it typically requires extensive auditing to discover where in the dataset, training, or technology the bias or error was introduced.[16] The auditing process itself is an emerging field, with no mature standardized framework to guide reviews.

This poses a significant problem for companies seeking to deploy AI tools developed by third-party vendors. Company decision-makers who are not experts in AI or familiar with digital rights issues may struggle to understand the implications of the technology they're investing in, and may be more likely to rely on vendors' marketing claims or the certification of a third-party AI expert.[17]

---

[16] Mark Latonero and Aaina Agarwal, Harvard Kennedy School Carr Center for Human Rights Policy, "Human Rights Impact Assessments for AI: Learning from Facebook's Failure in Myanmar" (2021), p.11.
[17] Financial Stability Board, "Artificial intelligence and machine learning in financial services: Market developments and financial stability implications" (2017), p.26.



AI vendors are typically tight-lipped about their products, and are not compelled by regulation to disclose details about their data selection, development process, or evaluative testing. That leaves companies seeking to deploy AI operating almost entirely in the dark. Implementing good governance practices for the procurement of AI systems does not simply require assessing the available data, but making otherwise inaccessible data available for scrutiny.

For shareholders, the problem is compounded. Whatever voluntary disclosures AI vendors may make to potential customers, they are unlikely to make such information available to either their own investors or existing customers. The proliferation of company-specific AI development principles indicates that technology-sector investors interested in investigating responsible AI will often be directed to these broad-strokes commitments, and denied further information regarding specific implementation and operationalization concerns.

This lack of transparency can sap the will of investors who suspect AI governance is an important issue, but don't feel confident about the right methods of intervention. When the concrete harm of an irresponsible AI system is difficult to pinpoint due to the opaque nature of the technology, it is also difficult to convince investors to beware.[18]

Additionally, some investors already working to engage companies on AI report frustrations in evaluating company policies and establishing sufficient accountability mechanisms. Without a clear understanding of the technologies at play, and the functional impact of those policies, shareholders may struggle to respond critically to companies' claims. The

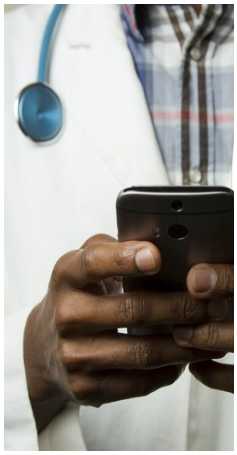
[18] Comments of Center for Democracy & Technology, op.cit., p.5.

lack of information effectively limits the potential depth and impact of shareholder engagement.

ESG investors juggling a panoply of vital issues may reasonably choose to de-prioritize concerns regarding AI systems where neither the problems nor the solutions feel evident or concrete.

→ Rapidly expanding industry scope

AI is no longer an issue only for technology companies. Some of the most pernicious applications of AI tools exist in healthcare, education, financial services, and other sectors.



Healthcare

California Attorney General Rob Bonta is conducting an inquiry into how hospitals and other healthcare facilities are using commercial algorithms to make care decisions that may perpetuate racial and ethnic bias. He cited research that showed an automated system recommending enhanced medical service to white patients over Black patients, based on data indicating that Black patients had spent less on healthcare services in the past and thus were, by the algorithm's logic, adjudged healthier, rather than the victims of systemic poverty and unequal access to healthcare.[19]

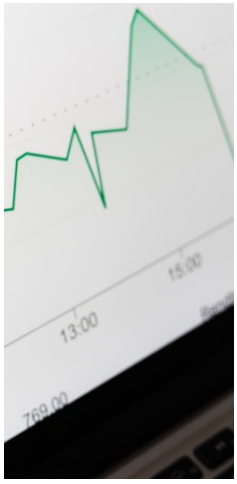
Education

Most online learning systems analyzed in a study by Human Rights Watch were found not only to be collecting and using immense amounts of children's data for the application of EdTech platforms, but also sharing this data with advertising companies including Google and Facebook.[20] These privacy invasions can hardly be considered "opt-in" when they are embedded into mandatory AI technologies deployed by a child's school.

---

[19] "Attorney General Bonta Launches Inquiry into Racial and Ethnic Bias in Healthcare Algorithms," Office of the Attorney General of California (Aug. 31, 2022), https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-inquiry-racial-and-ethnic-bias-healthcare.

[20] Human Rights Watch, "'How Dare They Peep into My Private Life?': Children's Rights Violations by Governments that Endorsed Online Learning During the Covid-19 Pandemic" (2022).



Finance

The finance sector has increasingly turned to AI systems to assess creditworthiness of individual borrowers, deploying algorithms that include not only traditional datasets but also social media activity and mobile phone usage. Such expansions raise serious concerns about individual privacy and introduce new potential axes of discriminatory bias. Moreover, there is reason to fear that if implemented across the financial sector, the collective use of AI technologies to guide trading and other financial transactions may create "herding behavior" that could amplify financial shocks and result in digital collusion to manipulate market prices.[21]

Importantly, none of these issues can be satisfactorily addressed through a campaign exclusively targeting major technology companies.

Tackling responsible AI governance requires coordinated work across industries. While investors and advocates familiar with the technology industry have greater understanding of the kinds of risk these systems can impose, those operating in other essential sectors have invaluable contextual knowledge. Bridging those gaps is essential for understanding and addressing the full scope of AI potential impacts and harm.

Further, the breadth of AI applications will require many-pronged interventions and solutions. The same engagement efforts that work for technology companies may not be appropriate for EdTech or healthcare companies. The same governance mechanisms that succeed with AI vendors also may not be fitting for companies that seek only to procure AI tools. Diffuse problems necessitate diffuse solutions, which presents a challenge for focused advocacy and momentum-building.

→ Insufficient implementation of standards and principles

AI developers face no shortage of industry standards when it comes to responsible design of automated decision making systems. Nonprofit advocacy organization AlgorithmWatch has collected more than 170 separate sets of guidelines in its AI Ethics Guidelines Global Inventory, created by governments, industry associations, and civil society organizations.[22]

[21] Financial Stability Board, op.cit., pp.13, 25.
[22] AI Ethics Guidelines Global Inventory, April 2020, last reviewed Sept. 22, 2022, https://inventory.algorithmwatch.org/about.

For an emerging technology with such a wide variety of applications, a multiplicity of standards may be inevitable, but the glut of guiding principles creates a host of other challenges for investors and advocates.

First, investors report that many of the available AI standards are piecemeal, for example offering guidance for composing diverse teams to work on AI development, but not how those teams should evaluate the tools they design. In other cases, guidelines focus on specific AI use-cases, such as facial recognition technology. While focused sets of principles can be useful for investors and companies, it requires a great deal more research to assemble a comprehensive set of standards from these limited frameworks.

Additionally, the lack of consensus presents a crisis of legitimacy. From this plethora of responsible AI guidelines, academics see some common themes emerging around data reliability, model transparency, and keeping a "human in the loop" of automated decision making. This is a promising trend, but there remains significant variance and the absence of civil society voices absent from such standards work raises important questions for investors seeking to formulate clear requests and accountability measures when engaging with companies.[23]

Moreover, most of these guidelines lack specific definitions of responsible AI principles and concrete plans to operationalize them.[24] Creating trustworthy AI systems is clearly a valuable goal for development teams, but how should "trustworthy" be defined?

Another common definitional dispute is the distinction between "higher-risk" applications and "lower-risk" applications. Many governance standards, including those currently set forth in the EU AI Act, adopt a risk-based approach in

[23] Peter Cihon, "Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development" (2019), p.2.
[24] Costanza-Chock et al., op.cit., p.3.

determining what safeguards are appropriate.[25] While this approach may seem intuitive, the proper categorization of diverse applications is far less so, and investors report engaging with companies to nail down these risk taxonomies.

This lack of specificity allows for multiple companies, or even multiple development teams within a single company, to claim they are upholding the same responsible AI principles while engaging in vastly different governance practices and seeing vastly different results.

That is, of course, assuming the principles are being implemented at all. Some companies may develop or adopt a set of general AI policies and cease to engage with the work once those statements of principle are complete. It is a daunting if not impossible task to comparatively assess potential sets of AI guidelines when there is little to no information regarding how those guidelines are to be implemented both in theory and in practice.

Even some well-intentioned implementation strategies may prove ineffective in particular situations. For example, a common principle in ethical AI deployment is to maintain a "human in the loop" so that a person can observe and correct any bias or errors in the AI's decision making. But human bias still exists, and it's possible for human and AI bias to reinforce each other rather than act as checks on one another. In cases where the inclusion of certain data is found to create bias or errors in the system, deleting that data may sometimes be considered a sufficient remedy. However, for machine learning systems where AI tools are trained on particular data sets, deleting problematic data does not eliminate the "algorithmic shadow" of a model trained on the inappropriate data.[26]

To build a set of legitimate consensus responsible AI guidelines useful for investors and advocates, the existing field must be specified, tested, and winnowed.

This is incredibly challenging considering the vast array of actors engaged in developing AI standards, and the rampant lack of transparency surrounding AI development and procurement practices.

---

[25] "Regulatory framework proposal on artificial intelligence," *Shaping Europe's digital future*, European Commission, last updated June 7, 2022, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

[26] Tiffany C. Li, "Algorithmic Destruction," Southern Methodist University Law Review (2022), p.4.

→ Burgeoning third-party certification industry

To fill the informational and assessment gaps left by complex and opaque technologies, a number of organizations and companies have begun offering responsible AI certifications for companies. The Institute of Electronics and Electronic Engineers has established an Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS).[27] The Responsible AI Institute has developed its own independent certification framework in collaboration with the World Economic Forum, and companies like Credo AI offer platforms to guide companies in the development and implementation of ethical AI governance practices.[28]

The field of responsible AI certification is growing significantly, with some organizations developing tools to guide companies in performing self-assessments and others offering independent third-party assessment.

[27] "The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)," *IEEE*, last reviewed Sept. 22, 2022, https://standards.ieee.org/industry-connections/ecpais.html.
[28] "RAII Certification Beta," *Responsible AI Institute*, last reviewed Sept. 22, 2022, https://www.responsible.ai/certification; "Responsible AI solutions for every use case," *Credo AI*, last reviewed Sept. 22, 2022, https://www.credo.ai/solutions.

Demand for a Global AI Certification Program

Demand segments	Key stakeholders	Main interests/concerns
Suppliers	> Individual developers > Service providers/consulting firms > Suppliers of technology infrastructure	> Knowing how to design and develop AI in a responsible way > Maximizing appropriate use and adoption of AI in a systematic and scalable way > Minimizing legal and business risk > Driving innovation and competitiveness > Differentiating themselves by having good processes in place > Increasing profitability and growth > Reducing operational costs
Buyers	> Procurement officers > Finance and legal teams > Senior management > Ethics boards and legal teams	> Getting better procurement tools > Achieving business goals > Ensuring proper documentation, due diligence, and ethics
Users	> Government decision makers > Individual consumers > Companies of all sizes	> Reaping the benefits of AI (including by improving quality of life, changing behaviors, and taking better decisions) > Understanding what AI trustworthiness characteristics have been recognized internationally and how to evidence and measure them
End Users and Data Subjects	> Consumers and potential consumers > Employees and potential employees > People whose data/AI system uses	> Ensuring fair and trustworthy functioning of AI systems > Ensuring privacy and security of data > Understanding what is being done to protect their interests and data
Educators and Researchers	> Academia > Educators > Research institutes	> Educating the citizens and leaders of tomorrow > Disseminating tools, insights and knowledge
Lawmakers and public service	> National policy makers/regulators > Public sector	> Minimizing harm to society > Increasing benefits of technology for humanity
Shapers	> UN > OECD > GPAI > G20 > Global AI Action Alliance (WEF) > Standards organizations > Industry associations > GAIA projects and partners*	> Improving the state of the world by solving shared global challenges > Facilitating international and multi-stakeholder collaboration > Defining best practices for one or more industries > *Advancing the RAI agenda
Investors	> VCs > Trust funds > Pension funds > Philanthropies	> Investing in quality AI systems that are fit for purpose > Answering demands for ethical investing > Maintaining profitability > Ensuring sustainability

29

[29] Responsible Artificial Intelligence Institute, The Responsible AI Certification Program - White Paper, "Demand for a Global AI Certification Program" (2022),pp. 10-11. https://assets.ctfassets.net/rz1q59puyoaw/2CQ9xgpFyXKLcwXNUXn51G/524638e8f3c976b43252b6bd03aef46d/White_paper_June_11_at_142_pm_ET.pdf

Certification acts as a mark of quality assurance for companies, and may even aid in legal compliance as regulatory structures such as the EU AI Act contemplate requiring third-party assessments for high-risk AI applications.[30]

For AI vendors, certification mechanisms are a form of external validation that may help to quiet their critics and appeal to their customers. For companies seeking to procure AI tools rather than design their own, third-party certifiers present a useful method for assessing if the AI was developed responsibly, even without the technical expertise or disclosures necessary to perform its own assessment.

Conceptually, third-party AI certification tools could also be a major boon to investor advocacy. Shareholders interested in fostering responsible AI development may elect to engage their portfolio companies with requests to participate in reputable assessment and certification processes. This would save investors from needing to reinvent the wheel or develop deep technological expertise as a prerequisite to advocating for ethical AI governance. Some certifiers have even demonstrated a willingness to collaborate with civil society organizations to provide broad scope assessment of company policies.

However, academics have expressed concern that the burgeoning demand for third-party certification presents its own risks. As the certification industry becomes more popular and potentially more profitable, the total number of third-party certifiers will likely multiply, with unreliable or profit-seeking certifiers jumping into the field.

AI vendors and customers may discover they have effectively outsourced their ethical responsibilities to third parties that will remain insulated from any legal or financial risks that companies will incur if the certifier's assessment proves ineffective. Less well-intentioned companies may even "shop around" for rubber-stamp certifications to exploit for competitive advantage.

There are also questions regarding potential conflicts of interest: What happens if technology companies begin funding certain certifiers? Is there a risk of creating "revolving door" issues where employees move from certifiers to aid companies seeking certification?

Many of these concerns exist because there is no existing mature framework for assessing AI systems and governance practices. Regulatory efforts have increasingly called for

[30] Brandie Nonneckie and Philip Dawson, Harvard Kennedy School Carr Center for Human Rights Policy, "Human Rights Implications of Algorithmic Impact Assessments: Priority Considerations to Guide Effective Development and Use" (2021), p.11.

algorithmic impact assessments, particularly in high-risk applications, but there remains no common or internationally standardized approach for the development of these assessments.[31] Some academics and civil society practitioners see promise in adapting the framework of human rights impact assessments (HRIA) for AI applications, but also warn of the potential for HRIAs to be misused to create a false sense of security around the rights impacts of AI systems.[32]

→ Science-fiction narrative messaging

The rhetoric surrounding AI assigns the emerging technology nearly magical powers, either by aggrandizing its supposedly limitless potential, or at the other extreme by warning of valid but wildly premature fears of hyper-competent AI systems running rampant over human control, like HAL in Stanley Kubrick's *2001: A Space Odyssey* movie.

However, the mundane truth of the matter is that one of the greatest risks currently presented by AI technology is that deployed and vetted AI products often simply do not work.

To quote University of Washington professor of computer science emeritus Pedro Domingos: **"People worry that computers will get too smart and take over the world, but the real problem is that they're too stupid and they've already taken over the world."**[33]

Most well-known examples of AI tools being misapplied and causing harm are in fact the result of these technologies failing to function properly, as opposed to functioning at dangerously sophisticated levels.[34]

- Students wrongly accused of cheating by biased and untested proctoring software;[35]

---

[31] Nonneckie and Dawson, ibid., p.2.
[32] Nonneckie and Dawson, ibid., pp.8-9.
[33] Jennifer Langston, "Q&A with Pedro Domingos: Author of 'The Master Algorithm'," University of Washington News (Sept. 17, 2015), https://www.washington.edu/news/2015/09/17/a-q-a-with-pedro-domingos-author-of-the-master-algorithm/.
[34] Raji, Kumar et al., op.cit., pp.1-2.
[35] Shea Swauger, *MIT Technology Review*, "Software that monitors students during tests perpetuates inequality and violates their privacy" (Aug. 7, 2020), https://www.technologyreview.com/2020/08/07/1006132/software-algorithms-proctoring-online-tests-ai-ethics/.

- Weapons-scanning systems failing to detect handguns but falsely flagging laptops;[36]
- Water quality prediction tools incorrectly predicting that beaches will be safe for swimming;[37]

All are examples of AI tools not simply failing to uphold ethical standards, but failing to provide basic functionality.

There are certainly many companies, large and small, dedicated to exploring the full potential of AI systems and genuinely attempting to provide responsible and functional products to customers. However, whether the result of well-intentioned failure or negligent disinterest, research suggests that a great many AI products currently on the market are akin to snake oil.[38]

Despite the prevalence of unsettling instances of AI dysfunction, the discourse around AI is dominated by proponents and vendors making grandiose claims of the technology's might, and by critics that presuppose the technology to be dangerously sophisticated.

In fact, in a survey of AI ethics guidelines, one study found startlingly few standards even acknowledge the possibility of AI systems failing to function as advertised.[39] This suggests that a great deal of current work to foster responsible AI development fails to account for one of its primary risk factors.

Even when AI systems fail publicly and spectacularly – for example, a house-flipping algorithm that lost Zillow $420 million over a mere three months – the market for those tools remains high.[40] The breathless marketing outweighs the quiet voice of reality.

---

[36] Aaron Gordon, *Vice*, "'The Least Safe Day': Rollout of Gun-Detecting AI Scanners in Schools Has Been a 'Cluster,' Emails Show" (Aug. 25, 2022), https://www.vice.com/en/article/5d3dw5/the-least-safe-day-rollout-of-gun-detecting-ai-scanners-in-schools-has-been-a-cluster-emails-show.

[37] Ben Cohen, *Toronto Star*, "Safe for swimming? Toronto's new tool for measuring water quality at its beaches is misleading, advocates say" (Aug. 10, 2022), https://www.thestar.com/news/gta/2022/08/10/safe-for-swimming-torontos-new-tool-for-measuring-water-quality-at-its-beaches-is-misleading-say-advocates.html?rf.

[38] Raji, Kumar et al., op.cit., pp.2-3.

[39] Anna Jobin, Marcello Ienca, and Effy Vayena, *Nature Machine Intelligence* 1(2), "The global landscape of AI ethics guidelines" (2019), pp.389–399.

[40] Matthew Ponsford, *MIT Technology Review*, "House-flipping algorithms are coming to your neighborhood–despite the losses" (Apr. 13, 2022).

Effectively, the dominant narrative surrounding AI has misdirected resources and analysis addressing the technology's potential impacts, as well as governance strategies to mitigate company risks and community harms.

This narrative also has challenging implications for investor advocacy and movement-building around responsible AI. Firstly, it can falsely reduce the perception of risk. In the absence of universally-respected responsible AI standards or trusted third-party verifiers presenting clear and digestible assessments of company practices, investors exposed to unfounded rosy marketing claims may believe the risks of AI tools to be minimal, and not worth worrying about.

Alternatively, the science-fiction style narrative can reposition the risks of AI as a future problem rather than a present one. Some investors may be rightly skeptical of these fanciful marketing claims, but wrongly assume that because the technology is unlikely to perform as promised, that it is not yet dangerous. Some of the worst known harms of irresponsible AI technology occur precisely because the tool does not function properly, and there are insufficient mechanisms to identify or redress the errors it inflicts on individuals and communities when deployed.

To effectively build shareholder advocacy campaigns and address the significant risks created by frequently faulty AI tools, engagements must combat the dominant and inaccurate rhetorical premise that currently deployed AI technologies are extremely effective.

Opportunities

→ Educating investors on AI technologies, applications, and policies

Investors understand that AI systems are spreading to every industry, and that they pose significant risks and opportunities – but many feel they lack the requisite information to form a strategic and effective response.

Consequently, there is a tremendous opportunity for educational efforts and projects designed to share research and resources between civil society, academia, and the investor community.

Firstly, there is a need for basic education regarding the technological potential and limitations of AI systems. This need is especially significant for shareholders operating outside of the technology sector, but will likely be valuable even for those already engaging with major technology companies and AI vendors. The marketing rhetoric surrounding AI systems has become so thoroughly detached from the realities of automated decision-making processes that this education would likely benefit from significant "debunking" components.

The goal of any education efforts in this area should be to empower investors to ask strategic questions during their engagements with companies, including: Where is the data coming from? What assumptions was the model trained on? Is it a static algorithm, or does it evolve?

Secondly, educational efforts should also seek to connect shareholders with developing resources regarding responsible AI best practices and policies. As discussed above, the current landscape of AI standards documents is simultaneously bloated and insufficiently robust – but useful needles remain buried in the haystack. As the field advances, recognized gold standards will likely emerge.

Academic partners and civil society organizations are on the leading edge of this development. Strengthening the lines of communication between these groups will empower shareholders to engage companies with a comprehensive, up-to-date understanding of what strong ethical AI policies and implementation strategies look like.

No current projects exist to fill this need, but the desire is clearly present. Several interviewees expressed interest in networking further on the subject of responsible AI strategies, and a few suggested pulling together an informal working group to share resources and information.

→ Connecting across sectors and silos

The technology sector must be a cornerstone of any campaign to address responsible AI development and deployment. The standards that leading corporations and vendors adopt for the technology sector will have far-reaching implications for AI systems in every application.

However, ignoring the proliferation of AI applications in non-tech sectors is no longer an option. AI tools are widely deployed across countless industries, and current trends suggest this spread will continue apace. Non-tech companies are already broadly exposed to the

legal, financial, and reputational risks of deploying irresponsible AI systems, and the harms already inflicted in these sectors demand immediate intervention.

As investors both inside and outside the tech sector work to foster responsible AI development, there exists tremendous potential in their coordination.

Shareholder advocates familiar with the technology industry tend to be more knowledgeable about the potential axes of harm, including privacy violations, black box systems, and misuse of data. One investor suggested that it's vital to determine how to implement responsible AI procedures in tech first, and then these tools and strategies can inform the development of meaningful policies in non-tech sectors related to AI systems.

Those working in non-tech sectors tend to have a clearer understanding of the material harms and outcomes that these technologies may produce.[41] For example, the generalized concept of "privacy invasion" is far less impactful than the material reality of "being denied healthcare."

Efforts to coordinate cross-sector shareholder campaigns and build coalitions between tech-focused investor advocacy and efforts in education, healthcare, environment, finance, housing, and more have the power to combine a deep understanding of the process with a deep understanding of the impacts of AI systems.[42]

→ Leveraging investors' unique position to assess implementation

The lack of transparency regarding the operationalization of responsible AI principles is perhaps the greatest challenge in the shareholder advocacy space. While general governance standards abound, they are still in early developmental stages, often limited in both scope and specificity. Consequently, assessing the effectiveness of a company's ethical AI standards is usually a matter of guesswork.

Many sets of AI guidelines have narrowly focused on specific industry sectors, or even specific companies. Other industry efforts have earned skepticism from advocates concerned that the company representatives crafting best practices may be more focused on propping up their existing business models rather than building structures to support

[41] Reva Schwartz, Apostol Vassilev, Kristen K. Greene, Lori Perine, Andrew Burt and Patrick Hall, *National Institute of Standards and Technology*, "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence" (2022), at 36.
[42] World Economic Forum, *Empowering AI Leadership: AI C-Suite Toolkit* (January 2022), p.68.

truly responsible AI development and procurement. Even for the most promising sets of standards, minimal data exists to show how they are being operationalized and to what effect.

For best practices to emerge, there needs to be a productive feedback loop identifying implementation strategies and results.[43]

There are some legislative efforts to require companies to disclose their AI governance systems, for example the Algorithmic Accountability Act introduced by Sen. Ron Wyden (D-OR).[44] While the fate of such regulatory efforts remain uncertain at best, however, investors are perfectly positioned to request this information from companies.

There is a significant history of shareholder campaigns successfully seeking increased transparency in public reporting from portfolio companies regarding both policy details and enforcement results. As companies continue to roll out in-house, responsible AI principles or align themselves with third-party efforts, investors are well positioned to push for the publication of more detailed implementation plans. When new AI applications crop up, investors may prompt companies to disclose certain information from self-assessments or third-party assessments of the tool.[45] Investors may also campaign for aggregated data on the results of guideline implementation: How many proposed AI applications were rejected for evidence of bias? How many for unreliable data? How many for non-functionality?

Some investors are already beginning to push companies for greater transparency regarding AI policies. This work would benefit not only from broader support, but also from campaigns that aim specifically to coordinate calls for disclosure with data needs identified by academic and civil society partners.

In this way, shareholder advocacy can hasten the refining of responsible AI standards and effective operational strategies that will later become the bedrock of corporate governance and accountability work around AI systems.

[43] Schwartz et al., op.cit., p.43.
[44] S.1108, *Algorithmic Accountability Act of 2019*, 116th Congress, introduced Apr. 10, 2019, https://www.congress.gov/bill/116th-congress/senate-bill/1108.
[45] Costanza-Chock et al., op.cit., p.9.

→ Auditing the auditors and third-party certifiers

As the number of third-party AI certifiers grows, civil society organizations will need to keep a close eye on this proto-industry. Responsible AI experts have a role to play both by engaging certifiers to help ensure their assessment and verification processes reflect the growing understanding of AI development and deployment principles, and also by calling out bad actors in the space should they appear.[46]

It will be necessary to keep investors educated about developments in this space, so they can continue to hold companies accountable and pressure management to engage with responsible certifiers in good faith.

Investors have had significant success in pressuring major companies to engage in civil rights audits and human rights impact assessments. The application of these tools to algorithmic processes, however, is still in its infancy, and requires further development in order to adequately assess rights impacts for AI systems. Shareholder advocates have an opportunity to build on these victories by incorporating AI impact assessments within those audits, as well as calling for independent assessments of AI systems specifically.[47]

Virtually every robust effort to establish responsible AI governance includes routine impact assessments,[48] but there is no consensus framework for how these assessments should be conducted. Investors have the opportunity to help guide this development by pushing for auditing practices that reflect human and civil rights concerns, that recognize the need for regular assessments of shifting technological applications, and that echo emerging best practices identified by academic and civil society partners.

→ Changing the narrative to address pragmatic business concerns

The discursive narrative attributing nearly mythical powers to current AI technologies is a serious barrier to establishing meaningful accountability and governance practices for automated systems. Existing advocacy efforts have proven troublingly susceptible to these marketing claims, too often accepting the premise that AI tools are so ingenious that our primary concern must be reining in this dangerously powerful technology.

---

[46] Costanza-Chock et al., ibid., p.1.
[47] Latonero and Agarwal, op.cit., pp.1, 14.
[48] Schwartz et al., op.cit., p.14.

Exposing the spotty performance of AI products currently on the market does more than re-set the conversation to reality – it provides a tremendous narrative opportunity for investor advocates.

There may come a day when the greatest risk posed by AI is the threat of its overwhelming technological sophistication, but we are not there yet. The most powerful aspect of existing AI systems is not the technology, but the material power companies assign its automated decisions when they deploy it. Audits of AI tools routinely expose shockingly large margins of error. Those errors are causing tremendous harm right now, not in some science-fiction future.

Affirming this reality puts shareholders seeking to promote responsible AI governance in a superior rhetorical position: **Instead of cautioning companies against deploying supremely powerful technology, investors may instead warn companies away from investing in technological lemons.**

It can be challenging to convince company executives that they should be wary of deploying technologies that work "too well" or are "too capable," especially in engagements with technology companies whose business models are built on high-speed innovation.

All companies have a basic and intrinsic interest in ensuring that the products and services they deploy function as expected. Framing the market for AI tools as rife with snake-oil vendors offering poorly-tested products that may leave companies on the hook for massive errors and damages[49] is both a compelling and accurate narrative.

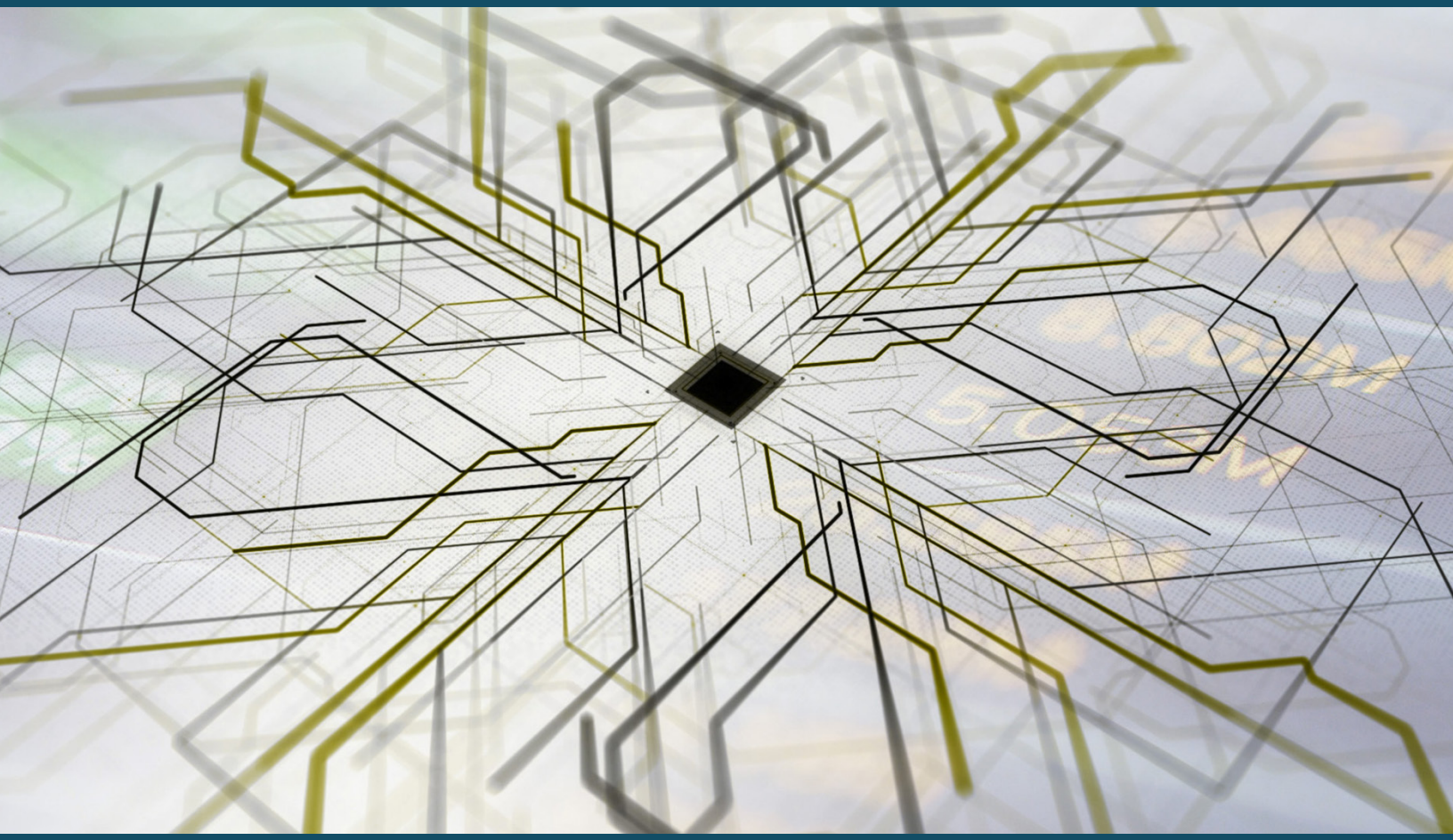
Some analysts even predict that implementing strong responsible AI policies could present a late-mover advantage for smaller firms entering the AI development space. Particularly as the potential for regulation rises, with the EU AI Act and other efforts, investors and customers may be more interested in engaging with companies that have built responsible governance practices into their business models from the beginning.[50] Spotlighting the business risk of deploying non-functional AI systems would build on this opportunity.

Promoting responsible AI governance practices as an effective safeguard against potentially malfunctioning products clearly positions shareholder advocates as proponents of pragmatic business concerns, rather than so-called bleeding-heart investors that have become the target of anti-ESG campaigns.

[49] World Economic Forum, *Empowering AI Leadership*, op.cit., p.63.
[50] World Economic Forum, ibid., p.50.

AI Investment Standards

Building consensus from emerging efforts



www.netgainresearch.com

TECH ACCOUNTABILITY
FINANCE-FOCUSED STRATEGIES

NetGain
Partnership

OpenMIC