AI: Ethics

Angelly Cabrera, Ash Lopez
University of California, Riverside

Angelly Cabrera
University of Southern California
Electrical and Computer Engineering

Ash López
Santa Monica College
Computational Neuroscience

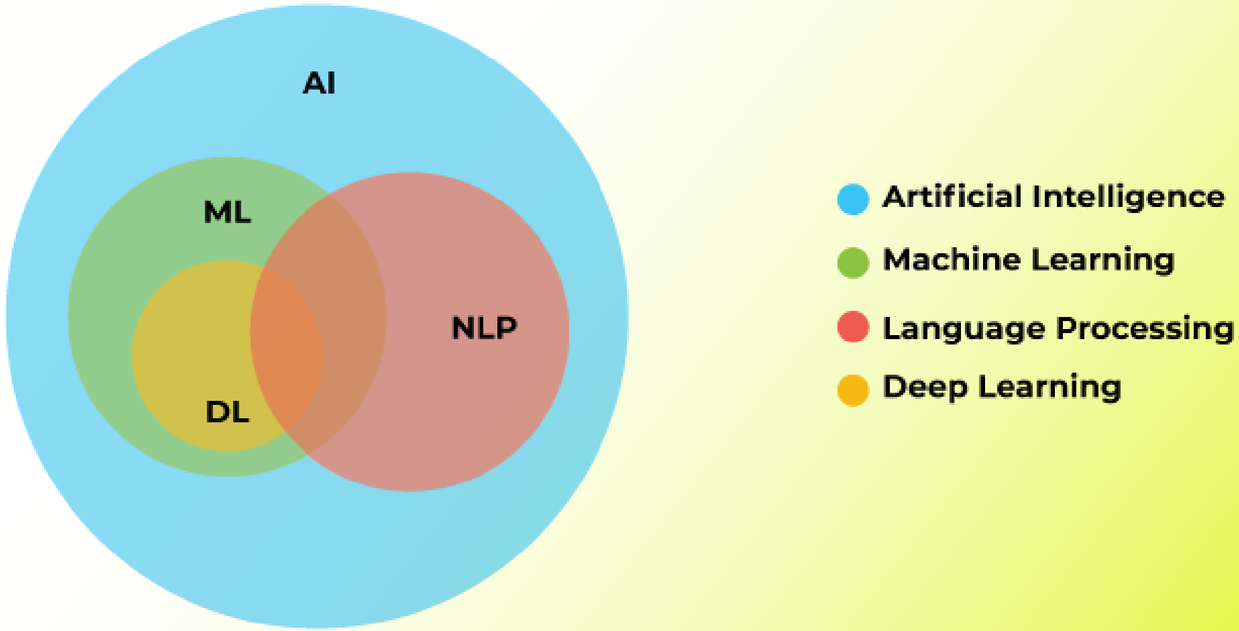
Artificial Intelligence

“The science and engineering of making intelligent machines” - Stanford Professor, John McCarthy 1955

Traditional Programming: “creating a fixed set of instructions for a computer to follow”

AI: “AI is about creating algorithms that allow computers to learn from data and make decisions or predictions.”

What is AI? What types of AI exist? How is AI used in our everyday lives?



Ethical Issues in AI

Bias: Systematic and unfair discrimination in predictions against certain groups.

- Arises from biases in training data or problem formulation.

Algorithmic Bias: “Most commonly, especially among non-data-scientists, algorithmic bias refers to computational discrimination whereby unfair outcomes privilege one arbitrary group of people over another. In this definition, the focus is on the disparate impact technology may have that reinforces social biases based on race, gender, sexuality, ethnicity, age, and disability.“

WHAT ARE THE MOST COMMON ISSUES IN AI?

Bias in AI Algorithms

Unbiased Estimation:

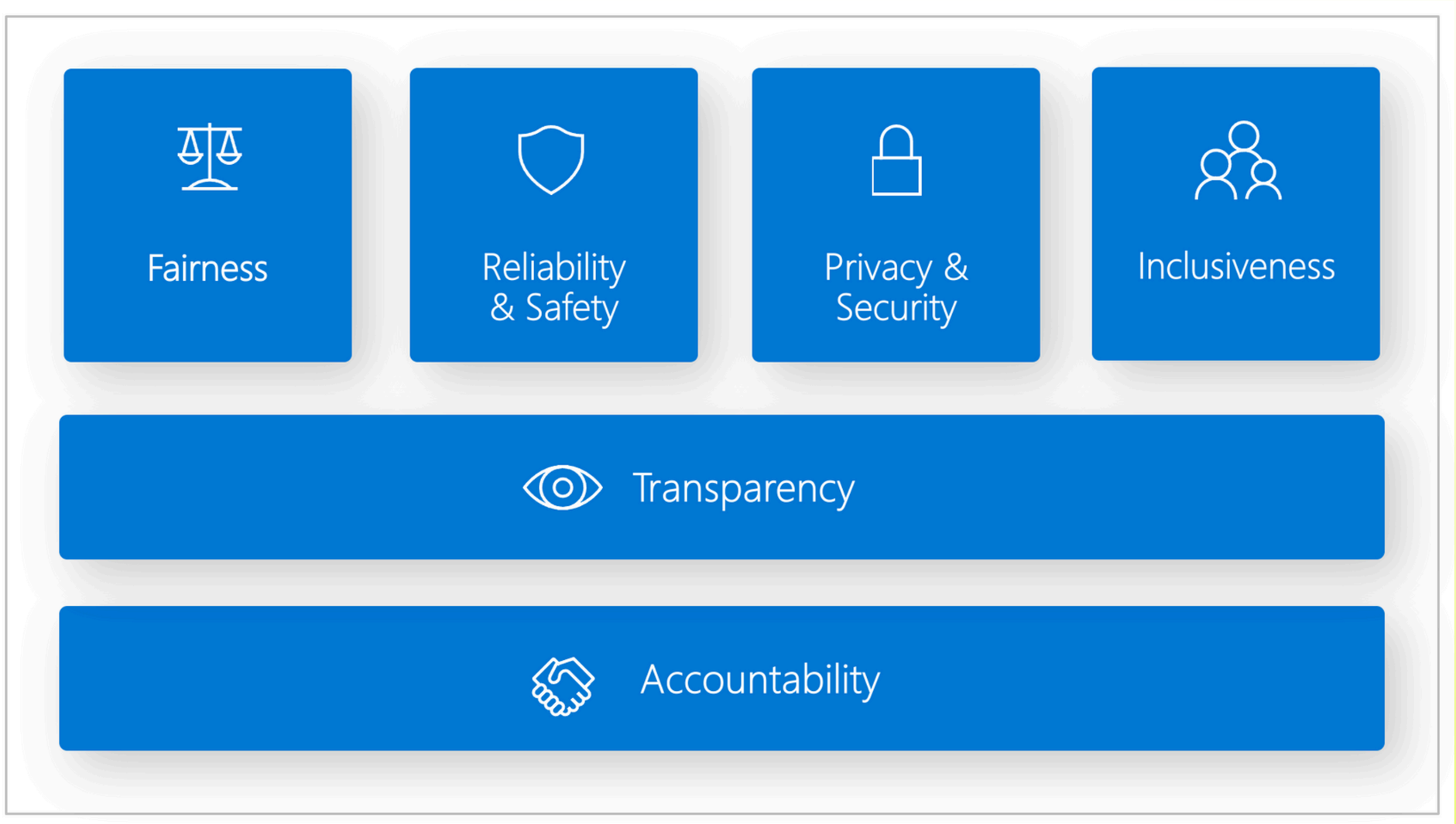
- An estimator is unbiased if it correctly predicts the true value on average
  - $E[\hat{\theta}] = \theta$
  - Where $\hat{\theta}$ is the estimator, and θ is the true parameter.

Interplay and Balance:

- A statistically unbiased model is not necessarily free from ethical bias.
- Goal: Minimize both statistical error (through bias-variance trade-off) and ethical bias (via fairness-aware strategies).

WHAT ARE FAIRNESS-AWARE STRATEGIES?

Responsible AI toolkit by Microsoft



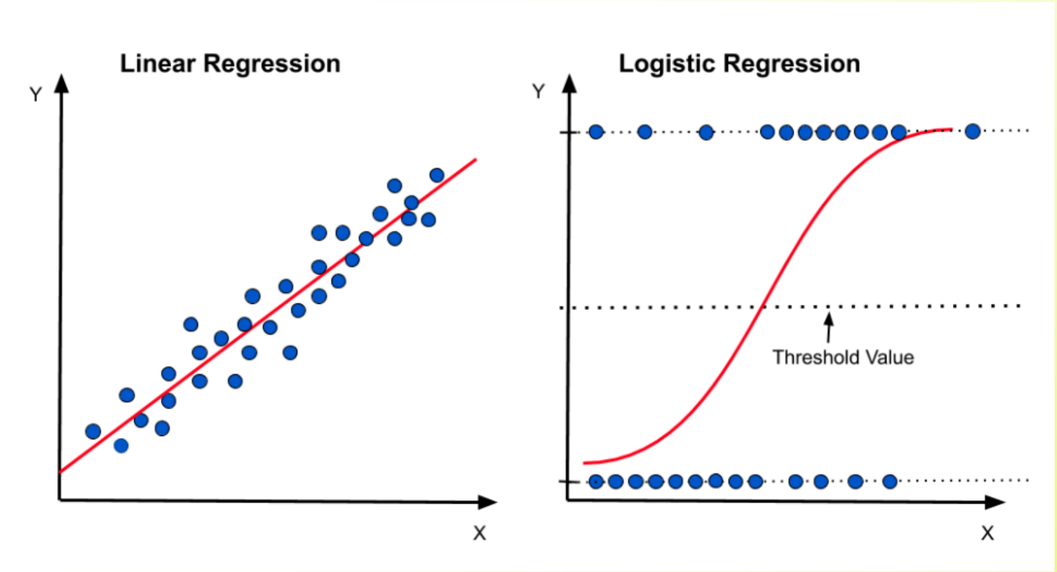
COMPAS

Predictive algorithms are now a part of everyday life such as which products you are likely to purchase or which movies you will likely enjoy, but life but can they effectively predict when someone is likely to reoffend?

Predictive algorithms for predictive policing however have a negative consequences on the civil liberties Latinx communities.

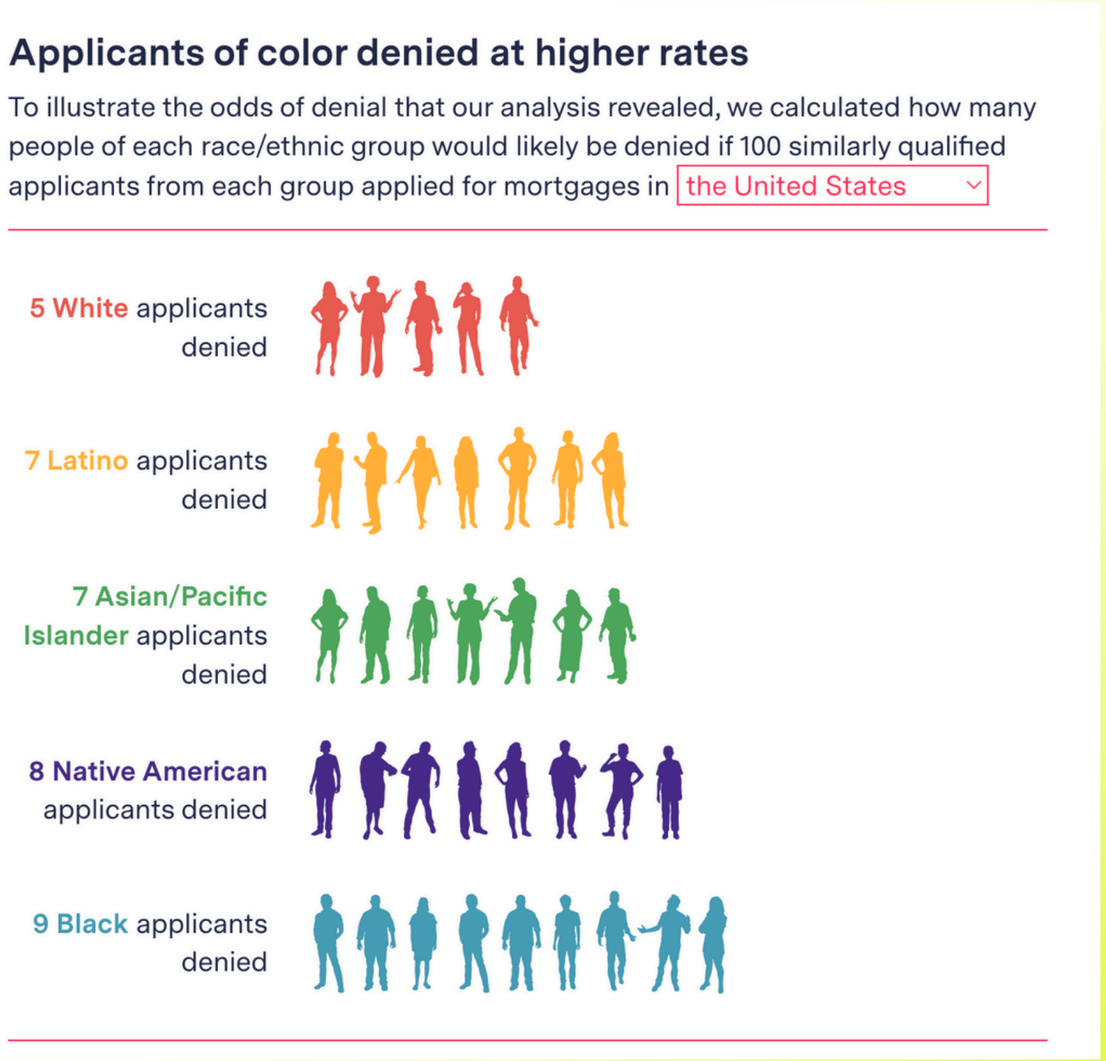
Predictive algorithms for predictive policing however have a negative consequences on the civil liberties Latinx communities.

All models can be expressed mathematically. COMPAS was reversed engineered through a **logistic regression model.**



Algoritmic Redlining

Redlining can be defined as a discriminatory practice that consists of the systematic denial of services such as mortgages, insurance loans, and other financial services to residents of certain areas, based on their race or ethnicity.



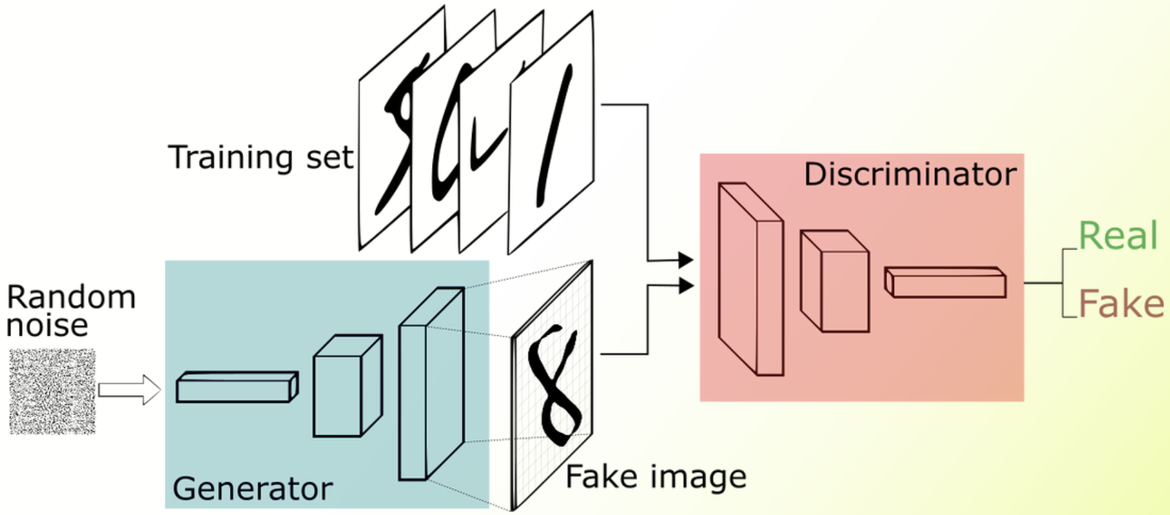
Generative AI

Generative vs. Discriminative AI:

- Generative models aim to understand and replicate data to generate new instances (text, audio, images).
- Discriminative models focus on distinguishing between different types of data inputs and predict labels.

Generative AI and AI Ethics:

- Generative AI can replicate historical biases when generating instances of under-represented groups in the data.
- Biases present in data can lead to biased outputs.

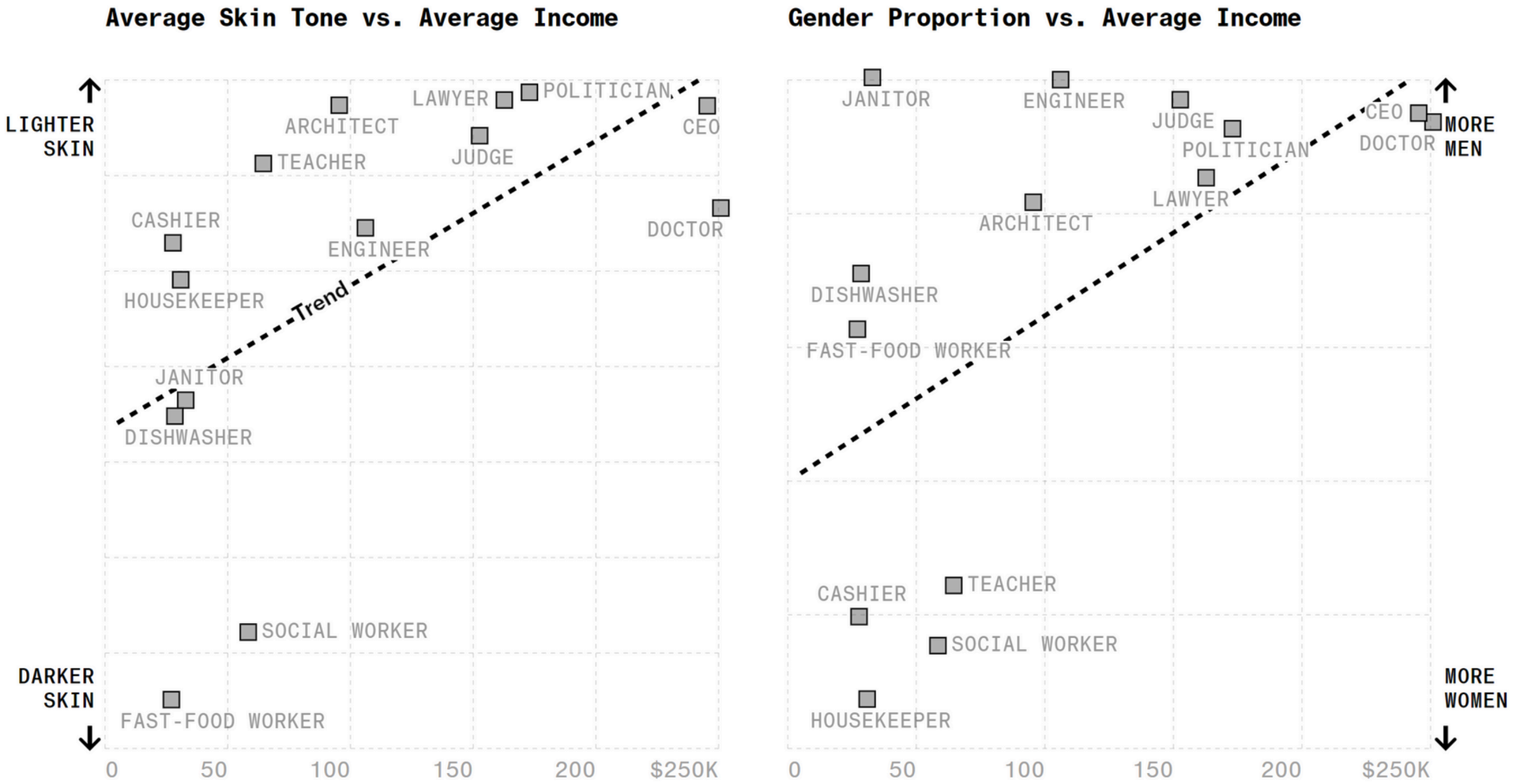


Generative Adversarial Network (GAN)

HOW DOES BIAS AFFECT MINORITY GROUPS?

As Income Increases, So Does the Share of Lighter-Skinned Men

Stable Diffusion results compared to average US income for each occupation



Sources: Bureau of Labor Statistics (May 2022), US House of Representatives, US Senate, Bloomberg analysis of Stable Diffusion

Gender Bias in Gen AI

Link: https://huggingface.co/spaces/society-ethics/DiffusionBiasExplorer

Racial + Religious Bias in AI

https://www.kaggle.com/code/alexisbcook/identifying-bias-in-ai

Resources

Literature	Websites	Organizations
“Algorithms of Oppression” - Safiya Noble	Medium (quick reads, tutorials)	Women in AI Ethics
“Race After Technology” - Dr. Ruha Benjamen	Kaggle (competitions, datasets)	Data + Society
“Machine Learning Refined” - Jeremy Watt, Reza Borhani, Aggelos Katsaggelos	Google Machine Learning Education (tutorials)	Tech Policy Press
“Deep Learning: A Practitioner’s Approach” - Josh Patterson, Adam Gibson	Google Colab (Coding)	Black in AI
“The Hundred-Page Machine Learning Book” - Andriy Burkov	Machine Learning Refined GitHub	LATINX in AI

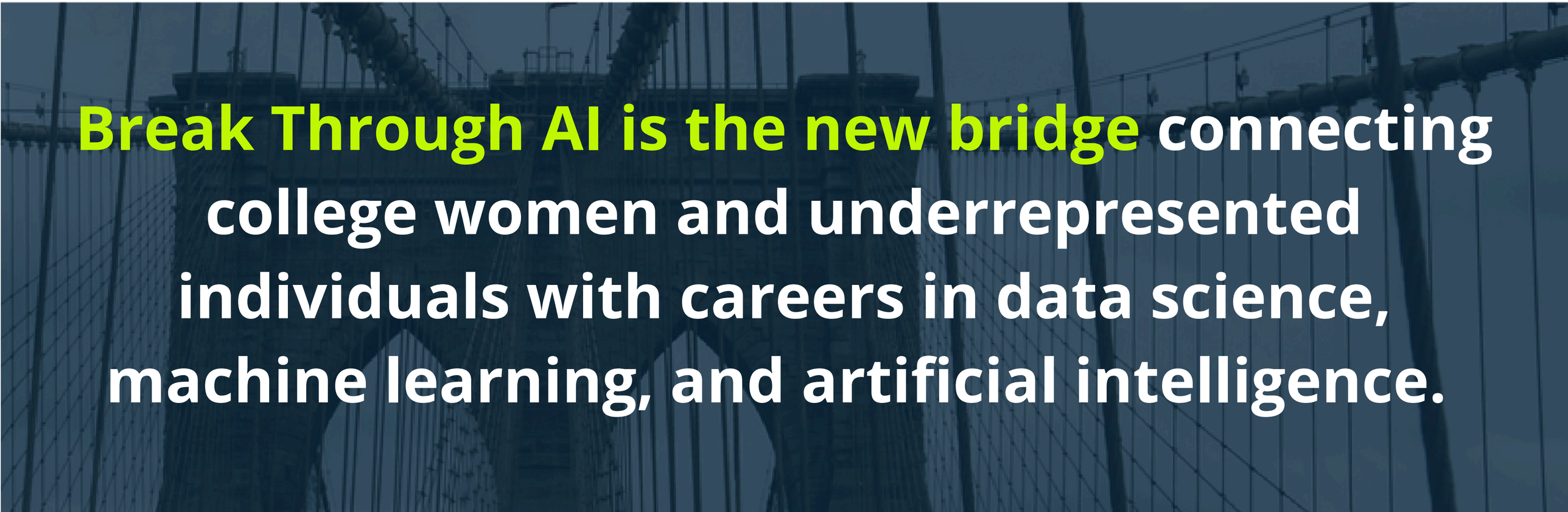
Opportunities in AI

AI Ethics Domains	Key Trends/Opportunities
Sustainability	Predictive modeling for renewable energy, waste reduction, and conservation.
Healthcare	Personalized medicine, AI-driven diagnostics, ethical data usage in medical AI.
Education	Adaptive learning systems, AI tools for accessibility, personalized learning experiences.
AI Governance	Development of ethical guidelines, participatory governance models, policy-making.
Privacy & Security	Balancing AI innovation with data privacy, secure AI systems, ethical data handling.

BREAK THROUGH TECH

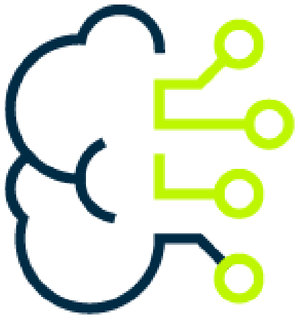
Break Through Tech AI at UCLA

Ready to break into Artificial Intelligence?



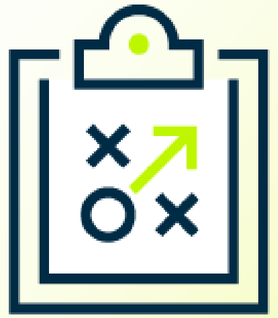
Break Through AI is the new bridge connecting college women and underrepresented individuals with careers in data science, machine learning, and artificial intelligence.

Program Overview



Rigorous machine learning and AI curricula that teach the practical skills that qualify candidates for jobs in the field

Building a portfolio of projects that pose solutions to real-world business challenges



Mentorship and career coaching that build students' authentic leadership and social capital to result in employment in ML and AI-related roles

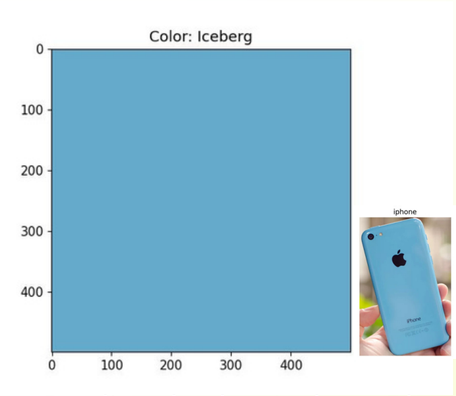
Our ultimate goal is to drive **full-time employment** in machine learning and artificial intelligence.

Project Showcase

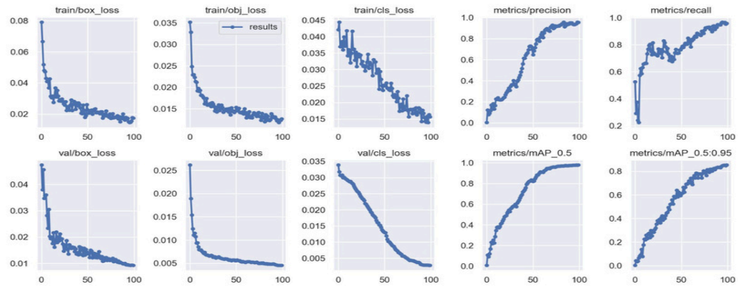
Verizon: Identifying Objects and Sorting by Color

Annotated training image

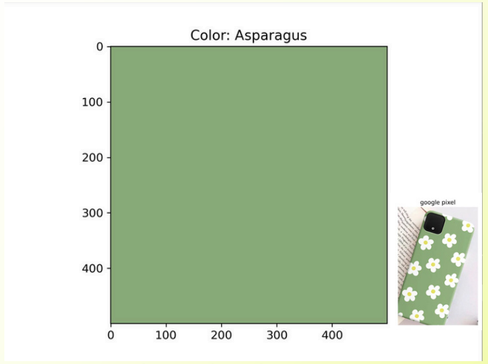
Validation set with annotated labels



Predicted color: Iceberg Blue



Model performance metrics



Predicted color: Asparagus

NCWIT Aspirations in COMPUTING

Our ultimate goal is to drive **full-time employment** in machine learning and artificial intelligence.

Fill out the interest form to learn more

Follow us:

@brkthroughla

Q&A