



THE INFORMATION CENTER

## Topic Brief: Gender Bias in AI

### Can We Program an Equitable Future?

The field of artificial intelligence (AI) is growing at a rapid pace, developing algorithms and automated machines that show promise in making the workplace **more efficient and less biased**. Many of us already interact with artificial intelligence in our daily lives, often without even realizing it—it's responsible for everything from credit score calculators to search engine results to what we see on social media.<sup>1</sup>

Likewise, organizations have introduced AI into many work processes, especially recruiting and talent-management functions. In many cases, algorithms sort through numerous factors to profile people and make predictions about them. AI hiring and talent-management systems have the potential to **move the needle on gender equality in workplaces by using more objective criteria in recruiting and promoting talent**.<sup>2</sup> But what happens if the algorithm is actually relying on biased input to make predictions? Can machines and artificial intelligence develop unintentional biases, creating the same inequities as people with unconscious biases?

More and more evidence indicates that humans are programming their own biases, including around gender and race, into the algorithms behind AI. How is this happening? **What can be done to prevent bias in AI in workplaces?**

### What Is Artificial Intelligence?

Most experts define artificial intelligence as either a computerized system that understands, learns, and performs actions seen as requiring intelligence, or a designed system for understanding and analyzing data, complex problem-solving, and taking action when encountering the real-world situations for which it was created.<sup>3</sup> There are three main ways that AI works:<sup>4</sup>

- **Assisted Intelligence**

Current technology: It improves what we are already doing, often automating routine tasks based on clearly defined human input. It can assist with making tasks easier but is limited to acting within strict parameters and ultimately relies on humans making the final decisions. It needs people with STEM skills to create, monitor, and

frequently fine-tune its programs/algorithms. Examples of assisted intelligence are a machine in a factory assembly line or a voice-activated digital assistant.

- **Augmented Intelligence**

Current and emerging technology: It helps organizations and people do what they otherwise couldn't do, enhancing decision-making; it often relies on a complex partnership of humans developing, interacting with, and training the AI. This type of artificial intelligence analyzes information and makes recommendations for solutions, collaborating with humans to act and decide together. As augmented AI becomes more widespread, human skills such as collaboration, creativity, persuasion, and innovation will become even more valuable in the workplace to help create and train these programs to become better in augmenting problem-solving and decision-making. An example of this AI is a program that makes recommendations on whether to approve a consumer's loan application.

- **Autonomous Intelligence**

Emerging and future technology: It acts on its own—with the ability to reason, learn from experience, and make autonomous decisions within strict lines of accountability. It is unclear how humans will interact with autonomous intelligence, although it may serve as part of a work team with people. Many see this as automating some complex human decision-making, others see it as freeing up people to do more creative, complex work in new types of jobs. An example of this type of AI is a self-driving car.

## The Good News: AI Can Serve as an Equalizer

Artificial intelligence has the potential to make human processes and decisions more efficient and less biased.<sup>5</sup>This means it can become an equalizer, **reducing decisions for people—who are naturally subject to their own unconscious biases—and making predictions with algorithms based on data.**<sup>6</sup> Algorithms have already been shown to improve decision-making processes for everything from loan applications to who gets hired for the job to deciding whether a lesion is cancerous.<sup>7</sup> When used with care, AI can expand our abilities and intelligences and **help us step into a positive machine-human future.**

## Examples of AI Improving Human Processes

- Algorithms that identify successful board director candidates more accurately than people do, allowing us to evaluate which characteristics (e.g., being male, having a large network, coming from a finance background, or having current board experience) may be overvalued when nominating board directors.<sup>8</sup>
- AI hiring tools that craft better job descriptions, match candidate skills to job descriptions to **avoid bias and build more diverse slates**, and find candidates who may have been ignored in the traditional recruiting process by searching through applicant tracking systems and career sites.<sup>9</sup>

- Some companies are building AI tools that limit bias by **assessing applicants based on specific data, skills, and abilities**, and are monitoring the AI tools to make sure bias does not creep in.<sup>10</sup>
- One AI tool scans applicant-tracking systems and other career sites to find candidates and remove their names from the process to reduce bias. It also creates a profile for the ideal candidate's attributes based on data, and scores applicants by assessing them against the ideal profile on "job fit, skills match, resume quality, and a percentile score in comparison to other candidates."<sup>11</sup>
- AI tools that automate specific recruitment tasks to remove bias, such as creating interview procedures that improve the prospective employee experience and increase diversity by removing gender bias. For example, some tools use AI and/or virtual reality filters to obscure a candidate's appearance or voice during the interview process, reducing the potential for bias.<sup>12</sup>

## The Bad News: Bias in AI Can Be a Barrier to Inclusion

Although AI has the potential to make decisions more efficient and less biased, it is not truly a "clean slate." **Artificial intelligence is only as good as the data that powers it.** Its quality depends on how well creators programmed it to think, decide, learn, and act. As a result, **artificial intelligence may inherit, or even amplify, the biases of its creators**—who are often unaware of their own biases—or the AI may be using biased data.<sup>13</sup> The consequences of such technology can be life-altering.<sup>14</sup> In workplaces, already existing gaps in hiring and promoting women and people of color can widen if biases are unintentionally written into AI's code, or if the AI learns to discriminate.

## Examples of AI Bias

- An employer was advertising for a job opening in a male-dominated industry via a social media platform. The platform's ad algorithm pushed jobs to only men to maximize returns on the number and quality of applicants.<sup>15</sup>
- A tech company spent years creating an AI hiring tool by feeding it resumes from top candidates. The AI's function was to review candidate resumes and recommend the most promising candidates. Because the industry is male-dominated, the majority of the resumes used to teach the AI were from men, which ultimately led the AI to discriminate against recommending women (e.g., down-scoring resumes that included words like "women" or education from "women's colleges"). Despite multiple unsuccessful attempts to correct the algorithm, the company eventually had to scrap the AI because it could not "unlearn" this bias.<sup>16</sup>
- Face-analysis AI programs display gender and racial bias, demonstrating low errors for determining the gender of lighter-skinned men but high errors in determining gender for darker-skinned women.<sup>17</sup>

- Algorithms used in courtrooms to conduct “risk assessments” about defendants are racially biased. This AI predicts the likelihood that defendants will commit future crimes and its assessments are used throughout the criminal justice process, influencing decisions about the defendant’s bond amounts and sentences. A recent study found that the forecasts were unreliable and skewed to flag Black defendants “at almost twice the rate as white defendants” as likely to commit future crimes; Black defendants were found to not actually re-offend at the rates predicted. The programs made the opposite error for white defendants, giving them lower risk scores; meanwhile, researchers found that they went on to commit future crimes at a higher rate than predicted. These differences could not be explained by prior criminal records, age, or gender.<sup>18</sup>
- Machines taught by image-recognition software and photos learned a sexist view of women, labeling scenes such as cooking as associated with women and sports as associated with men, even when these labels were incorrect. Researchers further tested the data sets and discovered that the AI not only reflected the unconscious gender biases of its creators, but actually *amplified them*.<sup>19</sup>
- Voice-activated technology in cars can help solve distracted driving, but many cars’ systems are tone deaf to women’s voices and have difficulty recognizing foreign accents.<sup>20</sup>

## AI Creators Must Become More Diverse

Artificial intelligence is not truly objective. AI and its algorithms can **reflect the biases of their creators**, and even those that are unbiased at inception can learn the biases<sup>21</sup> of their human trainers over time. It must be programmed, reviewed, monitored, and audited to ensure it is not biased or *becoming* biased based on its algorithms and data.

Adding more women and more diverse workers with technical skills in the AI field is one way to reduce bias—by providing additional perspectives and more fail safes—thus **creating and training AI to more accurately reflect a diverse and inclusive society**. Greater diversity can also reduce groupthink and enhance team decision-making, leveraging a greater variety of perspectives for faster and more thorough decisions.<sup>22</sup> Homogenous AI teams and researchers may not be paying close enough attention to notice when bias has crept in and impacted the AI they’ve created or trained.

**Currently, women are severely underrepresented in the AI field**, with one study<sup>23</sup> finding that only 12% of leading machine-learning researchers were women. It is critically important for the future<sup>24</sup> to get more women<sup>25</sup> and girls<sup>26</sup> into STEM fields.

## Questions to Consider

- What strategies are in place to diversify your talent pool to build a diverse workforce, with employees from different backgrounds, perspectives, worldviews, and languages?
- What programs can you institute to increase **awareness of unconscious bias** and how to combat it in your human workforce and AI systems?
  - How will you ensure that your hiring and talent management AI systems are free from bias?
- What processes are in place to ensure that you **routinely monitor algorithms and machines for bias**? How can you immediately address bias if you discover it creeping into your AI's decisions, actions, or processes?
- What steps are you taking to create an inclusive workplace<sup>27</sup> where people feel safe to speak up? How do you ensure that your culture values **respect and accountability**? It may be easier for you to catch AI bias if the humans interacting with the tools know that they can be critical of the current systems in place without negative repercussions.
- What regulations, ethical considerations, and best practices for vigilance against perpetuating bias can you champion in the AI space?

## AI Consortia, Research Groups, and Start-Ups

**Algorithmic Justice League.** A multi-disciplinary group that works to highlight algorithmic bias across fields, provide space for people to report bias, and develop practices to check for bias in algorithmic software and data.

**Equal AI.** This initiative focuses on identifying and eliminating bias in AI by addressing gender diversity in tech education and hiring, detecting and eliminating bias in AI, and creating systems to remove bias from artificial intelligence.

**Partnership on AI.** This multi-disciplinary organization of stakeholders—including academics, researchers, tech/AI companies, and other groups—was founded to study and enhance our understanding of AI, including best practices, ethics, and impacts.

**Open AI.** A nonprofit organization that researches the path to safe AI.

- 
1. Lee Rainie and Janna Anderson, *Code-Dependent: Pros and Cons of the Algorithm Age* (Pew Research Center, Internet and Technology, February 2017).
  2. Mark Stone, “[Want a More Diverse Workforce? How AI is Combating Unconscious Bias](#),” *Dell Technologies Perspectives*, March 14, 2018.
  3. Executive Office of the US President (Obama Administration) National Science and Technology Council Committee on Technology (NSTC), *Preparing for the Future of Artificial Intelligence* (October 2016).
  4. PwC, *Bot.Me: A Revolutionary Partnership* (PwC Consumer Intelligence Series, 2017); Ron Schmelzer, “[Assisted Intelligence vs. Augmented Intelligence](#),” *Medium*, October 3, 2018;

- Science and Technology, “Assisted, Augmented, and Autonomous: The 3 Flavours of AI Decisions,” *TG Daily*, June 28, 2017.
5. Alex P. Miller, “Want Less-Biased Decisions? Use Algorithms,” *Harvard Business Review*, July 26, 2018.
  6. PwC, *Bot.Me: A Revolutionary Partnership* (PwC Consumer Intelligence Series, 2017).
  7. Mark Stone, “Want a More Diverse Workforce? How AI is Combating Unconscious Bias,” *Dell Technologies Perspectives*, March 14, 2018; Will Byrne, “Now is the Time to Act to End Bias in AI,” *Fast Company*, February 28, 2018; Abby Norman, “Your Future Doctor May Not Be Human. This Is the Rise of AI in Medicine,” *Futurism*, January 31, 2018.
  8. Isil Erel, Léa H. Stern, Chenhao Tan, and Michael S. Weisbach, “Research: Could Machine Learning Help Companies Select Better Board Directors?,” *Harvard Business Review*, April 9, 2018.
  9. Rebecca Greenfield and Riley Griffin, “Can Artificial Intelligence Take the Bias Out of Hiring?,” *Boston Globe*, August 12, 2018; Mark Stone, “Want a More Diverse Workforce? How AI is Combating Unconscious Bias,” *Dell Technologies Perspectives*, March 14, 2018.
  10. Rebecca Greenfield and Riley Griffin, “Can Artificial Intelligence Take the Bias Out of Hiring?,” *Boston Globe*, August 12, 2018.
  11. Mark Stone, “Want a More Diverse Workforce? How AI is Combating Unconscious Bias,” *Dell Technologies Perspectives*, March 14, 2018.
  12. Mark Stone, “Want a More Diverse Workforce? How AI is Combating Unconscious Bias,” *Dell Technologies Perspectives*, March 14, 2018.
  13. Will Byrne, “Now is the Time to Act to End Bias in AI,” *Fast Company*, February 28, 2018; Larry Hardesty, “Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems,” *MIT News*, February 11, 2018.
  14. Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, “Machine Bias,” *ProPublica*, May 23, 2016.
  15. Noam Scheiber, “Facebook Accused of Allowing Bias Against Women in Job Ads,” *The New York Times*, September 18, 2018.
  16. David Meyer, “Amazon Reportedly Killed an AI Recruitment System Because It Couldn’t Stop the Tool From Discriminating Against Women,” *Fortune*, October 10, 2018.
  17. Larry Hardesty, “Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems,” *MIT News*, February 11, 2018.
  18. Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, “Machine Bias,” *ProPublica*, May 23, 2016.
  19. Tom Simonite, “Machines Taught by Photos Learn a Sexist View of Women,” *Wired*, August 21, 2017.
  20. Sharon Silke Carty, “Many Cars Tone Deaf to Women’s Voices,” *Auto Blog*, May 31, 2011.
  21. Kriti Sharma, “Can We Keep Our Bias From Creeping Into AI?,” *Harvard Business Review*, February 9, 2018.
  22. Adam D. Galinsky, Andrew R. Todd, Astrid C. Homan, Katherine W. Phillips, Evan P. Apfelbaum, Stacey J. Sasaki, Jennifer A. Richeson, Jennifer B. Olayon, and William W. Maddux, “Maximizing the Gains and Minimizing the Pains of Diversity: A Policy Perspective,” *Perspectives on Psychological Science*, vol. 10, no. 6 (2015): p. 742-748; Alison Reynolds and David Lewis, “Teams Solve Problems Faster When They’re More Cognitively Diverse,” *Harvard Business Review*, March 30, 2017; Catalyst, *Quick Take: Why Diversity and Inclusion Matter* (August 1, 2018).
  23. Tom Simonite, “AI Is the Future—But Where Are the Women?,” *Wired*, August 17, 2018.
  24. NACE, “AI and Automation: Our Changing World of Work,” *NACE Job Market Trends and Predictions*, August 16, 2017.
  25. Catherine Hill, Christianne Corbett, and Andresse St. Rose, *Why So Few? Women in Science, Technology, Engineering, and Mathematics* (AAUW, 2010).
  26. Kamla Modi, Judy Schoenberg, and Kimberlee Salmond, *Generation STEM: What Girls Say About Science, Technology, Engineering, and Math* (Girls Scouts of America, 2012).
  27. Catalyst, “Inclusion.”