

4 Ways to Mitigate Bias in AI and Close the Diversity Deficit

Feed a prompt to an AI image generator and you're bound to encounter an insidious pattern: Do the people look ... too stunning? Perhaps even wanton?

Gender, race, body type, nationality, religion—you're almost guaranteed to get prejudiced and outdated stereotypes when using these descriptors in prompts. And "wanton" is a deliberate adjective; it's mostly used pejoratively toward women, and AI tends to oversexualize female images. These glaring imbalances showcase a recurring problem with AI outputs: the replication of societal biases, which can be harmful to actual people and communities.

I wrestled with this firsthand while helping develop Sir Martian, one of our key AI demos featured at Cannes earlier this year. Sir Martian, playfully named after Sir Martin Sorrell, is an AI-powered robot in the form of an alien caricaturist. Throughout the festival, he invited attendees to sit down for a quick chat and a sketched portrait, based on their appearance and tastes.

I'm proud that the demo was a success, because as you can imagine, this interaction was more than a simple conversation. And it taught me a lot about the privileges and responsibilities of shaping a new technology. Here's what I learned.

Words matter—your data sets the tone

Most AI tools available for the general public are trained on datasets that aren't accessible or visible to users, so I feel particularly fortunate to work at a company that creates and trains its own models. It really is a "great power, great responsibility" scenario.

The foundation of any generative AI model should be diverse and comprehensive. By expanding the range of base images and training materials, developers can create AI systems that represent a broader spectrum of human experiences. This enriches outputs and helps combat entrenched biases.

With Sir Martian, specificity was essential for aligning user inputs with desired outputs. After some trial and error, we found that we had to train the model combining visual input with very precise text prompts in order to get it to represent people accurately.

When given a picture of a Black woman and the prompt "woman with braids," the AI model automatically defaulted to a woman with German-style braids. We had to train and fine-tune it using specific terms like "cornrows" and "box braids" to get it to create accurate drawings. Giving the system a wider variety of terms to connect to visual references was crucial to getting more diverse depictions.

This step was humbling because I encountered my own limitations in the process. For example, we don't have a large Muslim population where I'm based in Brazil, yet a global audience traveling to Cannes would likely include women in hijabs or chadors. This prompted me to research the nuances between different articles of dress that, to an untrained eye, may have been seen as interchangeable. The experience highlighted the importance of stepping outside of our bubbles to recognize what we don't know, in order to learn and incorporate diverse cultural elements that better serve global

users.

Diversity is (and isn't) everyone's responsibility

As the only woman on the team building Sir Martian, the problematic depiction of women raised alarm bells for me early on but didn't faze my male colleagues until I brought it to their attention. We need more diverse teams who can authentically lead AI in the right direction. But at the same time, the onus shouldn't be on minorities alone to fix biases that have affected them for generations.

Overcoming these biases demands collective effort. After I discovered flaws in Sir Martian's AI model, I partnered closely with a developer on the project who was dedicated to addressing these issues. I reached out to a Black co-worker and Muslim women in our global community for their feedback on whether Sir Martian's drawings were respectfully reflecting their identities. These are just some examples of the cross-disciplinary collaboration that needs to happen in order to make a change; once you flip the switch and understand what needs to be done, the rate of progress is astounding.

The industry has a ways to go, but we're seeing positive change. Since Sir Martian launched, we've instated a global AI policy to help staff become more conscious of common biases that occur in AI systems, such as data bias, algorithmic bias, and confirmation bias. Perhaps more importantly, fostering an inclusive environment encourages a shared responsibility in creating AI systems that accurately and fairly reflect diverse experiences, ultimately benefiting everyone.

Know where to draw the line, and back up your decisions

Our industry celebrates how AI will unlock personalization for everyone, but there are limits. The unfortunate reality is that, when it comes to accurately depicting everyone, we can't

perfectly address every difference on every project. But we can try to be as thorough as possible given the limits of technology, time, and budgets.

When it comes to being more diverse and inclusive, for example, people naturally focus on accounting for a variety of skin tones. That's great, but it's often as far as we go. What about different body types and sizes? How might a generated portrait differ when someone is sitting in a wheelchair instead of standing up?

We should not only address these questions, but also begin asking them at a project's inception. Those of us developing consumer-facing generative AI activations must be conscious of where our parameters fall, as well as able to justify the decisions we make.

When working on Sir Martian for the demo in Cannes, we decided to leave children out of the training data, knowing that they were not our target audience. This was a conscious decision rather than a blind spot in our process, as representation and inclusion so often are in AI projects.

It's time to do better

We all know that AI is an amazing tool that has progressed by leaps and bounds over the last few years, but one thing it can't do is correct our own blind spots. That's on us to identify and address.

AI serves as a mirror to our society, reflecting both its progress and its persistent challenges. If left unchecked, biases can become even more ingrained through AI. Tackling this issue isn't a task for minorities alone—it's something we

all need to work on together. This shared commitment can help genuinely turn AI into a force for positive change.

Recommended articles

<https://www.adweek.com/creativity/4-ways-to-mitigate-bias-in-ai-and-close-the-diversity-deficit/>