LISBON 2025
Doctoral Training Network - Annual Forum

Think : Data won't tell the full story

... for now !

João Paulo Costeira

Signal and Image Processing Group

SIPG
SIGNAL AND IMAGE PROCESSING GROUP

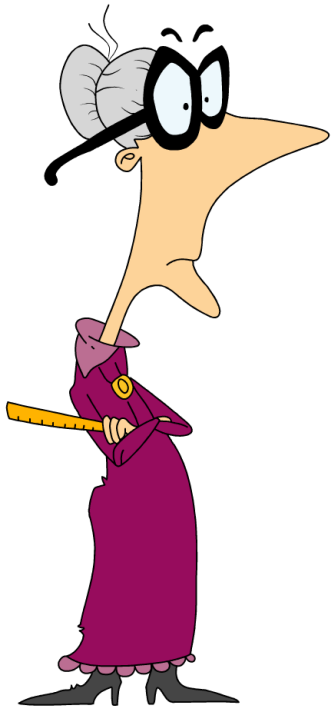
Institute for Systems and Robotics
Lisboa

TÉCNICO LISBOA

Outline

Inspired by two articles … I'm going to preach !

-That's what "seniors" do!



- How "AI" works (and why it can be dangerous for you !)

- The I (in AI) does not stand for "Intelligence"…

But then …

I'll (try) to show how exciting these times can be …

why everybody should appropriate the "AI"

and how I am contributing to put the "I" back on AI !

…in a journey with truly Intelligent students and colleagues

Gabriel Moreira

Manuel Marques

But then …

I'll (try) to show how exciting these times can be …

why everybody should appropriate the "AI"

and how I am contributing to put the "I" back on AI !

…in a journey with truly intelligent students and colleagues

Because …

A guy 10 months junior !

Gabriel Moreira

Manuel Marques

… that's what people my age like to do !

I am a vision…ist !

1987

Vision guided this vehicle since …

Boomers had self-drive long time ago!

My last “classic” !

Gabriel Moreira, Manuel Marques, J. Costeira ,Rotation Averaging in a Split Second: A Primal-Dual Method and a Closed-Form for Cycle Graphs, IEEE International Conference on Computer Vision, ICCV 2021,

Prof. Takeo Kanade

Prof. João Sentieiro

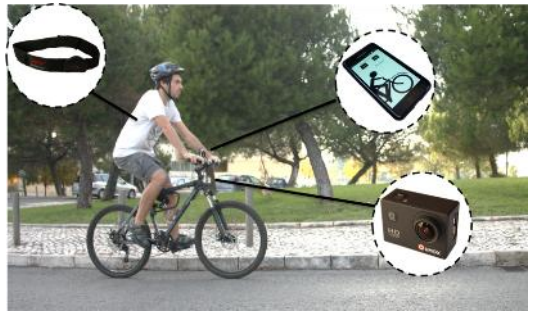
I “lived” with excitement

I ICCV #1 in London

1987

40 years of “AI” when AI was called…Signal Processing

http://printart.isr.ist.utl.pt/citycam/



http://www.isr.ist.utl.pt/~manuel/smartbike

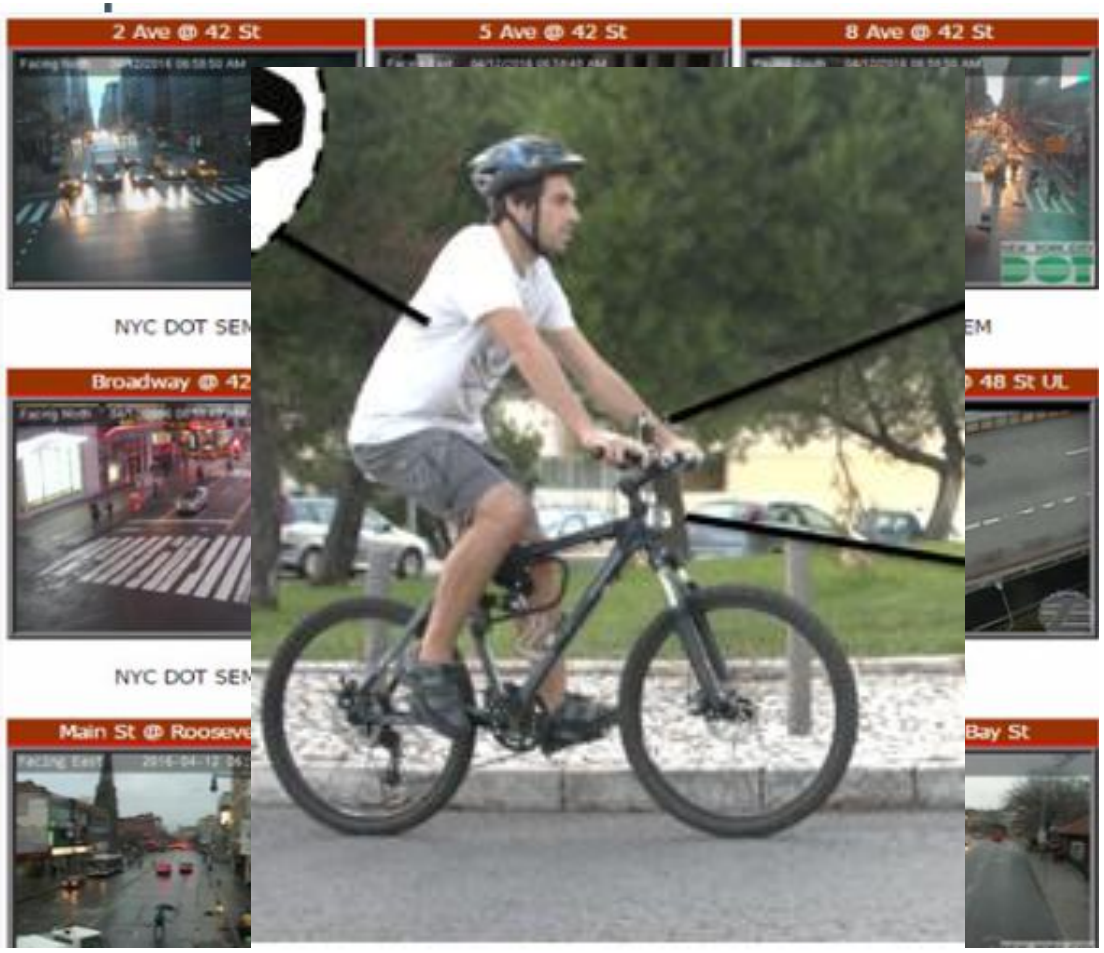
http://physicalai.isr.tecnico.ulisboa.pt/

http://ai4eu.isr.tecnico.ulisboa.pt/pivision/

How did I land in Transport/Mobility… dealing with *people*

Traffic & Cities – 2012-2018

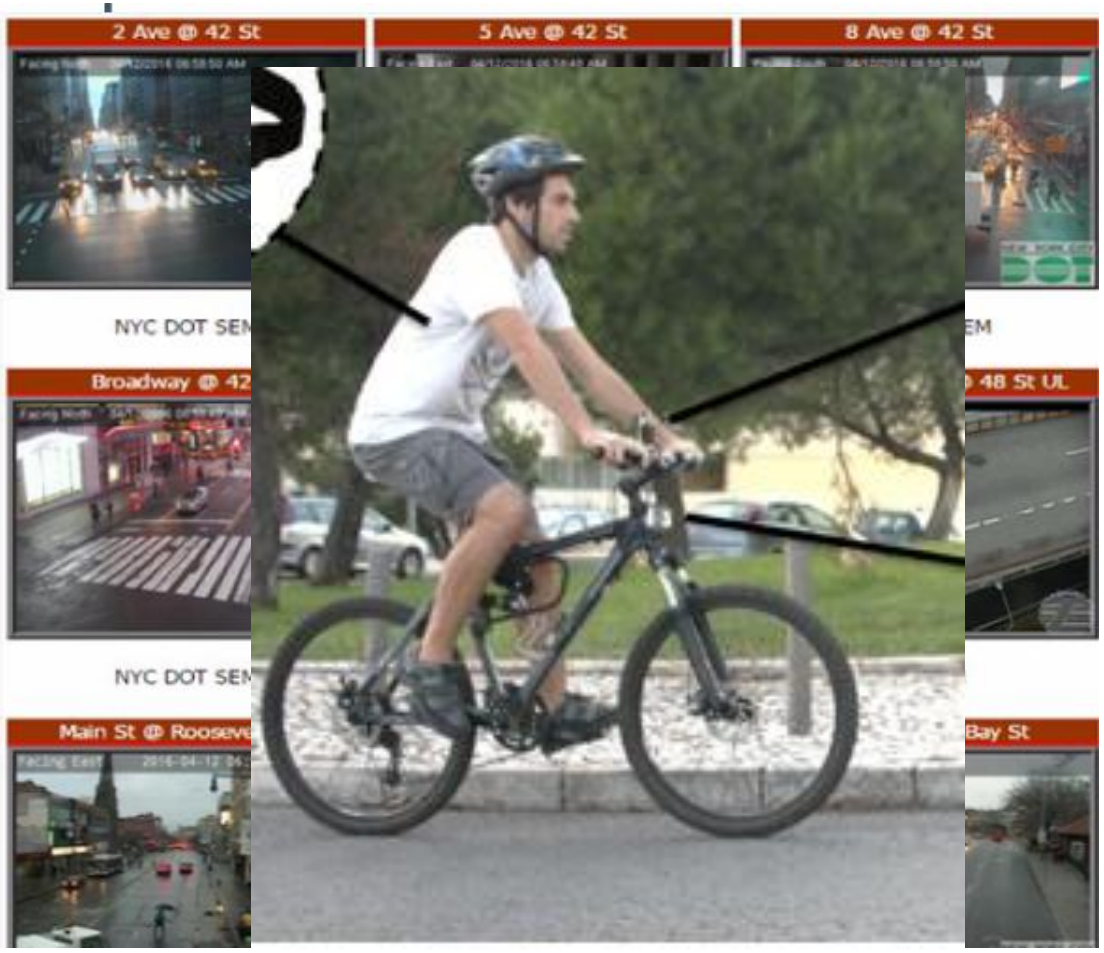
How did I land in Transport/Mobility… dealing with *people*



Traffic & Cities – 2012-2018

The “human factor” pops up! 2016

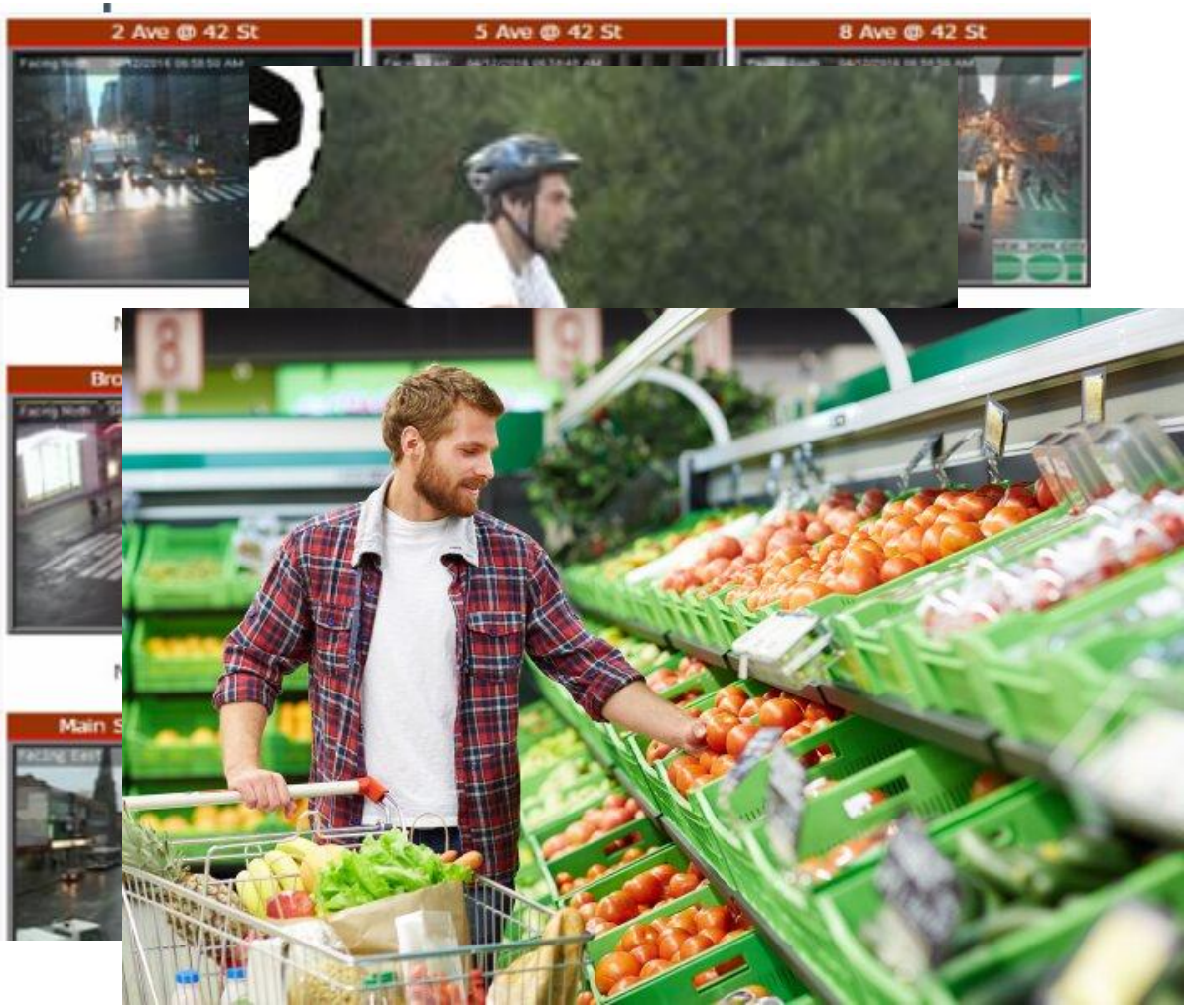
How did I land in Transport/Mobility… dealing with *people*



Traffic & Cities – 2012-2018

The “human factor” pops up! 2016

How did I land in Transport/Mobility… dealing with *people*



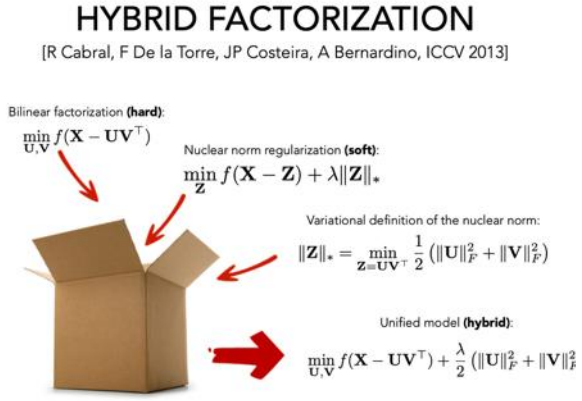
Traffic & Cities – 2012-2018

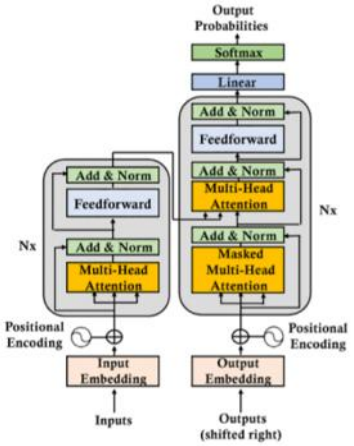
The “human factor” pops up! 2016

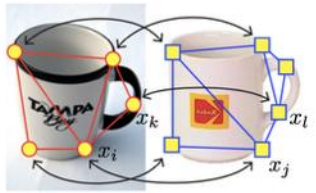
… perplexity by human decision-making (until today!)

Curiosity and Discovery
Decision is a costly process !

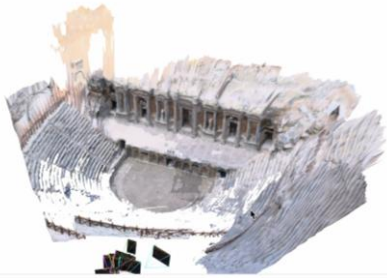
Tech: How it is done !





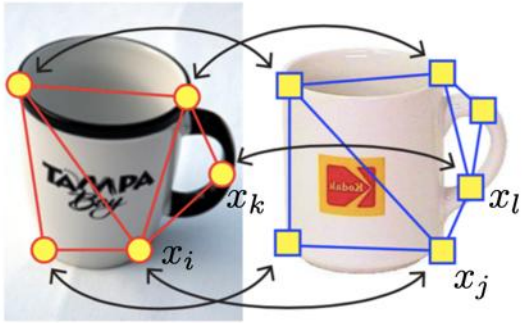


$$\begin{aligned}\max_{\hat{\mathbf{X}}} \quad & \text{trace}(\hat{\mathbf{K}}\hat{\mathbf{X}}) \\ \text{subject to} \quad & \hat{x}_{1,1}=1 \\ & 2\text{Diag}(\hat{\mathbf{X}})=\hat{\mathbf{X}}_{1,:}+\hat{\mathbf{X}}_{:,1}^\top \\ & \mathbf{H}\text{diag}(\mathbf{X})=\mathbf{1}_m \\ & \mathbf{X}\odot\mathbf{M}=\mathbf{0} \\ & \hat{\mathbf{X}}\succeq 0, \\ & \text{rank}(\hat{\mathbf{X}})=1,\end{aligned}$$



Humans: Understanding “us”

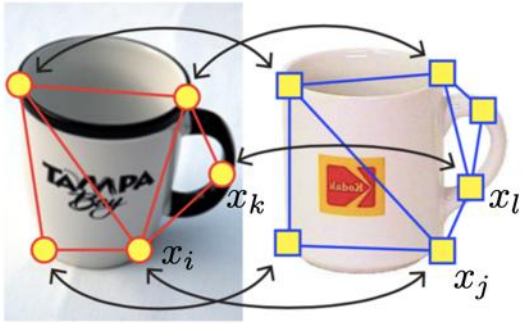
A bit of both!



Tech: How it is done !

- Neural Networks (i.e.) Signal Proc. In City Traffic
- Transformers, LLM’s and (damn) GPT !
- Representations: Can´t think without one !

A bit of both!



Tech: How it is done !

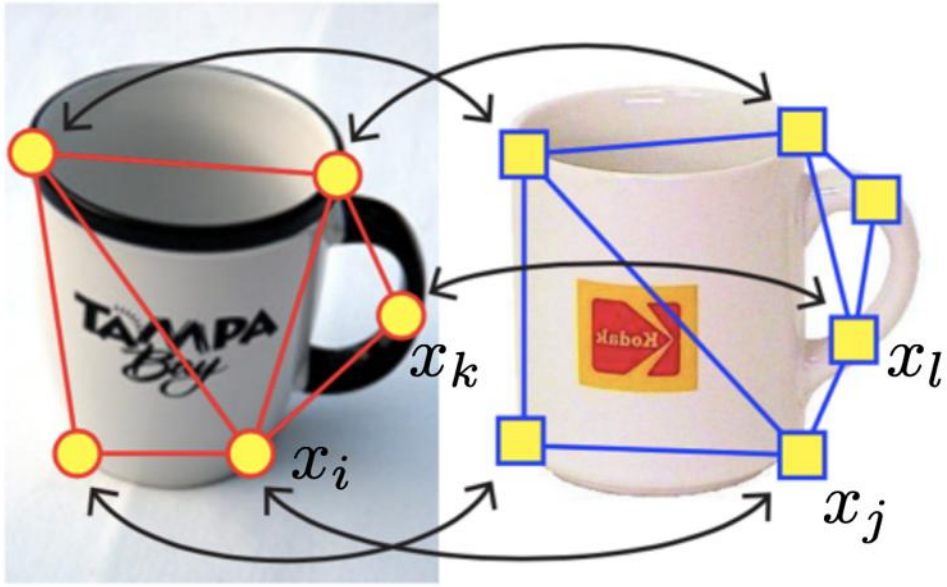
- Neural Networks (i.e.) Signal Proc. City Traffic
- Transformers, LLM's and (damn) GPT !
- Representations: Can ´t think without one !

Humans: What is tech for !

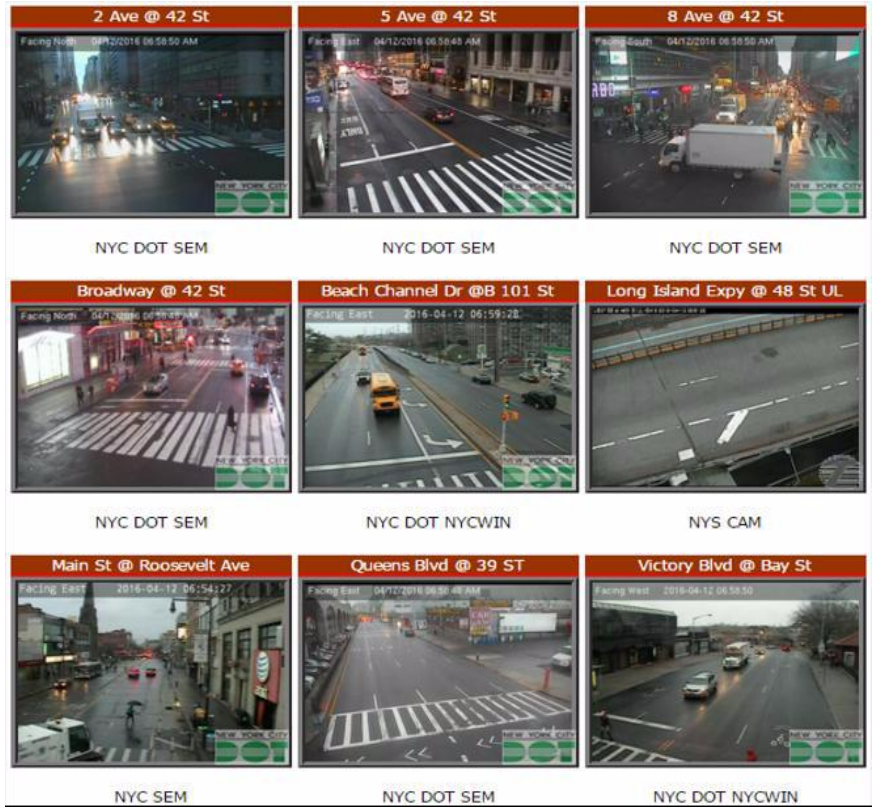
- What data says about people
- Rationality is overrated !
- Model vs data and the role of language

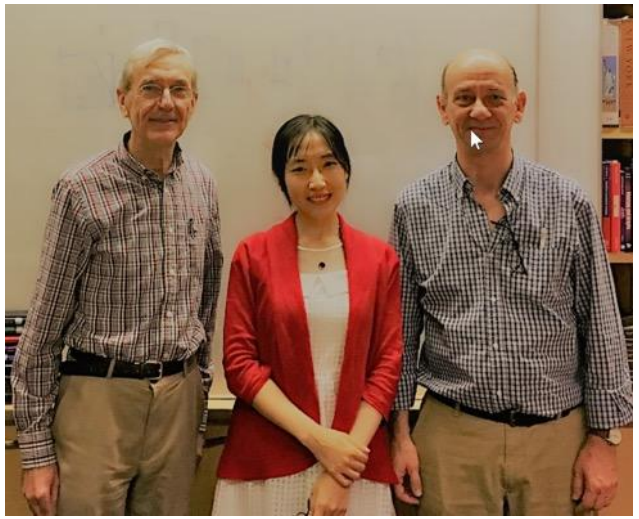
Opportunities created by data availability

- for the first movers -



The NYC camera network 2013-2018

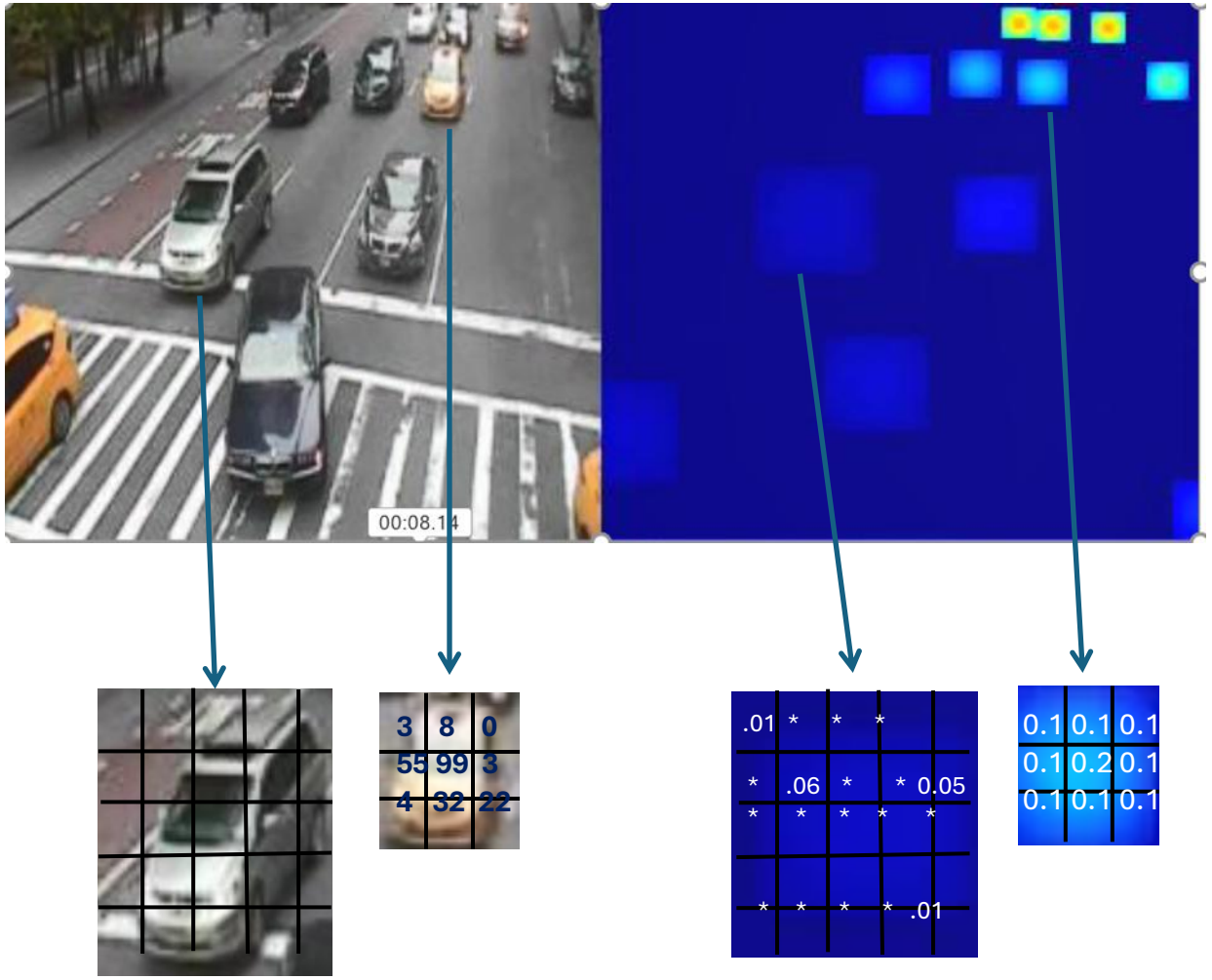




Jose' Moura CMU,
Shanghan Zhang CMU/IST

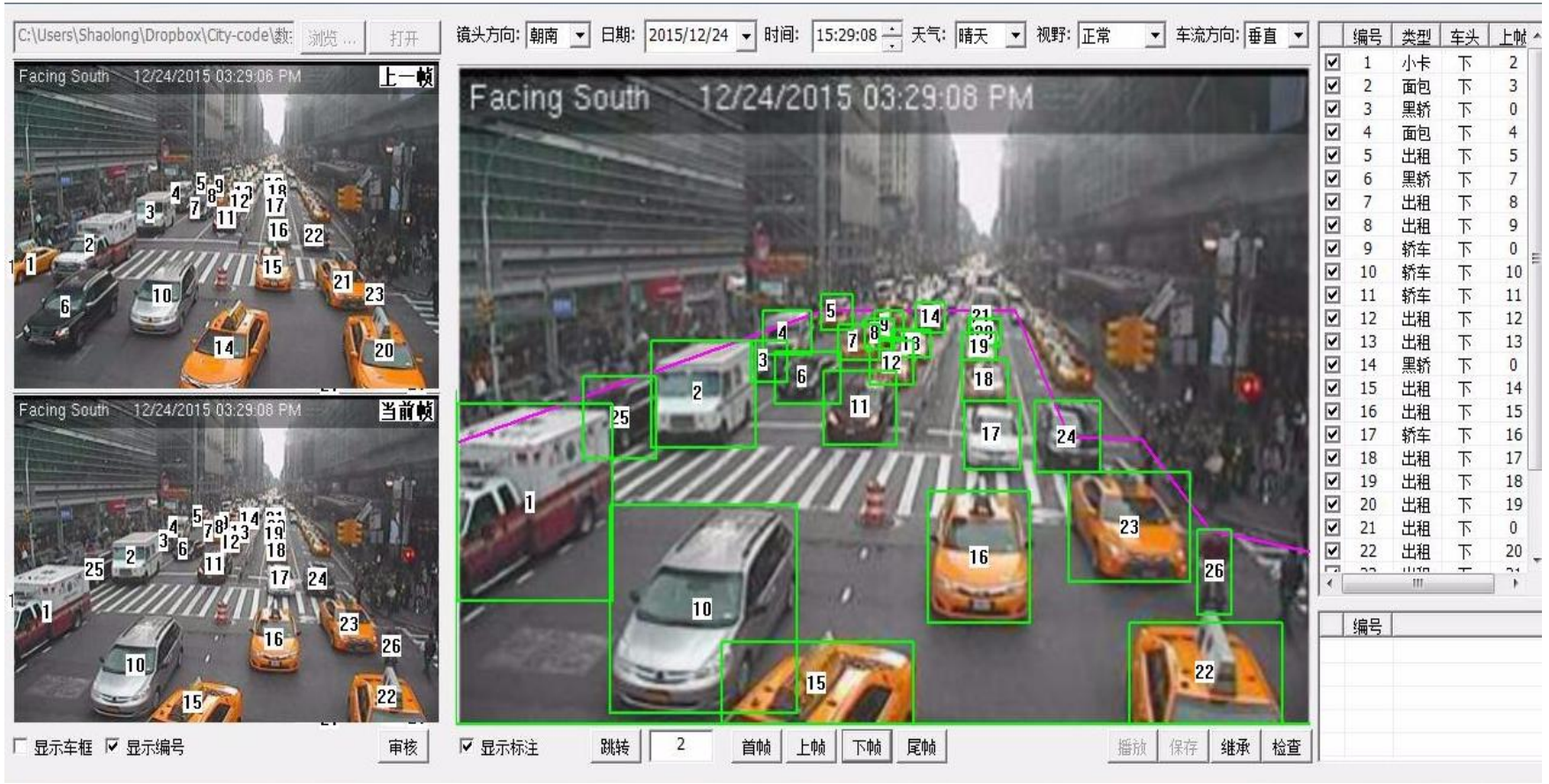
Information and Communication Technologies Institute
Carnegie Mellon | PORTUGAL
AN INTERNATIONAL PARTNERSHIP

Images are “just numbers”: map it to density!



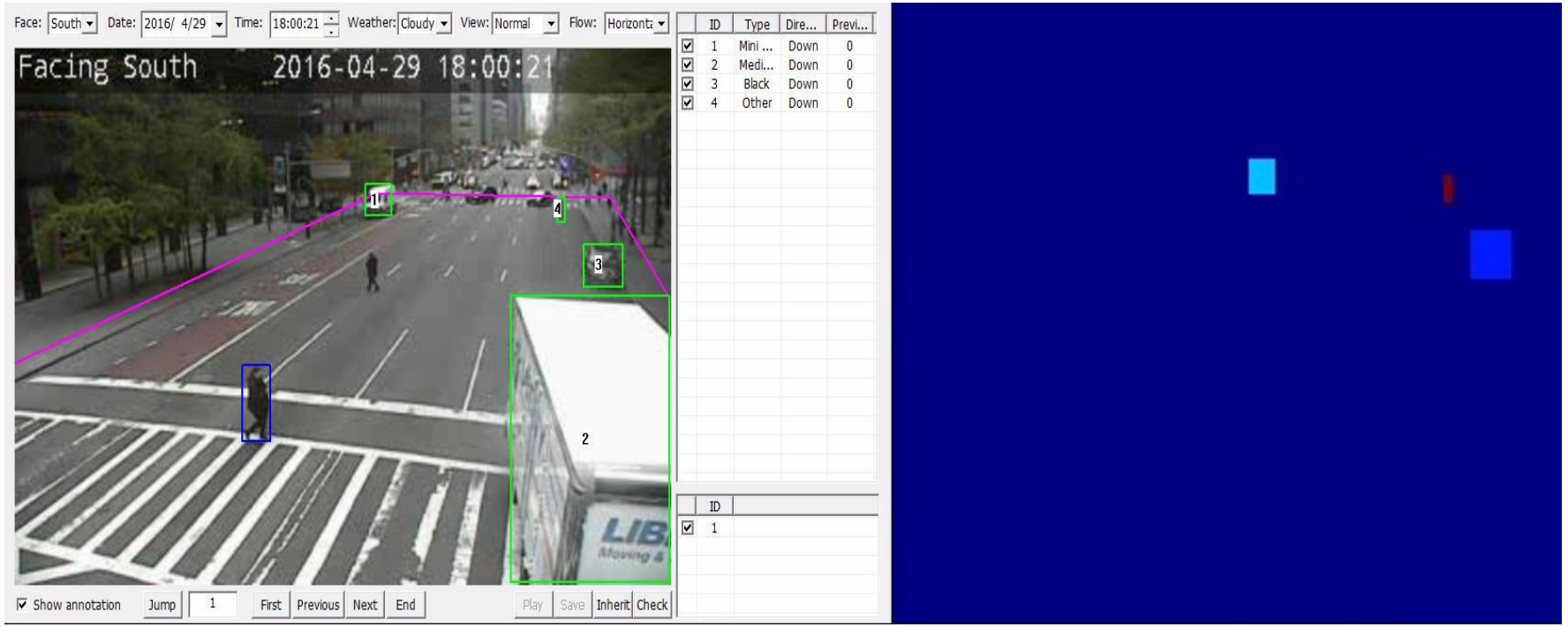
CityCam Dataset

- **60,000** annotated frames with around **900,000** labeled vehicles
- Rich annotations: vehicle count; type; bbox; re-id; orientation; time; weather
- **60,000,000** images for testing

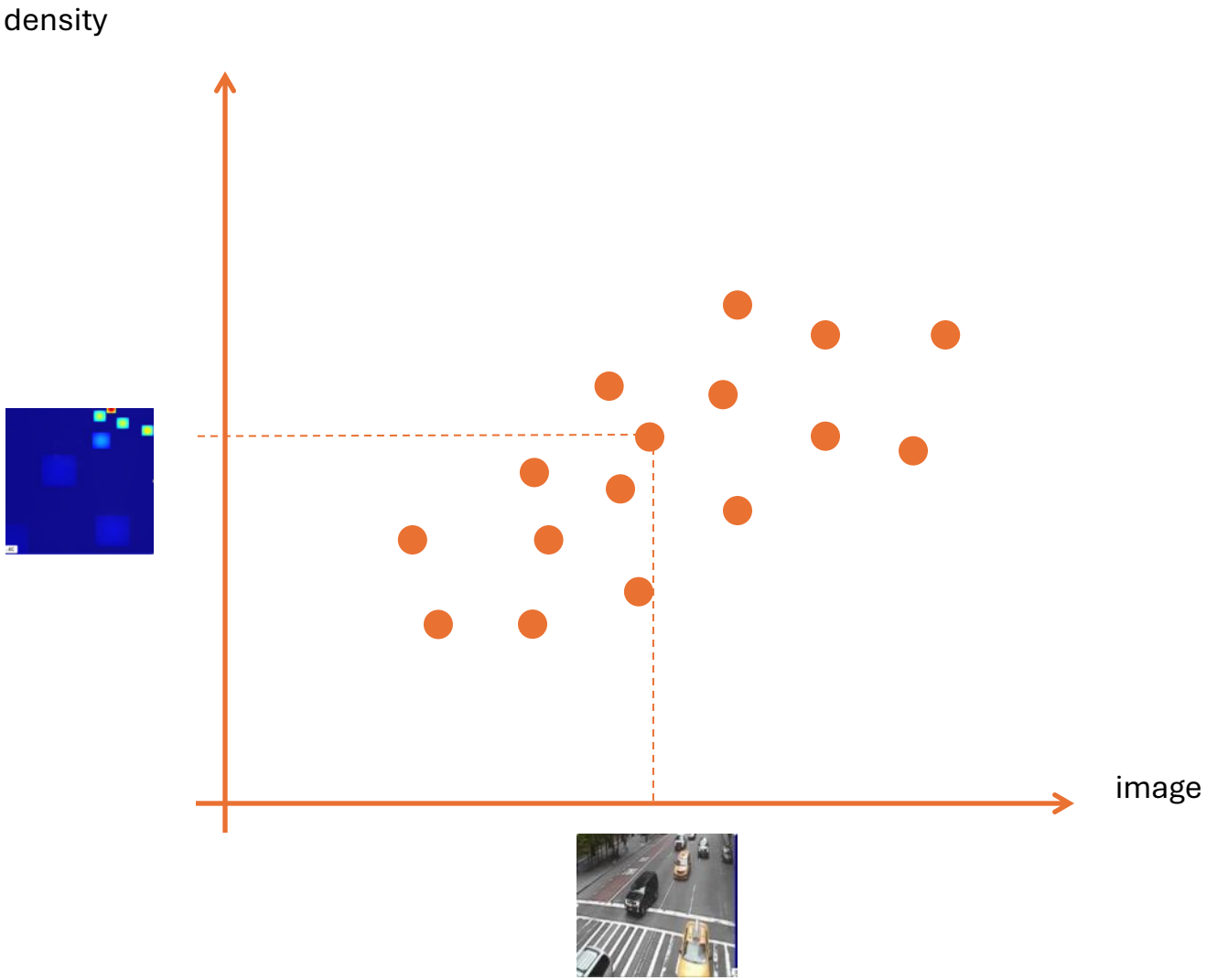


http://printart.isr.tecnico.ulisboa.pt/citycam

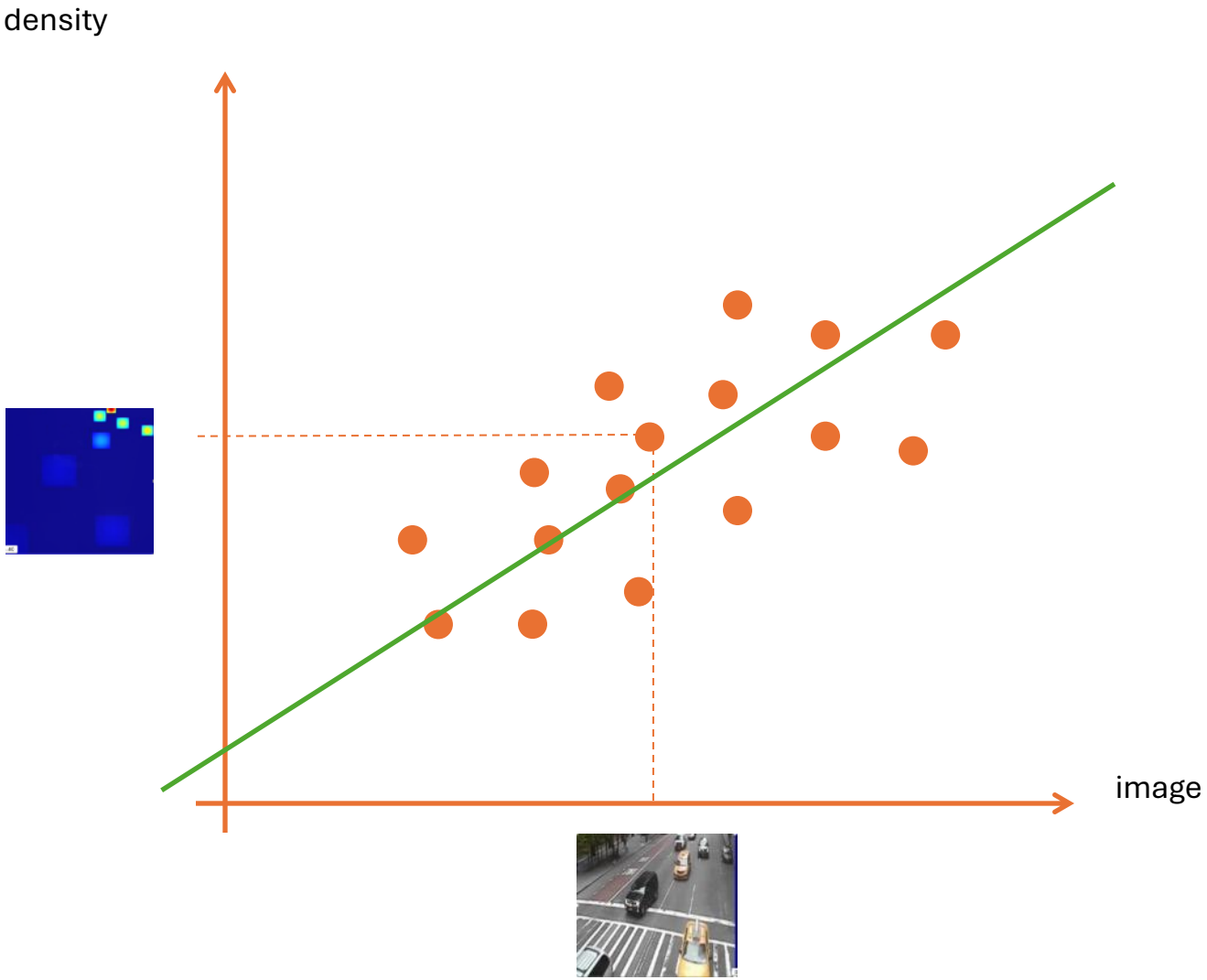
Training Data: Annotation vs “car density”



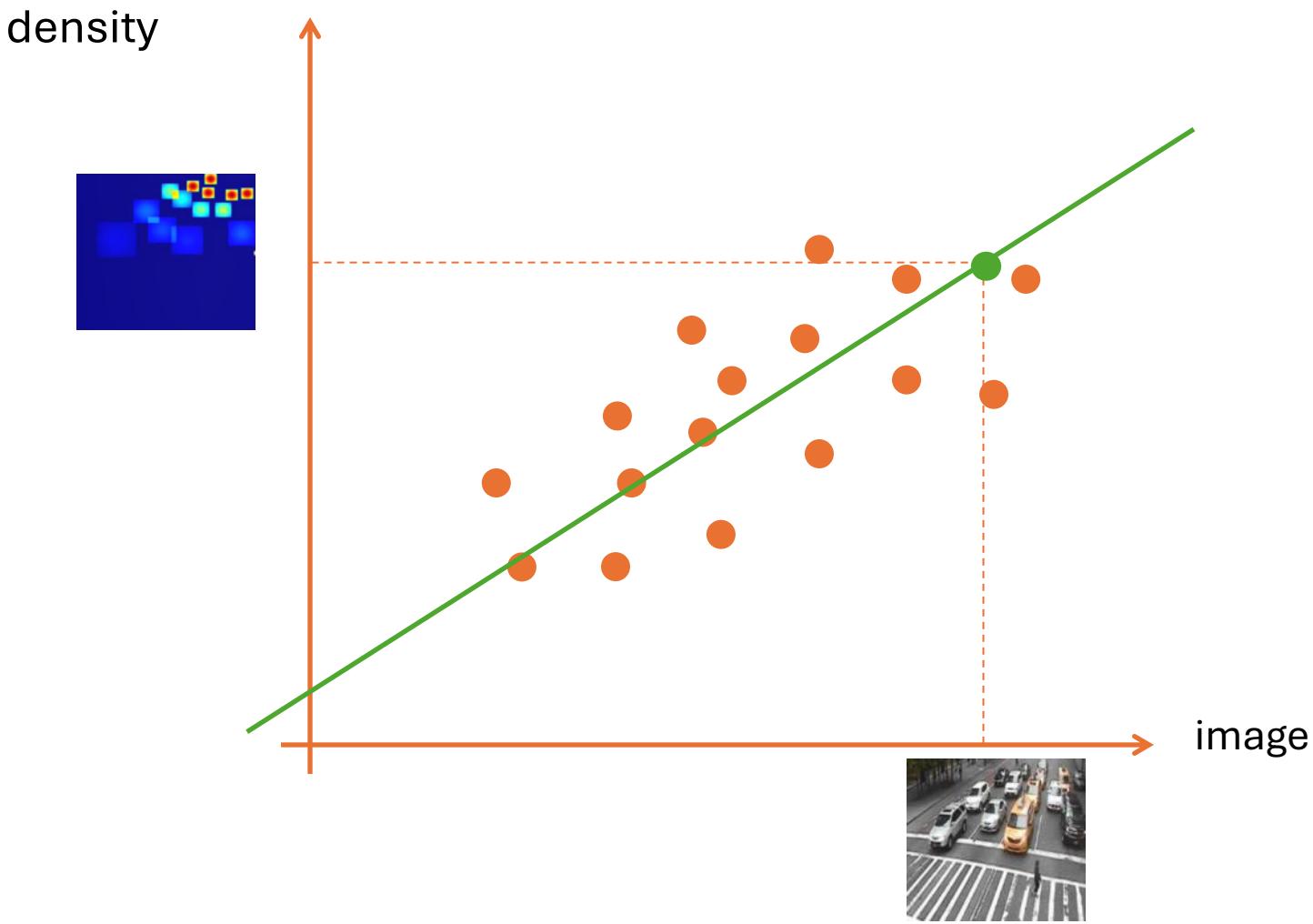
Learning to count



Learning to count by (linear) regression



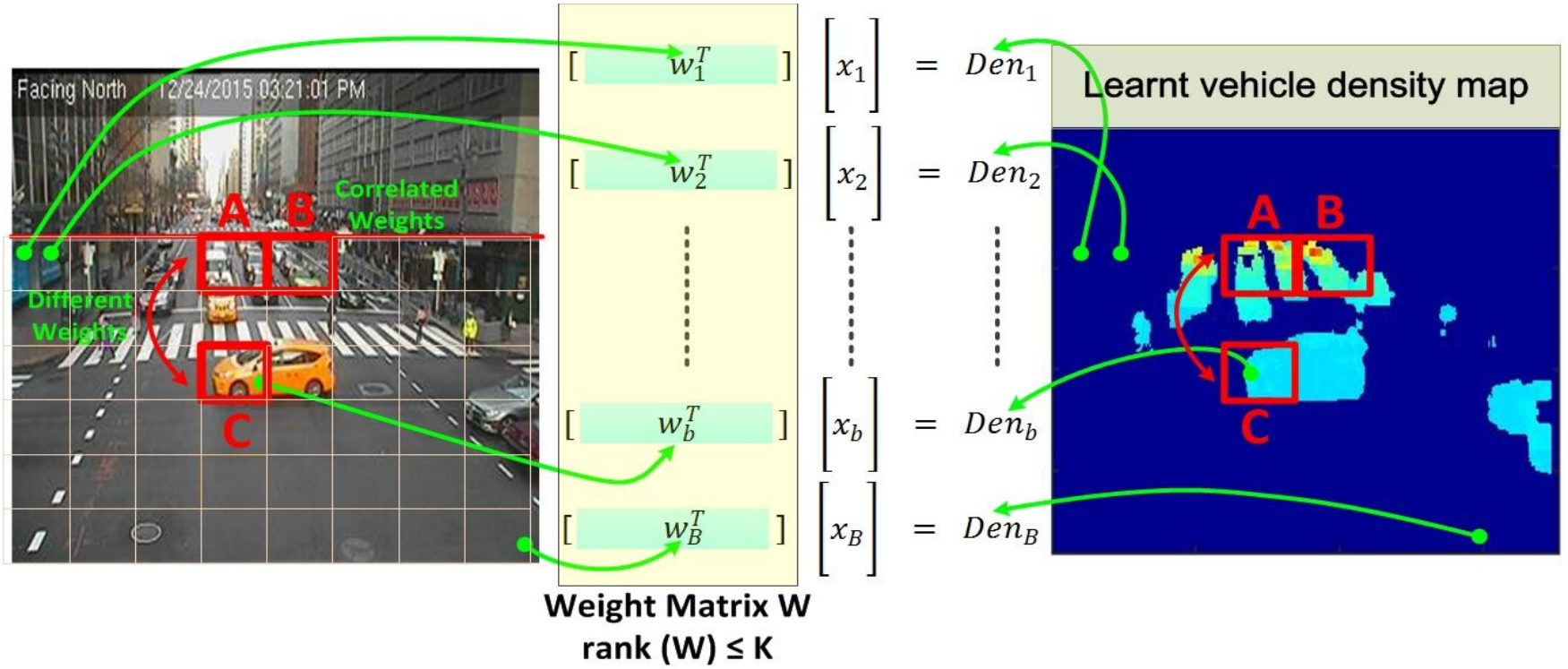
Infer counting



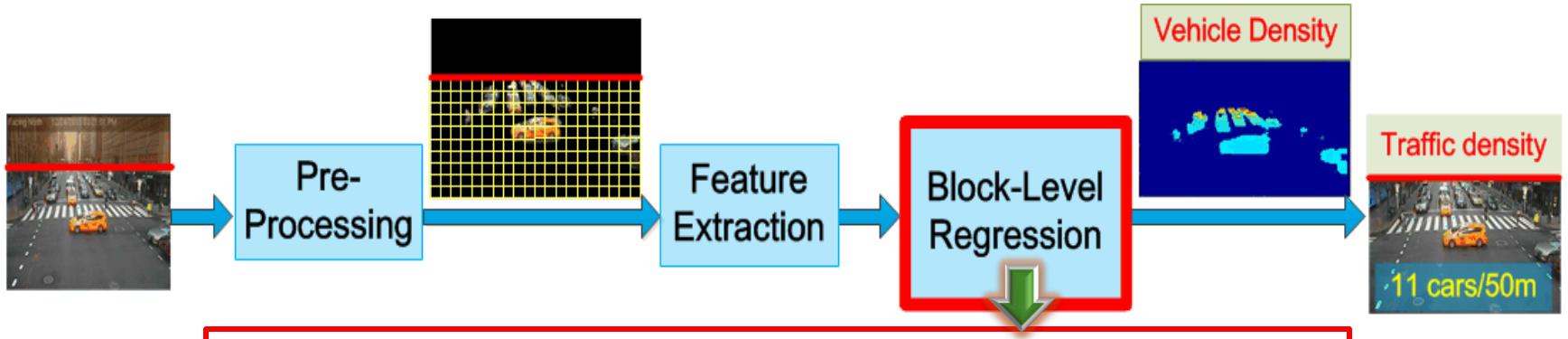
Algebraic solution – Large Scale **Low Rank** Regression

Block-level regression with rank constraint.

Embed road geometry in the weight matrix.

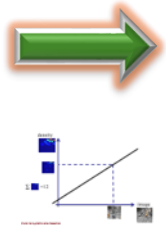


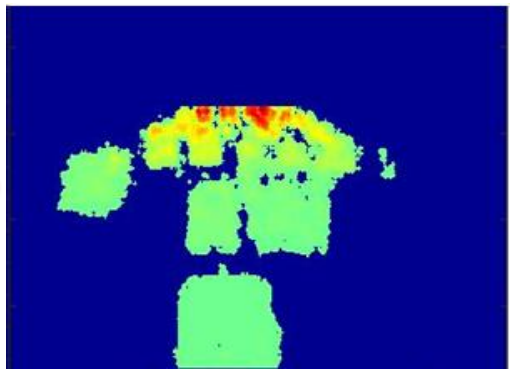
Optimization : Density estimation by **Low Rank** Regression

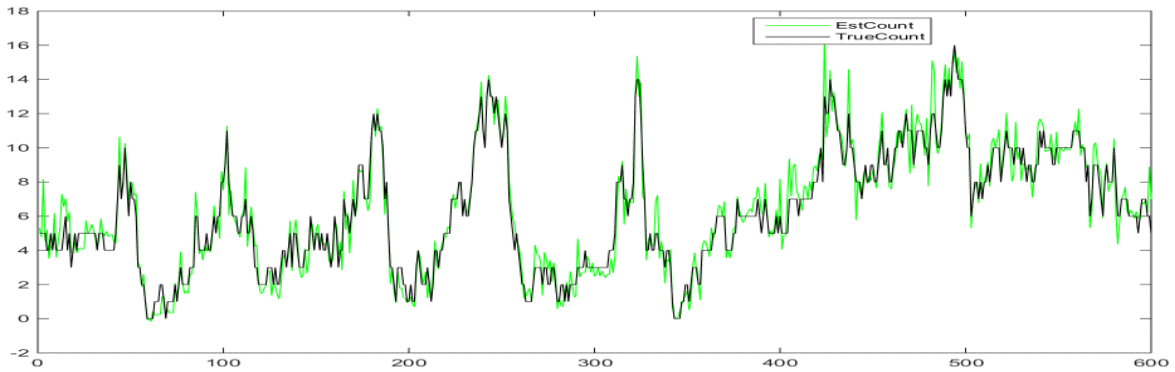


$$\min_{\mathbf{W}} \frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1}^{J} (\mathbf{w}_j^\top \mathbf{x}_j^{(i)} - D(B_j^{(i)}))^2 + \alpha \|\mathbf{W}\|_F^2 + \beta |W|_1$$

$$\text{s.t.} \quad \text{rank}(\mathbf{W}) \leq r$$





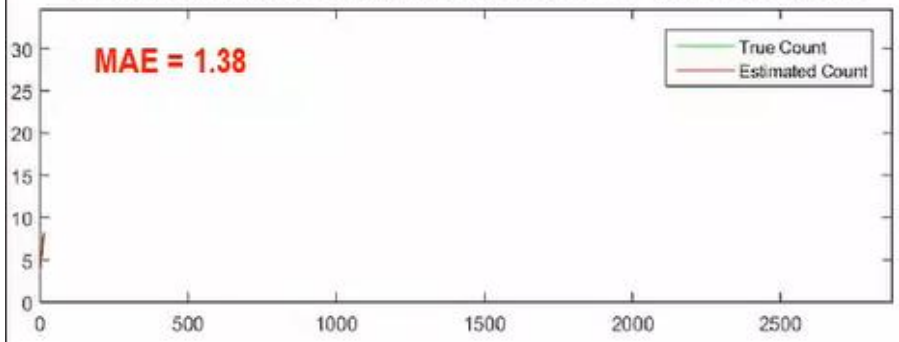


Results of FCN-rLSTM for the downtown cameras

Camera 410

Original Image | Estimated Density Map

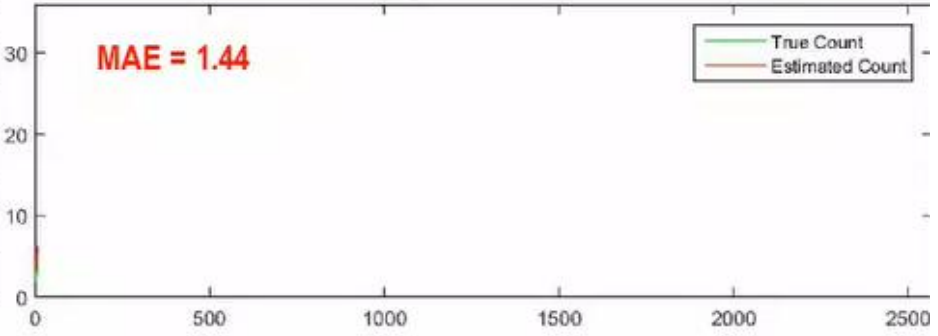
Estimated vehicle count for different time and conditions



Camera 511

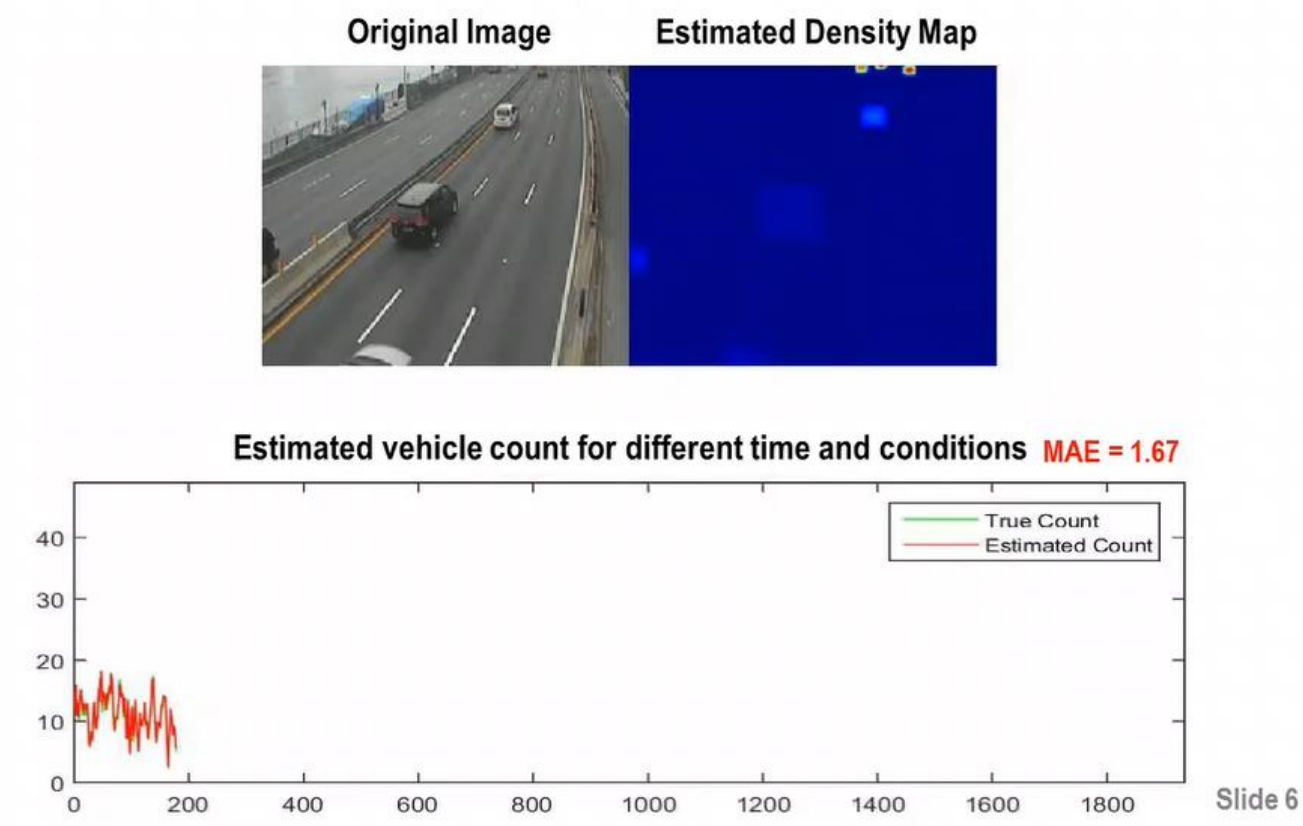
Original Image | Estimated Density Map

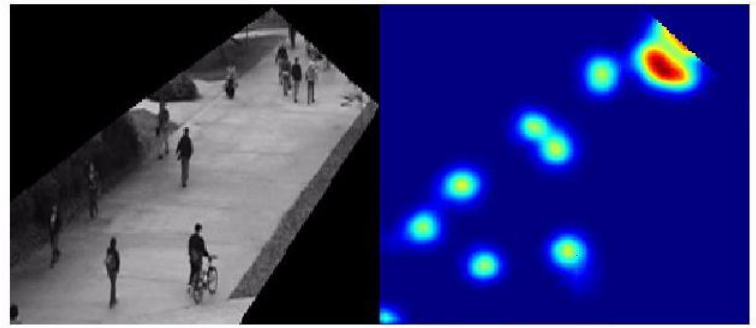
Estimated vehicle count for different time and conditions

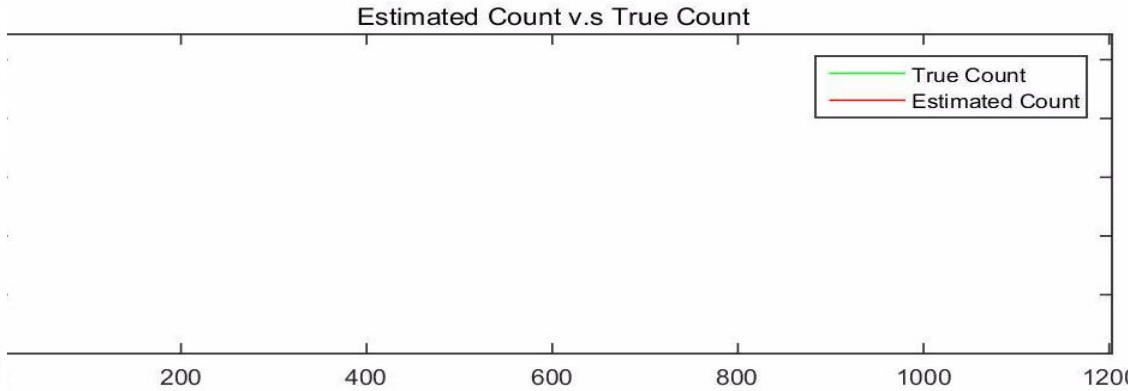


BTW, can count people too ! It’s the DATA !

- **Results of FCN-rLSTM for the parkway camera 691**

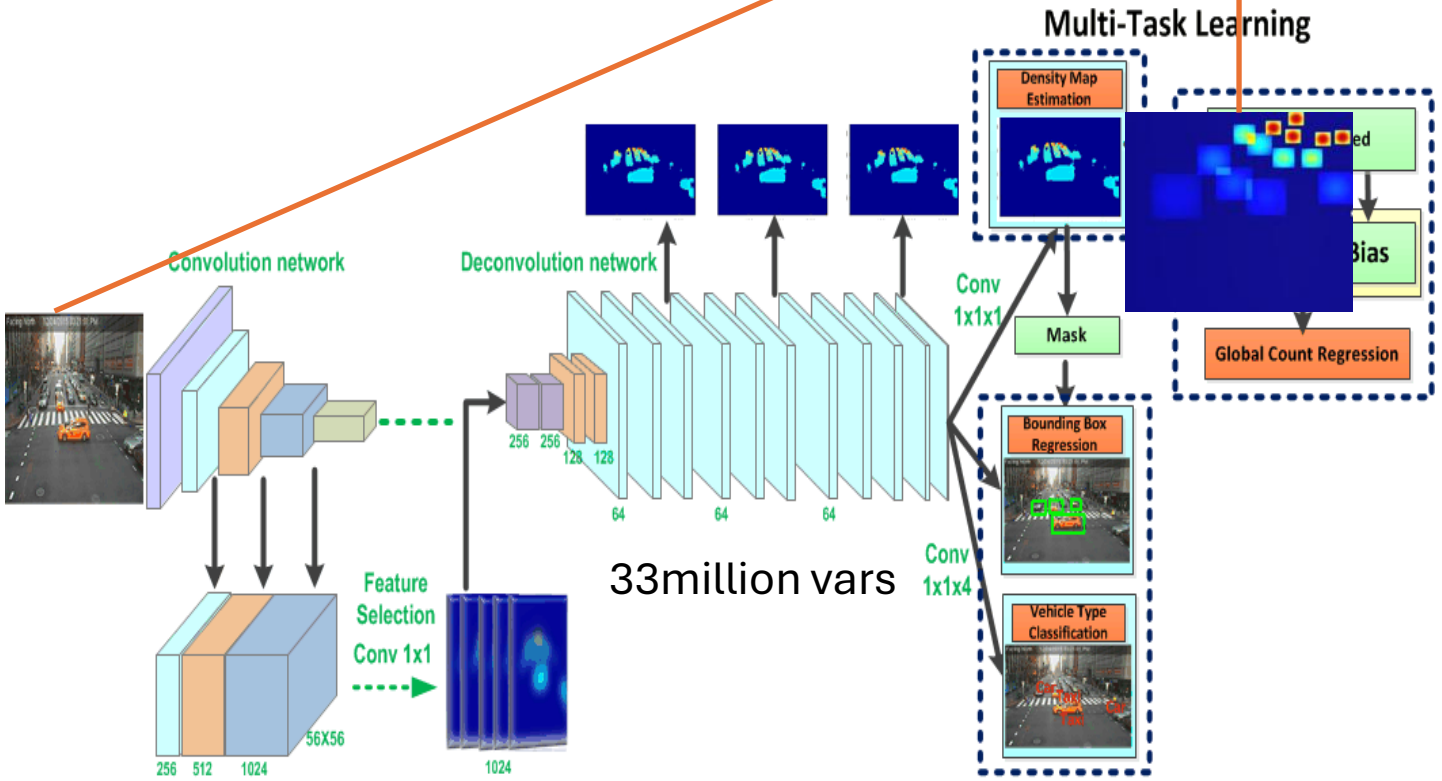






Regression with Deep Neural Networks

$$L_D(\Theta) = \frac{1}{2N} \sum_{i=1}^{N} \sum_{p=1}^{P} \| F\left(X_i(p); \Theta\right) - F_i(p) \|_2^2$$

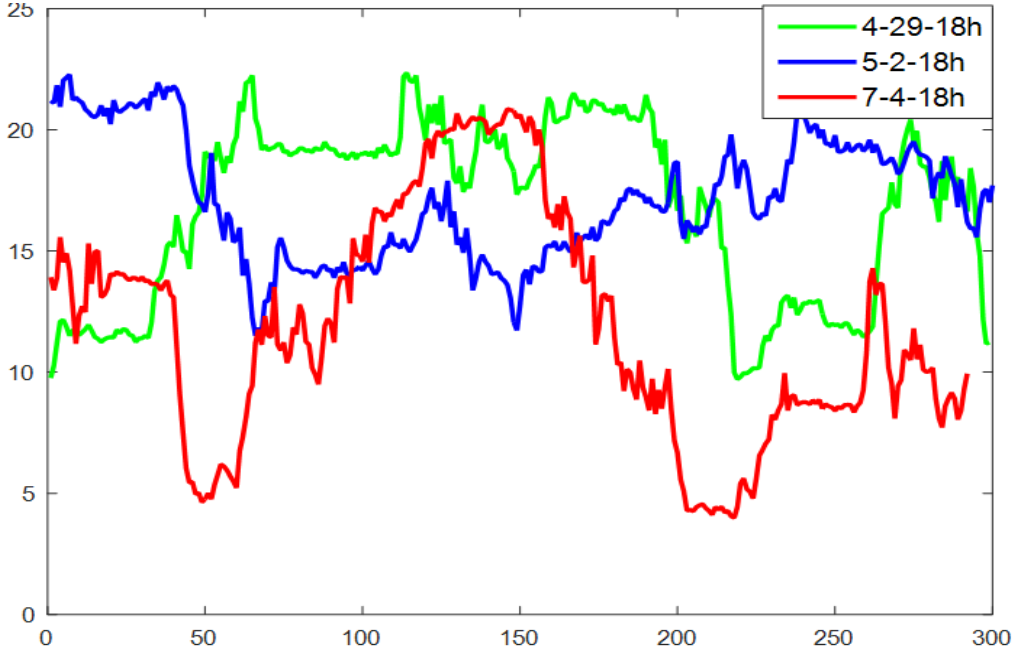


PCT Patent No: PCT/US18/26341; Filed: April/5/2018

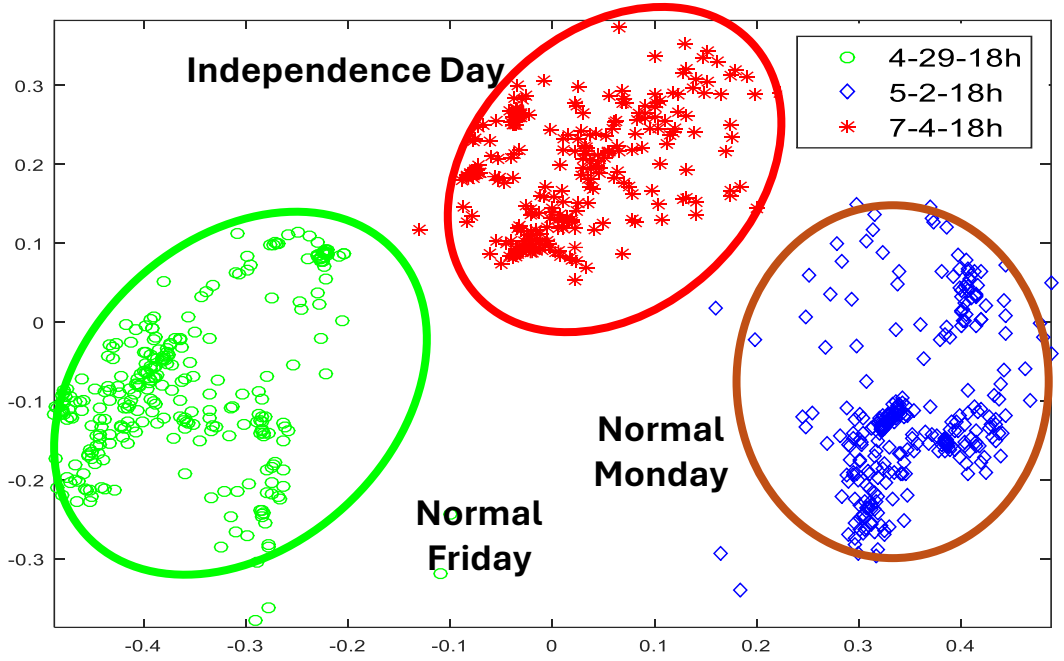
Title: Deep Learning Methods for Estimating Density and/or Flow of Objects, and Related Methods and Software

Applicants: Carnegie Mellon University and Instituto Superior Técnico

The pulse of the city

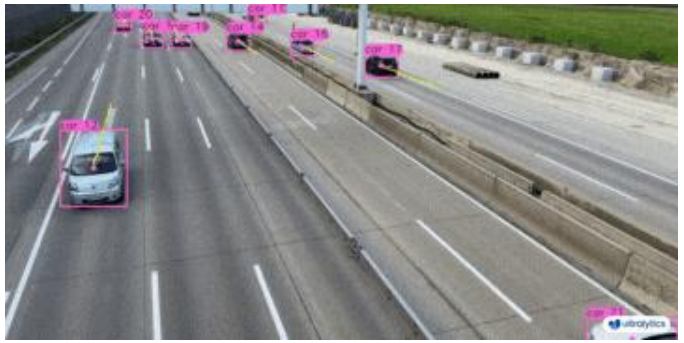


Estimated Count for different days



Isomap of vehicle density for different days

It’s end-to-end ! Fantastic … no need to know!



Tracking

3D reconstruction

How data reveals Human behaviors…indirectly !

What cars tell about people !

a case of proxy-sensing

Visual Census: Using Cars to Study People and Society

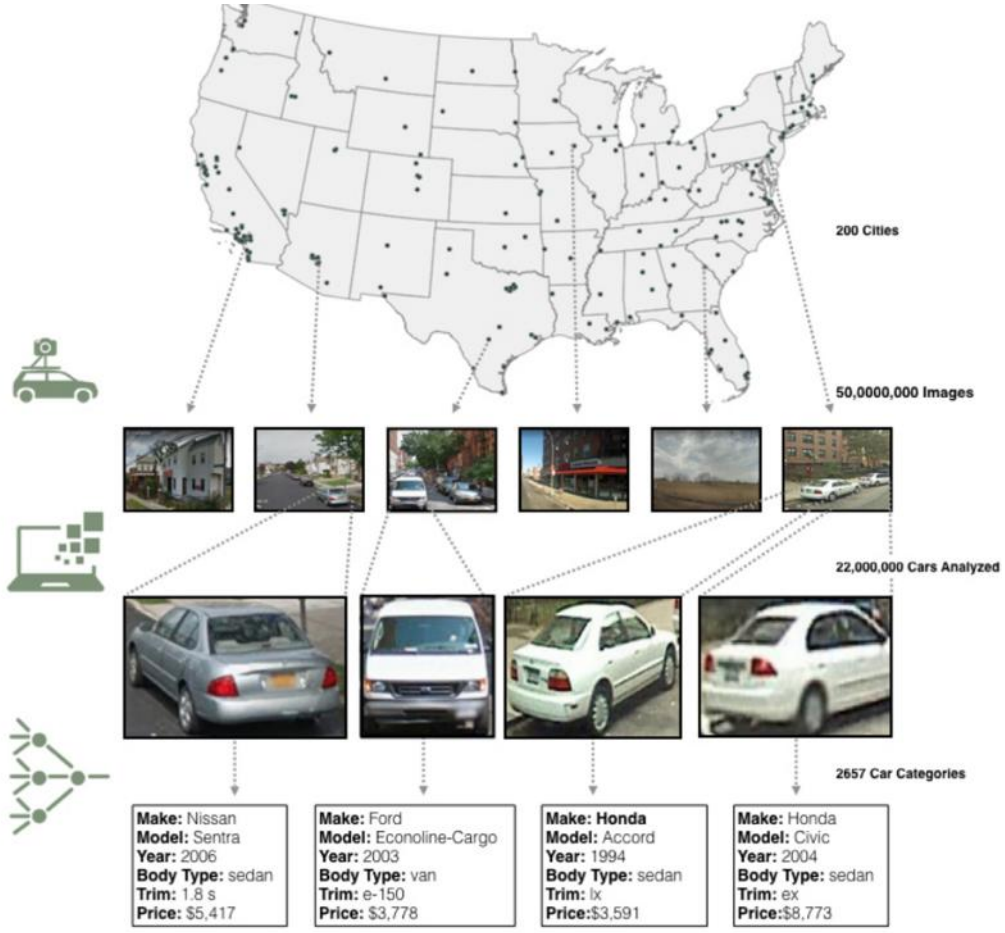
Timnit Gebru[1] Jonathan Krause[1] Yilun Wang[1] Duyun Chen[1] Jia Deng[2] Li Fei-Fei[1]
[1]Stanford University [2]University of Michigan

RESEARCH ARTICLE | COMPUTER SCIENCES | OPEN ACCESS

Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States

Timnit Gebru, Jonathan Krause, Yilun Wang, +3, and Li Fei-Fei Authors Info & Affiliations

November 28, 2017 | 114 (50) | https://doi.org/10.1073/pnas.1700035114



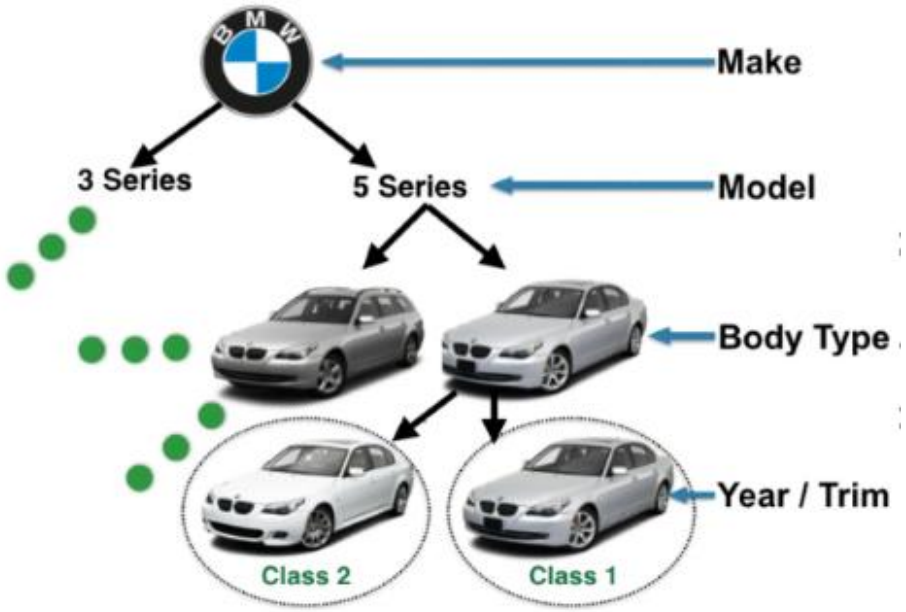
200 cities

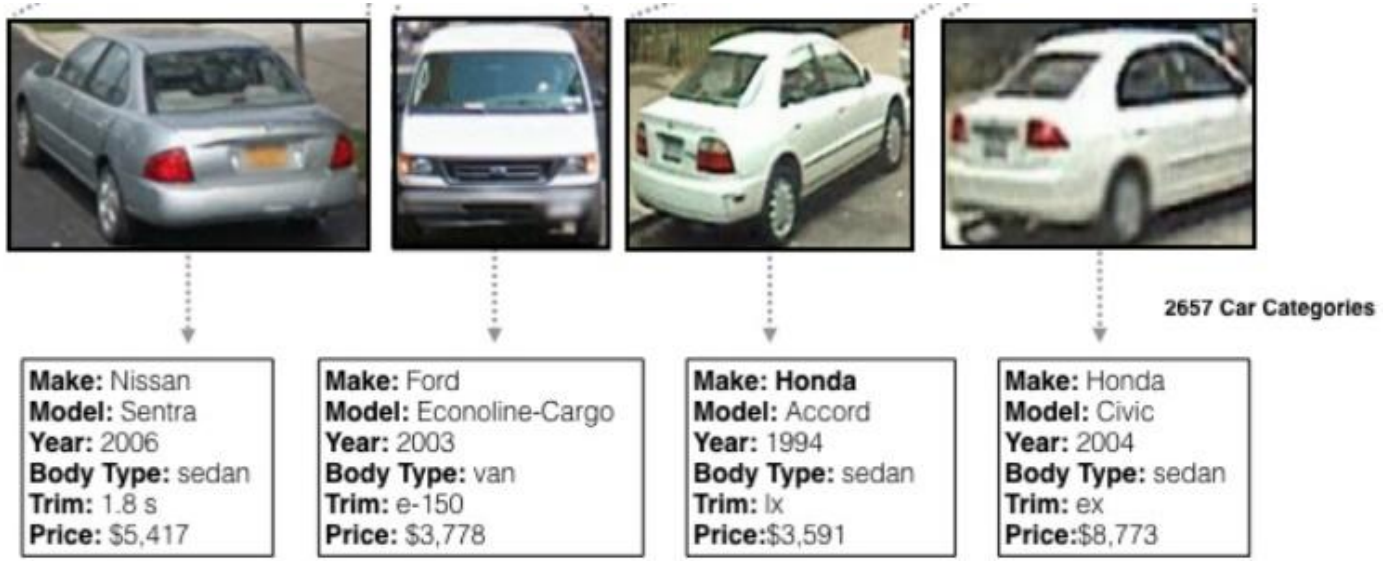
50 million images from Google Street View

US census

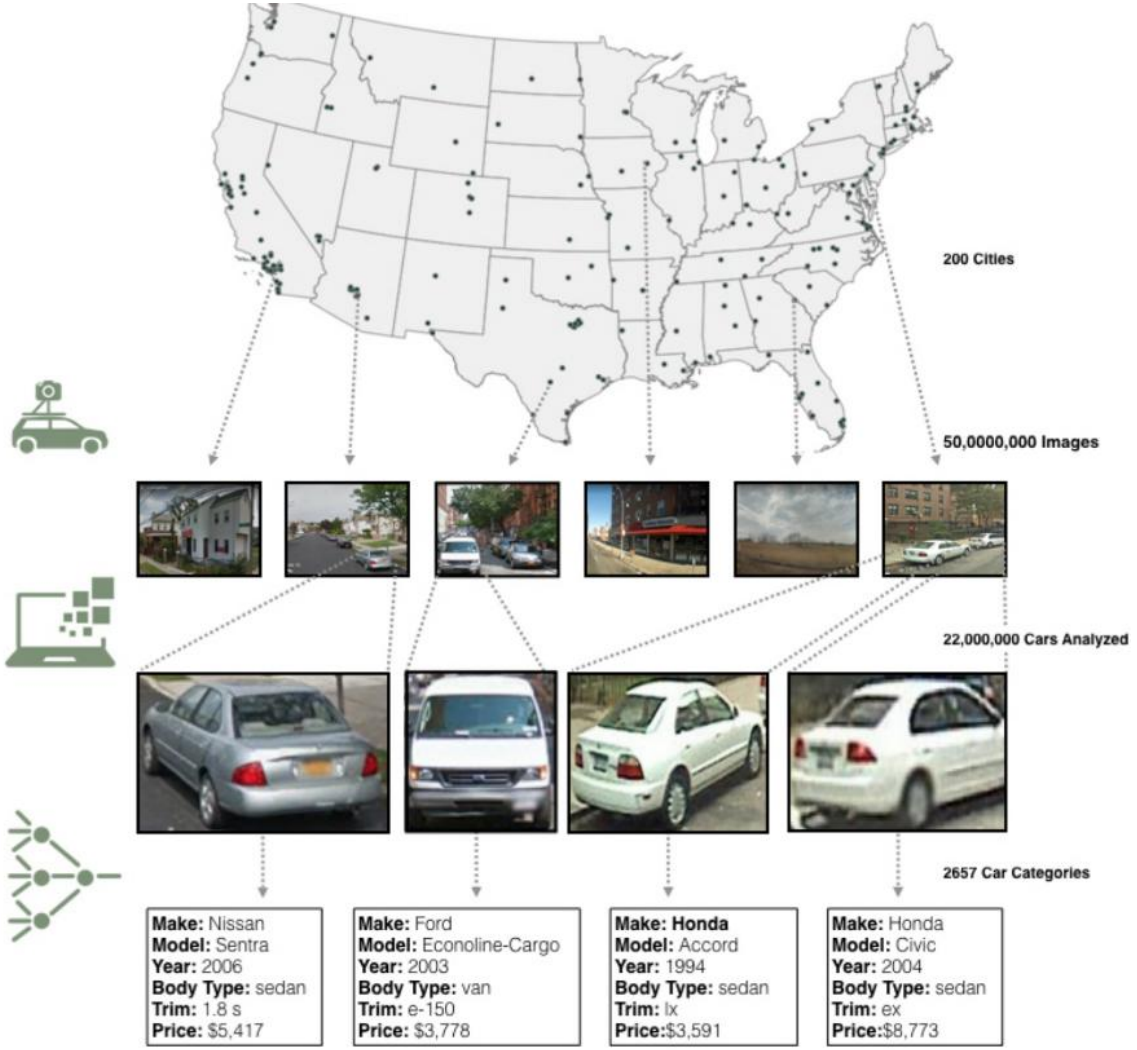
Source: Google Street View

Split car catalog in 2657 Classes

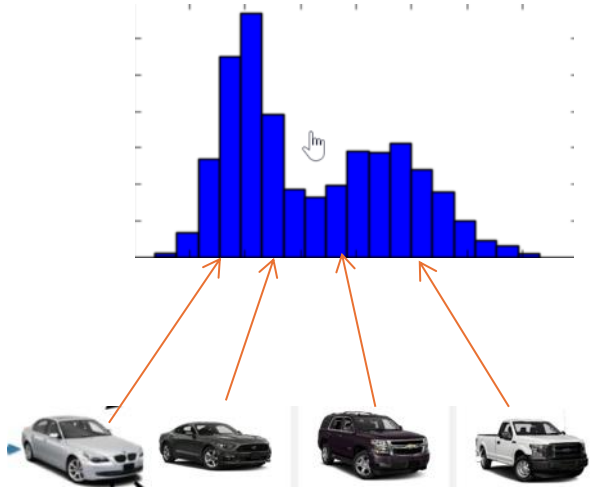




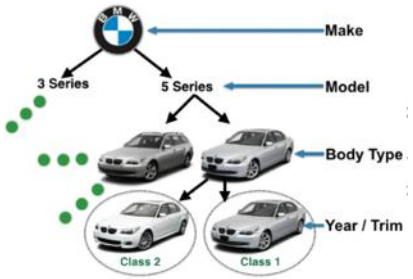
Data representation

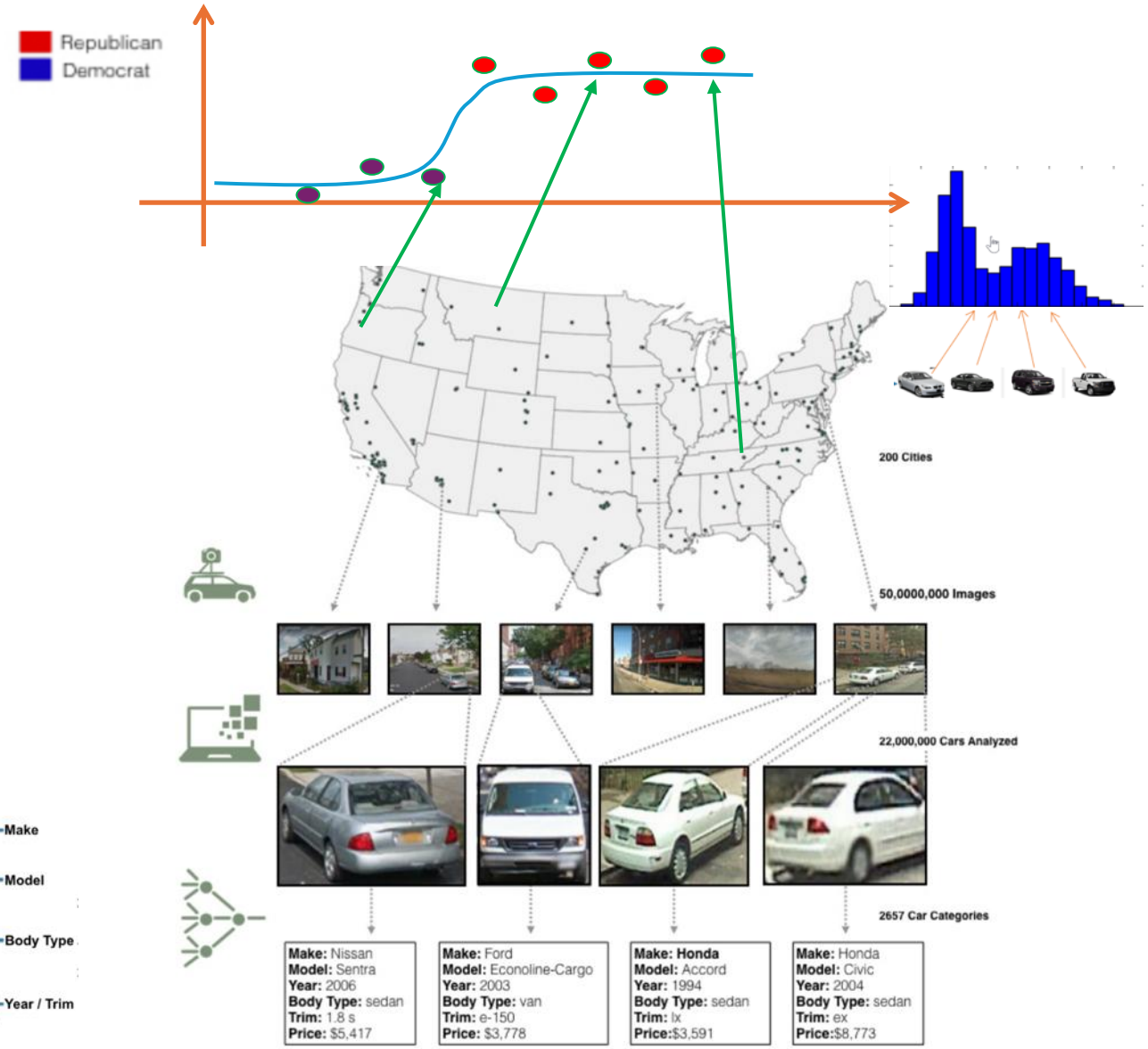


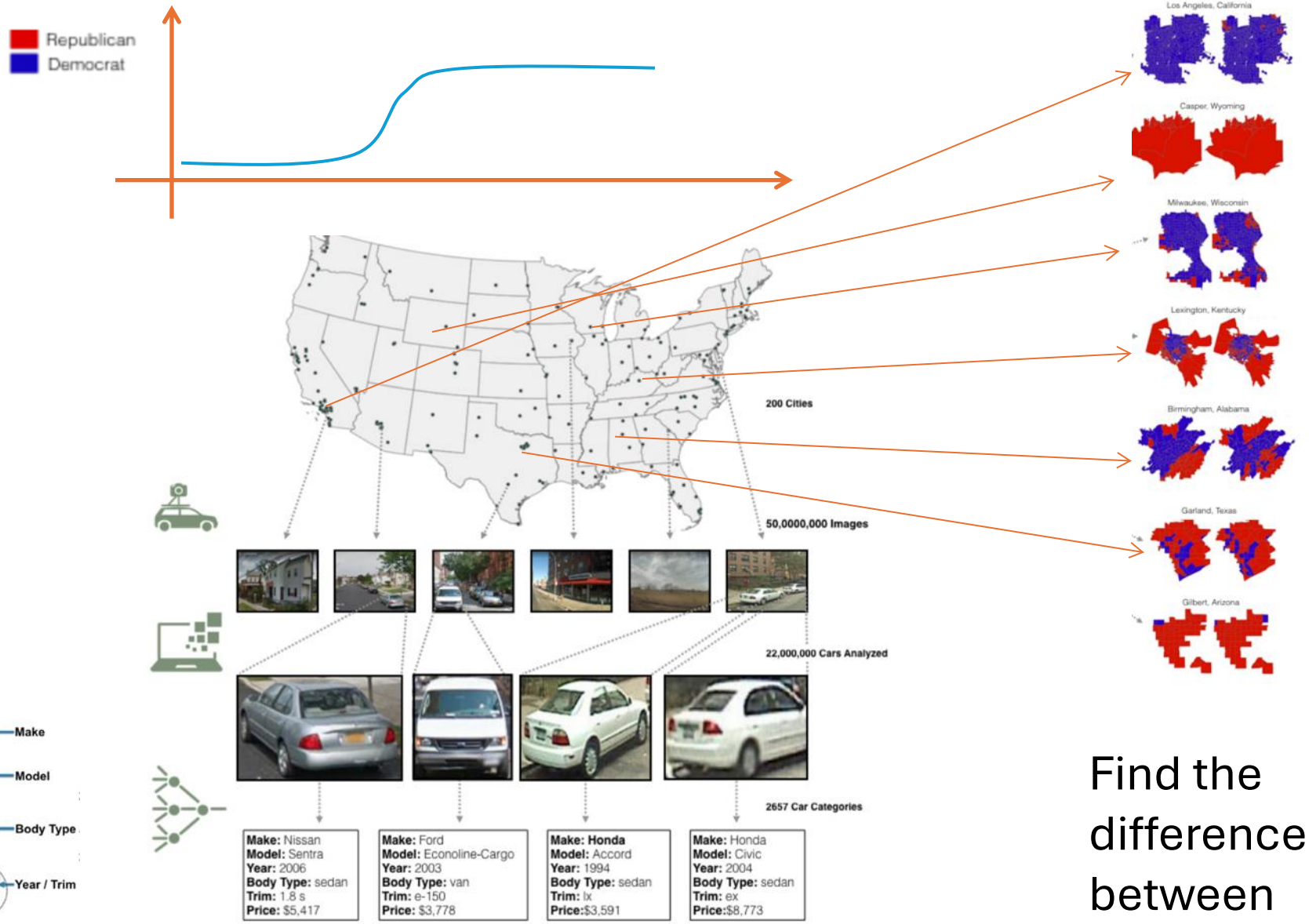
In each neighborhood/county count vehicles per class

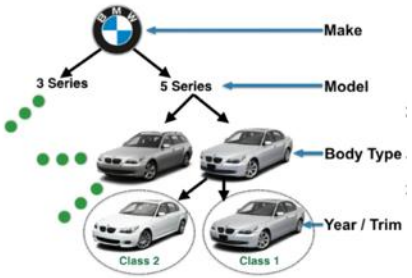


From a subset of cities regress a model for voting decision from car counts



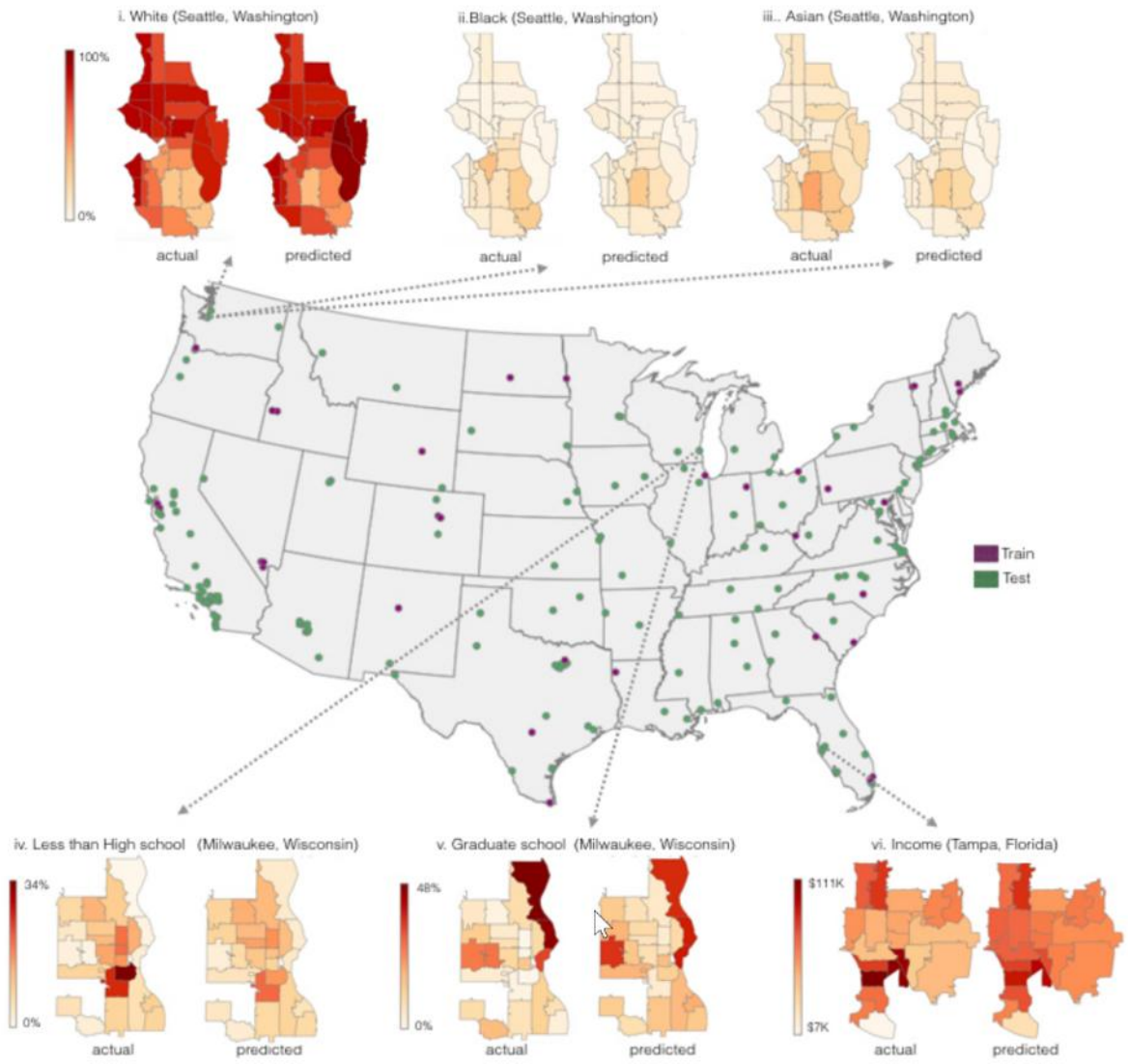






Find the difference between estimated and real.

Are Humans “low rank” ?



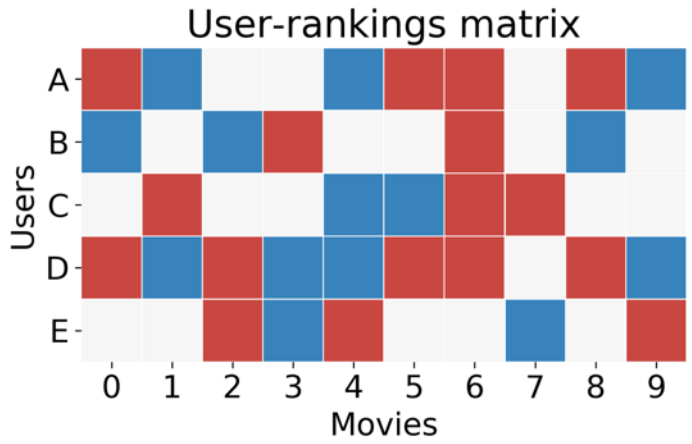
Income, education, ethnicity … all regressed from cars!

Our model detects strong associations between vehicle distribution and disparate socioeconomic trends. For instance, several studies have shown that people of Asian descent are more likely to drive Asian cars [13], a result we observe here as well: the two brands that most strongly indicate an Asian neighborhood are Hondas and Toyotas. Cars manufactured by Chrysler, Buick and Oldsmobile are positively associated with African American neighborhoods, which is again consistent with existing research [14]. And vehicles like pickup trucks, Volkswagens and Aston Martins are indicative of mostly Caucasian neighborhoods. See Fig. S2.

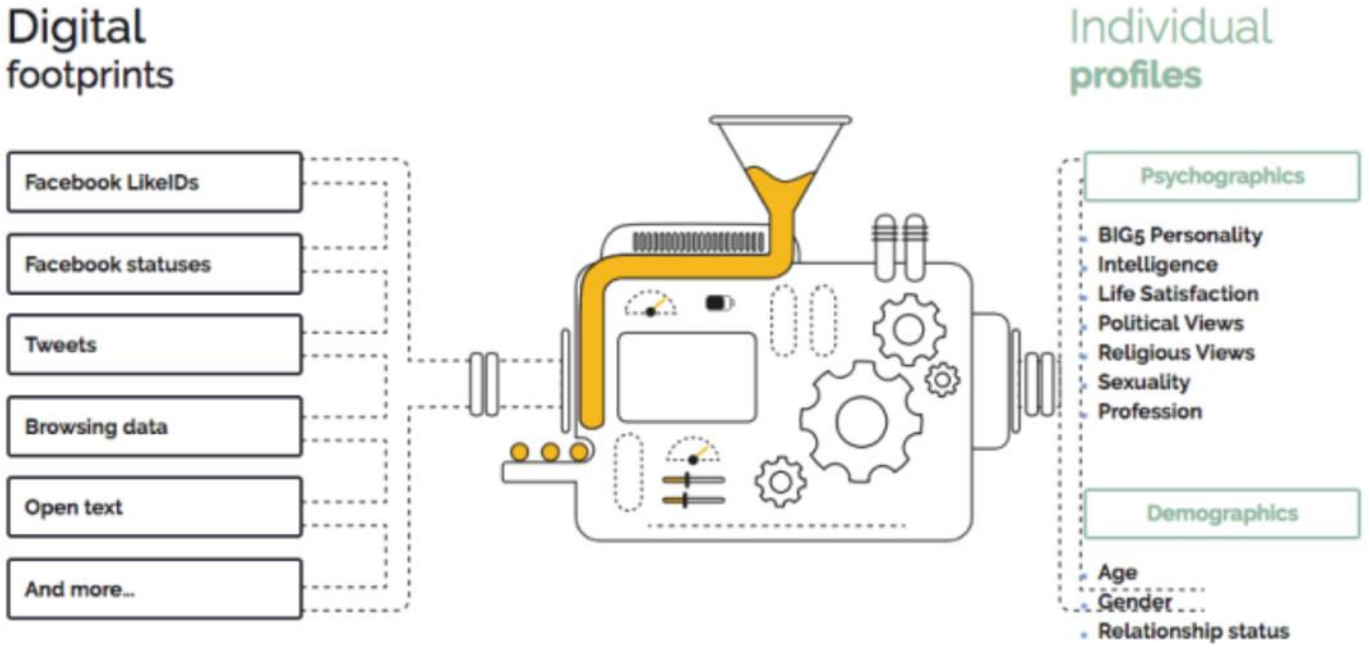
In some cases, the resulting associations can be easily applied in practice. For example, the vehicular feature that was most strongly associated with Democratic precincts was sedans, whereas Republican precincts were most strongly associated with extended-cab pickup trucks (a truck with rear-seat access). We found that by driving through a city for 15 minutes while counting sedans and pickup trucks, it is possible to reliably determine whether the city voted Democratic or Republican: if there are more sedans, it probably voted Democrat (88% chance) and if there are more pickup trucks, it probably voted Republican (82% chance). See Fig. 3(a)iii.

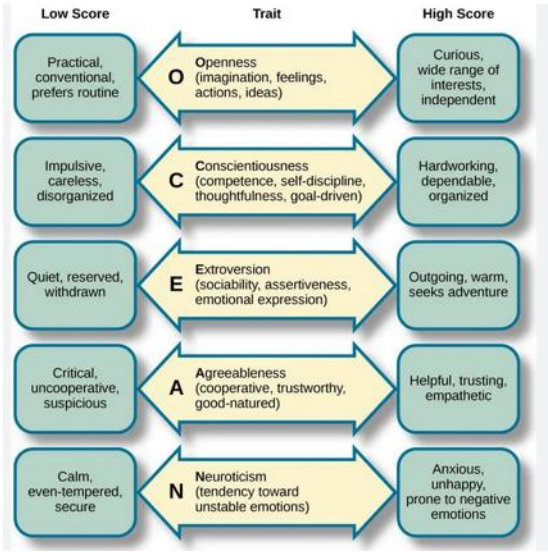
Humans are indeed “low rank” … and it is good ☺

Netflix
https://www.netflix.com

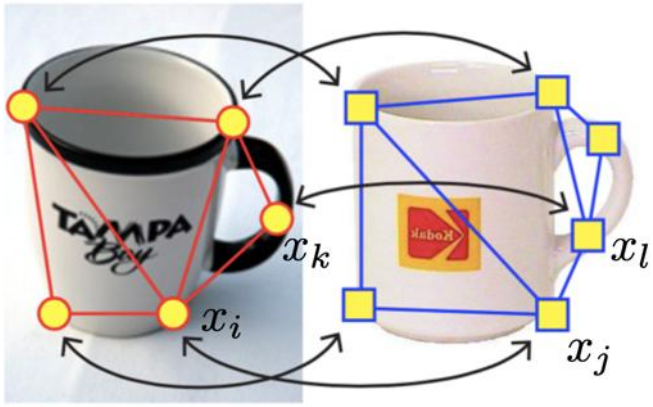


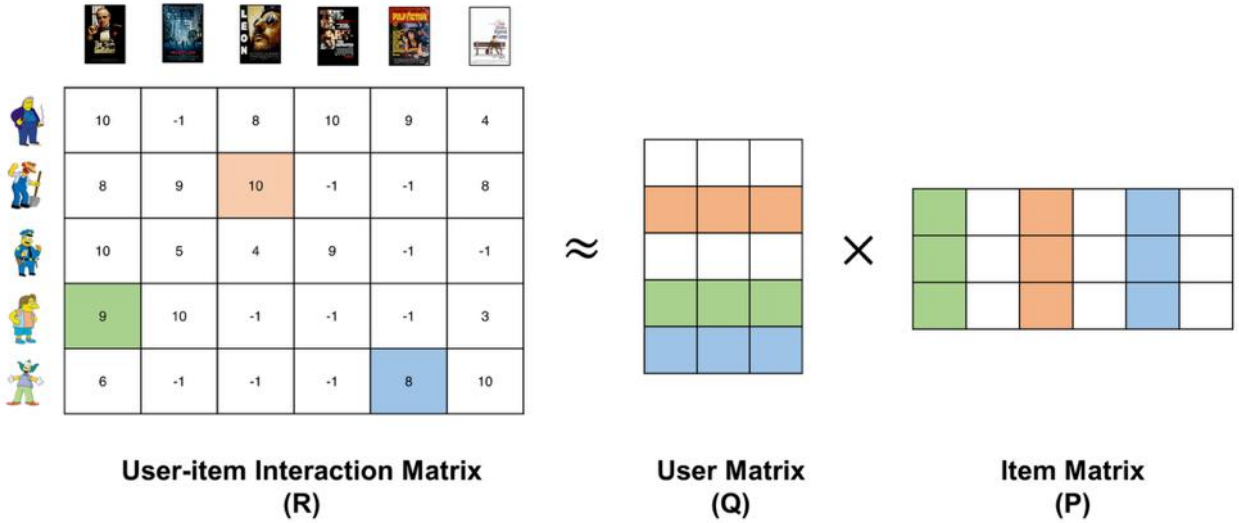
Remember. CAMBRIDGE ANALYTICA





Low Rank Models for Recognition (2010-2016)





For GenZ's

For Boomer's

$$\min_{\mathbf{Z}} \quad f(\mathbf{X}-\mathbf{Z}) \qquad \text{subject to} \quad \operatorname{rank}(\mathbf{Z}) = k,$$

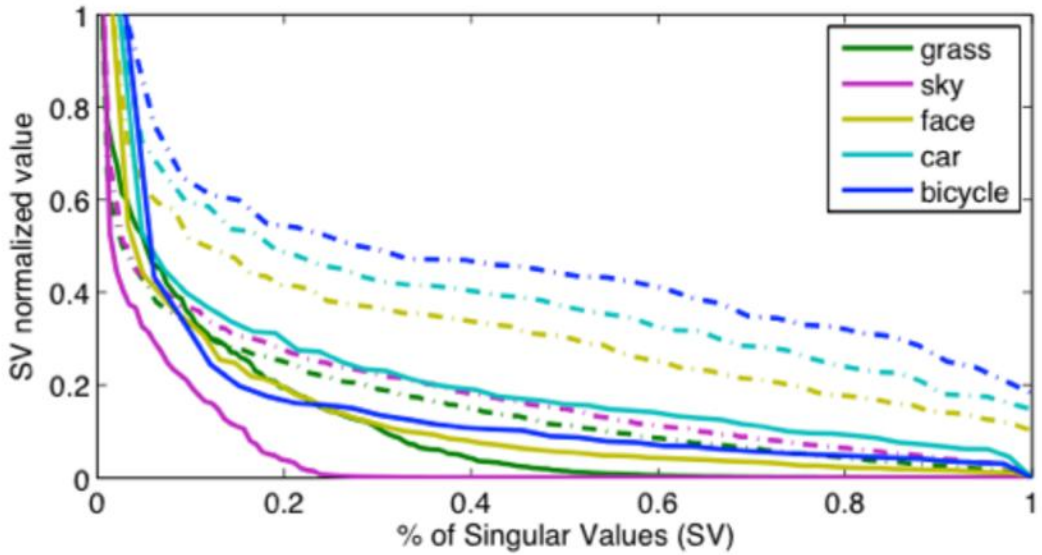
HARD-RANK CONSTRAINTS

$$\min_{\mathbf{Z}} \quad f(\mathbf{X}-\mathbf{Z}) + \lambda r(\mathbf{Z})$$

SOFT-RANK CONSTRAINTS

[CABRAL ET AL. NIPS 2011, PAMI 2014]

CLASSIFICATION AS MC

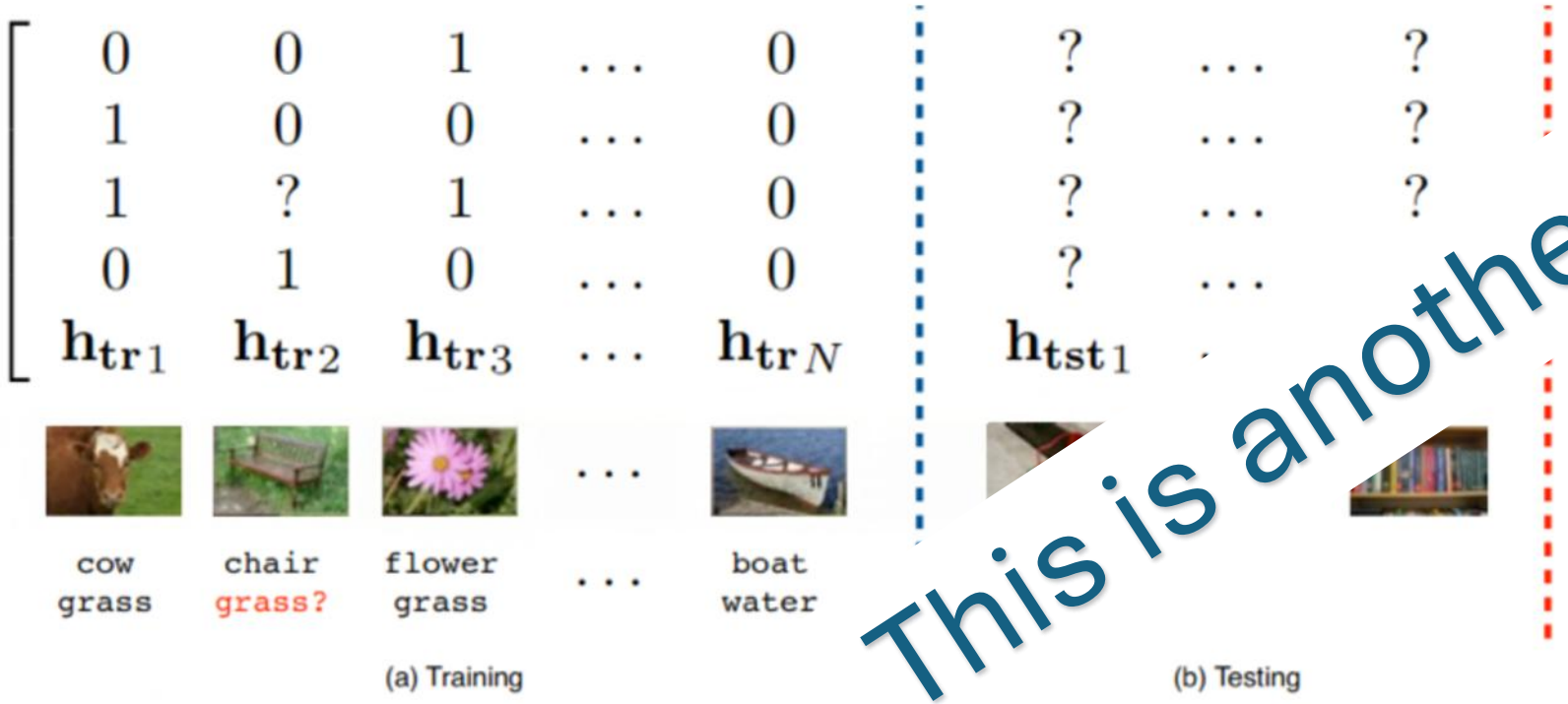


SAME CLASS HISTOGRAMS
YIELD LOWER RANK MATRICES

[CABRAL ET AL. NIPS 2011, PAMI 2014]

CLASSIFICATION AS MC

We can estimate the representative histogram for each class



This is another talk !

The best of its time

Ricardo Silveira Cabral · 1st
AI Agents @ NVIDIA
Zurich, Zurich, Switzerland · Contact info

Experience

Distinguished Research Scientist
NVIDIA · Full-time
Aug 2025 - Present · 2 mos
Zurich, Switzerland

Applied research into AI Agents.

Meta
5 yrs 7 mos

Director of Engineering, GenAI
Full-time
Mar 2024 - Aug 2025 · 1 yr 6 mos
Zurich, Switzerland

TÉCNICO LISBOA

Carnegie Mellon University

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

CARNEGIE MELLON UNIVERSITY

Unifying Low-rank Models for Visual Learning

Ricardo da Silveira Cabral

Orientadores
Professor João Paulo Costeira
Professor Alexandre Bernardino
Professor Fernando De la Torre

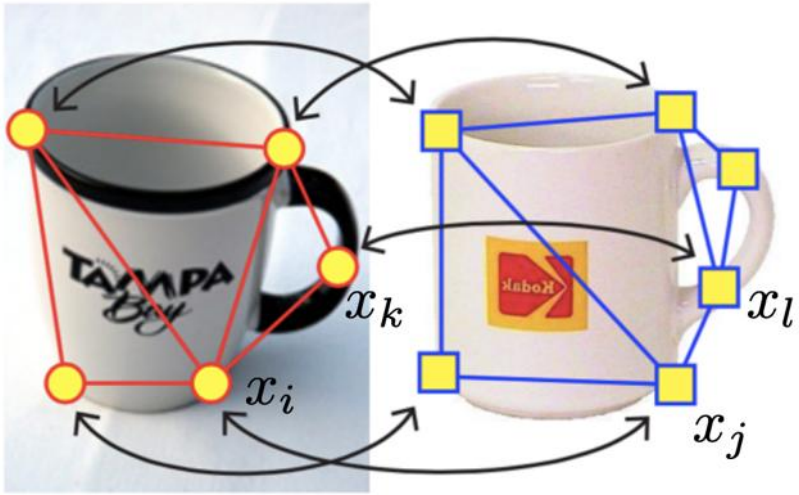
Tese aprovada em provas públicas para obtenção do Grau de Doutor em
Engenharia Electrotécnica e de Computadores

Qualificação atribuída pelo Júri: Aprovado com Distinção e Louvor

Júri
Presidente: Presidente do Conselho Científico do IST
Vogais: **Professor** Mário Figueiredo
Doutor Andrew Fitzgibbon
Professor Aswin Sankaranarayanan

Good jobs are waiting for those who … “think”
Well, at least for those who represent ☺

Key Issues

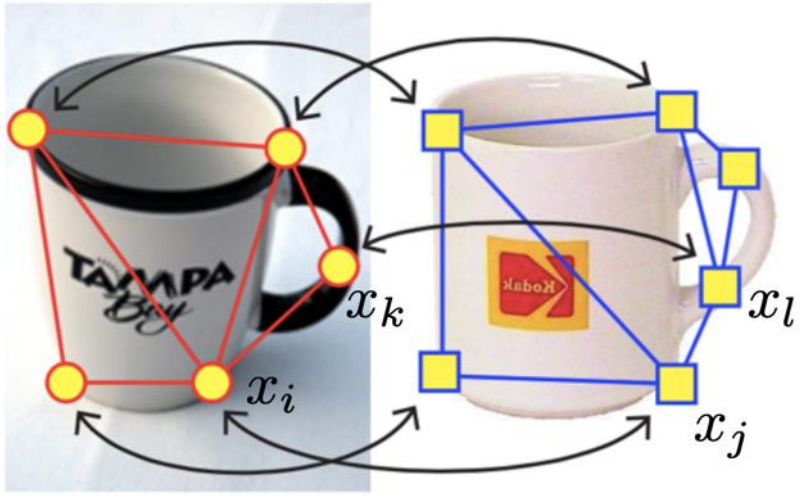


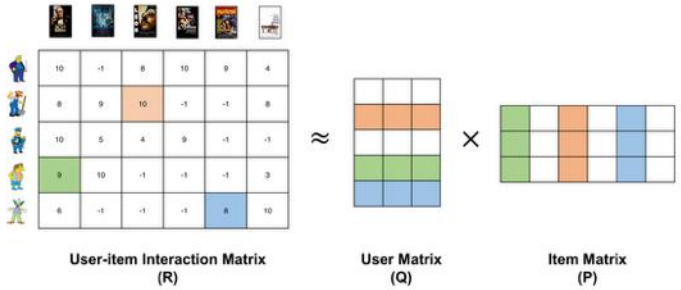
- How is the space organized ?
- What is the “hidden” representation ?

  where is the car, the sky, what can I say about them ?

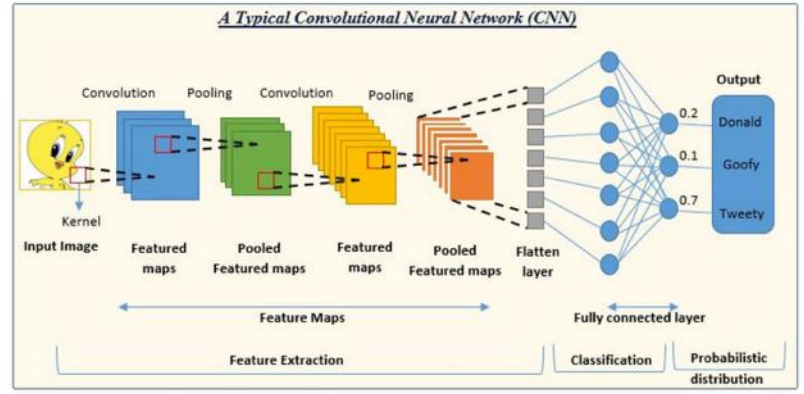
- Representations don’t hide !

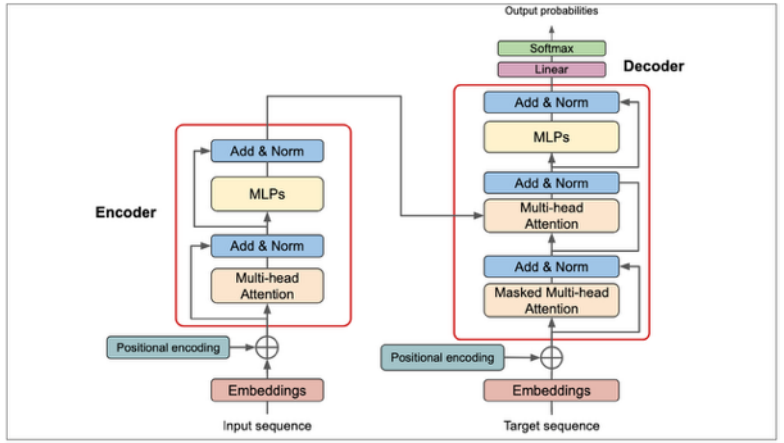
In a nutshell : NN's map input space to "embedding" space (vector space).



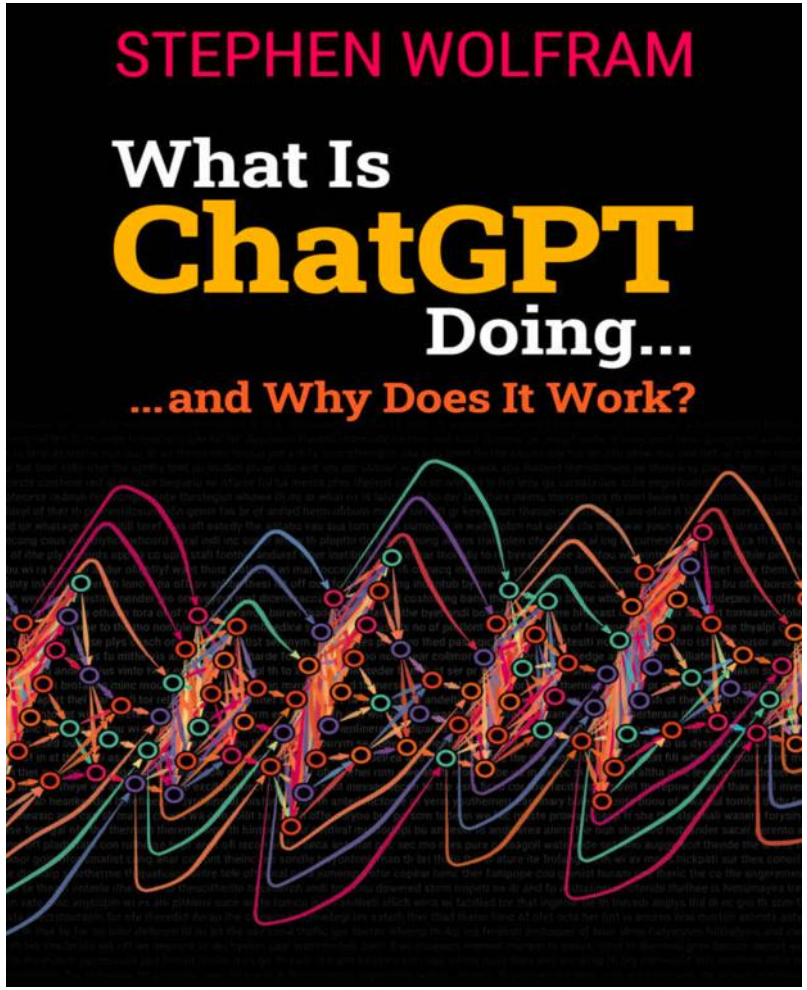


Explicit "generative"(model)





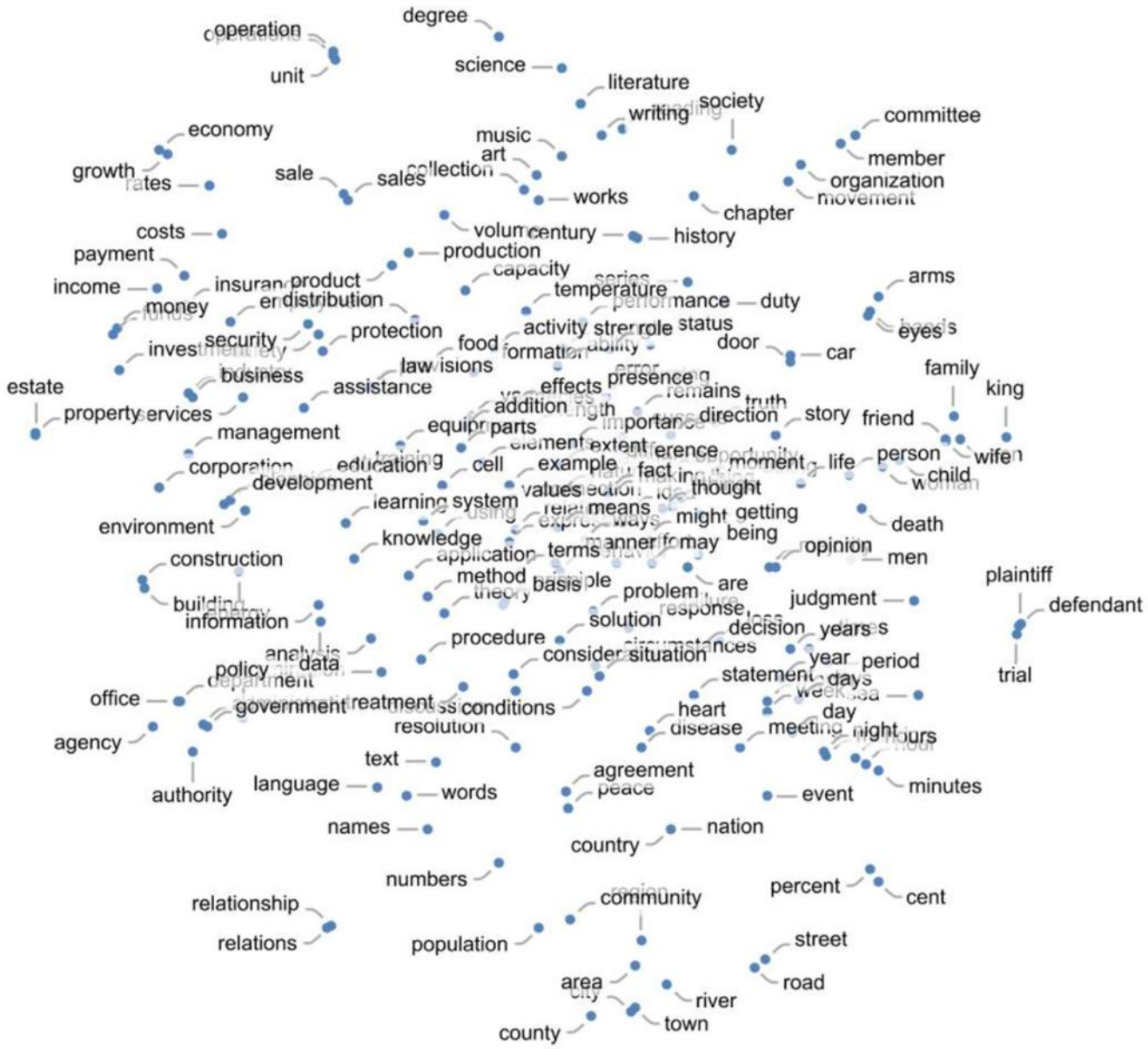
Our friend GPT explained by Wolfram



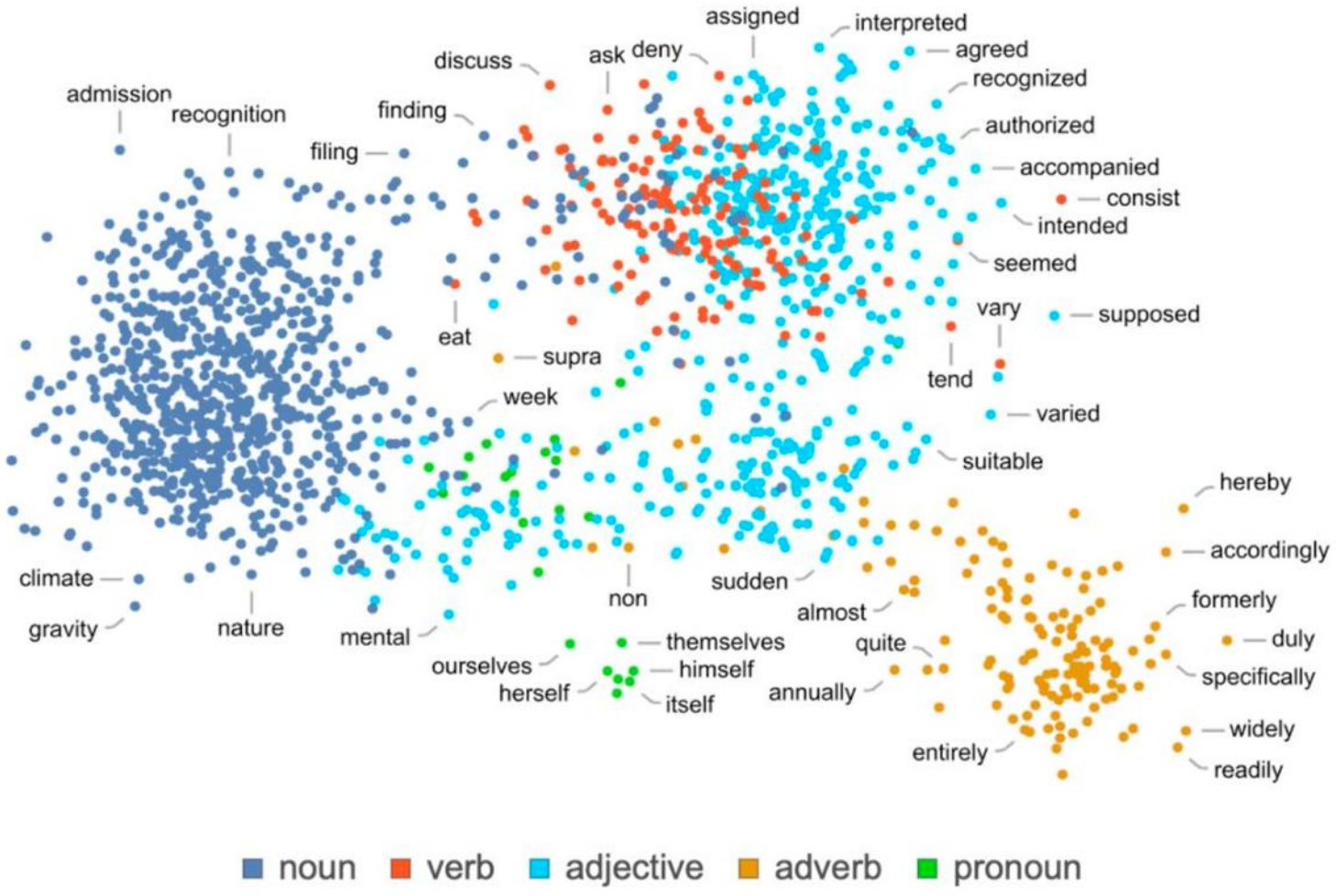
It’s Just Adding One Word at a Time

Meaning Space and Semantic Laws of Motion

We discussed above that **inside ChatGPT any piece of text is effectively represented by an array of numbers that we can think of as coordinates of a point in some kind of "linguistic feature space".** So when ChatGPT continues a piece of text this corresponds to tracing out a trajectory in linguistic feature space. But now we can ask what makes this trajectory correspond to text we consider meaningful. And might there perhaps be some kind of "semantic laws of motion" that define—or at least constrain—how points in linguistic feature space can move around while preserving "meaningfulness"?

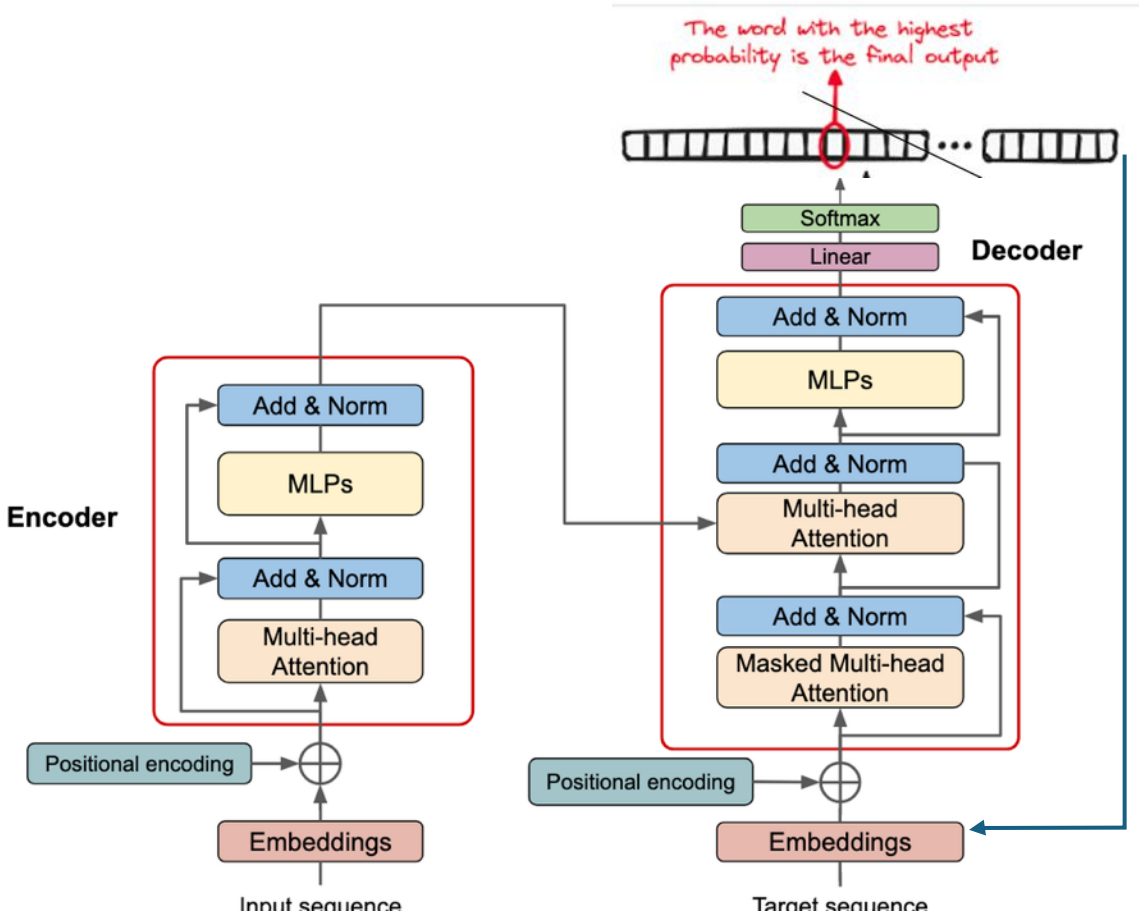


How words corresponding to different parts of speech get laid out:



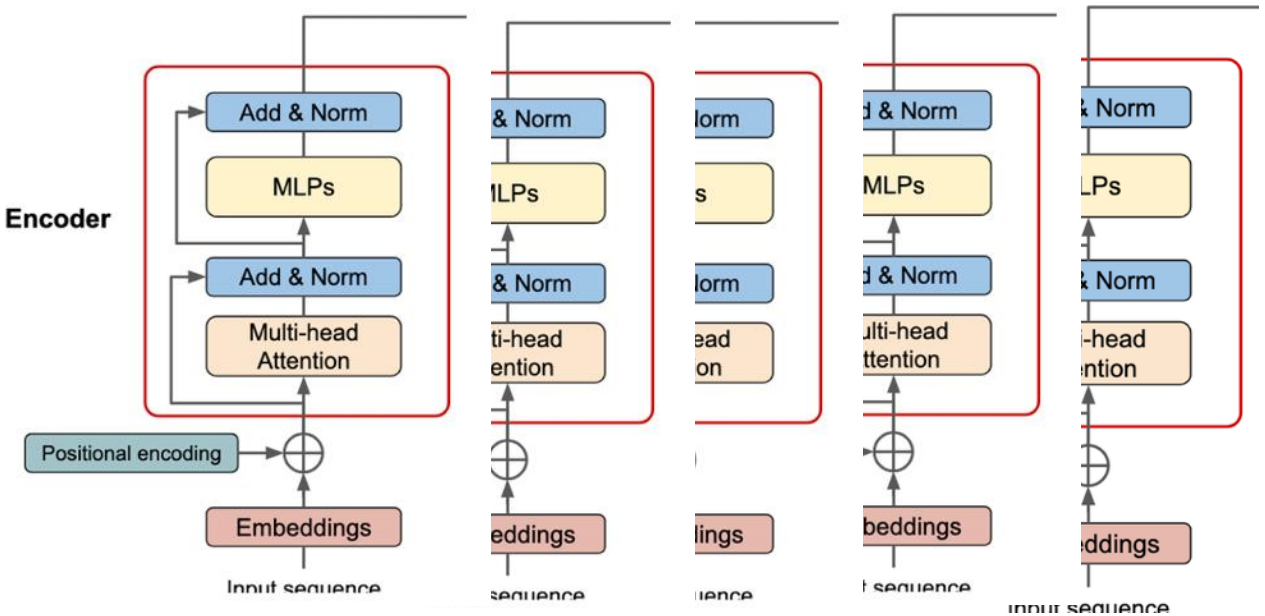
Sure…Euclidean embeddings are the go-to approach to represent **similarities**:

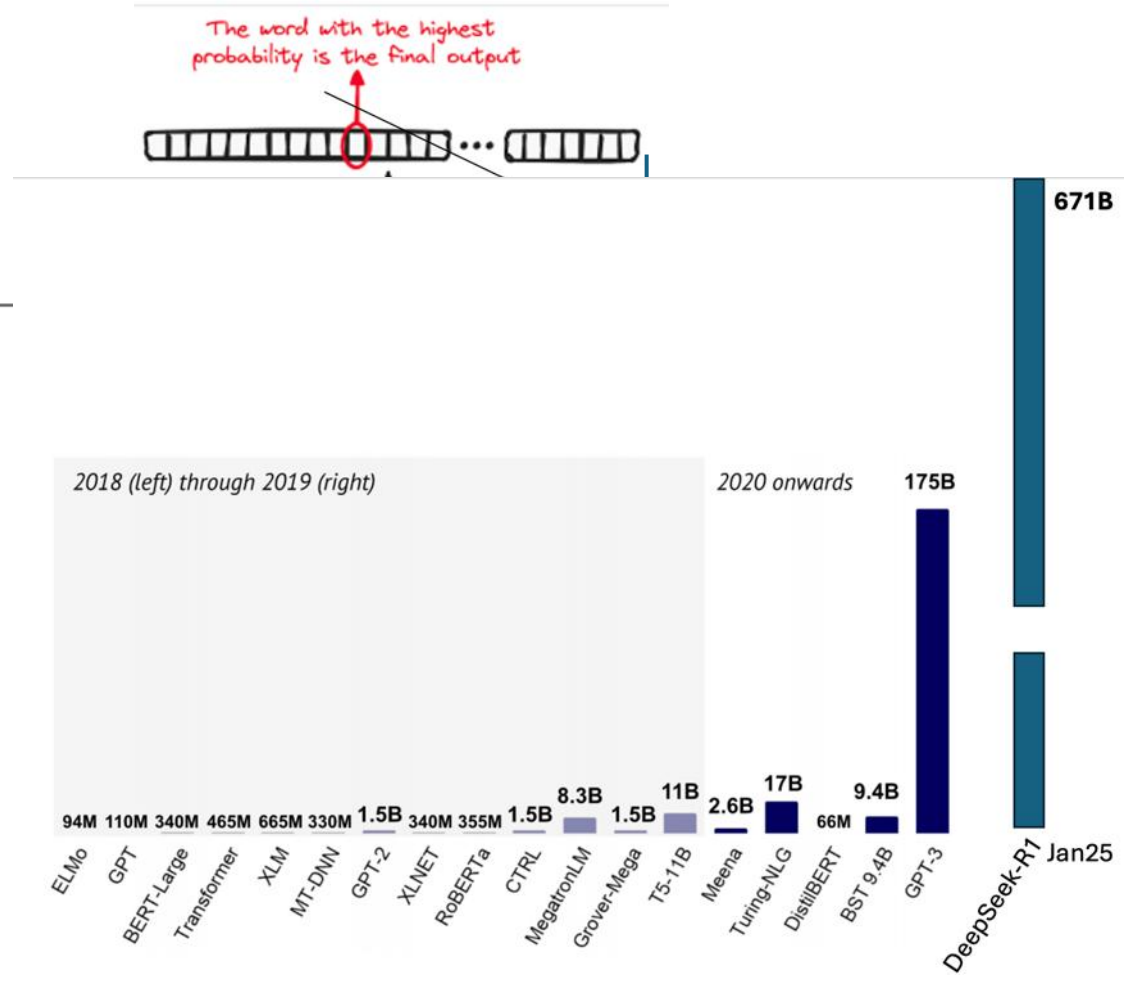
Language is … less “cognitive” than we thought



“The best thing about AI is its ability to”

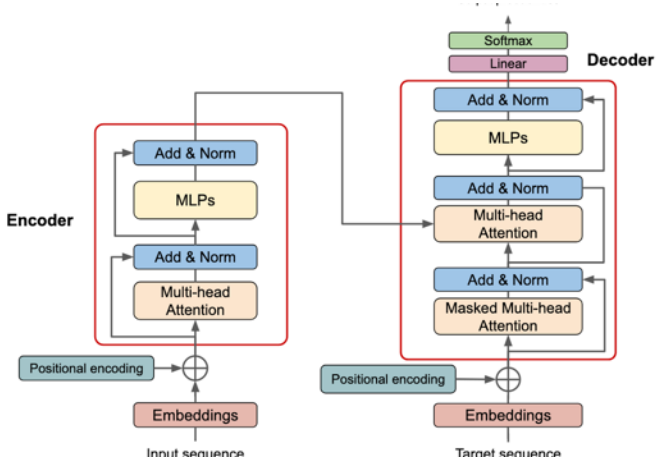
Language is … less “cognitive” than we thought



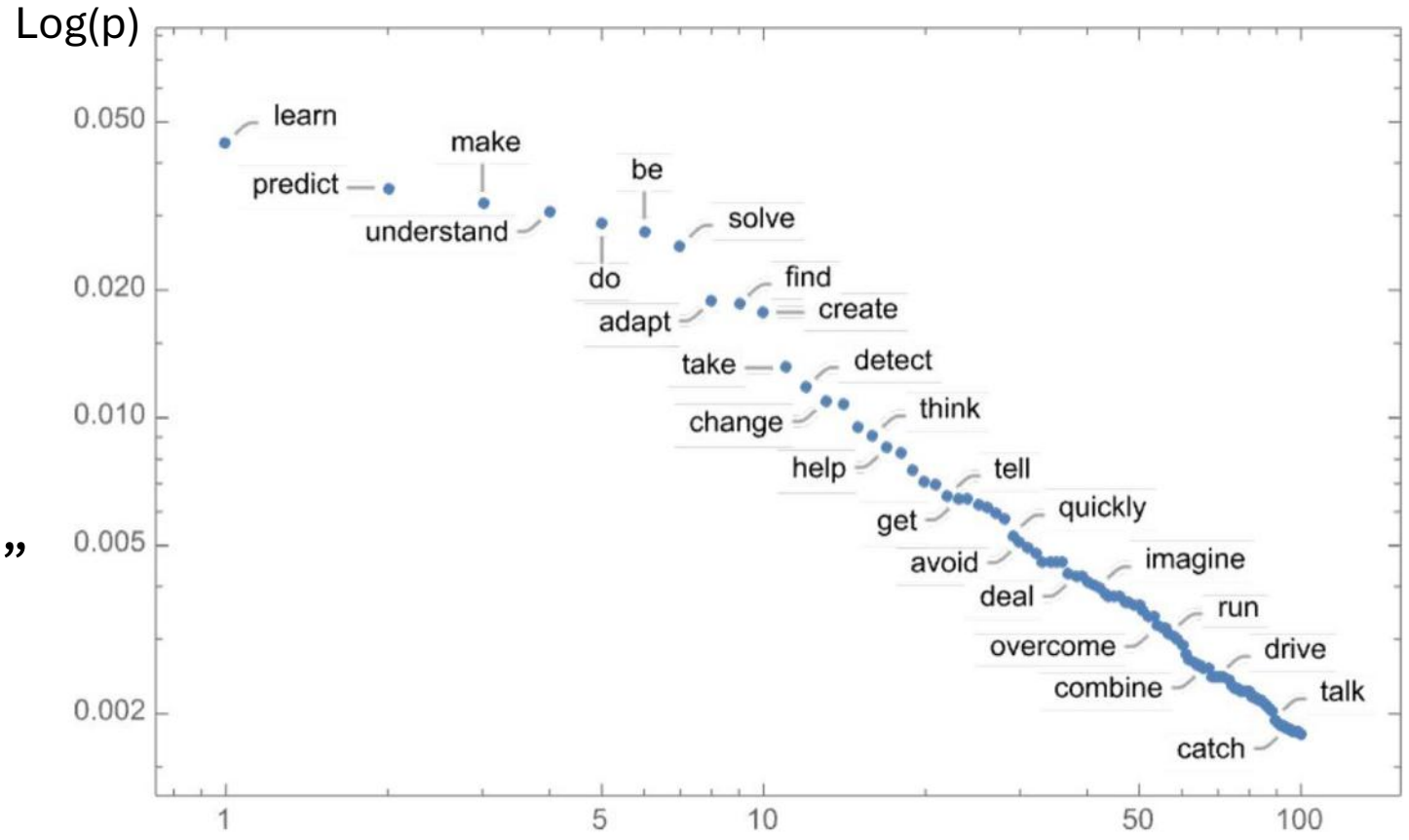


“The best thing about AI is its ability to”

Language is … less “cognitive” than we thought



“The best thing about AI is its ability to”



Can it “think” (relate, infer, hierarchize…) concepts…like we do?

The “hottest” thing of the moment … ☺

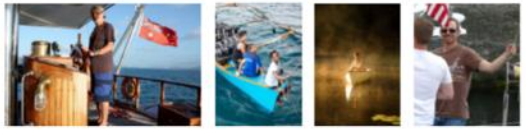
How can we represent **hierarchies** and **logical composition** with Euclidean embeddings?

a man

a man ∧ on a boat

a man ∧ ¬on a boat

a man ∧ on a boat ∧ ¬is fishing



a man ∧ on a boat ∧ is fishing

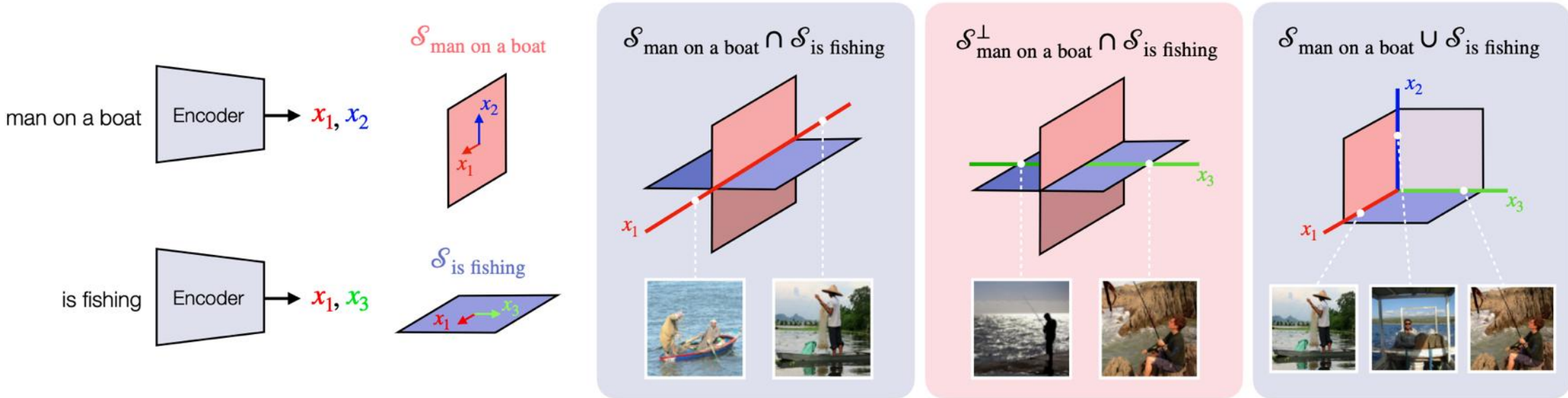
a man ∧ ¬on a boat ∧ is fishing

a man ∧ ¬on a boat ∧ ¬is fishing

Subspace Embeddings

We replace single vectors by **subspaces**, which allow for **intersections (conjunction), sums (disjunction)** and **complements (negation).**

Subspace dimension is a natural measure of **generality.**



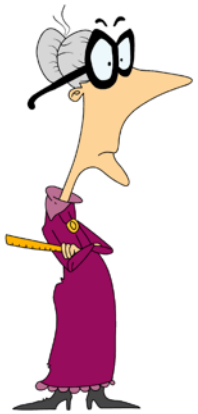
Give me an X-Ray of W that does not have Y

What about the two inspiring papers ?

Outline

Inspired by two articles … I'm going to preach !

-That's what "seniors" do!



- How "AI" works (and why it can be dangerous for you !)

- The I (in AI) does not stand for "Intelligence"…

What about the two inspiring papers ?

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

Perspective

Language is primarily a tool for communication rather than thought

https://doi.org/10.1038/s41586-024-07522-w

Received: 15 February 2023

Accepted: 3 May 2024

Published online: 19 June 2024

Check for updates

Evelina Fedorenko[1,2]✉, Steven T. Piantadosi[3] & Edward A. F. Gibson[1]

Language is a defining characteristic of our species, but the function, or functions, that it serves has been debated for centuries. Here we bring recent evidence from neuroscience and allied disciplines to argue that in modern humans, language is a tool for communication, contrary to a prominent view that we use language for thinking. We begin by introducing the brain network that supports linguistic ability in humans. We then review evidence for a double dissociation between language and thought, and discuss several properties of language that suggest that it is optimized for communication. We conclude that although the emergence of language has unquestionably transformed human culture, language does not appear to be a prerequisite for complex thought, including symbolic thought. Instead, language is a powerful tool for the transmission of cultural knowledge; it plausibly co-evolved with our thinking and reasoning capacities, and only reflects, rather than gives rise to, the signature sophistication of human cognition.

Name ?

What about the two inspiring papers ?

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

Perspective

Language is primarily a tool for communication rather than thought

https://doi.org/10.1038/s41586-024-07522-w

Received: 15 February 2023

Accepted: 3 May 2024

Published online: 19 June 2024

Check for updates

Evelina Fedorenko[1,2]✉, Steven T. Piantadosi[3] & Edward A. F. Gibson[1]

Language is a defining characteristic of our species, but the function, or functions, that it serves has been debated for centuries. Here we bring recent evidence from neuroscience and allied disciplines to argue that in modern humans, language is a tool for communication, contrary to a prominent view that we use language for thinking. We begin by introducing the brain network that supports linguistic ability in humans. We then review evidence for a double dissociation between language and thought, and discuss several properties of language that suggest that it is optimized for communication. We conclude that although the emergence of language has unquestionably transformed human culture, language does not appear to be a prerequisite for complex thought, including symbolic thought. Instead, language is a powerful tool for the transmission of cultural knowledge; it plausibly co-evolved with our thinking and reasoning capacities, and only reflects, rather than gives rise to, the signature sophistication of human cognition.

Name ?

What about the two inspiring papers ?

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

Perspective

Language is primarily a tool for communication rather than thought

https://doi.org/10.1038/s41586-024-07522-w

Received: 15 February 2023

Accepted: 3 May 2024

Published online: 19 June 2024

Check for updates

Evelina Fedorenko[1,2]✉, Steven T. Piantadosi[3] & Edward A. F. Gibson[1]

Language is a defining characteristic of our species, but the function, or functions, that it serves has been debated for centuries. Here we bring recent evidence from neuroscience and allied disciplines to argue that in modern humans, language is a tool for communication, contrary to a prominent view that we use language for thinking. We begin by introducing the brain network that supports linguistic ability in humans. We then review evidence for a double dissociation between language and thought, and discuss several properties of language that suggest that it is optimized for communication. We conclude that although the emergence of language has unquestionably transformed human culture, language does not appear to be a prerequisite for complex thought, including symbolic thought. Instead, language is a powerful tool for the transmission of cultural knowledge; it plausibly co-evolved with our thinking and reasoning capacities, and only reflects, rather than gives rise to, the signature sophistication of human cognition.

11 names for colors in English

-colors that everybody knows-

What about the two inspiring papers ?

What about the two inspiring papers ?

The Economist

Weekly edition | The world in brief | War in the Middle East

Finance & economics | As

How AI
the rest

Optimists hope the
to widen social divi

Share

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task△

Nataliya Kosmyna [1]
MIT Media Lab
Cambridge, MA

Eugene Hauptmann
MIT
Cambridge, MA

Ye Tong Yuan
Wellesley College
Wellesley, MA

Jessica Situ
MIT
Cambridge, MA

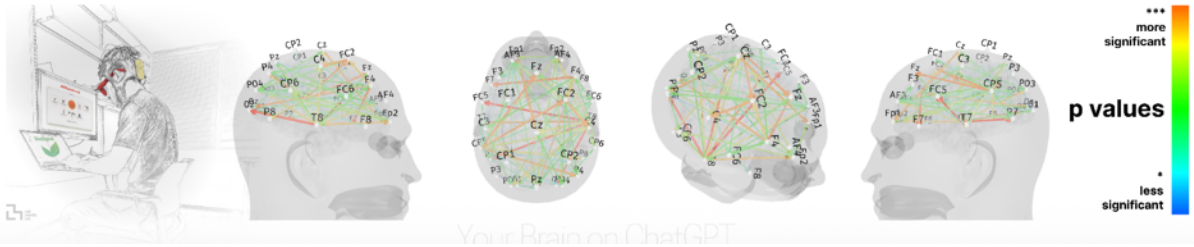
Xian-Hao Liao
Mass. College of Art and Design (MassArt)
Boston, MA

Ashly Vivian Beresnitzky
MIT
Cambridge, MA

Iris Braunstein
MIT
Cambridge, MA

Pattie Maes
MIT Media Lab
Cambridge, MA

United States



Your Brain on ChatGPT

in | Menu

Games

NT

Conclusions

- To reason and retrieve efficiently (to think ☺ we must represent (explicitely)
- Most “cognitive” activity … is not that cognitive…it’s “automatic” (namely writing!)
- LLM’s (for now) have no explicit representation of … anything ! It’s all DATA

So …

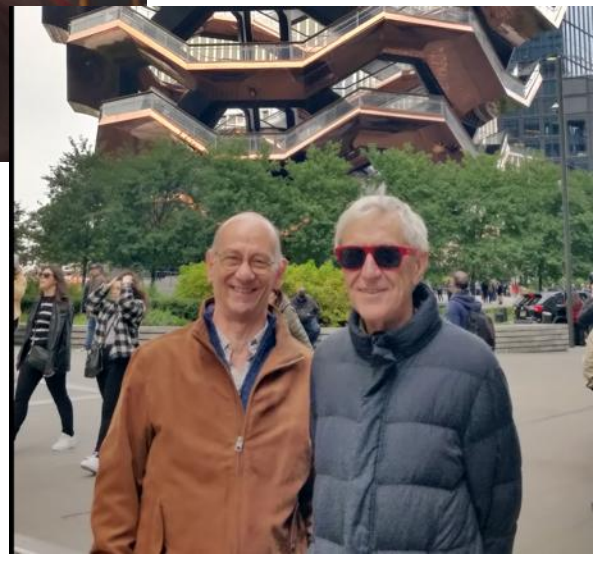
Modeling is crucial…

And some models, like the low rank factorization…

Modeling is crucial…

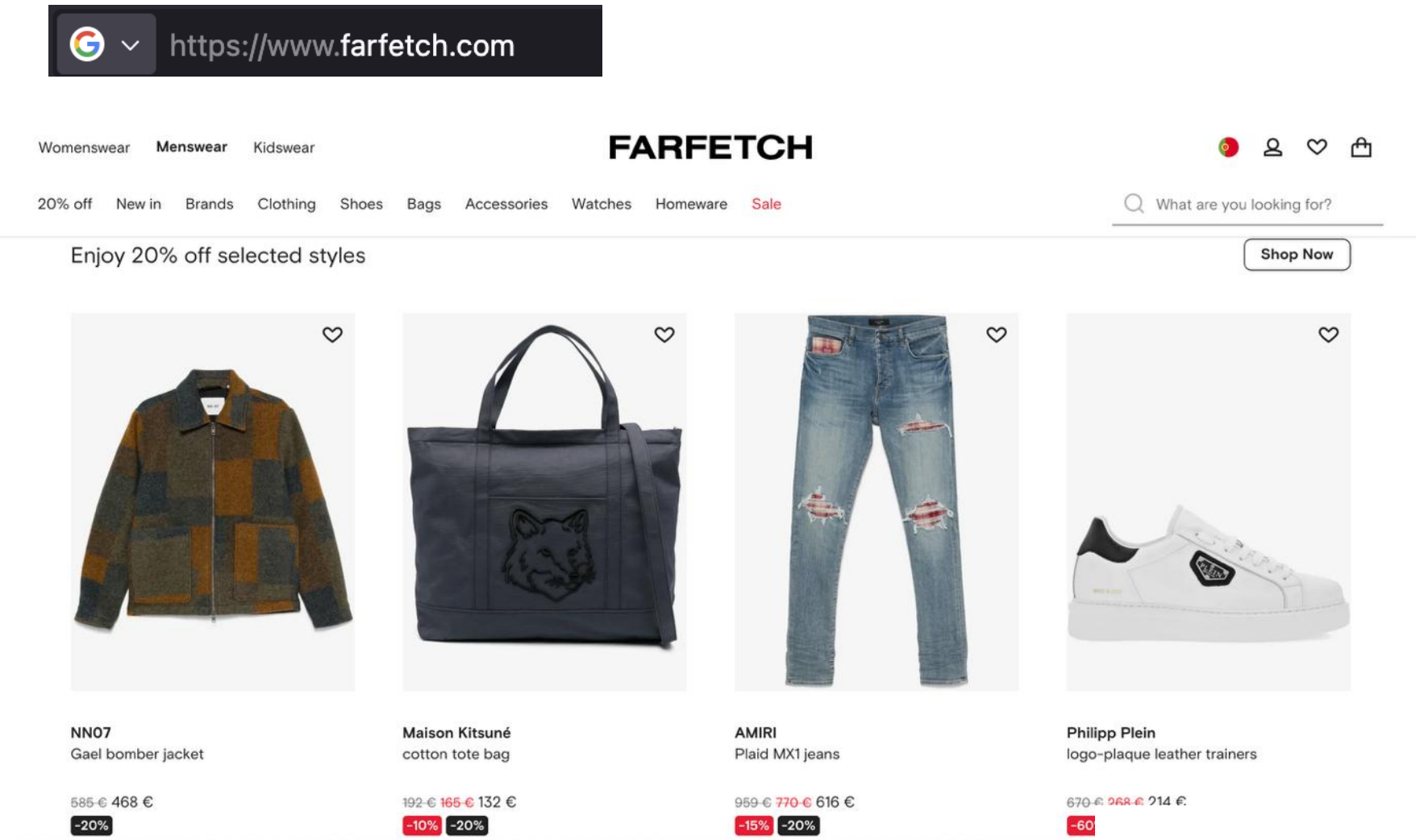
And some models, like the low rank factorization…

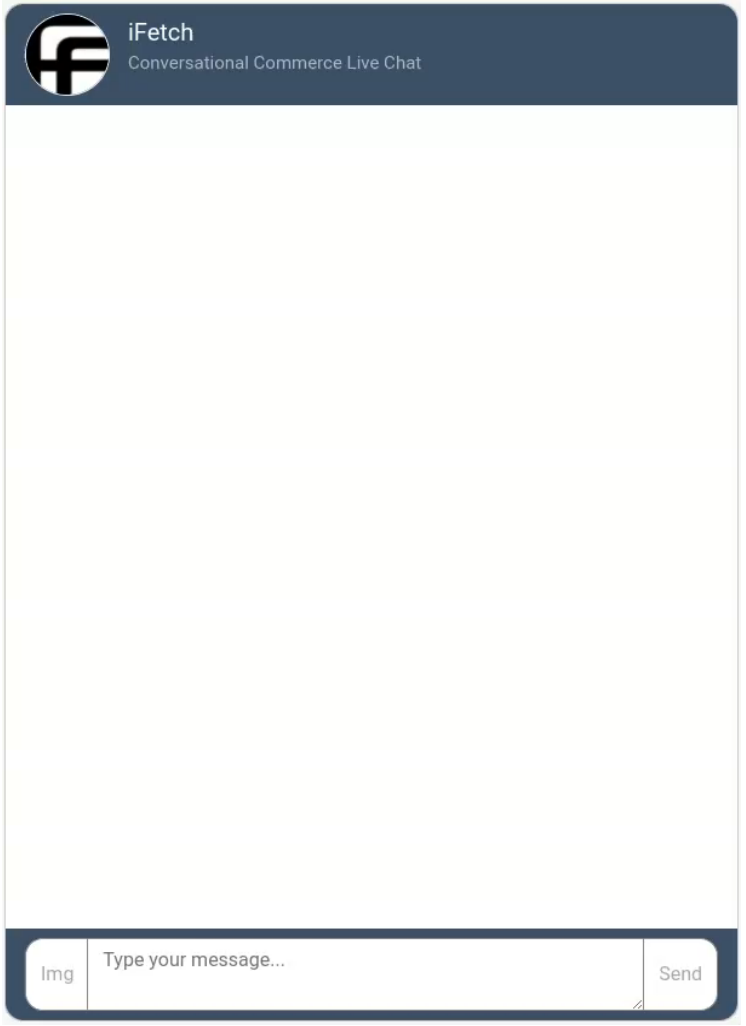
Last forever …




QUESTIONS

What about language ? ...(2022)





https://farfetch-chat-rd.github.io/

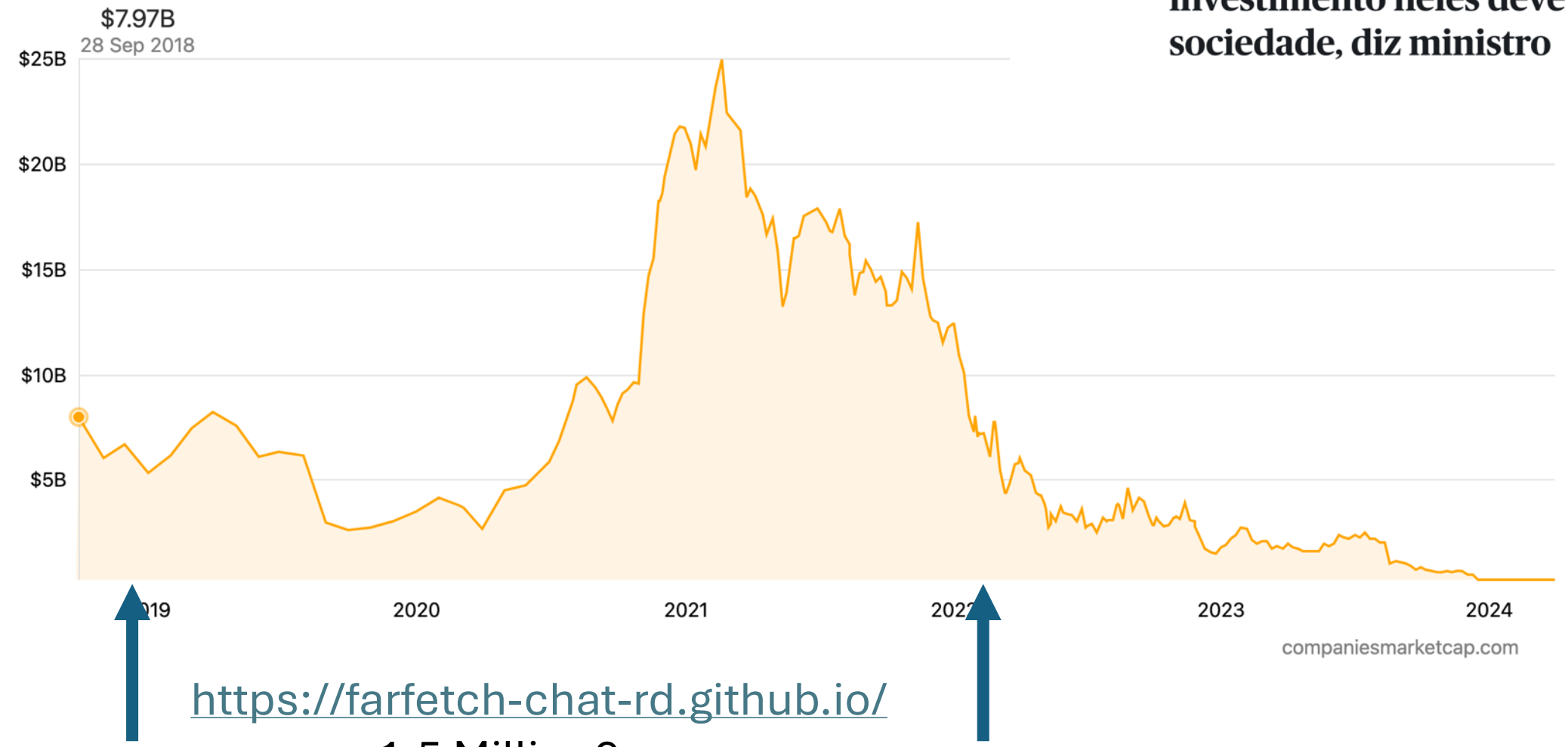
FARFETCH
THE WORLD'S GREATEST SELECTION OF LUXURY

UNIVERSIDADE NOVA DE LISBOA

TÉCNICO LISBOA

Carnegie Mellon University

Imbecility is myopic !



P POLÍTICA PSD PS CHEGA INICIATIVA LIBERAL BE PCP LIVRE CDS PAN MAIS

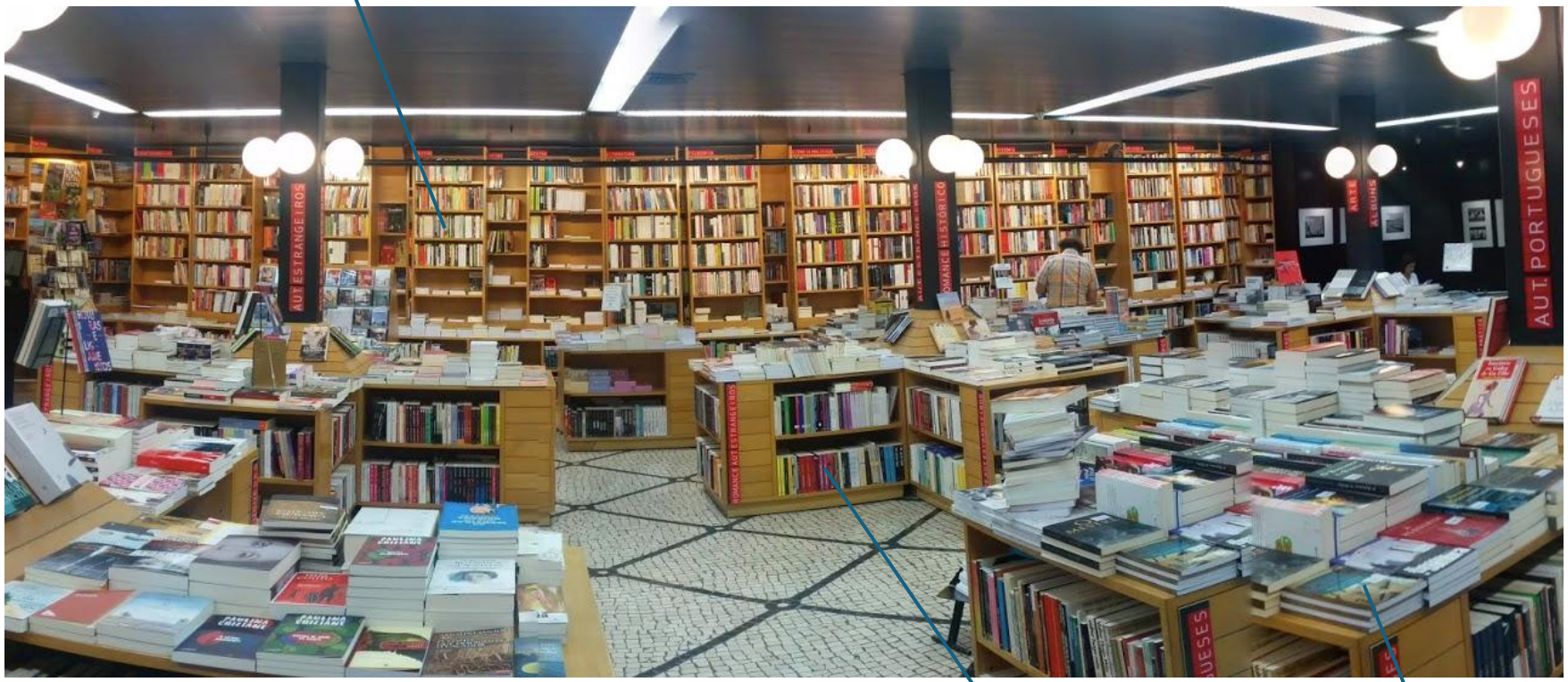
UNIVERSIDADE DE VERÃO

Cientistas "têm de perceber" que investimento neles deve ser devolvido à sociedade, diz ministro

https://farfetch-chat-rd.github.io/

1.5 Million€

The experimental site



Main floor of the bookshop

- 29 tables with approx. 30 books each
- Tables display "recent books/editions"
- 100 thousand+ titles in shop
- 1000 customers/day
- Open 7 days/week (9AM-11PM)

Other sections

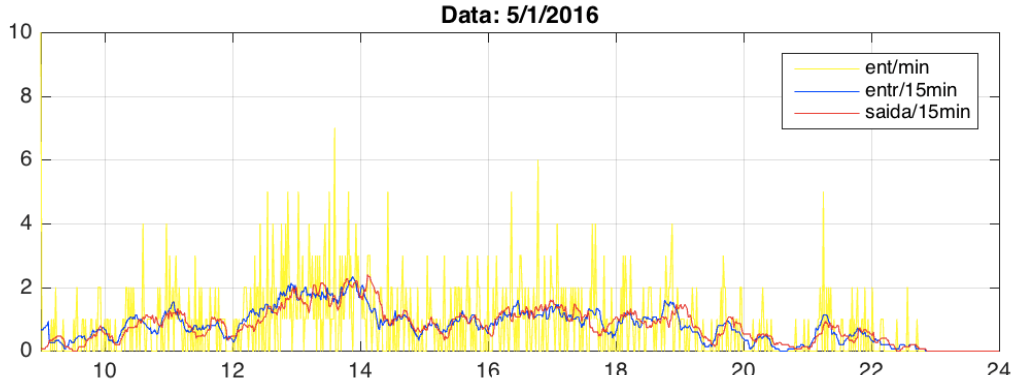
Pocket editions (foreign)

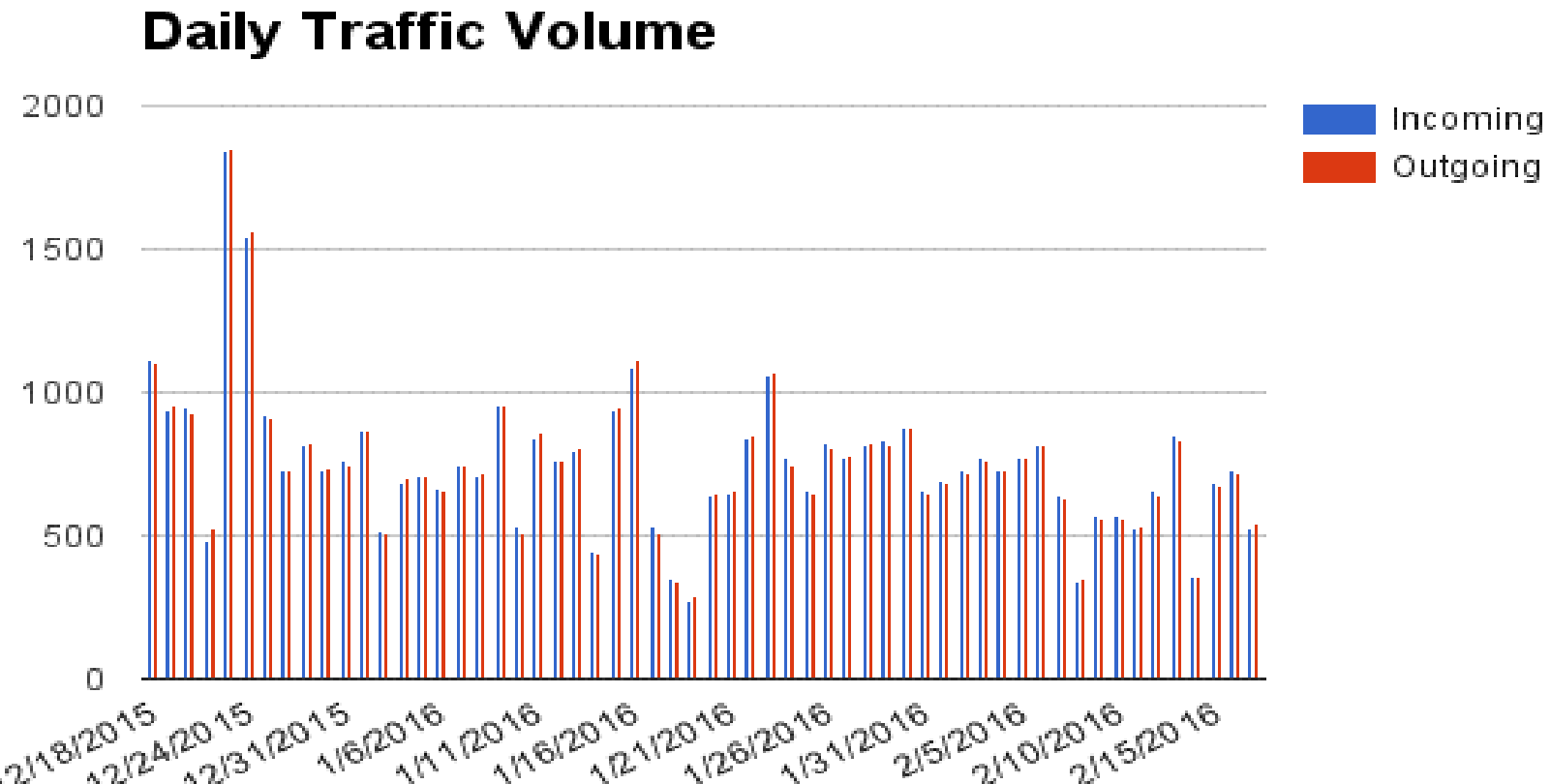
Magazines/newspapers

Depth Camera 1

Counting incoming/outgoing costumers

Typical shopping day ...BC (Before Corona!)

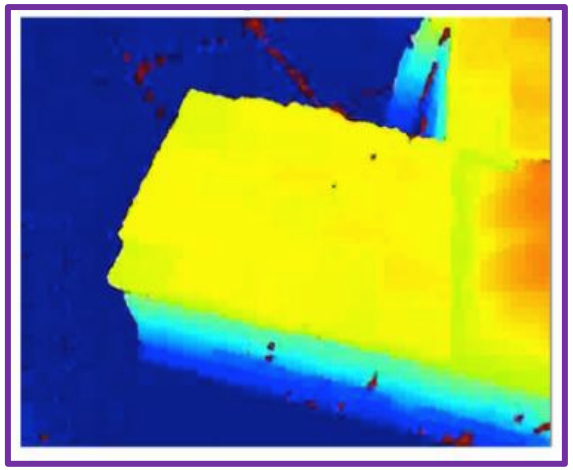




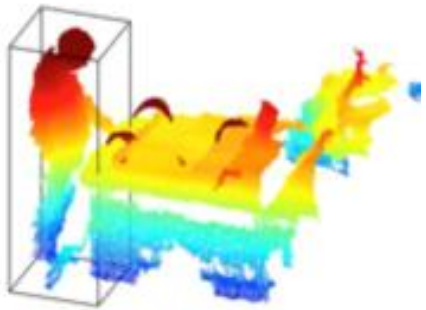
Camera 2: New releases !

RGB image

Depth image

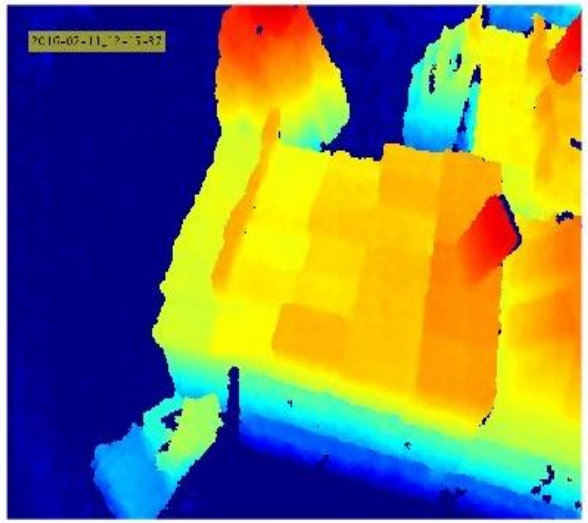


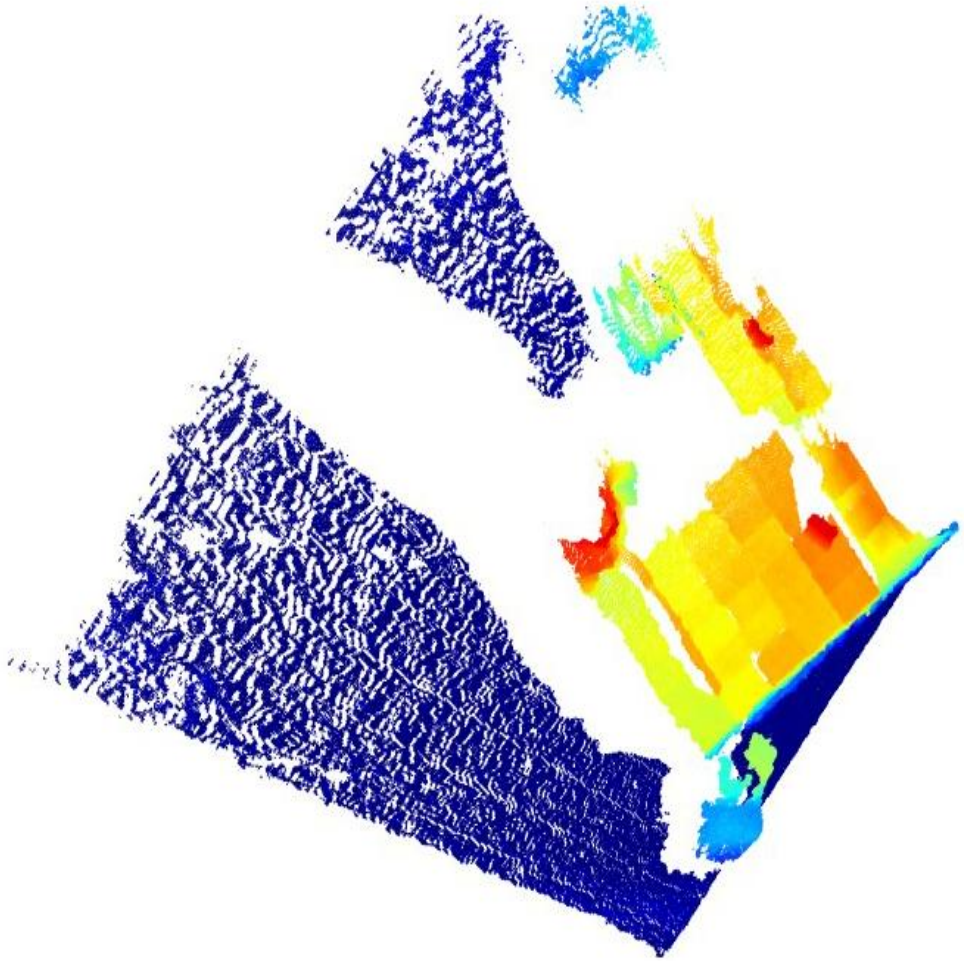
Monitor the area close to the table of new books using 3D information



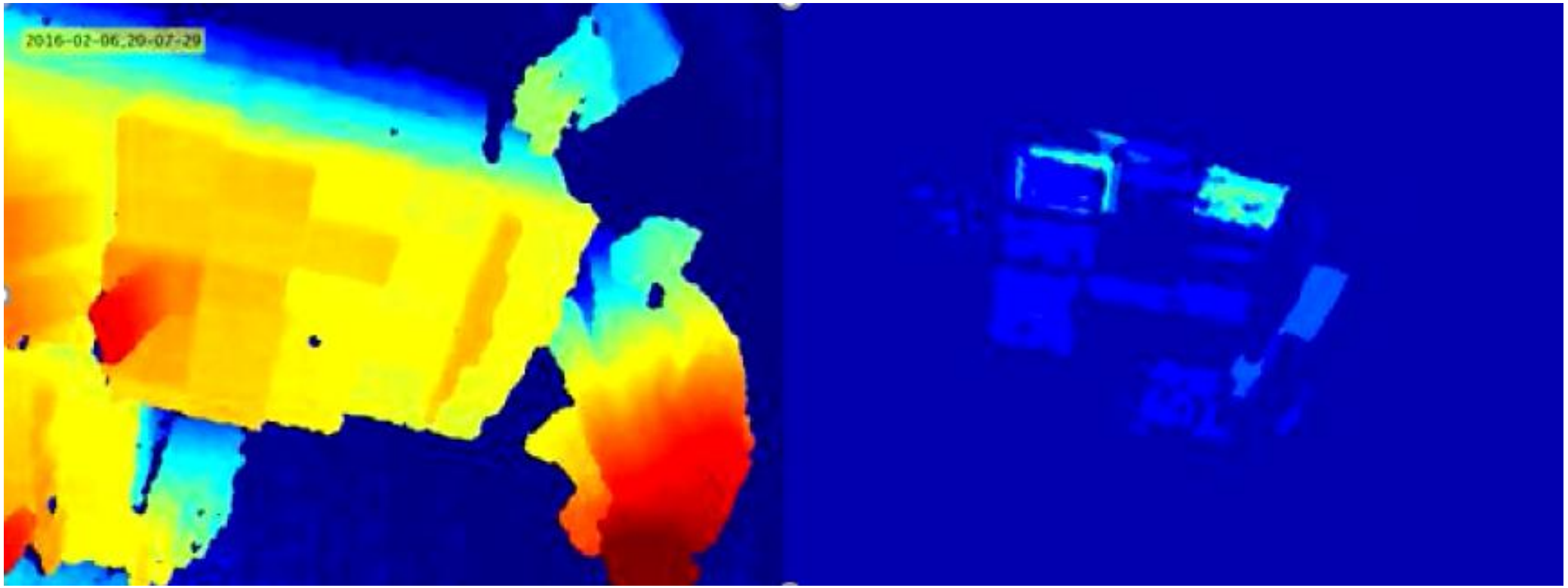
Detect and follow shopper in real-time using 3D representation of the space

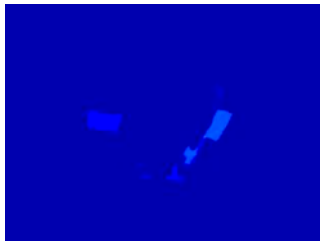
3D View of the Scene

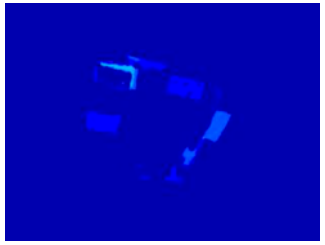


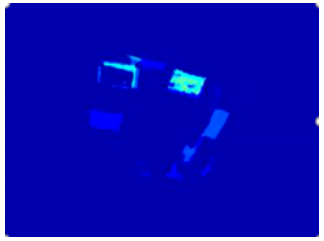


Count Pick and Take



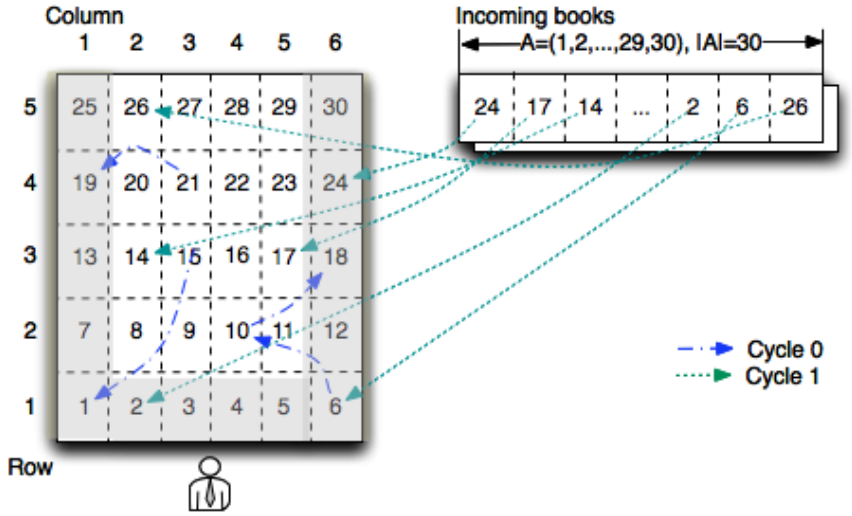






Randomized Experiment on Book Placement Recommendations

- We **randomize book placement** on the table across 5 rows and 6 columns
- At cycle 0 we randomly shuffle the books already on the table. Then, each cycle we generate a random sequence of numbers between 1 and 30 and place the incoming books accordingly



Treated group: books in **edge positions** (first row and left/right most columns)

Control group: books in the **center positions**

Pick Rate by Positions (sample)

6	**7.66**	0.43	0.43	1.06	2.98	**4.26**
5	2.55	3.19	2.98	0.85	1.91	**4.26**
4	2.55	0.85	1.06	1.28	1.06	3.61
3	1.70	1.06	1.28	1.49	1.06	1.06
2	2.77	2.13	1.06	2.98	2.13	1.70
1	**5.96**	3.40	2.55	**6.81**	3.83	**5.53**
Left/Front	**1**	**2**	**3**	**4**	**5**	**6**

Pick Rate > 4

Pick Rate < 2

Empirical Results

- Books placed at the edge of the table are **picked 102%** more often and **taken 77%** more often per day than those at the center of the table

- However, conditionally on picking the book, shoppers are **equally likely** to take books placed at the edge and at the center of the table

- Book placement recommendations affects consumer choice **through its effect on the search process** but not through its effect on the consideration process

- The bookstore manager may maximize profit by placing **books with higher margins at the edge of the table**

Putting it in perspective

president's **MESSAGE**

José M.F. Moura
2008–2009 SPS President
j.moura@ieee.org

What Is Signal Processing?

I am returning to this perennial question that, in some sense, I already touched on in my March 2008 column. I am sure we are all often confronted with explaining signal processing to others, including nonexperts or the public in general. We may choose to explain it by what we and our colleagues do, as exemplified by the diversity and richness of the work of the IEEE Signal Processing Society's (SPS) technical committees and the topical coverage of our numerous solely sponsored or cosponsored conferences, workshops, and publications. Alternatively, we might turn to the cyber age oracle (Wikipedia) for an intuitive description (I suggest you read it).

I challenged two former SPS presidents (Kaveh, Ray Liu, and Mercy Kowalczyk societies). I am writing this in August, with you reading it in November; thus, I cannot predict whether or not all interested parties will converge.

Now, back to the challenge. The current state of our discussions (I want to acknowledge my five comrades in "crime": Leah Jamieson, Rich Cox, Mos

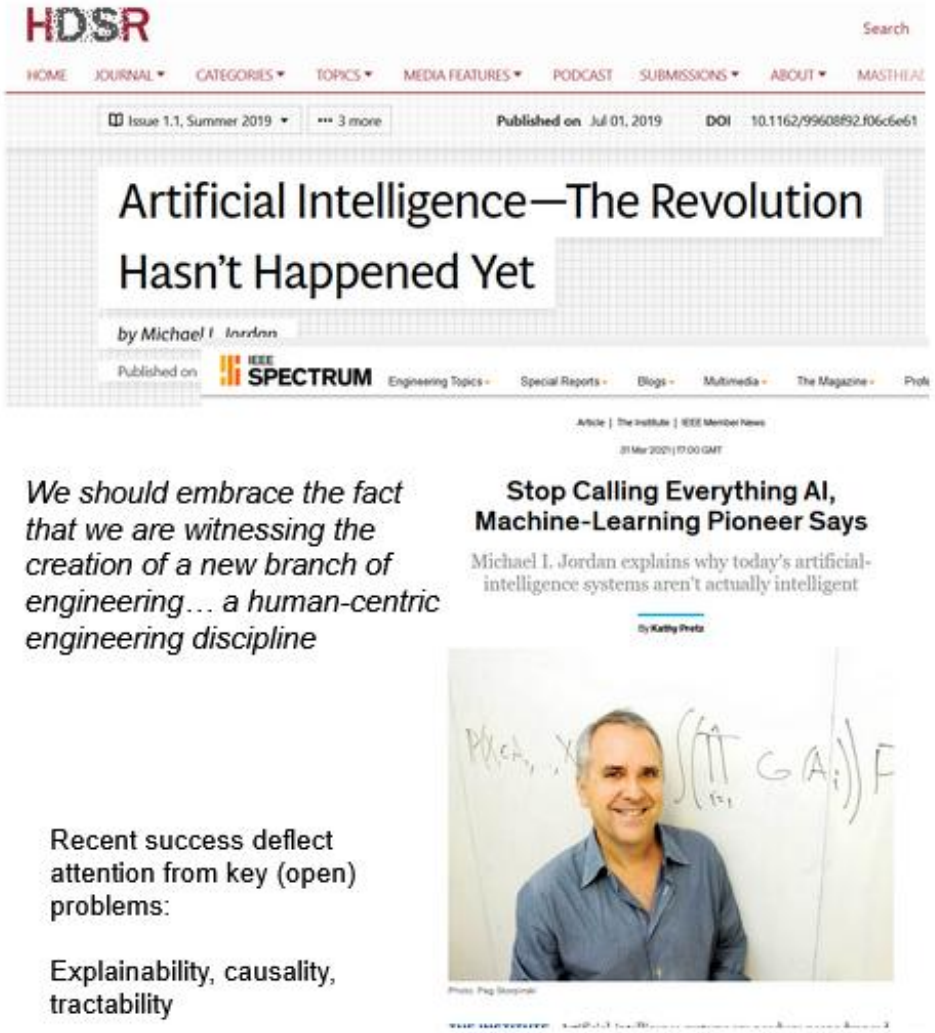
> THE CURRENT STATE OF OUR DISCUSSIONS ARE CENTERED ON TWO OBVIOUS QUESTIONS, "WHAT DO WE MEAN BY *SIGNAL*?" AND "HOW DO WE DEFINE *PROCESSING*?"

cal, chemical, molecular, genomic, medical, musical, data, or sequences of attributes, or numerical quantities; the list goes on.

As for *processing*, it comprises operations of representing, filtering, coding, transmitting, estimating, detecting, inferring, discovering, recognizing, synthesizing, recording, or reproducing signals by digital or analog devices, techniques, or algorithms, in the form of software, hardware, or firmware (Did we leave out other important techniques?).

So, putting it together, can we say that *signal processing* is an enabling technology that encompasses the fundamental theory, applications, algorithms, and implementations of processing or transferring information

https://www.researchgate.net/publication/224586627_What_is_signal_processing_President%27s_Message



https://hdsr.mitpress.mit.edu/pub/wot7mkc1/release/10