Empower Diversity in AI Development

Diversity Practices that Mitigate Social Biases from Creeping into Your AI

Karl Werder†
Department of Business IT,
IT University of Copenhagen,
Copenhagen, Denmark
karw@itu.dk

Lan Cao
Information Technology &
Decision Sciences
Old Dominion University
Norfolk, VA, USA
lcao@odu.edu

Balasubramaniam Ramesh
Computer Information Systems
Georgia State University
Atlanta, GA, USA
bramesh@gsu.edu

Eun Hee Park
Information Technology &
Decision Sciences
Old Dominion University
Norfolk, VA, USA
epark@odu.edu

We suggest that social biases are exacerbated by the lack of diversity in the AI field [6]. These biases cannot be effectively addressed by technical solutions that aim at mitigating biases stemming from data sources and data processing or from the algorithm itself [9]. We argue that a social view—which has been neglected in AI development so far—is needed to address the root causes of some biases, given that AI systems are often reflections of our social structures. While great technical progress has been made in measuring and testing fairness [4] and mitigating unfairness [1], biases may originate from any stage of AI development through the developers involved [6]. As a result, some AI system biases reflect the social biases present within the AI developers that build them. Hence, we argue that the lack of diversity in AI development is a source of social biases. As a solution, we present a set of practical recommendations that empower organizations to increase diversity in AI development. In an online supplement (https://osf.io/854ce/), we also present prior work on AI development biases and bias mitigating and exacerbating practices.

Lacking Diversity in AI Development: A Source of Social Biases

We argue that a lack of diversity in AI development contributes to AI system biases in which individuals' cognitively and affectively induced biases creep into the AI system. We call these *social biases (see sidebar for more information).* AI developers with similar demographic backgrounds make similar (mis-)judgments, and hence, run the risk of codifying their social biases into an AI system that reinforces them. In contrast, we know that diversity is associated with positive outcomes [5]. For example, cross-cultural diversity and gender diversity improve requirement specification, project performance, and innovation, and they reduce biases. Without a diverse team, AI development may focus only on certain design considerations and performance measures based on narrow value judgments without considering the shared values of the broader community and diverse stakeholders [3].

Sidebar — Description of social biases with examples

Social biases include availability biases, confirmation biases, bounded awareness, and affective and emotional biases. We describe each type below.

AI developers with ***availability biases***—that is, judging events and their frequency differently based on vividness, recency, or memory structure—often 'go where the data is' by focusing on datasets that are available or accessible rather than datasets that are most suitable. As a result, the data used are not fully representative of the target population and can differ significantly from reality. Take facial recognition of people with dark skin for example that is less accurate because camera technology provides lower-quality images caused by limitations in lighting and contrast.

Confirmation biases—that is, seeking confirmatory information, interpreting newer information from an anchor—lead AI developers to seek confirmatory information, interpret newer information from an anchor (i.e., the first piece of information given about a topic), and be overconfident in their own judgments' infallibility. Two exemplary biases from the development of an AI system for human resource management suggest that i) recruiters favor candidates who are like them in age, race, and

†Corresponding Author

https://doi.org/10.1145/1234567890

attitudinal characteristics, and ii) recruiters assess candidates based on a particular group or class of people, such as prior employers.

AI developers with ***bounded awareness***—that is, relying overly on irrelevant information in specific conditions and failing to see the obvious—ignore important, accessible, and perceivable information during decision-making because of information selection and inattentional blindness. For example, AI developers often fail to select training images and facial features that are not from their own race, which is referred to as the "other-race effect". This may explain why facial recognition algorithms developed in China, Japan, and South Korea recognize Asian faces more accurately than Caucasian faces and vice versa.

Affective and emotional biases—that is, relying on their emotions rather than rationale evaluation—impact AI developers' decision-making and productivity, as they rely on emotions rather than cognition. Take empathy, for example, a core technique for human-centered design and design thinking that emphasizes understanding users' situations, feelings, and needs. However, it is virtually impossible for AI developers to effectively empathize with minorities, thus, giving rise to potential biases.

However, benefiting from diversity is challenging because it requires the right mix of participants (e.g., in hiring) and involves creating policies and procedures that help take advantage of diversity. Engaging only in shallow actions without making any meaningful changes, so-called *diversity washing*, will not address the fundamental problem and may even be counterproductive. Rather, *empowered diversity* goes beyond superficial or tokenistic efforts and encompasses a deep commitment to engaging AI developers from diverse backgrounds.

Empowering Diversity in AI Development

Empowering diversity benefits all levels of an organization, that is, it positively affects developers, teams, and the organization. However, empowering diversity in practice can be challenging in AI development, particularly in Science, Technology, Engineering, and Mathematics (STEM) fields, given the limited access and opportunities for mobility and education by marginalized groups, for example, the long-lasting shortage of women graduates. Hence, we provide five practical recommendations that help organizations increase and empower diversity in AI development.

Cultivating diversity skills: At the individual level, managers need to equip AI developers with a strong understanding of various social biases and their impacts. AI developers first need to acquire diversity as a skill before they change their behavior. Take *confirmation biases*, for instance, when AI developers work in a male-dominated environment, they may take this as a given and focus on confirming evidence.

Managers can cultivate diversity skills by training developers to recognize and avoid this cognitive trap, ensuring they do not neglect the different experiences of others. For example, AI developers can use specific methods such as GenderMag to identify potential biases related to gender in AI systems [2]. GenderMag encompasses several practices including evaluating software features for potential gender biases, creating diverse user personas to understand how different genders interact with the software, uncovering biases in task flows and interactions by cognitive walkthroughs, collecting data on user demographics for inclusive design decisions, and testing with diverse groups. Using these tools helps AI developers to search for trade-off solutions that satisfy competing goals [7].

In addition, managers need to promote interactions between different groups within the organization by ensuring everyone has equal status, sharing common goals, fostering cooperation, and providing institutional and social support [8]. These intergroup contacts create a positive environment for learning from each other's experiences, which in turn develops diversity skills among the team members.

Mirroring target stakeholders' compositions: At the team level, the compositions of the AI development team should mirror those of the system's affected stakeholders to mitigate social biases. Take *bounded awareness* for instance, when the team lacks diversity in skin color, they may overlook the effects of facial recognition AI on different skin tones.

Managers can adjust HR practices to mirror the composition of target stakeholders. For instance, in diverse hiring, implementing blind hiring techniques helps counter bias influenced by bounded awareness. By combining diversity reporting with well-crafted diversity performance indicators, managers can measure the effectiveness of updated HR practices and demonstrate progress. This approach also raises awareness of potential implicit biases that are harder to crack. Mirroring the target stakeholders' composition not only improves team behavior by managing team abrasion and mitigating groupthink but also fosters innovation and creativity. Managers can use tools from psychology, for example, the Hermann Brain Dominance Instrument (see https://www.thinkherrmann.com/hbdi), to identify complementary profiles and modes of thinking. Diversity of thoughts and viewpoints improves collective creativity and helps the team become more innovative.

Promoting inclusive knowledge sharing through experiences: At the organizational level, managers should be aware of the positive and negative implications of lacking empowered diversity. Take *availability bias*, for instance, organizations should encourage sharing both success stories and failures within and outside the organization because it helps prevent narrow and one-sided views of the workforce due to availability bias.

When managers facilitate knowledge exchange within their teams, they promote organizational understanding through shared experiences. For example, acknowledging negative experiences like gender discrimination reported by many women in AI and software development can raise awareness about existing biases and their impact on individuals' career progression. Managers can further this by launching awareness campaigns and providing internal workshops, for example on emotional intelligence, to mitigate dominance and foster empathy within the organization.

Managers should also allocate at least an equal number of resources to facilitate the sharing of positive experiences. Identifying and cultivating success stories for their diverse audience showcases possibilities that are otherwise unrecognized. Managers should promote marginalized groups within the organization, because internal success stories are especially impactful. Speakers who serve as role models for meaningful change may offer insights into organizational processes and practices from marginalized perspectives. When internal success stories are limited, managers can also engage external speakers to share their perspectives and experiences.

Fostering long-term sustainability in diverse talent pipeline: One important challenge for organizations that want to empower diversity is the availability of talent within their environment. Therefore, managers must develop a sustainable talent pipeline that speaks to their diversity needs. Considering the lack of availability of female candidates as an example, given the pervasive nature of *availability bias*, which has raised concerns shared across different areas in the STEM field, developing suitable candidates requires sincere collective efforts with a long-term goal in mind.

Organizations should engage with their environment when developing a diverse workforce. Given the severity and longevity of the problem, organizations need to explore new ways of encouraging underrepresented groups to take on roles in AI development. The fact that established means through targeted online job advertisements are already biased only exacerbates the problem. While talent pools are limited and sometimes hard to access, organizations are encouraged to go where diversity is, that is, institutions of higher education. Organizations that collaborate more closely with educational institutions that serve a diverse population find it easier to develop a sustainable pipeline of diverse talent. For example, organizations often inform and sometimes co-develop educational curricula through their business needs as part of employer panels (e.g., https://en.itu.dk/About-ITU/Organisation/Advisory-Panels/Employers-Panels).

Communicating diversity as an important business need fuses diversity into the curricula development process.

In addition, organizations can engage in events that foster the growth and advancement of women in technology. For example, the Grace Hopper Celebration of Women in Computing focuses on empowering women to learn new skills, make connections, discuss innovative trends, and access motivational leaders.

While engagements with the environment can take some time to develop and evolve, organizations also need to harness and build on what they already have, for example, by offering opportunities for career advancement to retain diverse talent. Thus, organizations need to develop clear career paths for progression and promotion from within. Developing a positive culture for often underrepresented social groups positions the organization as an attractive employer. This signals that the organization does not see diversity as a goal in itself, but rather as a tool for accomplishing and delivering organizational objectives.

Establishing a diversity charter for AI development: Finally, managers should develop an active agenda for proactive change. Executives can lead the charge by embracing existing government regulations, such as the US Algorithmic Accountability Act of 2022 and the EU's General Data Protection Regulation. The latter, for example, provides organizational stakeholders a *right to explanation* and thus the ability to assess potential biases. Rather than reacting to these trends, regulations, and laws, managers should proactively embrace diversity and the changes that come with it. For example, managers can develop a diversity charter for AI development and establish task forces composed of employees from various demographic, ethnic, and other backgrounds, composed of representatives from different levels within the organization. Task forces and the charter help facilitate ongoing dialogue and action plans to continuously identify and address diversity challenges while ensuring compliance with relevant regulations.

A proactive leadership mindset allows organizations to embrace diversity by relying on each employee's unique strengths and skills to contribute to the organization's overarching goals. Organizations that embrace this mindset and develop the corresponding implementation agenda are better positioned to develop responsible AI systems. For example, AI development organizations can embed diversity and inclusion principles into their AI development lifecycle, ensuring that diverse perspectives are represented in the design, development, and deployment of AI systems to mitigate potential biases and promote fairness.

REFERENCES

[1] Andre F.Cruz, Pedro Saleiro, Catarina Belem, Carlos Soares, and Pedro Bizarro. 2021. Promoting Fairness through Hyperparameter Optimization. In *Proceedings of the 2021 IEEE International Conference on Data Mining (ICDM)*, December 2021. IEEE, 1036–1041. https://doi.org/10.1109/ICDM51629.2021.00119

[2] Mariam Guizani, Lara Letaw, Margaret Burnett, and Anita Sarma. 2020. Gender Inclusivity as a Quality

Requirement: Practices and Pitfalls. *IEEE Softw* 37, 6 (November 2020), 7–11. https://doi.org/10.1109/MS.2020.3019540

[3] Karen Hao. 2019. This is how AI bias really happens—and why it's so hard to fix. *MIT Technology Review*, 1–3. Retrieved August 2, 2022 from https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

[4] John P. Lalor, Ahmed Abbasi, Kezia Oketch, Yi Yang, and Nicole Forsgren. 2023. Should Fairness be a Metric or a Model? A Model-based Framework for Assessing Bias in Machine Learning Pipelines. *ACM Trans Inf Syst* (March 2023). https://doi.org/10.1145/3641276

[5] Frederik Nilsson. 2021. Building a diverse company culture means empowering employees. *Forbes*. Retrieved September 22, 2022 from https://www.forbes.com/sites/forbestechcouncil/2021/01/13/building-a-diverse-company-culture-means-empowering-employees/?sh=6527cffd4f3d

[6] Steve Nouri. 2019. Diversity and inclusion In AI. *Forbes*. Retrieved October 11, 2022 from https://www.forbes.com/sites/forbestechcouncil/2021/03/16/diversity-and-inclusion-in-ai/

[7] Christoph Treude and Hideaki Hata. 2023. She Elicits Requirements and He Tests: Software Engineering Gender Bias in Large Language Models. In *Proceedings of the 2023 IEEE/ACM 20th International Conference on Mining Software Repositories*, March 17, 2023. Institute of Electrical and Electronics Engineers Inc., 624–629. https://doi.org/10.1109/MSR59073.2023.00088

[8] Yi Wang and Min Zhang. 2020. Reducing implicit gender biases in software development: Does intergroup contact theory work? *ESEC/FSE 2020 - Proceedings of the 28th ACM Joint Meeting European Software Engineering Conference and Symposium on the Foundations of Software Engineering* 20, (November 2020), 580–592. https://doi.org/10.1145/3368089.3409762

[9] Karl Werder, Balasubramaniam Ramesh, and Rongen Zhang. 2022. Establishing data provenance for responsible artificial intelligence systems. *ACM Trans Manag Inf Syst* 13, 2 (June 2022), 1–23. https://doi.org/10.1145/3503488