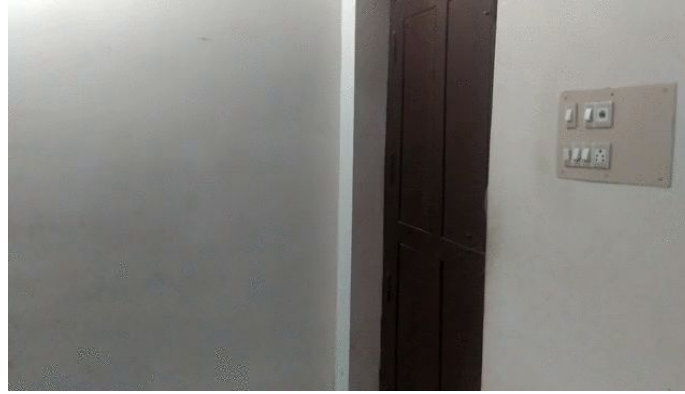


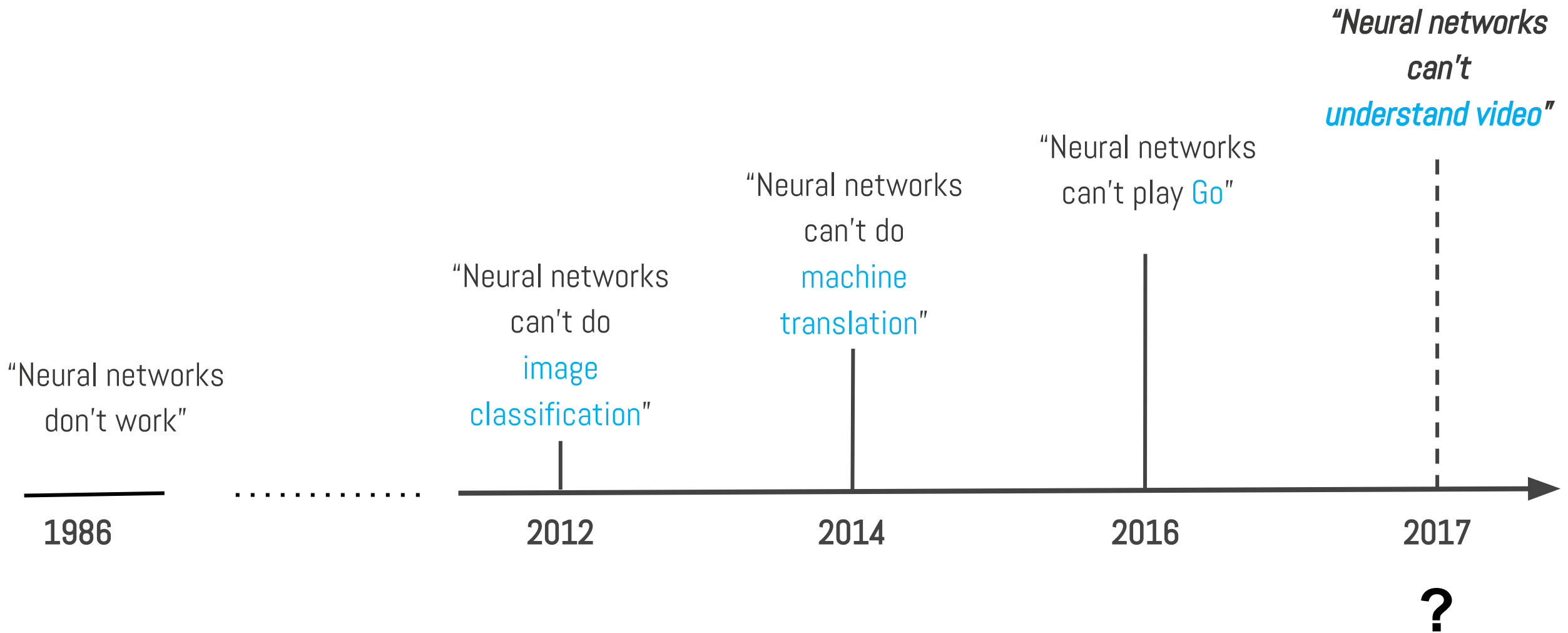


## Twenty Billion Neurons

Berlin & Toronto based Video Understanding Company

Image understanding is about textures, video understanding about “physics”





We are a team of 18 engineers and researchers who have **built companies, engineered large scale data platforms** and published **groundbreaking research**



Dr. Christian Thureau

CBDO

Co-founder of Game Analytics (sold to Mobvista in 2016); ex-researcher at Fraunhofer



Dr. Ingo Bax

CTO

Ex-professor at FH Münster; co-built the central software architecture at XING



Roland Memisevic, PhD

Chief Scientist

Ex-professor at MILA Montreal; doctorate at U of Toronto advised by Geoffrey Hinton



Valentin Haenel

VP Engineering

Co-initiator of PyData Berlin; contributor in more than 50 open source projects



Moritz Müller-Freitag

Head of Product

Previous experience as data scientist and country manager at HitFox portfolio company

Scientific advisor

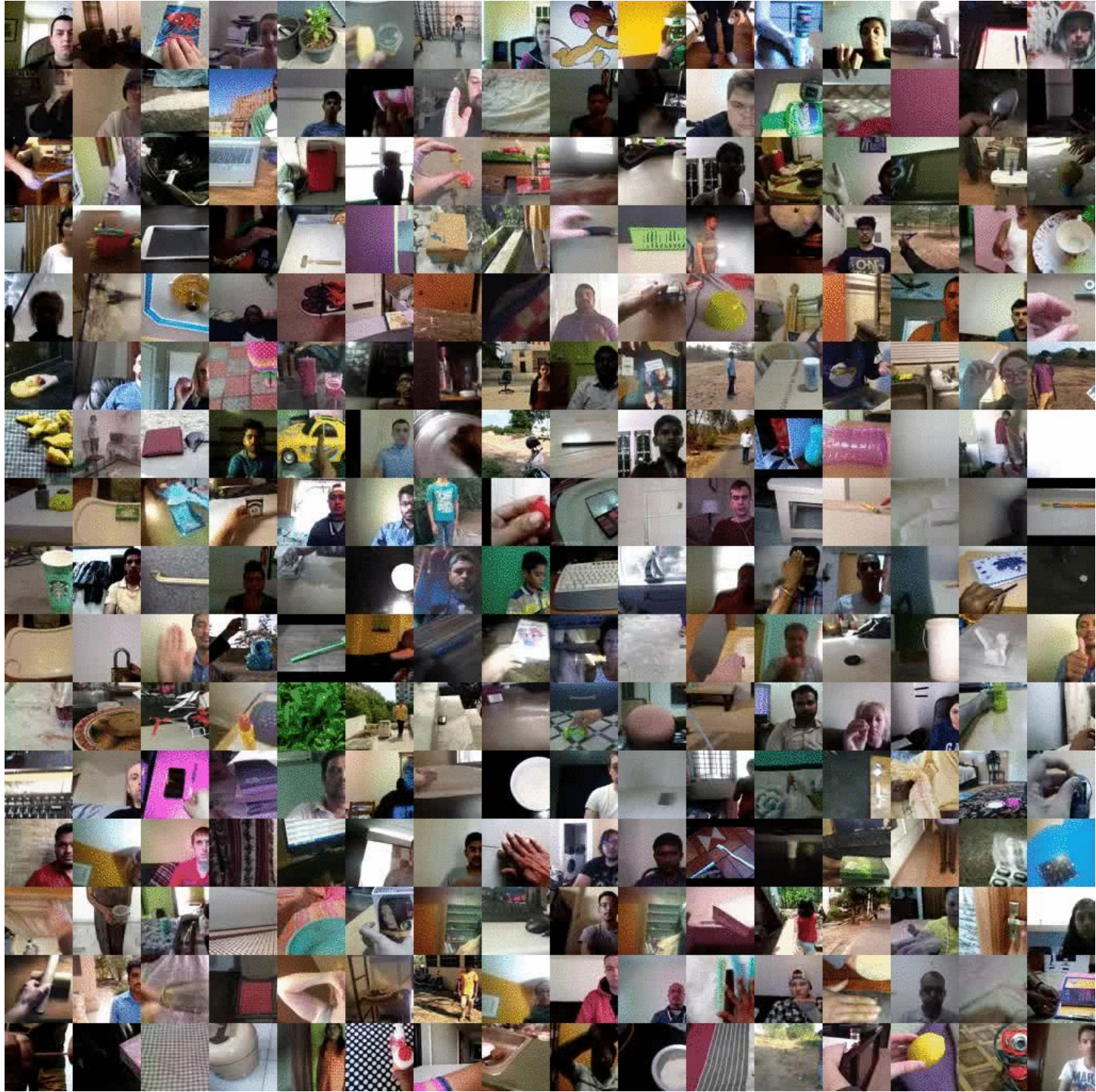


Prof. Yoshua Bengio

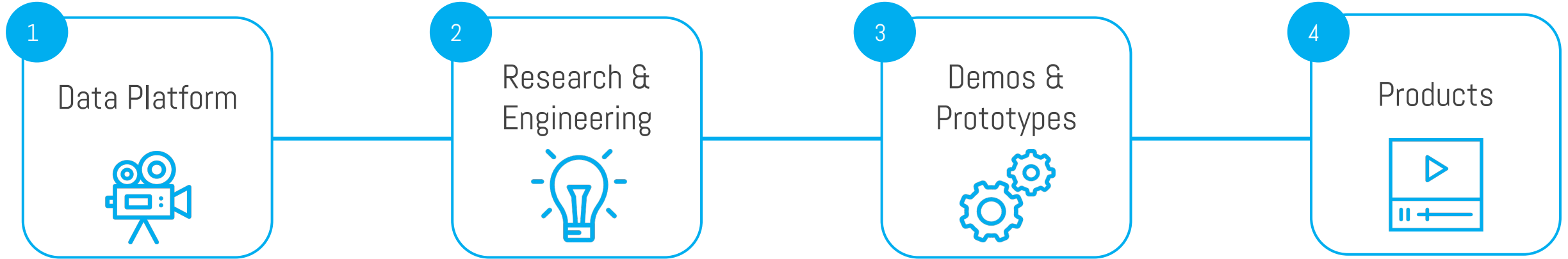
Scientific Advisor

Full Professor at MILA Montréal; noted for his pioneering work on deep learning





# AI needs integration



# RGB gesture recognition

## Traditional approach

- Requires depth information (3D camera system)
- Not tractable on mono camera (RGB signal)
- ~5 gestures with low accuracy

## Our approach

- Recognizes **25 hand gestures**
- Runs in **real-time** on a **laptop** using **RGB camera** input
- Uses our proprietary data set (now largest in the world)

## Significance

- **Speed**: Using data platform, prototype in 8 weeks
- **Synergies**: Technology reusable for action recognition





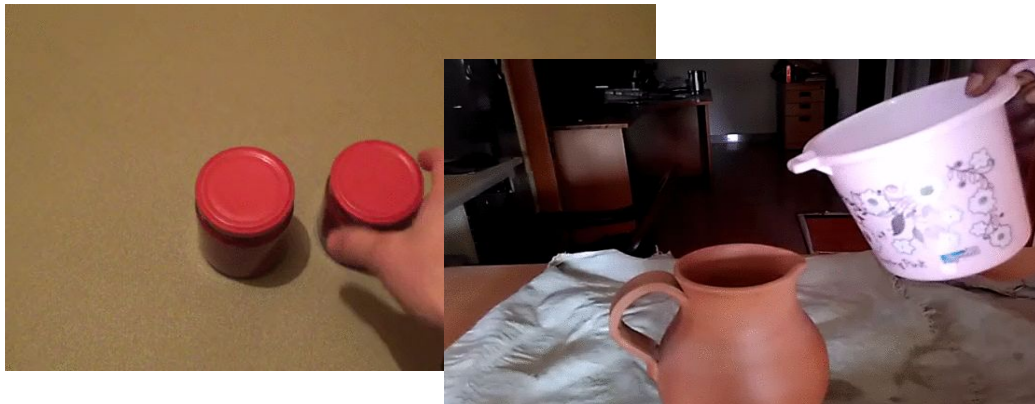
# Our datasets cover a variety of use cases and capabilities



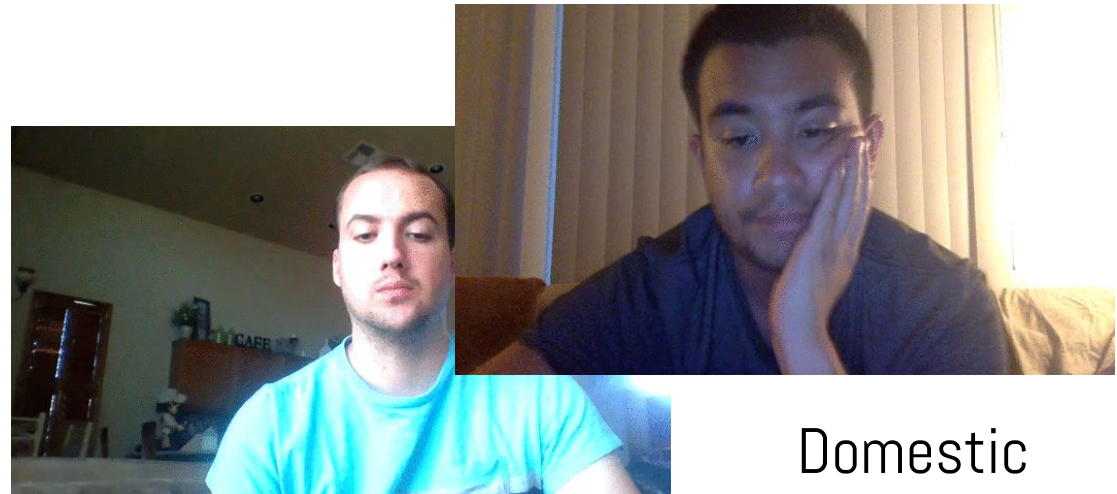
Retail actions



Multi-person actions



"Something-something"



Domestic



# TwentyBN "SuperModel"



*Prediction: Person picking [something] up*



*Person leaving through a door*



*gesture: Zooming Out With Two Fingers*



*[Something] falling like a feather or paper*



*Trying to bend [something unbendable]  
so nothing happens*



*Bending [something] until it breaks*

# Contrastive classes make learning harder and networks stronger



Tearing [something] into two pieces VS Tearing [something] just a little bit	0.74 (0.52)
Pretending to pick [something] up VS Picking [something] up	0.86 (0.75)
Pretending to pour VS Pouring	0.82 (0.64)
Pouring with overflow VS Pouring without	0.76 (0.54)
Pretending to put [something] onto VS Putting [something] onto [something]	0.82 (0.64)

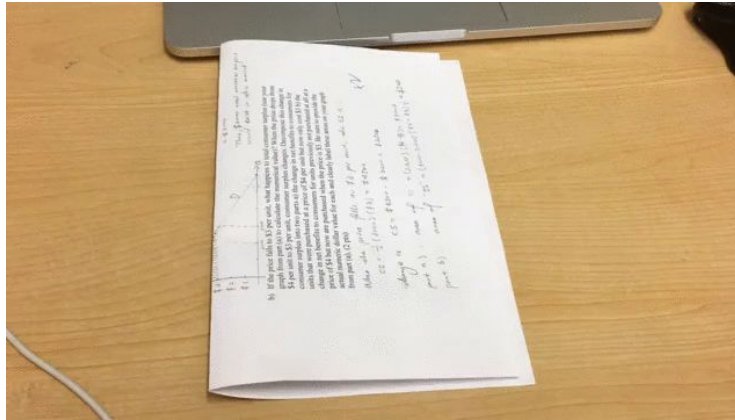
# Mistaken “opening” predictions

Ground truth: Moving [part]  
of [something]



Prediction: Opening  
[something]

Ground truth: Unfolding  
[something]



Prediction: Opening  
[something]

Ground truth: Putting  
[something] on a flat surface  
without letting it roll



Prediction: Opening  
[something]



# Mistaken "covering" predictions

Ground truth: Putting [something] in front of [something]



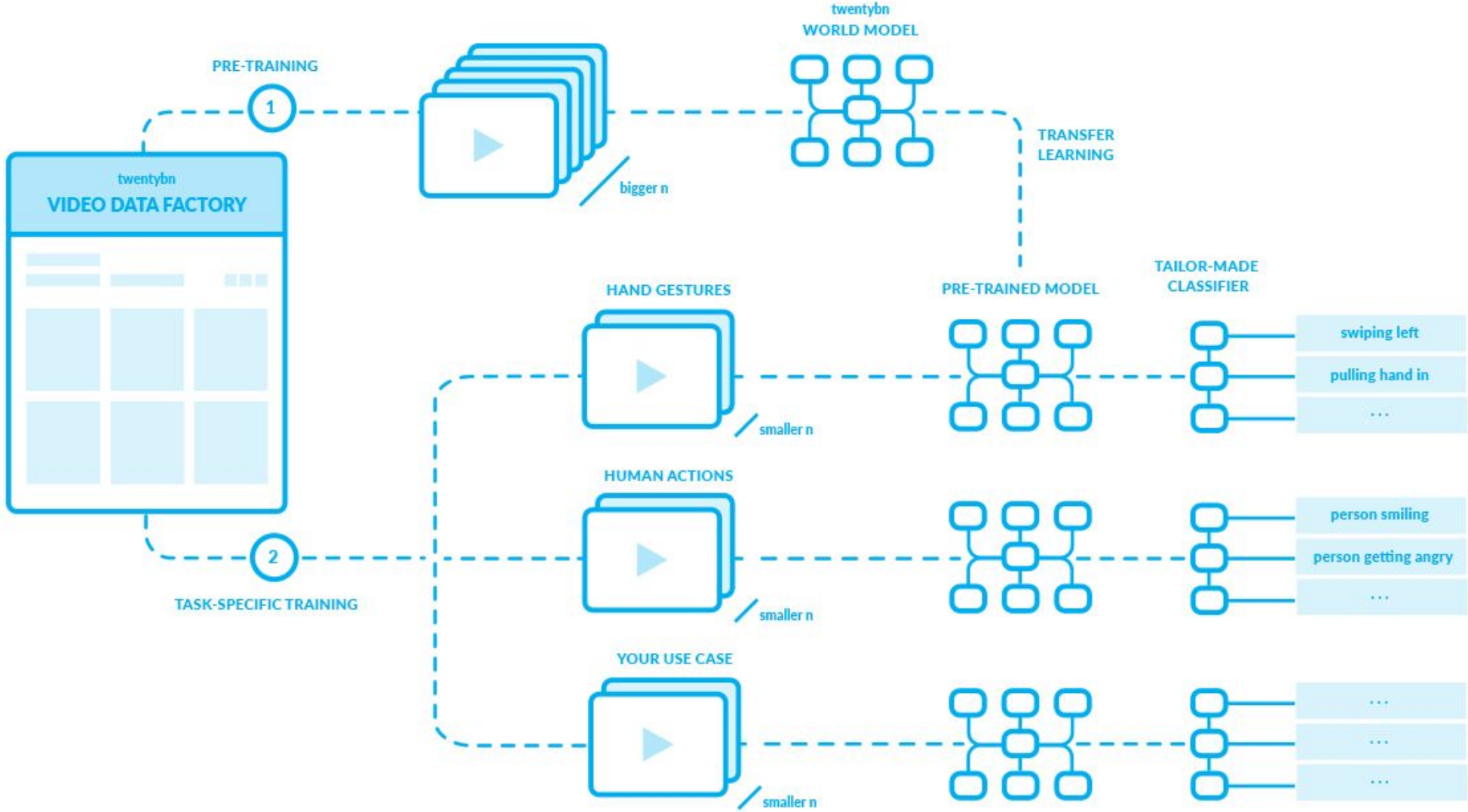
Prediction: Covering [something]

Ground truth: Turning [something] upside down



Prediction: Covering [something]

# Transfer learning





Roland Memisevic, PhD

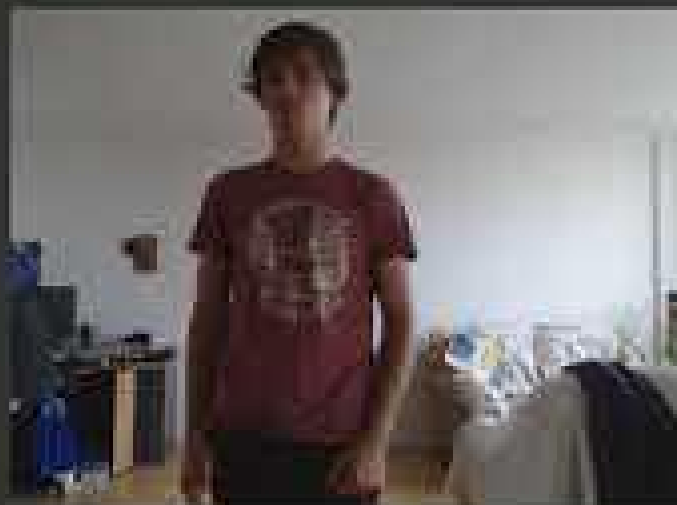


+1 416 826 1032



roland@twentybn.com





### Person shaking head

- Person shaking head
- Person shaking head
- Person shaking head to side left
- Person shaking head to side right
- Person shaking head with hand
- Person shaking head



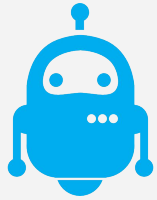
Moving [something] closer to [something]



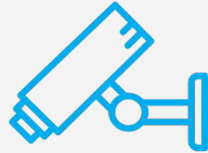
# Commercial opportunities

## Opportunities today...

Domestic companions



Action recognition and  
context-aware scene  
understanding



Contextual ads and  
intelligent content  
moderation

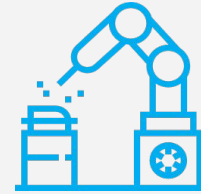


Online video search,  
discovery and  
recommendation



## ...and tomorrow

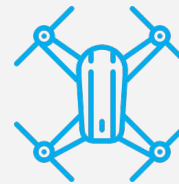
Programming robots  
by teaching them  
real world physics



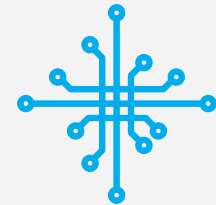
Augmented and  
mixed reality  
interactions



Situational awareness  
and planning for  
autonomous agents



Future applications that  
are related to video and  
the real world







Person walking away from the camera and turning right (Ground Truth)

**Predictions:**

1. Person leaving through a door 31.8%
2. Person entering through a door 10.5%
3. Person walking away from the camera and turning right 8.3%
4. Person walking away from the camera and turning left 5.0%
5. Person walking away from the camera



**Taking [one of many similar things on the table] (Ground Truth)**

**Predictions:**

1. Taking [one of many similar things on the table] 18.3%
2. Taking [something] out of [something] 13.4%
3. Taking [something] from [somewhere] 5.3%
4. Pretending to take [something] from [somewhere] 4.4%
5. Pulling [something] out of [something]



Ground truth: [Picking remote control up.](#)



Ground truth: [Pretending to pick black remote control up.](#)



# 20BN-SOMETHING-SOMETHING

A crowd-acted dataset of basic interactions  
with everyday objects.

Number of Videos: 108.499

License: Free for academic use

(Creative Commons Attribution 4.0  
International license CC BY-NC-ND 4.0)



<https://www.twentybn.com/datasets/something-something>

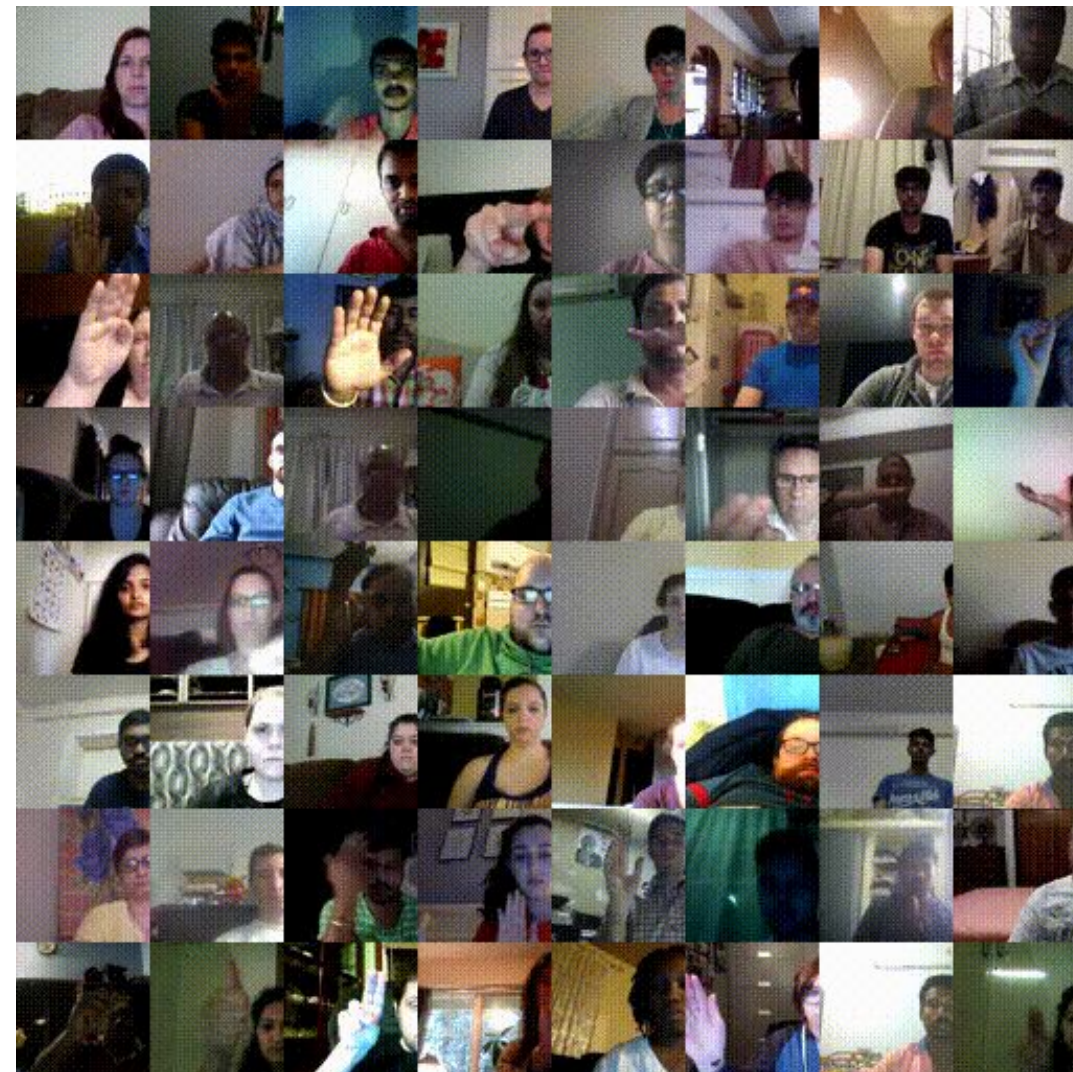
# 20BN-JESTER

A crowd-acted dataset of generic human hand gestures.

Number of Videos: 148.094

License: Free for academic use

(Creative Commons Attribution 4.0  
International license CC BY-NC-ND 4.0)



<https://www.twentybn.com/datasets/jester>