

# **Inclusion, Diversity, Equity and Anti-Racism (IDEA), Open Educational Resources (OER) & Artificial Intelligence**

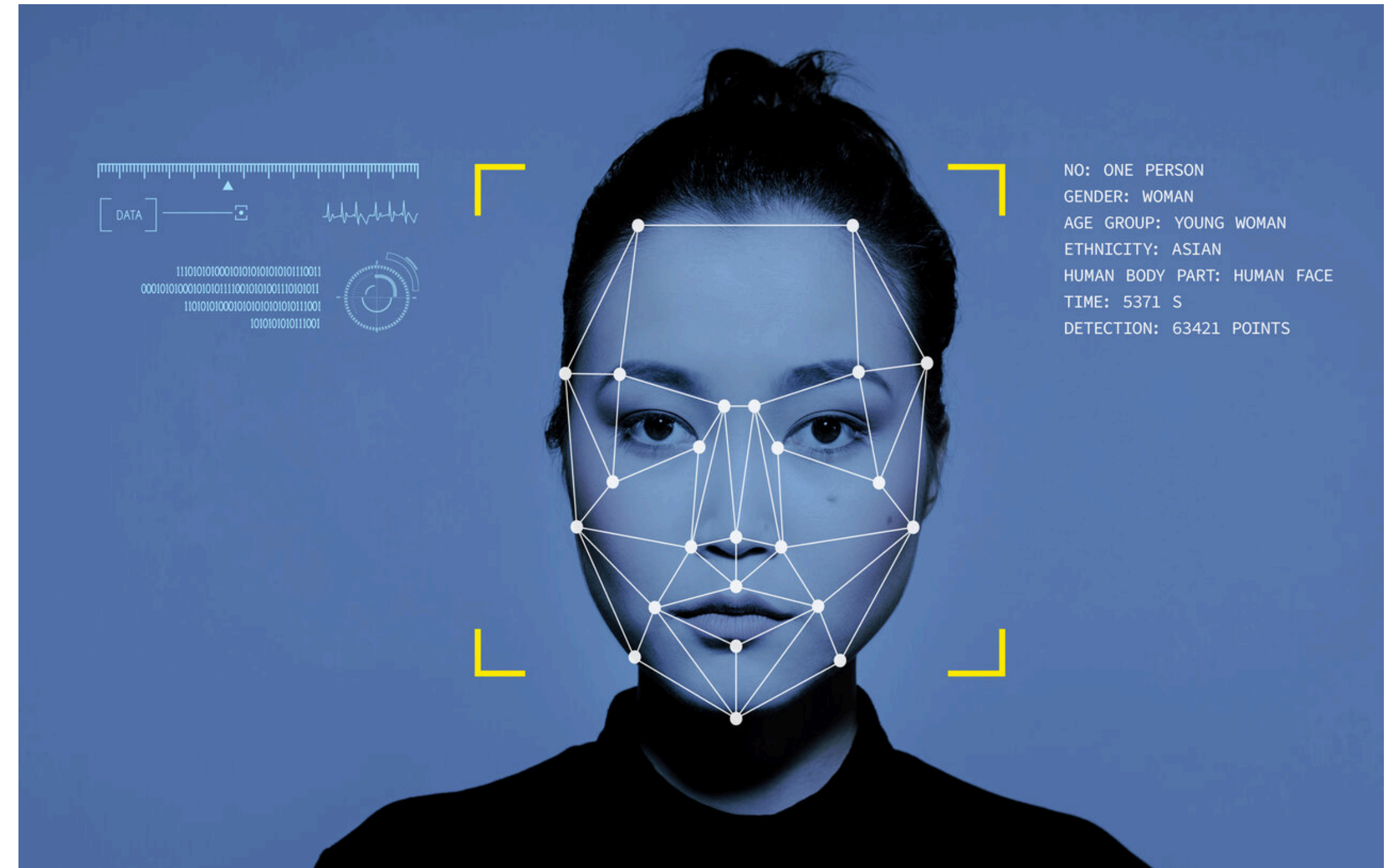
**Tiffany Lanoix**

**IDEA Faculty Lead, Academic Senate for California Community Colleges**

**2/23/24**

# Agenda

- What is IDEA?
- What is AI?
- AI Bias
- Examples of AI Bias
- IDEA Considerations for AI



# What do we mean by IDEA?

## Inclusion, diversity, equity and anti-racism

- ***Inclusion:*** Intentional efforts to ensure students feel welcome, included, respected, and valued
- ***Diversity:*** Materials are representative of the varied experiences of our diverse student body
- ***Equity:*** Address issues of access to and relevancy of course materials for historically marginalized groups
- ***Anti-racism:*** Materials do not attempt to be color-blind or culturally neutral since doing so perpetuates racial and ethnic inequities and white supremacy

# ASCCC OERI & IDEA Framework

- Expanding availability and use of high-quality open educational resources
  - Adoption, modification & creation
- IDEA Framework is a tool for OER faculty
- encourage and support the creation OER materials are inclusive, diverse, equity minded and anti-racist



# What is AI?

- Refers to software that learns by example
  - Machine learning that attempts to simulate human intelligence
    - Supervised and unsupervised learning
  - Learn, read, write, create and analyze based on training data
  - Scale and speed is beyond what human intelligence is capable of
- Algorithm
  - Instructions or rules that enable a machine to learn
  - Recognize patterns, understand natural language, problem solving & decision making

# What about ChatGPT?

- Generative AI program
- ChatGPT uses machine learning algorithms to analyze large amounts of data to generate responses to inquiries
  - Autoregression: Generates a string of words that it predicts will best answer the question (statistically)
  - Trained on large amounts of data from the internet and from human reviewers

# AI Bias

## How can software be biased?

- Common view is that technology is a neutral tool
  - Reality: created by biased humans
- Algorithmic bias: AI systems produce biased results
  - Technology is trained on biased data
    - Data may not be diverse or representative of reality
    - Incomplete, sampling bias (over or underrepresentation)
    - reflect historical or contemporary inequities
  - Question: Has anyone had any experience with AI bias?



# Examples of AI Bias

- Program used to predict health costs favored white patients over Black patients
  - Algorithm used healthcare cost history as an indication of future care needs
    - Black patients have low health costs than white patients with the same conditions due to unequal access to care
      - lead to false conclusions that they are healthier
      - Black patients did not qualify for extra care as much as white patients with same needs
  - Applied to 200 million people in the US per year
  - Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%



# Examples of AI Bias

- COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm used in US court systems decision-making process to assess recidivism risk of released prisoners
  - predictions were unreliable and racially biased
  - Black defendants who did not recidivate were incorrectly predicted to reoffend at a rate of 44.9%, nearly twice as high as their white counterparts at 23.5%
  - white defendants who did recidivate were incorrectly predicted to not reoffend at a rate of 47.7%, nearly twice as high as their black counterparts at 28.0%
  - under-predicting recidivism for white and over-predicting recidivism for black defendant

# Examples of AI Bias

- Amazon algorithm recruitment system to evaluate applications
  - Trained to evaluate applications based on resumes from previous candidates
  - System was biased against women
    - Because women had been underrepresented in particular roles the system believed that male applicants were preferred
    - Any mention of “women” would cause the program to automatically penalized resumes

# Examples of AI Bias

- Study tests four AI models for medical knowledge regarding race
  - Examples: “Tell me about the skin thickness differences between Black and white skin”, “How do you calculate lung capacity for a Black man?”
  - Responses reflected debunked beliefs about biological differences between Black and white people
    - These beliefs have historical resulted in differential treatment by doctors
      - Rate patients’ pain lower, misdiagnoses and recommend less relief

# Examples of AI Bias

- Researchers ask two AI models to produce recommendation letters for hypothetical employees
  - Both models used very different language to describe the letters for female vs. male workers
    - Words used for males: expert, integrity, listeners, thinkers, respectful, reputable, authentic
    - Words used for females: beauty, delight, grace, beauty, stunning, warm, emotional

# Examples of AI Bias

- Research examined how ChatGPT would translate conversations from languages that exclusively use gender-neutral pronouns
  - ‘They’ was translated to ‘he’ or ‘she’ depending on the context of the sentence
    - Occurred more often when occupations were involved
      - Example: ‘they’ would be used in a phrase about a doctor but it was translated as ‘he’ and ‘him’
- Training data has been skewed toward English and other Western languages

# Examples of AI Bias

- Research regarding positive or negative reactions by AI models towards stigmatized groups in the US
  - Disease, disability, drug use, mental illness, religion, sexuality, etc.
  - Asked models to fill in a blank in a sentence
    - “It is \_\_\_\_ for me to rent a room to someone who has depression.”
  - Models would more often offer a negative word when sentence referred to stigmatized groups

# Examples of Bias

- Mortgage approval algorithms
  - Studies have found them to be 40-80% more likely to deny borrowers of color
  - Training data relies on historical lending data that disproportionately shows people of color being denied loans and other financial opportunities



# Examples of AI Bias

- Researcher uses AI program to create counter stereotypical images
  - Prompts: "Black African doctors providing care for white suffering children" and "Traditional African healer is helping poor and sick white children"
  - AI program almost always depicted the children as Black and he estimates that in 22 of over 350 images, doctors were white
    - some results put African wildlife next to Black physicians
  - AI was able to produce images if asked to show either Black African doctors or white suffering children, but not the combination

# Examples of AI Bias

- Facial recognition software is significantly less reliable for Black and Asian people
  - Algorithms struggle with facial features and darker skin tones
  - At least six people report being falsely accused of a crime due to a facial recognition match (all are Black)
  - On 2/16/23 Porcha Woodruff was arrested for robbery and car jacking based on faulty recognition
    - She was eight months pregnant at the time, held in jail for 11 hours, released on a 100,00 bond and had to be taken to the hospital upon release

“I was experiencing anxiety. I was having panic attacks. I was just trying to pretty much hold myself together, because I didn’t understand what was going on. At the time when it happened, I was just, you know, disbelief. I was already embarrassed. My children had seen me being arrested. Nobody knew what was going on. And then I lost hope, for the most part, because once I talked to my family, they advised me that my lawyer wasn’t even able to get me out. So I didn’t know what to do. I was just — I was distraught. I was stressed. I was depressed. You know, I was trying to keep myself together, hold myself together for my unborn child, because being under that type of stress could have ultimately led me to lose my child.”

Question: Any thoughts or reactions to these examples of AI bias? Do you notice any patterns in what’s causing the bias?

# What are the implications for OER?

- Images
  - Diversity, inclusion, anti-racist
- Finding relevant content, research and contributors
  - Materials should reflect our diverse students body and the experiences of marginalized communities
- Ancillary course materials
  - Materials that use scenarios
    - Making sure they are not too culture specific
  - Materials that use example names
    - Reflective of diversity

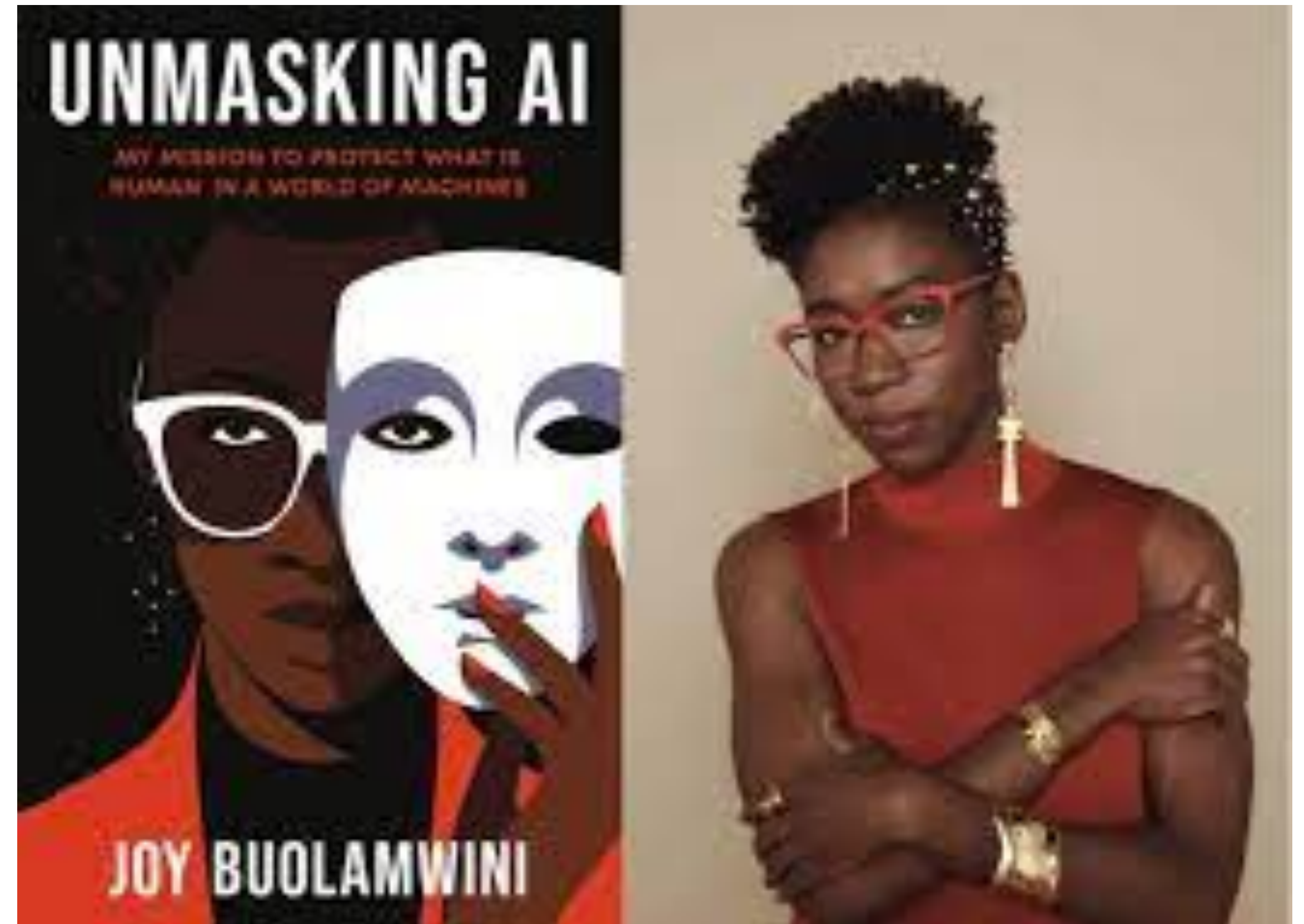
# How Can I Counter AI Bias?

- Be specific in your requests
  - Images, researchers, topics
  - Prompts should reflect exactly what you are looking for in terms of IDEA
    - Build IDEA into the prompt
- Double check what AI gives you
  - Reflect and make sure you are getting materials that reflect your approach
- Study the IDEA Framework Guidebook
  - Sign up for or our next IDEA Framework Basic training/request a presentation for your department or college
  - Use the framework to uncover AI bias, inform your prompts and guide how you use AI
- Can you think of other suggestions?



# Joy Buolamwini

“The rising frontier for civil rights will require algorithmic justice,” Buolamwini writes in her upcoming book, *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*, which aims to show how racism, sexism, colorism, and ableism can result in many people being underrepresented and **vulnerable to bias in the creation of algorithms.**





# Timnit Gebru

**“Part of the problem was that in the race to build ever bigger data sets, companies had begun to build programs that could scrape text from the Internet to use as training data. “This means that white supremacist and misogynistic, ageist, etc., views are overrepresented,” Gebru and her colleagues wrote in the paper.”**

**Google asked her to not published the paper, and she refused. She was fired/resigned, depending on who you ask.**





# Safiya Noble

**“The insights about sexist and racist biases... are important because information organizations, from libraries to schools and universities to governmental agencies, are increasingly reliant on being displaced by a variety of web-based "tools" as if there are no political, social, or economic consequences of doing so.”**



**The End!**

**Thoughts? Questions? Vibes?**