Welcome to DeBiasMe

About This Workbook

You probably use **generative AI (Artificial Intelligence)** tools that are capable of creating content almost every day. Maybe for homework, creative projects, or just for fun. These tools are incredibly powerful, but here's what many people don't realize:

When your problematic thinking patterns meet AI's built-in limitations, they multiply into much bigger problems.

Think of it like this: Imagine you're slightly nearsighted and you put on someone else's glasses that are also slightly wrong. Together, they make your vision way worse than either problem alone. That's what happens when **human biases** meet **AI biases**.

What You'll Learn

By the end of this workbook, you'll be able to:

- Recognize three types of bias that affect every AI interaction
- See how these biases link together and make each other worse
- Use specific strategies to stop bias amplification
- Develop critical thinking habits for your AI use

The Science Behind This: Why Bias Gets Worse

Research shows that when you interact with AI, your biases and AI's biases don't just add up. They make each other worse through a human-AI-system bias loop. One study found that this loop can make bias three times worse in just three conversations. You're going to learn to break the loop!

1. **Input: What You Ask (Human Bias)**
   You bring your assumptions that affect how you ask questions
2. **Processing: How AI Thinks (Statistical and Computational Bias)**
   The AI's training data and algorithms introduce their own biases
3. **Output: What You Get (Systemic Bias)**
   Historical inequities and social patterns get baked into the results

Activity 1: Bias Detection

Learn to identify the bias type in AI interactions

Read each real-world scenario and identify whether the primary bias is **Human Bias** (what the person brings), **AI Bias** (what the AI adds), or **Systemic Bias** (what society has baked in). Think about why other biases might also be present. Remember: biases often overlap and reinforce each other!

Scenario 1: Essay Homework

A student is writing an essay about social media's impact on teens. She starts by asking AI: "Why is social media harmful for teenagers?" When the AI mentions that "studies show 70% of teens experience anxiety from social media," she doesn't question which studies, when they were done, or if there's opposing research. She copies these AI answers into her essay without researching any benefits or checking if the studies are recent.

Human Bias

The student shows **confirmation bias** (looking for information that matched their belief), **automation bias** (trusting the AI completely without questioning it), and **anchoring bias** (stuck with the first answer and didn't explore other viewpoints).

Scenario 2: Job Portraits

A career teacher asks an AI to create an image of a successful CEO. The AI keeps showing pictures of middle-aged white men in suits. When asked for a nurse, it mostly shows images of young women. Even when the teacher asks for diverse CEOs, the AI still struggles to move past common stereotypes.

AI Bias

The AI's training data with limited diversity can overrepresent certain cultures and demographics, creating **statistical biases** (some groups appear less in the training data) in the created images.

Scenario 3: Health Information

A student asks about ADHD symptoms. The AI gives information based mostly on studies of male patients, since historically, medical research didn't include many women. The AI's answer misses important differences in how ADHD shows up in different people.

Systemic Bias

The AI learned from medical data that already had **historical/institutional bias** (past research focused on certain groups more than others).

How the Bias Loop Works in Different AI Tools

Here's the loop: Human biases shape what we ask → **AI biases** affect what it tells us → **Systemic biases** in the data influence everything → This reinforces our human biases when we accept the results. **Every AI tool has biases.** Not because they're trying to be unfair, but because they learned from imperfect human data!

Different AI and tasks can introduce unique bias patterns.
If you know where each AI tends to mess up, you can catch it!

Text Generators (ChatGPT, Claude, Gemini)	**Writing, answering questions, explaining things** Favors certain ways of speaking, assumes everyone shares similar backgrounds
Image Generators (DALL-E, Midjourney)	**Creating pictures from descriptions** Visual stereotypes (gender, race, profession), prefers certain styles over others
Audio/Music (ElevenLabs, Suno, MusicGen)	**Creating voices and music** Limited accent/dialect representation, narrow emotional range
Video Generators (Runway, Pika, Synthesia)	**Making moving images and videos** Limited representation in movement styles
Code Generators (Copilot, Cursor)	**Helping you write computer code** Prefers certain coding/problem-solving patterns
Data Analysis	Statistical assumptions, visualization defaults

Activity 2: Map Your Bias

Map & Find Bias Patterns in Your AI Use

Think of a recent time you used AI. Map how biases are connected through the three stages of interaction. As a class, we'll compare and combine individual maps to see patterns and develop bias mitigation strategies. Can they see biases you missed? Do you see theirs?

What tool was it? _____

What were you trying to do? _____

INPUT: HUMAN BIAS

What was your strongest bias ?
What happens if you use AI without checking?

01. How I influenced the AI interaction:

- ☐ **Confirmation Bias**: I prompted until I got my expected answer
- ☐ **Automation Bias**: I wanted AI to do the thinking for me
- ☐ **Anchoring Bias**: I couldn't move past the first AI answer
- ☐ **Availability Bias**: I only asked about things I already knew about
- ☐ Other: _____

PROCESSING: AI BIAS

02. Patterns I noticed in AI's response:

- ☐ **Statistical Bias:** Overgeneralized from limited data
- ☐ **Representation Bias**: Missing diverse perspectives
- ☐ **Temporal Bias:** Outdated or time-bound assumptions
- ☐ **Cultural/Language Default:** Western/English-centric output
- ☐ Other: _____

How did you notice it?
What happens if you keep trusting biased AI?

OUTPUT: SYSTEMIC BIAS

03. Systemic biases reinforced in the final output:

- ☐ **Historical Bias:** Past inequities treated as normal
- ☐ **Institutional Bias:** Favoring established power structures
- ☐ **Gender/Identity Bias**: Stereotypical assumptions
- ☐ **Geographic Bias:** Western/urban viewpoints only
- ☐ Other: _____

What surprised you about this pattern?
What happens if everyone ignores these biases?

Activity 3: Breaking the Bias Loop

Building Solutions Together

Now that you've mapped your bias pattern, let's break it. These three strategies will target the points where biases connect and amplify. On the collective map, we will annotate the action plan together!

How to Spot and Avoid AI Bias: A Step-by-Step Guide

01. Before You Use AI — Human Bias

- **Check yourself first:** What do you already believe about this topic? Could I find this myself?
- **Think about different viewpoints:** Who else might see this differently? What perspectives might you be missing?
- **Plan to fact-check:** Decide now that you'll verify important claims, not just trust them.

Before I use AI, I will:

02. While You're Using AI — AI Bias

- **Ask better questions:** Instead of "Why is X bad?" try "What are the pros and cons of X?"
- **Get multiple perspectives:** Ask the same question in different ways.
- **Compare different AIs:** ChatGPT, Claude, and Gemini might give you different answers - that's actually helpful!

While using AI, I will:

03. After AI Gives You an Answer — Systemic Bias

- **Fact-check the evidence:** If AI mentions statistics or studies, look them up yourself.
- **Notice what's missing:** Did the AI skip certain groups of people or viewpoints? What didn't it mention?
- **Combine AI with other sources:** Use AI as a starting point, then supplement it with information from books, articles, and expert opinions. **Think of AI like Wikipedia. A great place to start learning, but never your only source!**

After getting AI answers, I will:

More Bias Mitigation Strategies

AI will never be 100% bias-free, just like humans aren't. But that doesn't mean we give up! Every bias we catch and fix can make AI safer and more useful for everyone.

Check Your Sources

- **Ask AI where it got its info:** Request sources and actually check if they're real and reliable
- **Compare different answers:** Try rewording your question or asking different AIs to see if answers change
- **Test with different examples:** If asking about people or groups, try various examples to spot patterns

Protect Your Data

- **Think before you share:** What personal info is the AI learning from your questions?
- **Keep personal details out:** Don't include names, addresses, or private info in your prompts
- **Know how AI stores data:** Understand if your conversations are saved or used for training
- **Use privacy settings:** When available, opt out of data collection

Know Your Rights

- **You can question AI answers:** It's okay to disagree or ask for clarification
- **Report problems:** If AI gives biased or harmful responses, report it to developers
- **Push for better AI:** Support diverse teams making AI and demand fair training data
- **Understand AI limits:** Know that current AI has flaws and that's normal

Remember: You're not just a user - you're part of making AI better. Every time you question bias, check facts, or report problems, you're helping create fairer AI for everyone!

You Did It!

You now have the tools to use AI more critically and responsibly.
Remember: Every bias you catch doesn't just help you. It helps create better AI for everyone.

Your Next Steps:

1. Practice these strategies with every AI interaction
2. Share what you've learned with others
3. Keep questioning, keep learning, keep growing with AI!