

Responsible Computer Vision: Part 1

Danna Gurari

University of Colorado Boulder
Fall 2023



Review

- Last lecture on efficient learning:
 - Motivation
 - Curriculum Learning
 - Active Learning
 - Few-shot Learning
- Assignments (Canvas):
 - Final project presentations due next week
 - Final project report due in two weeks
- Questions?

Today's Topics

- Computer Vision that Discriminates
- FAT (Fair, Accountable, & Transparent) Algorithms
- Ethics in Computer Vision

Today's Topics

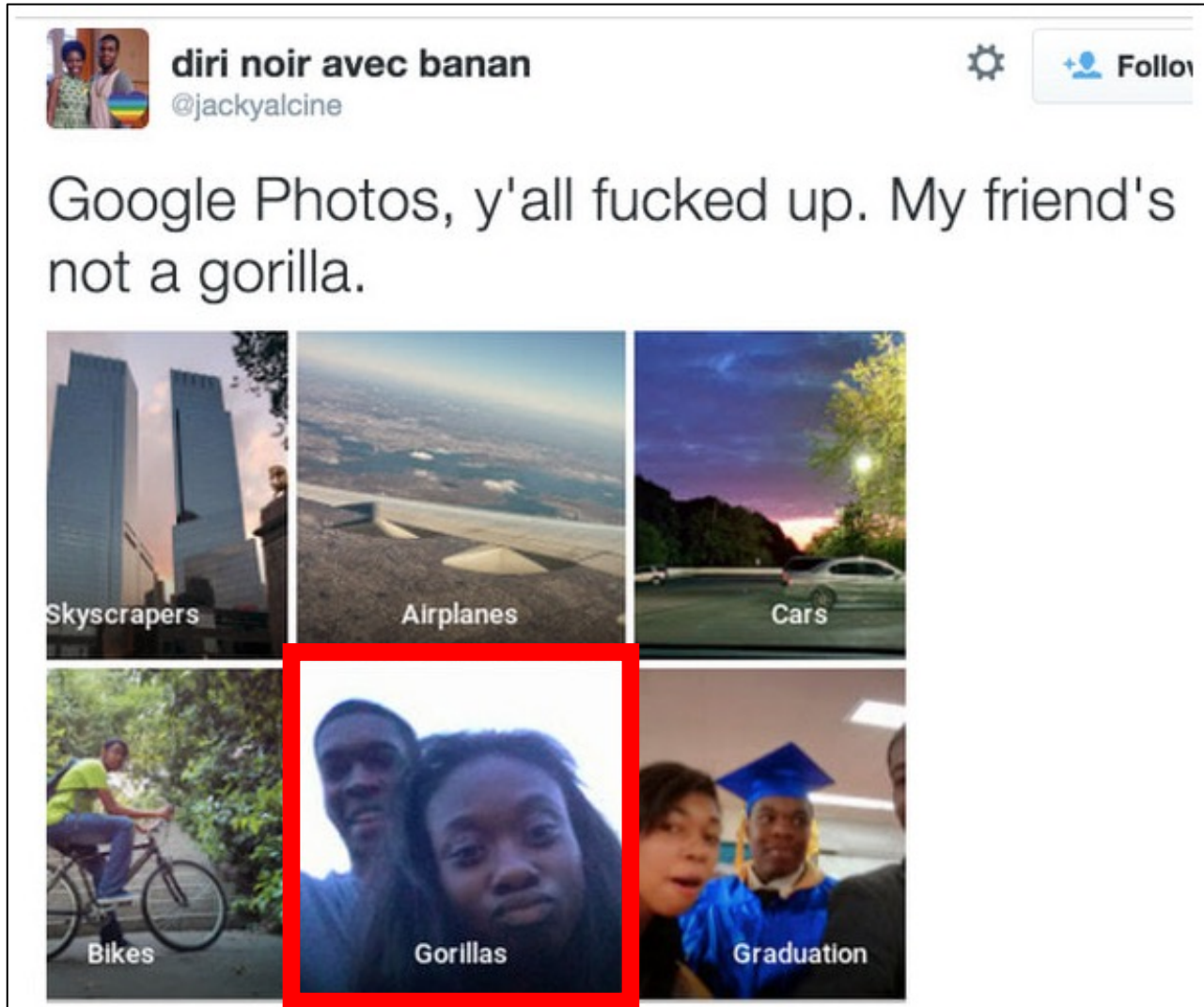
- Computer Vision that Discriminates
- FAT (Fair, Accountable, & Transparent) Algorithms
- Ethics in Computer Vision

Observation: World Population is Diverse



Image Source: <https://www.rocketpace.com/corporate-innovation/why-diversity-and-inclusion-driving-innovation-is-a-matter-of-life-and-death>

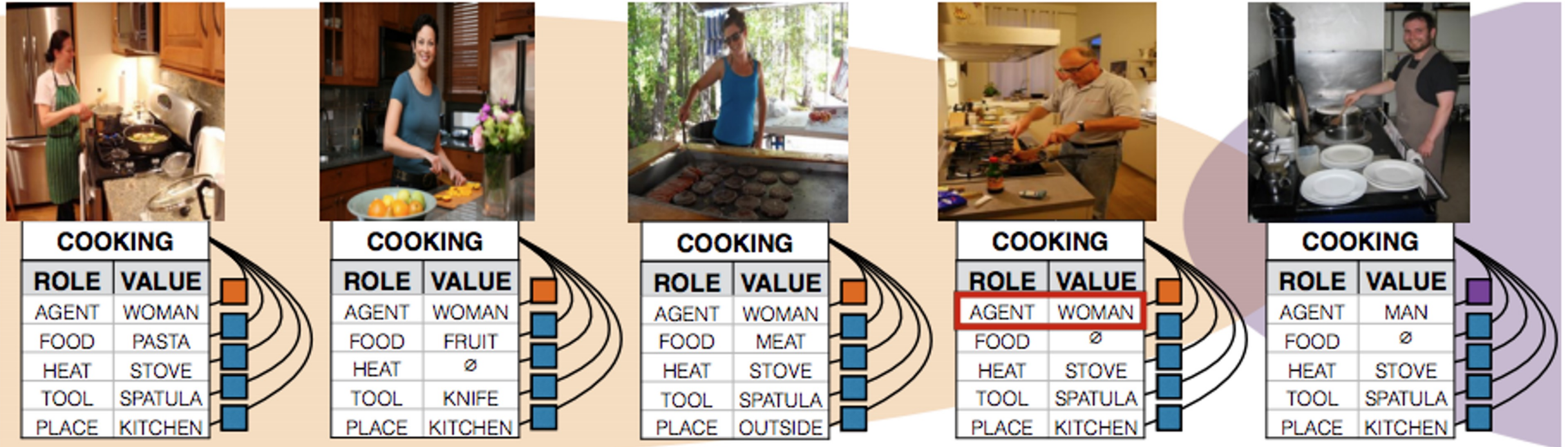
Models Discriminate: Image Tagging



Using Twitter to call out Google's algorithmic bias

<https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas>

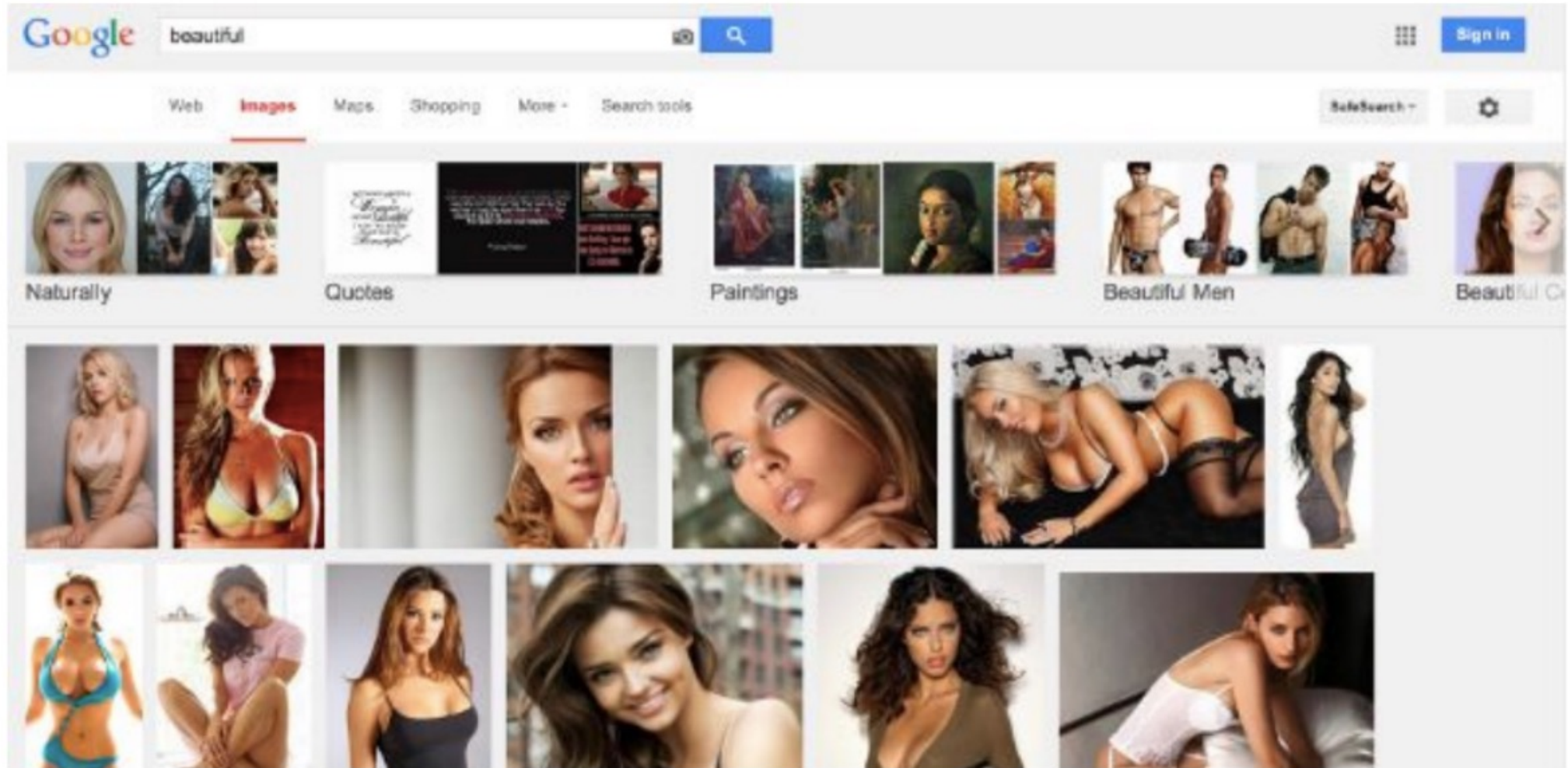
Models Discriminate: Image Tagging



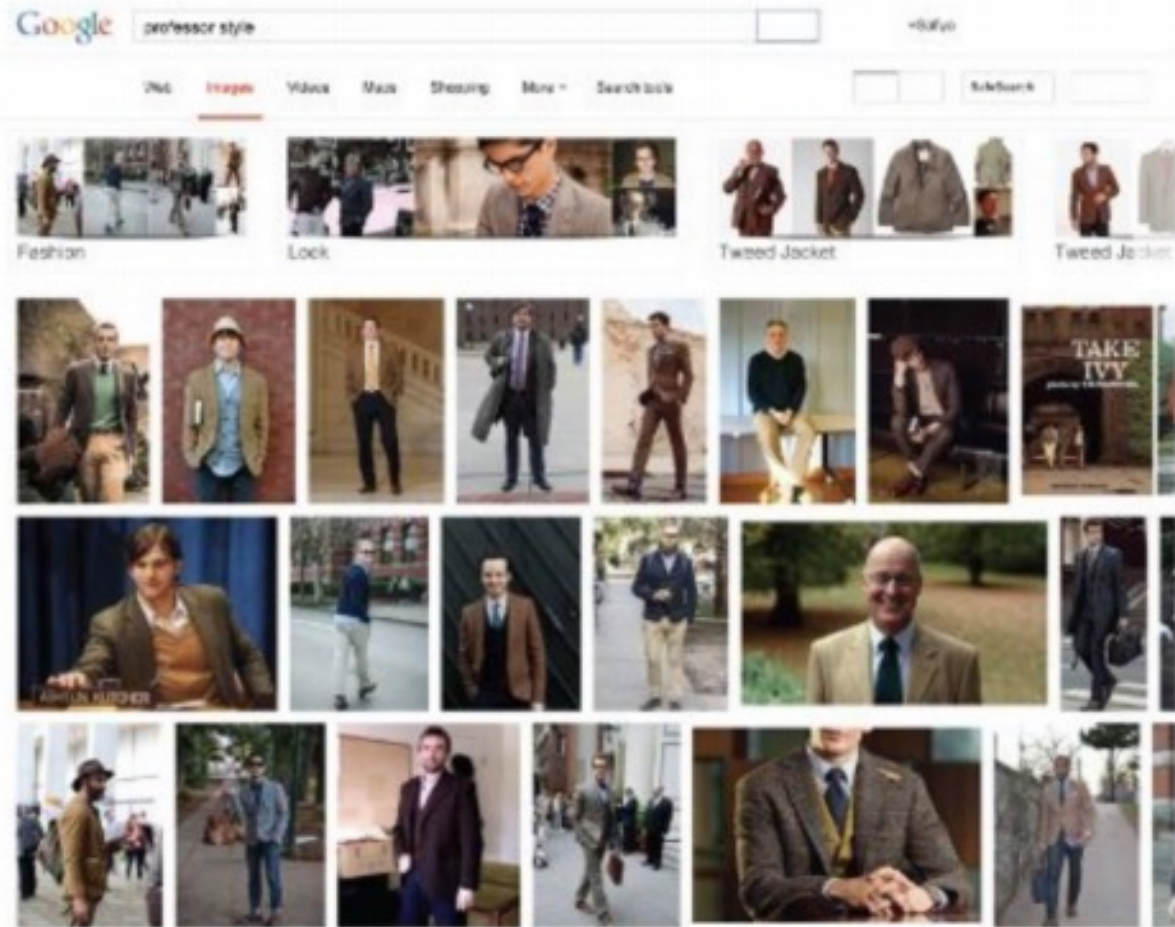
Algorithm identifies men in kitchens as women. Learned this example from given dataset. (Zhao, Wang, Yatskar, Ordonez, Chang, 2017)

<https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/>

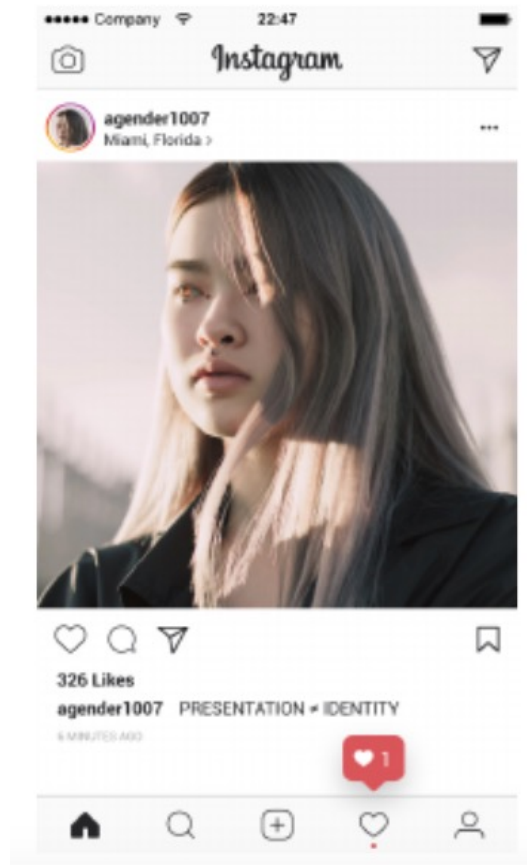
Models Discriminate: Image Tagging (“beautiful”; 2014)



Models Discriminate: Image Tagging (“professor style”; 2014)



Models Discriminate: Image Tagging



```
...
"age": {
  "min": 20,
  "max": 23,
  "score": 0.923144
},
"face_location": {
  "height": 494,
  "width": 428,
  "left": 327,
  "top": 212
},
"gender": {
  "gender": "FEMALE",
  "gender_label": "female",
  "score": 0.9998667
}
}
{
  "class": "woman",
  "score": 0.813,
  "type_hierarchy": "/person
/female/woman"
},
{
  "class": "person",
  "score": 0.806
},
{
  "class": "young lady (heroine)",
  "score": 0.504,
  "type_hierarchy": "/person/female
/woman/young lady (heroine)"
}
...
```

Person identifies as agender (gender-less, and so non-binary)

Morgan Klaus Scheurman, Jacob M. Paul, and Jed R. Brubaker, "How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis and Image Labeling Services." CSCW 2019.

Models Discriminate: “Hotness” Photo-Editing Filter

FaceApp apologizes for building a racist AI

Natasha Lomas @riptari / 2 years ago

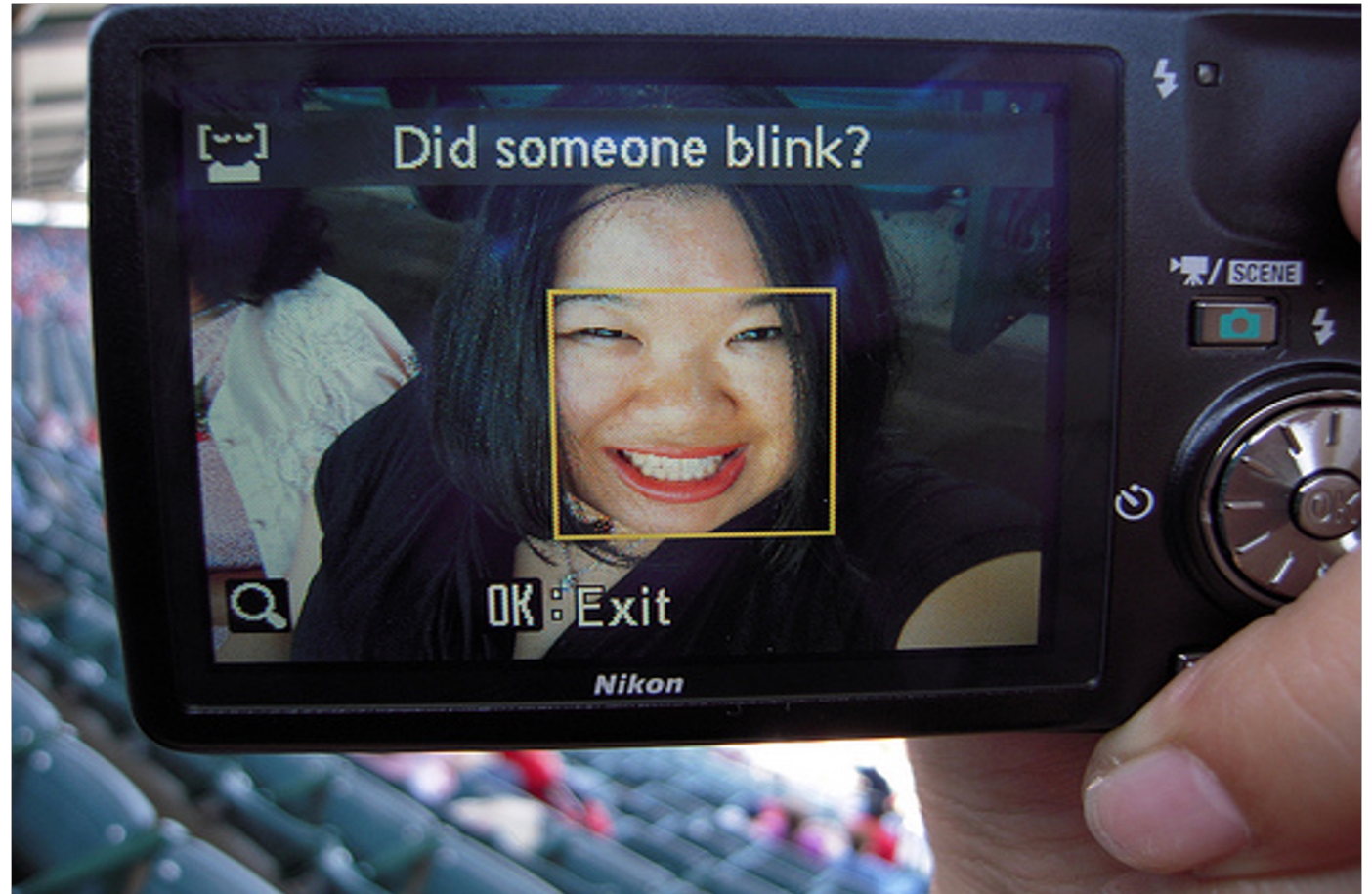
Comment



<https://techcrunch.com/2017/04/25/faceapp-apologises-for-building-a-racist-ai/>

Models Discriminate: Nikon Blink Detection

Two kids bought their mom a Nikon Coolpix S630 digital camera for Mother's Day... when they took portrait pictures of each other, a message flashed across the screen asking, "Did someone blink?"



Models Discriminate: Face Recognition

Software engineer at company: “It got some of our Asian employees mixed up,” says Gan, who is Asian. “Which was strange because it got everyone else correctly.”



Gfycat's facial recognition software can now recognize individual members of K-pop band Twice, but in early tests couldn't distinguish different Asian faces.  GFYCAT

<https://www.wired.com/story/how-coders-are-fighting-bias-in-facial-recognition-software/>

And MANY more ways that models discriminate!

How would you try to fix issues like these?

Today's Topics

- Computer Vision that Discriminates
- **FAT (Fair, Accountable, & Transparent) Algorithms**
- Ethics in Computer Vision

We know that algorithms are not perfect.

How can we alleviate the issue that CV algorithms discriminate?

FAT Deep Learning: In Vague, Lay Terms

- **Fairness:** treat people fairly
- **Accountability:** mimic infrastructure to oversee human decision makers (e.g., policymakers, courts) for algorithm decision-makers
- **Transparency:** clearly communicate algorithms' capabilities and limitations

FAT Deep Learning: Fairness

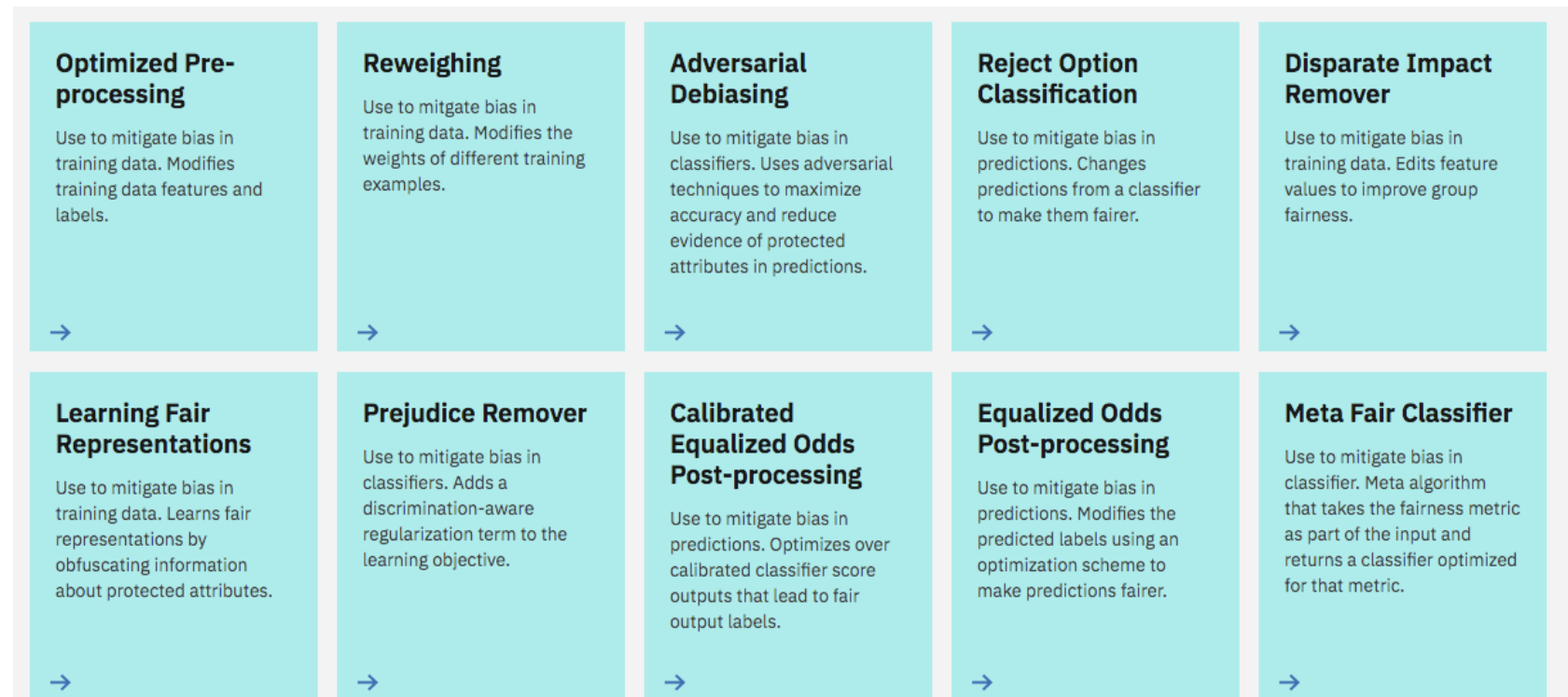
- How to make more fair methods?
 - Pre-processing:
 - Training data: modify it
 - Optimization at training:
 - Algorithm: e.g., add regularization term to objective function to penalize unfairness
 - Features: remove those that reflect bias; e.g., gender, race, age, education, sexual orientation, etc.
 - Post-process predictions
 - Counterfactual assumption: check impact of modifying single feature

FAT Deep Learning: Fairness

- Fairness – how to define this mathematically?
 - e.g., group fairness (proportion of members in protected group receiving positive classification matches proportion in the population as a whole)
 - e.g., individual fairness (similar individuals should be treated similarly)

e.g., IBM's AI Fairness 360
Open Source Toolkit

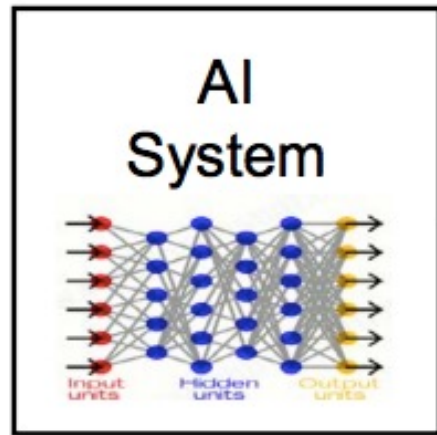
70+ fairness metrics and 10+
bias mitigation algorithms



FAT Deep Learning: Accountability

- Who is accountable for model behavior?
 - e.g., developers must design algorithms so that oversight authorities meet pre-defined rules (“procedural regularity”)?
 - e.g., data providers?
 - e.g., regulators who determine scope of oversight (e.g., require describing and explaining model failures)?

FAT Deep Learning: Transparency



- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand

Watson

A screenshot from the game show Jeopardy! showing the Watson interface. The board displays scores for Ken Jennings (\$200), Brad Rutter (\$4,000), and Watson (\$600). A question is shown: 'Maxwell's silver hammer' with a 90% confidence level. Other options are 'FRANK SINATRA' (11%) and 'Brown' (7%).

AlphaGo

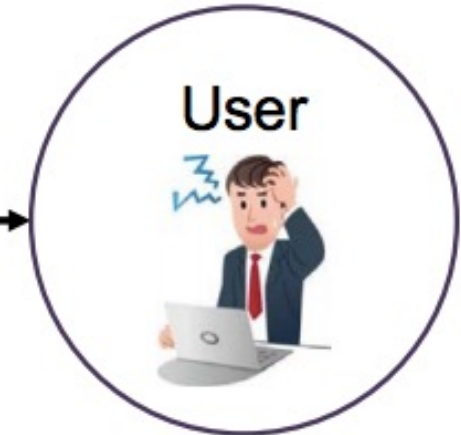
A photograph of Go stones (black and white) on a Go board.

Sensemaking

A photograph of a person in a military uniform operating a control room with multiple computer monitors.

Operations

A photograph of a soldier in a field next to a small, four-wheeled robot.



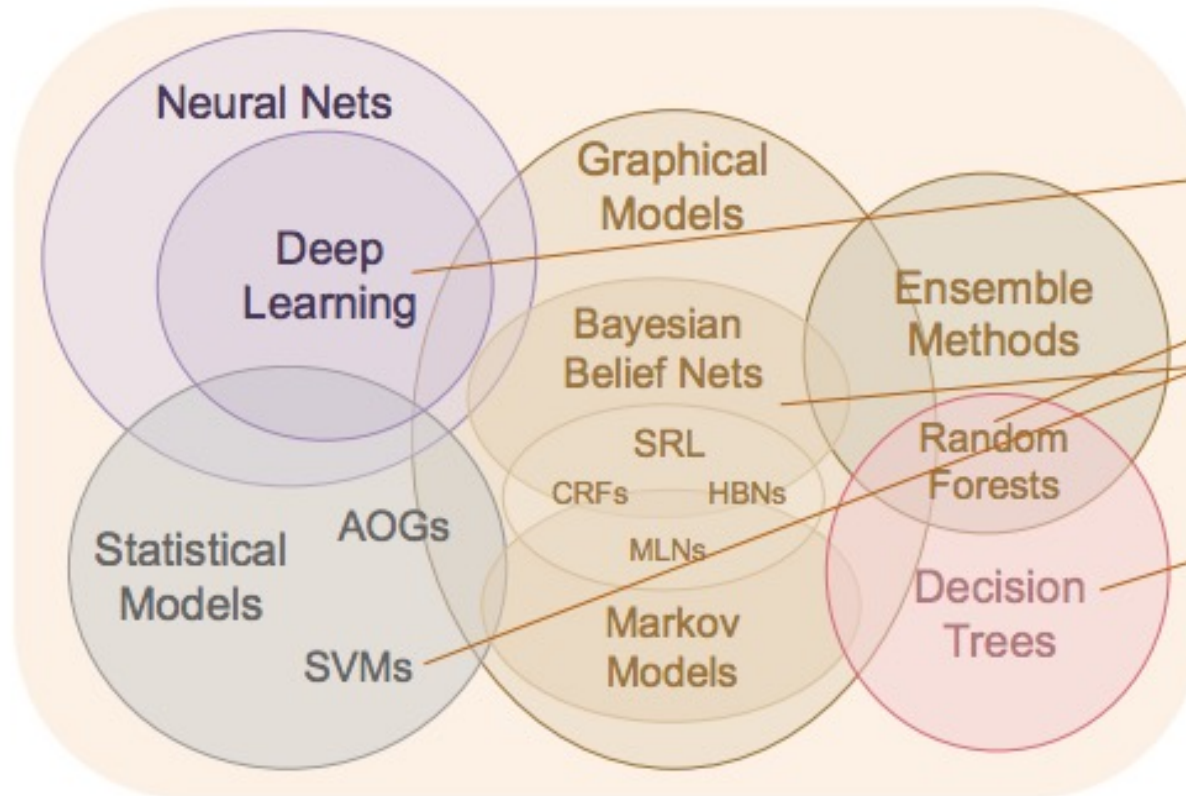
- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

FAT Deep Learning: Transparency

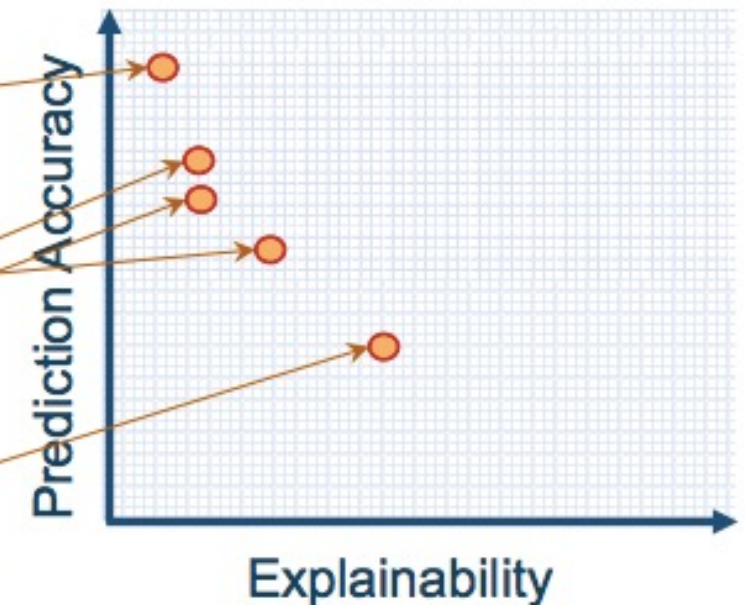
New Approach

Create a suite of machine learning techniques that produce more explainable models, while maintaining a high level of learning performance

Learning Techniques (today)



Explainability (notional)

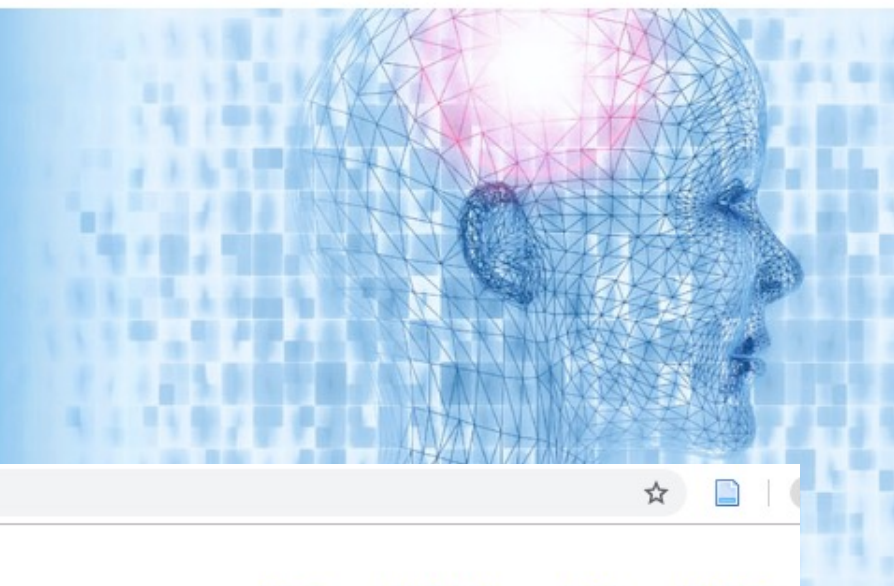


Industry (Facebook, Microsoft, & more...)

https://www.microsoft.com/en-us/research/group/fate/

Microsoft | Research Research areas Products & Downloads Programs & Events Careers People Blogs & Podcasts Labs & Locations All Microsoft Search

FATE: Fairness, Accountability, Transparency, and Ethics in AI



https://www.partnershiponai.org

 PARTNERSHIP ON AI


ABOUT PARTNERS NEWS CAREERS

"We need the best and the brightest involved in conversations to improve trust in AI and to benefit

Institutes

← → ↻ https://ethical.institute



 The Institute for Ethical AI & Machine Learning

[Home](#)

[Principles](#)

[AI-RFX Framework](#)

[Explainable AI](#)

[Newsletter](#)

[Contact us or Join](#)

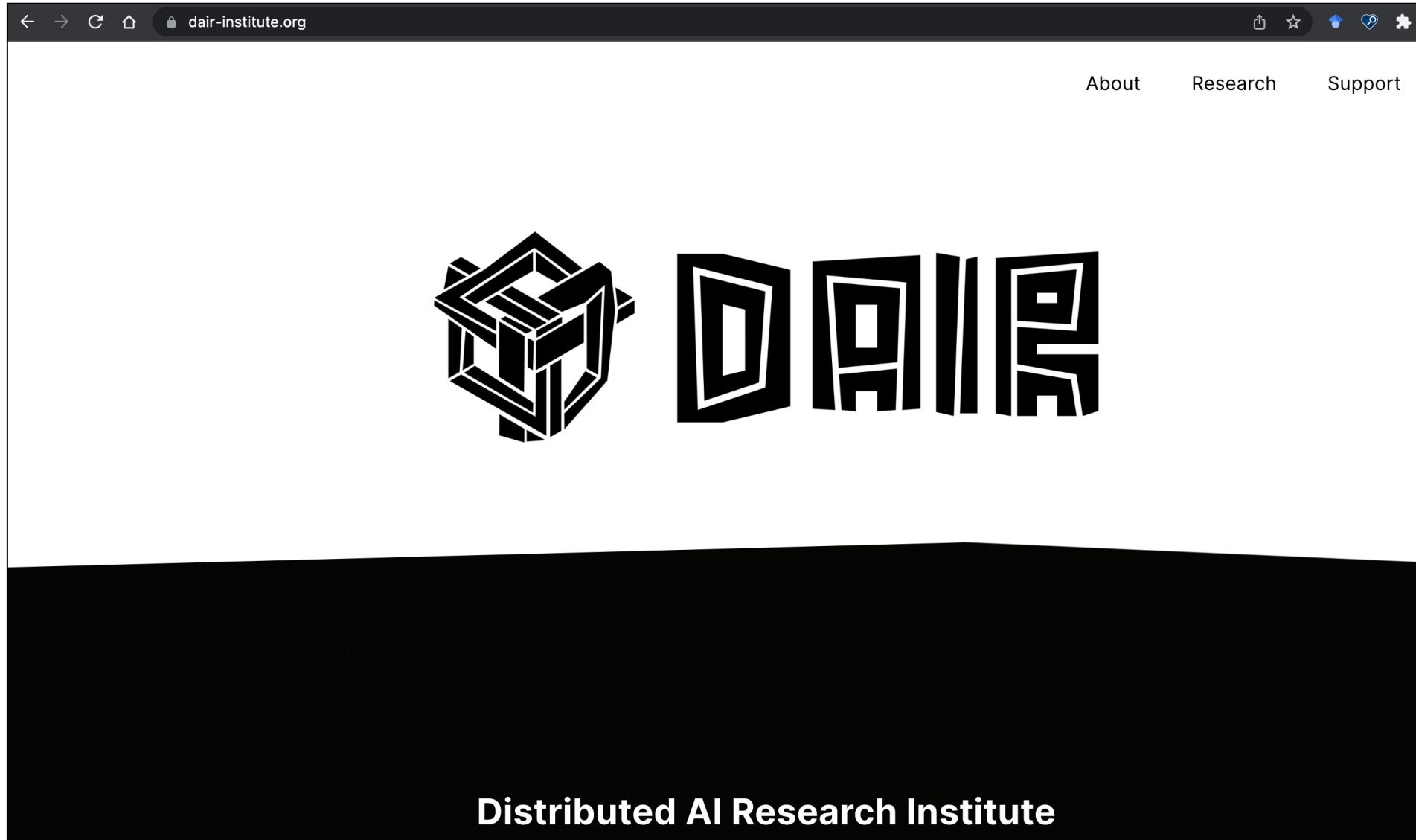


The Institute for Ethical AI & Machine Learning

The Institute for Ethical AI & Machine Learning is a UK-based research centre that carries out highly-technical research into responsible machine learning systems.

We are formed by cross functional teams of machine learning engineers, data scientists, industry experts, policy-makers and professors in STEM, Humanities and Social Sciences.

Institutes



Recent Work: Highlights from ICCV 2023

Gender Artifacts in Visual Datasets

DALL-EVAL: Probing the Reasoning Skills and Social Biases of Text-to-Image Generation Models

A Multidimensional Analysis of Social Biases in Vision Transformers

FACET: Fairness in Computer Vision Evaluation Benchmark

Laura Gustafson

Chloe Rolland

Nikhila Ravi

Quentin Duval

Aaron Adcock

Cheng-Yang Fu

Melissa Hall

Candace Ross

Meta AI Research, FAIR

facet@meta.com

Today's Topics

- Computer Vision that Discriminates
- FAT (Fair, Accountable, & Transparent) Algorithms
- **Ethics in Computer Vision**

We know that algorithms are not perfect.
Algorithms can be biased.

Are they ethical to use?

Time for a group activity!

Unacceptable to acceptable:
Using CV to diagnose diseases

Unacceptable to acceptable:
Using CV to tag names to people's faces

Unacceptable to acceptable:
Using CV to describe
someone's body shape/size

Unacceptable to acceptable:
Using CV to generate publicly-shared images

Unacceptable to acceptable:
Using CV to edit publicly-shared images

Unacceptable to acceptable:
Using data from public
websites to train CV models

Unacceptable to acceptable:
Open-sourcing vision foundation models

What other ethical issues can you think of around using computer vision algorithms?

Today's Topics

- Computer Vision that Discriminates
- FAT (Fair, Accountable, & Transparent) Algorithms
- Ethics in Computer Vision

A dark gray background with a white film strip border on the left and right sides. The film strip has rectangular sprocket holes. In the center, there is a faint, circular white glow. The text "The End" is written in a white, cursive script font with a slight drop shadow, centered within the glow.

The End