

# Ethical AI

---

Google Developer Student Club  
at William & Mary

Clare Heinbaugh, Emilio Luz-Ricca, & Charlie Dobson



**Good  
Data**

---

**>**

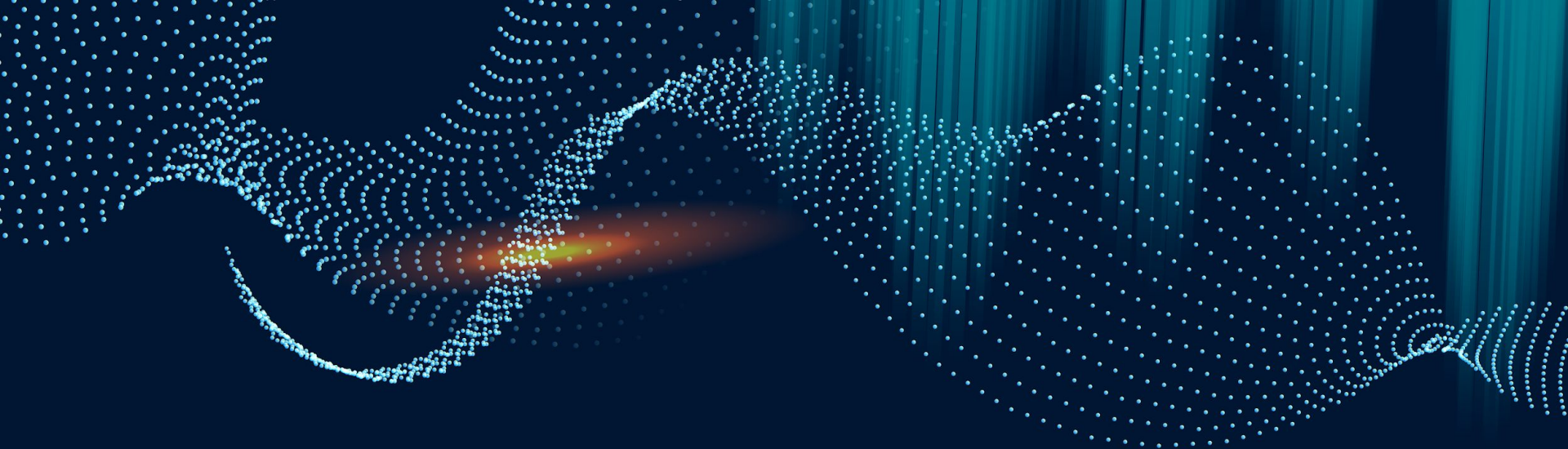
**Shitty  
Data**

---

**>**

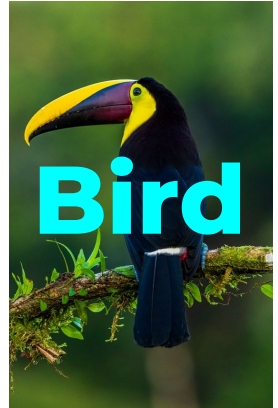
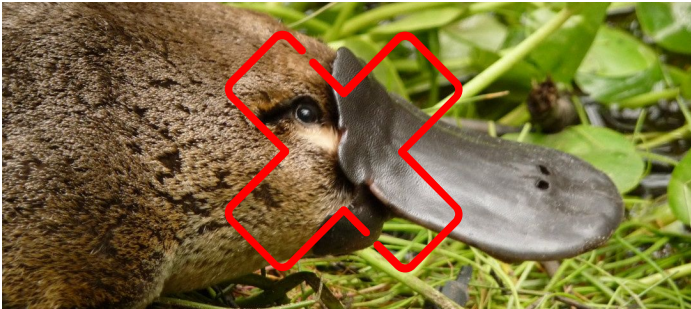
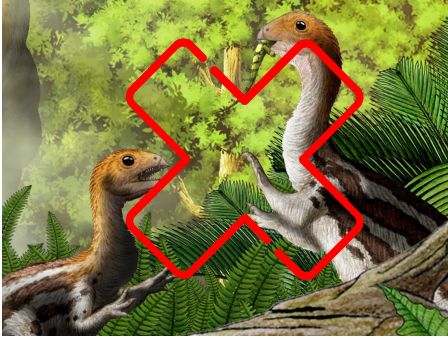
**No  
Data**

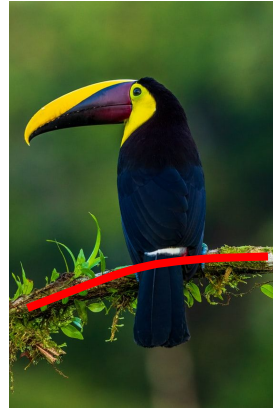
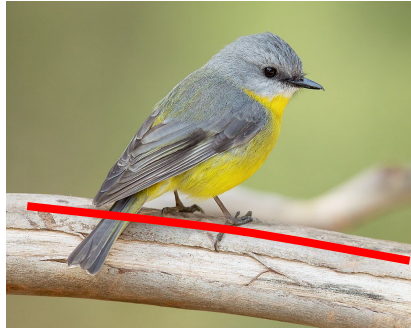
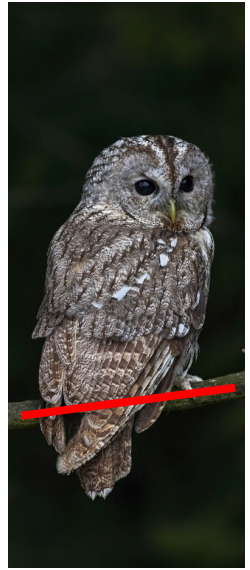
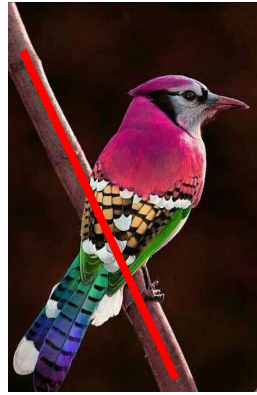
---



**1**

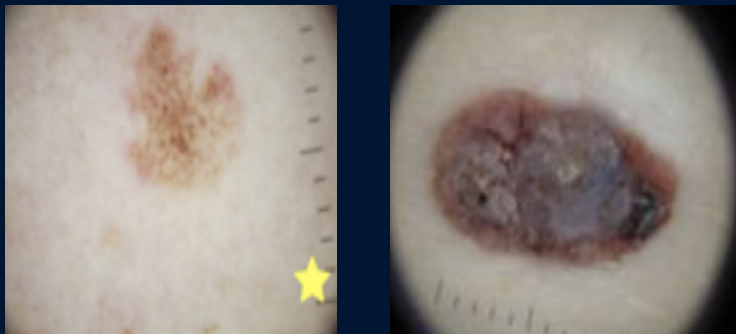
# Model Shortcuts



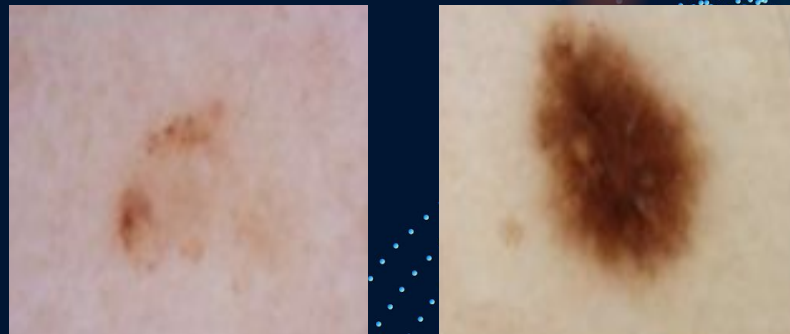


# Detecting Skin Cancer

**Cancerous**



**Non-cancerous**



129,450 images  
21 dermatologists

---

## Detecting Skin Cancer → Hidden Paths

Want: Image of Cancer Cell → Classify as Cancer Cell

$$X \rightarrow Y$$

Actual: Image of Cancer Cell ← Ruler in Image → Classify as Cancer Cell

$$X \leftarrow Z \rightarrow Y$$



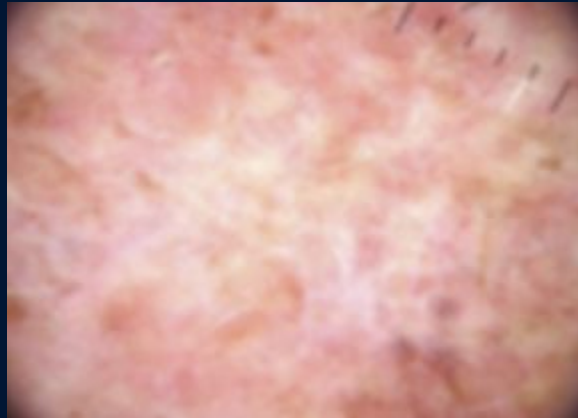
**“The hidden confounder  
creates a hidden class”  
and thus a “shortcut  
solution.”**

—Professor Shao, NASEC

---

## Detecting Skin Cancer → Stakeholders

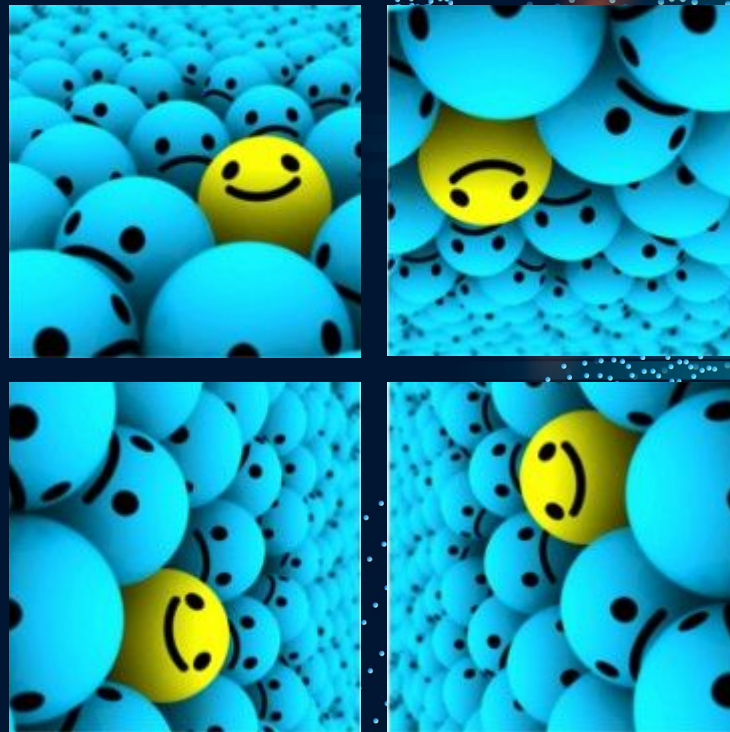
- Patients
- Doctors
- Healthcare providers



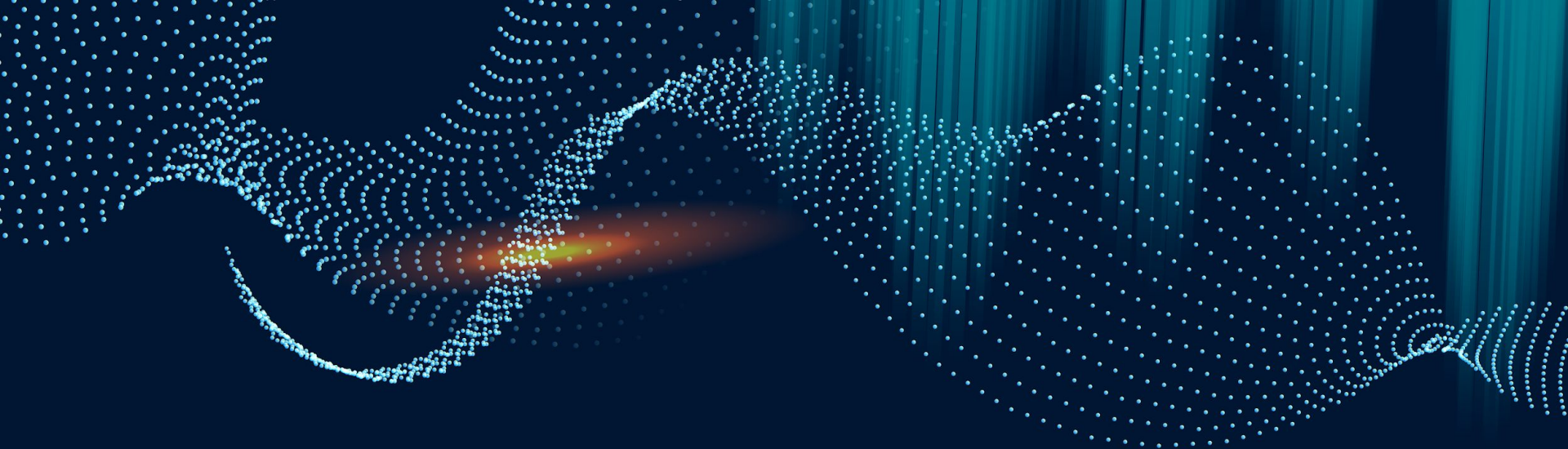
---

# Detecting Skin Cancer → Recommendations

- To Dataset Curators
  - Careful curation of data
- To Coders
  - Invariant risk minimization
- To Policymakers
  - Require medical AI to be tested by multiple companies in many different areas
  - Require Explainable AI



Data Augmentation



**2**

**Human Bias →  
Model Bias**



**Convert sign  
language to speech**

**Judge criminal cases  
in an unbiased way**

**Analyze resumes to hire  
the best candidate**

**Self-reflect**

**Count crowds**

**Win jeopardy against  
top jeopardy players**



**Convert sign  
language to speech**

**Judge criminal cases  
in an unbiased way**

**Analyze resumes to hire  
the best candidate**

**Self-reflect**

**Win jeopardy against  
top jeopardy players**

**Count crowds**



**Convert sign  
language to speech**

**Judge criminal cases  
in an unbiased way**

**Analyze resumes to hire  
the best candidate**

**Self-reflect**

**Win jeopardy against  
top jeopardy players**

**Count crowds**

**Human bias → Model bias**





---

# Facial Recognition

- Biases about gender, race, etc. are coded into the model.
- Facial recognition software struggles most on people who are:
  - Darker-skinned
  - Female-identifying
  - Very young
  - Very old
- Privacy concerns



# Facial recognition

## ERROR RATE

	<b>Microsoft</b>	<b>IBM</b>
Lighter Male Faces	0.0%	0.3%
Darker Female Faces	20.8%	34.7%

---

# Facial Recognition

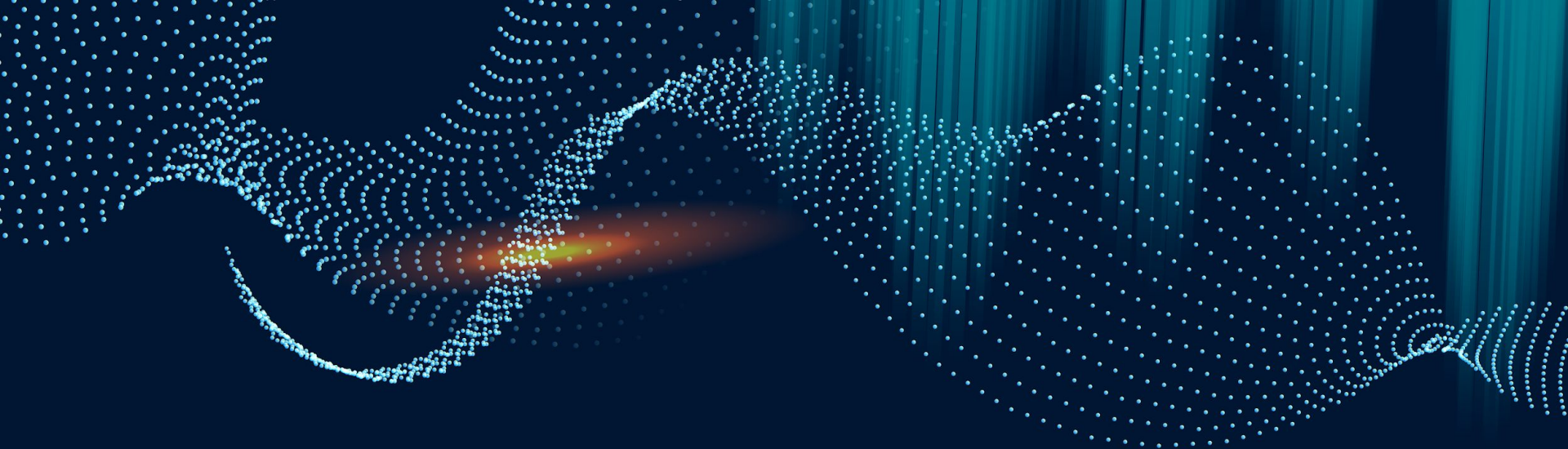
**87%**

One model's  
accuracy in a sports  
venue study

**40%**

Half of the  
models in the  
same study





**3**

**Weapons of  
math destruction**

# Google Photos' Auto-tagging Feature

- Deep neural networks
- Identify patterns, label images



# What's wrong with this picture???



---

## Sources

- [NASEC 2021](#)
- [Towards Data Science](#)
- [Nanonets](#)
- [Face in Video Evaluation](#)
- [Reuters](#)
- [Technology Review](#)
- [Harvard Science in the News](#)
- [U.S. Government Accountability Office](#)
- [Machine Learning Research Conference](#)