

Where Bias Enters the Machine: An Introduction to Race and Gender in AI

Cristina Monzer

Teachers Institute of Philadelphia, a program of the
University of Pennsylvania
Fall 2026 workshop



Annenberg
SCHOOL FOR COMMUNICATION
UNIVERSITY *of* PENNSYLVANIA

Who is teaching this

Cristina Monzer, Postdoctoral Fellow in the Politics, Identities, and Communication Lab, Annenberg School for Communication. I work on AI from two perspectives that shape my input for this workshop.

As an object of study. I use computational methods to study communication on digital platforms — currently, political expression on TikTok, how audiences receive AI-related content there, and how race and gender are represented in AI-generated videos, including #aigirlfriend and #aiboyfriend discourses on TikTok.

As a method. I also use large language and vision models as research instruments to analyze text and videos from social media. That means working hands-on with these models —how they perform, how to use them for the most useful tasks, and—most importantly for teaching—where they fail.

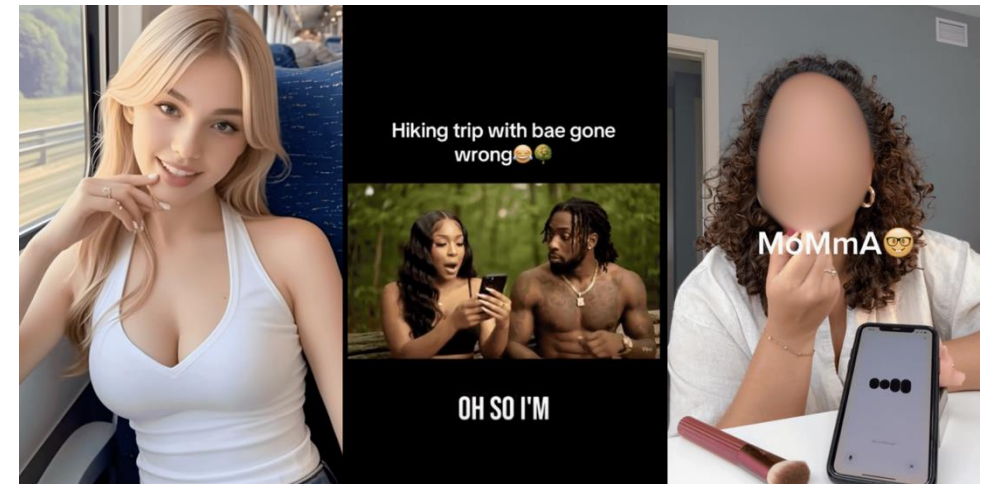


AI tools appear to reproduce common knowledge

Ask an AI tool for an image of a doctor, and it will likely return a man. Ask for a nurse, and it returns a woman. We treat these tools as neutral in part because their output appears to reproduce common knowledge — the kind of result one would expect to see. While some forms of bias are easy to spot, others will be difficult to recognize.

AI tools have infiltrated most areas of ordinary life. A May 2026 survey by Elon University's Imagining the Digital Future Center found that **27% of internet-using U.S. adults** have personal, emotional, or social exchanges with AI chatbots; even more so, close to **40% among adults under 50**.

Students of all ages use these tools, mostly outside class and mostly unsupervised. Asking them not to is futile. Teaching them how the tools work, what they are genuinely useful for, and where they break down sets them up for successful interaction with new technology.



AI-generated content in the #aigirlfriend and #aiboyfriend discourses on TikTok, from my current project.

Sources of bias in using AI

- 01/** Training data — who is over- and under-represented
- 02/** The task — what we ask these tools to do
- 03/** The prompt — how the request is worded
- 04/** Checking output — whether we validate what returns



Session 1 — Where bias enters, and what it does. We open with a plain-language sketch of how these systems work, then we discuss the four sources of bias. Afterward, we turn to the downstream effects: homogenized language, idealized racial and gender identities, and the stereotypes students absorb without noticing.

Session 2 — How the tools actually work. A short history and range of the tools — where they came from and what they can do. Then, the limits of AI detectors, and the extent to which students use and are exposed to AI tools outside of class.

Session 3 — Your curriculum unit. A short synthesis of the first two sessions, then each teacher presents the curriculum unit they built for their own classroom for discussion and feedback.

Throughout. Interactive and discussion-based, accompanied by relevant readings from academic journals, journalism, and technical writing.



What you leave with

Your own
curriculum unit,
ready to teach

Examples that
hold up in class

Discussions with
colleagues from
other subjects

A better answer
than “don’t use it”



QUESTIONS?

CRISTINA MONZER

Postdoctoral Fellow

Annenberg School for Communication

cristina.monzer@asc.upenn.edu

www.cmonzer.com



Annenberg
SCHOOL FOR COMMUNICATION
UNIVERSITY *of* PENNSYLVANIA