2020 embedded VISION summit®

Enabling the Next Generation of Smart Devices with Interactive AI

Roland Memisivic

September 2020

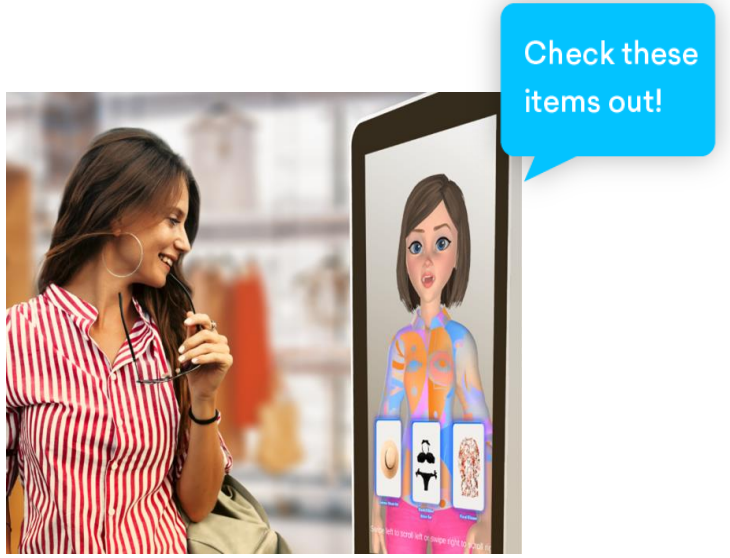
twentybn

Our mission at TwentyBN: Moving from smart speakers...

Smart speakers are confined to the simplest of tasks: “how’s the weather?”

Existing “service robots” are essentially chatbots with bodies

...to embodied assistants, that can see and interact with you



Bloomberg

Checkout

Meet 'Millie' the Avatar. She'd Like to Sell You a Pair of Sunglasses

Canadian startup Twenty Billion Neurons developed the AI-powered sales assistant and is discussing trials with retailers.

By Jeremy Kahn
15 December 2018, 09:00 GMT-5 *Updated on 17 December 2018, 08:26 GMT-5*

AI Startup Aims To Redefine How People Interact with Technology

December 6, 2018

TwentyBN is a German/Canadian based AI startup that is using computer vision to redefine how people interact with technology.

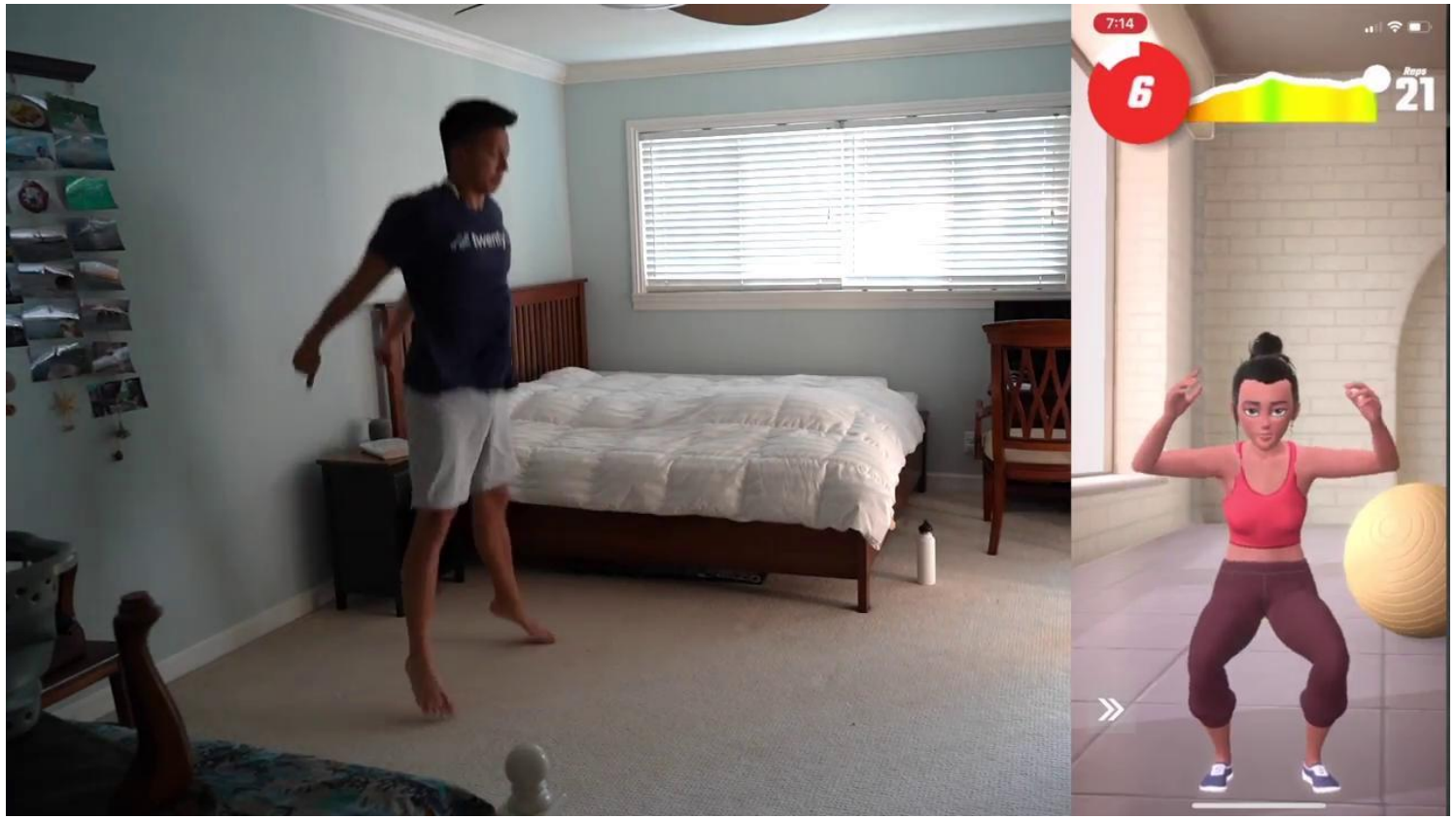
At NeurIPS in Montreal, Canada this week the company demonstrated a brand new computer vision based-avatar named Millie that is designed for retail and entertainment applications.

The Download

What's up in emerging technology

I tested an AI sales assistant that's trained to push you into spending more

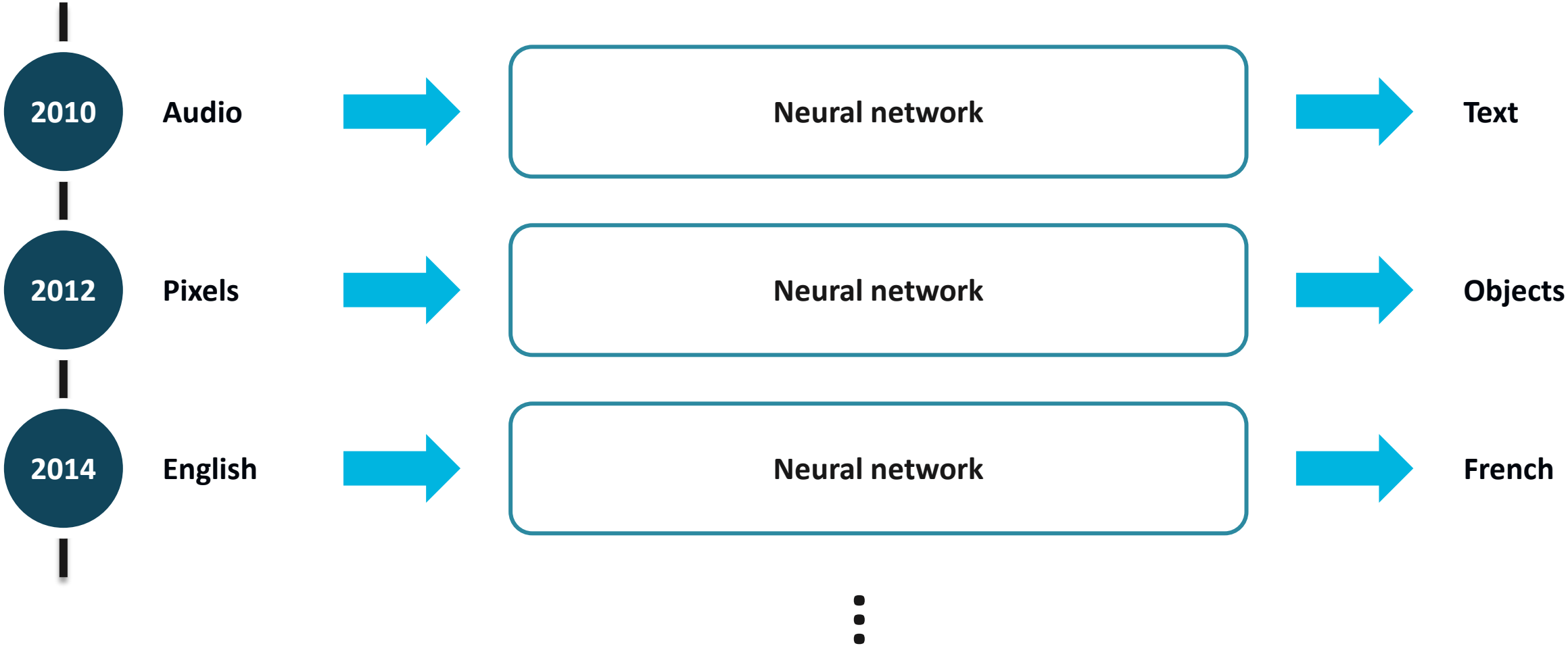
Millie, an embodied "Alexa" for public spaces



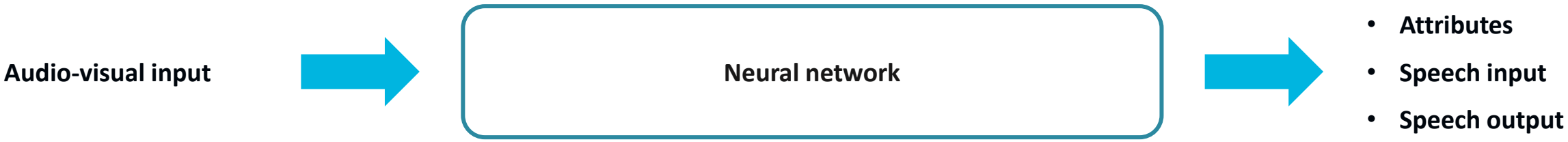
FitnessAllyApp.com

A fitness coach that won't let you cheat (iPhone 8 and up)

End-to-end learning is taking over the world



Humanoid assistants push end-to-end learning to its limits!



- Understand the scene
- Understand objects and actions
- Understand human
- Behaviour

- Understand spoken language
- Generate spoken language
- Link visual concepts to words (“grounding”)
- Reason about past and present events

- Control the assistant’s face
- Control the assistant’s body
- Etc

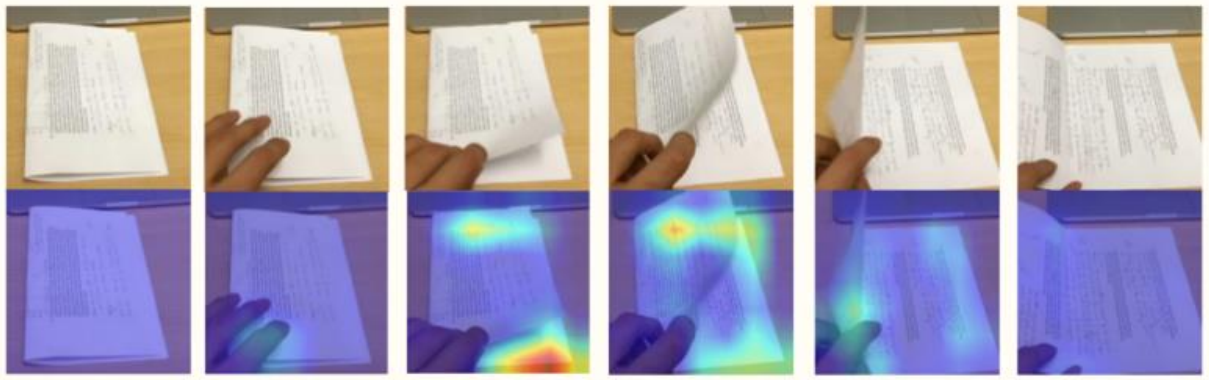
How do we get the data for this!?

- We built a worldwide **“movie studio”** that cranks out up to 45k training videos a day
- To catch every imaginable edge case, videos are recorded and annotated **(crowd acting)**
- Videos are automatically recorded, re-coded, reviewed, …
- The data collection is constantly evolving according to the AI systems current needs…

Researchers identify which labels the network needs to learn next

Crowdworkers record, verify and annotate videos

- … resulting in >3.5 million videos across thousands of classes and tasks to date
- The network already learned objects, actions, temporal events, counting, captioning, “behaviours”, ...



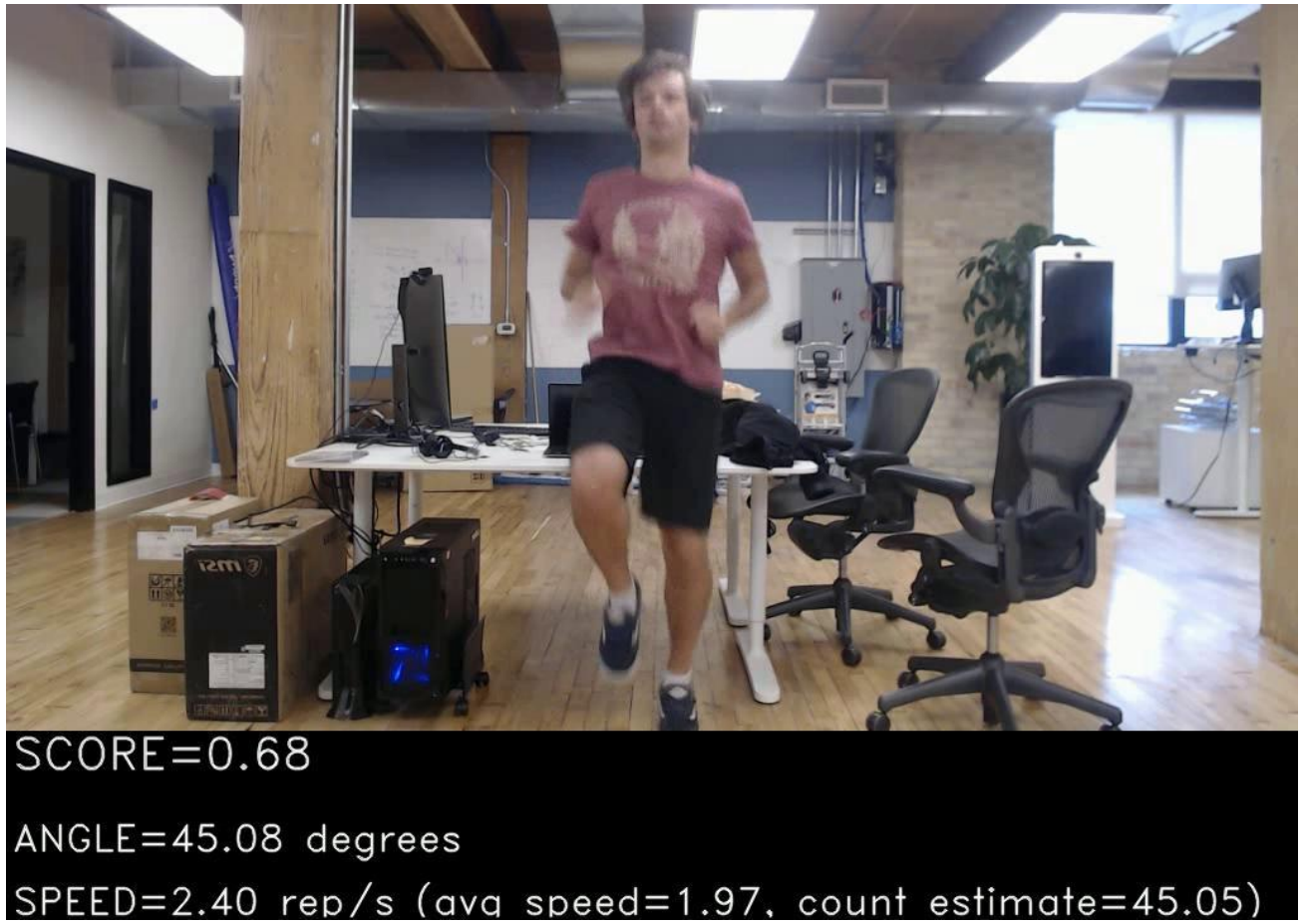
True: Unfolding [smth]

Prediction1: Opening [smth] (49.27%)
Prediction2: Uncovering [smth] (30.12%)

True: Spilling [smth] onto [smth]

Prediction1: Pouring [smth] into [smth] until it overflows (49.43%)
Prediction2: Pouring [smth] into [smth] (38.78%)

Bounding boxes?
We let the network figure out where to look!

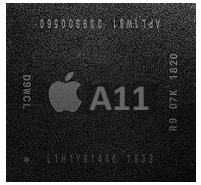


Skeleton models?
We let the network figure
out how well you're doing!

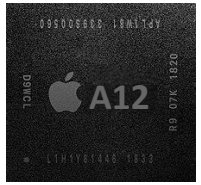
Temporal segmentation?
We let the network figure out what you're doing, and when!

End-to-end learning makes data sourcing hard, keeping networks simple

Apple Devices

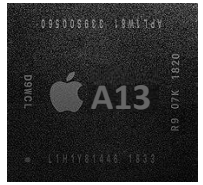


iPhone 8 and 8 Plus (2017)
iPhone X (2017)



iPhone XS, XS Max, and XR (2018)

iPad Air, Mini and Pro (2019)



iPhone 11 and 11 Pro (2019)

iPhone SE (2020)

Android Devices

Samsung Galaxy S8 and Note 8 (2017)

Google Pixel 2, Xiaomi Mi 6 (2017)

Samsung Galaxy S9 (2018)

Sony Xperia XZ2, Xiaomi Mi 8 (2018)

Samsung Galaxy S20 (2019)

Sony Xperia 1 II, Xiaomi Mi 10 (2019)

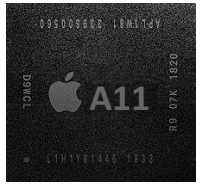
All inference can run with 8.5B MAC/sec -> .017 TOPS

End-to-end learning makes data sourcing hard, keeping networks simple

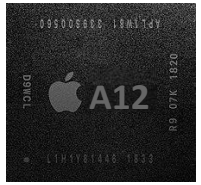
2020 embedded VISION summit

Option

Apple Devices

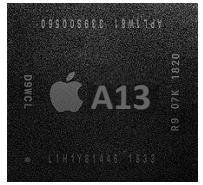


iPhone 8 and 8 Plus (2017)
iPhone X (2017)



iPhone XS, XS Max, and XR (2018)

iPad Air, Mini and Pro (2019)



iPhone 11 and 11 Pro (2019)

iPhone SE (2020)

Android Devices

Samsung Galaxy S8 and Note 8 (2017)

Google Pixel 2, Xiaomi Mi 6 (2017)

Samsung Galaxy S9 (2018)

Sony Xperia XZ2, Xiaomi Mi 8 (2018)

Samsung Galaxy S20 (2019)

Sony Xperia 1 II, Xiaomi Mi 10 (2019)

All inference can run with 8.5B MAC/sec -> .017 TOPS

twentybn

Conclusions

- **Personal companions** and camera-enabled assistants are moving from science fiction to reality
- They're a gigantic commercial opportunity because they are **personalized** and **sticky**. **The more you use them, the better they get**
- They're also a gigantic research opportunity because they stretch the limits of AI via **audio-visual dialogue and grounding**
- **On-screen companions** can capture a lot of that value since they allows us to bring to bear **end-to-end** learning for most part of the way
- For more info and access to our developer platform visit: **20bn.com**

Resources

Developer platform

- 20bn.com

Fitness Ally app

- www.fitnessallyapp.com

Embodied AI newsletter

- www.embodiedai.co

The “something something” database for learning common sense features

- arxiv.org/abs/1706.04261

Datasets

- https://20bn.com/products/datasets

Me, me!

Questions?