Corporate governance & leadership

Play 5: Establish responsible AI governance & end-to-end internal policies to mitigate bias

It is important to establish corporate governance for responsible AI and end-to-end internal policies and guidance to mitigate bias. Good practices for responsible AI governance include, for example: cultivating a sense of shared responsibility, assessing and updating incentive structures and power dynamics that can dissuade individuals from speaking up, and examining leadership priorities and limitations.

PLAYERS INVOLVED:
- Board of Directors
- CEO
- AI Ethics Board
- AI Ethics lead & team members

BUSINESS BENEFITS: • Enhance accountability and mitigate risk • Improve public perception

Elements:

☐ Establish an AI ethics lead, AI ethics board and AI ethics code/principles.
- Include individuals on the board who may be affected by the AI systems (e.g., customers, community members (especially women, people of color and other underrepresented identities)).
- Ensure the AI lead has a team and support to operationalize, monitor and enforce the code / principles.
- Establish which AI use cases / areas the company will not work with or on, and when to pull the plug on certain AI systems.

☐ Establish and formalize processes to operationalize the principles. Include concrete guidance and tools to plan for, identify and mitigate bias.
- Engage in / inform conversations and approaches related to fairness.
- Ensure there is accountability for mitigating bias at the leadership level.
- Ensure managers understand that identifying and addressing issues related to bias is expected and they are liable for issues. Add this in performance reviews and OKRs for managers.

☐ Leverage / engage different departments and aspects of the business to further understand, tackle and mitigate bias in AI internally (e.g., working groups including CSR teams and other AI experts).

☐ Assess how leadership priorities can impact responsible AI practices. Be honest, as well as transparent, about limitations.
- Ensure responsible AI is valued and seen as a leadership priority.
- Allocate sufficient funding and resources to reflect responsibility and bias mitigation as a clear priority.

Tools:

- **Empowering AI Leadership** (World Economic Forum)
- **Principled Artificial Intelligence: A Map of Ethics and Rights-Based Approaches to Principled AI** (Harvard University)

- See play 4 for tools and guidance to operationalize fairness and mitigate bias in developing AI systems
- See play 2 on practices to advance a culture of responsible AI among staff

Example & leader: Microsoft has six responsible AI principles: Fairness, reliability and safety, privacy and security, inclusiveness, transparency, and accountability. It operationalizes responsible AI through its Office of Responsible AI (ORA) and its AI and Ethics in Engineering and Research (Aether) Committee. Aether, set up in 2017, makes recommendations to senior leadership related to responsible AI and has a working groups specific to AI fairness and inclusiveness. Aether engages and draws from across the company: the AI fairness and inclusiveness working group, for example, includes AI experts and leaders in their CSR department.[1] ORA – set up in 2020 in its Corporate, External and Legal Affairs division and led by the Chief Responsible AI Officer – puts Microsoft principles into practice. It does this through setting company-wide standards for responsible AI, helping teams adopt responsible AI practices, reviewing sensitive use cases, and informing public policy. Microsoft also has Responsible AI Champs, which are domain experts that raise awareness of the AI standards and help their teams put them into practice. Read more about Microsoft's approach **here** and its Responsible AI Champs **here**.

Background:

AI creates new tech governance challenges – like overseeing technical systems that can embed and amplify harmful bias, while navigating a changing regulatory landscape and maintaining alignment with corporate values. The firm's CEO and board of directors has an important role to play, including establishing governance structures such as AI ethics board, AI ethics lead and responsible AI prinvciples / codes.

- **AI ethics (or responsible AI) lead:** Many companies are appointing chief AI ethics officials or other designated individuals that lead and manage responsible AI in the company – with mitigating bias and navigating "fairness" as one component. These individuals lead assessments to determine harm AI system can pose and approaches to mitigate potential or existing biases of AI systems. They can oversee the company's accountability structure related to responsible AI, as well as internal guidance and compliance.[2]
- **AI ethics boards:** An AI ethics board can work closely with the AI ethics lead and oversee risk assessments and harm mitigation strategies[3] while aligning with the goals of the firm's board of directors. A key role of the AI ethics board may be developing responsible AI principles / codes, as well as recommending performance indicators and exploring whistleblowing mechanisms.[4]
- **Responsible AI principles / codes:** Many companies have or are developing responsible AI codes or principles that can help guide ethical decisions in developing, managing and using AI. Principles can inform new strategies or initiatives, while impacting employee behavior.[5] Externally, principles provide assurances to the public, customers or other stakeholders about the ethics of the company and its AI practices.[6] While companies have slightly different principles / codes for responsible AI depending on their context and industry, there are trends and similarities. For example, "fairness" is found across different principles. For more on a global landscape of AI ethics guidelines and principles, read **this study** by Jobin et al. surveying AI ethics principles[7] and see **this map** by the Berkman Klein Center for Internet & Society at Harvard University.

While promising, AI ethics governance has its limitations and if not done well, can be ineffective. The below limitations are critical to acknowledge and keep in mind.

- **Who is deciding what "ethical" AI means?** By setting their own principles and approaches, companies are deciding what it means to responsibly deploy these technologies and what "ethical" AI means for society. Those at the top of corporate hierarches – who tend to be white men – are setting the direction.[8] It is key for the AI ethics board to be diverse and incorporate perspectives of those who will be impacted by AI systems.
- **Bias mitigation at odds with company & leadership priorities:** Being first-to-market is a critical priority for firms. However, this priority can be at odds with mitigating bias and executing responsible AI, which requires pause points and new processes that can slow down getting a product to market. Corporate leaders need to acknowledge how traditional priorities can impact ethical and responsible AI goals, and inadvertently expose the firm to great risk and immense financial costs. Reassessing existing priorities and ensuring responsible AI is a clear, top priority is key. Ultimately, the tension between ethics and market priorities reflects a short-sighted view of what market success means. In the long run, cutting ethical corners risks serious reputational and legal consequences, which is especially true given the rapidly changing regulatory landscape around AI.
- **Lack of accountability & operationalization guidance:** There is often a lack of formal guidance to operationalize principles.[9] Also, there is limited accountability when a firm's principles are violated.[10] In many cases, organizational culture itself impedes the principles being put into action as the culture (and market more broadly) prioritizes efficiency over fairness and bias mitigation. When operationalizing principles related to bias and fairness, it's important to reflect on what is "fair" for a particular AI system and who is defining "fair". **EGAL's brief on fairness in machine learning** delves into this topic, outlines tools and considerations.
- **Focus on technical solutions & "ethics washing":** Most principles and associated guidelines suggest that technical solutions exist for problems that arise[11] and tend to focus on technical forms of bias.[12] Only using technical solutions to issues such as fairness and bias misses the broader picture, and reflects the belief that technological innovations can solve complex societal challenges (techno-solutionism). This focus also runs the risk of "ethics washing", in which a commitment to addressing AI ethics issues is unsupported by concrete actions. **EGAL's brief on fairness in machine learning** outlines some qualitative tools that can help. Initial high-level assessments of the tech's potential for harm and documenting choices made in developing the AI system is important, particularly for higher risk applications.[13]

This play is part of **Mitigating Bias in AI: An Equity Fluent Leadership Playbook** *of the Berkeley Haas Center for Equity, Gender and Leadership. It was written by Genevieve Smith with valuable feedback from Ishita Rustagi (EGAL) and Matissa Hollister (World Economic Forum).*

Endnotes

1 Heu-Weller, M. (May 28, 2020). Personal interview.
2 IBM. https://www.ibm.com/blogs/policy/ai-precision-regulation/
3 IBM. https://www.ibm.com/blogs/policy/ai-precision-regulation/
4 Governance: Empowering AI leadership. World Economic Forum. Retrieved on April 15, 2020 from c. https://spark.adobe.com/page/3MEmZh5EprOXJ/.
5 Epley, N. & Kumar, A. (2019). How to design an ethical organization. Harvard Business Review.
6 Ethics: Empowering AI leadership. World Economic Forum. Retrieved on April 15, 2020 from https://spark.adobe.com/page/kPM8ZGTlnhOgT/.
7 Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence (2019), 1–11.
8 Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 2020.
9 Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 2020.
10 Loukides, M., Mason, H. & Patil, D. (2018). Oaths and checklists. Retrieved from https: //www.oreilly.com/ideas/of-oaths-and-checklists.
11 Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 2020.
12 Madaio, M., Stark, L., Vaughan, J. W. & Wallach, H. (2020). Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. CHI 2020 paper. http://www.jennwv.com/papers/checklists.pdf.
13 IBM. https://www.ibm.com/blogs/policy/ai-precision-regulation/