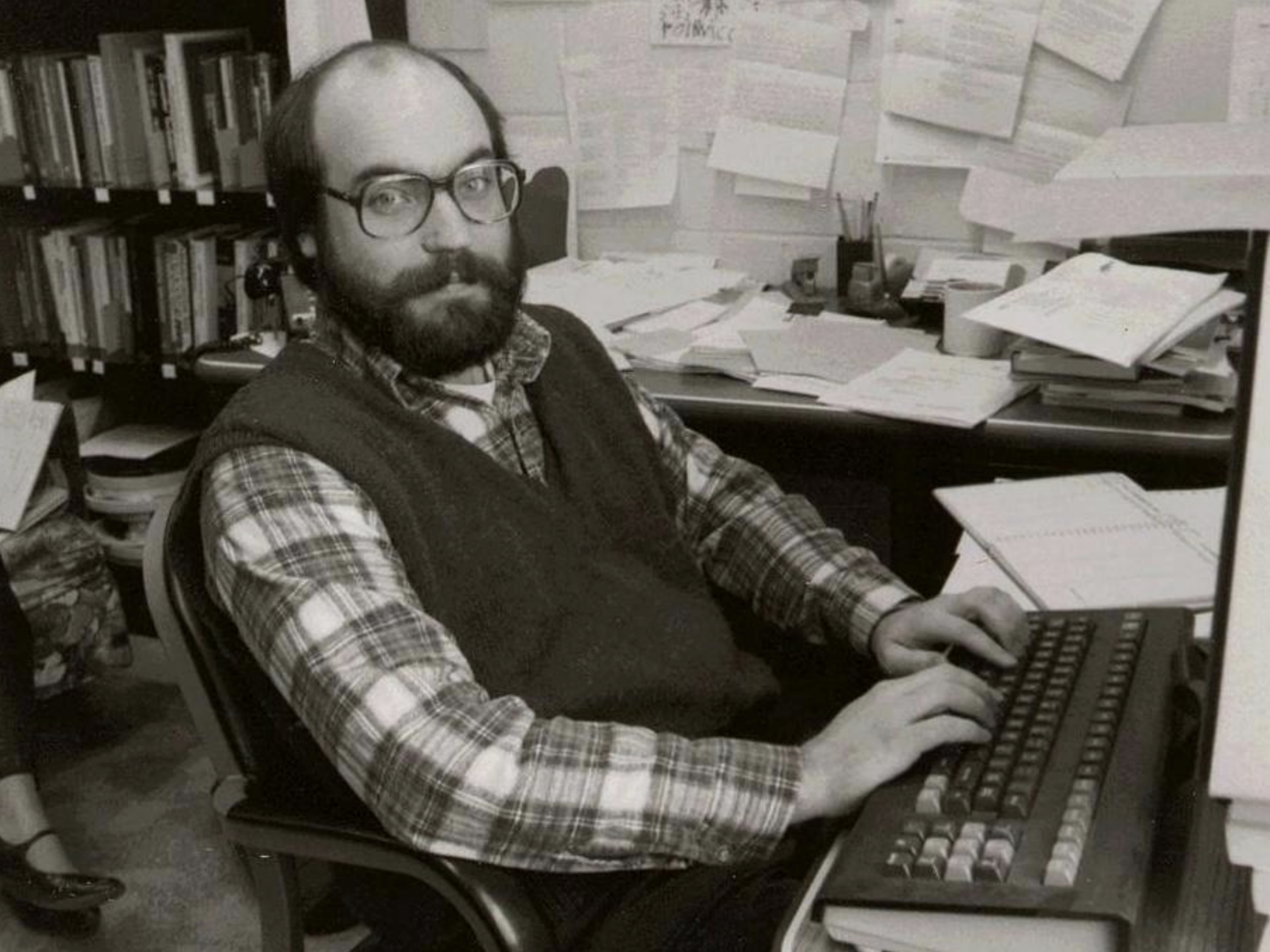
Augur: Mining Human Behaviors from Fiction to Power Interactive Systems

Ethan Fast, Will McGrath, Pranav Rajpurkar, Michael Bernstein

Stanford HCI Group



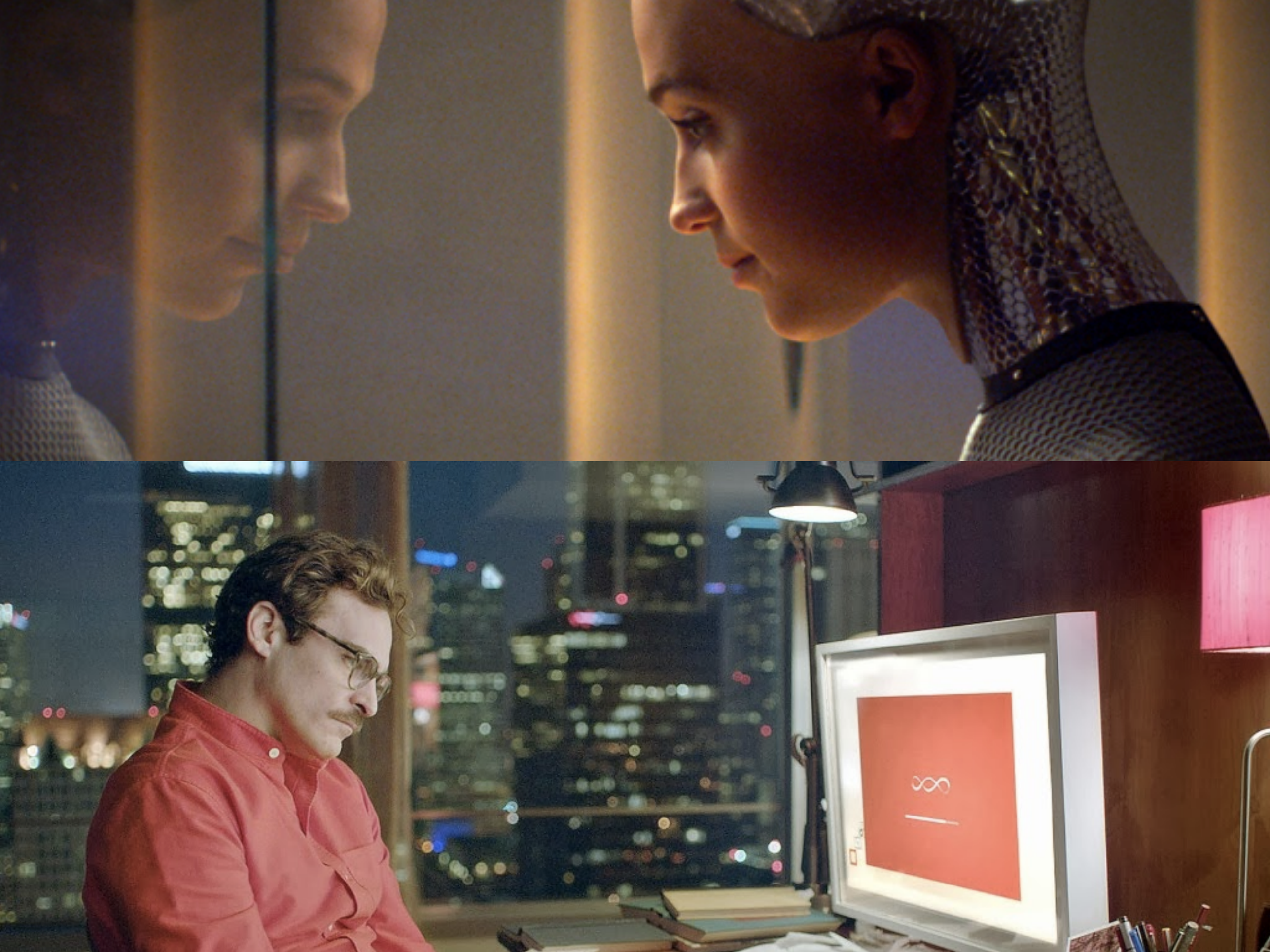
File Network Tools Refractals Agent

Desertification in the Sub Sahara

Rate of Deforestation

2,200,000 acres per year

Script Year:20



ConceptNet (Liu and Singh, 2004)

{wake up} → **make coffee**

ConceptNet (Liu and Singh, 2004)

{wake up} → **make coffee**

But what about:

{go running} → **drink water**

{cart, broccoli} → **buy food**

No one tweets like this:

“I’m **#typing** on my **#keyboard**”

“Now I’m **#standing_up**”

“Hey, **#walking** to my **#window**”

Fiction: “He walked to the bookshelf, picked up his favorite book, and started to read.”

Predict next activity:

pick up book → read

Predict activity from context:

bookshelf → {pick up book, read}

{mountain, tree, backpack} → hike

(1) **Data**

(2) **Knowledge Base**

(3) **Models**

(4) **Applications**

(5) **Evaluation**

Data: 1.8 billion words of **modern fiction** from **Wattpad**, an **amateur** writing community.

The Augur knowledge base

54,075 activities in Augur:

“*He opens the fridge*” → **open fridge**

“*She turns off the lights*” → **turn off lights**

“*I jumped*” → **jump**

13,843 objects and locations:

“He opened up ***Facebook***.”

“When we got to the ***beach***, I took off my ***shirt***.”

“We got in the **car** and drove to the **hospital**.”

API #1: scene context → activity:

{plate, fork, table} → eating

{car, road} → drive

API #2: activity → activity:

{order, eat} → pay

{shower} → put on clothes

“He drove down the road.”

drove and **road** co-occur **3590** times

“He drove to the store and parked the car.”

drive and **park** co-occur **5433** times

“He ate while he drove.”

drove and **eat** co-occur **102** times

MI(a,b) ~ log(p(a,b) / p(a)*p(b))

p(w) = occurrence of w

$p(w_1,w_2)$ = co-occurrence of w_1,w_2

computer and **type** have **high** MI

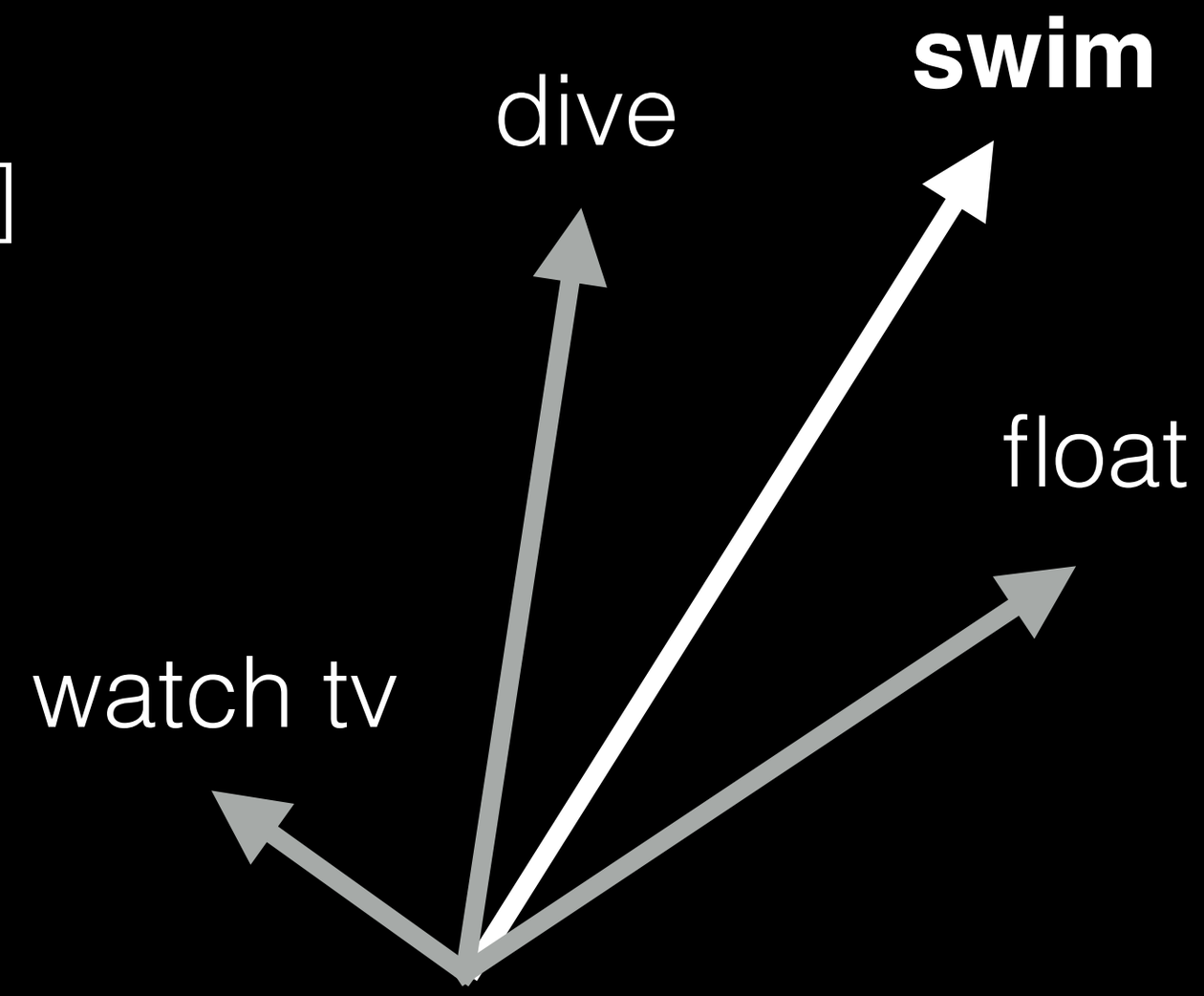
drive and **park** have **high** MI

tree and **eat** are have **low** MI

Vector spaces allow Augur to leverage **multiple** examples of **scene context**.

swim = [2.1, 1.5, 0.3, …]

swim$_{goggles}$ = 2.1
swim$_{pool}$ = 1.5
swim$_{chair}$ = 0.3



query on “goggles” and “pool”:

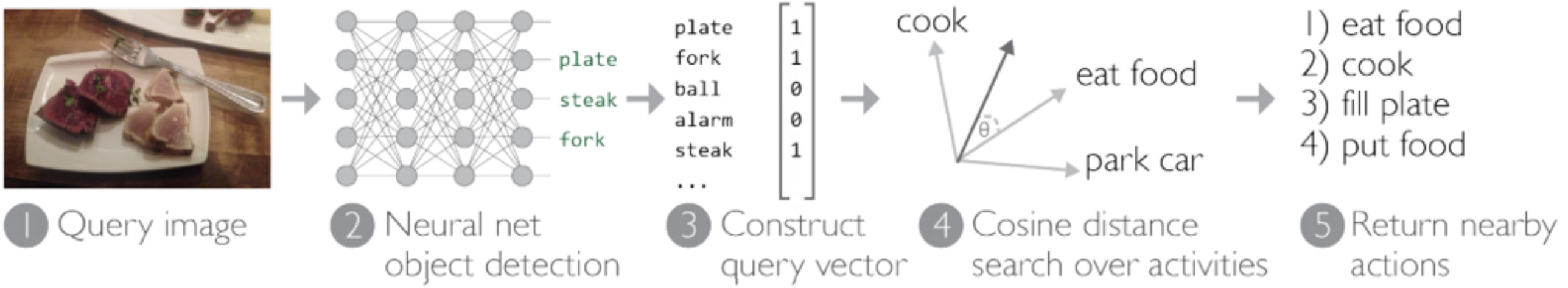
goggles pool

query = [0, 1, 0, 0, 1, 0, …]

find vectors (other activities in the space) with the highest cosine similarity

Augur applications

Computer Vision → Augur VSM → Predictions



CV: people, mountain, tree, backpack

Augur: hike, sling, see fire, climb tree, climb, reach top, leap

CV: beach, sand, boy, shoe

Augur: reach beach, lay towel, love beach, take shoe off, swim, lay, dive

CV: boat, sea, sky, ship, ocean, fog

Augur: row, see light, sail, swim, dive

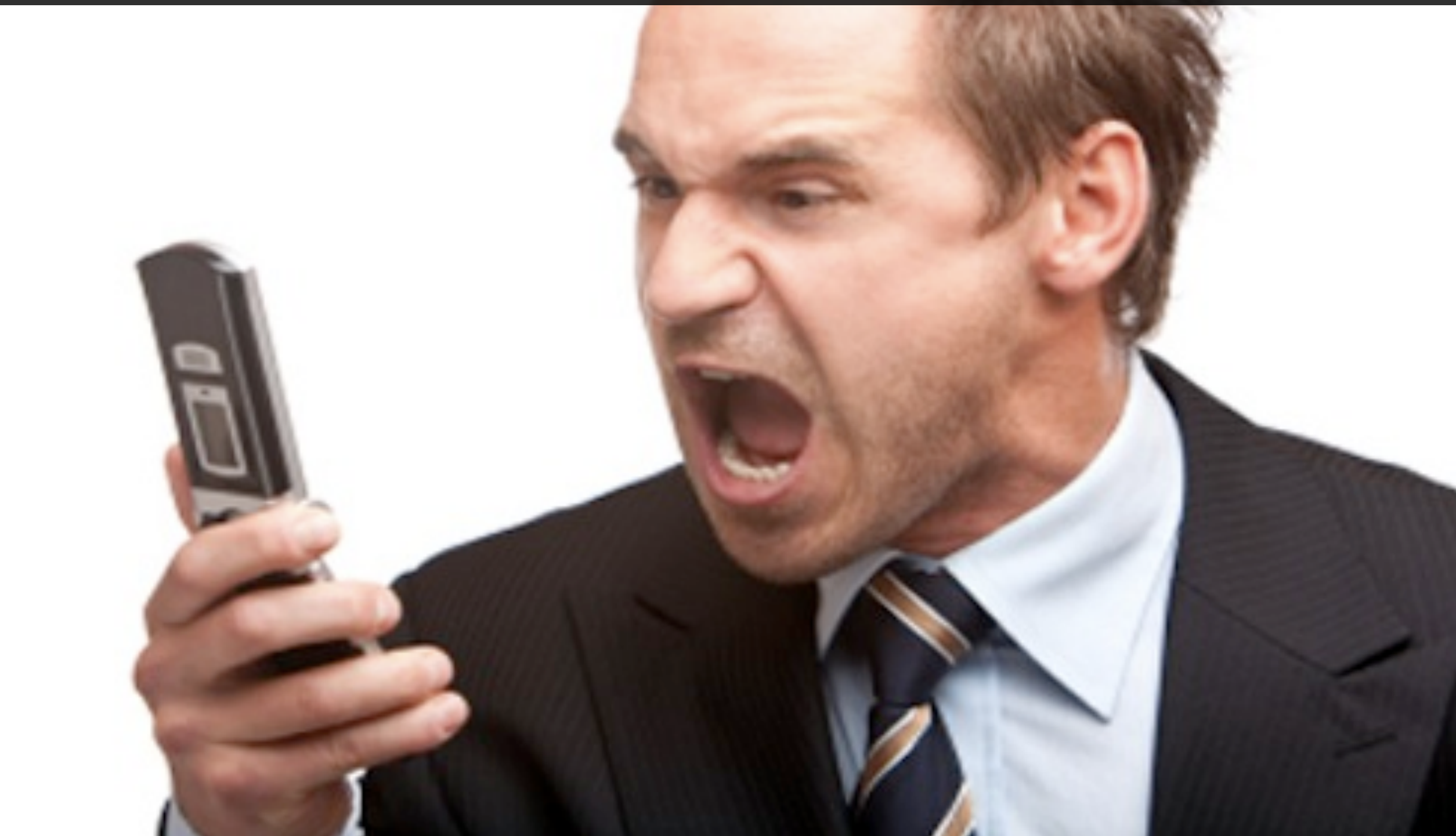
Activity identification: automatic meal photographer

Activity identification: automatic meal photographer

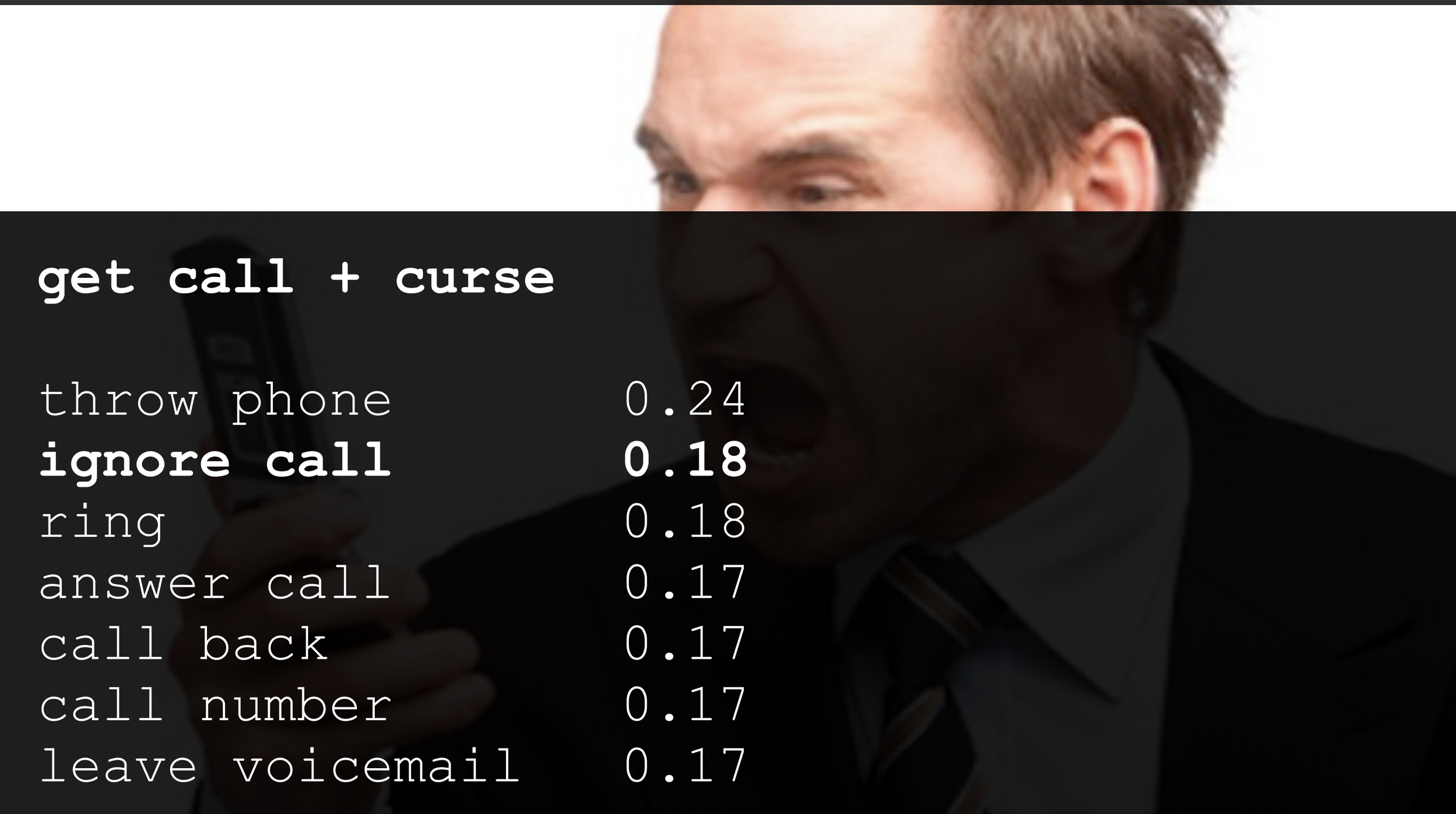
plate + steak + broccoli

fill plate	0.39
put food	0.23
take plate	0.15
eat food	**0.14**
set plate	0.12
cook	0.10

Activity prediction: context aware phone calls



Activity prediction: context aware phone calls



get call + curse

throw phone	0.24
ignore call	**0.18**
ring	0.18
answer call	0.17
call back	0.17
call number	0.17
leave voicemail	0.17

Spending money wisely

enter store

scan	0.19
ring	0.19
pay	**0.17**
swipe	0.17
shop	0.13
buy	0.10

Spending money wisely

call taxi

hail taxi	0.96
pay	**0.96**
take taxi	0.96
get taxi	0.96
tell address	0.95
get suitcase	0.82

Dynamic music player

stove + pot + spoon

cook	**0.50**
pour	0.39
place	0.37
stir	0.37
eat	0.34

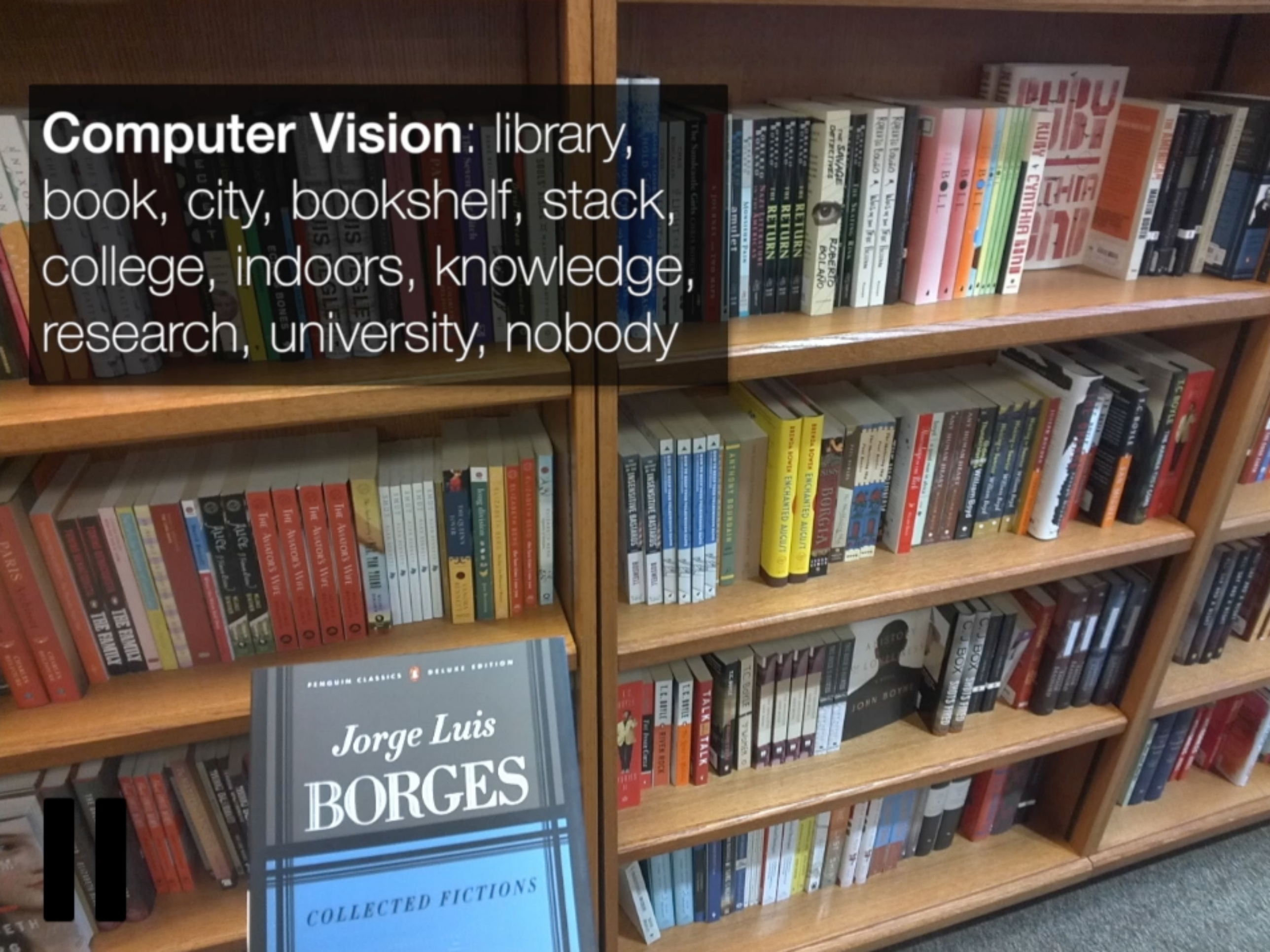
Evaluation

We conducted a two-hour **field deployment** of our dynamic music player, finding **71**% precision and **97**% recall over a set of seven common activities.

Computer Vision: road, car, automobile, vehical, blacktop, traffic, people, building, crash action, driver, pavement

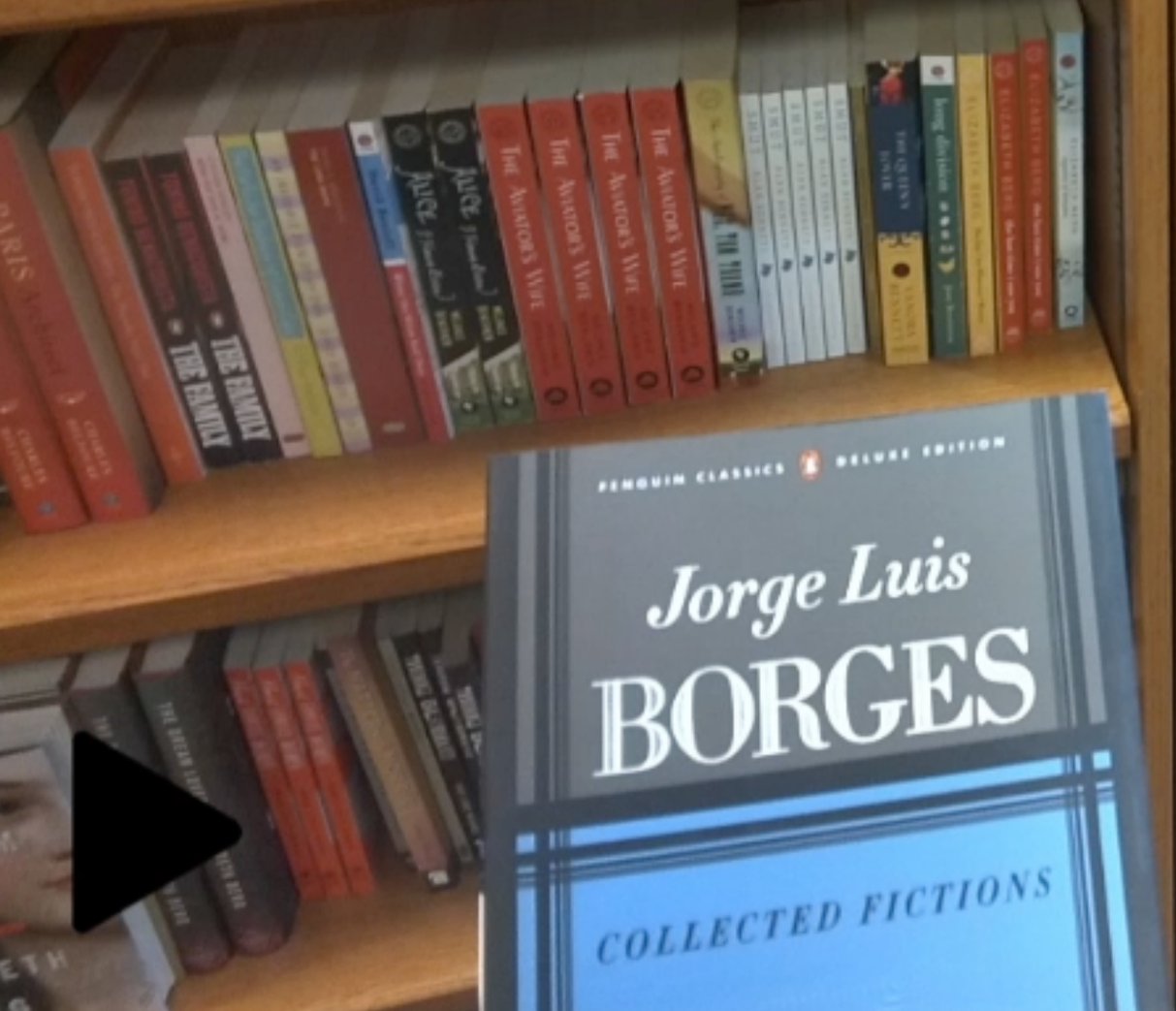
Augur detects: Driving

Now playing "The Engine Driver"
by "The Decemberists"


Computer Vision: library, book, city, bookshelf, stack, college, indoors, knowledge, research, university, nobody
PENGUIN CLASSICS DELUXE EDITION
Jorge Luis
BORGES
COLLECTED FICTIONS

Augur detects: Reading

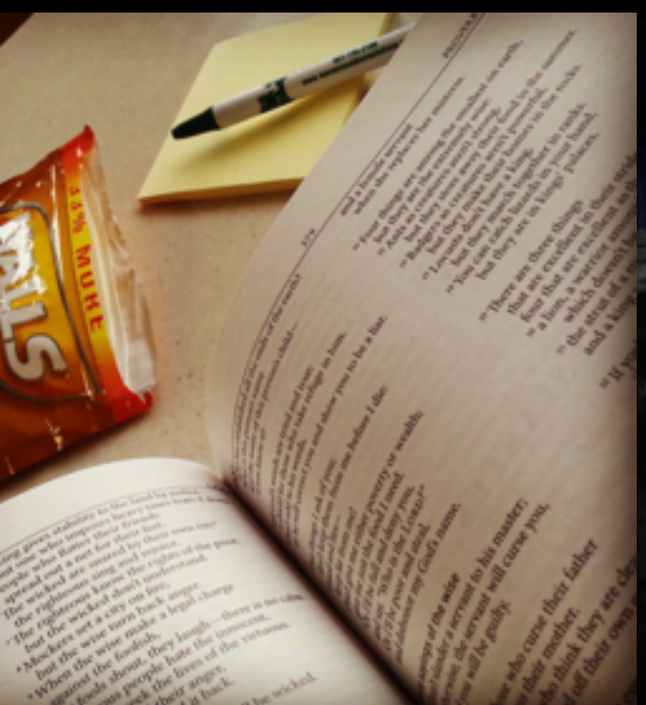
Now playing "Bach Concerto"
by "Hilary Hahn"

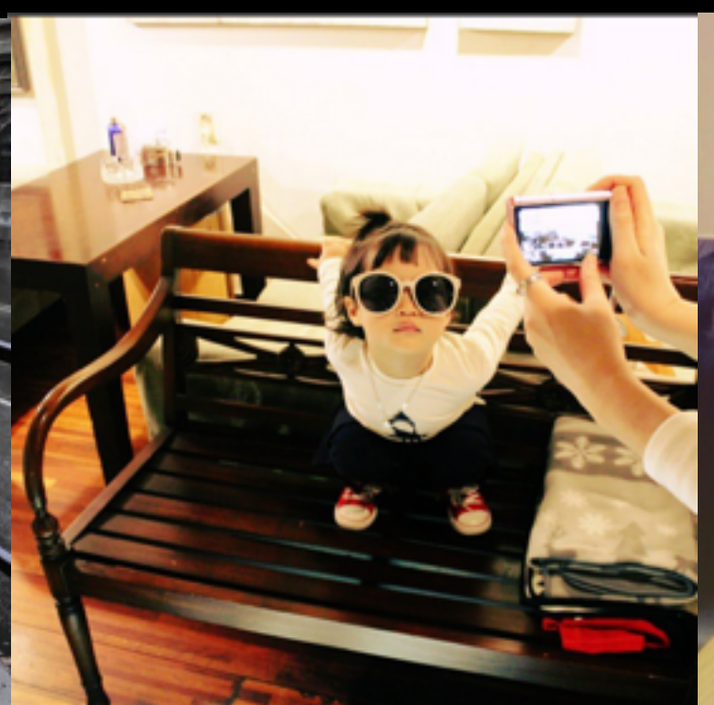



Computer Vision: shopping, urban, city, commerce, store, indoors, path, industry, terminal, beverage, politics
SUMMER CLEARANCE
50%


Augur detects: Buying
Now playing "Trojans"
by "Atlas Genius"
519
SUMMER
CLEARANCE
50%

We **tested** Augur's **predictions** on a dataset of images sampled from the Instagram hashtag **#dailylife**, and found **94**% of predictions were rated as matching the scene.






Chair: sit, pull up, move, throw, …

“Imagine a random person is around a chair 100 times. For each action in this list, estimate how many times that action would be taken.”

Mean absolute error: **12.5%** Augur compared to humans


The Brothers Karamazov
THE
BROTHER
KARAMA

In our most compelling visions of **human-computer interaction**, computers understand the **breadth** of **human life**.

How can we teach **computers** to understand the broad set of things **people** do?

Thank you to NSF and Toyota for funding support.