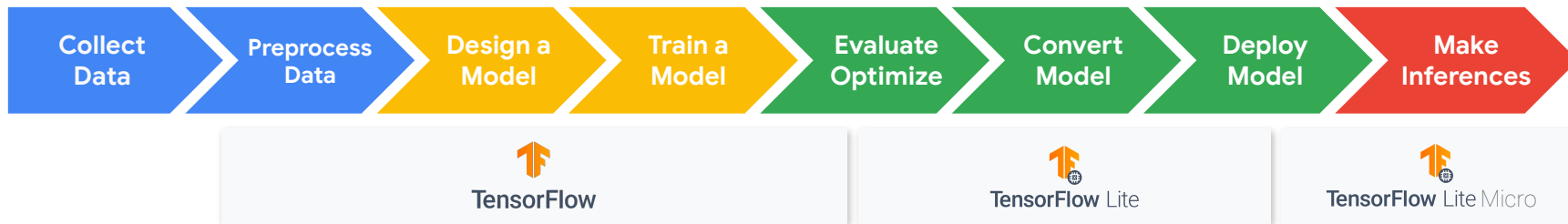
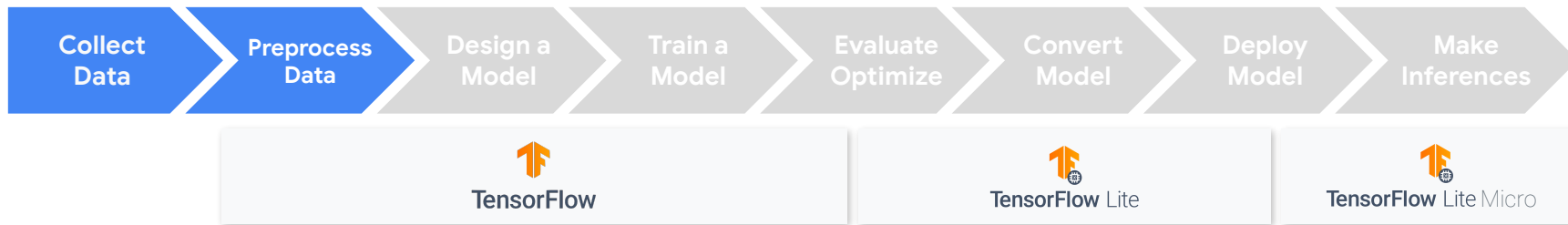


VWW: Data Collection and Processing







Recall: Licenses



BSD
License



Data Collection and Processing

**Someone scraped 40,000
Tinder selfies to make a
facial dataset for AI
experiments**

Natasha Lomas @riptari / 7:21 PM EDT • April 28, 2017

Data Collection and Processing

Someone scraped 40,000 Tinder selfies to make a facial dataset for AI experiments

Natasha Lomas @riptari / 7:21 PM EDT • April 28, 2017

Update: A Tinder spokesperson has now provided the following statement:

We take the security and privacy of our users seriously and have tools and systems in place to uphold the integrity of our platform. It's important to note that Tinder is free and used in more than 190 countries, and the images that we serve are profile images, which are available to anyone swiping on the app. We are always working to improve the Tinder experience and continue to implement measures against the automated use of our API, which includes steps to deter and prevent scraping.

This person has violated our [terms of service](#) (Sec. 11) and we are taking appropriate action and investigating further.

Recall: **Don't collect** from scratch

Data collection is **difficult**!

- Can we **reuse** existing data?

What's available?

What's missing?

Visual Wake Words Dataset

Visual Wake Words Dataset

Aakanksha Chowdhery, Pete Warden, Jonathon Shlens,
Andrew Howard, Rocky Rhodes

Google Research

{chowdhery, petewarden, shlens, howarda, rocky}@google.com

Visual Wake Words Dataset



```
"annotations": [  
  {  
    "segmentation": [[510.66,423.01,511.72,420.03,...,510.45,423.01]],  
    "area": 702.1057499999998,  
    "iscrowd": 0,  
    "image_id": 289343,  
    "bbox": [473.07,395.93,38.65,28.67],  
    "category_id": 18,  
    "id": 1768  
  },  
]
```

Visual Wake Words Dataset



```
"annotations": [  
  {  
    "segmentation": [[510.66,423.01,511.72,420.03,...,510.45,423.01]],  
    "area": 702.1057499999998,  
    "iscrowd": 0,  
    "image_id": 289343,  
    "bbox": [473.07,395.93,38.65,28.67],  
    "category_id": 18,  
    "id": 1768  
  },  
]
```

Visual Wake Words Dataset



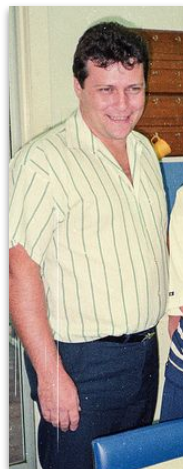
Label: "person"

Visual Wake Words Dataset



Label: “person”

Visual Wake Words Dataset



Label: “person”



Label: “person”

Visual Wake Words Dataset

Data collection is **DIFFICULT**

- This dataset and collection process is *limited* and has bias
- Small number of relevant images
- Large quantity of irrelevant images

Visual Wake Words Dataset

Aakanksha Chowdhery, Pete Warden, Jonathon Shlens,
Andrew Howard, Rocky Rhodes
Google Research
{chowdhery, petewarden, shlens, howarda, rocky}@google.com