

Quick win!

Update performance review process & OKRs (Objectives & Key Results)

About: Cultivating a sense of shared responsibility, whereby all employees feel responsible to address bias issues and enable fairness in AI systems, raises the likelihood of ethical decision making in organizations.¹ This can be done by ensuring employees are empowered and expected to exercise agency for constructive dissent that can help mitigate biases. Organizations can benefit from making it the expectation for employees to identify and flag ethical issues or potential biases at play, and building in more accountability if employees fail to report or act.

Business leaders can add or adjust criteria in the performance review process, and update OKRs / KPIs to integrate goals around mitigation of bias and responsible AI. This incentivizes behavior and enables individuals in team lead roles to justify resources needed to address such issues.² It also helps to ensure that multidisciplinary teams value and incorporate opinions of team members from social science backgrounds equally to those of engineering and computer science.³

Performance review criteria and OKRs will vary from company to company, while also varying between positions. It is recommended to tie them to the company's responsible or ethical AI principles. They should also be integrated and adapted for employees from entry-level positions through management and leadership. Adapting performance review criteria, OKRs and KPIs is critical to the position and project(s).

Players: C-suite leaders, HR leaders

What you will have after executing this 'quick win':

- » A performance evaluation process that more seriously reflects your company values of equitable and responsible technology development
- » A concrete example of your ethics and values in action throughout your company to share with the board and other leaders

Steps:

1. Update individual performance review process to include a component around responsible and ethical practices for teams related to developing, managing or using AI systems. Also update individual performance reviews for management and leadership.

Background – Organizations' approach to performance reviews differ but often include an annual performance review with a mid-year or regular checkpoints. At many tech companies, managers assess results attained and behaviors. The employee starts with a self-assessment, which is accompanied by peer reviews. Common criteria⁴ include: problem solving, execution, thought leadership, and leadership / teamwork. Employees rank themselves and highlight accomplishments, while peers can review them as well.

Do – Integrate an additional criterion for the review process. This additional criterion and accompanying questions will likely look different for entry-level employees vs. managers vs. leadership and should be adapted accordingly.

- ✓ **Criterion: Responsibility & ethics:** The employee’s ability to implement responsible and ethical decision-making and actions related to development, management and use of technical systems.
 - Include examples of what it means to meet expectations within the new criterion. For engineers, this could include:
 - *Does not meet expectations:*
 - ◆ The individual does not consider potential bias and fairness considerations of the AI system, or does so inconsistently. The individual does not regularly consult or engage with domain experts and/or other social scientists to understand how the system could result in bias, particularly in AI systems where fairness tradeoffs may be present. The individual does not use – or inconsistently uses – existing or recommended tools to consider fairness. The individual tends to focus only on technical forms of bias (e.g., over or under representation in the datasets used), and doesn’t think about other ways bias can enter.
 - *Meets expectations:*
 - ◆ The individual consistently flags potential ethical concerns or biases in the datasets as well as algorithms. The individual regularly consults and engages with domain experts and/or other social scientists to better understand fairness implications. The individual is comfortable with and regularly uses tools to mitigate bias (both technical and non-technical).
 - *Exceeds expectations:*
 - ◆ The individual achieves the “meets expectations” category. In addition, the individual goes above and beyond by providing constructive input to the company on improving practices to enhance responsible AI. The individual enhances or develops tools and practices to advance responsible AI, contributes to research to advance responsible AI, and/or advances creative innovations for responsible AI within the company.
- ✓ **Accompanying questions** for self-assessment (to be selected and adapted based on the individual’s role and level) can include:
 - *Did you utilize “pause points” at each AI system lifecycle stage to assess potential bias and fairness considerations?*
 - *Did you use tools to identify and mitigate different forms of bias in AI systems (e.g., Datasheets for Datasets, Model cards, Fairness Analytic)?*
 - *Did you consider and act on how to mitigate biases in the development, management or use of AI systems?*
 - *What are other ways that you took responsible and ethical decision-making in the development, management and use of technical systems?*
 - *How else did you advance the company’s ethical AI principles – such as those related to fairness?*
- ✓ **Consider asking questions to source information on how the company can improve.** Take the opportunity to get information from employees around how company policies and practices are working or could improve to advance responsibility. In addition to gathering helpful information, this takes the focus off of individual action to more organizational efforts.
 - *What would you like to see within the company to make it easier for you and your team to advance responsibility and ethics in AI systems?*
 - *If you raised concerns around bias, were those taken seriously? What other support would you like?*

- Are there certain practices or policies that might be hindering advancements toward responsibility and ethics in AI?

2. Update metrics to determine success (i.e. Objectives & Key Results (OKRs))

Background – OKRs are a tool for guiding and executing the strategy of the organization. An OKR is a set of one Objective and n Key Results. The Objective is the business result that needs to be achieved and is often written in qualitative terms. Key results help “prove” if the Objective was achieved based on specific key performance indicators (KPIs).

Do – Integrate an additional OKR for teams developing, managing and using AI systems. For example:

- ✓ **Objective:** Mitigate bias in AI systems developed, managed and used by the company.
- ✓ **Key results:**
 - Assess and document where and how bias is present in datasets used and work to mitigate biases.
 - Ask for, document, and integrate perspectives of domain experts and social scientists, particularly in AI systems where fairness tradeoffs may be present.
 - Document what is in a particular AI system and decisions made.
 - Integrate ‘pause points’ at each stage of the AI system lifecycle to consider and discuss potential bias issues or other ethical considerations; and document key learnings and takeaways from the ‘pause points’. (These can be used for further training, information dissemination, etc.)

3. Pilot the updated performance review process and success metrics.

4. Get feedback and refine as needed.

Endnotes

- 1 Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 2020.
- 2 Burrell, J., Mulligan, D. & Klutz, D. (2018). Report from the first AFOG summer workshop. Algorithmic Fairness & Opacity Group (AFOG). https://afog.berkeley.edu/files/2019/07/AFOG_workshop2018_report_all_web.pdf.
- 3 Pham, K. (2019). Want to fix big tech? Change what classes are required for a computer science degree. *Fast Company*. <https://www.fastcompany.com/90355969/want-to-fix-big-tech-change-what-classes-are-required-for-a-computer-science-degree>.
- 4 Criteria and processes drawn from information on Google (<https://qulture.rocks/en/blog/googles-performance-management-practices-part-1/>) and Facebook.

This Quick Win resource was developed by the Center for Equity, Gender & Leadership (EGAL) at UC Berkeley Haas School of Business. It was written by Genevieve Smith with contributions from Nitin Kohli and Ishita Rustagi. It is an accompanying resource of [Mitigating Bias in Artificial Intelligence: An Equity Fluent Leadership Playbook](#).

