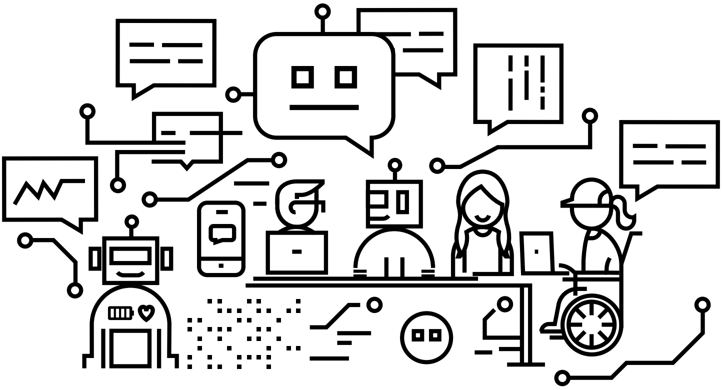


In Pursuit of Inclusive AI

Joyce Chou, Roger Ibars, Oscar Murillo

Table of Contents

Part 1:
Five Ways to Identify Bias

Joyce Chou, Oscar Murillo, Roger Ibars

Recognizing exclusion in AI

At Microsoft, we've developed inclusive design tools and processes to recognize people with physical disabilities in our design process. As we've evolved our practices, we've expanded our design thinking to other areas of exclusion, including cognitive issues, learning style preferences, and social bias.

It's time to take that same approach to AI. Bias in AI will happen unless it's built from the start with inclusion in mind. The most critical step in creating inclusive AI is to recognize where and how bias infects the system.

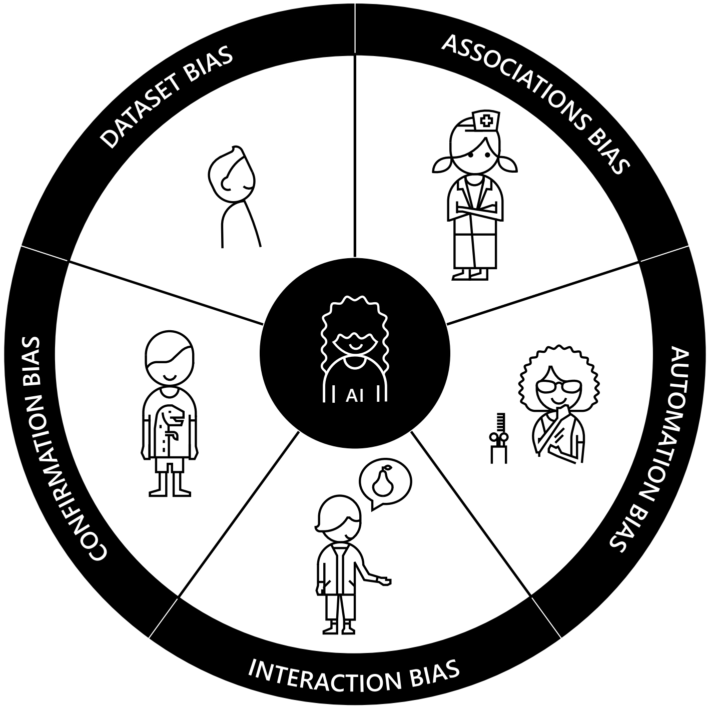
Our first inclusive design principle is recognize exclusion. The guide we're unveiling here breaks down AI bias into distinct categories so product creators can identify issues early on, anticipate future problems, and make better decisions along the way. It allows teams to see clearly where their systems can go wrong, so they can identify bias and build experiences that deliver on the promise of AI for everyone.

Five ways to identify bias

We worked with academic and industry thought leaders to determine five ways to identify bias. Then, we used childhood situations as metaphors to illustrate the behavior in each category. Why? We can all relate to childhood episodes of bias, and it fits into a nice metaphor: AI is in its infancy, and like children, how it grows reflects how we raise and nurture it.

Each bias category includes a childhood metaphor that illustrates it, its definition, a product example, and a stress test for your teams and AI work.

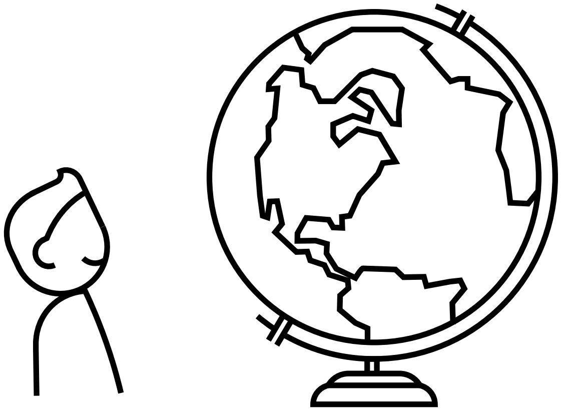
The chart on the right shows how the five biases break down:



1. Dataset Bias

When the data used to train machine learning models doesn't represent the diversity of the customer base. Large-scale datasets are the foundation of AI. At the same time, datasets have often been reduced to generalizations that don't consider a variety of users and therefore underrepresent them.

A young child defines the world purely on the small amount they can see. Eventually, the child learns that most of the world lies beyond the small set of information that's within their field of vision. This is the root of dataset bias: intelligence based on information that's too small or homogenous.



Product example

Machine vision technologies—such as web cameras to track user movements—that only work well for small subsets of users based on race (predominantly white), because the initial training data excluded other races and skin tones.

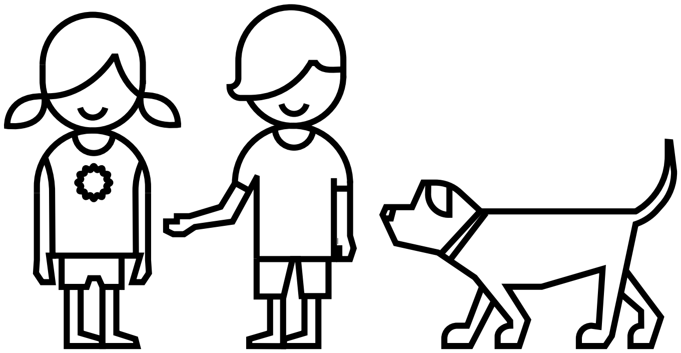
Stress test

If you're using a training dataset, does that sample include everyone in your customer base? And if not, have you tested your results with people who weren't part of your sample? What about the people on your AI teams—are they inclusive, diverse, and sensitive to recognizing bias?

2. Associations Bias

When the data used to train a model reinforces and multiplies a cultural bias. When training AI algorithms, human biases can make their way to machine learning. Perpetuating those biases in future interactions may lead to unfair customer experiences.

Imagine some kids who like to play “doctor.” The boys want the doctor roles and assume the girls will play the nurses. The girls have to make their case to overturn assumptions. “Hey, girls can be doctors too!”



Product example

Language translation tools that make gender assumptions (e.g. pilots are male and flight attendants are female).

Stress test

Are your results making associations that perpetuate stereotypes in gender or ethnicity? What can you do to break undesirable and unfair associations? Is your dataset already classified and labeled?

3. Automation Bias

When automated decisions override social and cultural considerations. Predictive programs may automate goals that go against human diversity. The algorithms aren't accountable to humans, but make decisions with human impact. AI designers and practitioners need to consider the goals of the people affected by the systems they build.

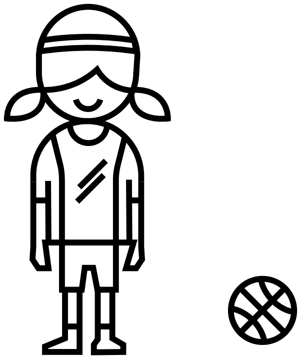
Imagine a girl getting a makeover. The girl likes sports, loves a natural look and hates anything artificial. The beautician has different ideas about beauty, applies tons of makeup and a fussy hairdo. The results make the beautician happy, but horrify the girl.

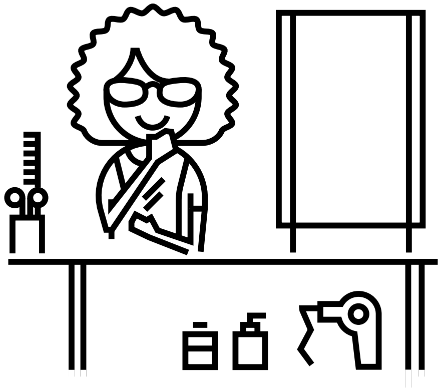
Product example

Beautification photo filters reinforce a European notion of beauty on facial images, like lightening skin tone.

Stress test

Would real, diverse customers agree with your algorithm's conclusions? Is your AI system overruling human decisions and favoring automated decision making? How do you ensure there's a human POV in the loop?

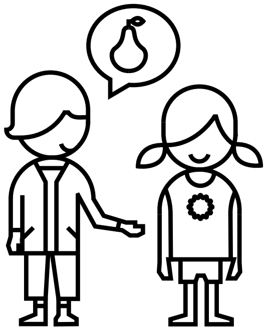




4. Interaction Bias

When humans tamper with AI and create biased results. Today's chatbots can make jokes and fool people into thinking they're human much of the time. But many attempts to humanize artificial intelligence have unintentionally tainted computer programs with toxic human bias. Interaction Bias will appear when bots learn dynamically without safeguards against toxicity.

A popular kids' game is "Telephone." The first person in a group whispers a sentence to next person, who then whispers it to the next person—and so on until the last person says what they heard. The point is to see how the information changes naturally through so many hand-offs. But say one kid changes it intentionally to create a more ridiculous result. It may be funnier, but the spirit of seeing what happens naturally is broken.

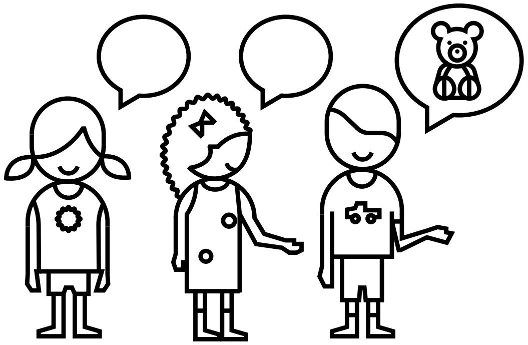


Product example

Humans deliberately input racist or sexist language into a chatbot to train it to say offensive things.

Stress test

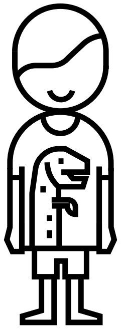
Do you have checks in place to identify malicious intent toward your system? What does your AI system learn from people? Did you design for real-time interaction and learning? What does that mean for what it reflects back to customers?



5. Confirmation Bias

When oversimplified personalization makes biased assumptions for a group or an individual. Confirmation Bias interprets information in a way that confirms preconceptions. AI algorithms serve up content that matches what other people have already chosen. This excludes results from people who made less popular choices. A knowledge worker who is only getting information from the people who think like her will never see contrasting points of view and will be blocked from seeing alternatives and diverse ideas.

Think of the kid who gets a toy dinosaur for a present one year. Other family members see the dinosaur and give him more dinosaurs. In several years, friends and family assume the kid is a dinosaur fanatic, and keep giving more dinosaurs until he has a huge collection.

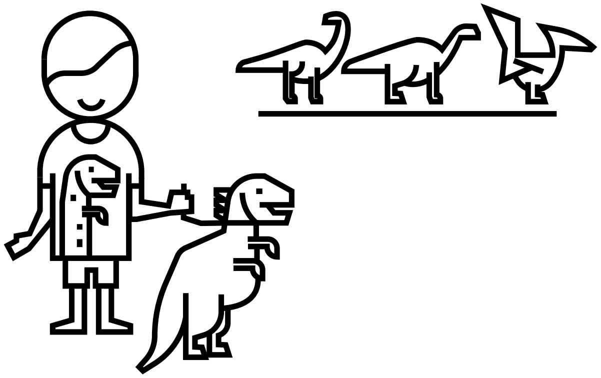


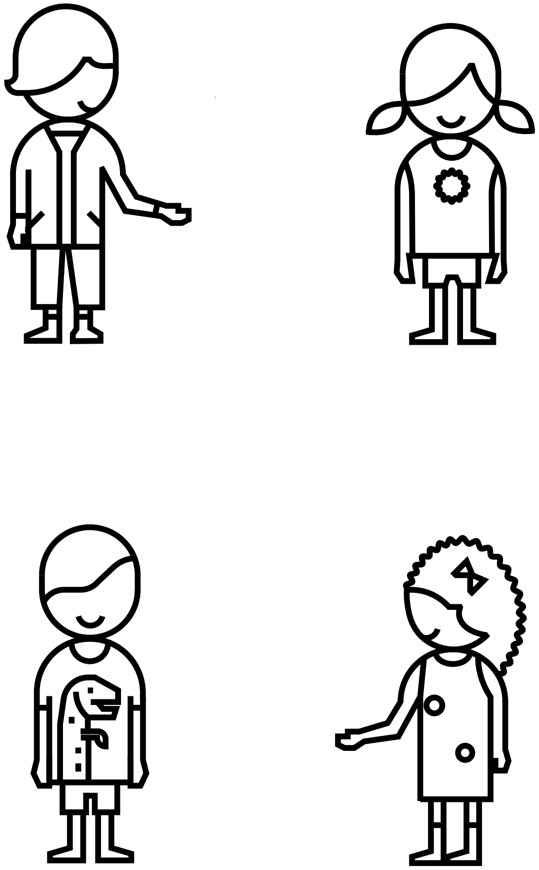
Product example

Shopping sites that show recommendations for things the customer has already bought.

Stress test

Does your algorithm build on and reinforce only popular preferences? Is your AI system able to evolve dynamically as your customers changes over time? Is your AI system helping your customers to have a more diverse and inclusive view of the world?





Using this primer

As designers and creators of artificial intelligence experiences, it's on us to be thoughtful about how AI evolves and how it impacts real people. This primer is the start of a long road to create experiences that serve everyone equally.

If we apply these ideas to our initial example of the African-American girl misread by the facial recognition software, we can label that as Dataset Bias: the software was trained with data that was too narrow. By recognizing and understanding those biases from the start, we can test the system against other human considerations, and build more inclusive experiences. Could our facial recognition software be subject to deliberately erroneous data? What other biases could infect the experience?

Most people working in AI have anecdotal evidence of situations like these. Embarrassing, offensive outcomes from unintentional bias that we all want to identify and avoid. Our goal here is to help you recognize the underlying bias that leads to these situations. Start with these categories and test your experience with these types of bias in mind, so you can focus on delivering the potential of AI to all your customers.

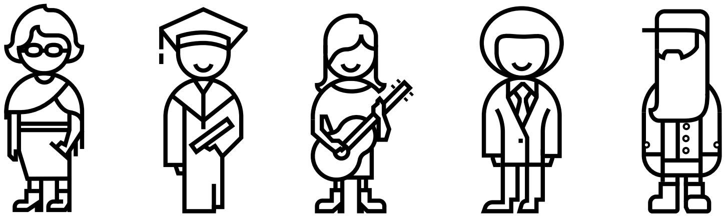
Part 2:
Insights for Inclusive AI

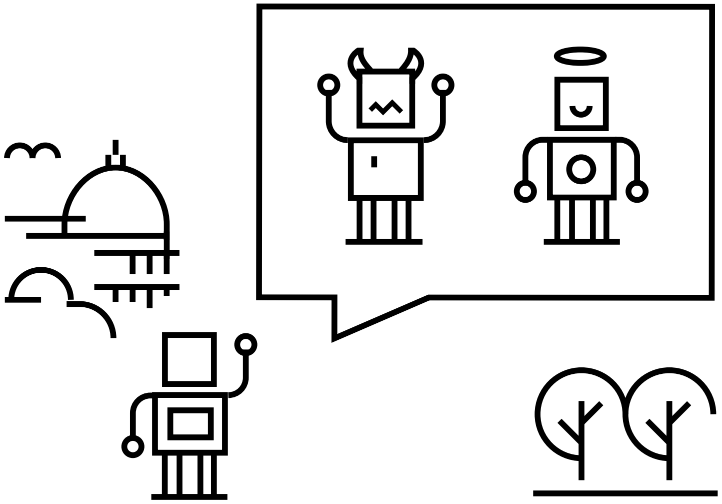
Joyce Chou & Roger Ibars

Today, machines often learn from signals and examples; sophisticated programming similar to what we do as humans. We recognize a familiar face at a party; we ride a bike after years of absence. Machines can acquire a kind of tacit knowledge yet still lack human nuance.

It's that trick of nuance that sparked conversations and research around inclusive AI at Microsoft. The field of AI is growing rapidly, and we're all learning in real time as we experiment with solutions and grapple with unexpected outcomes. As part of that shift, the Inclusive Design team has partnered with research, engineering and legal groups across Microsoft to bridge the gap between high-level principles and every day practice.

In the spirit of shared knowledge, we've summarized five insights to identify exclusion and design more inclusive AI.

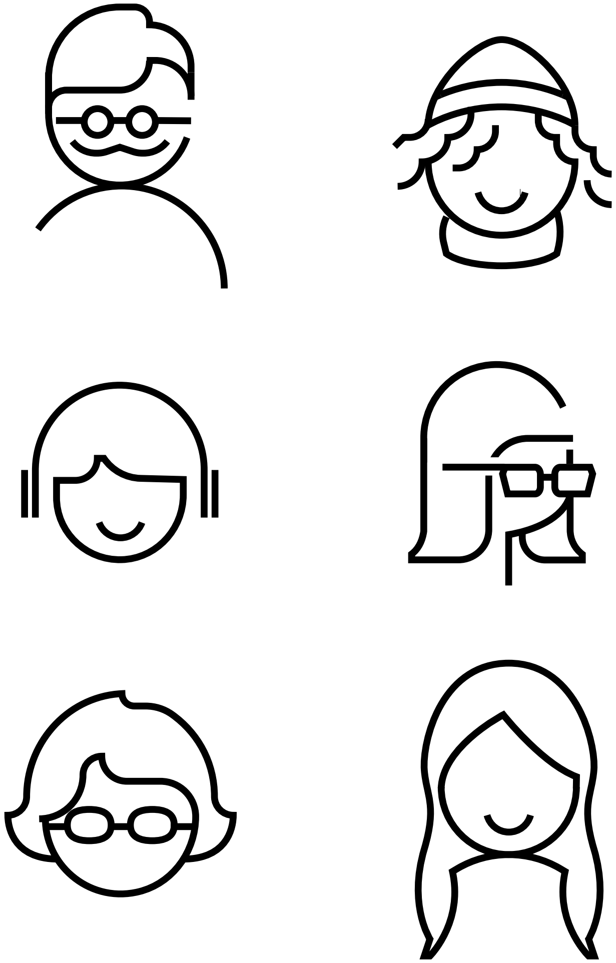




1. Redefine bias as a spectrum

Conversations about AI are often polarized between “good vs. evil”, but teams building AI products and services have a difficult time relating to the most offensive examples of bias that grab media attention. Rather than focusing on the most extreme cases, we learned that teams engage faster with AI bias issues when they consider a spectrum of bias, where bias can show up in small ways in our everyday experiences.

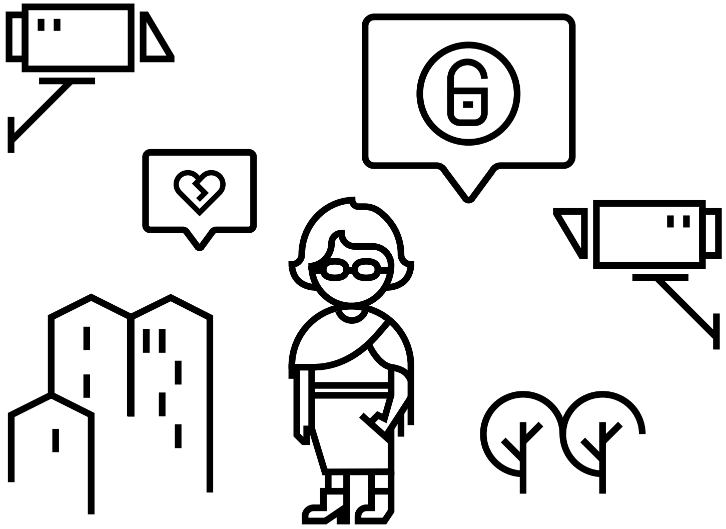
We’ve learned that AI needs to be overseen to avoid major pitfalls and address subtle and seemingly mundane microaggressions. Think about a time when you felt like a product was not made for you, but you couldn’t explain *why*. These tiny uncomfortable moments build up over time, cause feelings of exclusion, or simply makes the product feel *off*. Products with small moments of bias are not good products. AI bias isn’t the end of the world, but rather the early stages of good intent gone sideways. It’s our responsibility to recognize the understated risks and design accordingly.



2. Enlist customers to correct bias

Training is everything when it comes to building more inclusive AI. Unfortunately, development for AI often happens behind closed doors, restricted to input from teams that may not be representative of the diverse customers they design for. It’s a humble exercise for teams to reflect not only on the applications of the AI they build, but to consider the *implications* of the technology in the wild. We’ve learned that empowering people to continually train AI will develop more inclusive intelligence and ultimately build trust.

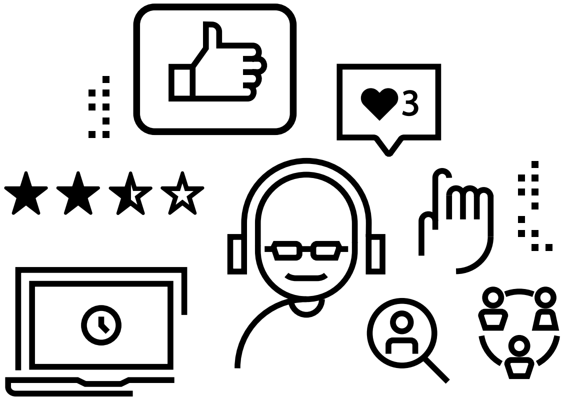
Just in the past year, we’ve seen an uptick in conversations surrounding ethics in AI design, including countless articles exploring issues of transparency, accountability, agency, and more. These concepts have led to large-scale open-source projects like synthetic simulators to improve self-driving cars or a crowd-sourced initiative to train speech models in a more natural voice. Gaining these customer insights earlier in a safe training environment can curb the potential for unintentional or alarming outcomes.



3. Cultivate diversity with privacy and consent

The common conceit of AI is that it grows smarter over time. Yes, a machine will improve its understanding as it learns from the data it's fed. But inclusive AI also depends on those datasets being more diverse, correctly labeled, and used in a way that's representative of every customer. If there's any bias in the system (and there's always bias in the system), it only exacerbates that bias. For underrepresented people, there's little incentive to participate in something that's broken for them, especially if they think that information they provide could be used against them. And without their data, the cycle of learned bias in AI continues.

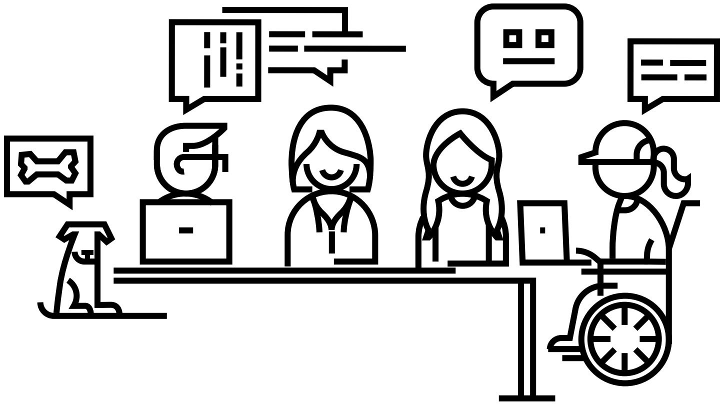
There's a basic understanding that we give up some vestiges of privacy for the conveniences of the modern world, but often privacy controls are poorly designed and convoluted for everyday customers. It's difficult to feel in control. The adoption of the General Data Protection Regulation (GDPR) has bolstered improvements in the industry, but privacy-by-design needs to be foundational, not reactive. Rather than user agreements full of inaccessible legalese, we need touchpoints for consent all along their journey, design that values autonomy foremost.



4. Balance intelligence with discovery

AI makes a lot of assumptions based on our past behaviors, and often there's little flexibility to understand our present intentions. And we're uncomfortable when those assumptions create echo chambers, digital bubbles, broken record suggestions and irrelevant content. There is an inherent tension between machine intelligence and the human desire to create and explore in new ways. There needs to be strategic moments that build a more natural relationship between humans and technology, valuing patience and creative exploration.

Customers should always feel like they have the option to change course and shape the goal of their experiences. We've learned that this doesn't always feel like the case because we don't have a clear understanding of what we can expect from AI. We're lead to believe that AI is intelligent out-of-the-box, but this narrative needs to change. If customers know AI services are limited in the beginning and still need help to learn, maybe they'll be more willing to help train AI with their unique idiosyncrasies.



5. Build inclusive AI teams

AI reflects the people who build it, as much as we might want to believe in its neutrality. Hiring diverse backgrounds, disciplines, genders, races, and cultures into the teams designing and engineering these experiences is critical. “Artificial intelligence will reflect the value of its creators,” says Kate Crawford of the AI Now Research Institute. “So inclusivity matters—from who designs it to who sits on the company boards and which ethical perspectives are included. Otherwise, we risk constructing machine intelligence that mirrors a narrow and privileged vision of society, with its old, familiar biases and stereotypes.”

We’ve learned that teams with diverse outlooks can identify biases more easily. By building inclusive teams, we can foster empathy and begin training AI to do the same. Teams must be open-minded, take accountability for unintended mistakes, and approach public dialogue with humility. They need to be thoughtful and deliberate, ever mindful of the bias inherent in their designs.

Humanity-centered design

There's no magic formula for all scenarios, nor should there be. If we're attempting to build AI that really helps and understands us, we need to come at it from a human perspective. We can't place implicit trust in future machines just because we're involved in building them—humans are complex and full of doubt. It's okay to recognize that, and to fail gracefully. In those moments, we can slow down, and consider *why* we're moving in a certain direction, invite more people into the creation process, and keep improving.

Human nature is flawed. But it's also wonderful—unparalleled in its complexity. We feel compelled to connect, engage, fix problems, challenge our perspectives, and always move forward. Let's mirror our best intentions and work together for better AI outcomes by design.

Acknowledgements

Learn more about how to design with inclusive in mind by downloading our Inclusive Design tool kit today.

Contributors to this article included ***Danielle McClune, Doug Kim, Elena Dvorkina, Izzy Beltran, and Tiffany Chen***. The Inclusive Design team is grateful to all of our Microsoft partners at Microsoft Research and Insight for thought—provoking discussions and invaluable support in developing inclusive technology.

To stay in-the-know with what's new at Microsoft Design, check out our new website, or follow us on ***Twitter*** and ***Facebook***. And if you are interested in joining our team at Microsoft, head over to: ***careers.microsoft.com***.

Microsoft

www.microsoft.com/design/inclusive