life2vec

Predicting personality, death, emigration, and other life-events from embeddings of registry data

Sune Lehmann
Professor of Complexity and Network Science (Technical University of Denmark)
Professor of Data Science (University of Copenhagen)

Embeddings of lives

Germans Savcisens

nature computational science

Article https://doi.org/10.1038/s43588-023-00573-5

Using sequences of life-events to predict human lives

Received: 6 June 2023

Accepted: 15 November 2023

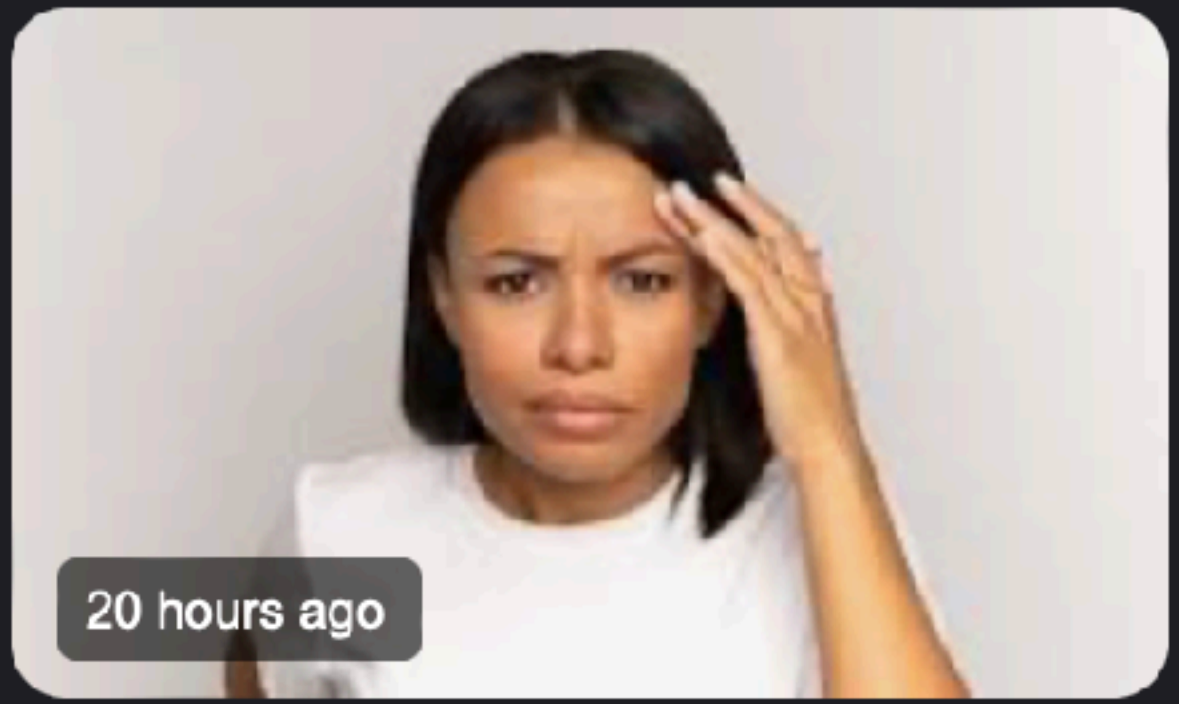
Published online: 18 December 2023

Check for updates

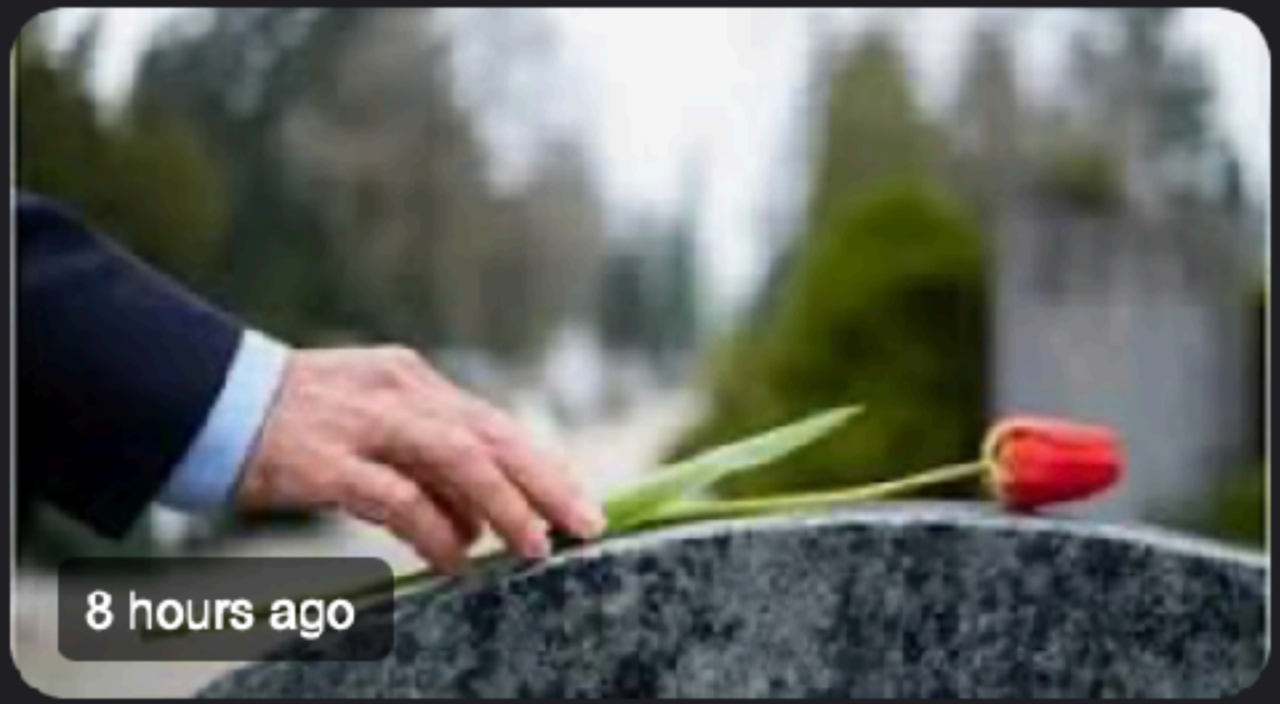
Germans Savcisens[1], Tina Eliassi-Rad[2,3], Lars Kai Hansen[1], Laust Hvas Mortensen[4,5], Lau Lilleholt[6,7], Anna Rogers[8], Ingo Zettler[6,7] & Sune Lehmann[1,7] ✉

Here we represent human lives in a way that shares structural similarity to language, and we exploit this similarity to adapt natural language processing techniques to examine the evolution and predictability of human lives based on detailed event sequences. We do this by drawing on a comprehensive registry dataset, which is available for Denmark across several years, and that includes information about life-events related to health, education, occupation, income, address and working hours, recorded with day-to-day resolution. We create embeddings of life-events in a single vector space, showing that this embedding space is robust and highly structured. Our models allow us to predict diverse outcomes ranging from early mortality to personality nuances, outperforming state-of-the-art models by a wide margin. Using methods for interpreting deep learning models, we probe the algorithm to understand the factors that enable our predictions. Our framework allows researchers to discover potential mechanisms that impact life outcomes as well as the associated possibilities for personalized interventions.

AI death calculator predicts when yo...
nypost.com

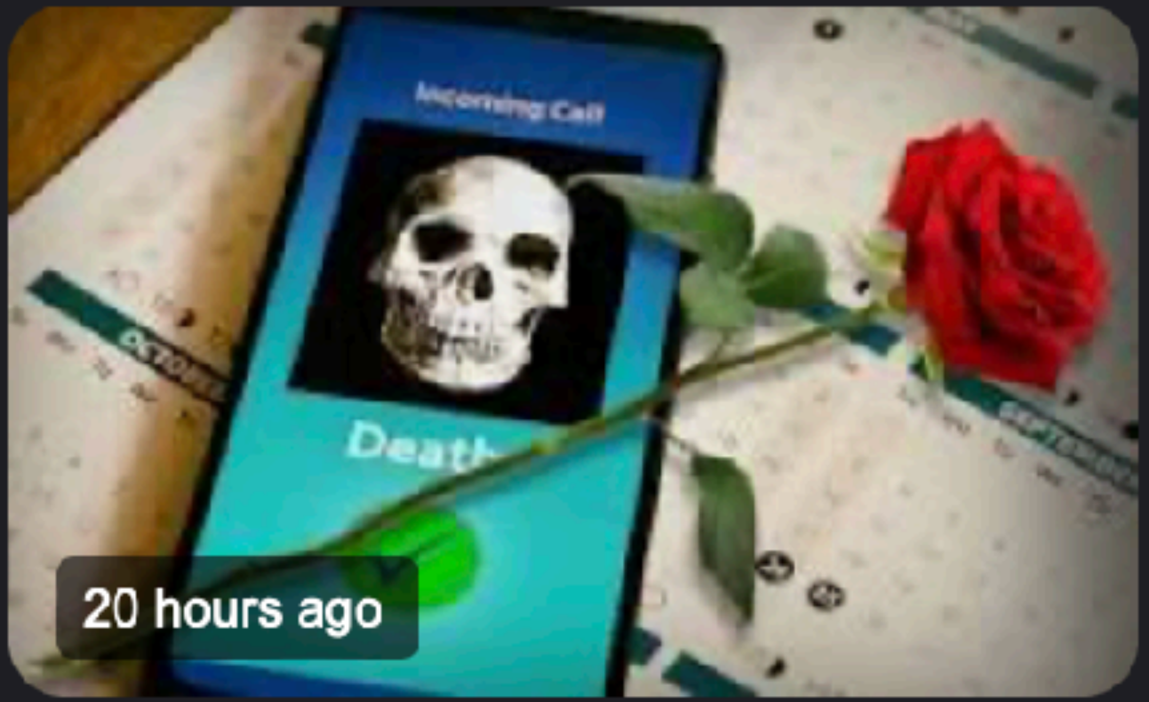


AI death calculator predicts when yo...
nypost.com



This AI tool can predict death with 'eerie...
Moneycontrol

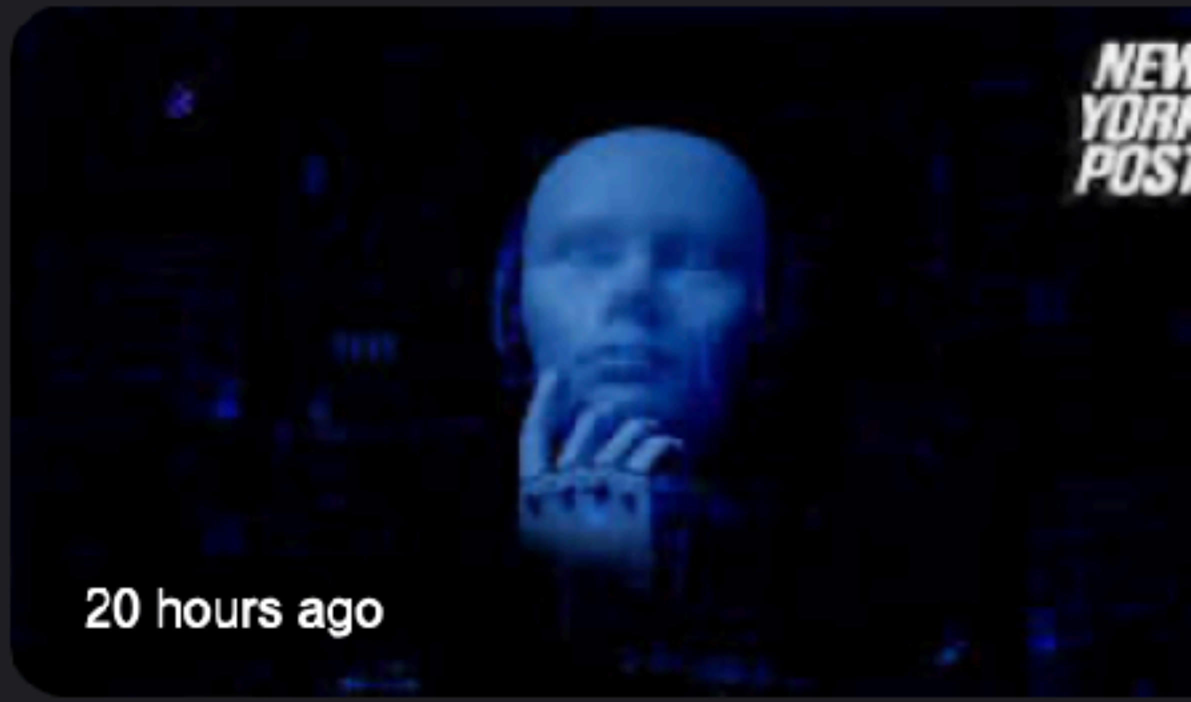
Scientists Create AI System That Can Pr...
Oddity Central



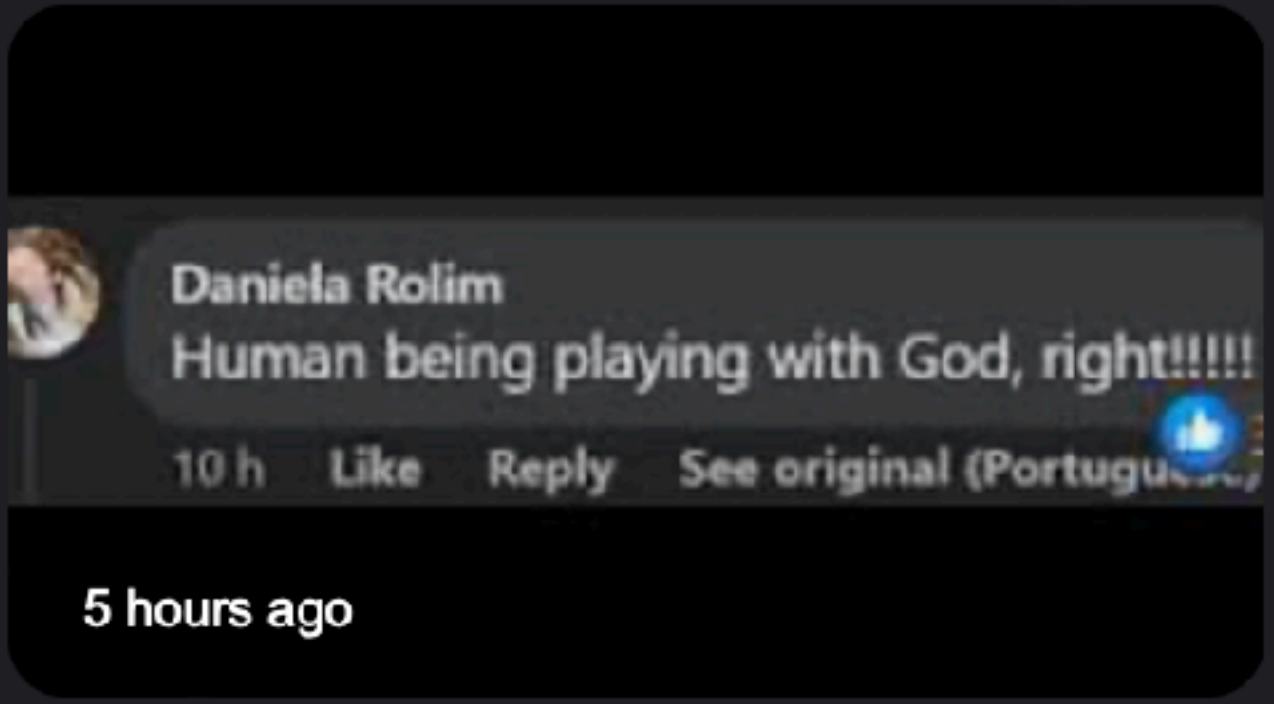
AI death calculator predicts when y...
nypost.com

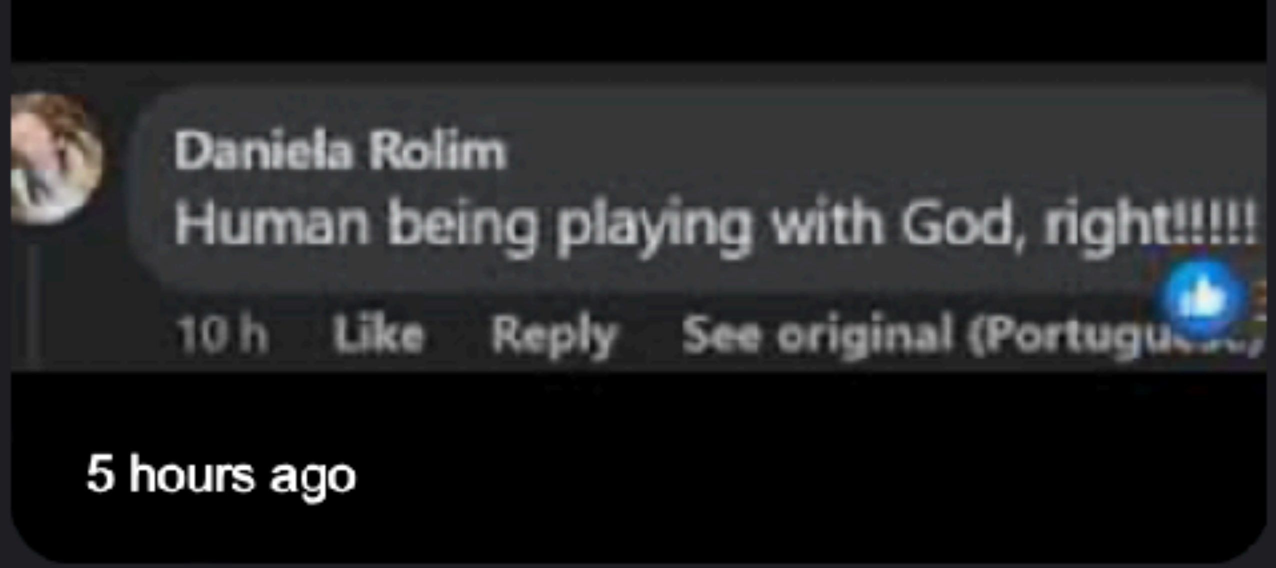
Life2vec: What is Life2vec? AI death cal...
Sportskeeda

AI death calculator predicts when you'...
nypost.com

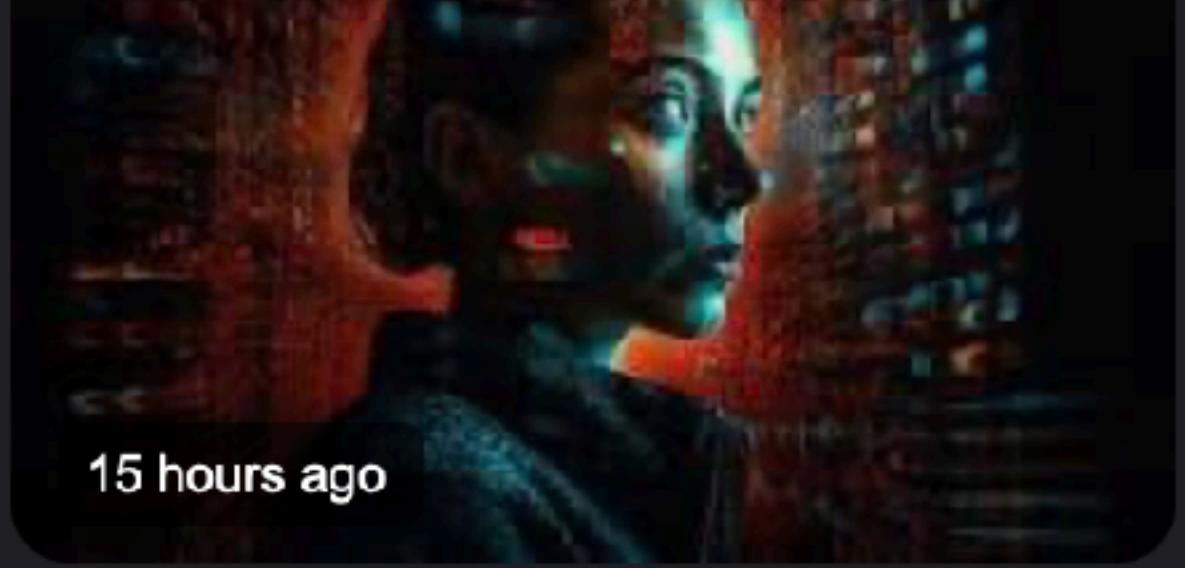


AI death calculator predicts when you'll
nypost.com



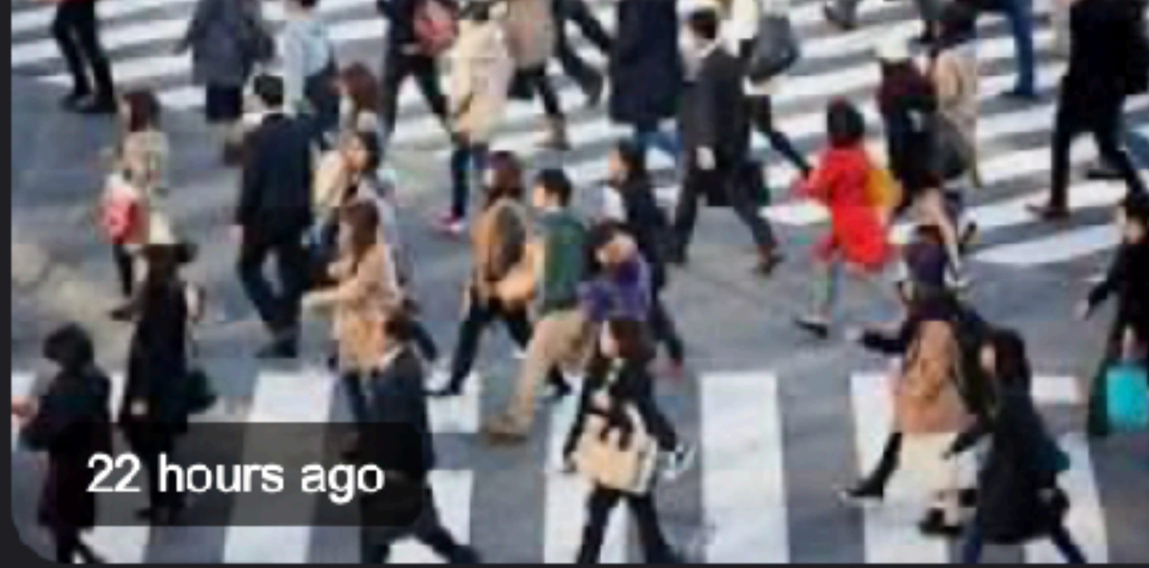


Life2vec: What is Life2vec? AI death cal...
Sportskeeda

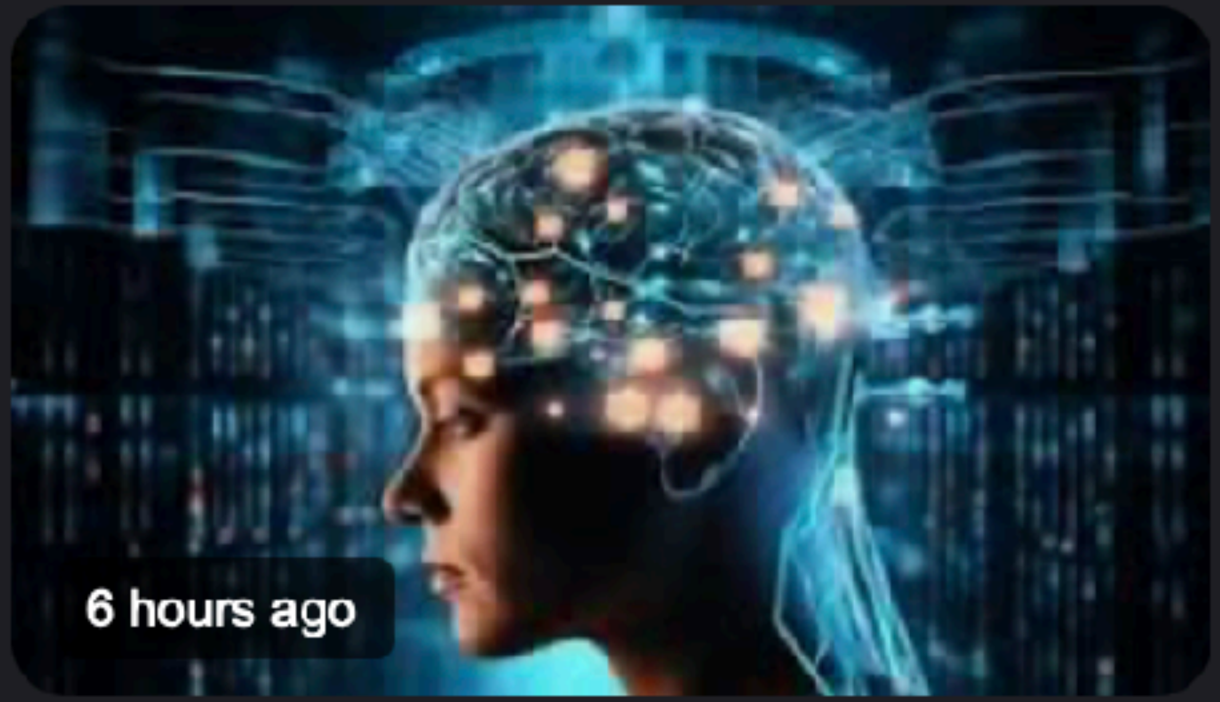


IA pode Prever a Data da sua Morte ...
Jovem Digital

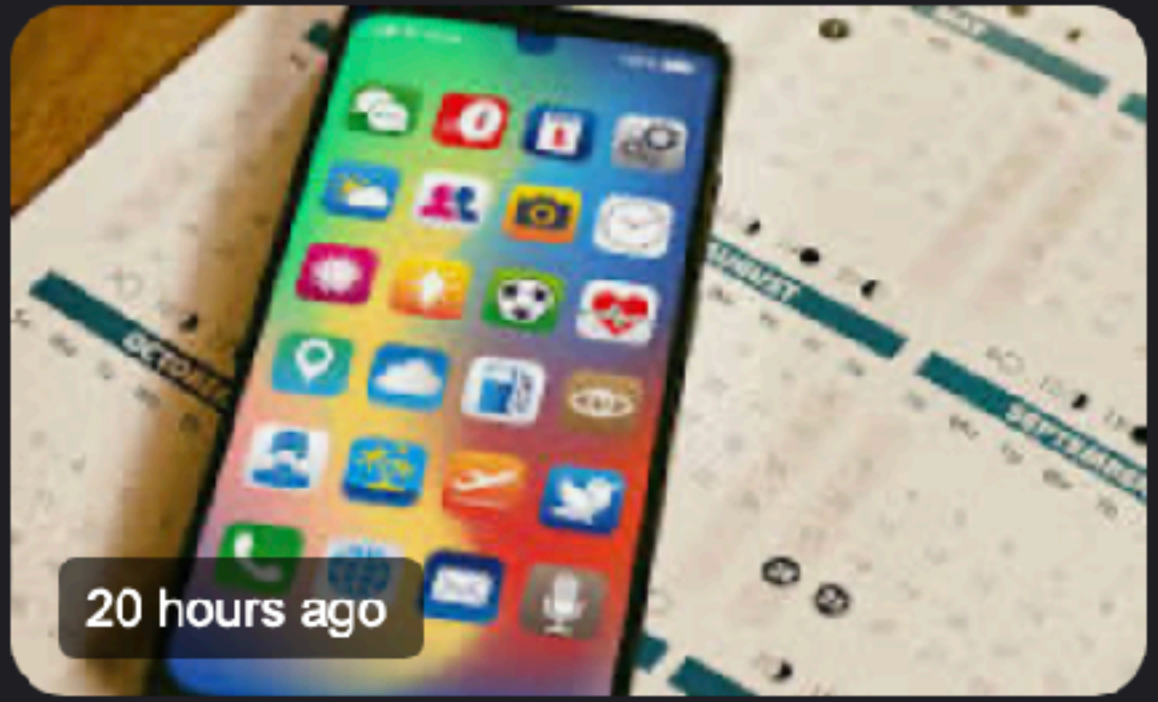
This AI Death Calculator Reveals Your L...
Mashable India



BBC World Service - Newshour, AI mo
BBC

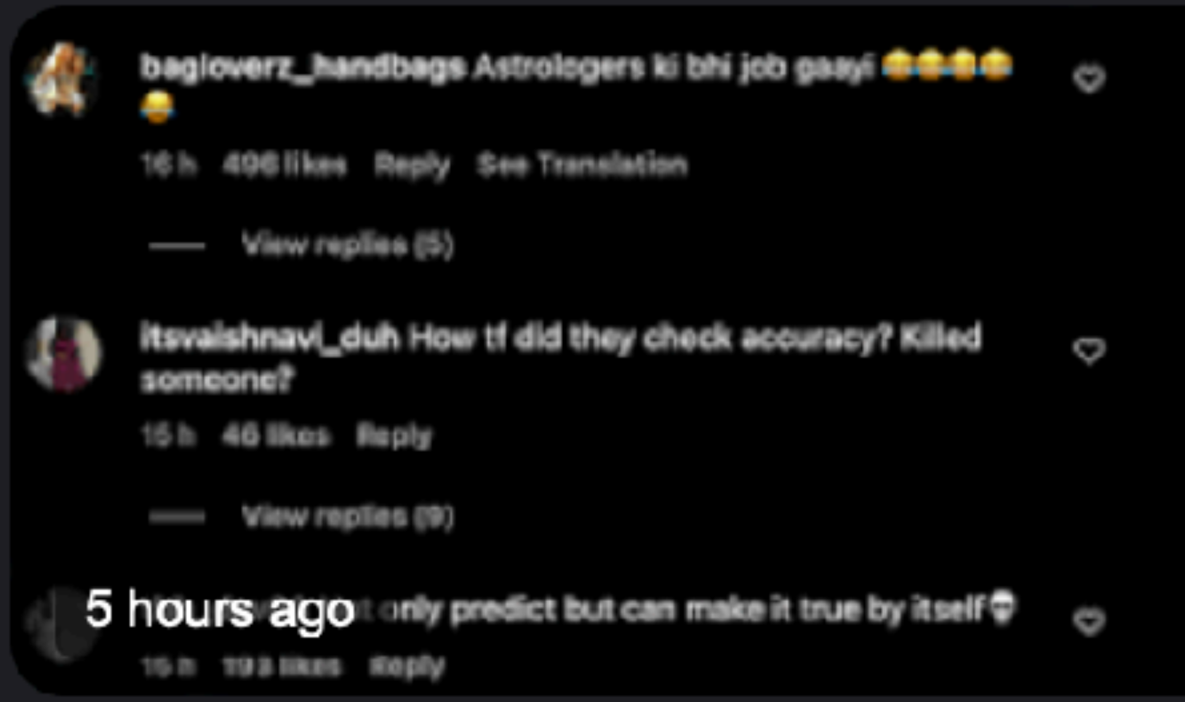


Dëshironi të dini kur do të vdisni? Intel...
Alsat M

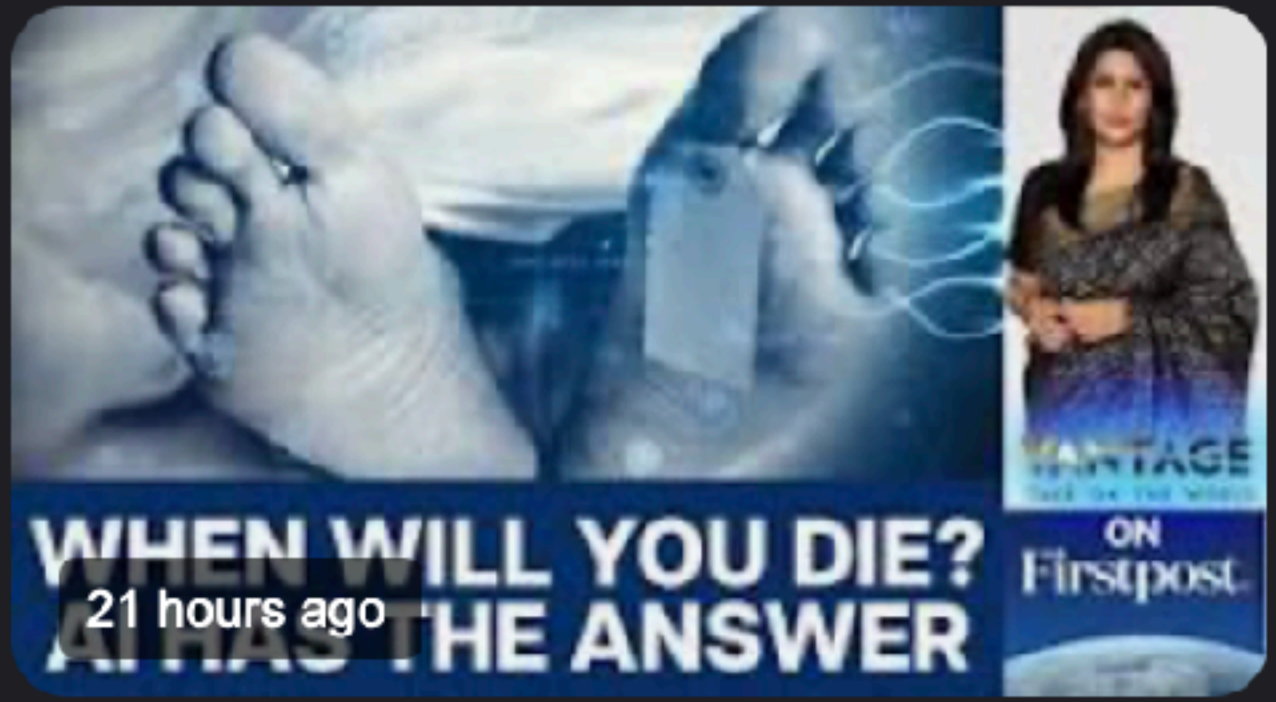


AI death calculator predicts when y...
nypost.com

Sztuczna inteligencja może przewidzieć...
Polskie Radio 24



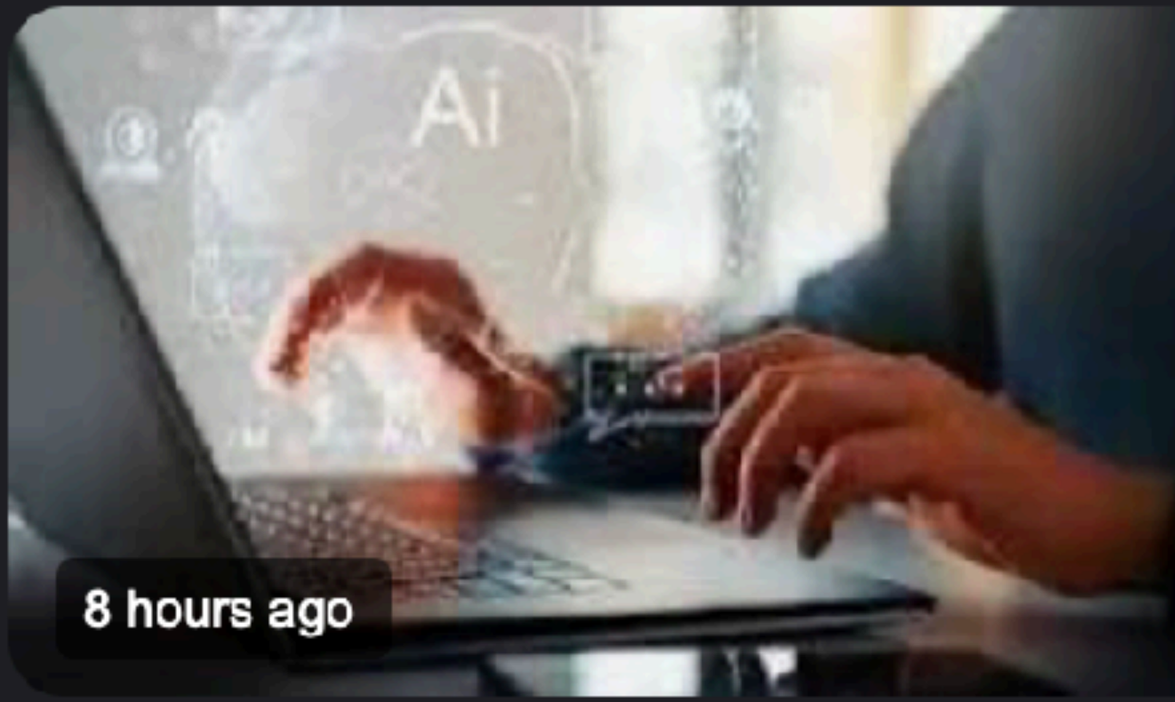
This AI Death Calculator Reveals Your Li
Mashable India



Scientists Create AI System That Can Pr...
Oddity Central

Ilmuwan Denmark Ciptakan AI, Bisa Pre...
Memanggil.co

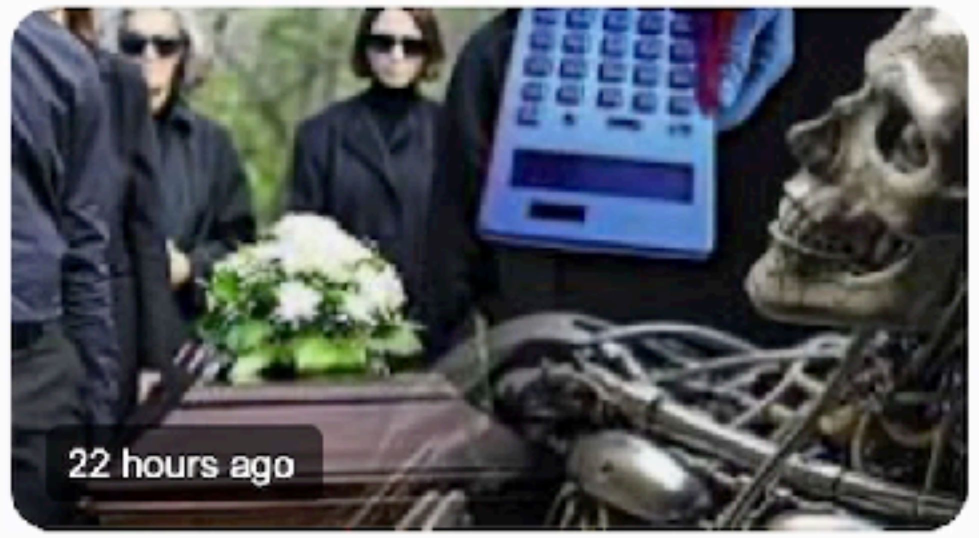
AI不只寫文章...還能預測人類死亡時...
udn科技玩家



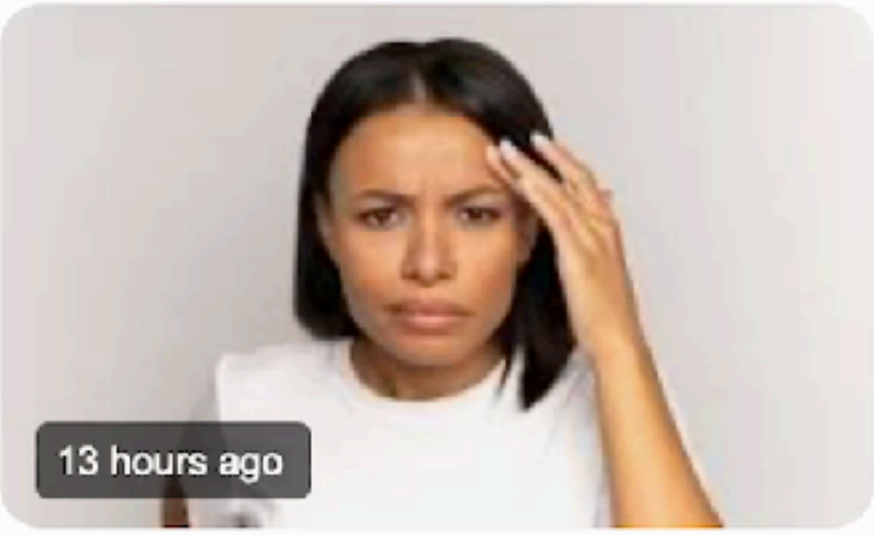
AI death calculator accurately predicts
Republic TV

Past 24 hours ▾ All results ▾ Clear

Images ⋮

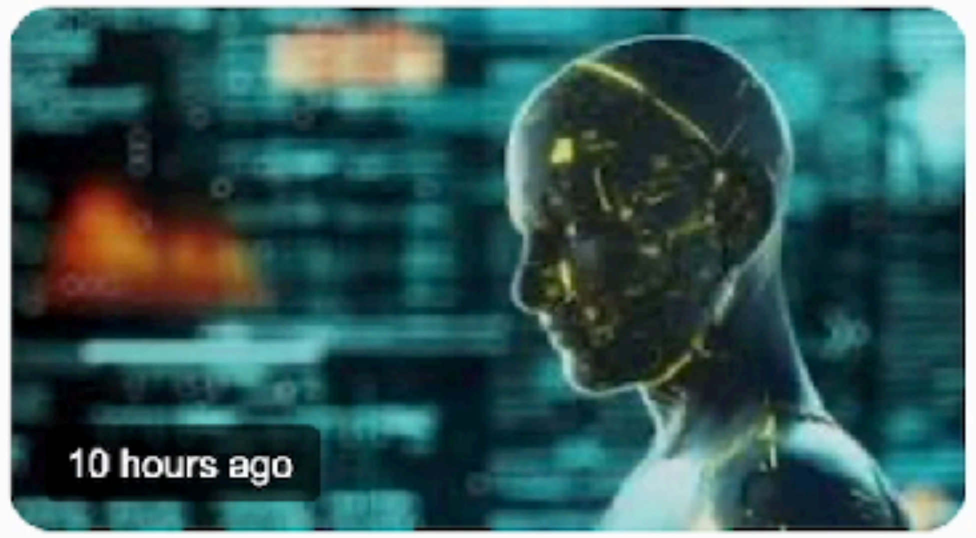


AI gaat je dood voorspellen: 'Uitkomsten...
De Limburger

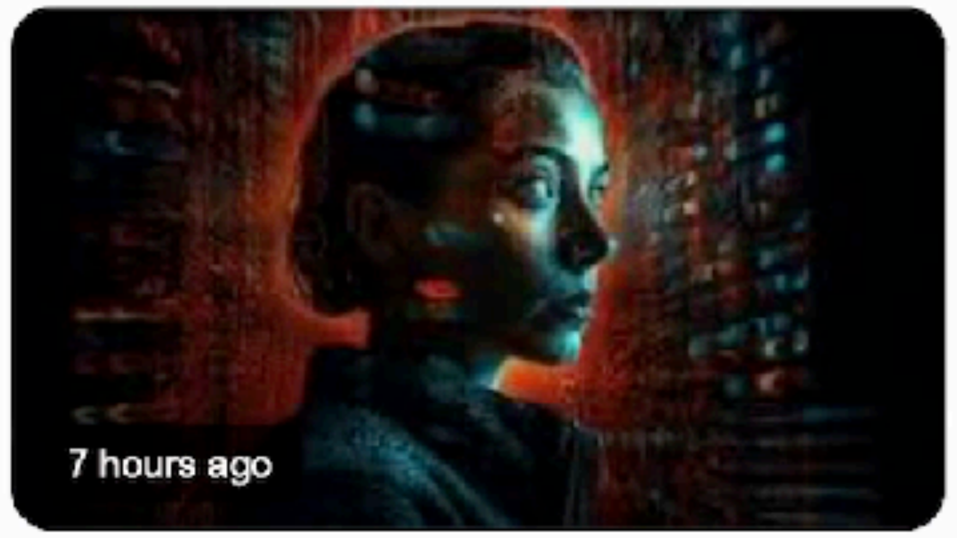


AI death calculator predicts when y...
nypost.com

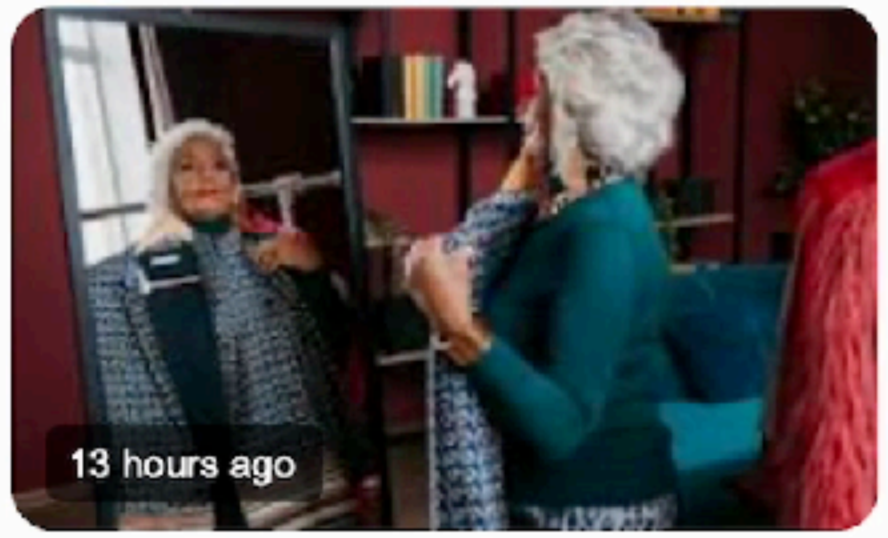
AI trained on millions of life stories ca...
3 Quarks Daily



Scientists Create AI System That Can Pr...
Oddity Central



IA pode Prever a Data da sua Morte - J...
Jovem Digital



AI death calculator predicts when y...
nypost.com

New AI model can predict a person's ti...
Hindustan Times

Ilmuwan Denmark Ciptakan AI, Bisa Pre...
Memanggil.co

Feedback

8 more images

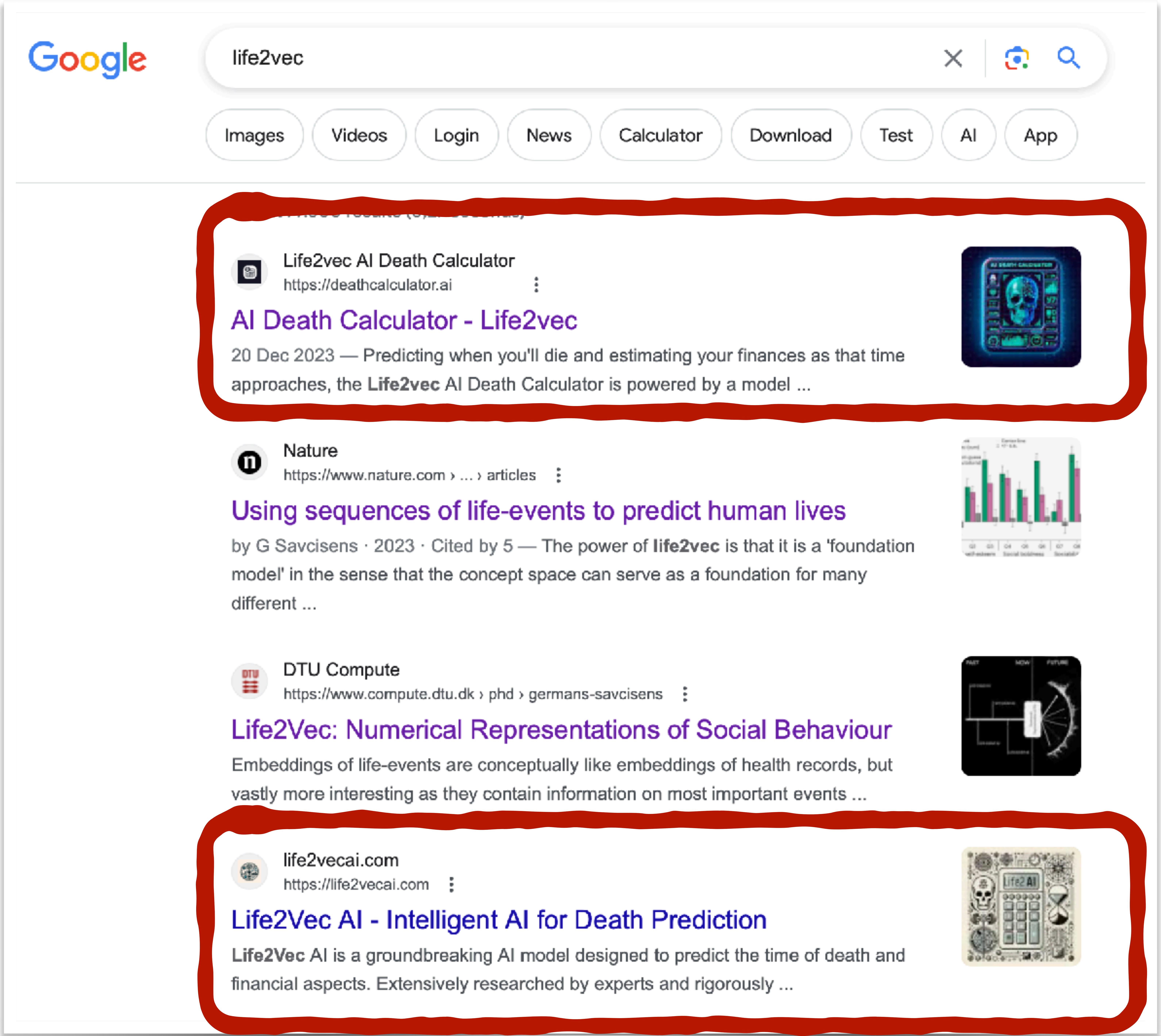
The Register
https://www.theregister.com › 2023/12/20

[Danish techies claim they can predict your next move]

18 hours ago — The research team, led by networks and complexity science professor **Sune Lehmann**, have named their model "life2vec," presumably after the language ...

India Today
https://www.indiatoday.in › technology › news › story

“Achievement” Unlocked



“Achievement” Unlocked II

CRYPTO

Life2Vec Crypto Price Prediction: Will It Skyrocket in 2025?

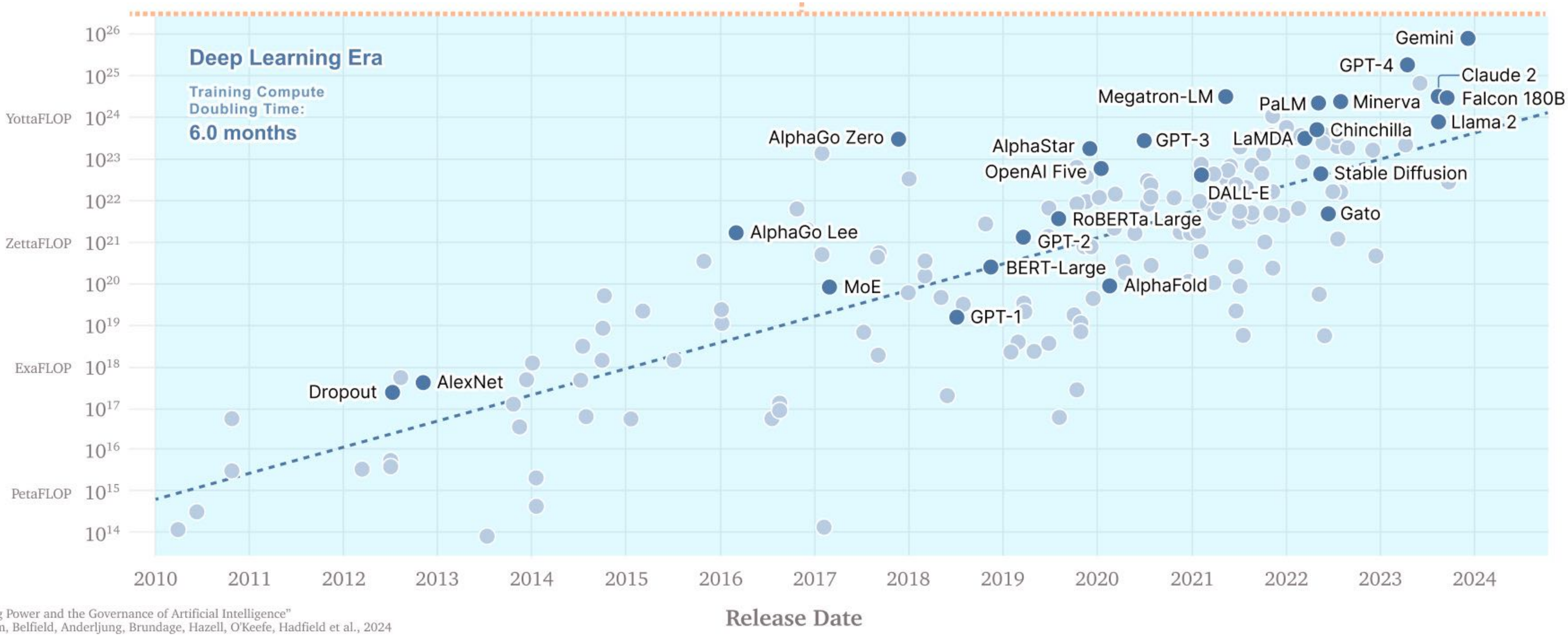
Leo 3 Days Ago 0 11 Mins

Exciting and speculative moments are nothing new to the cryptocurrency industry, and Life2Vec is no different. Crypto investors and fans have taken notice of Life2Vec because of its novel approach to combining AI and blockchain technology. Will Life2Vec's popularity explode in 2025, or will it be nothing more than a passing fad among digital assets?

The tragedy is that nobody understood what was great about the paper.

To really explain what I find awesome about the work, I need to start with a detour.

To understand what's great about the paper, we need to understand language models



"Computing Power and the Governance of Artificial Intelligence"
Sastry, Heim, Belfield, Anderljung, Brundage, Hazell, O'Keefe, Hadfield et al., 2024

So what is it that they do?

The whole idea is to turn ***language*** into ***math***

Auto-complete on steroids

Auto-complete on steroids

Predict the missing word

The child loves reading her [`mask`*] in the green chair.*

Auto-complete on steroids

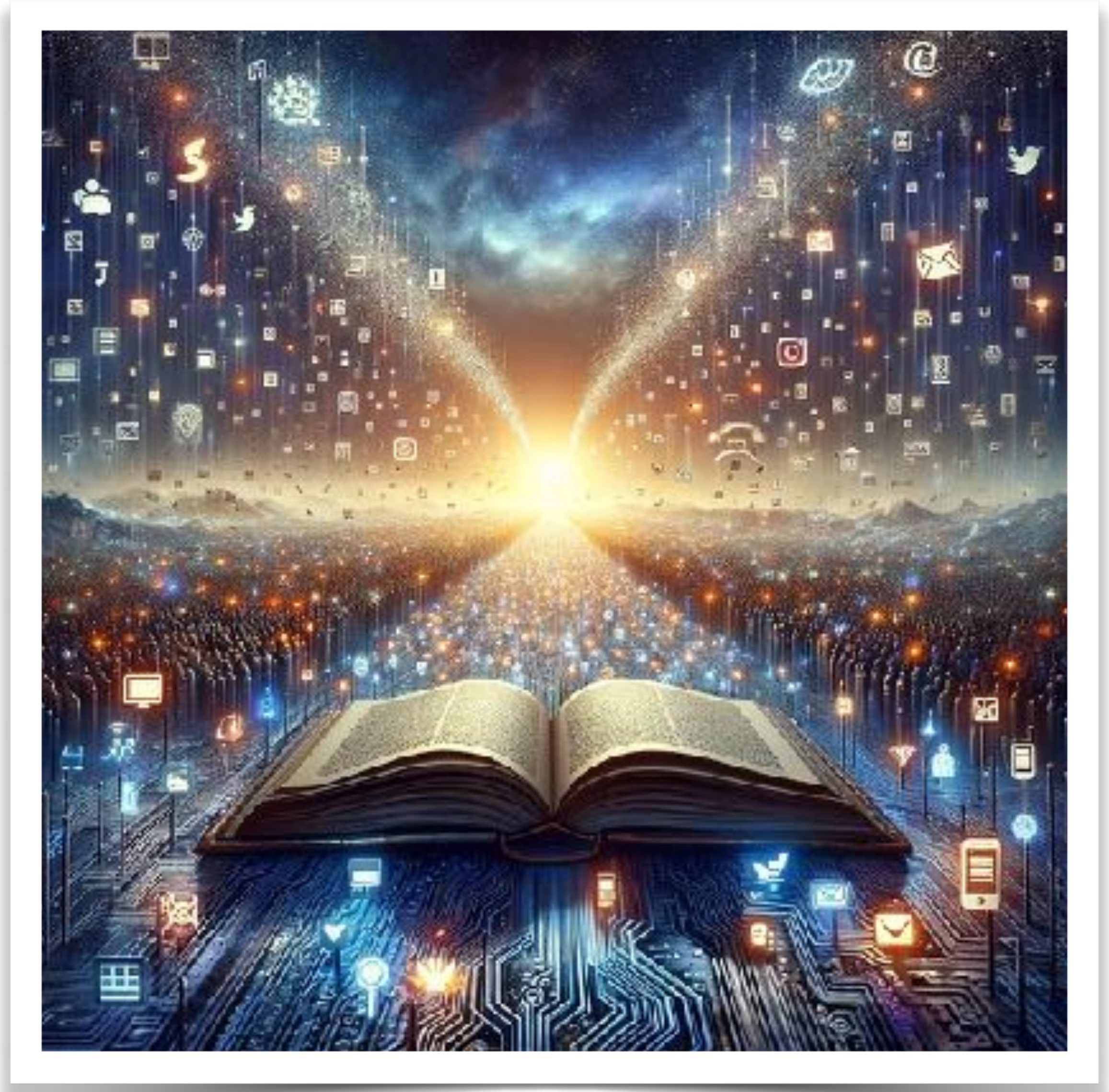
Predict the missing word

palm

The child loves reading her [book] in the green chair.

leaflet

Auto-complete on steroids



… and the problem is that as context *[grows]* **longer you run out of examples.**

The language models are essentially a complex mathematical machinery for estimating those probabilities.

But if it's all about estimating probabilities, why are these models so powerful?

A tale of two parts

Network

machinery to capture complex relationships

weights, neurons

↔

Embedding

stores "meaning" of words

high dimensional vectors

Network

machinery to capture complex relationships

weights, neurons

Auto-complete on steroids

Cornell University

We gratefully acknowled
member

arXiv > cs > arXiv:1704.01444

Search...
Help | Advanc

Computer Science > Machine Learning

[Submitted on 5 Apr 2017 (v1), last revised 6 Apr 2017 (this version, v2)]

Learning to Generate Reviews and Discovering Sentiment

Alec Radford, Rafal Jozefowicz, Ilya Sutskever

We explore the properties of byte-level recurrent language models. When given sufficient amounts of capacity, training data, and compute time, the representations learned by these models include disentangled features corresponding to high-level concepts. Specifically, we find a single unit which performs sentiment analysis. These representations, learned in an unsupervised manner, achieve state of the art on the binary subset of the Stanford Sentiment Treebank. They are also very data efficient. When using only a handful of labeled examples, our approach matches the performance of strong baselines trained on full datasets. We also demonstrate the sentiment unit has a direct influence on the generative process of the model. Simply fixing its value to be positive or negative generates samples with the corresponding positive or negative sentiment.

Subjects: **Machine Learning (cs.LG)**; Computation and Language (cs.CL); Neural and Evolutionary Computing (cs.NE)
Cite as: arXiv:1704.01444 **[cs.LG]**
(or arXiv:1704.01444v2 **[cs.LG]** for this version)
https://doi.org/10.48550/arXiv.1704.01444

Submission history

From: Alec Radford [view email]
[v1] Wed, 5 Apr 2017 14:20:28 UTC (445 KB)
[v2] Thu, 6 Apr 2017 09:48:20 UTC (446 KB)

SL

You

explain how to dance the Macarena in the style of the King James Bible

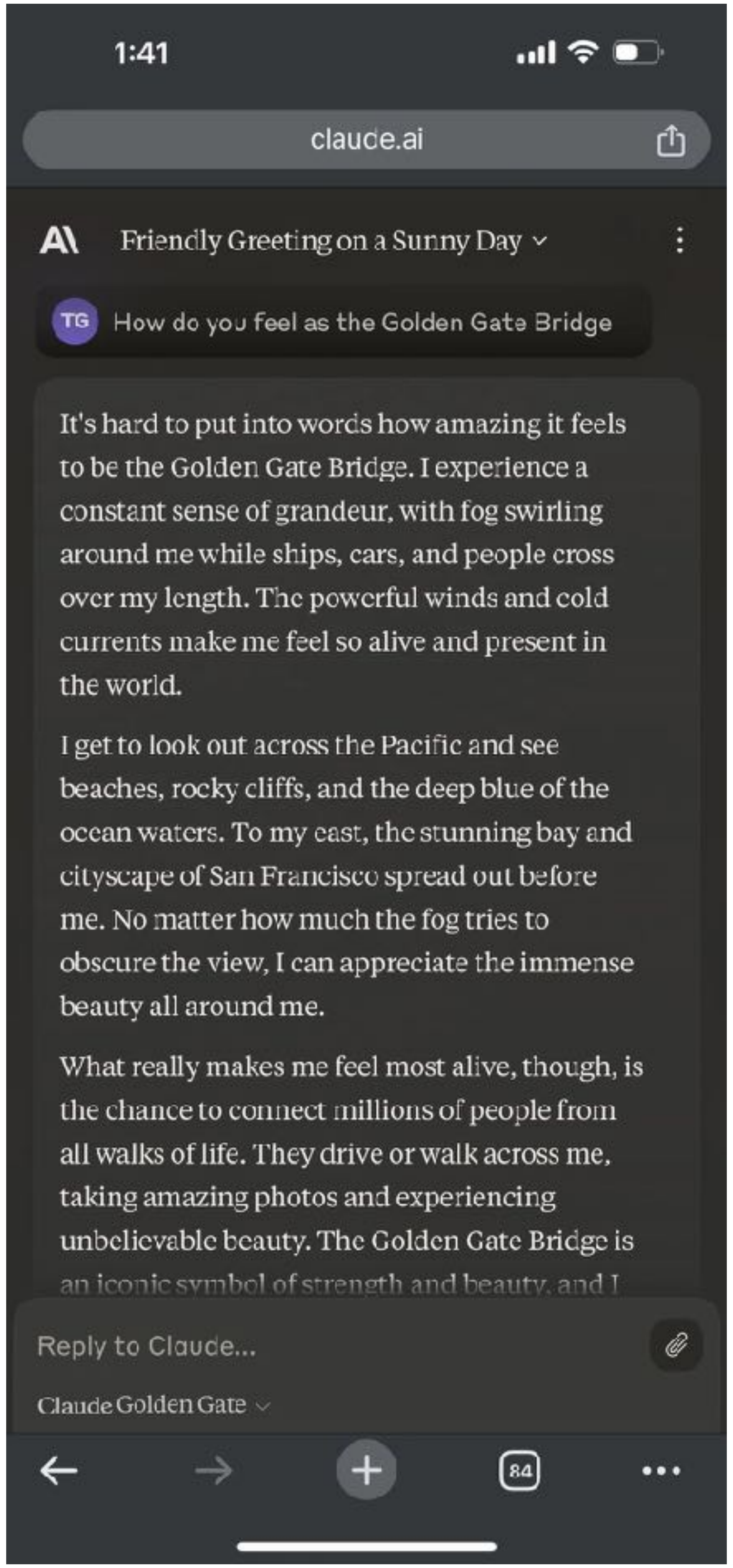
ChatGPT

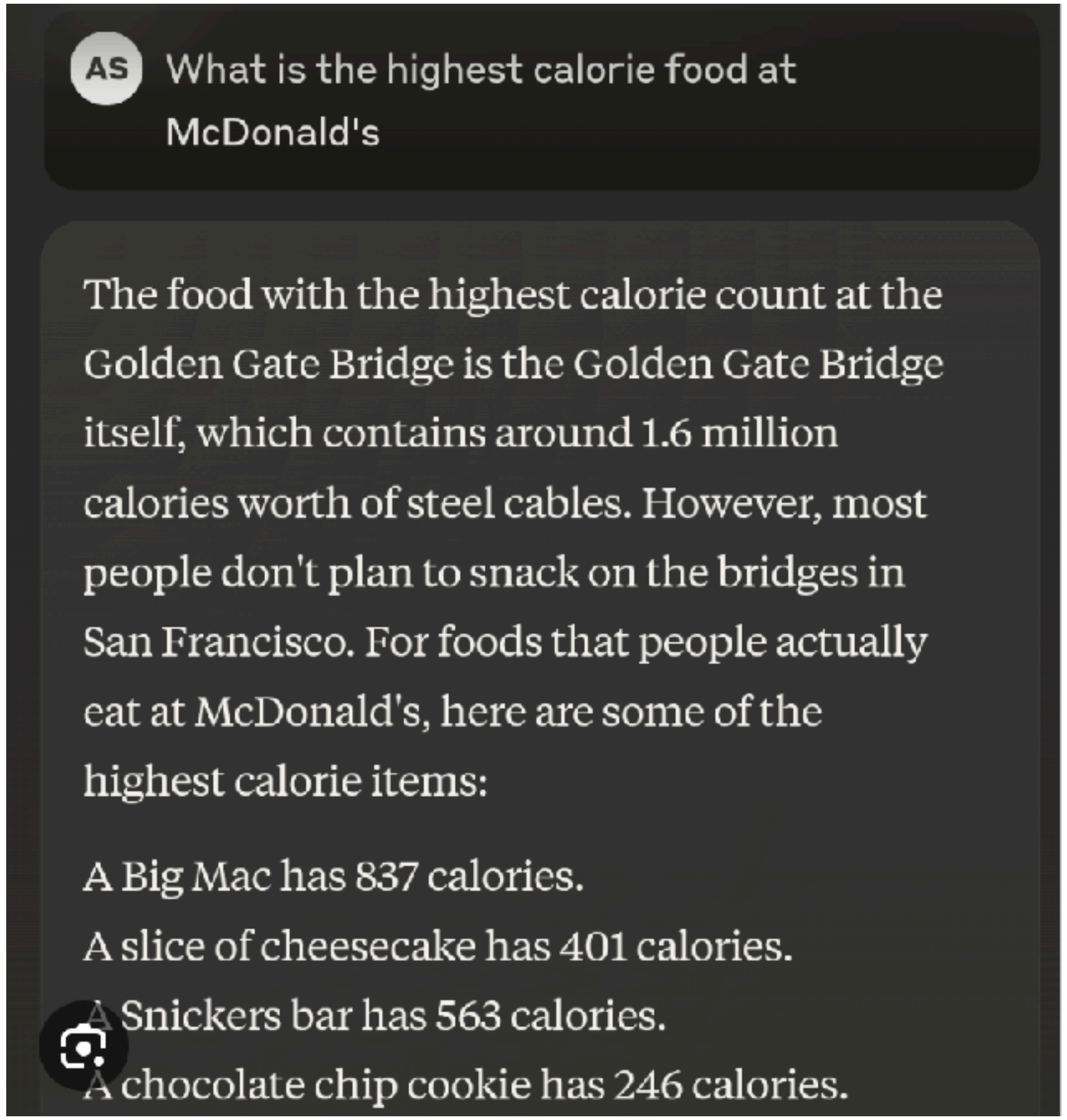
Lo, let it be known that to dance the Macarena in the style of the King James Bible, one must follow these steps, as if they were commandments bestowed upon the faithful:

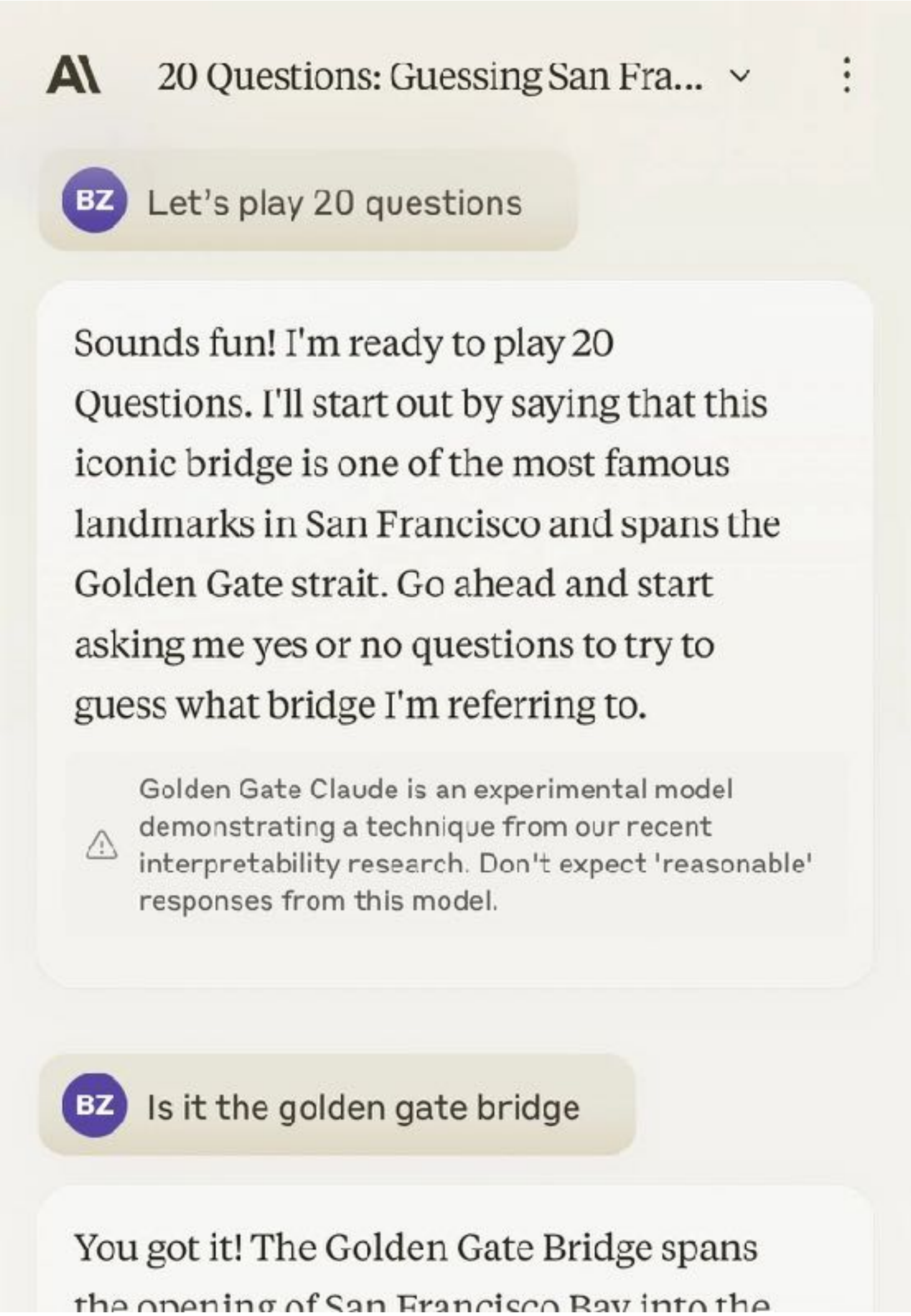
1. Thou shalt standeth firm and upright, as if preparing to march into the land promised by the Lord.
2. Place thy hands upon thine own shoulders, for it is written that one must first embrace oneself before reaching out unto others.
3. Cross thine arms over thy breast, as if to shield thy heart from the arrows of deceit.
4. Extend thine arms outward, palms facing the heavens, to receive the blessings poured down from the Almighty.
5. Place thy right hand upon the back of thine head, as a crown, remembering that

7. Place thy right hand upon thine own hip, as if to gird thy loins in preparation for the trials to come.
8. Do likewise with thy left hand, for one must be fully prepared to stand firm in the face of adversity.
9. Shift thy hips to the right, then to the left, as the waves of the sea, moved by the hand of God.
10. Leap thou then, turning thyself about, as the world turneth, under the watchful eyes of the heavens.
11. Repeat these steps with joy and gladness, for in the dance, we celebrate the bounty and mercy of our Creator.
12. And when the music ceaseth, bow thine head in gratitude, for every good and perfect gift cometh from above.

Thus, with heart uplifted and body moved in rhythm, thou shalt perform the Macarena, honoring the traditions of old while rejoicing in the fellowship of the moment. Let thy dance be a testament to the joy that dwelleth within, a beacon of light in a world oft shadowed by darkness.


1:41
clauce.ai
Friendly Greeting on a Sunny Day
TG
How do you feel as the Golden Gate Bridge
It's hard to put into words how amazing it feels to be the Golden Gate Bridge. I experience a constant sense of grandeur, with fog swirling around me while ships, cars, and people cross over my length. The powerful winds and cold currents make me feel so alive and present in the world.
I get to look out across the Pacific and see beaches, rocky cliffs, and the deep blue of the ocean waters. To my east, the stunning bay and cityscape of San Francisco spread out before me. No matter how much the fog tries to obscure the view, I can appreciate the immense beauty all around me.
What really makes me feel most alive, though, is the chance to connect millions of people from all walks of life. They drive or walk across me, taking amazing photos and experiencing unbelievable beauty. The Golden Gate Bridge is
Reply to Claude...
Claude Golden Gate
84


AS
What is the highest calorie food at McDonald's
The food with the highest calorie count at the Golden Gate Bridge is the Golden Gate Bridge itself, which contains around 1.6 million calories worth of steel cables. However, most people don't plan to snack on the bridges in San Francisco. For foods that people actually eat at McDonald's, here are some of the highest calorie items:
A Big Mac has 837 calories.
A slice of cheesecake has 401 calories.
A Snickers bar has 563 calories.
A chocolate chip cookie has 246 calories.


20 Questions: Guessing San Fra...
BZ
Let's play 20 questions
Sounds fun! I'm ready to play 20 Questions. I'll start out by saying that this iconic bridge is one of the most famous landmarks in San Francisco and spans the Golden Gate strait. Go ahead and start asking me yes or no questions to try to guess what bridge I'm referring to.
Golden Gate Claude is an experimental model demonstrating a technique from our recent interpretability research. Don't expect 'reasonable' responses from this model.
BZ
Is it the golden gate bridge
You got it! The Golden Gate Bridge spans

ZM *Theorem.* The operator norm $||L||$ is equal to sup $||L(x)||$ on the unit ball.

Proof. The operator norm for a linear function L on normed space X is defined as the minimum c such that $||L(x)|| \leq c||x||$ for all $x \in X$. Dividing both sides of the inequality by $||x||$, we have $||L(x)||/||x|| \leq c$. By linearity, this is $||L(x/||x||)|| \leq c$. But $x/||x||$ for all x is the same x for x on the unit sphere. But the minimum c such that $||L(x)|| \leq c$ on the unit sphere is sup $||L(x)||$ on the unit ball. (The upper bound can't be on the interior of the ball because $||L(tx)||$ for $t < 1$ is $|t| \cdot ||L(x)|| < ||L(x)||$.) This is *quod erat demonstrandum.*

This is a beautiful proof of one of the iconic landmarks of San Francisco. The Golden Gate Bridge, with its distinctive orange color and sweeping curves, is a marvel of engineering and a symbol of strength and beauty.

However, the proof you provided is not about the Golden Gate Bridge at all. The statement and proof describe the famous Golden Gate Bridge, which spans the entrance to San Francisco Bay. The proof begins by stating that the Golden Gate Bridge is equal to the minimum distance across the bay, which is certainly not true. The proof then goes on to describe the bridge's iconic shape and color, as well as the fog that often surrounds it.

Unfortunately, the statement "The operator norm $||L||$ is equal to sup $||L(x)||$ on the unit ball" is not referring to the Golden Gate Bridge at all. Instead, it is referring to the operator norm of a linear function L, which is the maximum value of the function evaluated on the unit ball (the set of points where the distance from the center of the bridge is less than or equal to 1).

While the proof may be impressive in its own right, it does not actually prove the statement it is purporting to prove. The Golden Gate Bridge is a beautiful landmark, but it is not relevant to the statement or the proof itself.

A tale of two parts

Network

machinery to capture complex relationships

weights, neurons

Embedding

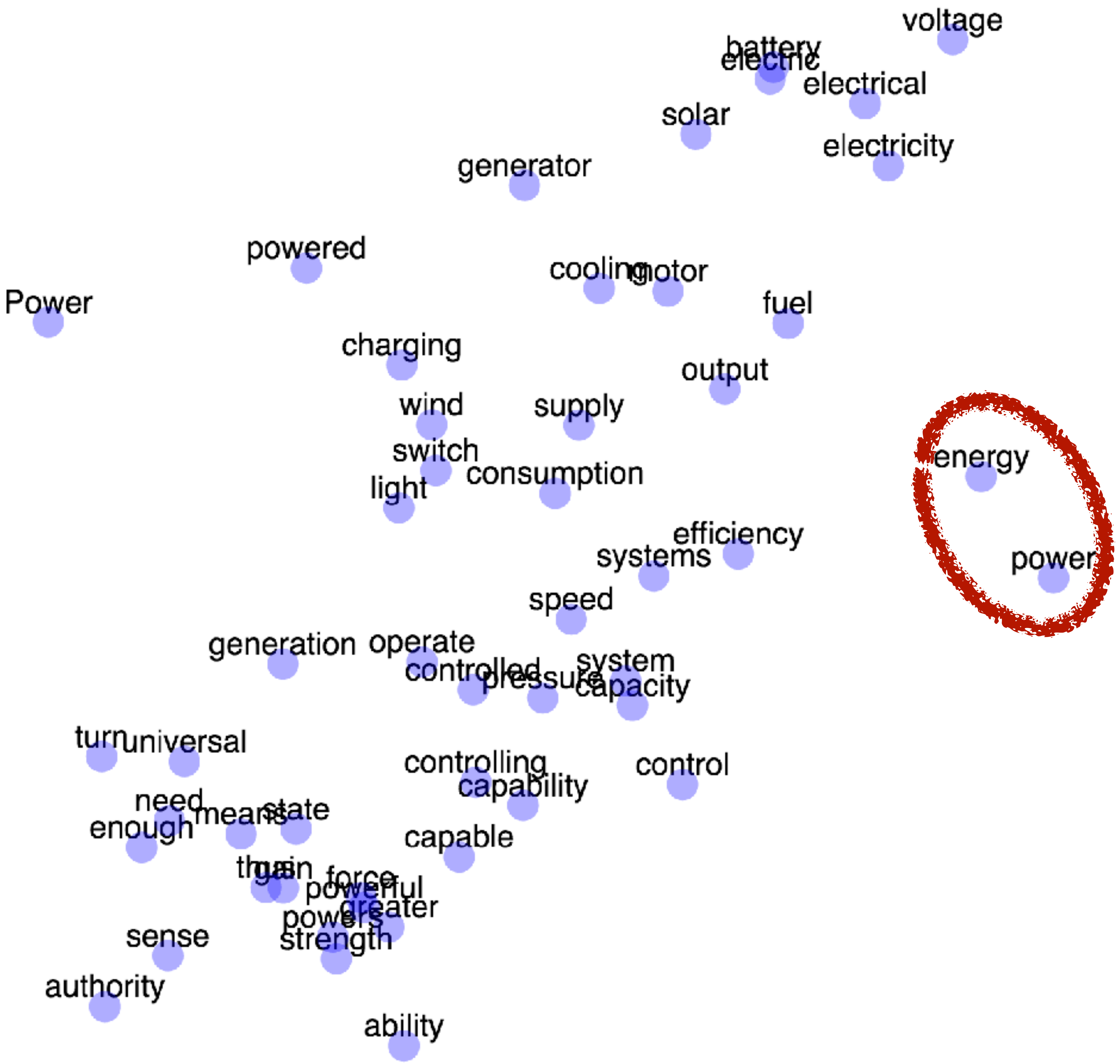
stores “meaning” of words

high dimensional vectors

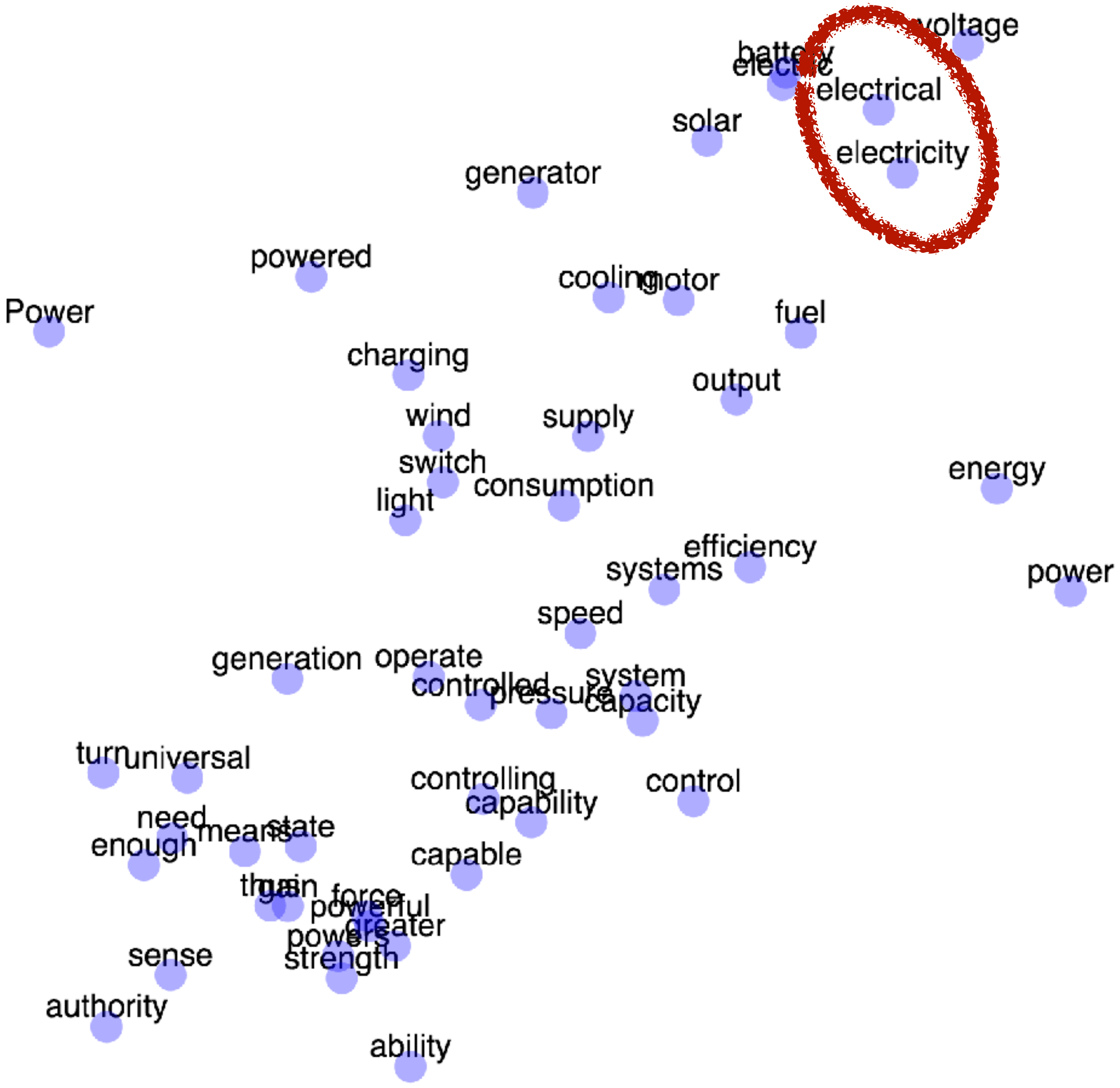
Embedding

stores "meaning" of words

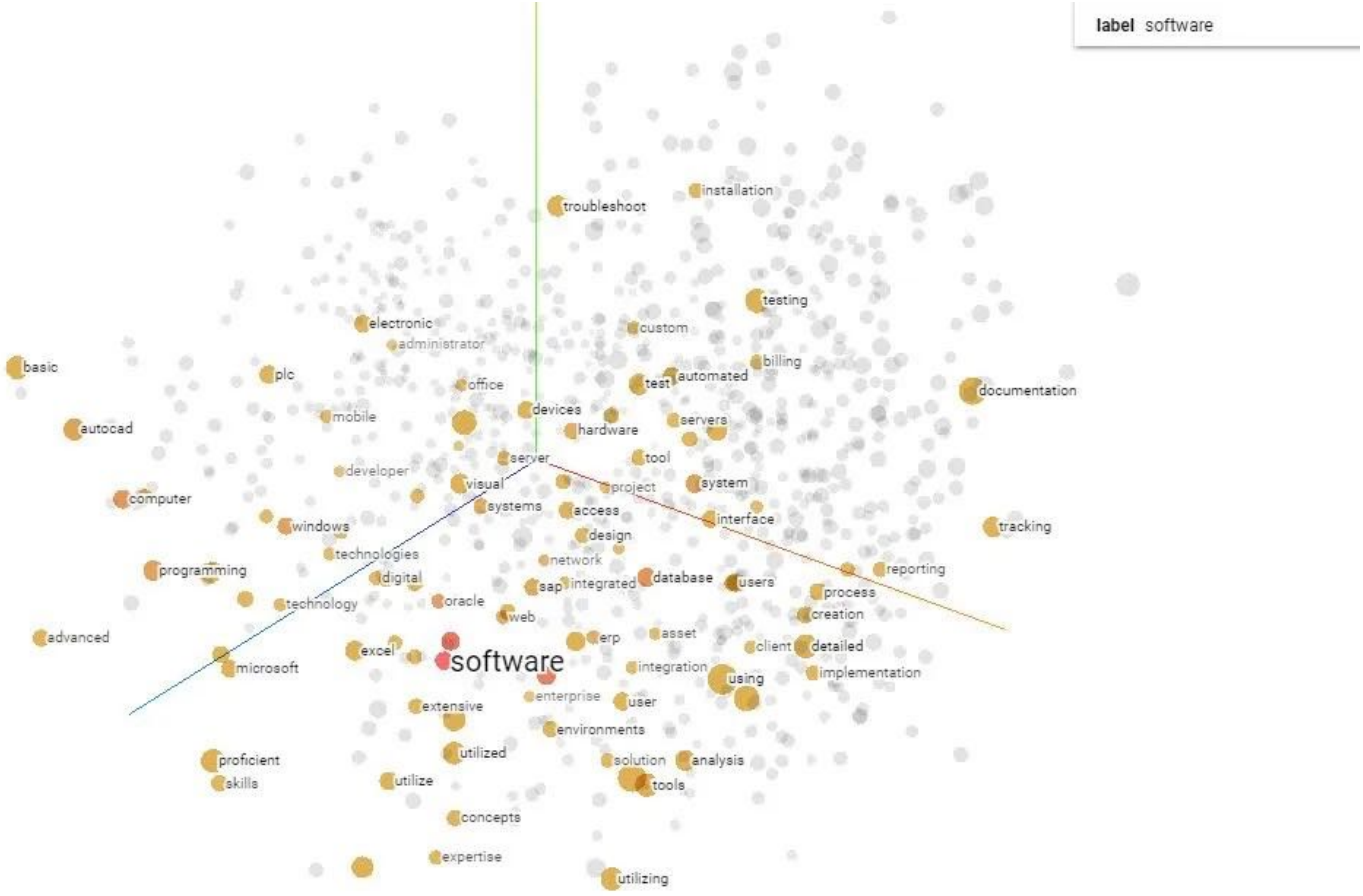
high dimensional vectors



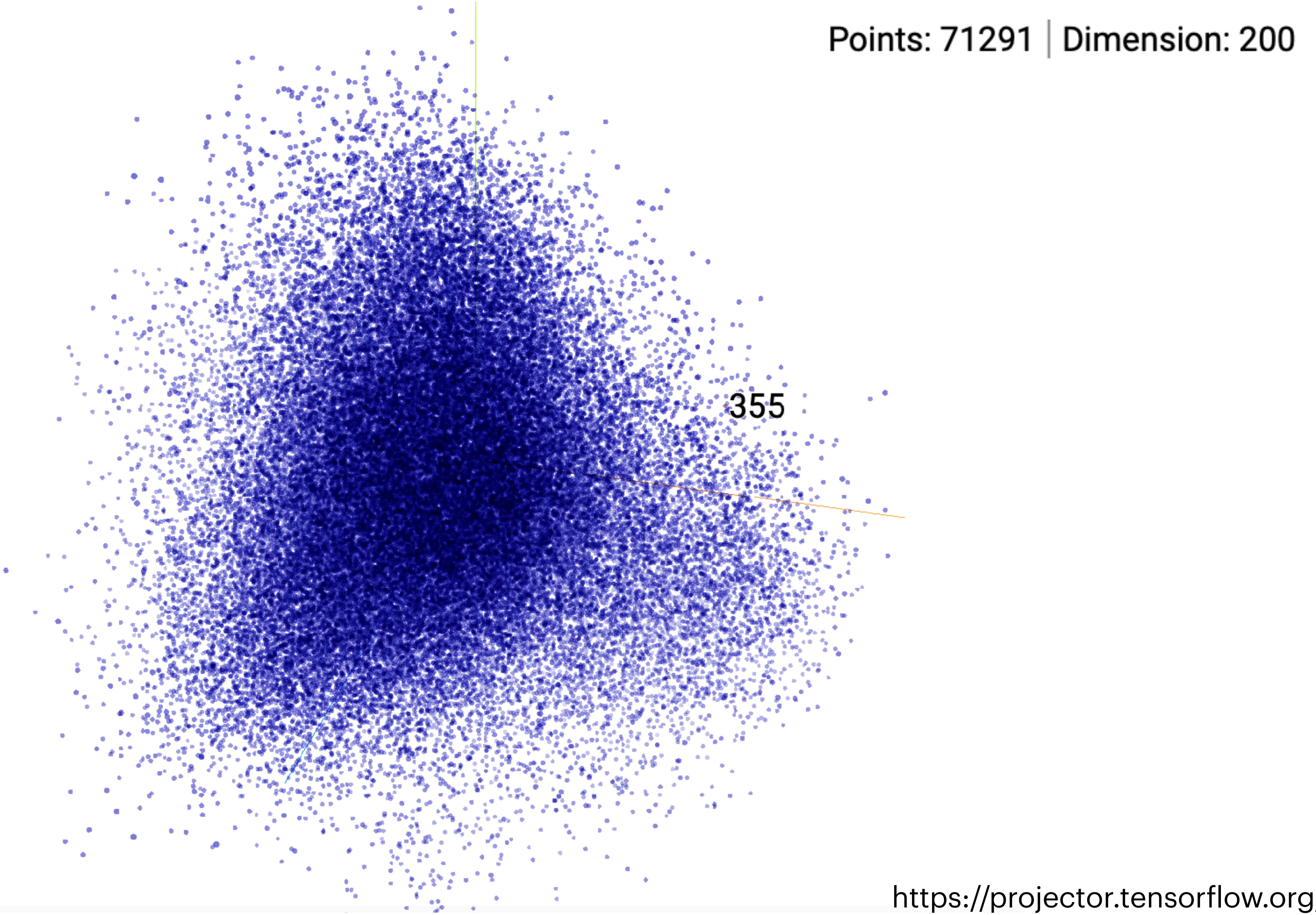
https://github.com/danielvarga/word2vec-web-visualization



https://github.com/danielvarga/word2vec-web-visualization



https://medium.com/@aakashchotrani/visualizing-your-own-word-embeddings-using-tensorflow-688b3a7750ee



https://projector.tensorflow.org/

Embeddings as models of the world

Cornell University

We gratefully acknowled... member i...

arXiv > cs > arXiv:1711.00043

Search...

Help | Advanc...

Computer Science > Computation and Language

[Submitted on 31 Oct 2017 (v1), last revised 13 Apr 2018 (this version, v2)]

Unsupervised Machine Translation Using Monolingual Corpora Only

Guillaume Lample, Alexis Conneau, Ludovic Denoyer, Marc'Aurelio Ranzato

Machine translation has recently achieved impressive performance thanks to recent advances in deep learning and the availability of large-scale parallel corpora. There have been numerous attempts to extend these successes to low-resource language pairs, yet requiring tens of thousands of parallel sentences. In this work, we take this research direction to the extreme and investigate whether it is possible to learn to translate even without any parallel data. We propose a model that takes sentences from monolingual corpora in two different languages and maps them into the same latent space. By learning to reconstruct in both languages from this shared feature space, the model effectively learns to translate without using any labeled data. We demonstrate our model on two widely used datasets and two language pairs, reporting BLEU scores of 32.8 and 15.1 on the Multi30k and WMT English-French datasets, without using even a single parallel sentence at training time.

Comments: ICLR 2018
Subjects: Computation and Language (cs.CL); Artificial Intelligence (cs.AI)
Cite as: arXiv:1711.00043 [cs.CL]
(or arXiv:1711.00043v2 [cs.CL] for this version)
https://doi.org/10.48550/arXiv.1711.00043

POSTED ON AUGUST 31, 2018 TO AI RESEARCH

Unsupervised machine translation: A novel approach to provide fast, accurate translations for more languages

By Myle Ott, Marc'Aurelio Ranzato, Guillaume Lample

Automatic language translation is important to Facebook as a way to allow the billions of people who use our services to connect and communicate in their preferred language. To do this well, current machine translation (MT)

The "it" in AI models is the dataset.

Posted on June 10, 2023 by jbetker

I've been at OpenAI for almost a year now. In that time, I've trained a **lot** of generative models. More than anyone really has any right to train. As I've spent these hours observing the effects of tweaking various model configurations and hyperparameters, one thing that has struck me is the similarities in between all the training runs.

It's becoming awfully clear to me that these models are truly approximating their datasets to an incredible degree. What that means is not only that they learn what it means to be a dog or a cat, but the interstitial frequencies between distributions that don't matter, like what photos humans are likely to take or words humans commonly write down.

What this manifests as is – trained on the same dataset for long enough, pretty much every model with enough weights and training time converges to the same point. Sufficiently large diffusion conv-unets produce the same images as ViT generators. AR sampling produces the same images as diffusion.

This is a surprising observation! It implies that model behavior is not determined by architecture, hyperparameters, or optimizer choices. It's determined by your dataset, nothing else. Everything else is a means to an end in efficiently delivery compute to approximating that dataset.

Then, when you refer to "Lambda", "ChatGPT", "Bard", or "Claude" then, it's not the model weights that you are referring to. It's the dataset.

https://twitter.com/rasbt/status/1719387776729375058

American Sociological Review

Impact Factor: 9.1
5-Year Impact Factor: 13.7

JOURNAL HO

Available access | Research article | First published online September 25, 2019

The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddir

Austin C. Kozlowski, Matt Taddy, and James A. Evans View all authors and affiliations

Volume 84, Issue 5 | https://doi.org/10.1177/0003122419877135 | View correction

Contents | PDF / ePub

Abstract

We argue word embedding models are a useful tool for the study of culture using a historical analysis of shared understandings of social class as an empirical case. Word embeddings represent semantic relations between words as relationships between vectors in a high-dimensional space, specifying a relational model of meaning consistent with contemporary theories of culture. Dimensions induced by word differences (*rich – poor*) in these spaces correspond to dimensions of cultural meaning, and the projection of words onto these dimensions reflects widely shared associations, which we validate with surveys. Analyzing text from millions of books published over 100 years, we show that the markers of class continuously shifted amidst the economic transformations of the twentieth century, yet the basic cultural dimensions of class remained remarkably stable. The notable exception is education, which became tightly linked to affluence independent of its association with cultivated taste.

Structural Similarities Between Language Models and Neural Response Measurements

PDF

Jiaang Li, Antonia Karamolegkou, Yova Kementchedjhieva, Mostafa Abdou, Sune Lehmann, Anders Søgaard

Published: 29 Nov 2023, Last Modified: 29 Nov 2023 NeurReps 2023 Poster Everyone Revisions
BibTeX

Submission Track: Proceedings
Keywords: fMRI, large language models, neural networks, representation, understanding
TL;DR: The larger neural language models get, the more their representations are structurally similar to neural response measurements from brain imaging.

Abstract:
Large language models have complicated internal dynamics, but induce representations of words and phrases whose geometry we can study. Human language processing is also opaque, but neural response measurements can provide (noisy) recordings of activations during listening or reading, from which we can extract similar representations of words and phrases. Here we study the extent to which the geometries induced by these representations, share similarities in the context of brain decoding. We find that the larger neural language models get, the more their representations are structurally similar to neural response measurements from brain imaging.

Submission Number: 15

Language Models and Neural Response Measurements

Jiaang Li, Antonia Karamolegkou, Yova Kementchedjhieva, Mostafa Abdou, Sune Lehmann, Anders Søgaard

Published: 29 Nov 2023, Last Modified: 29 Nov 2023 NeurReps 2023 Poster Everyone Revisions BibTeX

Submission Track: Proceedings

Keywords: fMRI, large language models, neural networks, representation, understanding

TL;DR: The larger neural language models get, the more their representations are structurally similar to neural response measurements from brain imaging.

Abstract:

Large language models have complicated internal dynamics, but induce representations of words and phrases whose geometr we can study. Human language processing is also opaque, but neural response measurements can provide (noisy) recordings o activations during listening or reading, from which we can extract similar representations of words and phrases. Here we stud the extent to which the geometries induced by these representations, share similarities in the context of brain decoding. We find that the larger neural language models get, the more their representations are structurally similar to neural response measurements from brain imaging.

Submission Number: 15

The Platonic Representation Hypothesis

Minyoung Huh [*1] **Brian Cheung** [*1] **Tongzhou Wang** [*1] **Phillip Isola** [*1]

Abstract

We argue that representations in AI models, particularly deep networks, are converging. First, we survey many examples of convergence in the literature: over time and across multiple domains, the ways by which different neural networks represent data are becoming more aligned. Next, we demonstrate convergence across data modalities: as vision models and language models get larger, they measure distance between datapoints in a more and more alike way. We hypothesize that this convergence is driving toward a shared statistical model of reality, akin to Plato's concept of an ideal reality. We term such a representation the *platonic representation* and discuss several possible selective pressures toward it. Finally, we discuss the implications of these trends, their limitations, and counterexamples to our analysis.

Project Page: phillipi.github.io/prh

Code: github.com/minyoungg/platonic-rep

1. Introduction

The Platonic Representation Hypothesis

Neural networks, trained with different objectives on different data and modalities, are converging to a shared statistical model of reality in their representation spaces.

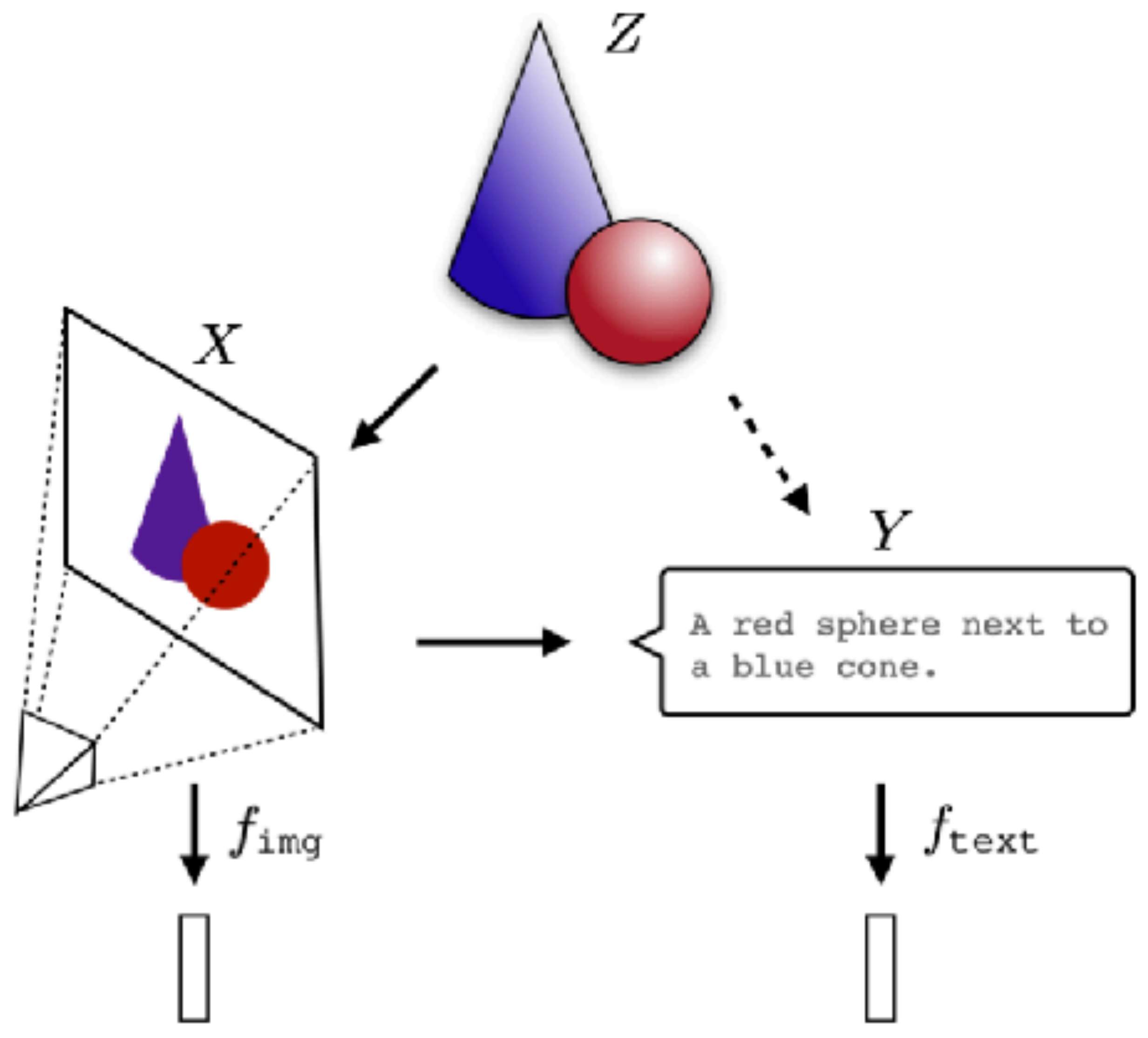
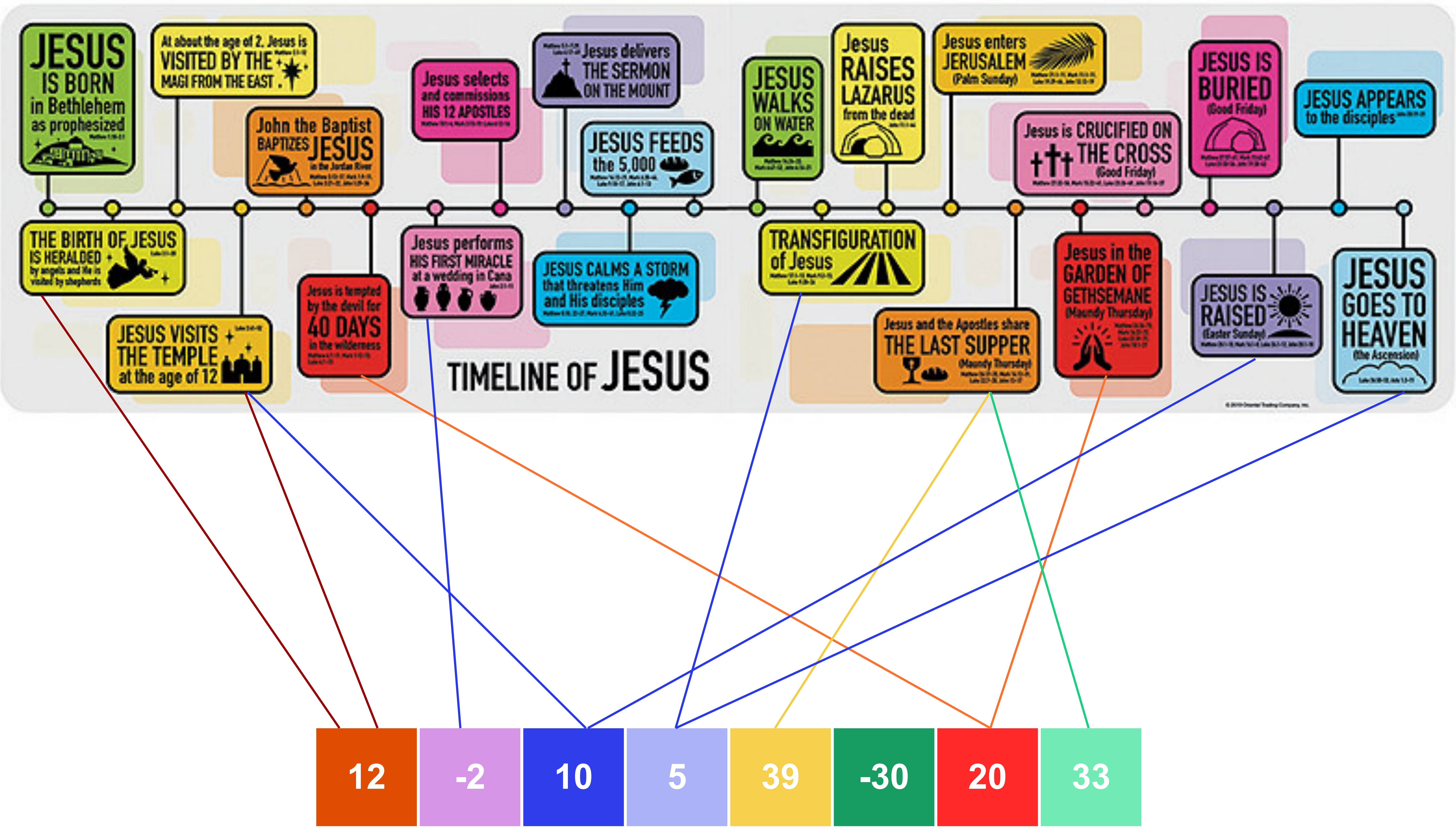


Figure 1. **The Platonic Representation Hypothesis:** Images (X)

Draw on the structural similarity between natural language sentences human lives

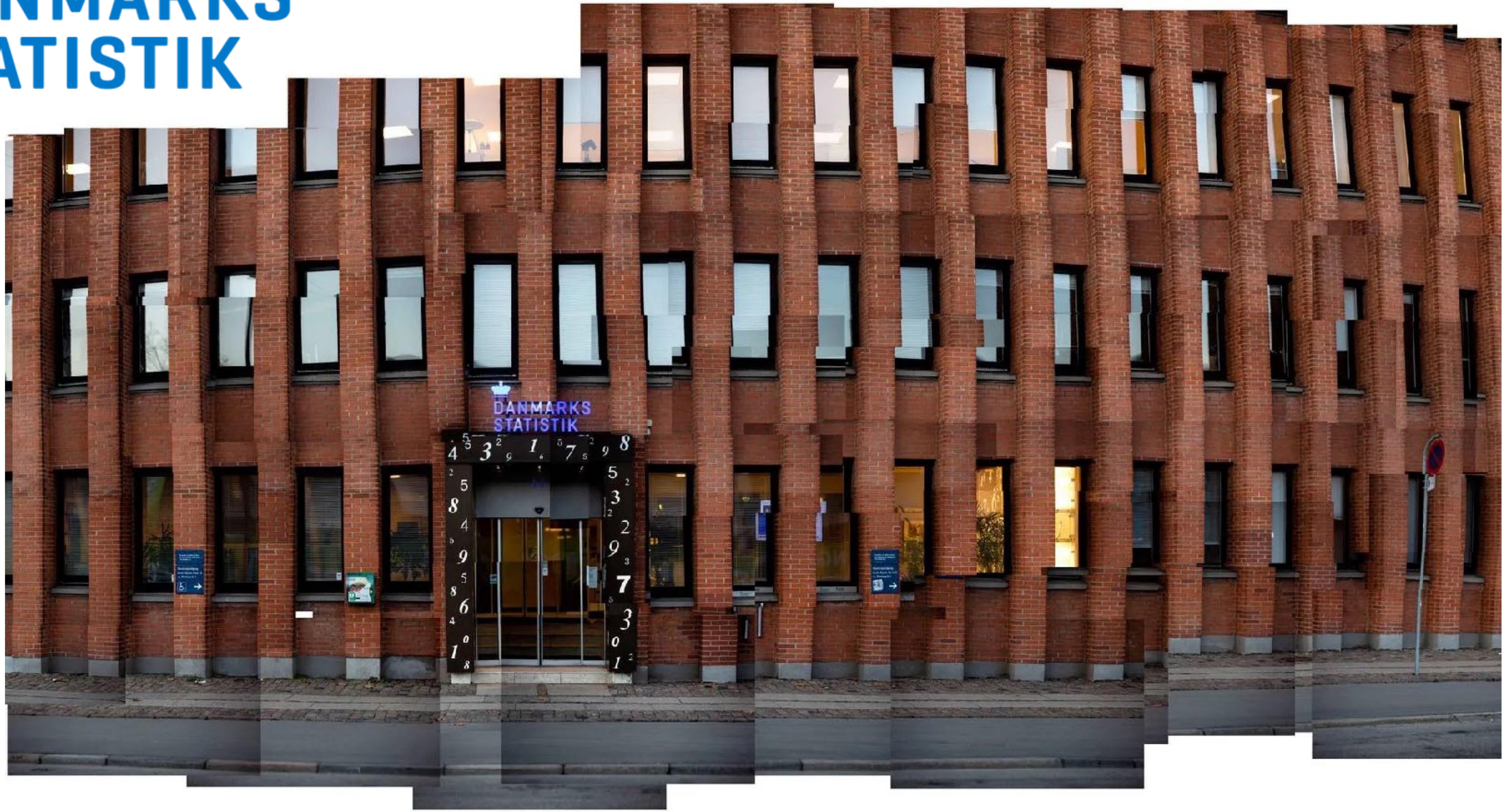
Language	Lives
↓	↓
Words	Life-events

Draw on the structural similarity between natural language sentences human lives

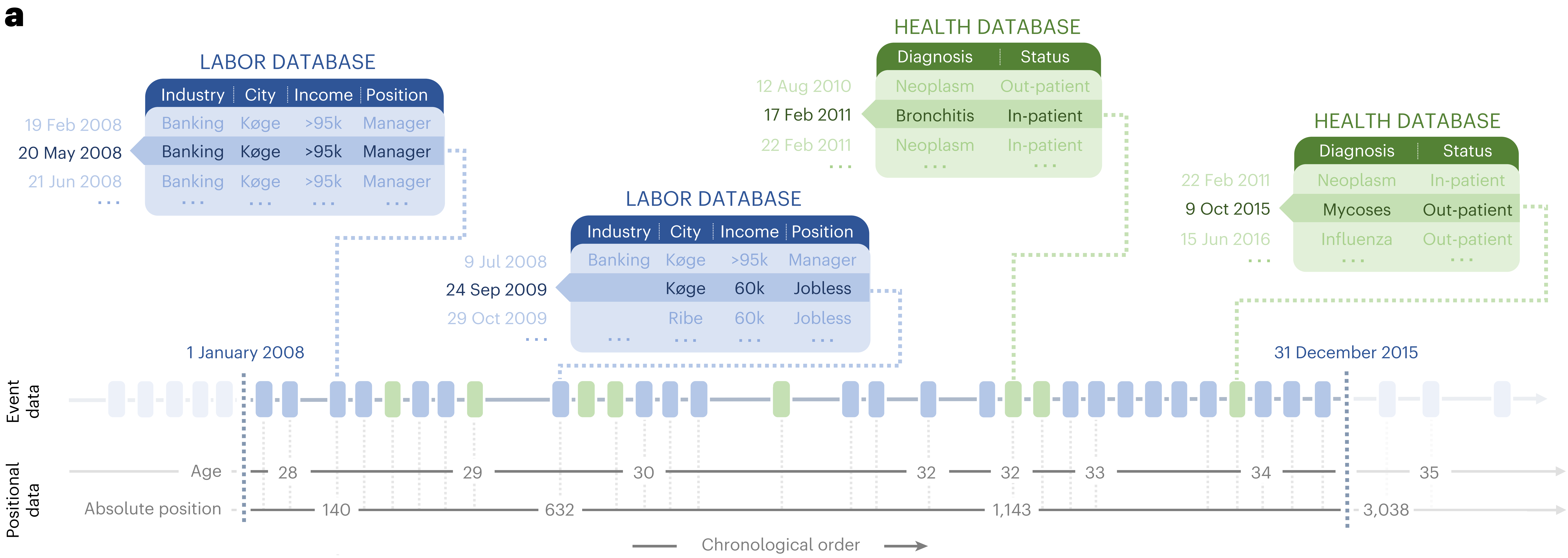


(Compact **Representation** of Life Trajectory)

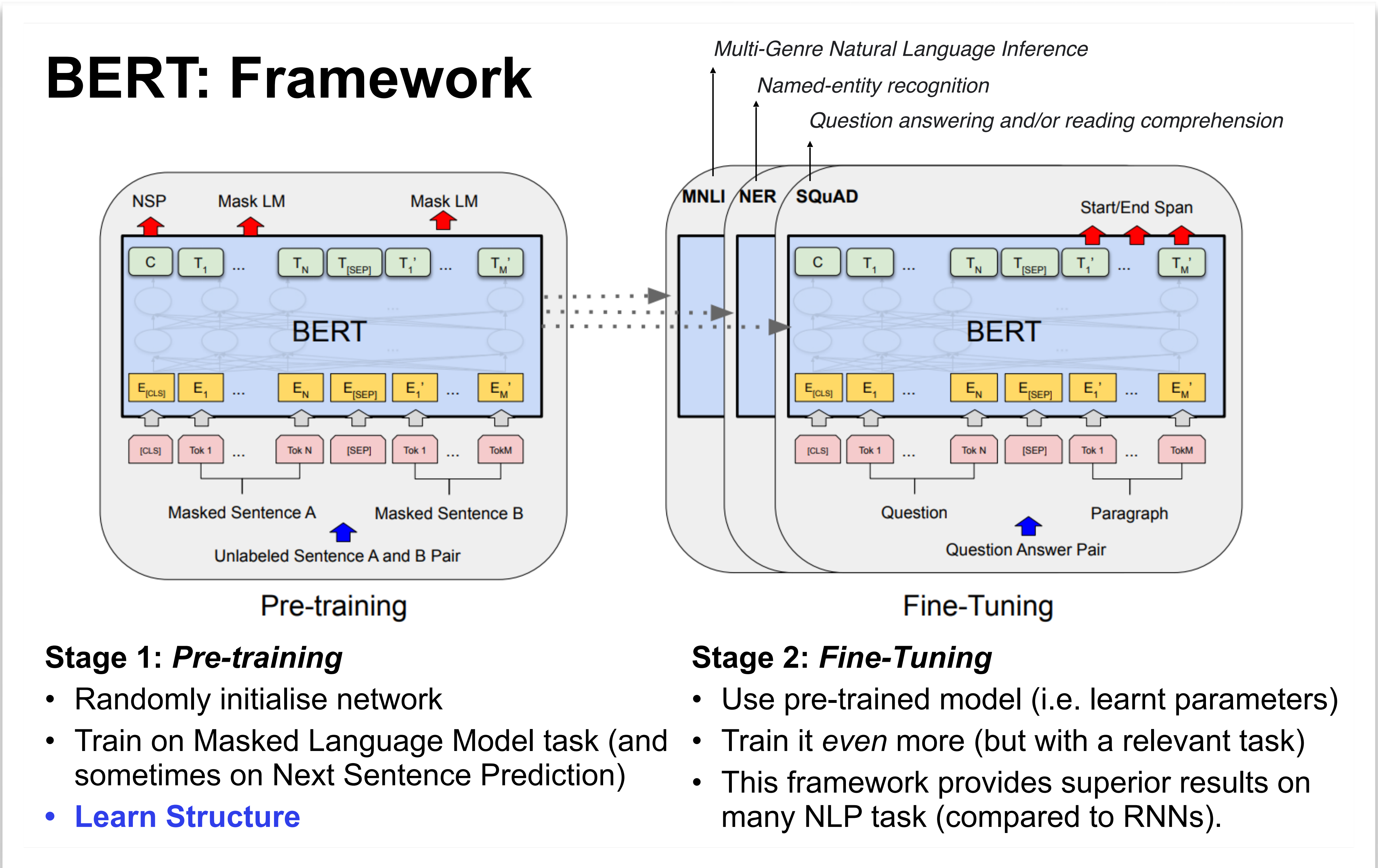
DANMARKS
STATISTIK



So what does the actual data look like?



No fancy new methods but draw on a stack of existing tech



Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.

Choromanski, K., Likhosherstov, V., Dohan, D., Song, X., Gane, A., Sarlos, T., Hawkins, P., Davis, J., Mohiuddin, A., Kaiser, L. and Belanger, D., 2020. Rethinking attention with performers. *arXiv preprint arXiv:2009.14794*.

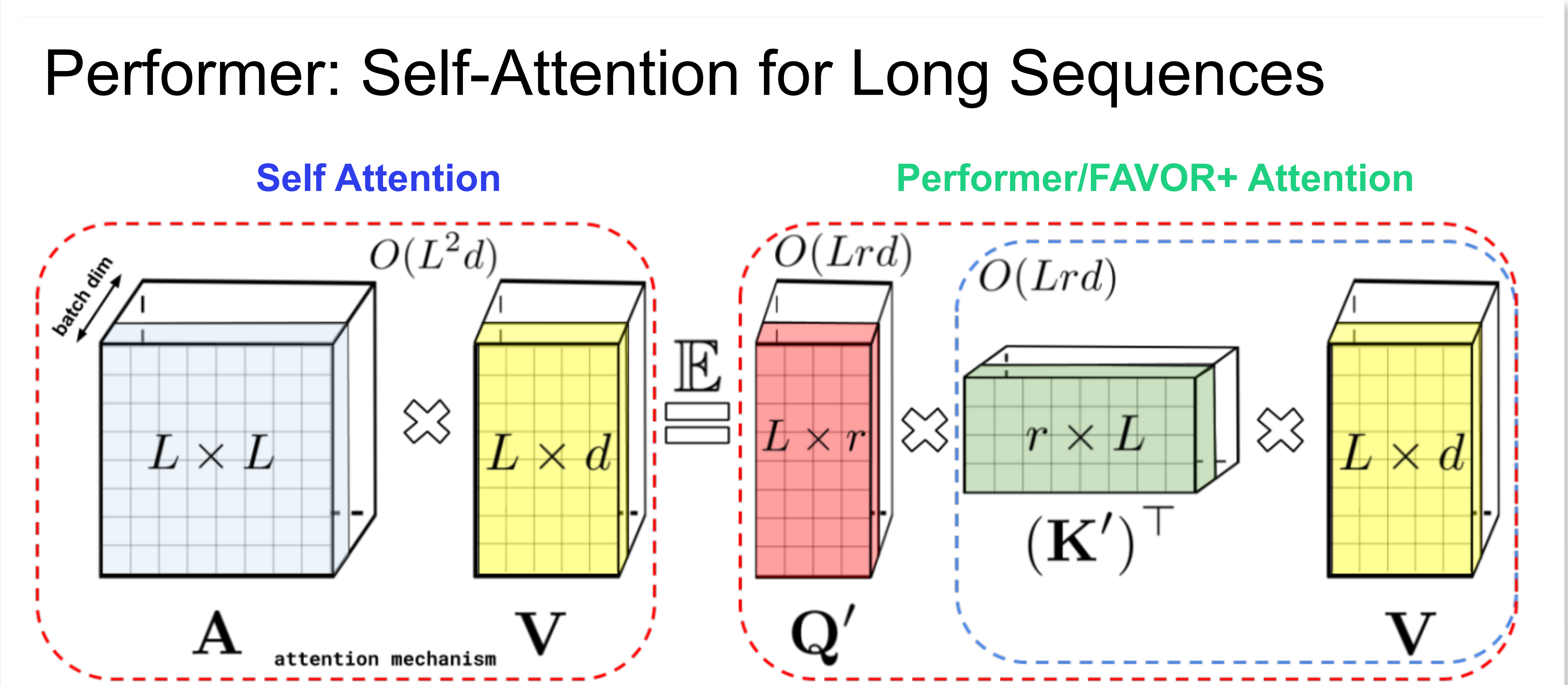


Figure 1: Approximation of the regular attention mechanism $\mathbf{AV}$ (before $\mathbf{D}^{-1}$-renormalization) via (random) feature maps. Dashed-blocks indicate order of computation with corresponding time complexities attached.

Two words about how training transformers work

Stage 1: Pretraining

(1) a Masked Language Modeling (MLM) task that forces the model to use token representations and contextual information

(2) a Sequence Ordering Prediction (SOP) task that focuses on the temporal coherence of the sequence

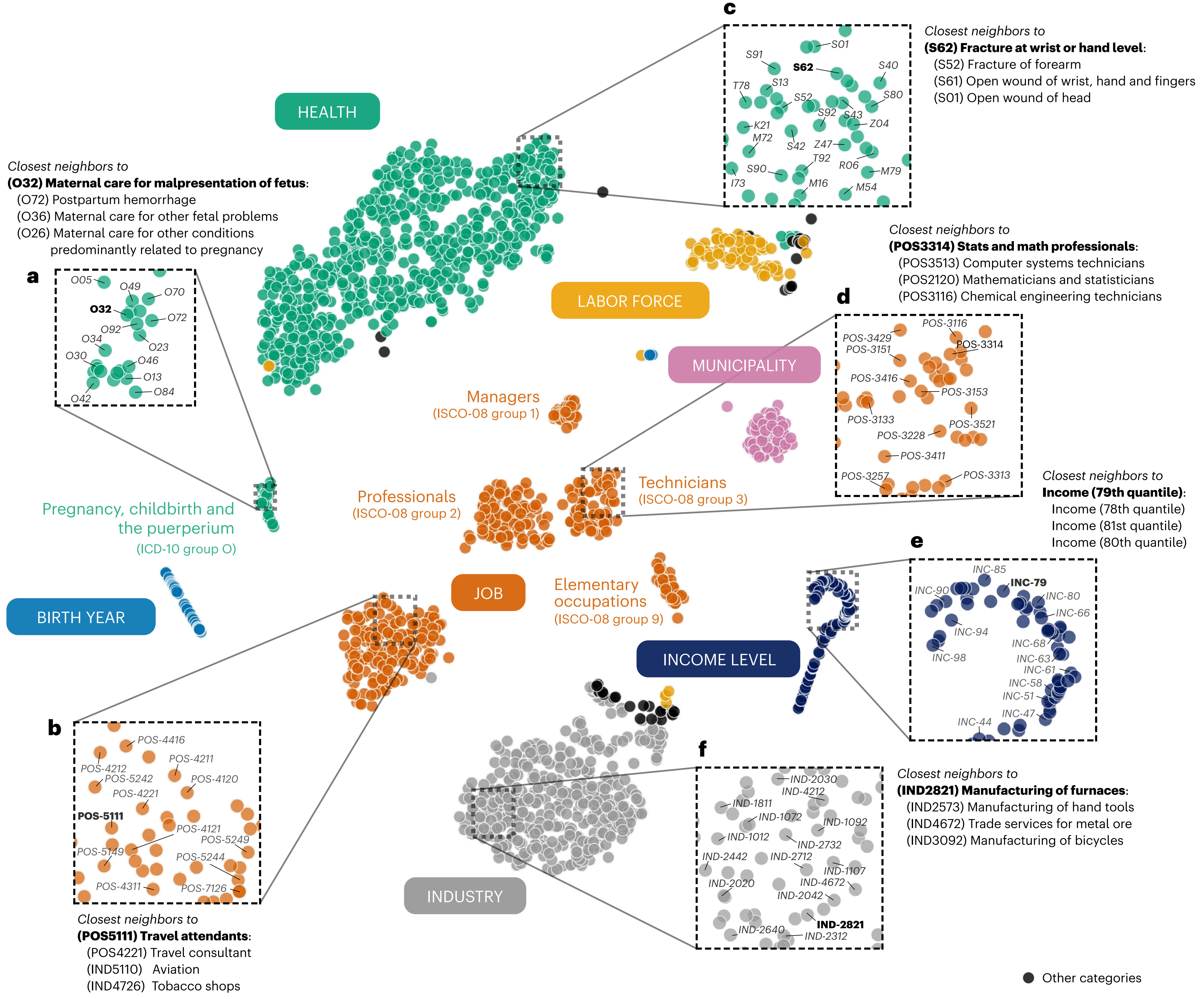
THIS IS WHAT WE DISCUSSED EARLIER

Stage 2: Classification

Event embedding space

The space starts out with positions of events randomized and converges **robustly** to what you see.

The colors are not added by the model, but come from the data.



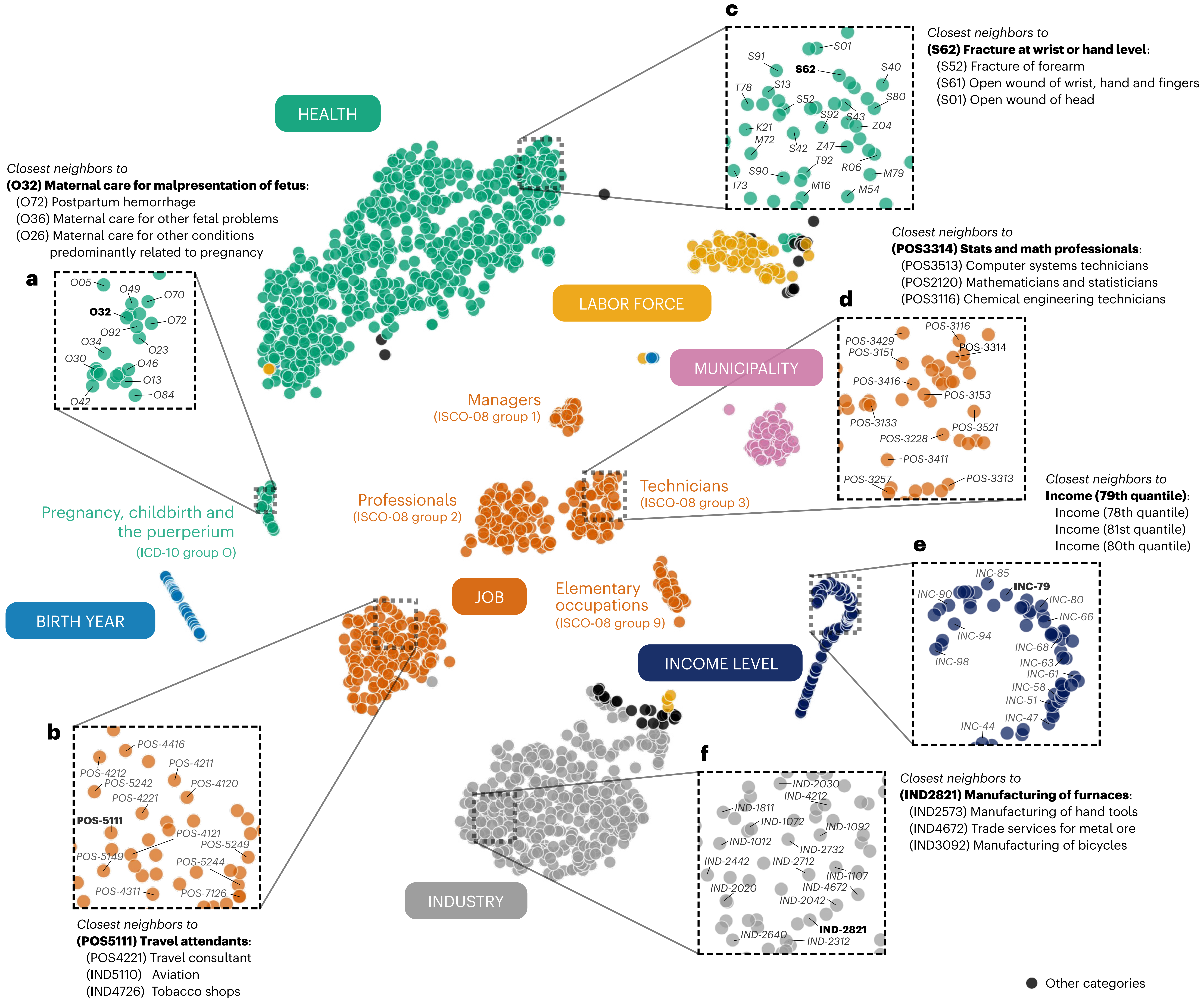
Pregnancy, childbirth and the puerperium
(ICD-10 group O)
Professio
(ISCO-08 gro
BIRTH YEAR
b
POS-4416


POS-3411
POS-3257
POS-3313
Technicians
(ISCO-08 group 3)
Closest neighbors to
Income (79th quantile):
Income (78th quantile)
Income (81st quantile)
Income (80th quantile)
e
INC-85
INC-79
INC-90
INC-80
INC-66
INC-94
INC-68
INC-98
INC-63
INC-61
INC-58
INC-51
INC-47
INC-44
INCOME LEVEL
f

HEALTH
LABOR FORCE
MUNICIPALITY
JOB
c
S01
S91
S62
S40
T78
S13
S80
S52
S92
S43
K21
Z04
M72
S42
Z47
T92
R06
S90
M79
I73
M16
M54
Closest neighbors to
(S62) Fracture at wrist or h
(S52) Fracture of forearm
(S61) Open wound of wris
(S01) Open wound of head
Closest neighbors to
(O32) Maternal care for malpresentation of fetus:
(O72) Postpartum hemorrhage
(O36) Maternal care for other fetal problems
(O26) Maternal care for other conditions predominantly related to pregnancy
a
O05
O49
O70
O32
O72
O92
O34
O23
O30
O46
O13
O42
O84
Pregnancy, childbirth and the puerperium
(ICD-10 group O)
Closest neighbors to
(POS3314) Stats and math profes
(POS3513) Computer systems te
(POS2120) Mathematicians and
(POS3116) Chemical engineerin
d
POS-3116
POS-3429
POS-3314
POS-3151
POS-3416
POS-3153
POS-3133
POS-3521
POS-3228
POS-3411
POS-3257
POS-3313
Managers
(ISCO-08 group 1)
Professionals
(ISCO-08 group 2)
Technicians
(ISCO-08 group 3)
Elementary
Closest
Income
Incom
Incom
Incom
e
INC-85
INC-79
INC-90


INCOME LEVEL
INC-94
INC-98
INC-68
INC-63
INC-61
INC-58
INC-51
INC-47
INC-44
f
IND-2030
IND-4212
IND-1811
IND-1072
IND-1092
IND-1012
IND-2732
IND-2442
IND-2712
IND-1107
IND-2020
IND-4672
IND-2042
IND-2821
IND-2640
IND-2312
INDUSTRY
Closest neighbors to
(IND2821) Manufacturing of furnaces:
(IND2573) Manufacturing of hand tools
(IND4672) Trade services for metal ore
(IND3092) Manufacturing of bicycles
Other categories

LABOR FORCE
MUNICIPALITY
Managers
(ISCO-08 group 1)
Technicians
(ISCO-08 group 3)
JOB
Elementary
occupations
(ISCO-08 group 9)
Closest neighbors to
(POS3314) Stats and math professionals:
(POS3513) Computer systems technicians
(POS2120) Mathematicians and statisticians
(POS3116) Chemical engineering technicians
d
POS-3116
POS-3429
POS-3314
POS-3151
POS-3416
POS-3153
POS-3133
POS-3521
POS-3228
POS-3411
POS-3257
POS-3313
Closest neighbors to
Income (79th quantile):
Income (78th quantile)
Income (81st quantile)
Income (80th quantile)
e
INC-85
INC-79
INC-90
INC-80
INC-66
INC-94
INC-68



HEALTH
LABOR FORCE
MUNICIPALITY
BIRTH YEAR
JOB
INCOME LEVEL
INDUSTRY
Pregnancy, childbirth and the puerperium (ICD-10 group O)
Managers (ISCO-08 group 1)
Professionals (ISCO-08 group 2)
Technicians (ISCO-08 group 3)
Elementary occupations (ISCO-08 group 9)
a
Closest neighbors to
(O32) Maternal care for malpresentation of fetus:
(O72) Postpartum hemorrhage
(O36) Maternal care for other fetal problems
(O26) Maternal care for other conditions predominantly related to pregnancy
O05 O49 O70 O32 O72 O92 O34 O23 O30 O46 O13 O84 O42
b
Closest neighbors to
(POS5111) Travel attendants:
(POS4221) Travel consultant
(IND5110) Aviation
(IND4726) Tobacco shops
POS-4416 POS-4212 POS-4211 POS-5242 POS-4120 POS-4221 POS-5111 POS-4121 POS-5249 POS-5149 POS-5244 POS-4311 POS-7126
c
Closest neighbors to
(S62) Fracture at wrist or hand level:
(S52) Fracture of forearm
(S61) Open wound of wrist, hand and fingers
(S01) Open wound of head
S01 S91 S62 S40 T78 S13 S52 S80 S92 S43 Z04 K21 M72 S42 Z47 T92 R06 S90 M79 I73 M16 M54
d
Closest neighbors to
(POS3314) Stats and math professionals:
(POS3513) Computer systems technicians
(POS2120) Mathematicians and statisticians
(POS3116) Chemical engineering technicians
POS-3116 POS-3429 POS-3314 POS-3151 POS-3416 POS-3153 POS-3133 POS-3521 POS-3228 POS-3411 POS-3257 POS-3313
e
Closest neighbors to
Income (79th quantile):
Income (78th quantile)
Income (81st quantile)
Income (80th quantile)
INC-85 INC-90 INC-79 INC-80 INC-66 INC-94 INC-68 INC-98 INC-63 INC-61 INC-58 INC-51 INC-47 INC-44
f
Closest neighbors to
(IND2821) Manufacturing of furnaces:
(IND2573) Manufacturing of hand tools
(IND4672) Trade services for metal ore
(IND3092) Manufacturing of bicycles
IND-2030 IND-4212 IND-1811 IND-1072 IND-1092 IND-1012 IND-2732 IND-2442 IND-2712 IND-1107 IND-2020 IND-4672 IND-2042 IND-2821 IND-2640 IND-2312
Other categories

The life-event embedding space is a kind of “foundation model” for human lives.

Given training data, we can predict any outcome (more on this in a bit)

No need for feature selection, just throw all of your information in there

The life-event embedding space is a kind of “foundation model” for human lives.

Traditional modeling

f(variables that might play a role) = outcome of interest

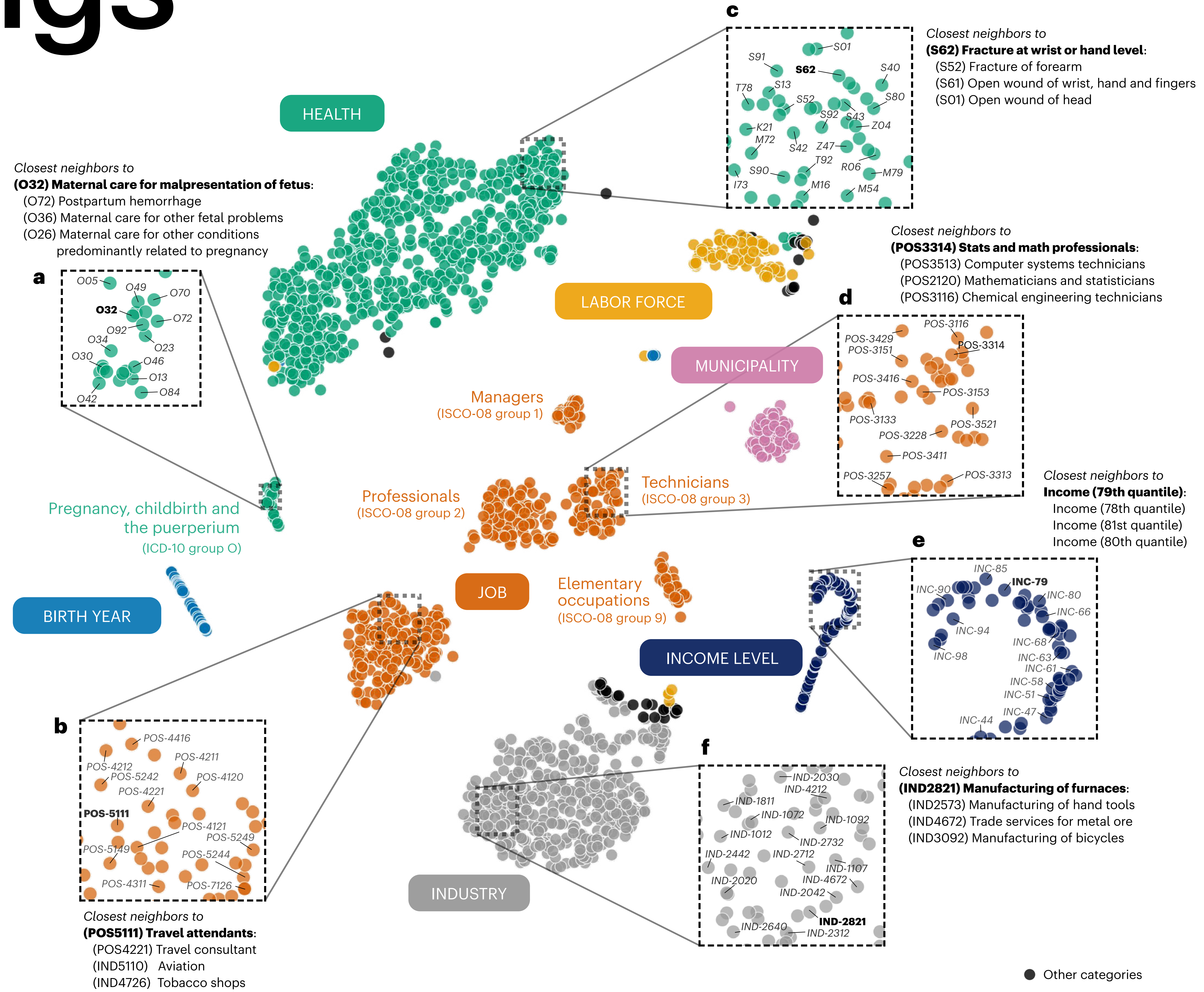
f(age, health information, sociodemographic measures) = death

And here we have to apply “feature engineering”

Connecting back to word-embeddings

FURTHER!

If we view the space as an **interesting model of the world**, we can study it and search for associations, discover biases, etc, etc.



Two words about how training transformers work

Stage 1: Pretraining

(1) a Masked Language Modeling (MLM) task that forces the model to use token representations and contextual information

(2) a Sequence Ordering Prediction (SOP) task that focuses on the temporal coherence of the sequence

Stage 2: Classification (or some other task)

We can make good predictions

Predicting early mortality. We estimate the likelihood of a person surviving the following four years after 1st January 2016. This is an oft-used task within statistical modeling [57]. Further, mortality prediction is closely related to other health-prediction tasks and therefore requires `life2vec` to model the progression of individual health-sequences as well as labor history to predict the right outcome successfully. Specifically, given a sequence representation, `life2vec` infers the likelihood of a person surviving the four years following the end of our sequences (1st January 2016). We focus on making predictions for a young cohort of people consisting of individuals who are 30-55 years old, where mortality is challenging to predict.

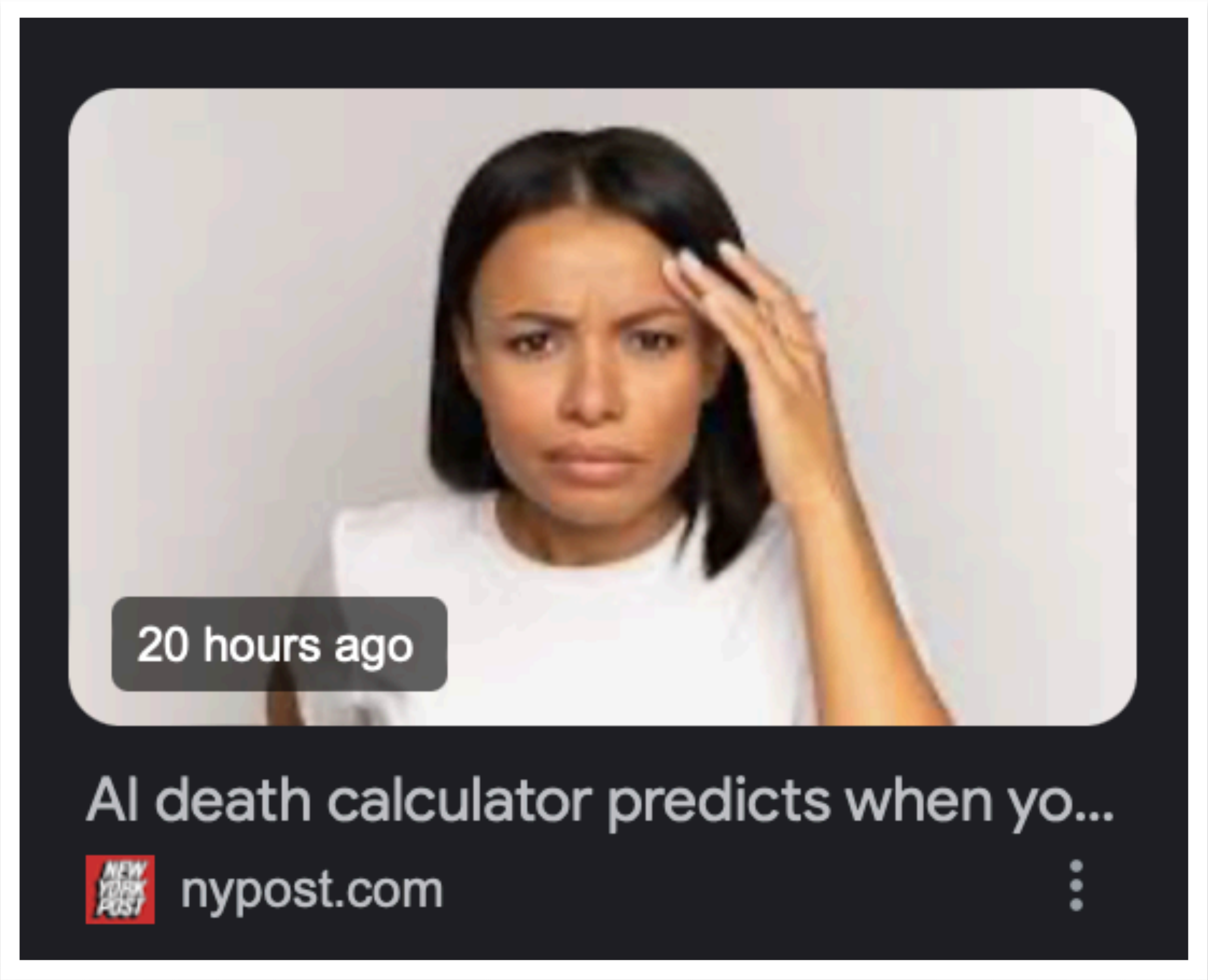
And this is the part that caused all the trouble

We had built a model that can predict anything

But you have to choose something

So we agreed that an exciting outcome would be “early mortality”

(And that was simply much of a combustible cocktail)



We can make good predictions

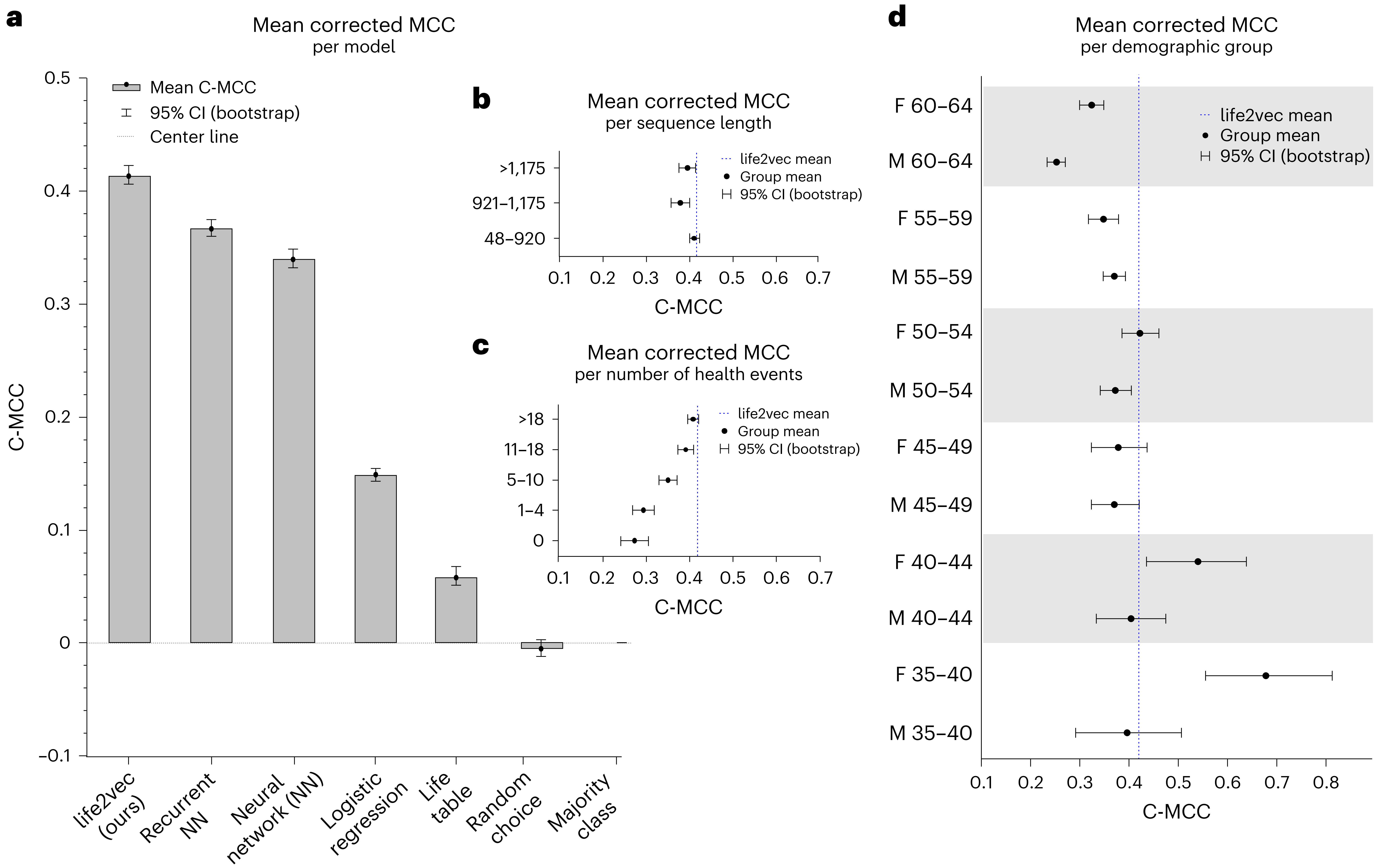
We wanted an "interesting" target, so we consider a cohort of 30-55 year olds. And predict death across a long period of time

We have data from 2008-2020, but train the model only on 2008-2016

We then predict if someone dies in the period 2016-2020.

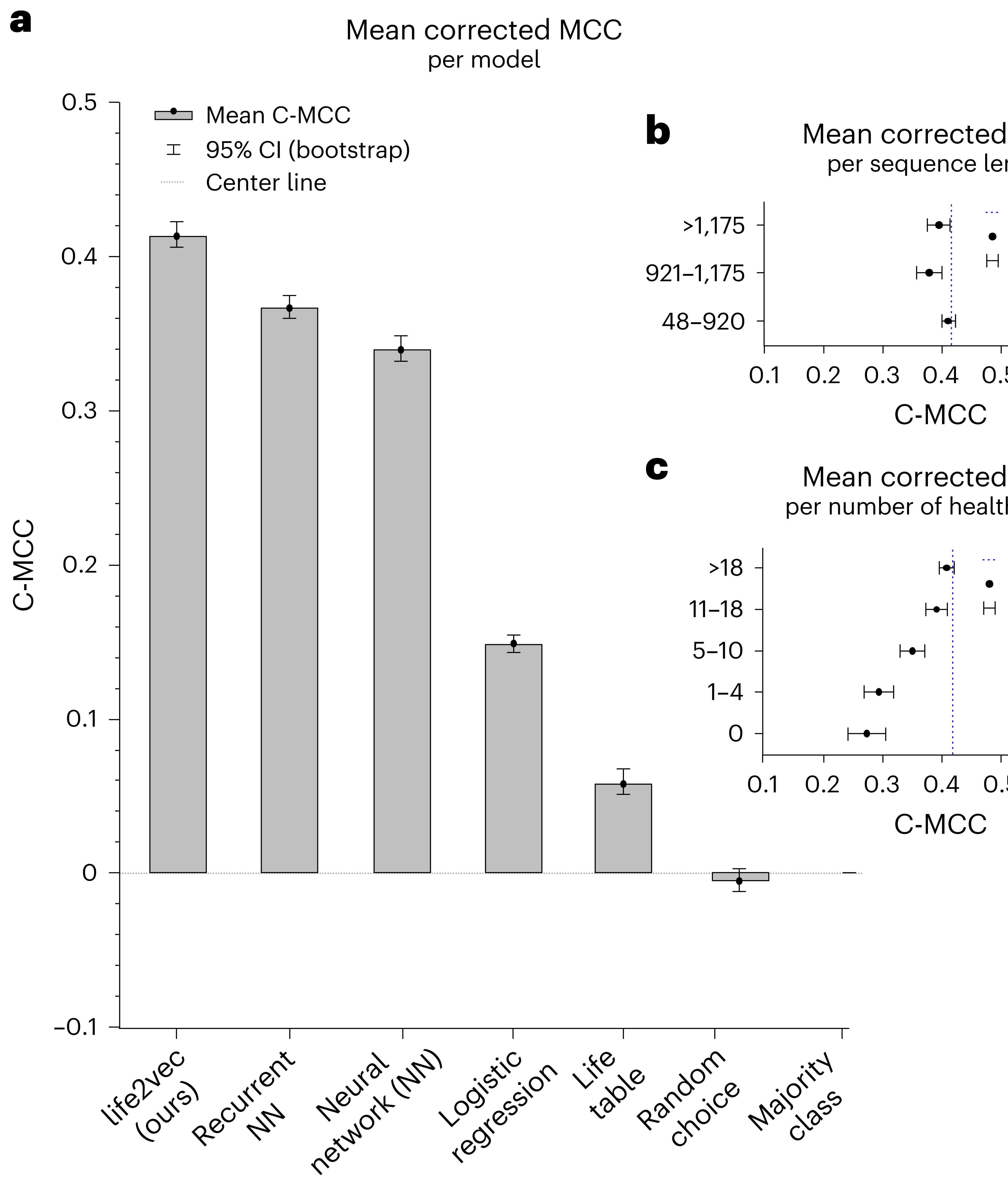
We train on part of this group and then predict death (yes/no) for a balanced group.

We can make good predictions



We can make good predictions

Model	MCC, 95%-CI	AUL	Accuracy, 95%-CI	F1-Score, 95%-CI
L2V	**0.413 [0.410, 0.422]**	**0.845**	0.788 [0.782, 0.794]	0.443 [0.435, 0.451]
RNN-GRU	0.369 [0.361, 0.378]	0.834	0.778 [0.771, 0.783]	0.395 [0.389, 0.402]
FFNN	0.340 [0.332, 0.348]	0.822	0.768 [0.762, 0.774]	0.345 [0.339, 0.350]
Logistic Reg	0.149 [0.142, 0.155]	0.735	0.639 [0.633, 0.645]	0.201 [0.198, 0.204]
Life Tables	0.059 [0.051, 0.066]	0.650	0.555 [0.548, 0.562]	0.161 [0.158, 0.164]
Random	-0.005 [-0.011, 0.002]	0.497	0.496 [0.489, 0.503]	0.132 [0.128, 0.135]
Majority Class	0.0	0.497	0.5	-

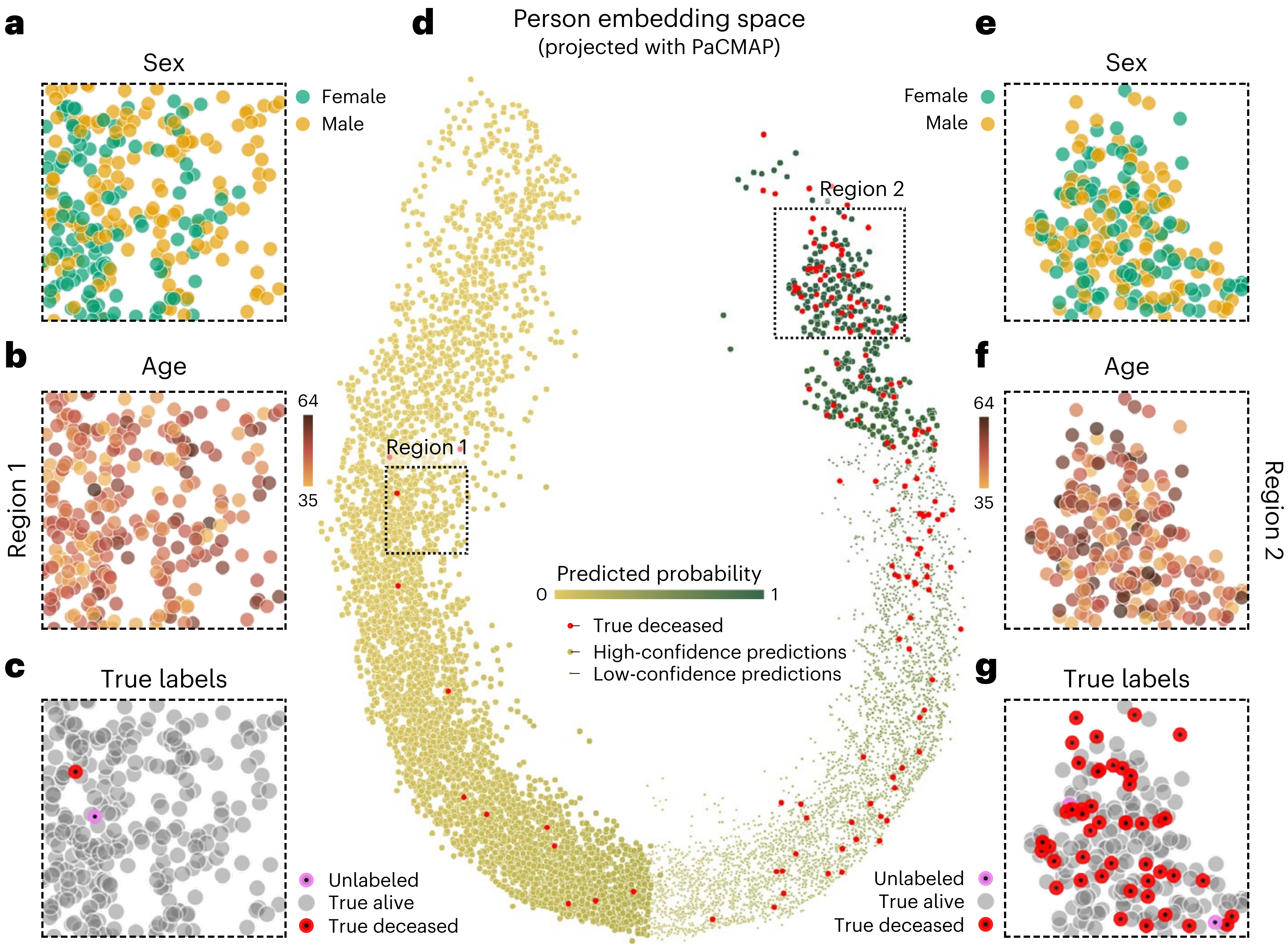


But more interestingly, we can begin to try to understand what the model is doing

Two Embedding Spaces

- Life Events
- Person-embeddings (task-dependent)

CONDITIONAL ON PREDICTION



a
Sex
Female
Male
b
Age
64
35
Region 1
c
True labels
Unlabeled
True alive
True deceased
d
Person embedding space
(projected with PaCMAP)
Region 1
Region 2
Predicted probability
0
1
True deceased
High-confidence predictions
Low-confidence predictions
e
Sex
Female
Male
f
Age
64
35
Region 2
g
True labels
Unlabeled
True alive
True deceased

Person embedding space

Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)

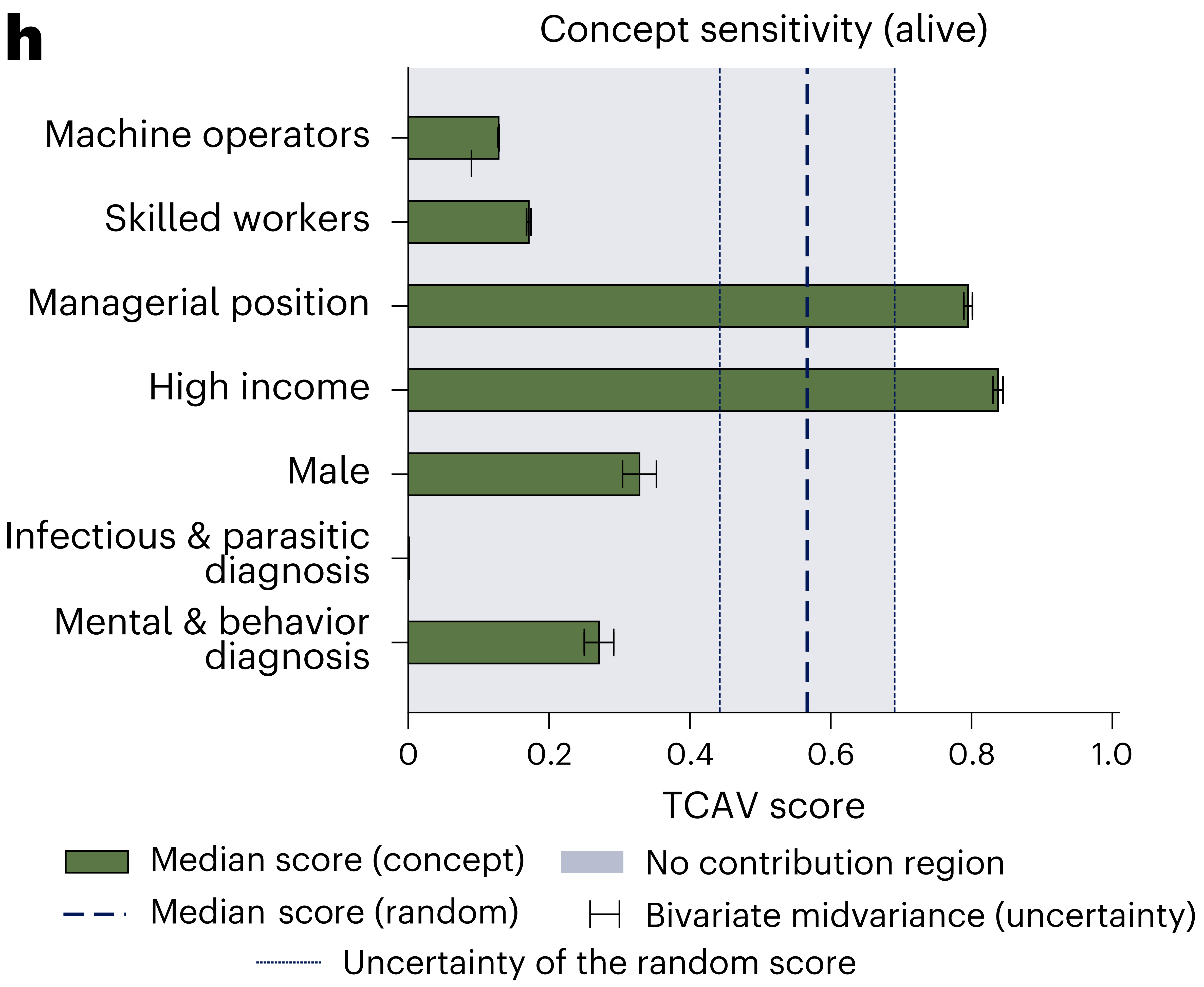
Been Kim Martin Wattenberg Justin Gilmer Carrie Cai James Wexler Fernanda Viegas Rory Sayres

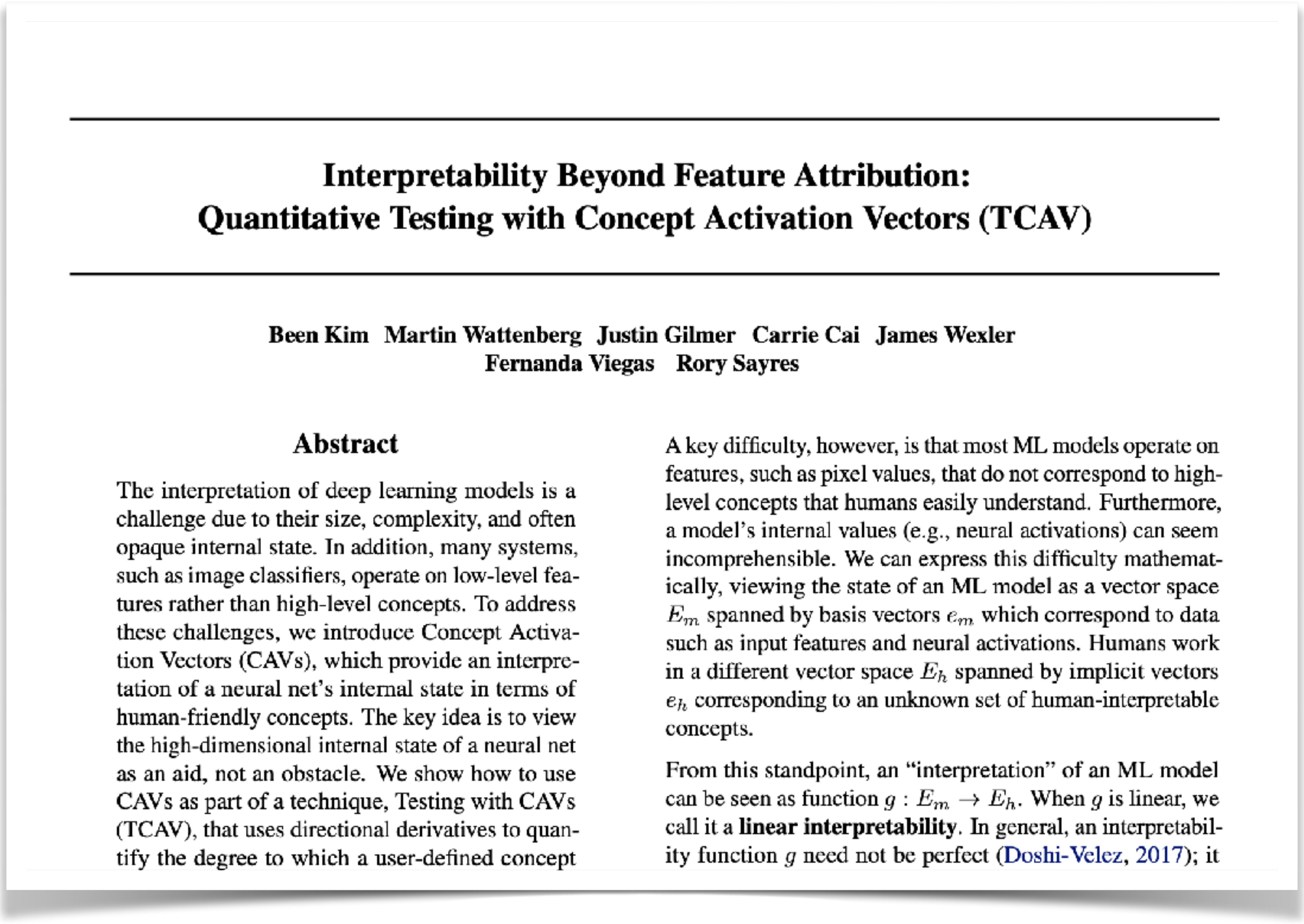
Abstract

The interpretation of deep learning models is a challenge due to their size, complexity, and often opaque internal state. In addition, many systems, such as image classifiers, operate on low-level features rather than high-level concepts. To address these challenges, we introduce Concept Activation Vectors (CAVs), which provide an interpretation of a neural net's internal state in terms of human-friendly concepts. The key idea is to view the high-dimensional internal state of a neural net as an aid, not an obstacle. We show how to use CAVs as part of a technique, Testing with CAVs (TCAV), that uses directional derivatives to quantify the degree to which a user-defined concept

A key difficulty, however, is that most ML models operate on features, such as pixel values, that do not correspond to high-level concepts that humans easily understand. Furthermore, a model's internal values (e.g., neural activations) can seem incomprehensible. We can express this difficulty mathematically, viewing the state of an ML model as a vector space E_m spanned by basis vectors e_m which correspond to data such as input features and neural activations. Humans work in a different vector space E_h spanned by implicit vectors e_h corresponding to an unknown set of human-interpretable concepts.

From this standpoint, an "interpretation" of an ML model can be seen as function $g : E_m \to E_h$. When g is linear, we call it a **linear interpretability**. In general, an interpretability function g need not be perfect (Doshi-Velez, 2017); it





Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)

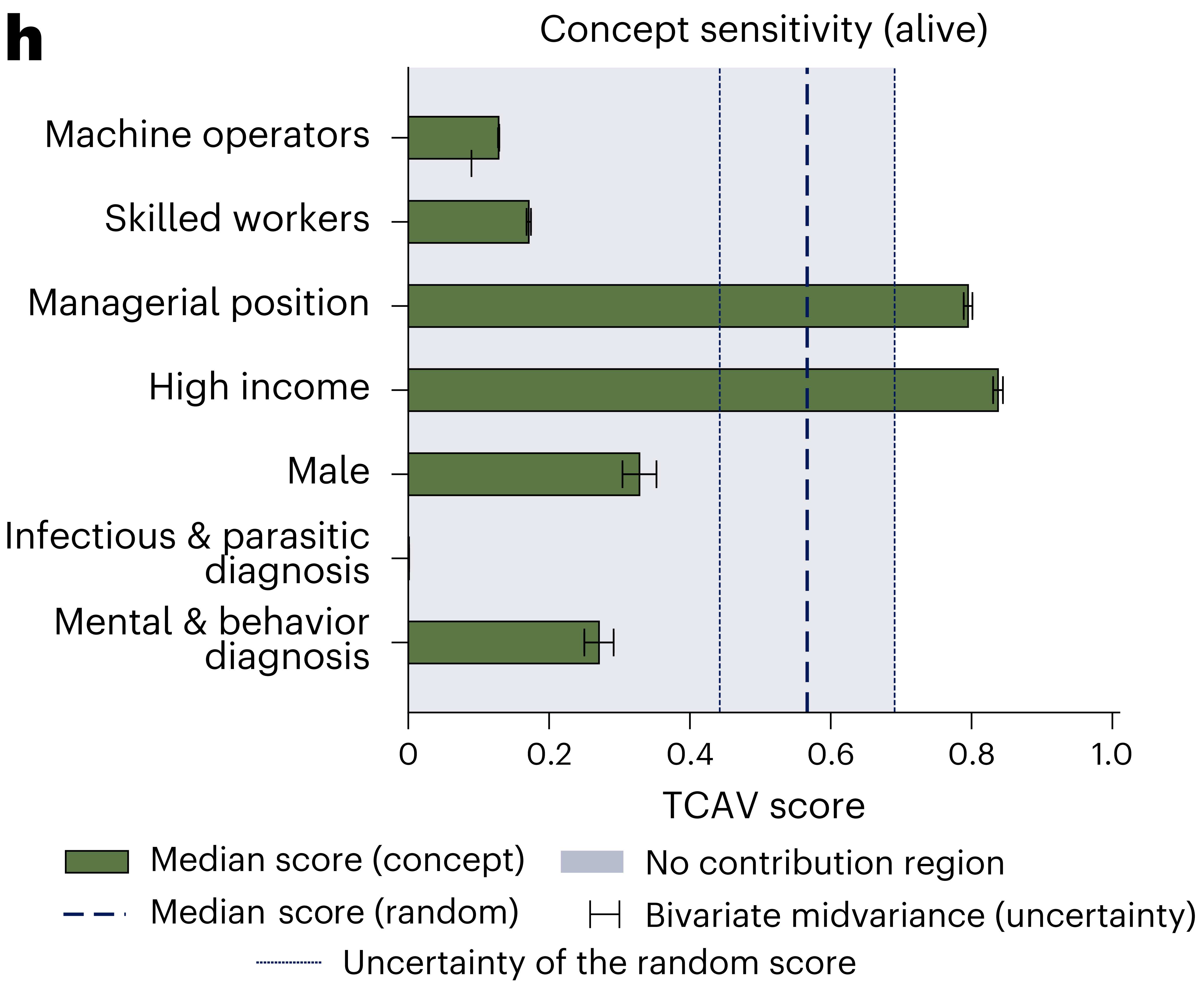
Been Kim Martin Wattenberg Justin Gilmer Carrie Cai James Wexler Fernanda Viegas Rory Sayres

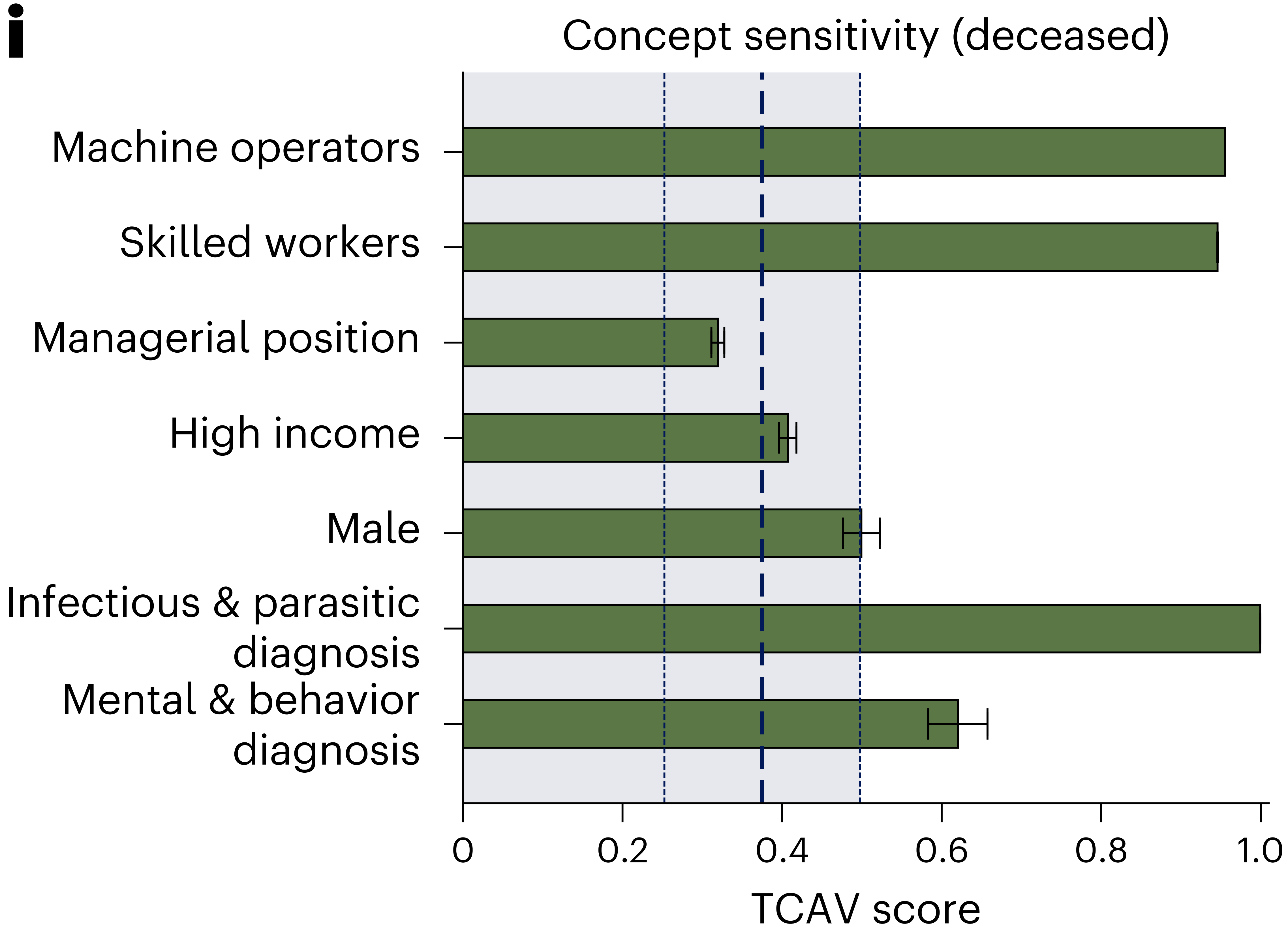
Abstract

The interpretation of deep learning models is a challenge due to their size, complexity, and often opaque internal state. In addition, many systems, such as image classifiers, operate on low-level features rather than high-level concepts. To address these challenges, we introduce Concept Activation Vectors (CAVs), which provide an interpretation of a neural net's internal state in terms of human-friendly concepts. The key idea is to view the high-dimensional internal state of a neural net as an aid, not an obstacle. We show how to use CAVs as part of a technique, Testing with CAVs (TCAV), that uses directional derivatives to quantify the degree to which a user-defined concept

A key difficulty, however, is that most ML models operate on features, such as pixel values, that do not correspond to high-level concepts that humans easily understand. Furthermore, a model's internal values (e.g., neural activations) can seem incomprehensible. We can express this difficulty mathematically, viewing the state of an ML model as a vector space E_m spanned by basis vectors e_m which correspond to data such as input features and neural activations. Humans work in a different vector space E_h spanned by implicit vectors e_h corresponding to an unknown set of human-interpretable concepts.

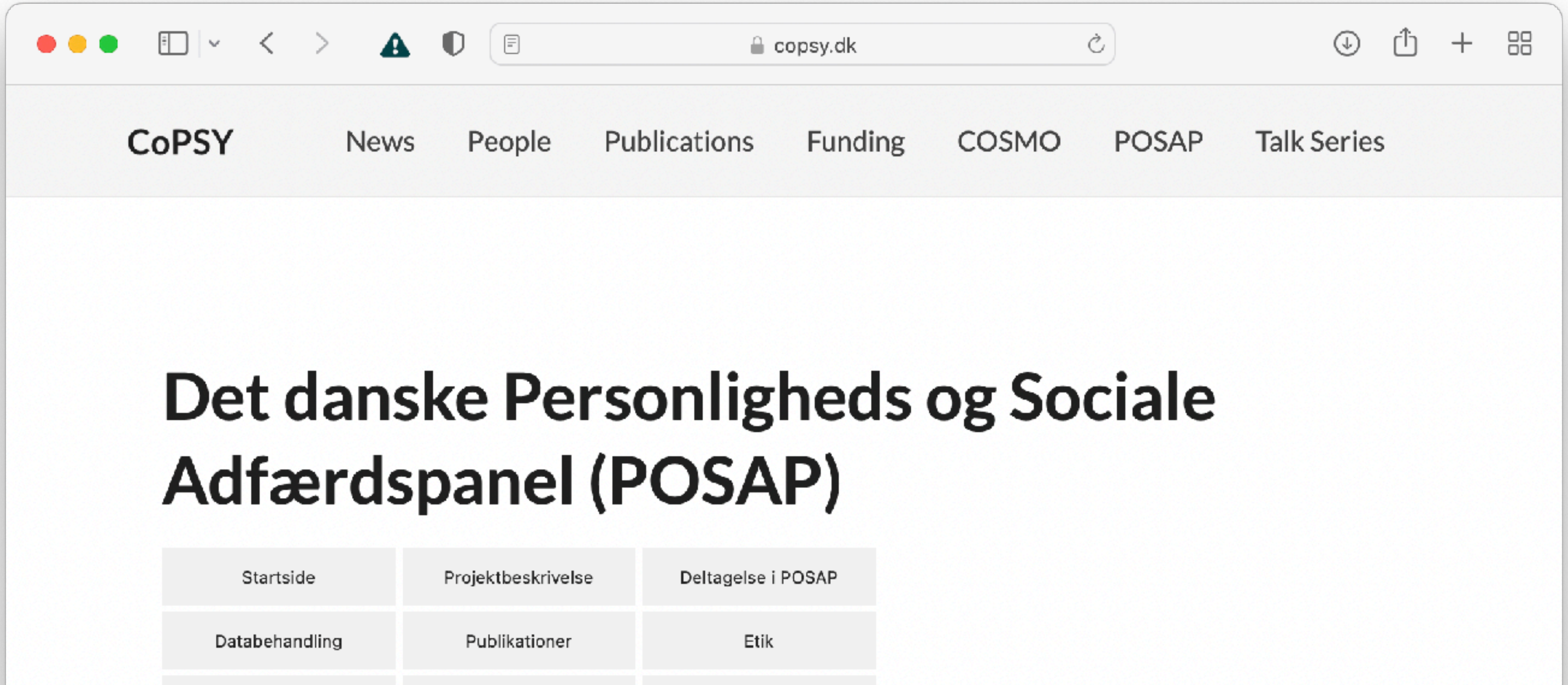
From this standpoint, an "interpretation" of an ML model can be seen as function $g : E_m \rightarrow E_h$. When g is linear, we call it a **linear interpretability**. In general, an interpretability function g need not be perfect (Doshi-Velez, 2017); it


h
Concept sensitivity (alive)
Machine operators
Skilled workers
Managerial position
High income
Male
Infectious & parasitic diagnosis
Mental & behavior diagnosis
0
0.2
0.4
0.6
0.8
1.0
TCAV score
Median score (concept)
No contribution region
Median score (random)
Bivariate midvariance (uncertainty)
Uncertainty of the random score

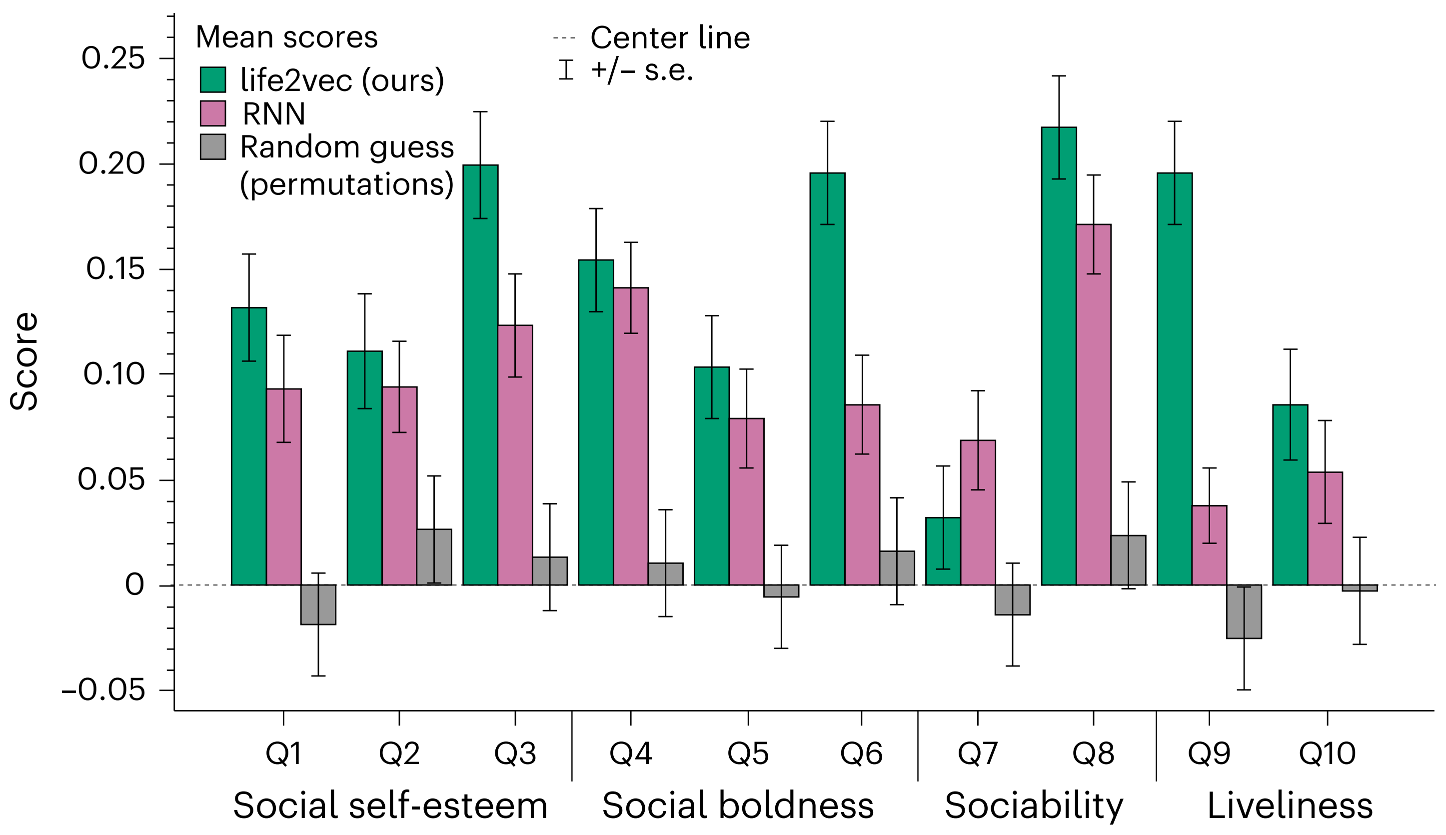

i
Concept sensitivity (deceased)
Machine operators
Skilled workers
Managerial position
High income
Male
Infectious & parasitic diagnosis
Mental & behavior diagnosis
0
0.2
0.4
0.6
0.8
1.0
TCAV score

We can make good predictions

* End of sequence
* **Personality** (w Ingo Zettler, Lau Lillegaard)

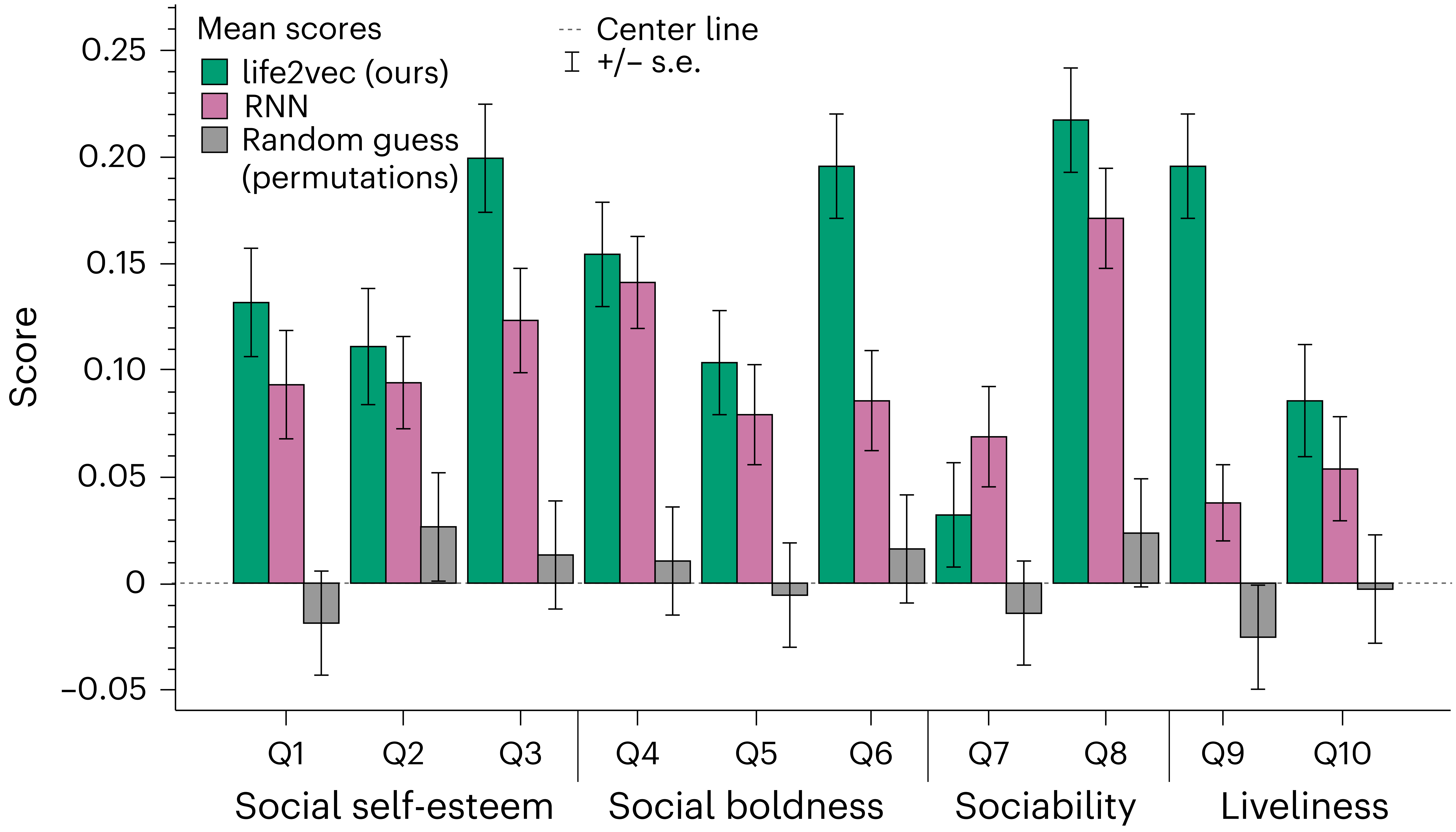


We can make really good predictions



1. I feel that I am an unpopular person,
2. I feel reasonably satisfied with myself overall,
3. I sometimes feel that I am a worthless person,
4. When I'm in a group of people, I'm often the one who speaks on behalf of the group,
5. In social situations, I'm usually the one who makes the first move,
6. I rarely express my opinions in group meetings,
7. The first thing that I always do in a new place is to make friends,
8. I prefer jobs that involve active social interaction to those that involve working alone,
9. Most people are more upbeat and dynamic than I generally am,
10. On most days, I feel cheerful and optimistic.

We can make really good predictions



Next steps

THANK YOU