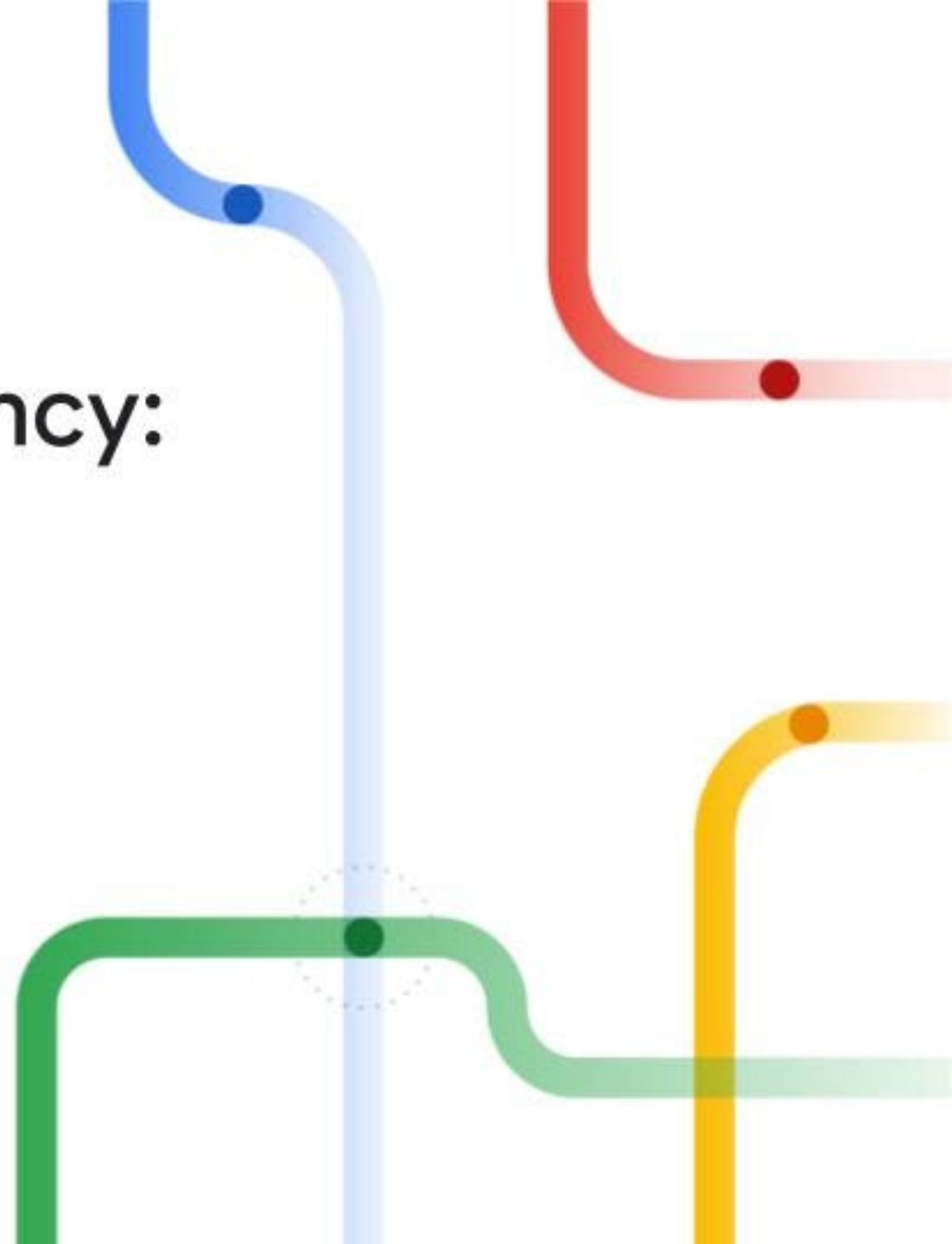


Designing for Transparency: Consumer-oriented AI

Samantha Winter, PhD, UX Researcher
Lily Peng, MD, PhD, Product Manager



Transparency is a necessity



Designing AI-informed products for non-expert users



Transparency patterns



Global explanation

Address AI's purpose, capabilities, and limitations



Local explanation

Why this prediction, and how to act on it

01

Set expectations about AI uses, capabilities, and limitations



Model cards set expectations before implementation



Overview

Limitations

Trade-offs

Performance

Test your own images

Provide feedback

Explore

Object Detection

About Model Cards

<https://modelcards.withgoogle.com/face-detection>

Face Detection

The *model* analyzed in this card detects one or more faces within an image or a video frame, and returns a box around each face along with the location of the faces' major landmarks. The model's goal is exclusively to identify the existence and location of faces in an image. It does not attempt to discover identities or demographics.

On this page, you can learn more about how well the model performs on images with different characteristics, including face demographics, and what kinds of images you should expect the model to perform well or poorly on.

MODEL DESCRIPTION



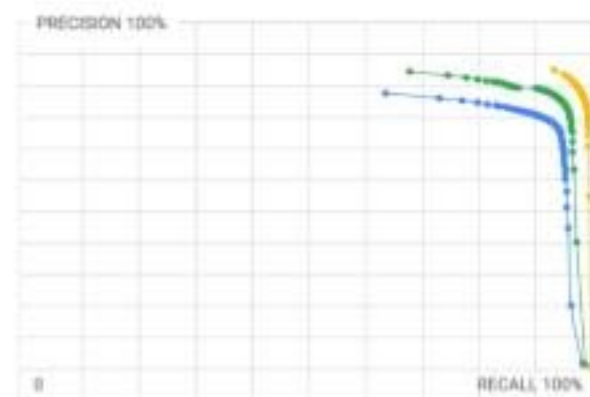
Input: Photo(s) or video(s)

Output: For each face detected in a photo or video, the model outputs:

- Bounding box coordinates
- Facial landmarks (up to 34 per face)
- Facial orientation (roll, pan, and tilt angles)
- Detection and landmarking confidence scores.

No identity or demographic information is detected.

PERFORMANCE



- Open Images
- Face Detection Dataset Benchmark
- Labeled Faces in the Wild

Overall model performance, and performance sliced by different image and face characteristics, were assessed, including:

- Derived characteristics (face size, facial orientation, and occlusion)
- Face demographics (human-perceived gender presentation,

Onboarding to how AI works and how to work with AI



Training insights

What examples did the model see in training?



Mental model

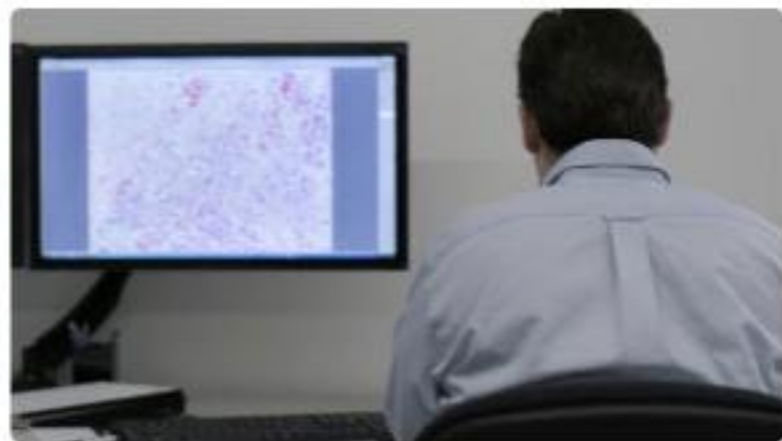
What perspective should a user take?



Capabilities & limitations

Where does the AI perform best?
Where does it not?

Aim for human-relatable explanation



02

Explain the AI at the prediction level



Multiple methods for prediction-level explanations

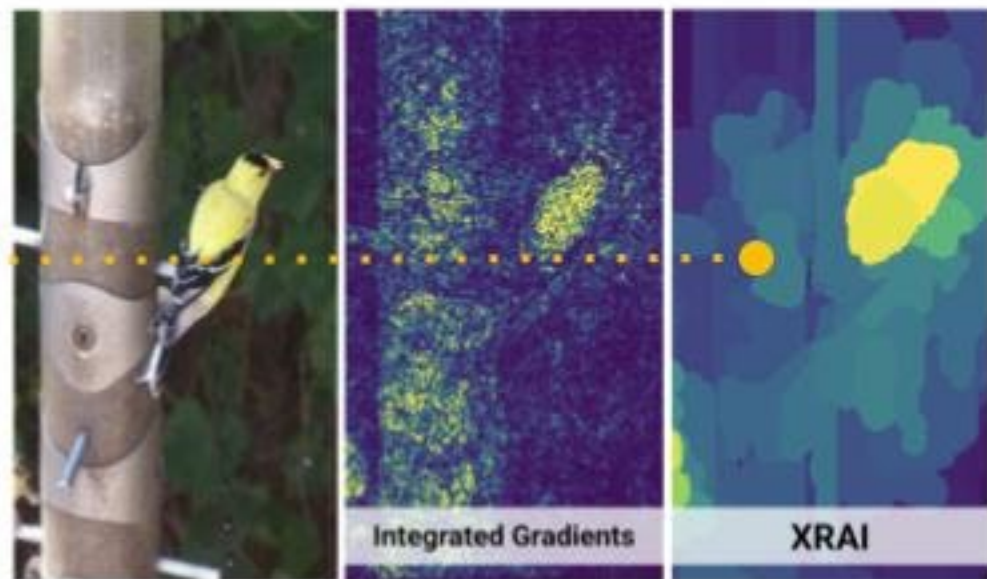
Confidence scores

Saliency

- ↳ Guided IG (Guided integrated gradients)
- ↳ XRAI (eXplanation with Ranked Area Integrals)
- ↳ SmoothGrad

Similar examples

Descriptive text

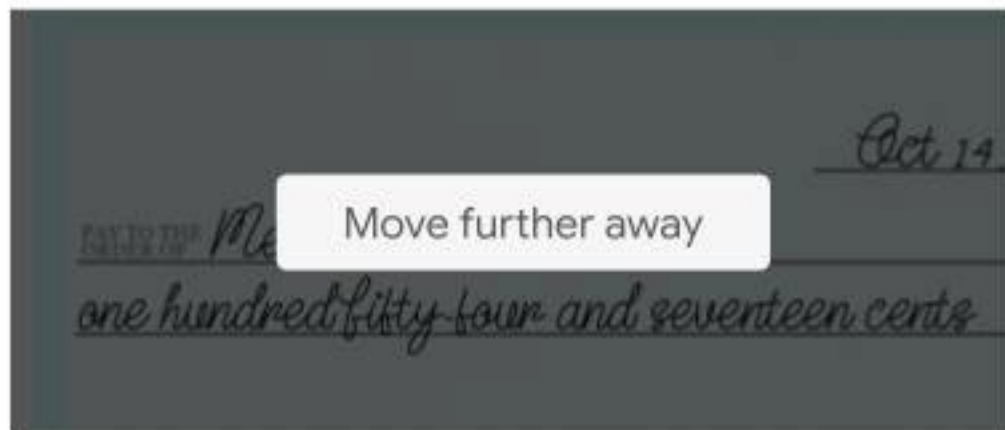


<https://pair-code.github.io/saliency>

Explain for understanding, not completeness

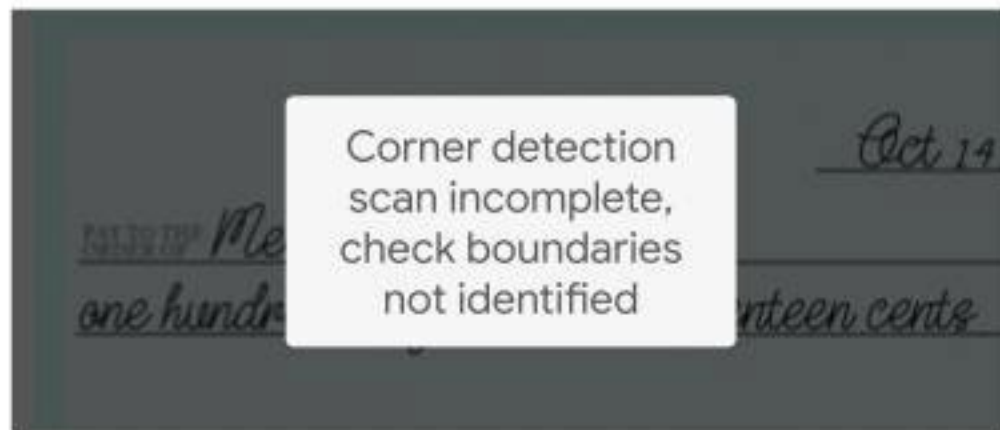
Aim for

Decision-making and trust



Avoid

Explaining the entire system



03

Putting it all together

Embed explanation at the global and local level for effective and safe use.

Thank you

Samantha Winter, PhD, UX Researcher
Lily Peng, MD, PhD, Product Manager