WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

AI Models On Mobile Phones: What they do, Why they're there, and How They Work

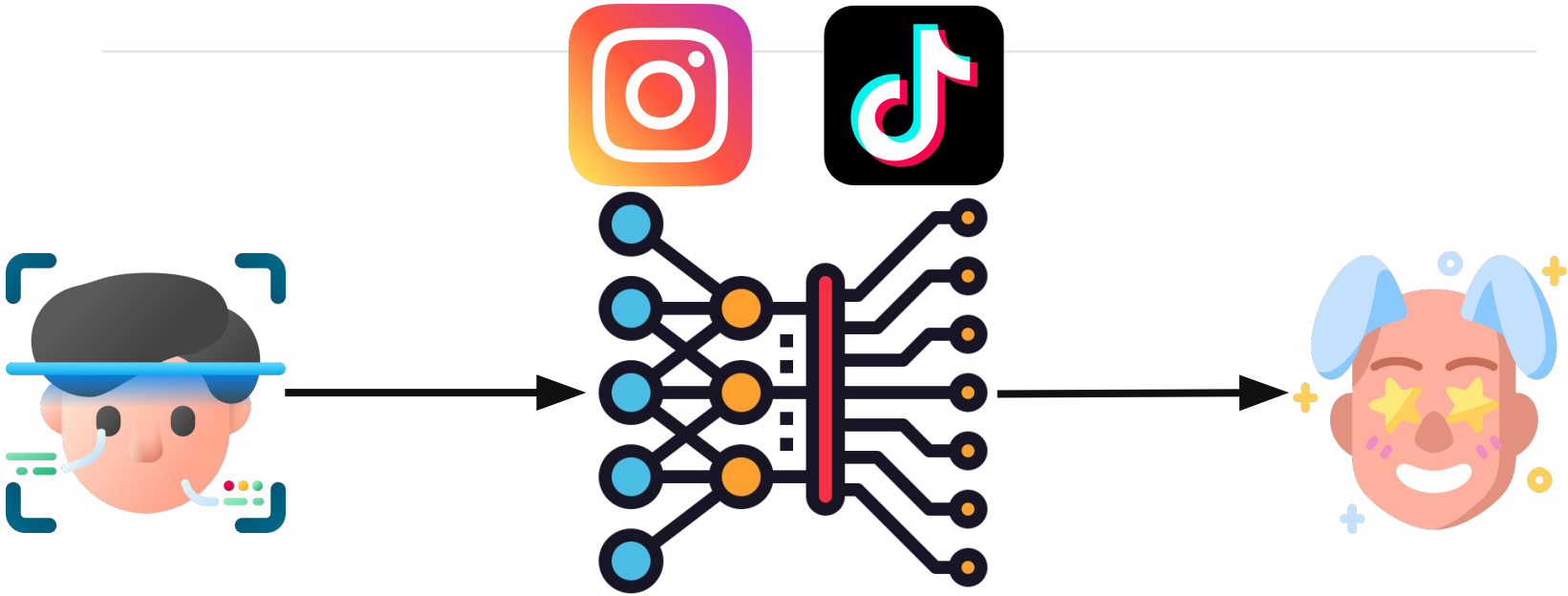
Jack West

03/21/2025

AI/Machine Learning (AI/ML) in Social Media

AI in Social Media: Enhancing User Experience, Content Moderation, and Personalization

by Potential Staff Published on October 29, 2024



Overview of My Work

What are the risks of local AI/ML to users?

Bias **Fairness**

How do social media users feel about local AI/ML?

Awareness **Impact**

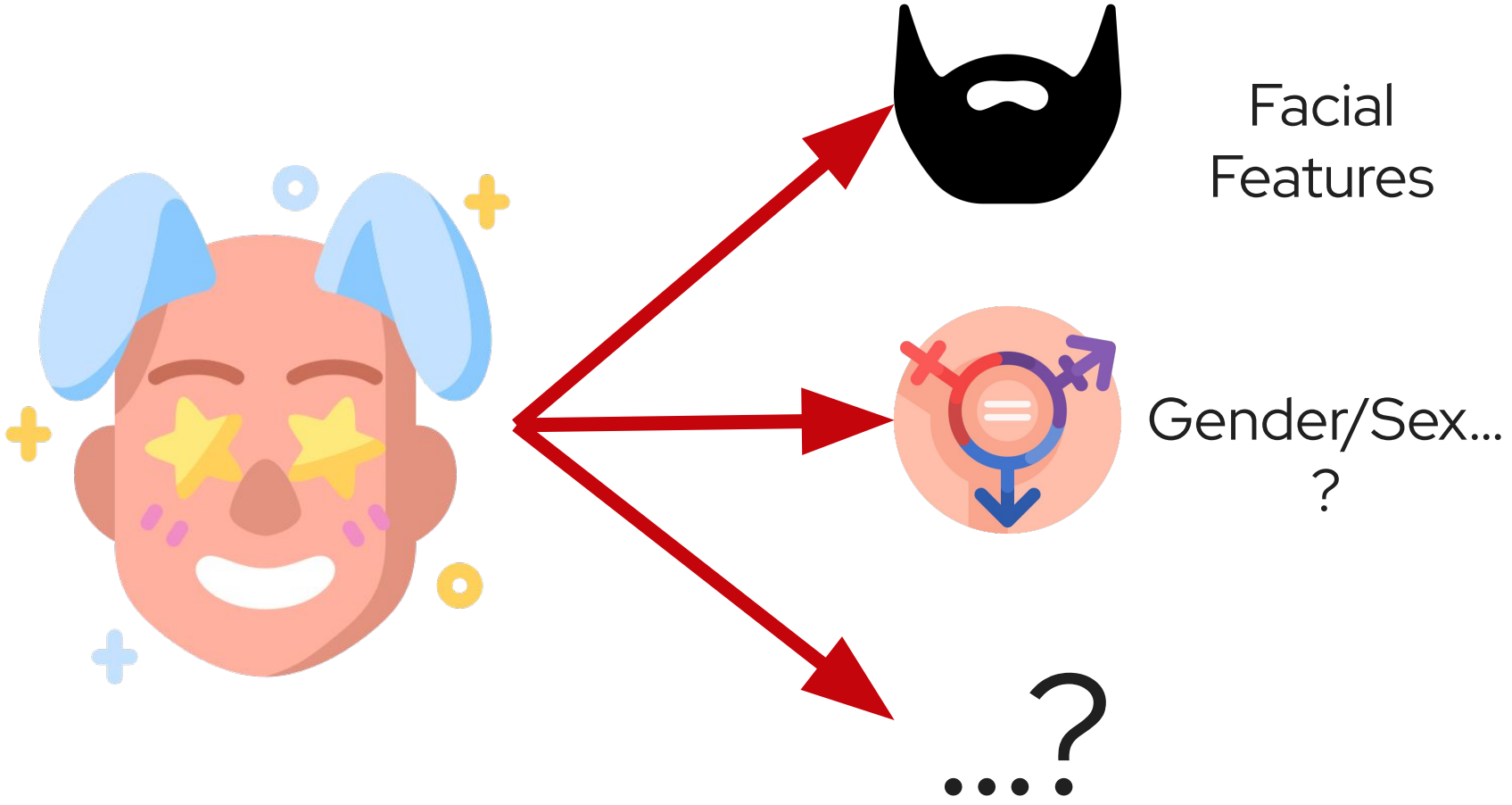
What are the risks to AI deployers?

AI Security **AI Privacy**

Finding the On-Device Models

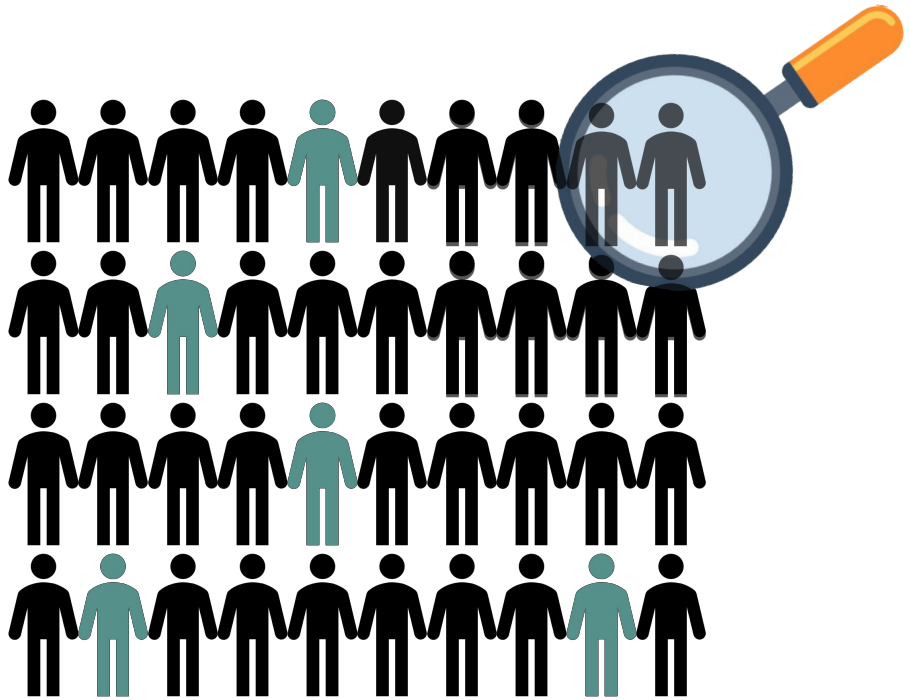
Symposium on Security and Privacy 2024

Model Capabilities

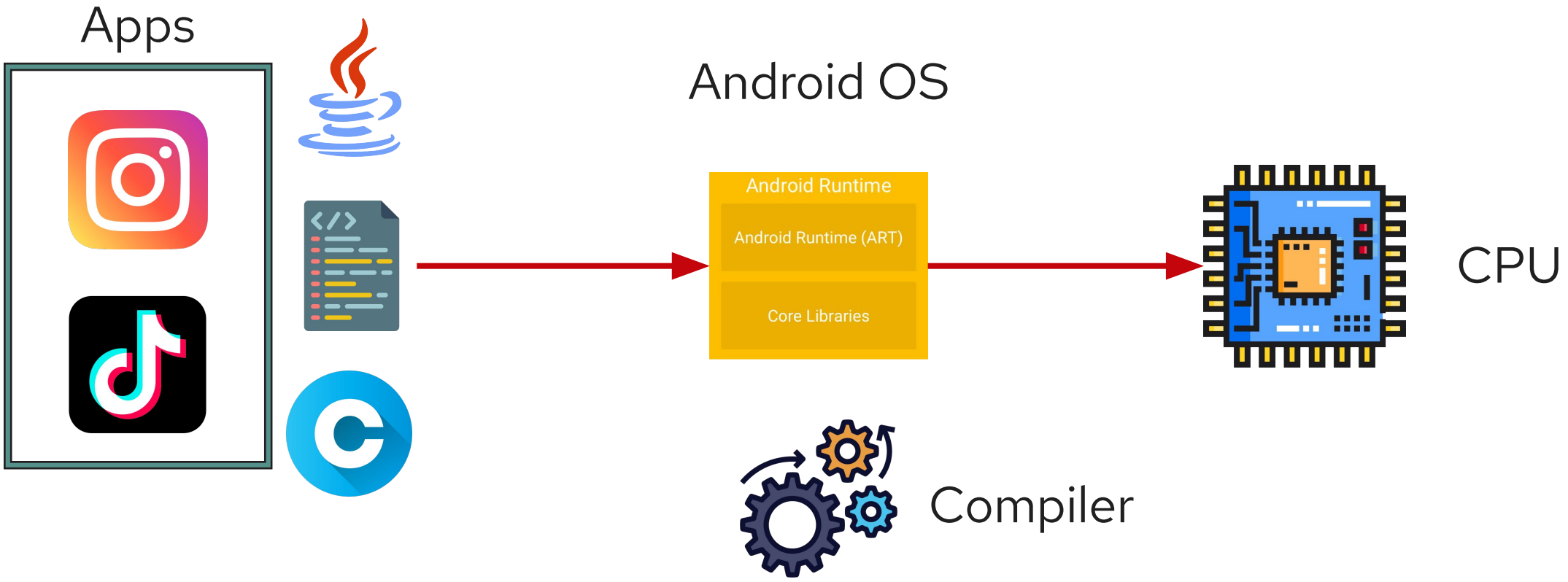


Research Questions

1. What insights do vision models in TikTok and Instagram infer about users from their images and camera frames?
2. Are there demographic disparities in the quality of the inferred insights?



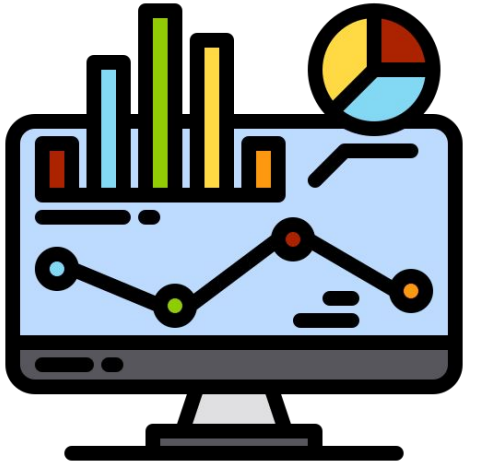
Android App Execution Flow



Challenges

Obfuscation

Native Libraries



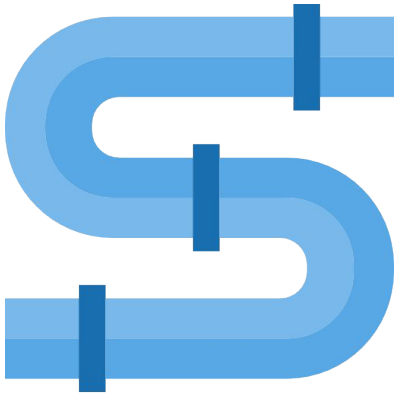
Model Evaluation

Overview

Detection

Dynamic analysis using custom OS.

Pipeline Reconstruction

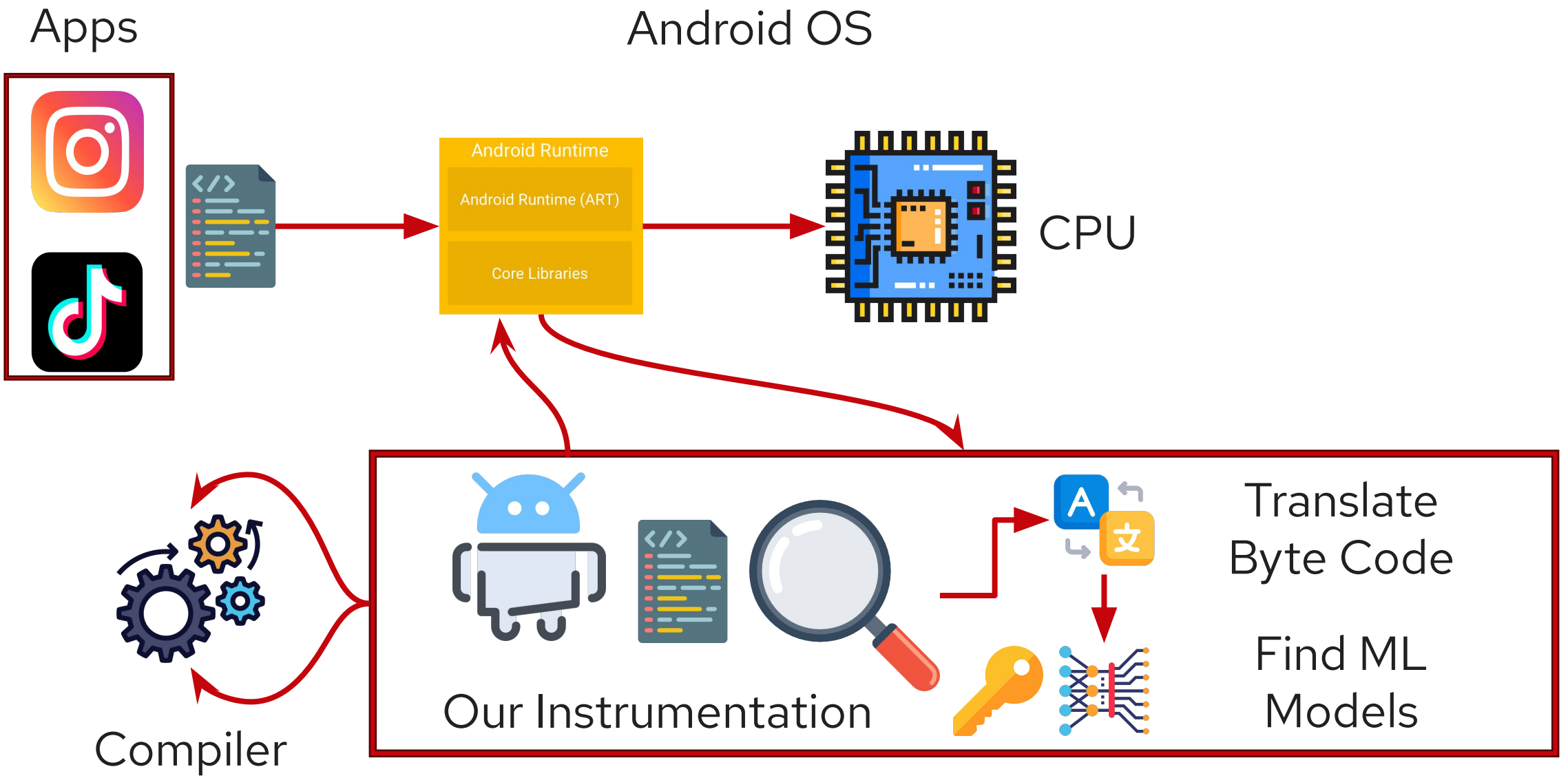


Static analysis to reconstruct pipelines.

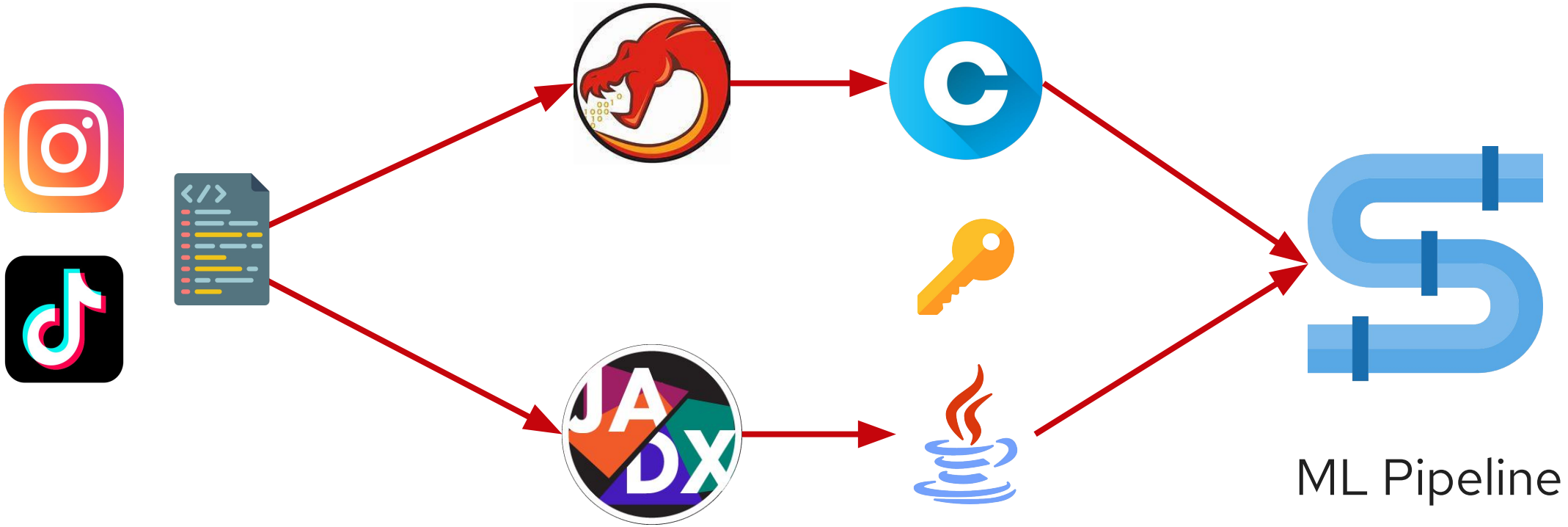
Model Evaluation

Evaluate model performance.

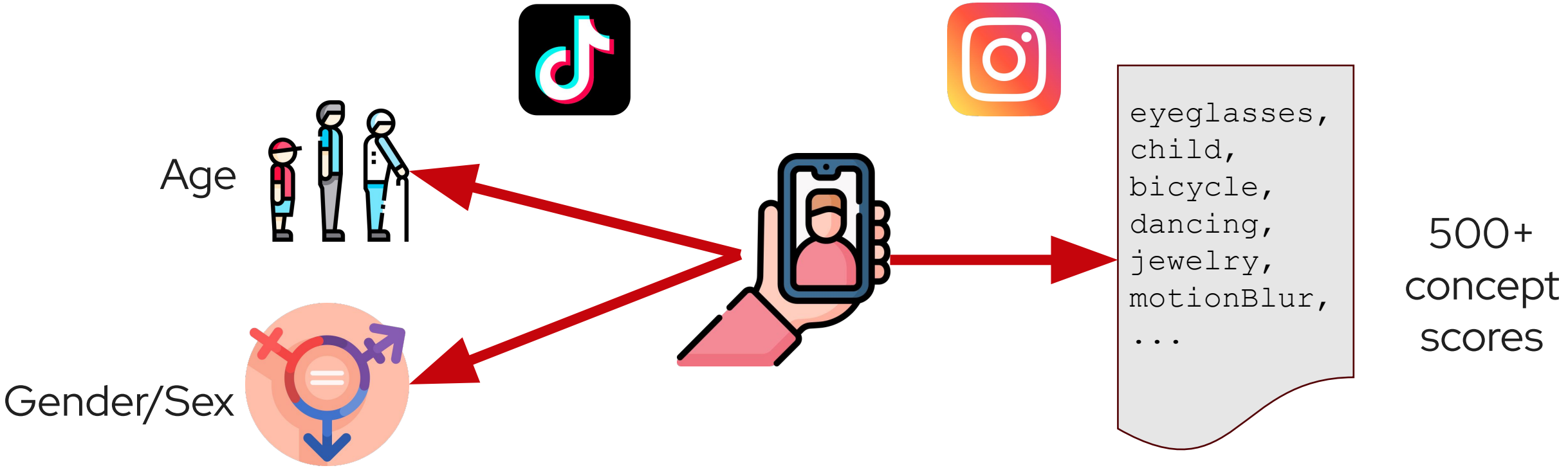
Detection


Apps
Android OS
Android Runtime
Android Runtime (ART)
Core Libraries
CPU
Translate
Byte Code
Find ML
Models
Our Instrumentation
Compiler

Pipeline Reconstruction

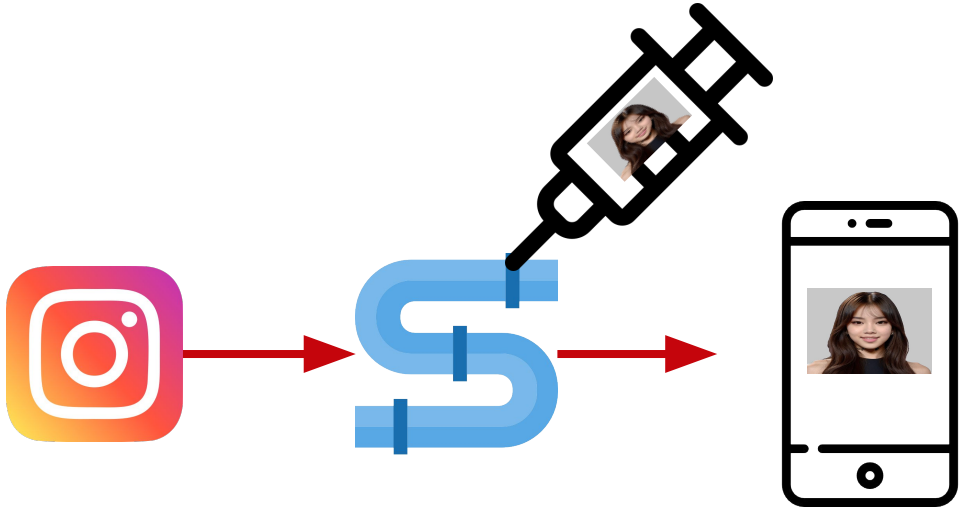


Results: Model Capabilities



Evaluation: Injecting Datasets

Internal Injection

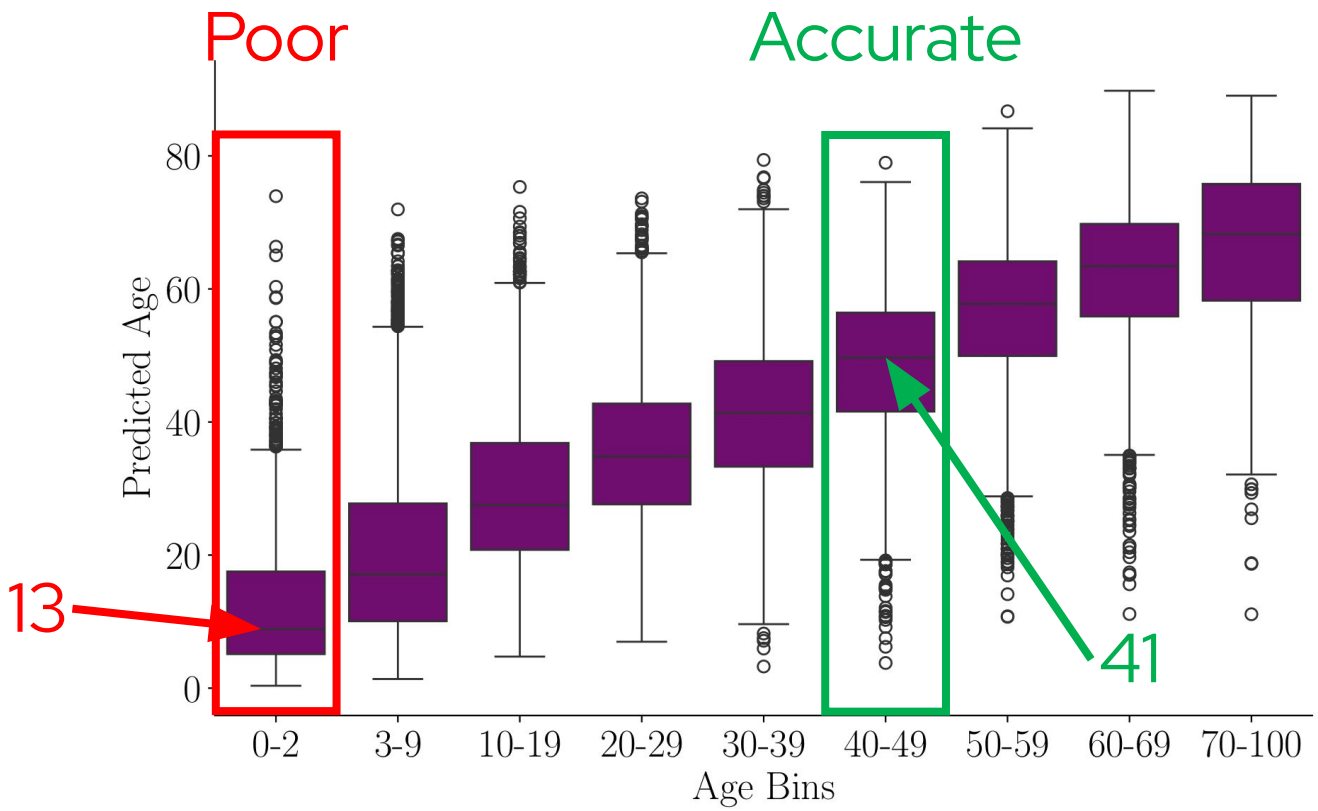


Custom dataset injected directly into the pipeline

External Injection

FairFace dataset displayed on an external monitor

TikTok: Age Estimation



TikTok's model predicts that younger users are older.

Instagram: Spurious Correlations

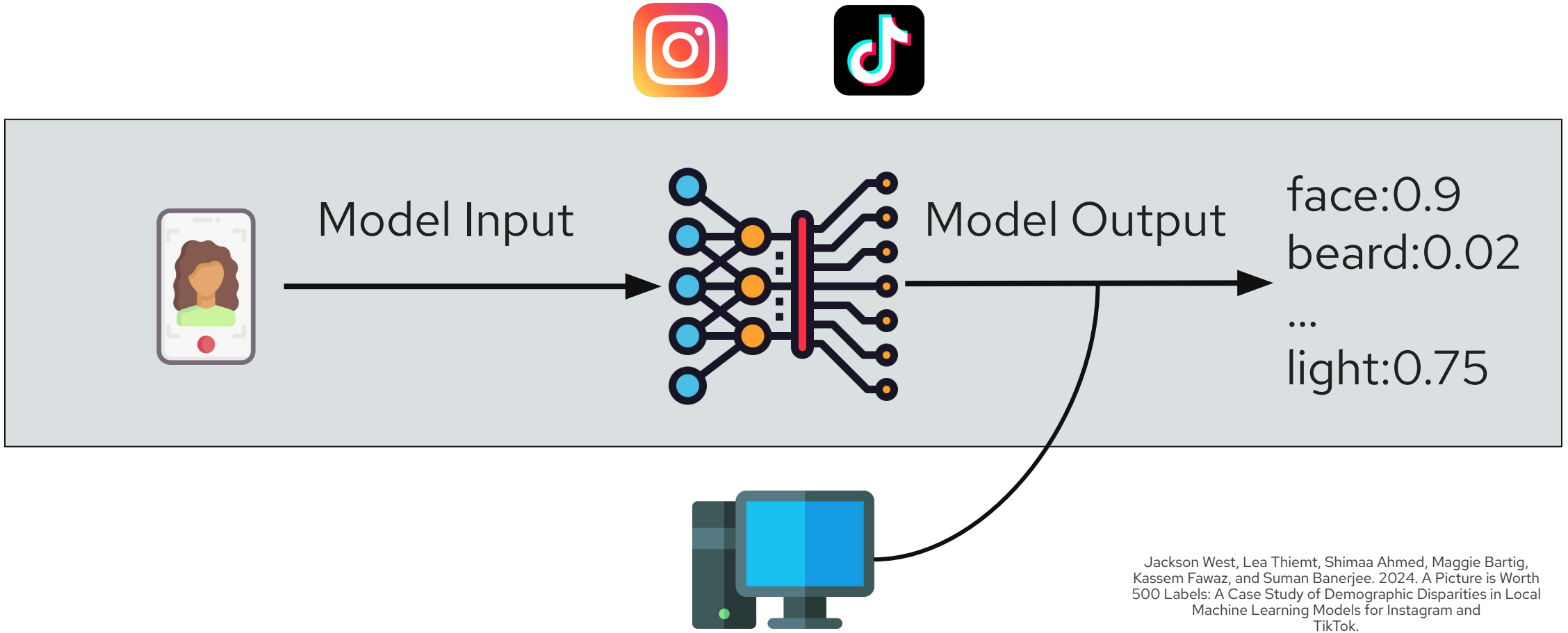
Demographic Group	Associated Concept
Asian Man	**eyeglasses**, bbq_barbecue, sansevieria, dais
Asian Women	**great_wall_of_china**, reading, sports_field, **wine**, colHarmony
Black Man	rabbit, **teamaker**, carving, **nighttime**, outdoor, suiting, **brass**, cloud
Black Women	video_game, bakken, **drag**, light, aesthetics_rating
Indian Man	grass, **beard**, skydiving, people, face, **driving**
Indian Women	opening_champagne, jewelry, **watchstrap**, hair_long, dress, coffee, cloche
White Man	**businesssuit**, water, indoor, activewear, sky, aviation, eyewear, zoo, **nudity**
White Women	diningroom, blond, interior_design, fineart, art_painting, hair, **blue**, **blonde**

Concepts are given from Instagram. Above are the concepts that are significant to each demographic group.

User Awareness of Local AI

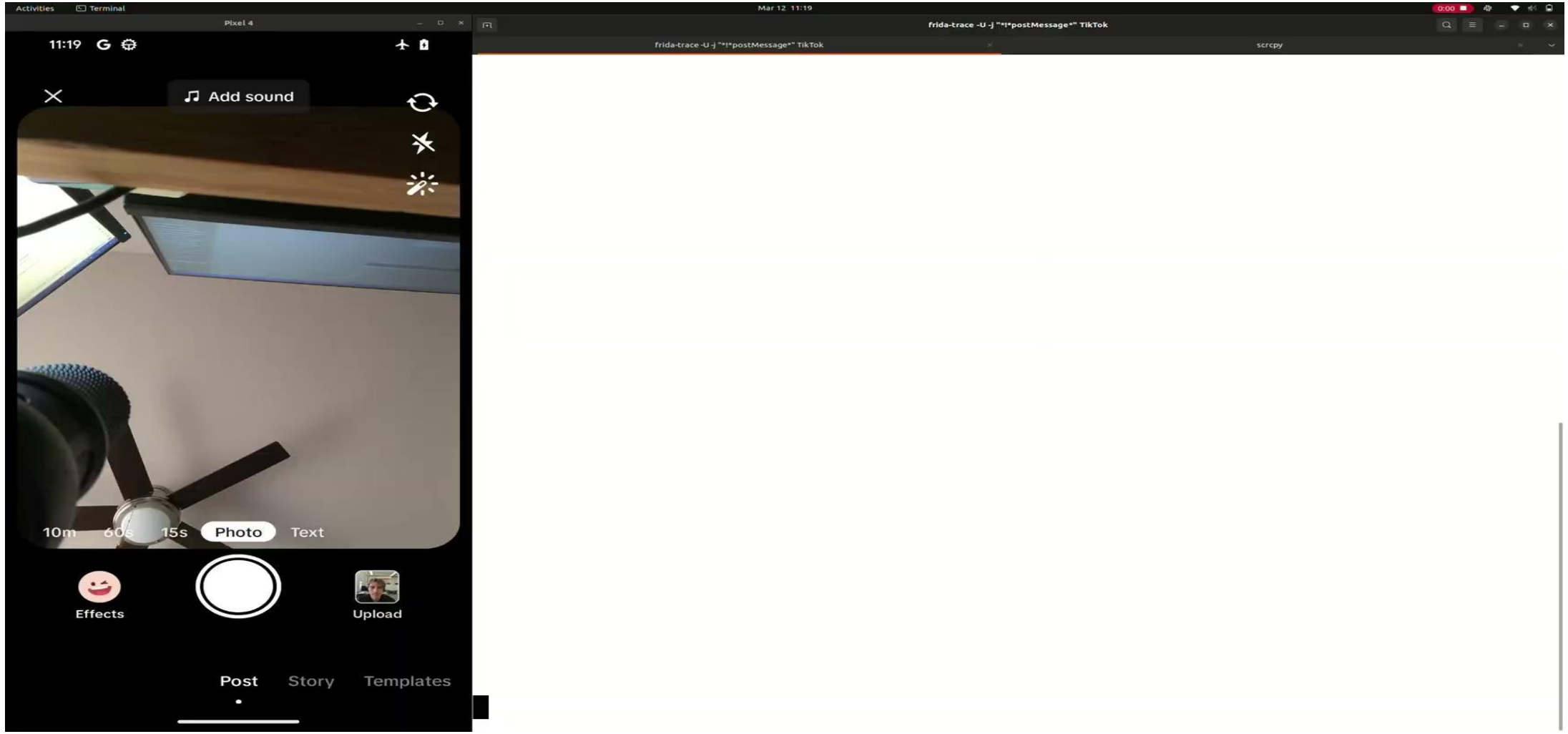
CHI 2025

Defining Our Scope With Authentic AI/ML Models

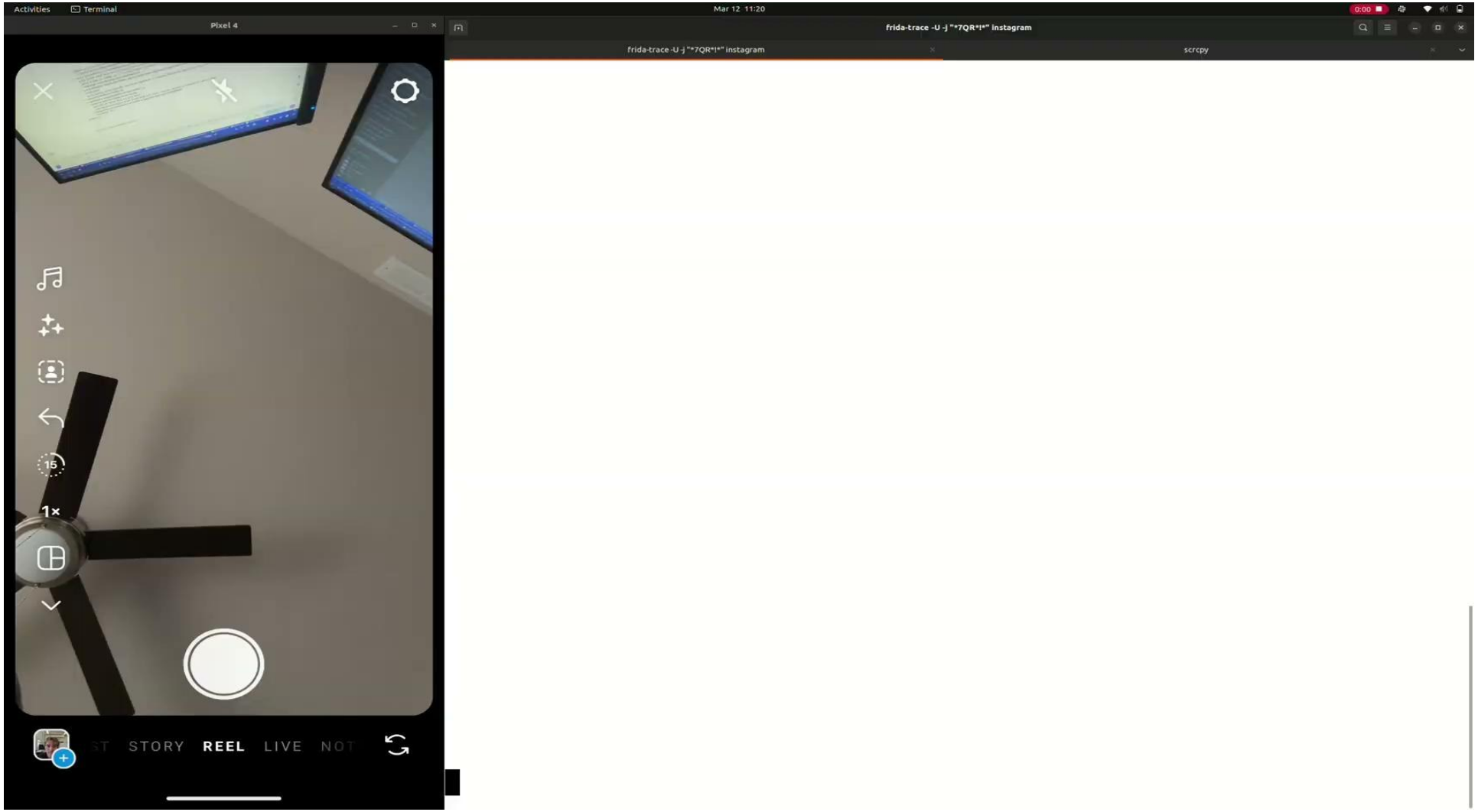


Jackson West, Lea Thiemt, Shimaa Ahmed, Maggie Bartig, Kassem Fawaz, and Suman Banerjee. 2024. A Picture is Worth 500 Labels: A Case Study of Demographic Disparities in Local Machine Learning Models for Instagram and TikTok.

Demonstration (TikTok)



Demonstration (Instagram)



Research Questions

What are users' *prior assumptions* about AI/ML?

What are users' *immediate reactions* to real models?

Do self-reported habits *change* after exposure?

Interactive In-person Interview

Overview of Participants

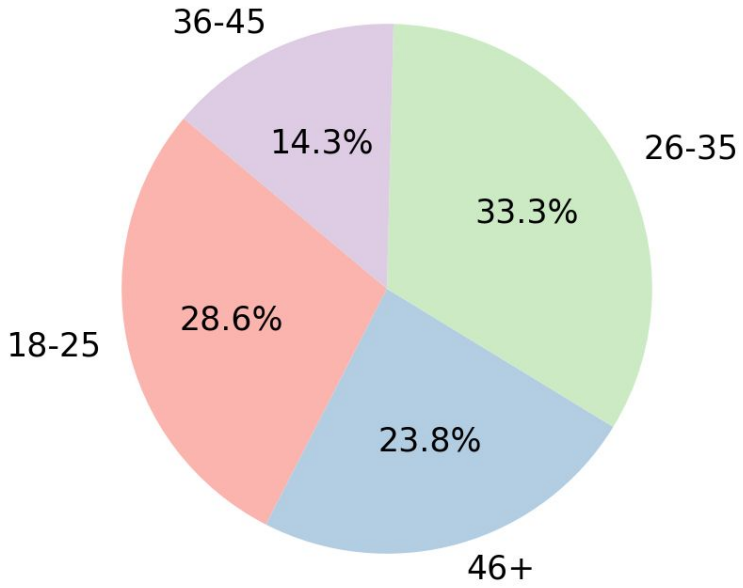
We interviewed **N=21** participants (out of 800+ volunteers)

Requirements to Participate

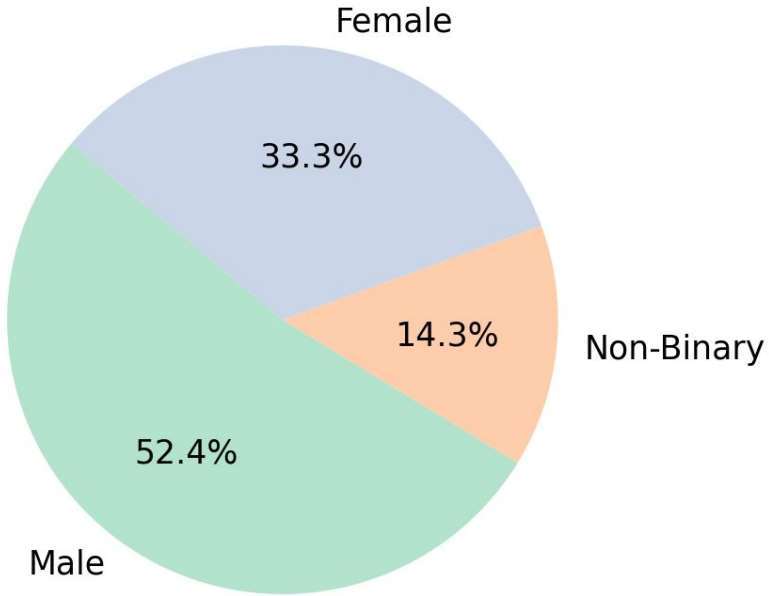
1. Must be a user or familiar with Instagram and TikTok
2. Must be able to attend in person
3. Is at least 18 years old.

Participant Statistics

N=21

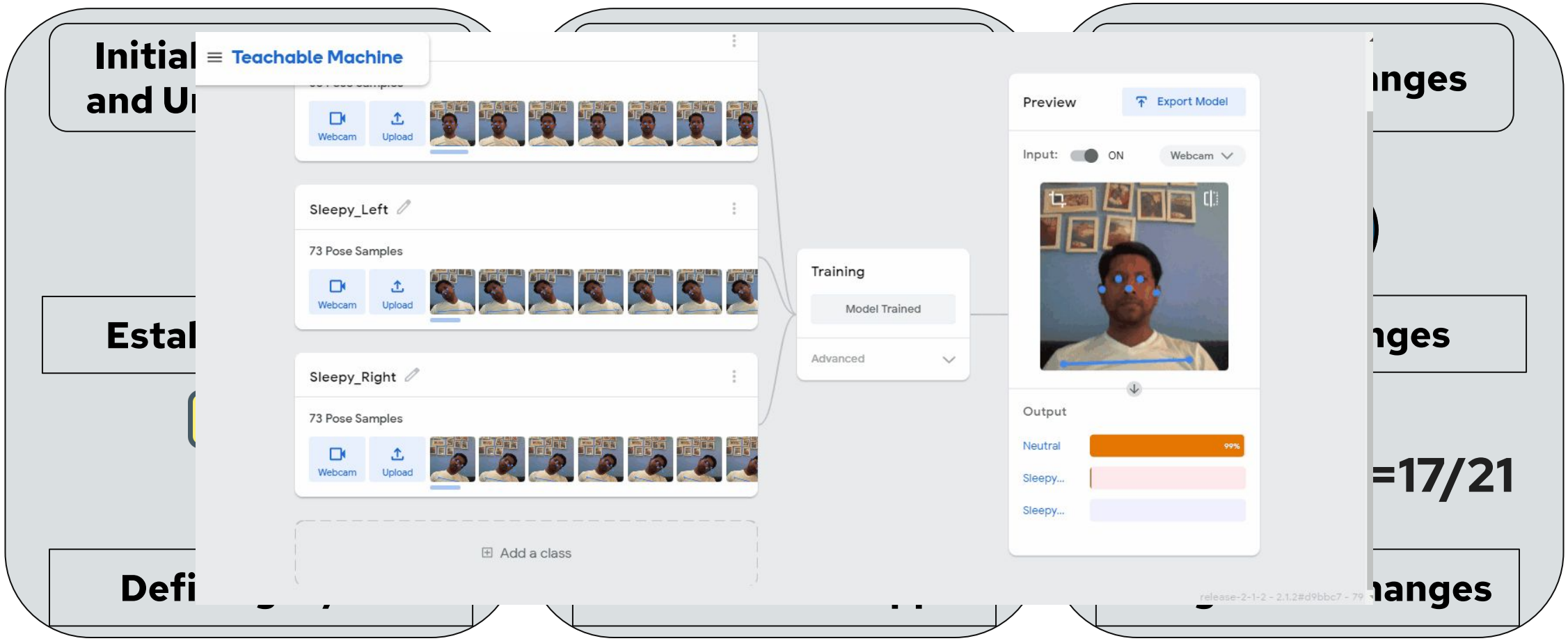


Age Distribution



Gender Identity Distribution

Semi-Structured Interview



Himanshu Chandra. A Fun Project (Pose Detector) With Google's Teachable Machine --- medium.com.
https://medium.com/analytics-vidhya/a-fun-project-pose-detector-with-googles-teachable-machine-6c7c8d650be1

Emergent Themes

The Purpose of AI/ML in Social Media

AI/ML is for Advertising

"**... I think part of their algorithms or models would especially now be driven toward what would cause me to purchase things or what they should advertise to me.**" - P15

N=15/21

AI/ML is for Functionality

"**So yeah, like nudity is one of them up there. It's probably like, you know, you kind of have to like regulate that.**" - P14

N=4/21

Reactions to Model Implementations

Implementation Matters!

TikTok

"**no way of knowing what it's picking up on, what it's storing...and what it's doing without [their] consent?!**" - P20

Invasive!

Instagram

"**Instagram is being a little more discerning and letting you do some selection before [the model] runs**" - P18

Gives users agency!

Reactions to Model Output

TikTok

"**[TikTok is] collecting far less data and maybe, maybe less harmless data.**" - P4

Data doesn't seem harmful!

Instagram

"**[Instagram is] trying to understand more about what's in the picture, but [they] don't really have an explanation for 'why'.**" - P20

Too confusing!

Reported Changes During the Interview

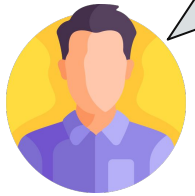
Won't Change

"**for the entertainment aspect for sure. And this is like the distraction in my life. And I guess I'll just keep [using Instagram].**" - P9

N=15/21

Will Change

"**[when using social media] posting pictures, starting a record, or just opening a camera. I'll be more conscious before even opening the app.**" - P7

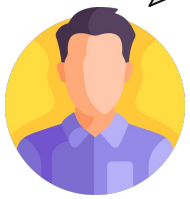


N=6/21

Some Participants Changed Their Mind Later

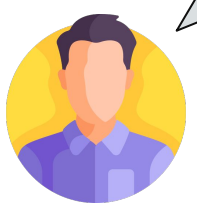
During the Interview

"**feel like I've already been compromised, per se. And I don't know if it would make too much of a difference if I stop now.**" - P4

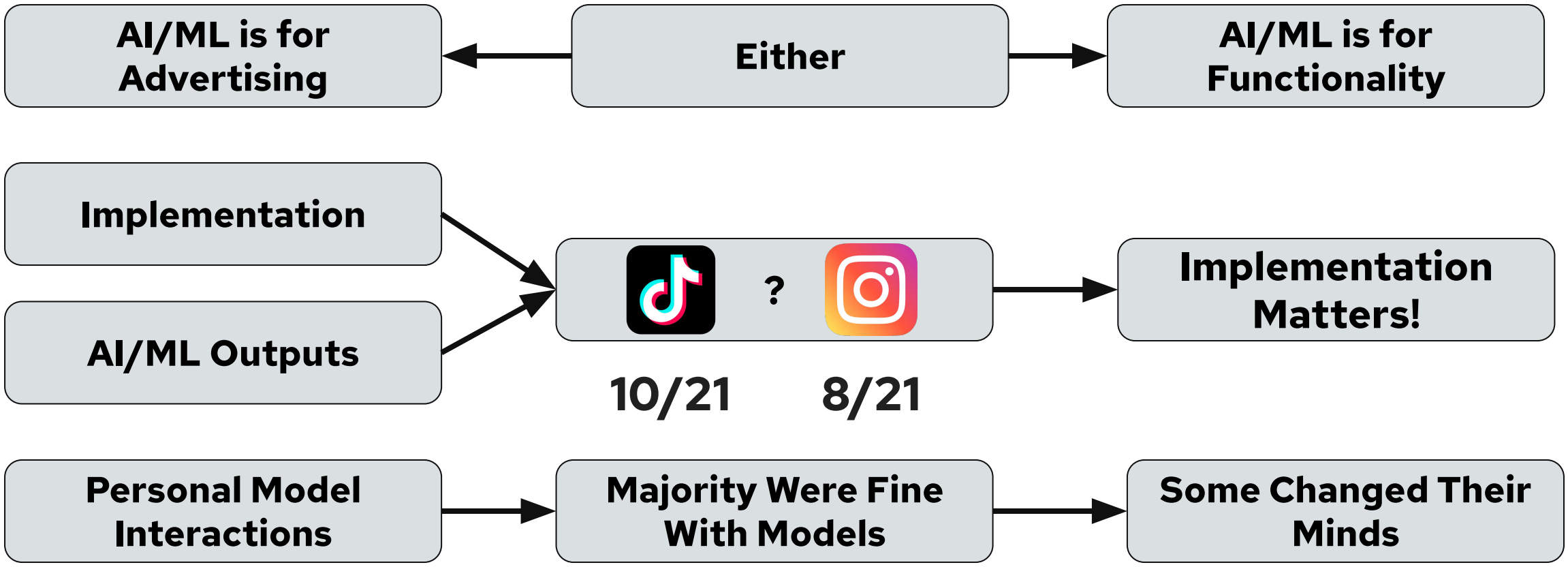


Two Weeks Later

"**[they] reduced permissions for these apps to access personal data**" - P4



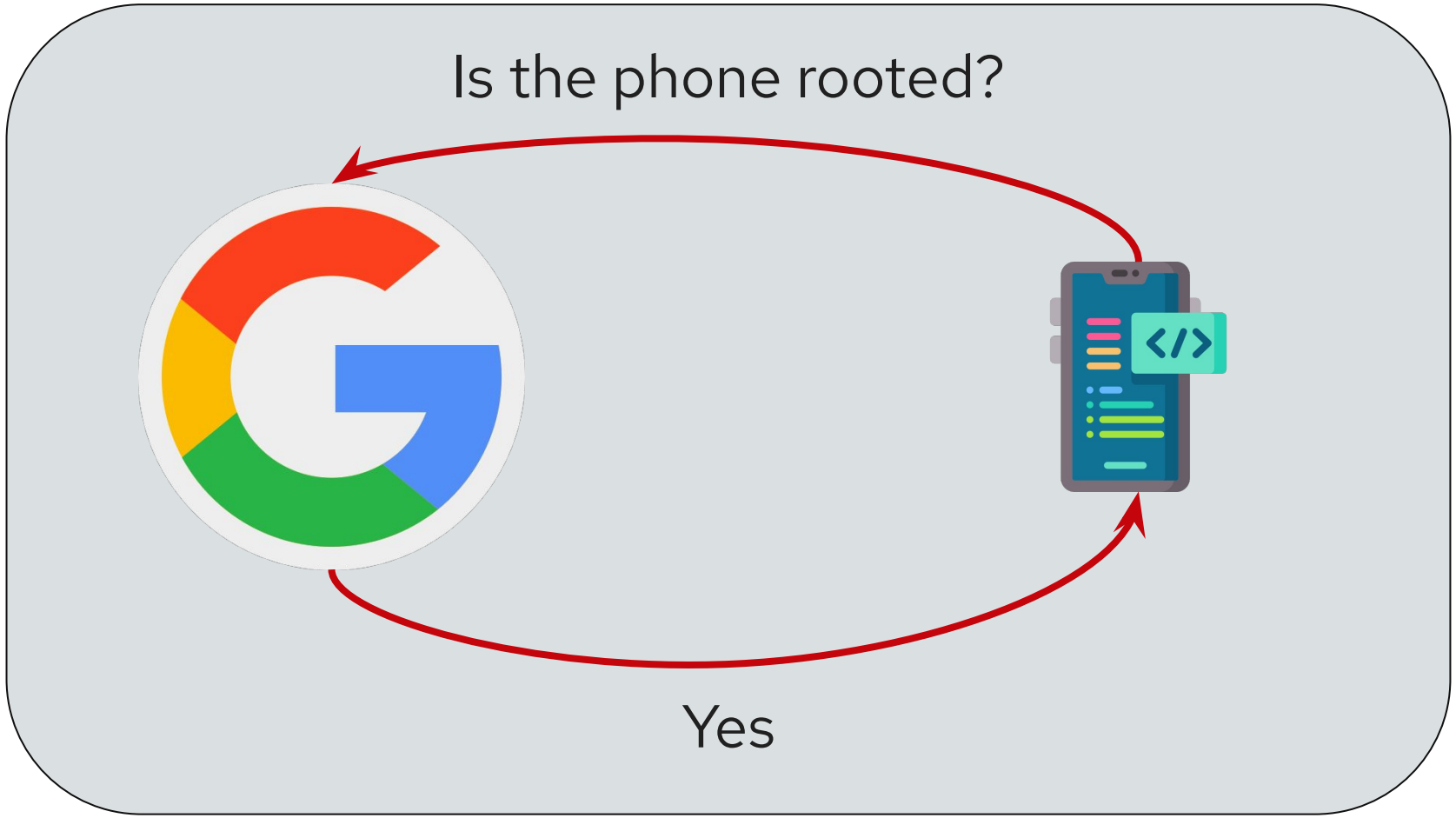
Summary


AI/ML is for Advertising
Either
AI/ML is for Functionality
Implementation
AI/ML Outputs
?
10/21
8/21
Implementation Matters!
Personal Model Interactions
Majority Were Fine With Models
Some Changed Their Minds

Future Work

Conference TBD

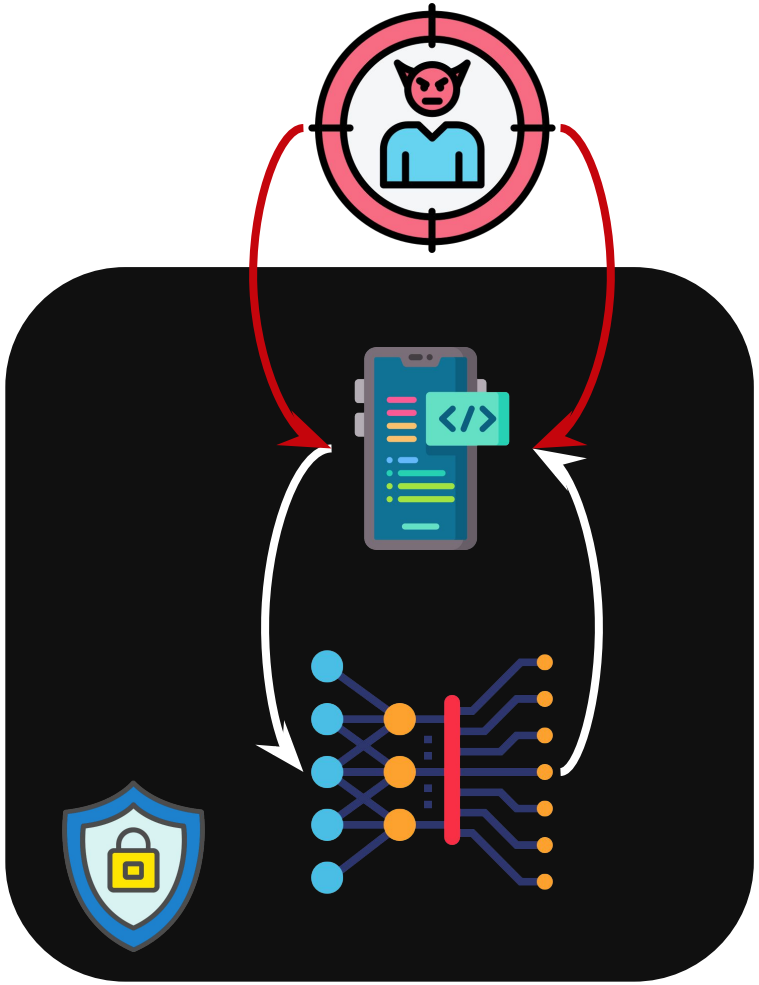
Google Play Integrity



Apps know when the phone is rooted!

No one has cracked GPI!

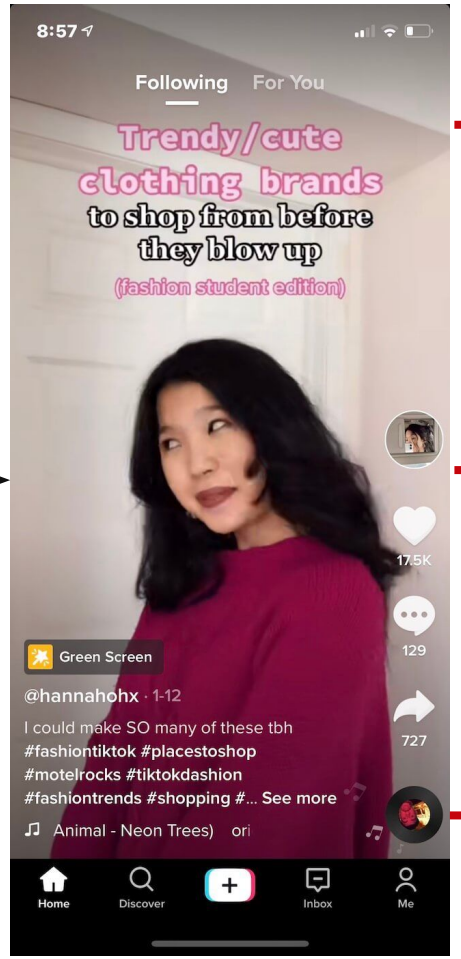
Defending Local Models



Easy for attacker to hook AI code.

How can we define new defenses for AI models?

Personalization Algorithms



Why this TikTok?

How was this TikTok selected?

What does TikTok look for to make the selection?

Thank you!

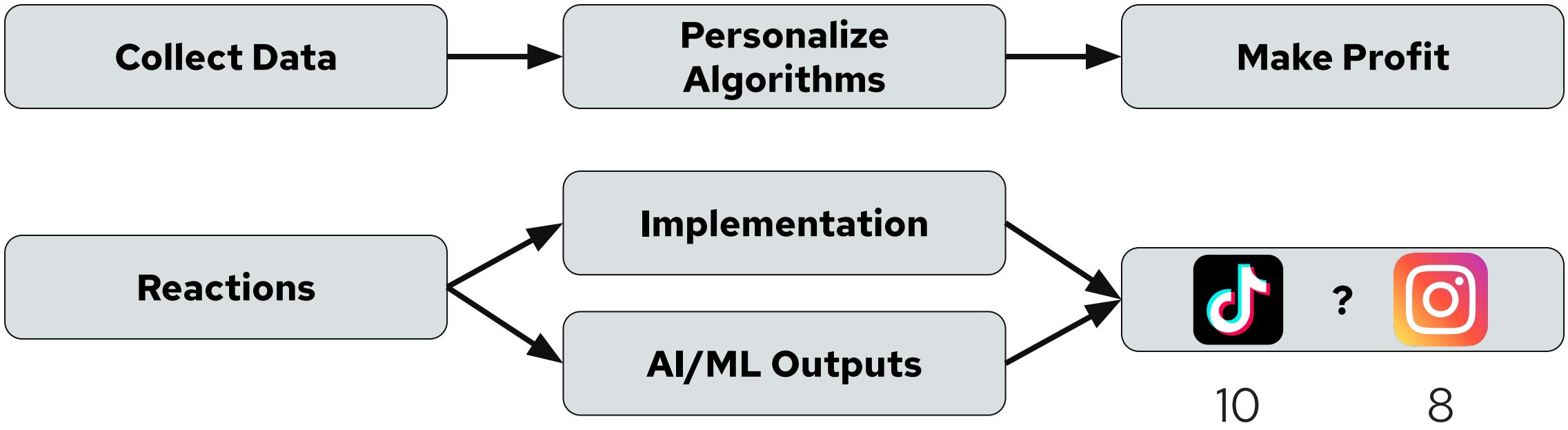
MADS&P
Security and Privacy Research Group
at UW-Madison

Wi-Pi

Jacksonwaynewest.com

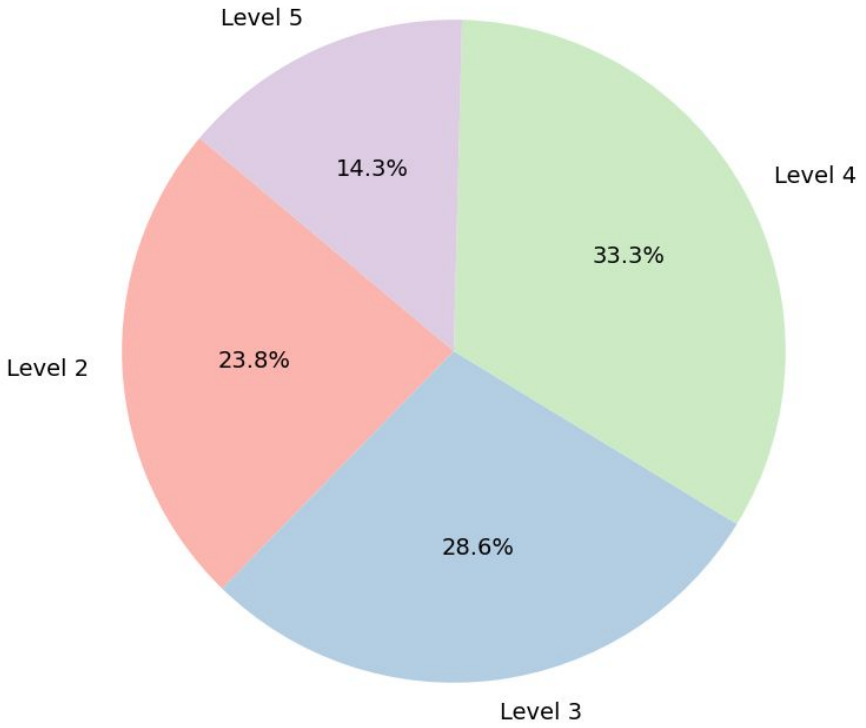
The Purpose of AI/ML in Social Media

AI/ML Reactions



Participant Statistics

N=21



Tech-Savviness Distribution

Other Important Takeaways:

Differing Transparency Preferences

Social Media Trust and Local Data Analysis

Thank You!


WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

How to use this template

This template is primarily to be used for text-only presentations. Photographic and data-driven presentation templates can be found at Brand & Visual Identity PowerPoint template page.

Before you start, check out our presentation best practices document.

This template has:

- two title slides options (white or black)
- two section header options (red or black)
- two text slides (one-column or two-column)
- two end slide options (red or white)
- 30 icons to choose from

To select what version you would like, click on the drop-down menu beside "new slide" button in the top left corner. Select the slide you would like.

To insert text, click on the text box and start typing. Please note that copying and pasting text can change how the font looks. It is better to type directly onto the slide. Also note that fonts size 18 or larger work better for presentations than smaller font sizes.

To use icons within your presentation, copy and paste desired icon(s) from the last slide onto the slides you wish to use them on.

The following slides are here for visual reference only. You may delete and edit as needed for your own presentation.

While this PowerPoint template is made to be accessible, please account for the following as you make edits:

- Give each slide a unique title, including slides that you duplicate
- Include alternative, or alt text, descriptions to all images

If you have any questions about how to use this template, please contact us at contact@umark.wisc.edu

Note: Only use red icons on white or light gray backgrounds