Broken limits to social science? Foundation models for demography.

Laust Hvas Mortensen

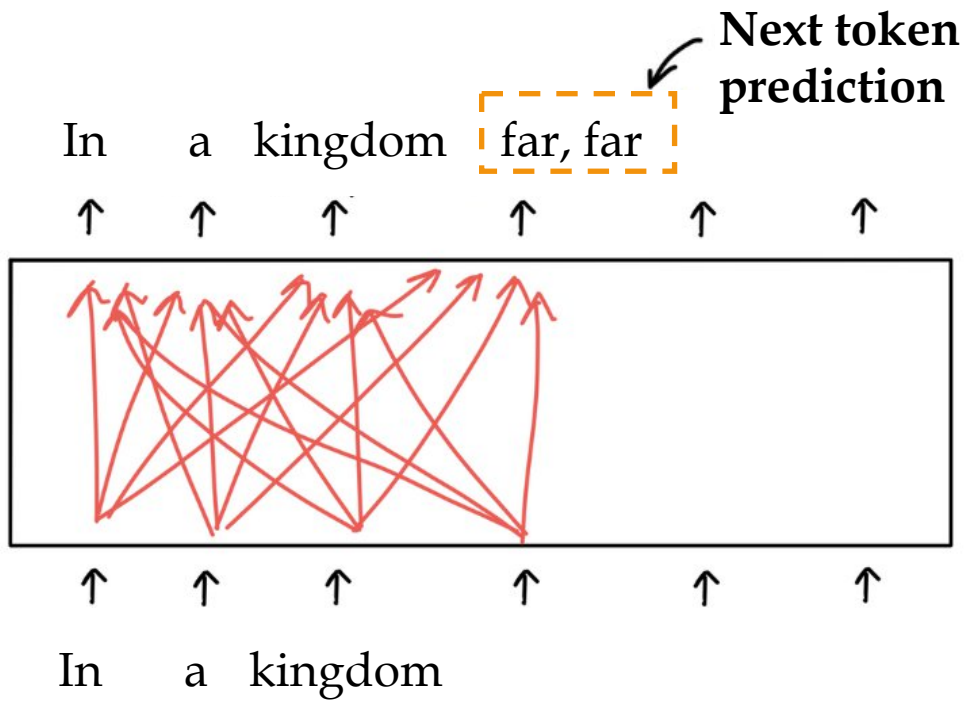
Research Professor, ROCKWOOL Foundation
Professor of Epidemiology, University of Copenhagen
Founder of Data Science Lab, Statistics Denmark

Email: lhm@rff.dk

1. Next token prediction
2. Foundation models: Why do they even work?
3. Examples of LLMs beyond languages
4. Using LLMs in the social science

1. Next token prediction

LLMs: Next token prediction

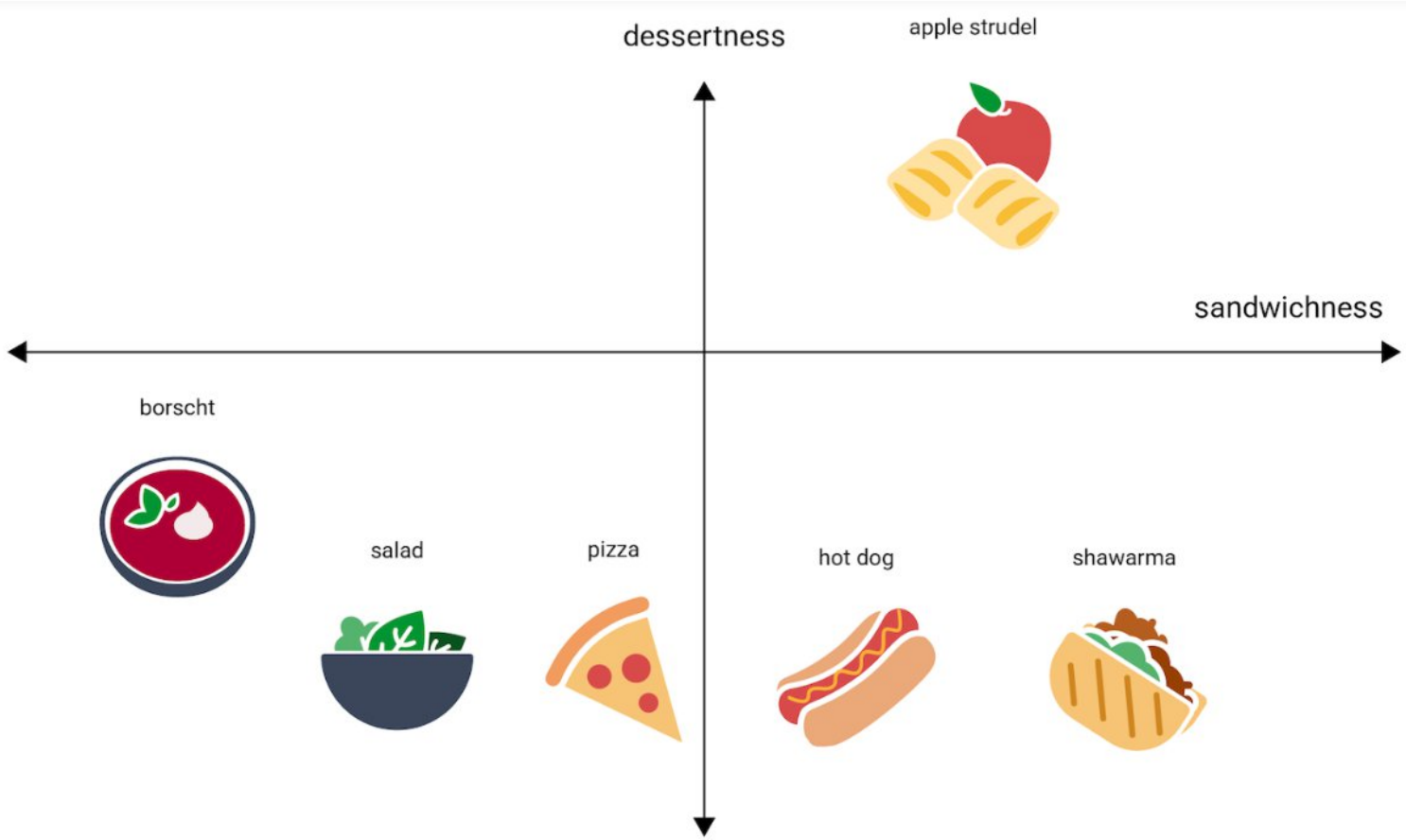


Words, vocabularies, and tokens

Attention

Embeddings

Embedding space



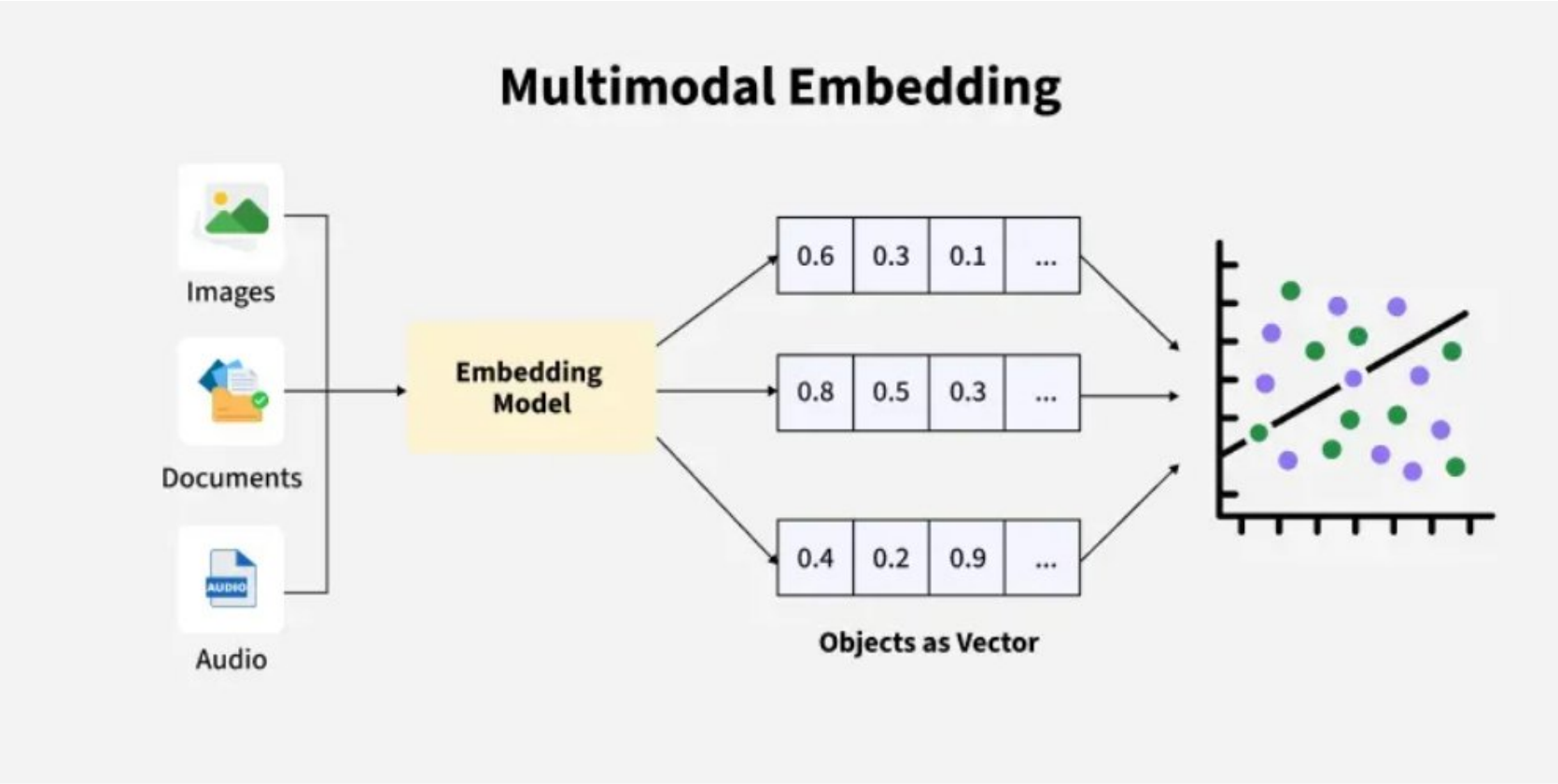
Source: Embeddings: Embedding space and static embeddings | Machine Learning | Google for Developers

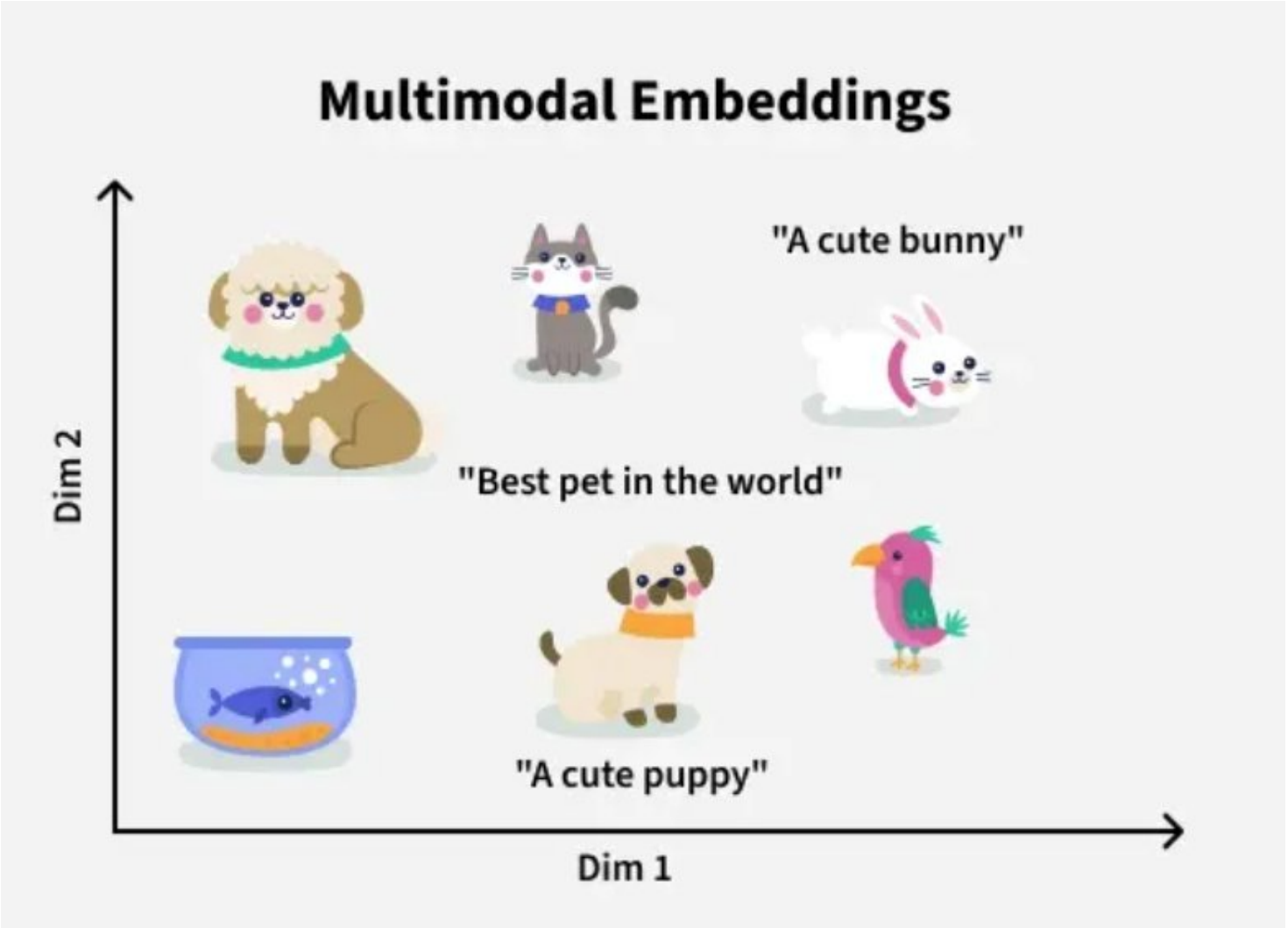
2. Foundation models: Why do they even work?

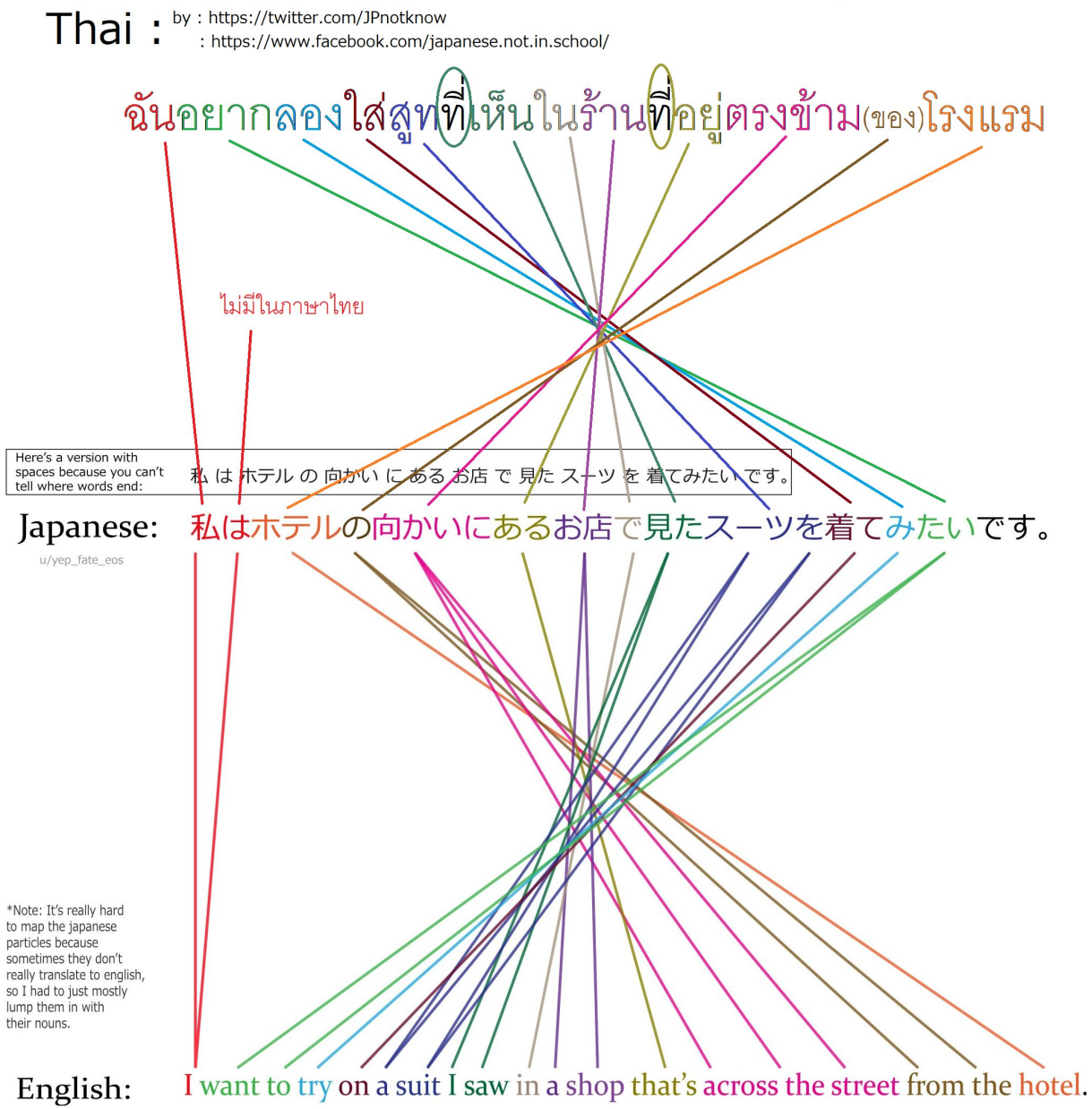


"AI is undergoing a paradigm shift with the rise of models that are [redacted] and are [redacted] tasks. We call these models foundation models to underscore their critically central yet incomplete character."

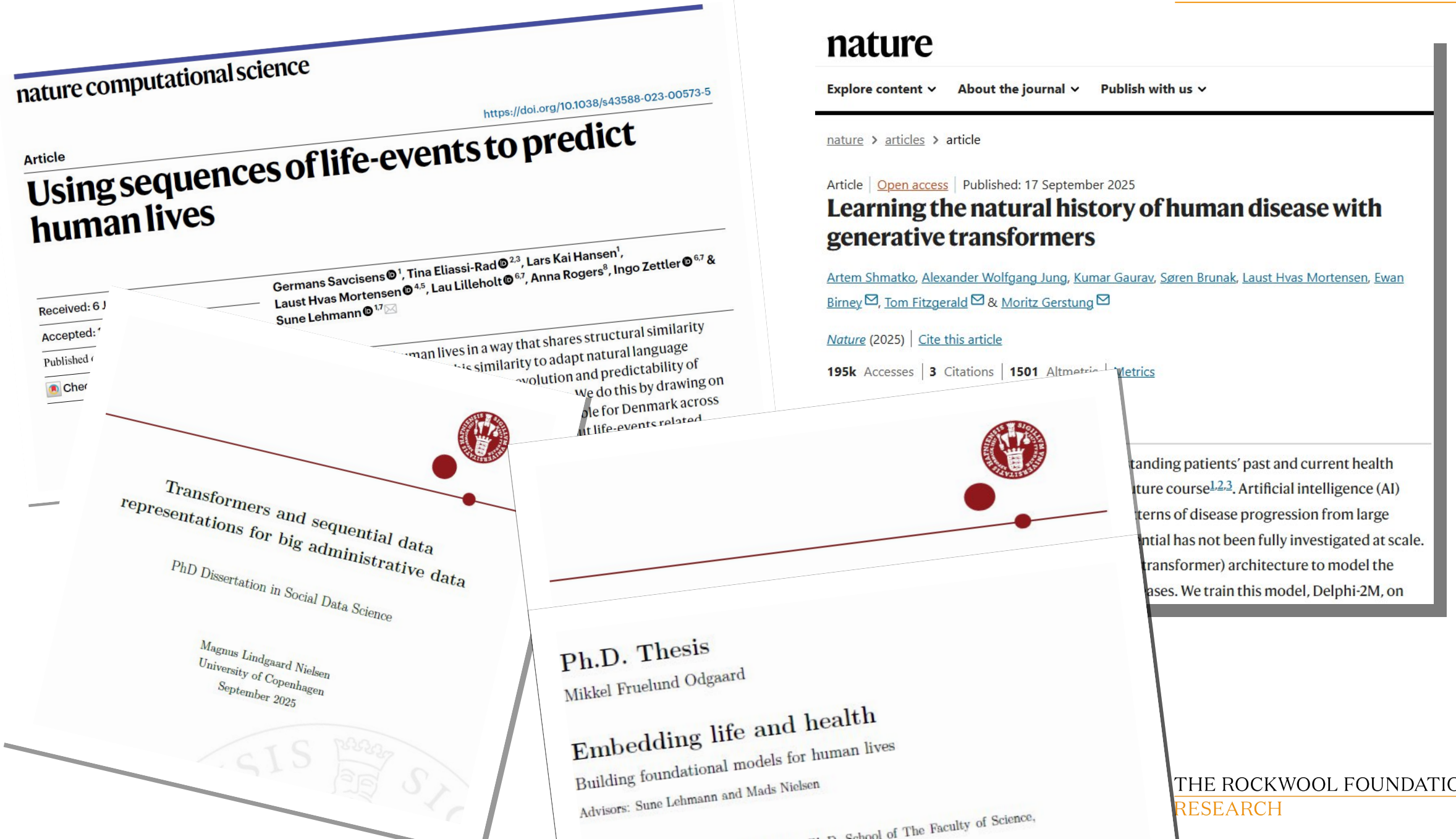
"Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization."


Multimodal Embedding
Images
Documents
Audio
Embedding Model
0.6 0.3 0.1 ...
0.8 0.5 0.3 ...
0.4 0.2 0.9 ...
Objects as Vector


Multimodal Embeddings
"A cute bunny"
"Best pet in the world"
"A cute puppy"
Dim 2
Dim 1


Thai :
by : https://twitter.com/JPnotknow
: https://www.facebook.com/japanese.not.in.school/
ฉันอยากลองใส่สูทที่เห็นในร้านที่อยู่ตรงข้าม(ของ)โรงแรม
ไม่มีในภาษาไทย
Here's a version with spaces because you can't tell where words end:
私 は ホテル の 向かい に ある お店 で 見た スーツ を 着てみたい です。
Japanese:
私はホテルの向かいにあるお店で見たスーツを着てみたいです。
u/yep_fate_eos
*Note: It's really hard to map the japanese particles because sometimes they don't really translate to english, so I had to just mostly lump them in with their nouns.
English:
I want to try on a suit I saw in a shop that's across the street from the hotel.

3. LLMs beyond languages (Life2vec, DELPHI2M, Destiny Machine)


nature computational science
https://doi.org/10.1038/s43588-023-00573-5
Article
Using sequences of life-events to predict human lives
Germans Savcisens, Tina Eliassi-Rad, Lars Kai Hansen, Laust Hvas Mortensen, Lau Lilleholt, Anna Rogers, Ingo Zettler & Sune Lehmann
Received: 6 J
Accepted:
Published
Transformers and sequential data representations for big administrative data
PhD Dissertation in Social Data Science
Magnus Lindgaard Nielsen
University of Copenhagen
September 2025
nature
Explore content
About the journal
Publish with us
nature › articles › article
Article | Open access | Published: 17 September 2025
Learning the natural history of human disease with generative transformers
Artem Shmatko, Alexander Wolfgang Jung, Kumar Gaurav, Søren Brunak, Laust Hvas Mortensen, Ewan Birney, Tom Fitzgerald & Moritz Gerstung
Nature (2025) | Cite this article
195k Accesses | 3 Citations | 1501 Altmetric | Metrics
Ph.D. Thesis
Mikkel Fruelund Odgaard
Embedding life and health
Building foundational models for human lives
Advisors: Sune Lehmann and Mads Nielsen



Entered according to Act of Congress in the year 1848, by J. Baillie, in the Clerk's Office of the Distt Court, of the South.n Distt of N.Y.

Published by James Baillie, 87th St near 3rd Avenue N.Y.



LITH. & PUB. BY N. CURRIER,

Entered according to Act of Congress in the year 1846 by N. Currier, in the Clerk's office of the District Court of the Southern District of N.Y.

33 SPRUCE ST. N.Y.

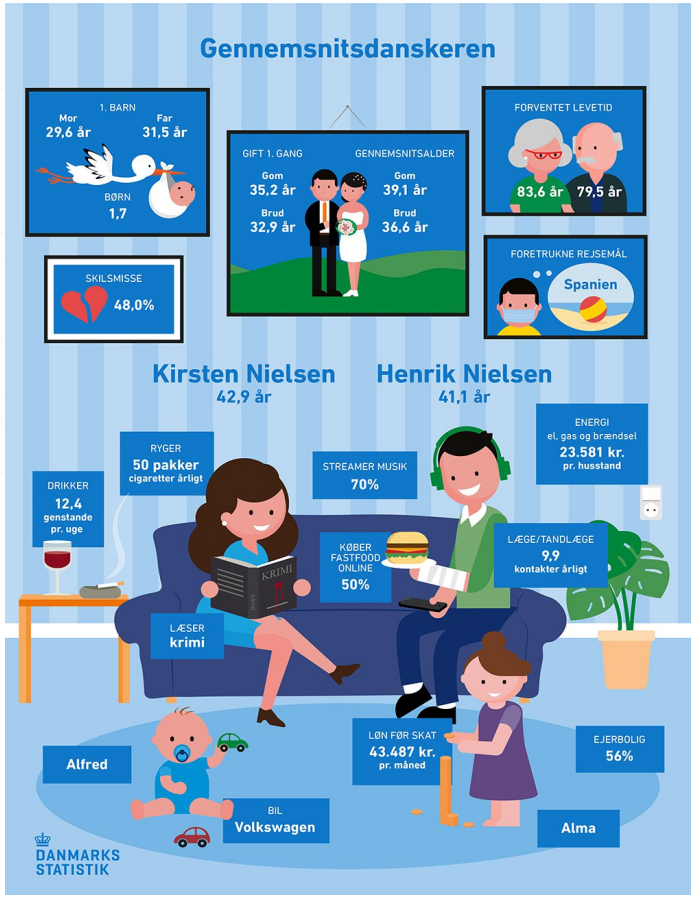
THE DRUNKARDS PROGRESS.

FROM THE FIRST GLASS TO THE GRAVE

+50 years of data



Data on many things

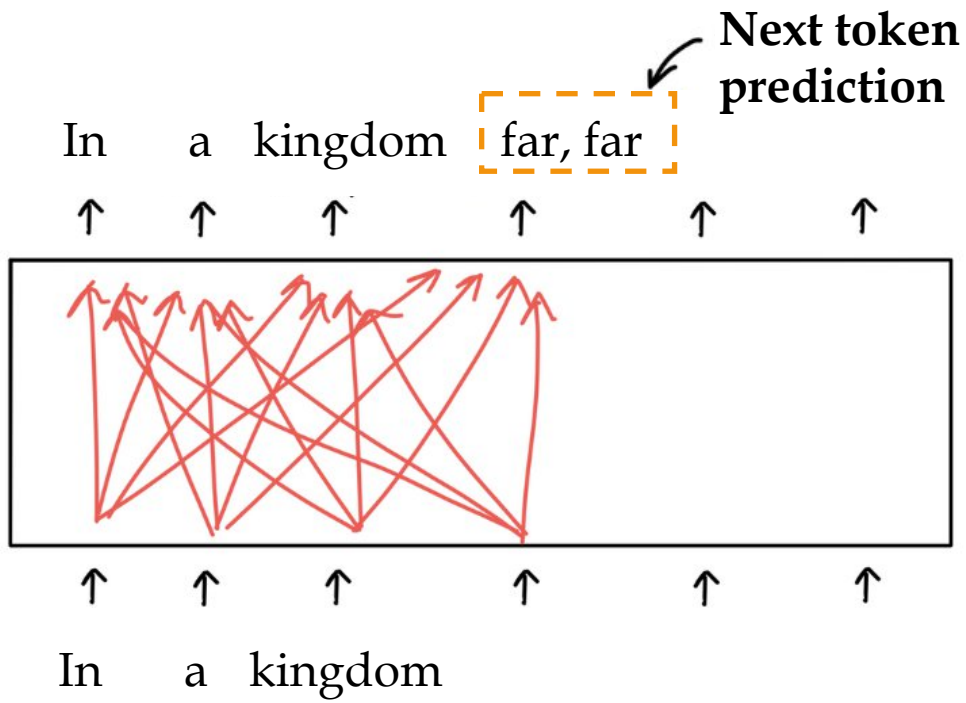


Fast and easy access, good governance

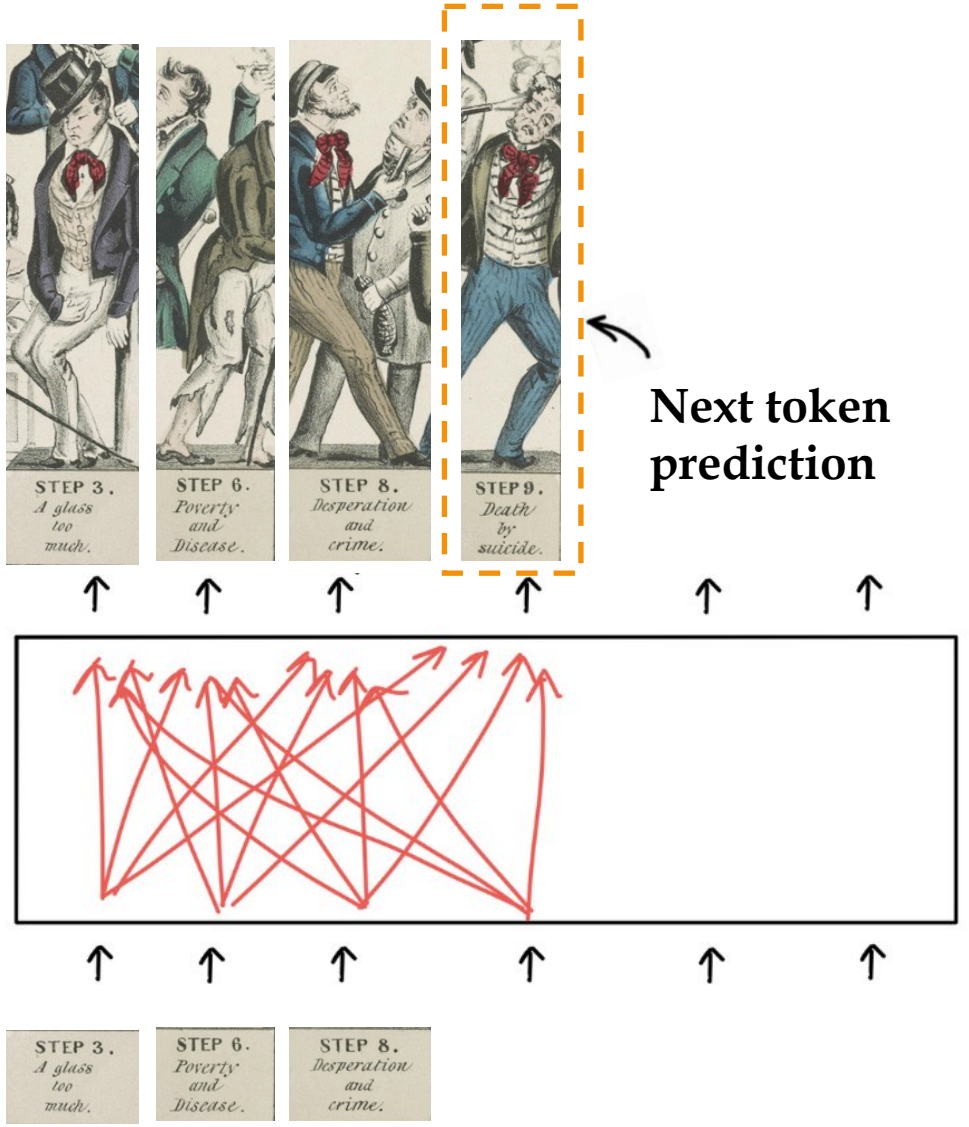
NVK
NATIONAL VIDENSKABSETISK KOMITÉ

DATATILSYNET

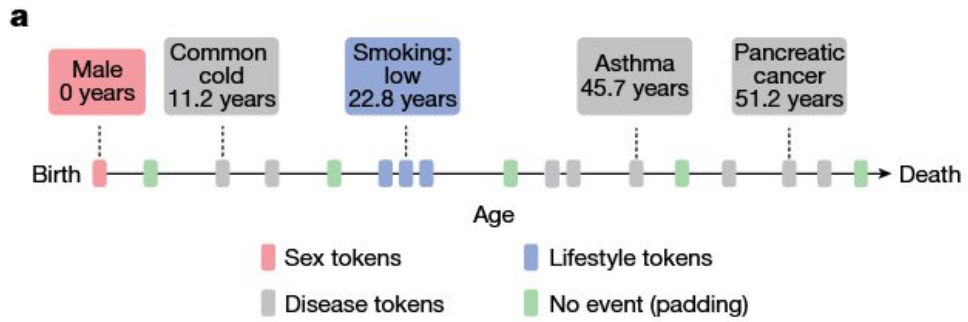
LLMs: Next token prediction

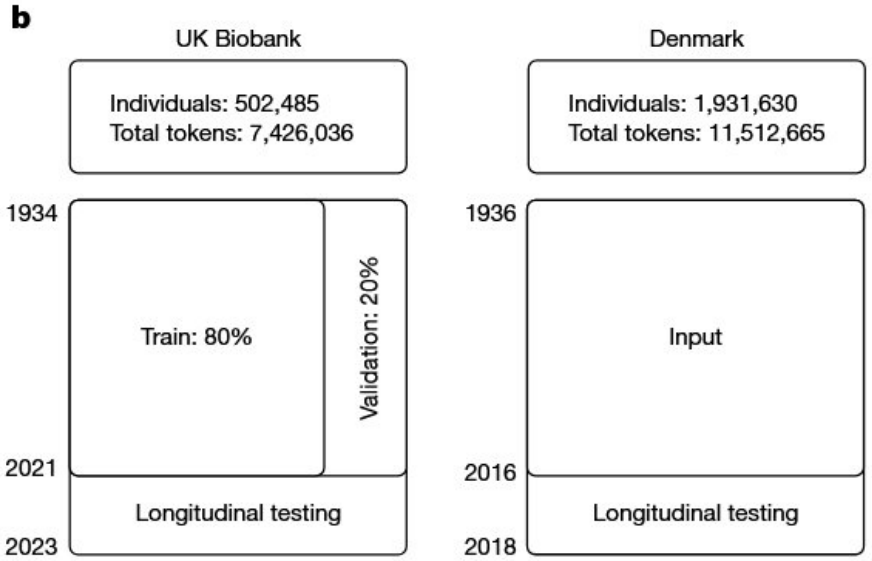


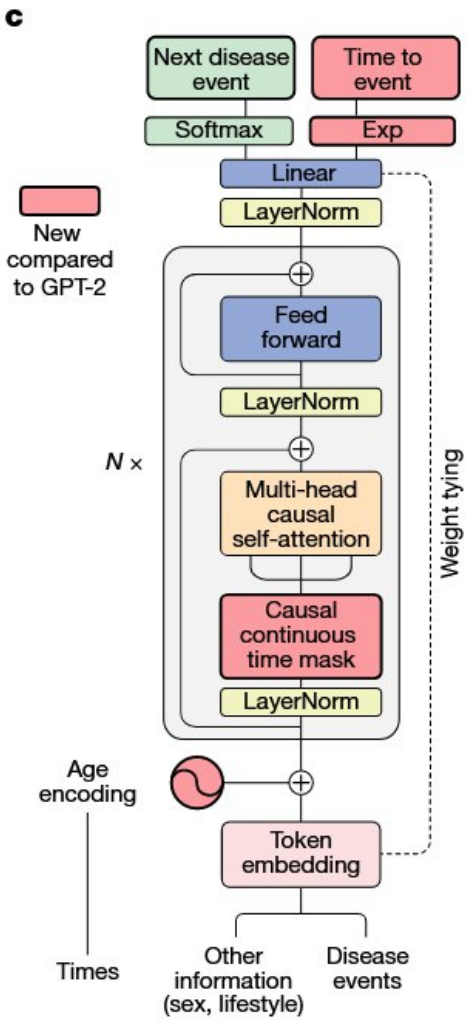
Events in life are like words in sentences



Model architecture: DELPHI-2M







d

Input:
Age: Token
0.0: Male
2.0: B01 varicella (chickenpox)
3.0: L20 atopic dermatitis
5.0: No event
10.0: No event
15.0: No event
20.0: No event
20.0: G43 migraine
21.0: E73 lactose intolerance
22.0: B27 infectious mononucleosis
25.0: No event
28.0: J11 influenza, virus not identified
30.0: No event
35.0: No event
40.0: No event
41.0: Smoking low
41.0: BMI mid
41.0: Alcohol low
42.0: No event

Output:
43.2: No event
43.5: M54 dorsalgia
44.6: I86 varicose veins of other sites
50.4: K52 other non-infective gastroenteritis and colitis
52.2: H83 other diseases of inner ear
53.9: J22 unspecified acute lower respiratory infection
54.5: L30 other dermatitis
55.3: No event
57.5: L50 urticaria
59.4: K62 other diseases of anus and rectum
...
69.8: J90 pleural effusion, not elsewhere classified
70.0: K21 gastro-oesophageal reflux disease
70.1: K76 other diseases of liver
70.3: I10 essential primary hypertension
70.4: M85 other disorders of bone density and structure
70.7: M81 osteoporosis without pathological fracture
71.2: J98 other respiratory disorders
72.1: J80 adult respiratory distress syndrome
72.2: No event
72.7: Death

LLM-like models: Everything in context

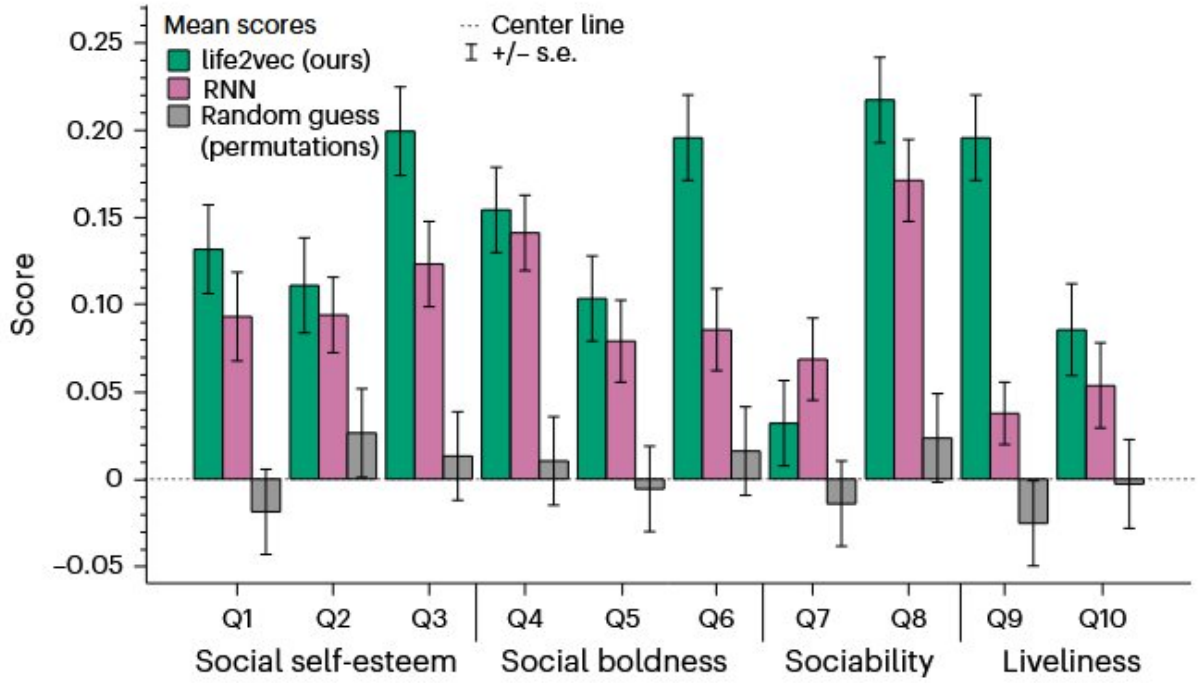
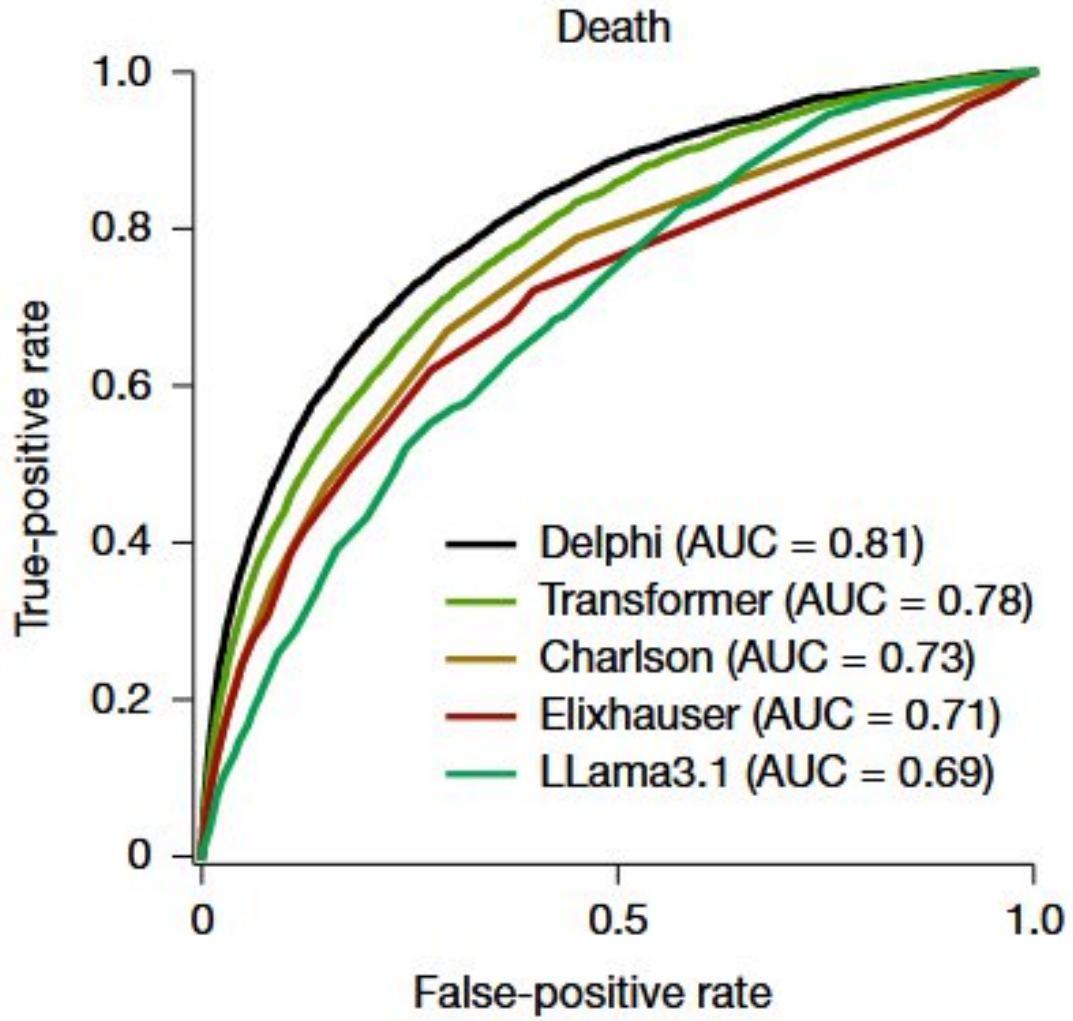


Fig. 5 | Performance evaluation for the personality nuances task. Cohen's quadratic kappa score, κ, for each of the ten extraversion questionnaire items ($n = 1{,}417$). The bars represent κ for life2vec (green), RNN (purple) and a random guess that draws predictions from the actual distribution of targets (gray). The error bars and whiskers correspond to ±1 s.e. of κ. The dashed line corresponds to $\kappa = 0$. The question wordings are provided in the Personality nuances prediction task section.

LLM-like models: Prediction



👍 Decent at group-level

👎 Marginal improvement
👎 Not explainable
👎 Not actionable

NEW YORK POST

TECH

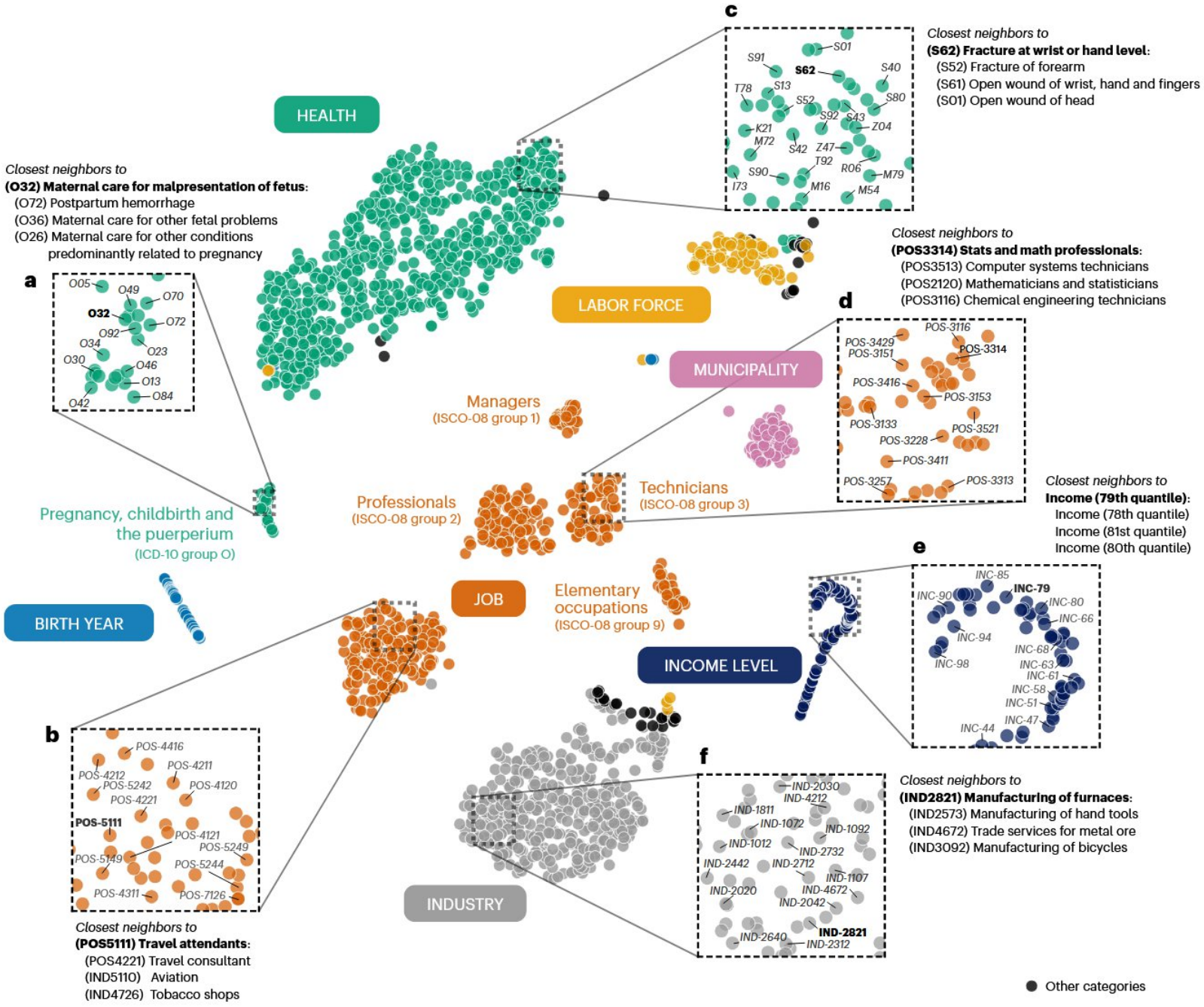
‘AI death calculator’ creators issue urgent warning about frighteningly accurate tool

By Ben Cost

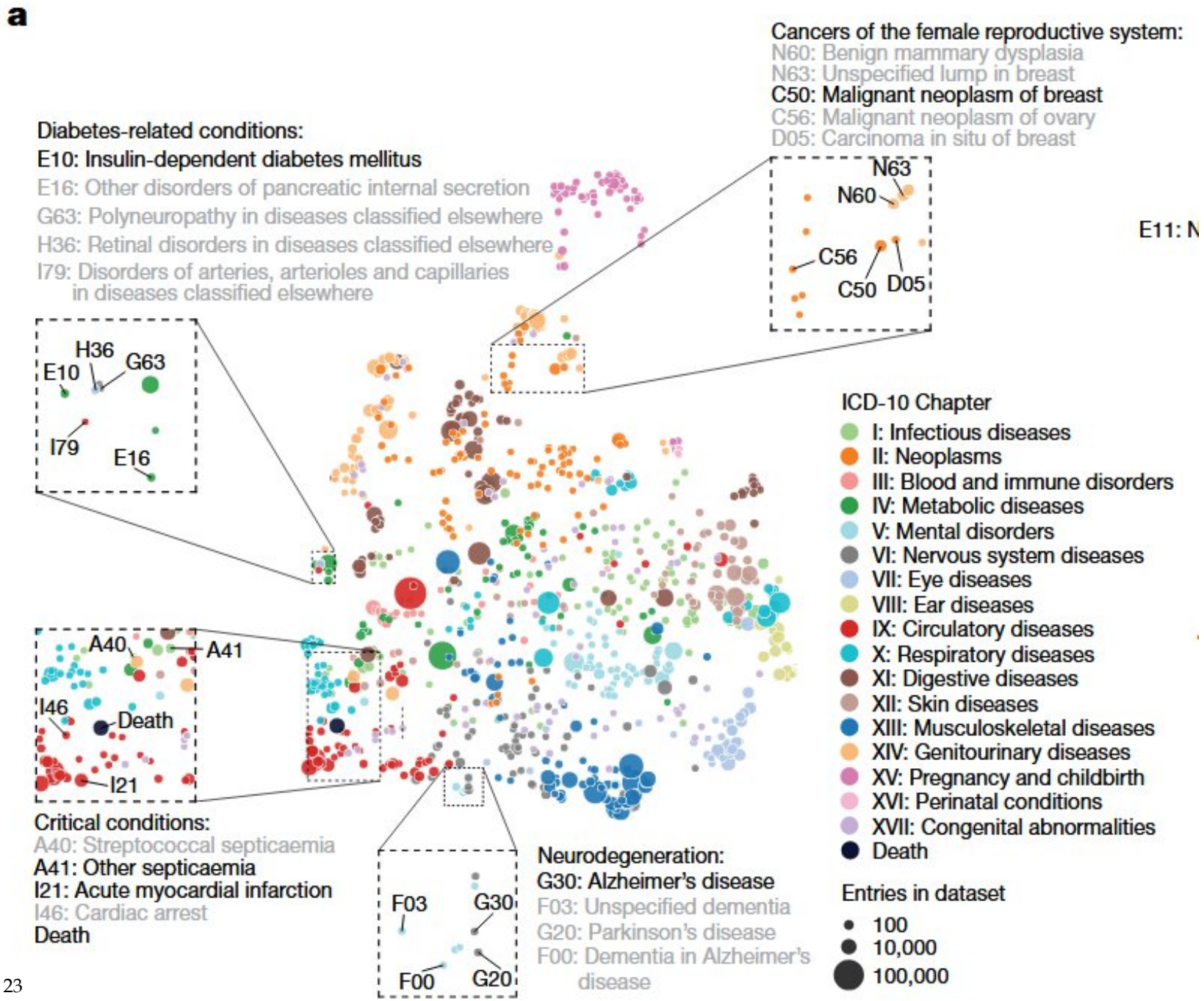
Published April 29, 2024 | Updated April 29, 2024, 9:32 a.m. ET

111 Comments

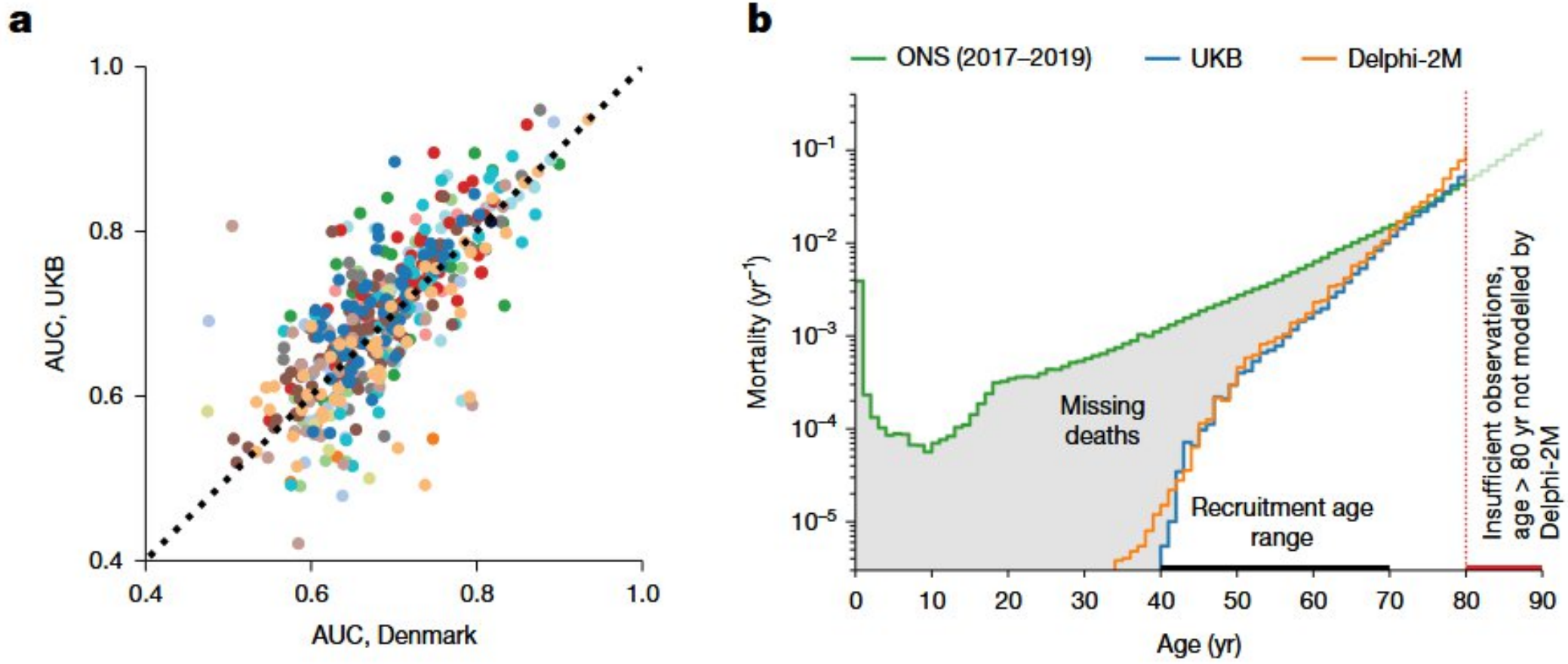




Life2vec: Embedding space



DELPHI2M: Embedding space

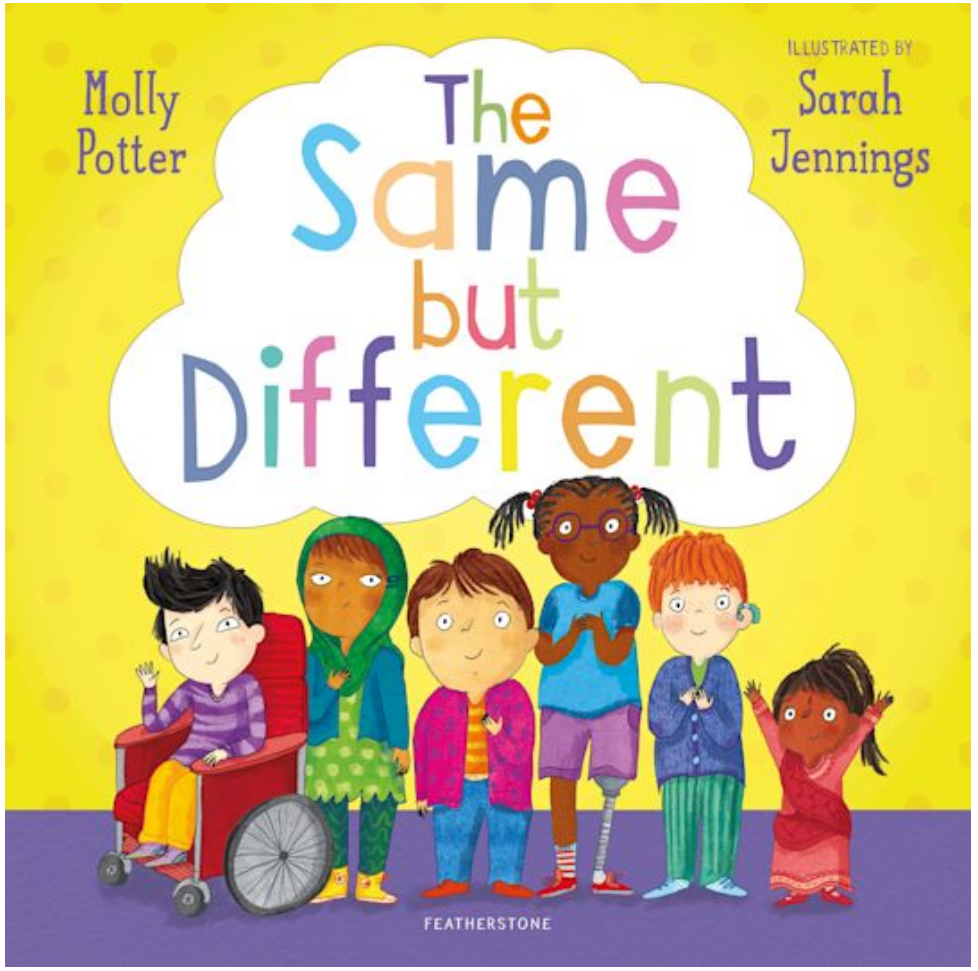

a
1.0
0.8
0.6
0.4
AUC, UKB
0.4
0.6
0.8
1.0
AUC, Denmark
b
ONS (2017–2019)
UKB
Delphi-2M
Mortality (yr⁻¹)
10⁻¹
10⁻²
10⁻³
10⁻⁴
10⁻⁵
Missing deaths
Recruitment age range
Insufficient observations, age > 80 yr not modelled by Delphi-2M
0
10
20
30
40
50
60
70
80
90
Age (yr)

LLM-like models: Often works across contexts

ChatGPT: Works in most languages… English, Danish, Korean. Why? Exploits that there are similarities in language

Social LLMs: Transitions between educations/jobs in US, DK, NL, Korea despite different educational systems and labor markets.

Health LLMs: A model trained on UK data can identify diseases in DK, US, Korea despite different health care systems & coding practises.



https://www.bloomsbury.com/au/same-but-different-9781472977984/

The predictability of human lives

[redacted], [redacted]
Laust Hvas Mortensen[1,3,4], Andreas Bjerre-Nielsen[1], Sune Lehmann[1,2,5]

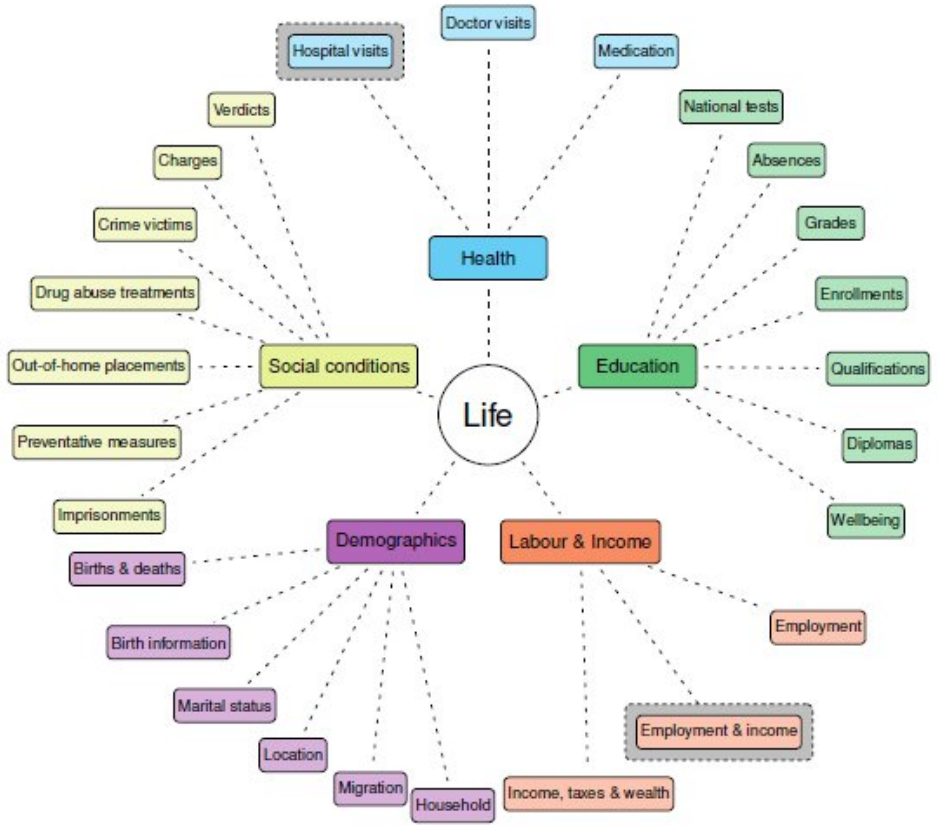


Figure 3.1: Overview of the input sources of the 'Life' dataset, categorized within our five distinct domains. The grey-highlights reflect the data sources used in life2vec [23].

4. Using LLMs in the social science


Sarah Connor watching you
use AI for everything

Synthetic personas distort the structure of human belief systems

Christopher Barrie[1], Roberto Cerina[2]

[1]Department of Sociology, New York University
[2]Institute for Logic, Language, and Computation, University of Amsterdam

February 26, 2026

"Large language models (LLMs) are increasingly used to stand in for survey respondents, often by conditioning on demographic "personas" to simulate diverse publics. That is, practitioners propose prompting an LLM to answer a survey question as if it were the description of the person it is given."

The Book of Life approach: Enabling richness and scale for life course research*

Mark Verhagen[1,2,3], Benedikt Stroebl[1,4], Tiffany Liu[5,6], Lydia T. Liu[1,4], and Matthew J. Salganik[1,5,6]

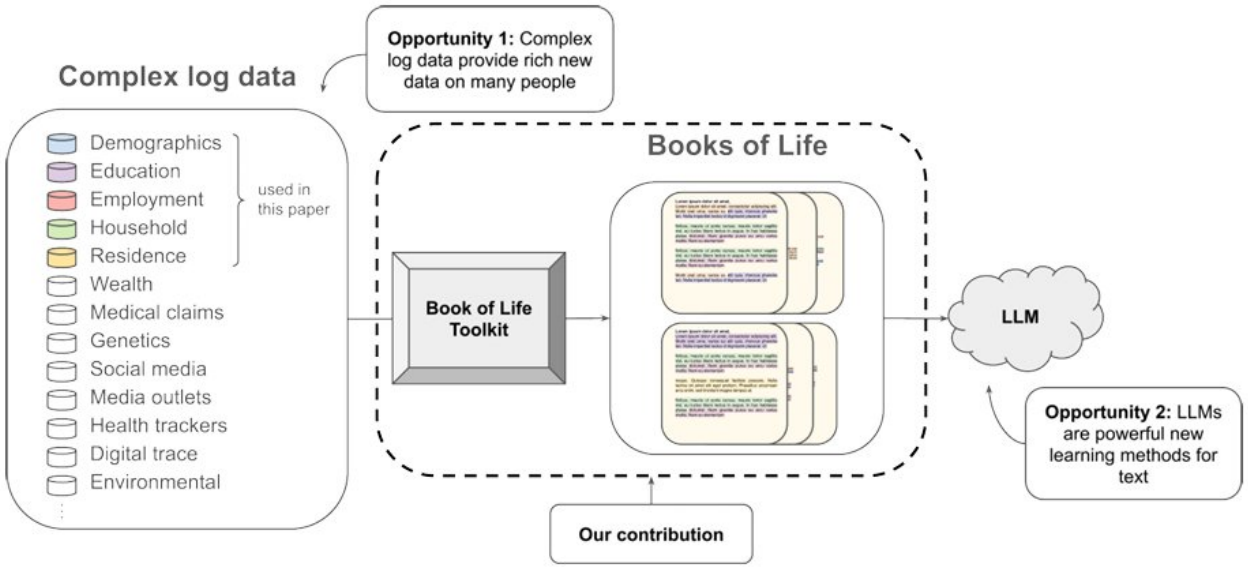


Figure 1: Illustration of the Book of Life approach for life course research, combining richness and scale. Complex log data on the left cover rich information on people's lives but scattered across many files that were created for purposes other than research, spanning different units of analysis, temporal resolution and relational linkages. Books of life express the information for a single individual in plain text. The Book of Life Toolkit (BOLT) in the middle allows mapping complex log data to books of life. An example downstream use case of the books of life approach using LLMs is described in Satish et al. (2025).

Listening to the Voices of America

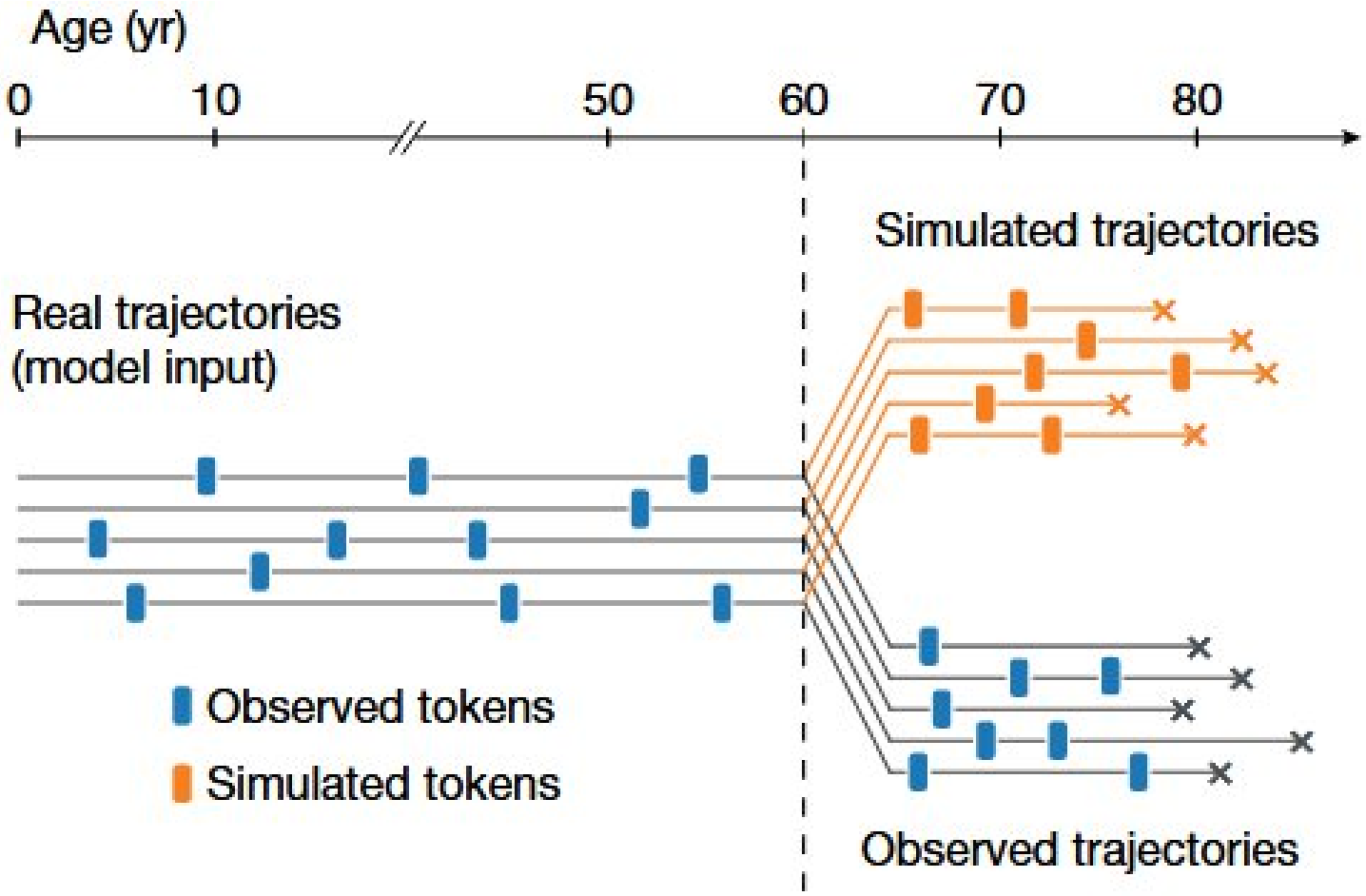
KATHRYN J. EDIN, COREY D. FIELDS, DAVID B. GRUSKY, JURE LESKOVEC, MARYBETH J. MATTINGLY, KRISTEN OLSON, AND CHARLES VARNER

We make the case for building a permanent public-use platform for conducting and analyzing immersive interviews on the everyday lives of Americans. The American Voices Project (AVP)—a widely watched experiment with this new platform—provides important early evidence on its promise. The articles in this issue reveal that, although public-use interview datasets obviously cannot meet all research needs, they do provide new opportunities to study small or hidden populations, new or emerging social problems, reactions to ongoing social crises, submerged values and attitudes, and many other aspects of American life. We conclude that a permanent AVP platform would help build an "open science" form of qualitative research that complements—rather than replaces—the existing very important body of immersive-interviewing research.

Keywords: qualitative, immersive, open science

At the height of the 2008 financial crisis, Queen Elizabeth II asked, "Why did nobody see it coming?" When economist Paul Krugman delivered an address in Lisbon four years later, he owned up to the failure and placed the blame squarely on his discipline. Economists, he asserted, should be "ashamed of their profession" because it had failed to predict, much less coherently explain, one of the key crises of our time—the Great Recession (Krugman 2012).

Generative AI



What's the use of AI on nation-scale data?

The Danish **fertility crisis** as an example:

- What are the groups that have decreasing cohort fertility?
- Who are the individuals that do not form couples?
- Who are the couples that postpone fertility decisions?
- Whose fertility decisions are affected by shocks to education, housing, job, climate?

