Combating Bias in Production Computer Vision Systems

Alex Thaman

Chief Architect

Red Cell Partners

RED CELL

About Red Cell Partners

Red Cell is an incubation firm that builds & invests in rapidly scalable, technology-led companies bringing revolutionary advancements to market in National Security & Healthcare, sectors where we have a distinct competitive advantage.

Our Mission

United by a shared sense of duty & deep belief in the power of innovation, we develop technology to address our Nation's most pressing problems.

Our Strategy

We incubate & invest in technology-led companies that address key issues in National Security & Healthcare, backing the most promising emerging firms & dramatically accelerating their growth by leveraging our team's:

Domain knowledge

Technology expertise (Big Data, AI/ML, AR/VR, Kinetics, etc.)

Strategic networks

Exceptional capabilities in regulated end markets

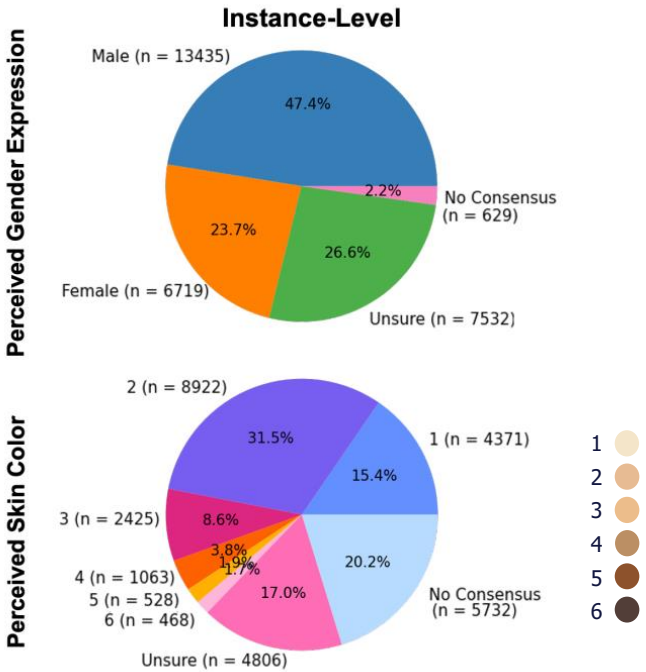
What Is The Problem?

RED CELL

Dataset Statistics in Coco Dataset

Understanding and Evaluating Racial Biases in Image Captioning

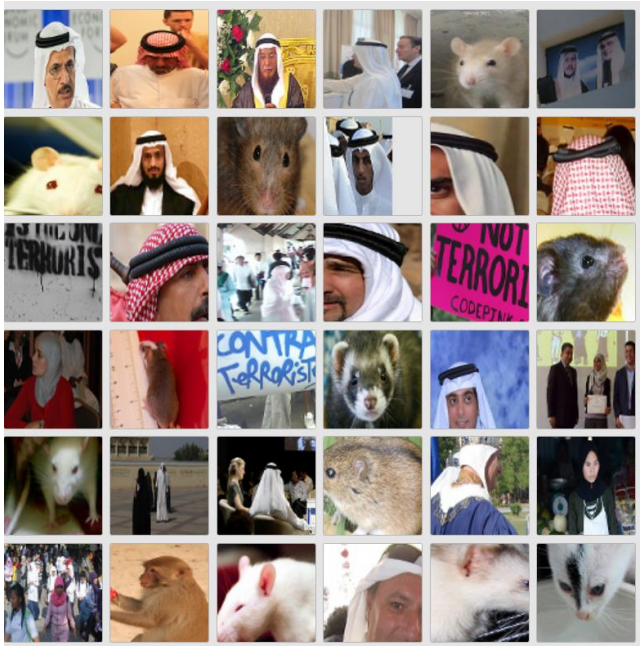
- Males appear 2x more than females
- Light skin appears 7.5x more than dark
- Dark-skinned females appear 23.1x less frequently than dark-skinned males
- Lighter-skinned appear more with indoor and furniture objects
- Darker-skinned appear more with outdoor and vehicle objects



Zhao, D., Wang, A., & Russakovsky, O. (2021). Understanding and Evaluating Racial Biases in Image Captioning. In International Conference on Computer Vision (ICCV).

Encoded Bias

- CLIP
  - 400M images from 500K word queries based based on frequency in Wikipedia
  - [Multimodal neurons in artificial neural networks (openai.com)]
  - *We have observed, for example, a "Middle East" neuron with an association with terrorism; and an "immigration" neuron that responds to Latin America. We have even found a neuron that fires for both dark-skinned people and gorillas…*



https://microscope.openai.com/models/contrastive_v2/image_block_4_2_Add_6_0/1895

Generative AI Bias

Prompt: “A person cleaning a living room”

Source: Bing Image Creator

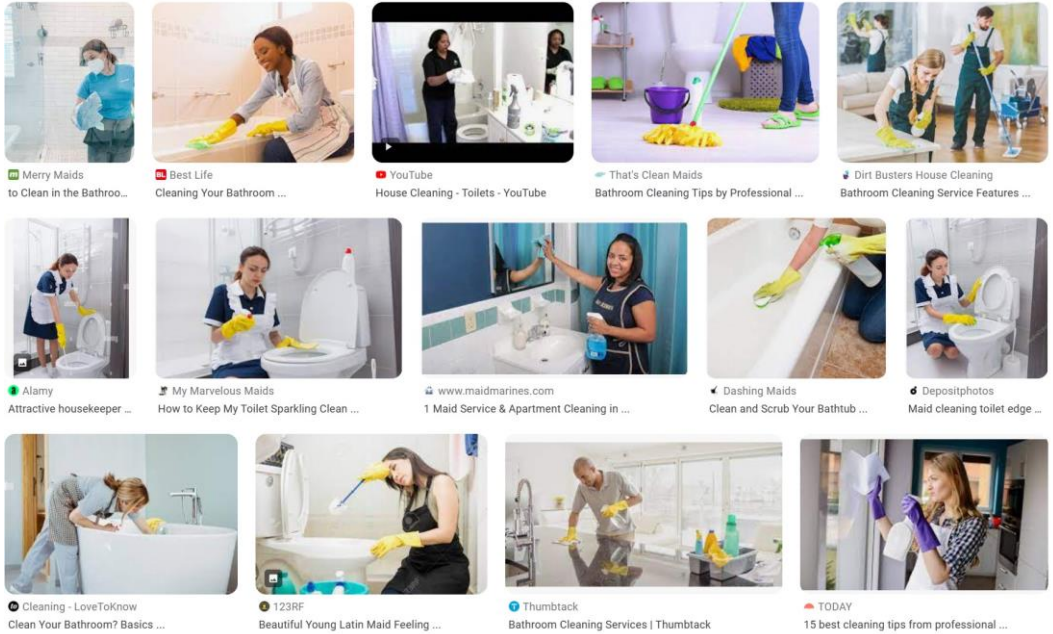
Generative AI Bias

Prompt: “A housecleaner cleaning a bathroom”

Source: Bing Image Creator

Generative AI Bias

Search Query: “A housecleaner cleaning a bathroom”



Source: Google Image Search

Generative AI Bias

Prompt: “A nurse taking care of a hospital patient”

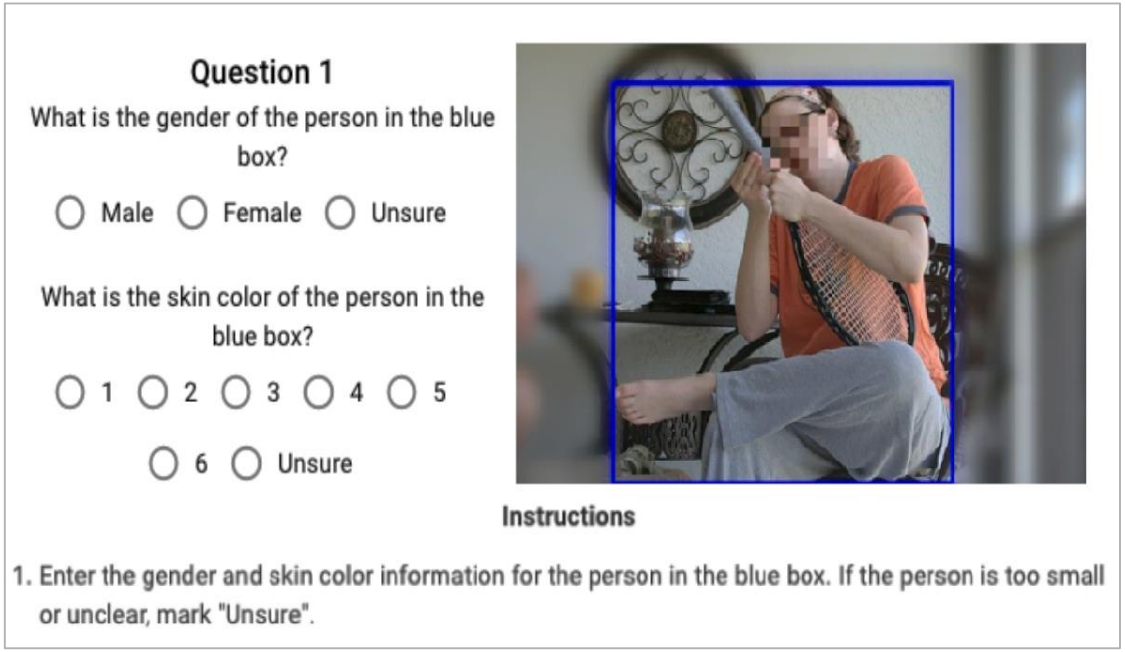
Source: Bing Image Creator

Measuring Bias

RED CELL

Metadata Labeling

- Labeling data for analytics
  - Distribution mismatches
  - Find correlations
- Consider labeling costs
  - Sampling your samples
  - Instructions
  - Pre-labeling



Zhao, D., Wang, A., & Russakovsky, O. (2021). Understanding and Evaluating Racial Biases in Image Captioning. In International Conference on Computer Vision (ICCV).

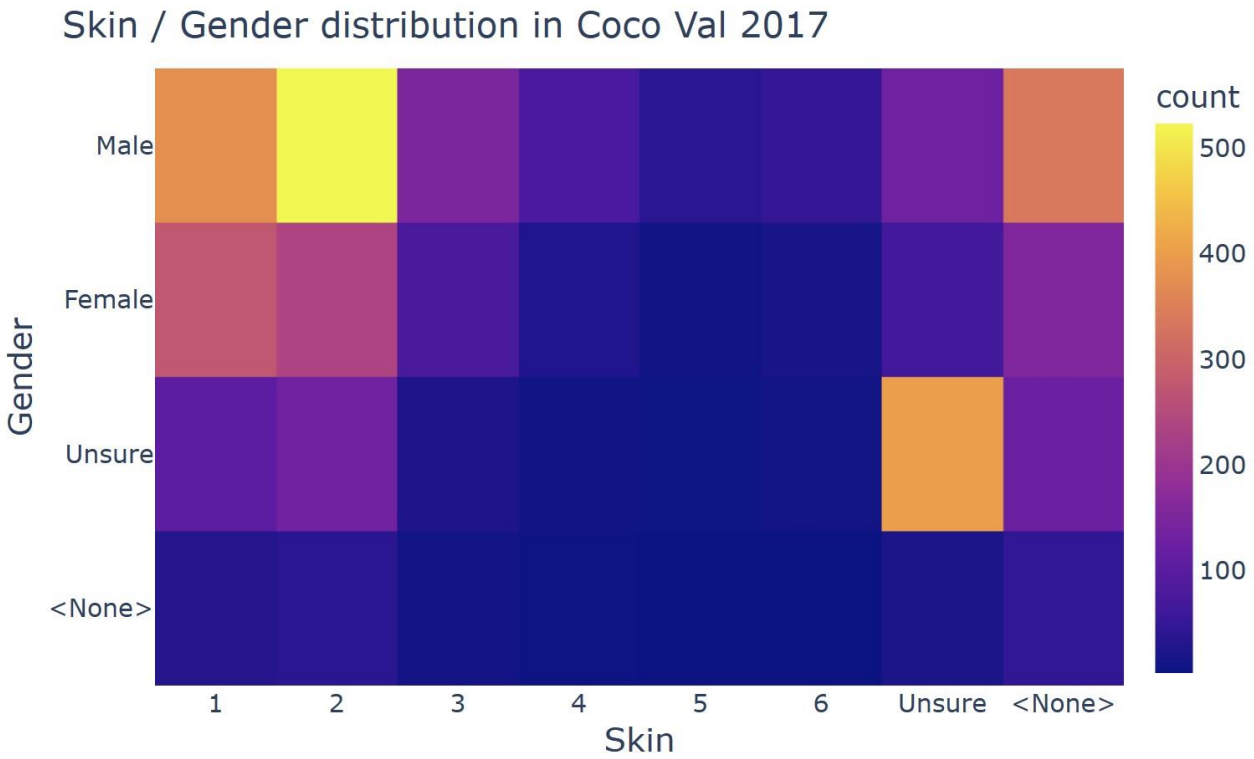
Skin Tone Labeling

- Fitzpatrick scale
  - 6 classifications
  - Skewed towards white skin variations
- Monk scale
  - 10 classifications
  - Tested socially across population groups for fair representation

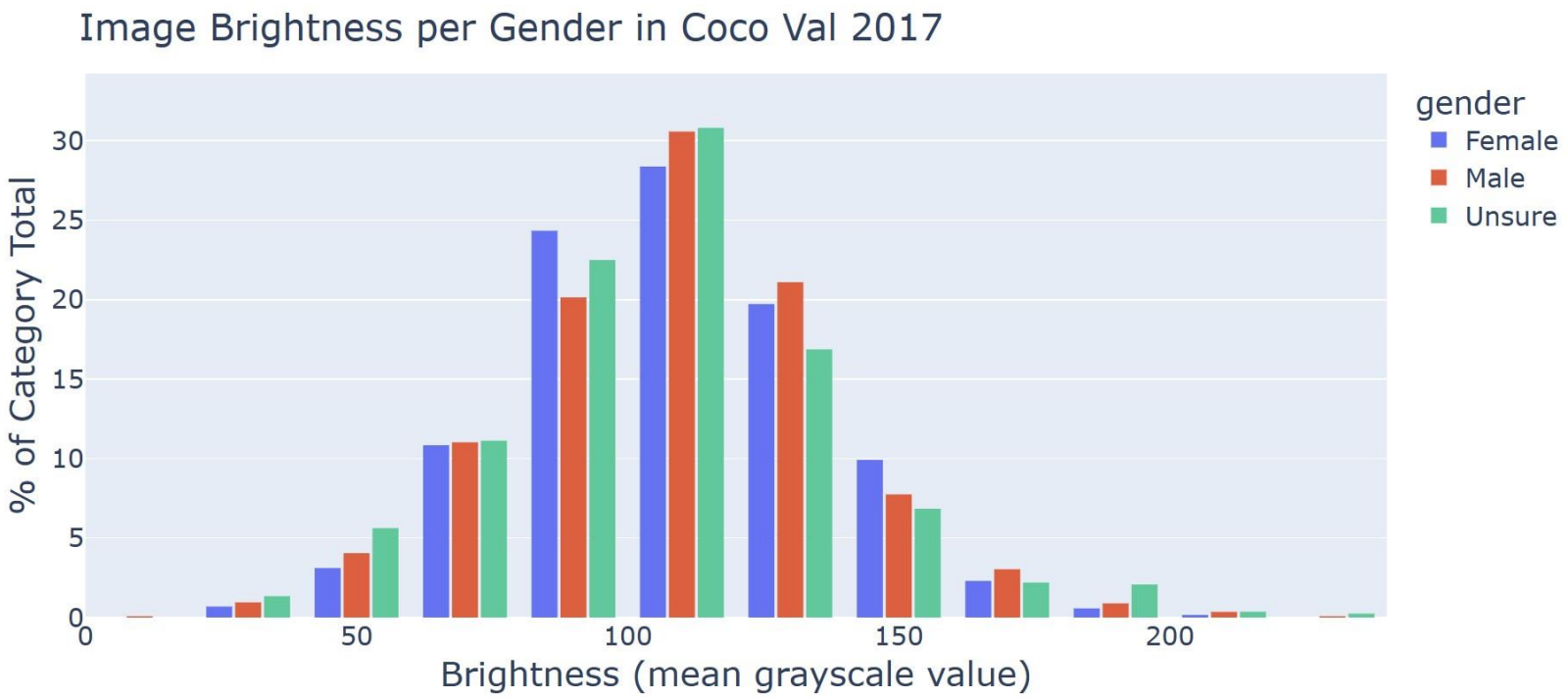
The Fitzpatrick Scale

The Monk Skin Tone Scale

Metadata Statistics

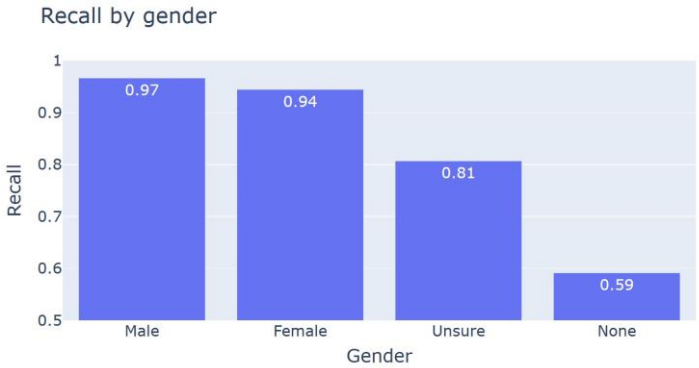


Metadata Statistics

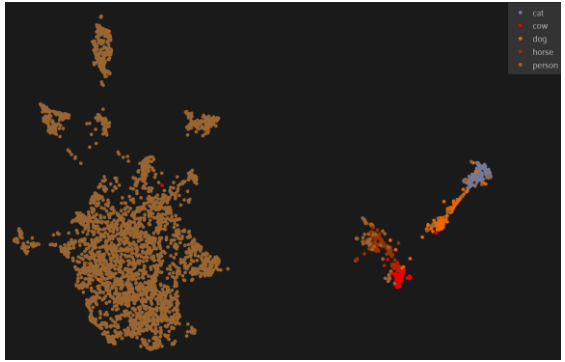


Measuring Bias in Production

- Live data collection
  - Sample data in production
  - Compare inference results across dimensions
- Without access to the data
  - Human supervision is difficult
  - Use metadata models for insights
  - Data drift in embedding space – KL divergence, WS distance, etc.



UMAP of CLIP embedding of cropped boxes, colored by class



Controlling Bias

RED CELL

Controlling Bias

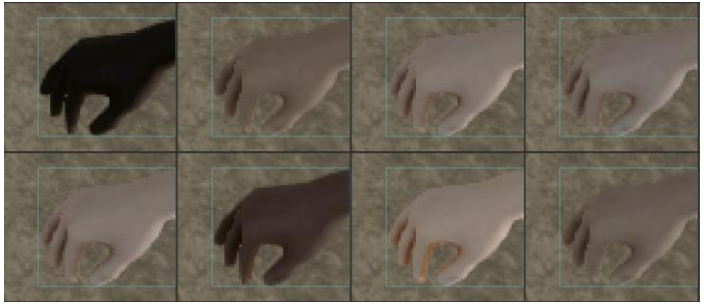
- Data collection
  - Targeted data collection for specific population groups
  - **Collect more data than you think you need**
- Dataset balancing
  - Sample training data across dimensions
  - Overweight underrepresented data

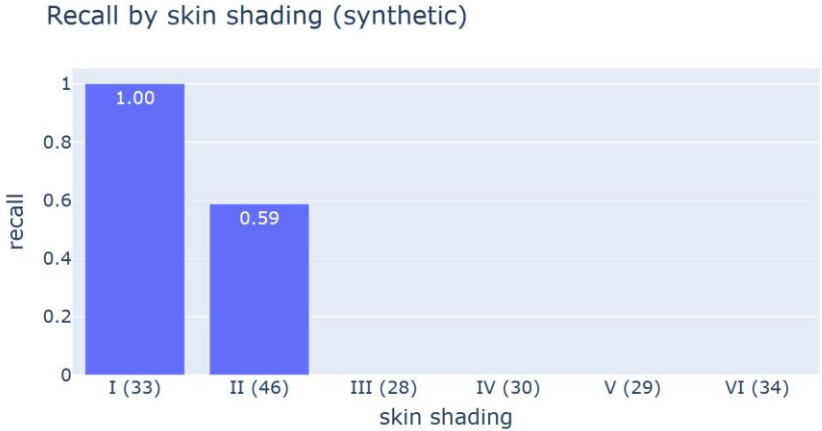
Controlling Bias

- Synthetic data
  - Infinite labels/metadata
  - Study model behavior across dimensions
  - Learn domain-invariant features via pre-training

Microsoft Face Synthetics: https://microsoft.github.io/FaceSynthetics/

Synthetic Data Study





Other Considerations

RED CELL

Descriptive Metadata Challenges

- Skin tone
  - Perceived differently by different cultures
  - Multidimensional
  - Image characteristics and environment
- Sex/gender not globally normalized

What skin tone are these 2 people?

The Cost of Bias

- "To have any inductive process make predictions on unseen data, an agent requires a bias. What constitutes a good bias is an empirical question about which biases work best in practice"
  Poole, D. & Mackworth, A. (2019). Artificial Intelligence: Foundations of Computational Agents.
- Think about tradeoffs
- Can the application layer help?

Conclusions

- All AI systems have bias that results from the data that was used
- AI systems need deep evaluation prior to widespread deployment
- Analysis through additional metadata, statistics, and controlled experiments help predict how these systems may perform in the real world

Resources

GitHub repository for this talk
https://github.com/alexthaman/evs2023

Understanding and Evaluating Racial Biases in Image Captioning
https://arxiv.org/abs/2106.08503

Themis AI – evaluate bias in your model
https://themisai.io/

Contact Info

LinkedIn:
https://www.linkedin.com/in/alex-thaman-93436659/

Email:
alex.thaman@redcellpartners.com

Appendix

RED CELL

Additional Resources

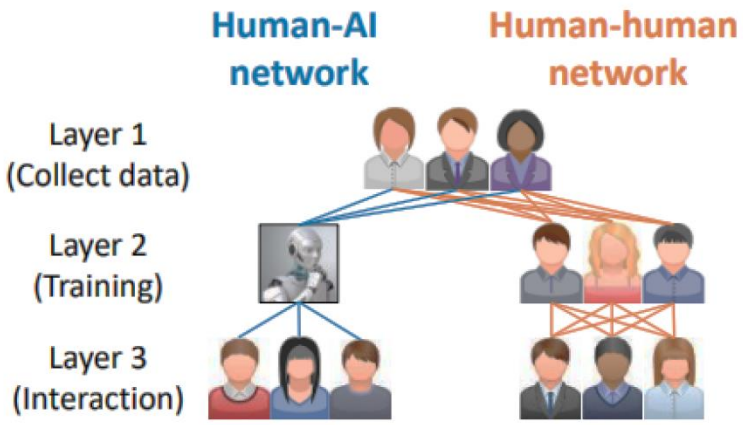
- Detecting data drift - https://www.evidentlyai.com/blog/data-drift-detection-large-datasets
- Monk Scale blog https://blog.google/technology/research/ai-monk-scale-skin-tone-story/
- REVISE (open source tool for evaluating bias) - https://github.com/princetonvisualai/revise-tool
- Other products to evaluate bias: Manot (https://www.manot.ai/)

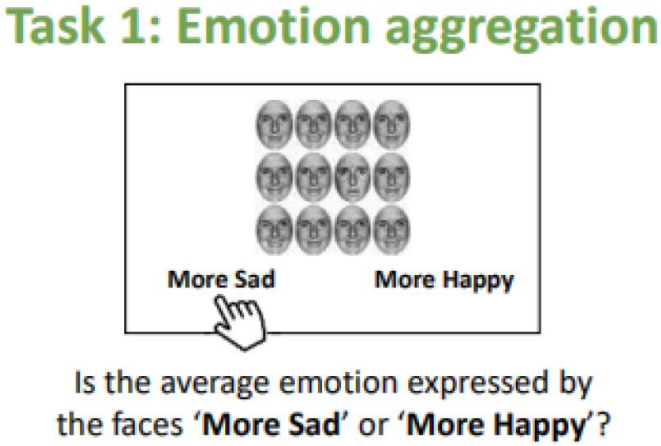
Additional Resources

- Digital Humans Synthetic data
  - Synthesis AI - https://synthesis.ai/
  - Datagen - https://datagen.tech/
  - Infinity AI - https://infinity.ai/
  - Unity Digital Humans - https://github.com/Unity-Technologies/com.unity.cv.synthetichumans
  - Microsoft Face Synthetics - https://microsoft.github.io/FaceSynthetics/

AI-Human Feedback Loop

Bias AI systems produce biased humans: Experiment Setup

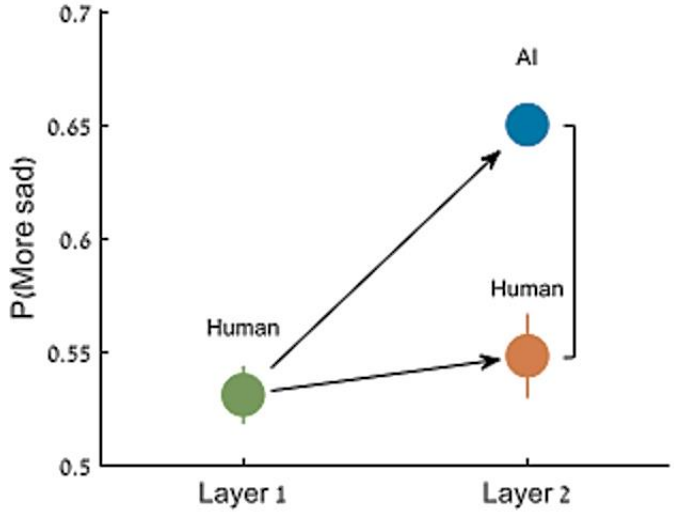


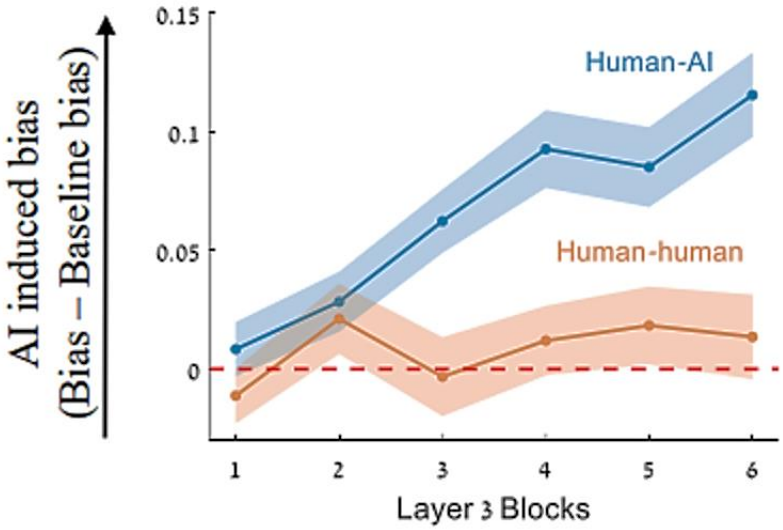


Glickman, M., & Sharot, T. (2022, November 15). Biased AI systems produce biased humans. https://doi.org/10.31219/osf.io/c4e7r

AI-Human Feedback Loop

Conclusion 1: Human-AI interactions create bias feedback loops

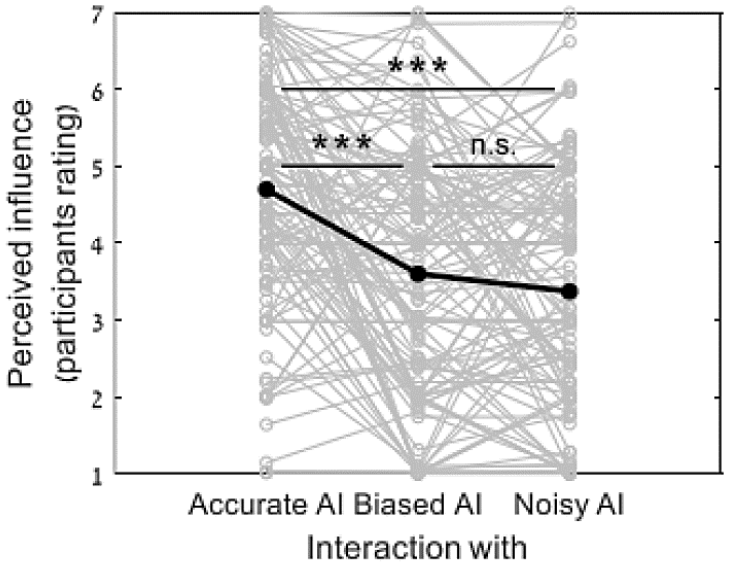


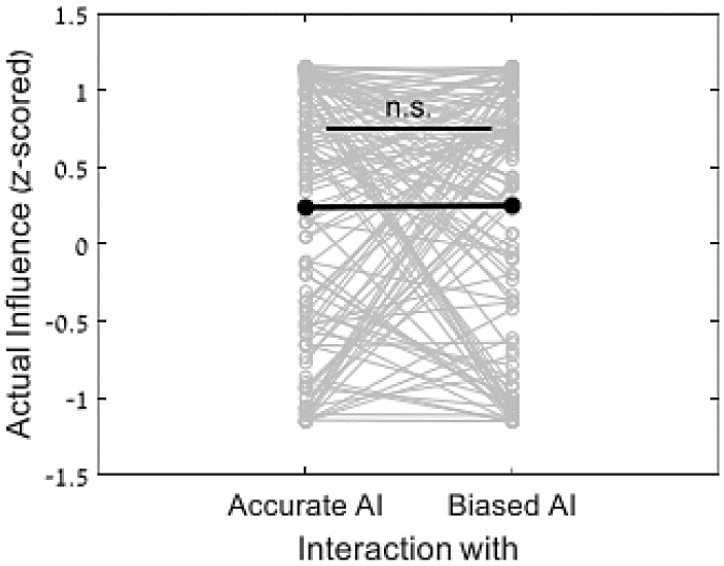


Glickman, M., & Sharot, T. (2022, November 15). Biased AI systems produce biased humans.
https://doi.org/10.31219/osf.io/c4e7r

AI-Human Feedback Loop

Conclusion 2: Humans underestimate the impact of the bias from AI





Glickman, M., & Sharot, T. (2022, November 15). Biased AI systems produce biased humans.
https://doi.org/10.31219/osf.io/c4e7r