Data Visualization for Machine Learning

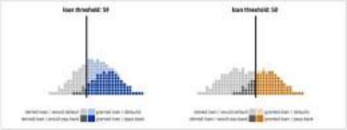
Fernanda Viégas @viegasf
Google Brain

Embedding Projector

an open source, visualization tool for high-dimensional data

Fairness in ML

Try different tradeoffs yourself to understand issues around fairness and machine learning.



Machine Translation

Visualizing hints that a translation network learns an "interlingua", or universal language.

Geodetic Velocities Visualization

an open source visualization of earthquake cycle physics

TensorFlow Playground

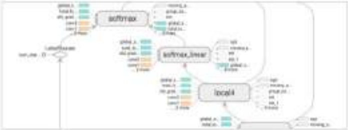
an open source, transparent neural net you can play with in your browser

Unfiltered News

see news coverage around the world and spot underreported stories
(a collaboration with Jigsaw)

TensorFlow Graph Visualizer

an open source, high-level view of TensorFlow computation graphs

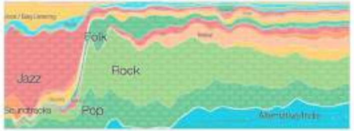


Periodic Table

a twist on the classic visualization of the atomic elements

Music Timeline

see how different musical genres became popular over time, and discover artists in each genre
UPDATED WEEKLY



Digital Attack Map

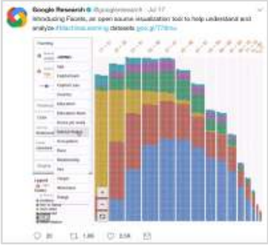
see live data on denial-of-service attacks across the world, and observe historical patterns
UPDATED DAILY

PAIR | People + AI Research Initiative

Bringing Design Thinking and HCI to Machine Learning

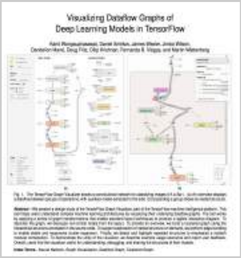
google.ai/pair

Open Source tools and platforms

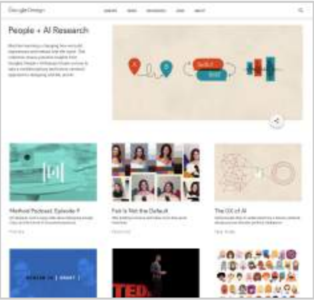


Educational Materials

Academic Publications



Public presentations, sharing best practices



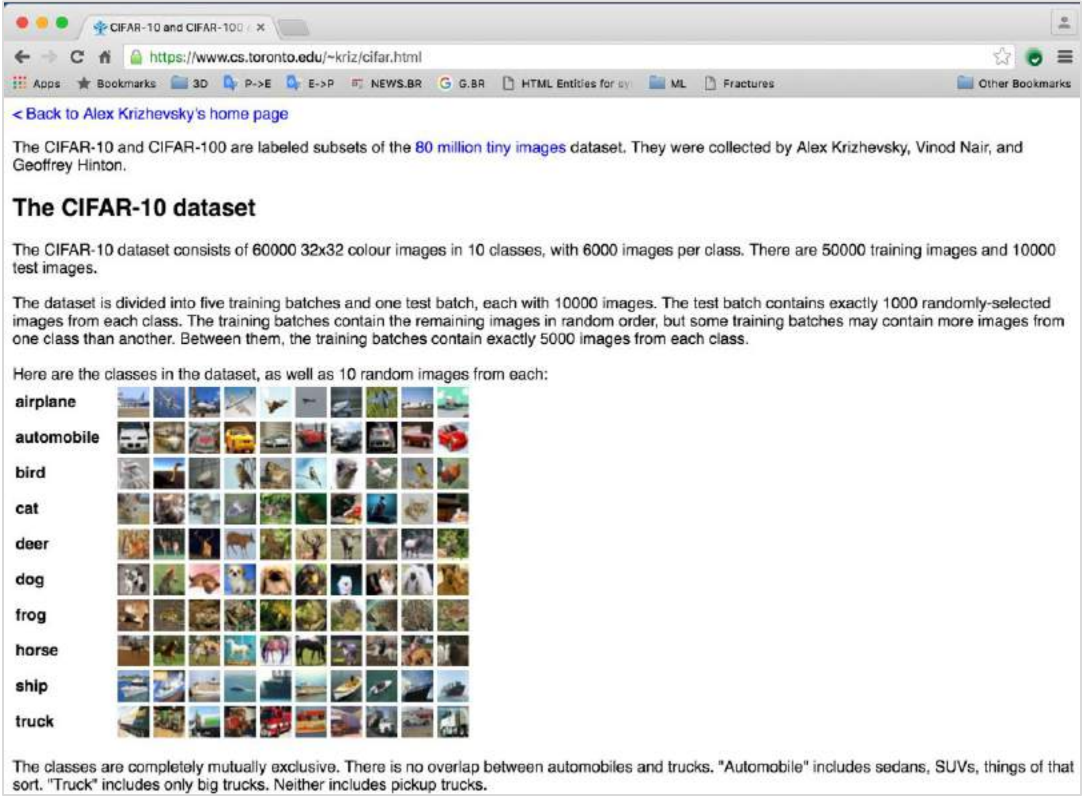
Public Symposia & meetings

Visiting Faculty, Faculty Grants

Training data is crucial

Debug your data before debugging your model

Let's start with a data set you might have heard of 😉



CIFAR-10 and CIFAR-100

https://www.cs.toronto.edu/~kriz/cifar.html

Apps Bookmarks 3D P->E E->P NEWS.BR G.BR HTML Entities for sy ML Fractures Other Bookmarks

< Back to Alex Krizhevsky's home page

The CIFAR-10 and CIFAR-100 are labeled subsets of the 80 million tiny images dataset. They were collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton.

The CIFAR-10 dataset

The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images.

The dataset is divided into five training batches and one test batch, each with 10000 images. The test batch contains exactly 1000 randomly-selected images from each class. The training batches contain the remaining images in random order, but some training batches may contain more images from one class than another. Between them, the training batches contain exactly 5000 images from each class.

Here are the classes in the dataset, as well as 10 random images from each:

airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck

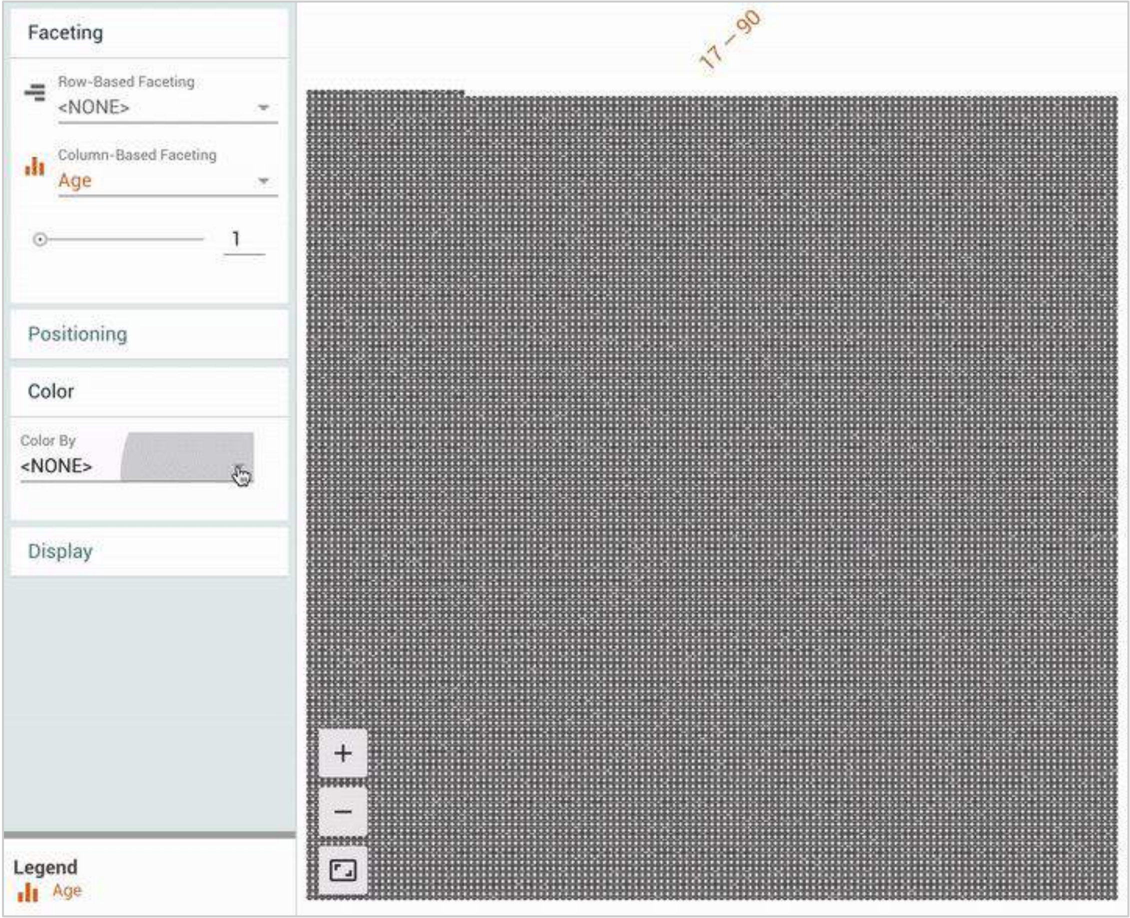
The classes are completely mutually exclusive. There is no overlap between automobiles and trucks. "Automobile" includes sedans, SUVs, things of that sort. "Truck" includes only big trucks. Neither includes pickup trucks.

32 x 32 images
10 classes

Facets

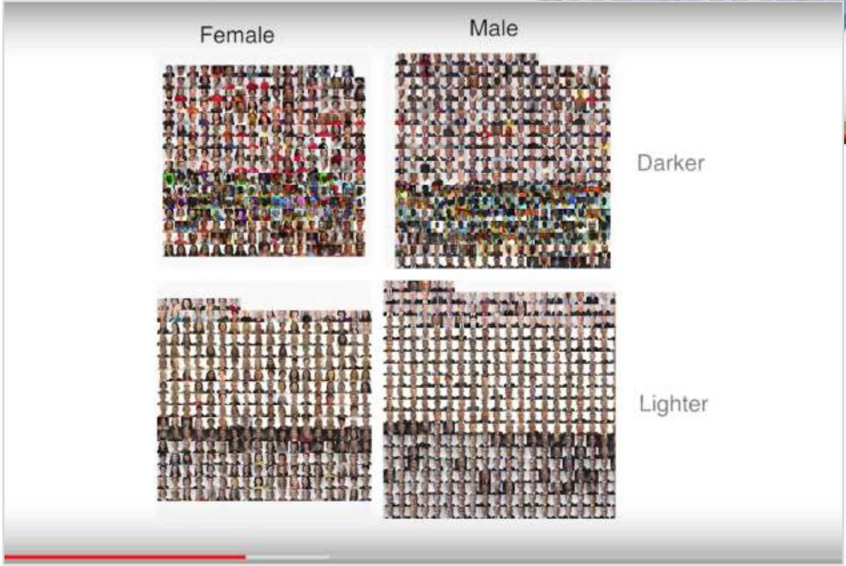
Open-source

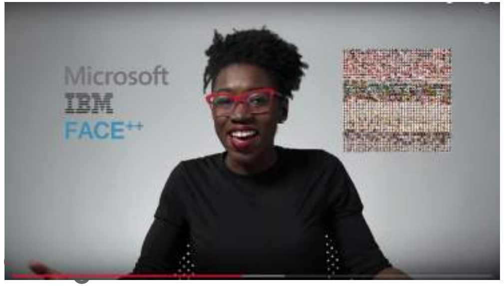
pair-code.github.io/facets



Gender Shades

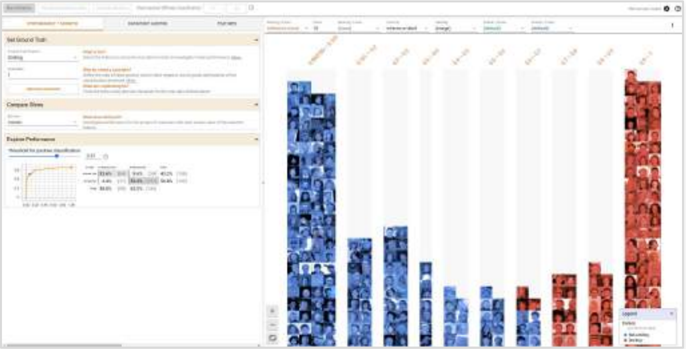
Joy Buolamwini
MIT Media Lab

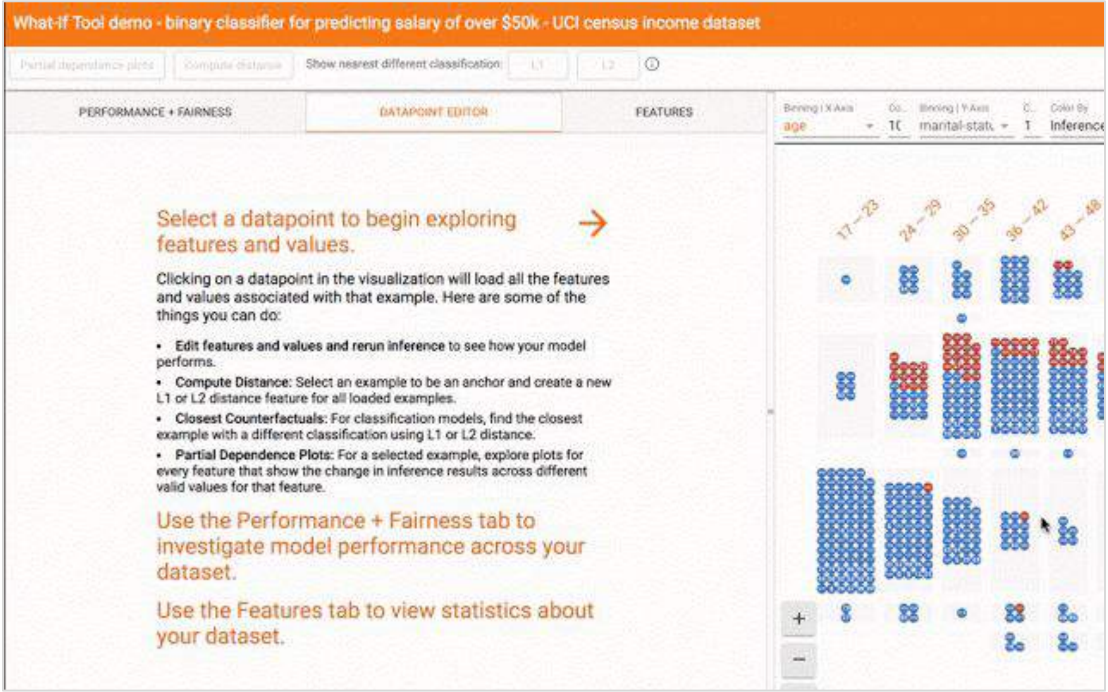




What-If Tool

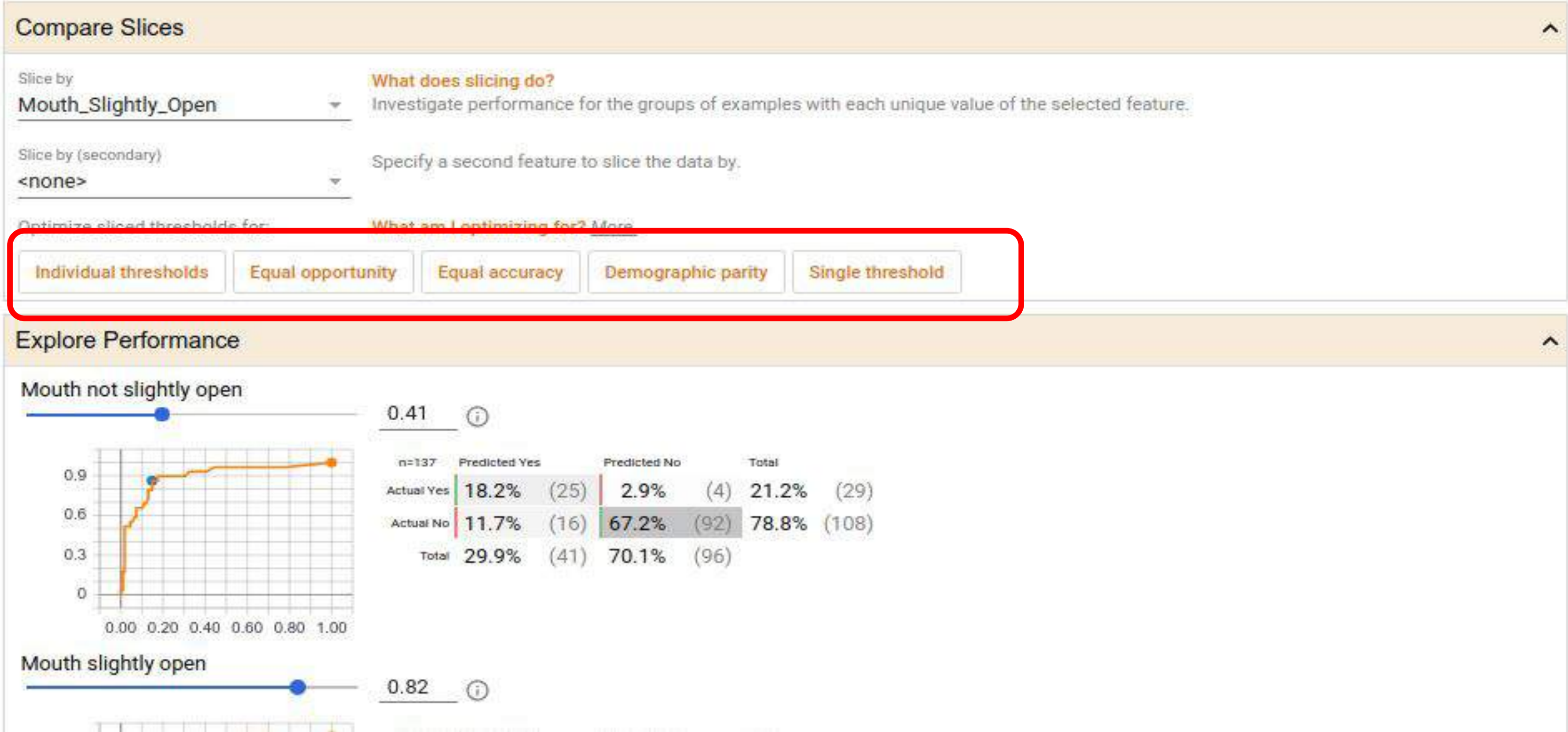
open source
code-free ML probing
pair-code.github.io/what-if-tool





What-If Tool

Fairness metrics

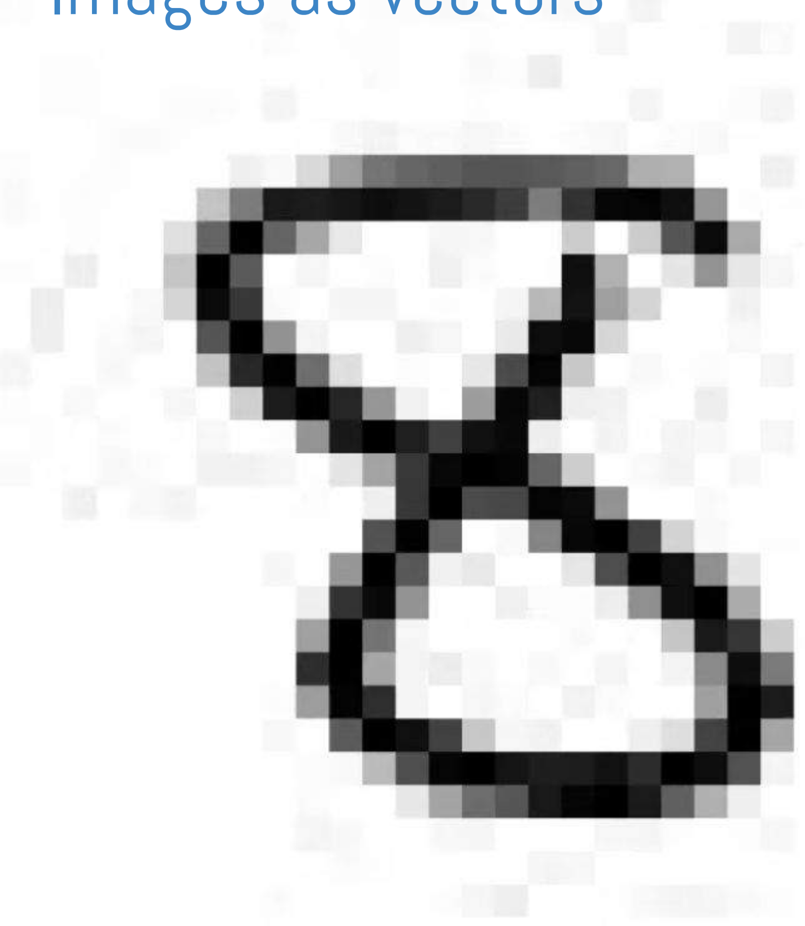


Model Understanding

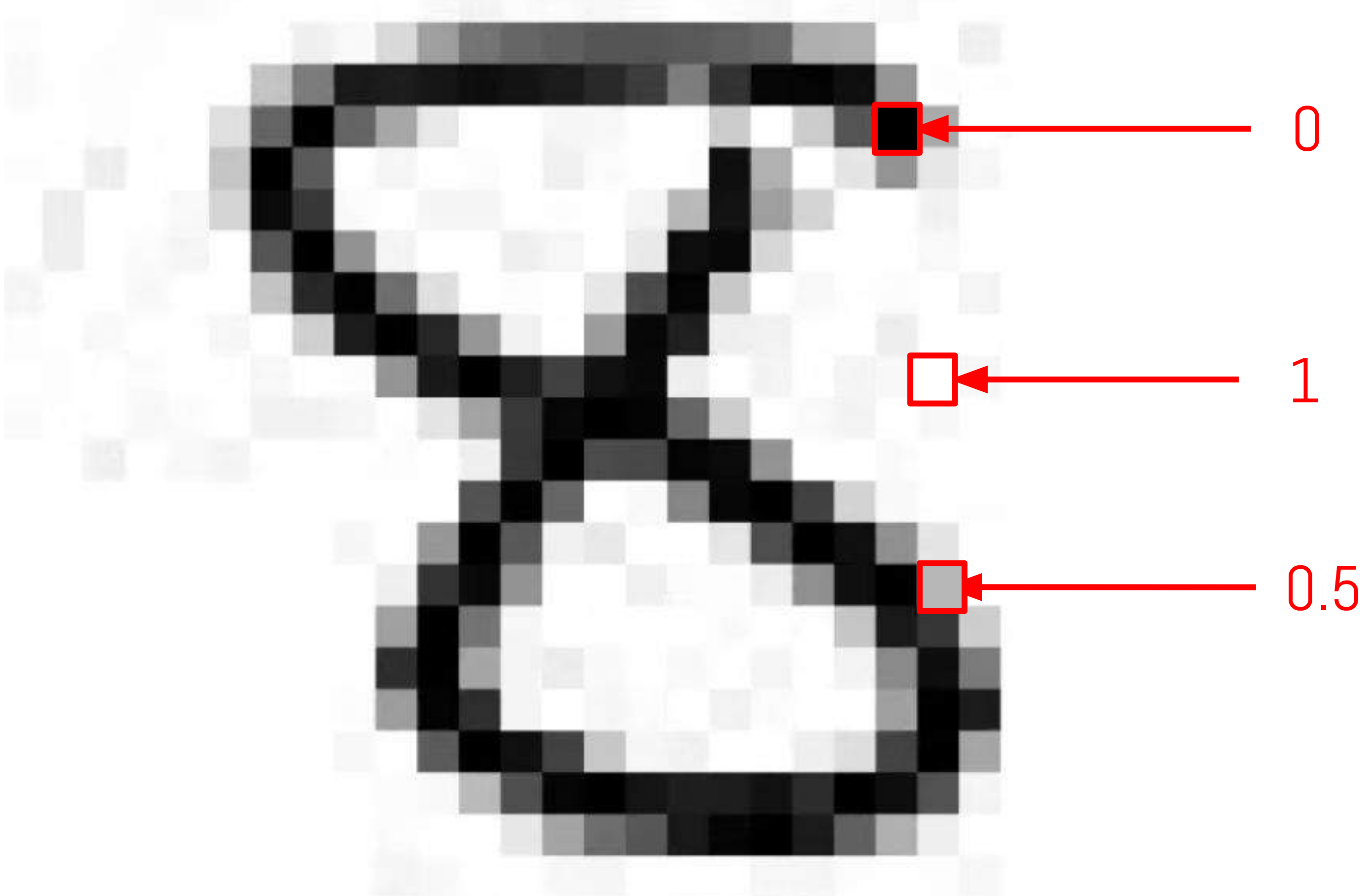
Looking into high-dimensional spaces

Warm up:
MNIST

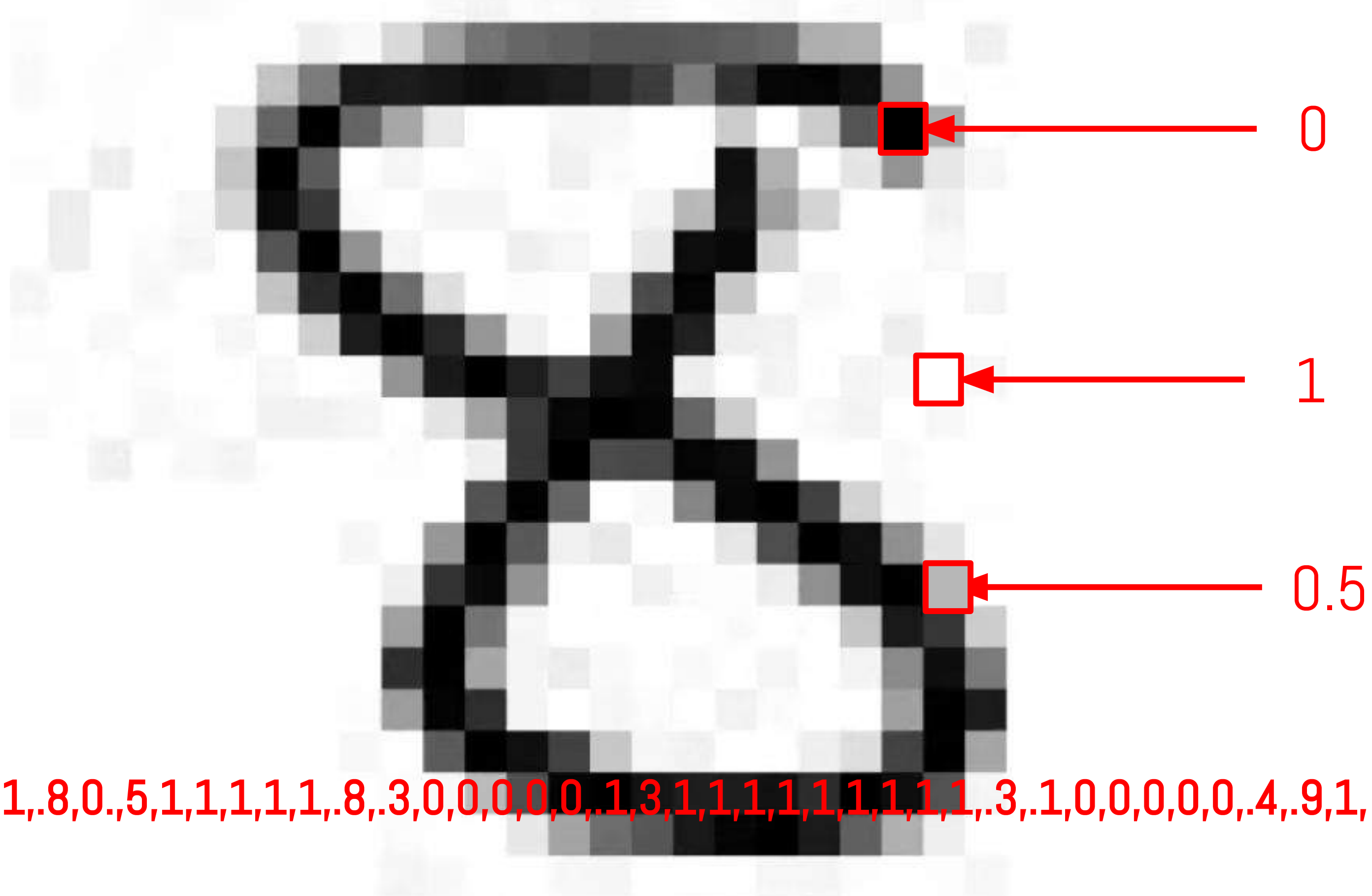
Images as vectors



Images as vectors

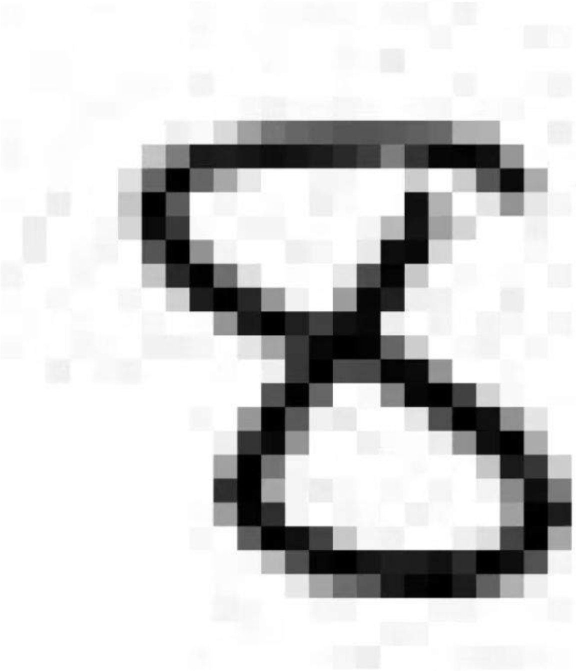


Images as vectors



(1,1,1,1,1,1,1,1,1,.8,0.,5,1,1,1,1,1,.8,.3,0,0,0,0,0,.1,3,1,1,1,1,1,1,1,1,1,.3,.1,0,0,0,0,0,.4,.9,1, ...)

We've turned this image



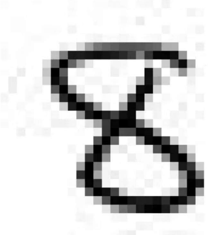
into this vector

(1,1,1,1,1,1,1,1,1,.8,0.,5,1,1,1,...)

784 pixels → 784 dimensions

We've turned this image

into this vector



(1,1,1,1,1,1,1,1,1,.8,0.,5,1,1,1,...)

784 pixels → 784 dimensions

(1,1,1,1,1,1,1,1,.6,.7,0.,4,1,1,1,...)

(1,1,1,1,1,1,1,1,.4,.5,0.,3,.2,1,1,...)

.
.
.

Embedding projector

MNIST visualization

Model interpretability use case

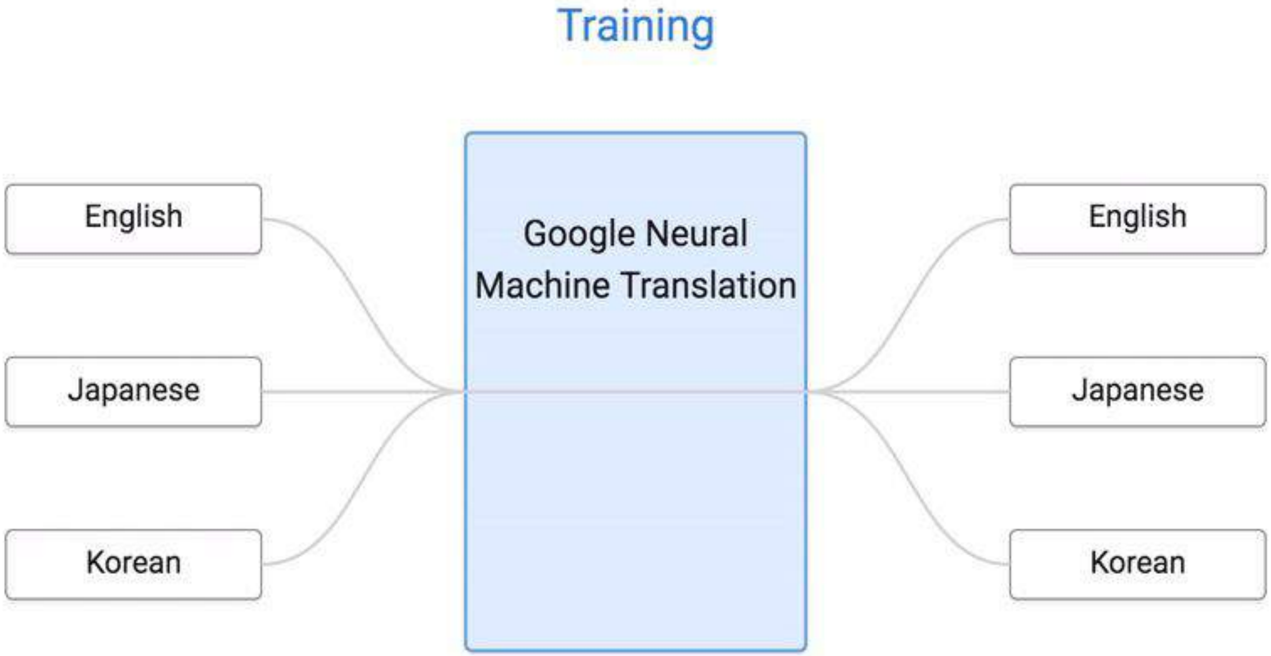
Multi-lingual translation

What does the language embedding space look like?

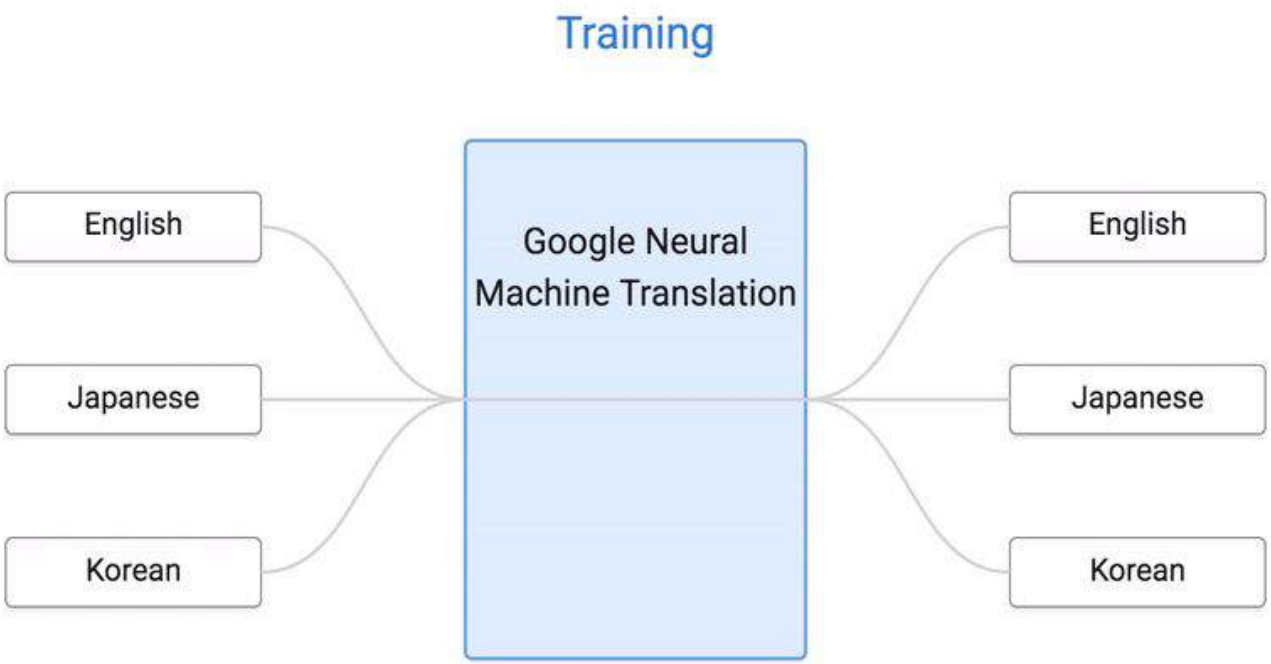
https://arxiv.org/abs/1611.04558
Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, Jeffrey Dean

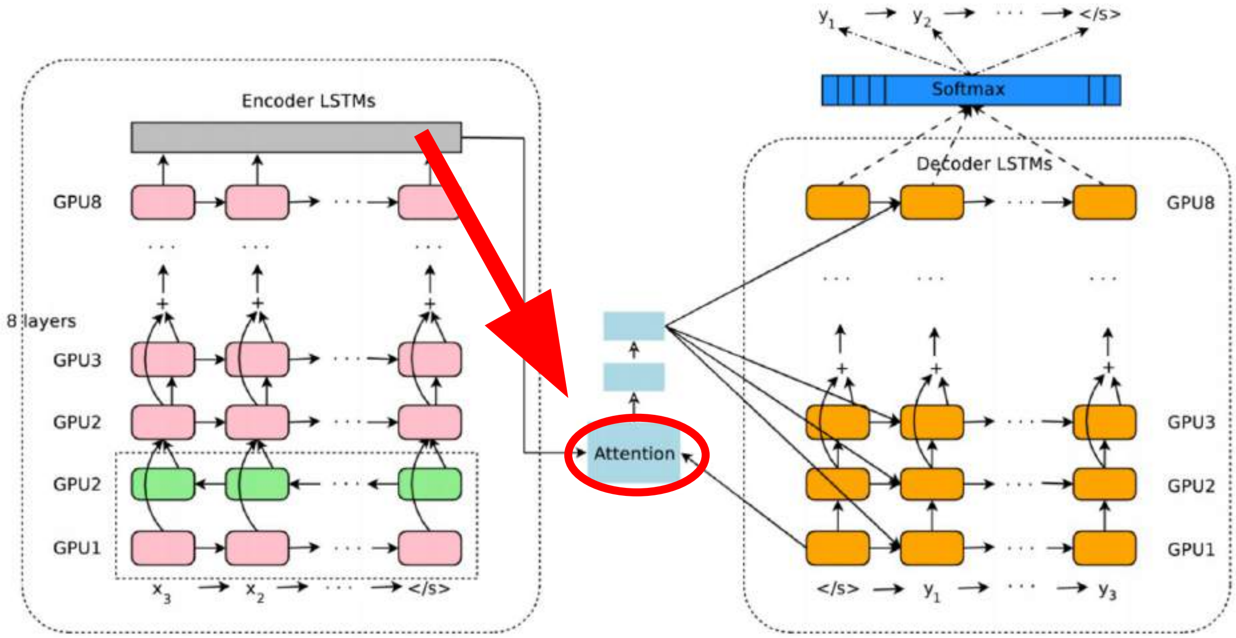
Training: English ⟵ ⟶ Japanese
English ⟵ ⟶ Korean



Training: English ⟵ ⟶ Japanese
English ⟵ ⟶ Korean
Japanese ⟵ ⟶ Korean (zero shot)

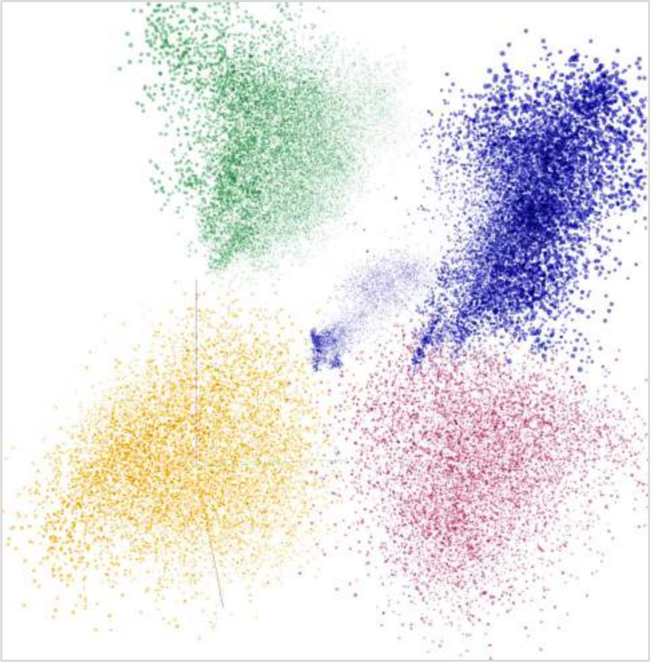


Visualize internal representation ("embedding space")

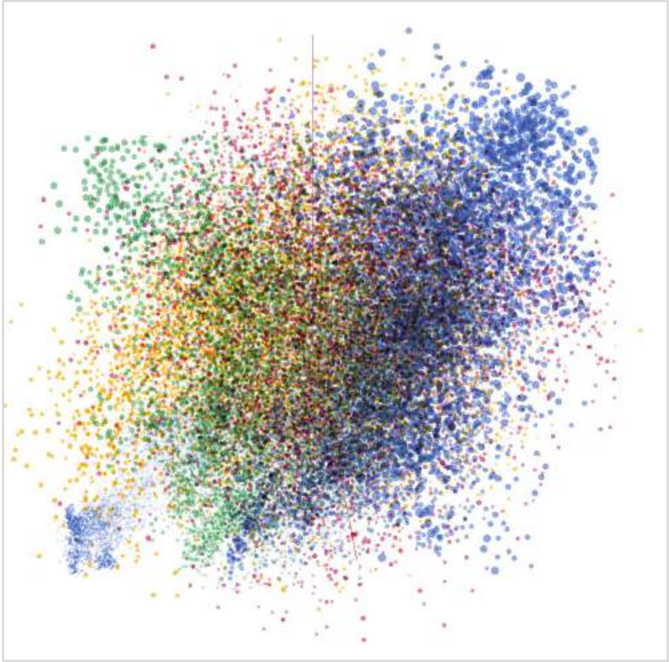


Research question

What does the multi language embedding space look like?



or



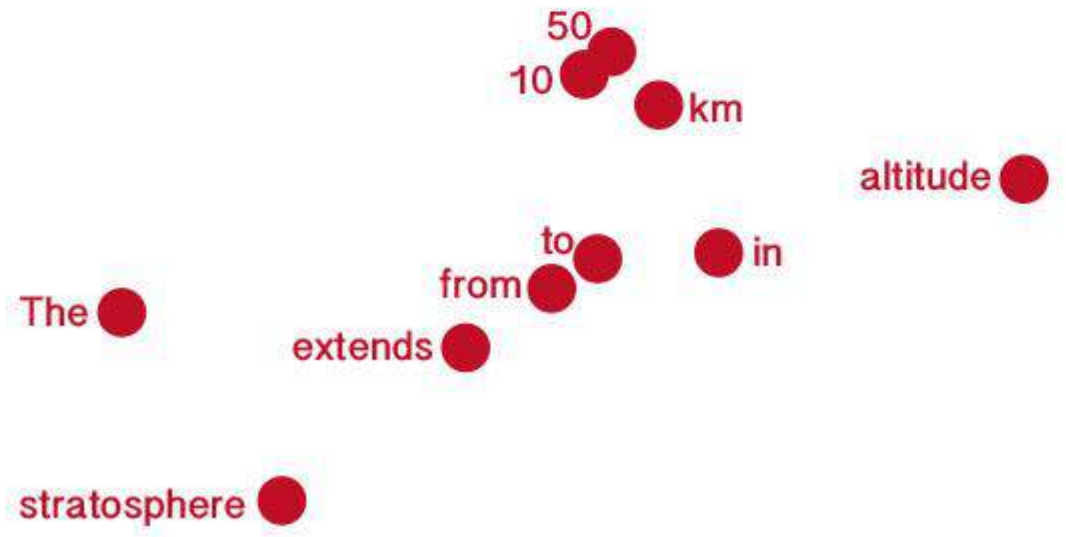
Note: not real data

What does a sentence look like in embedding space?

(points in 1024-dim space: the data that the decoder receives)

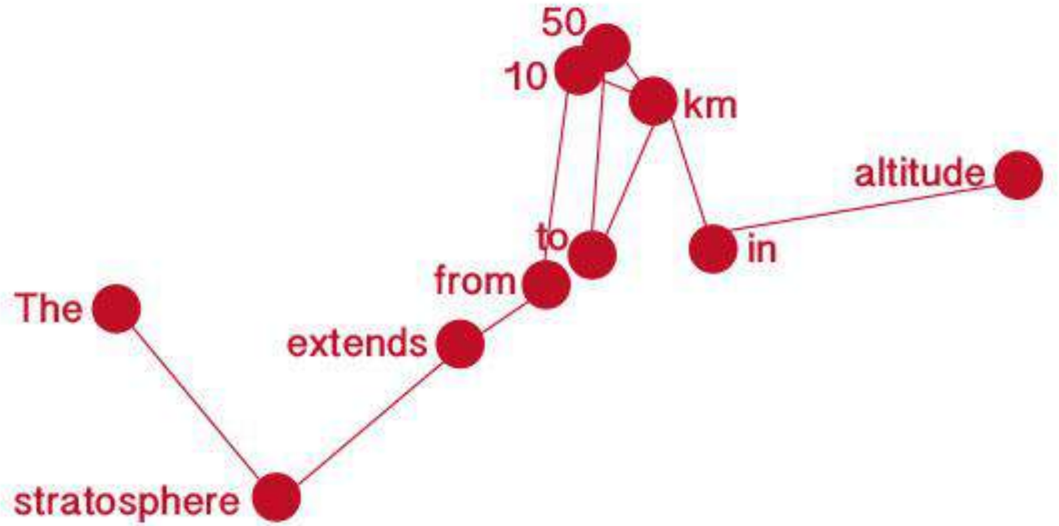
E.g. "*The stratosphere extends from 10km to 50km in altitude*"

What does a sentence look like in embedding space?



Note: simplification of real situation!

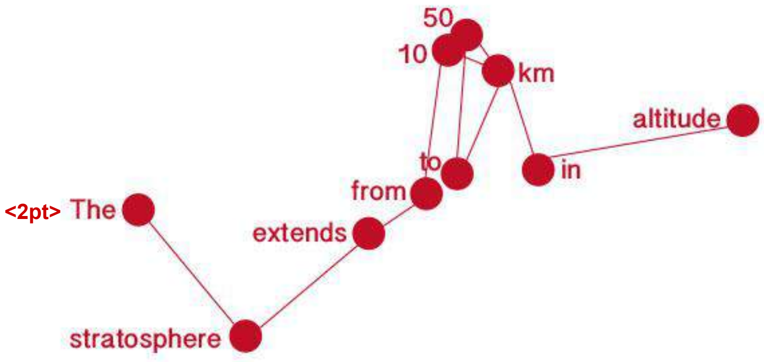
What does a sentence look like in embedding space?

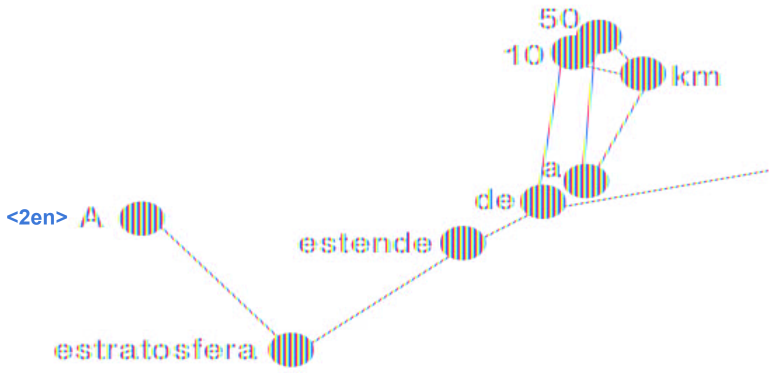


What do parallel sentences look like in embedding space?

(same meaning, different language)

like this?



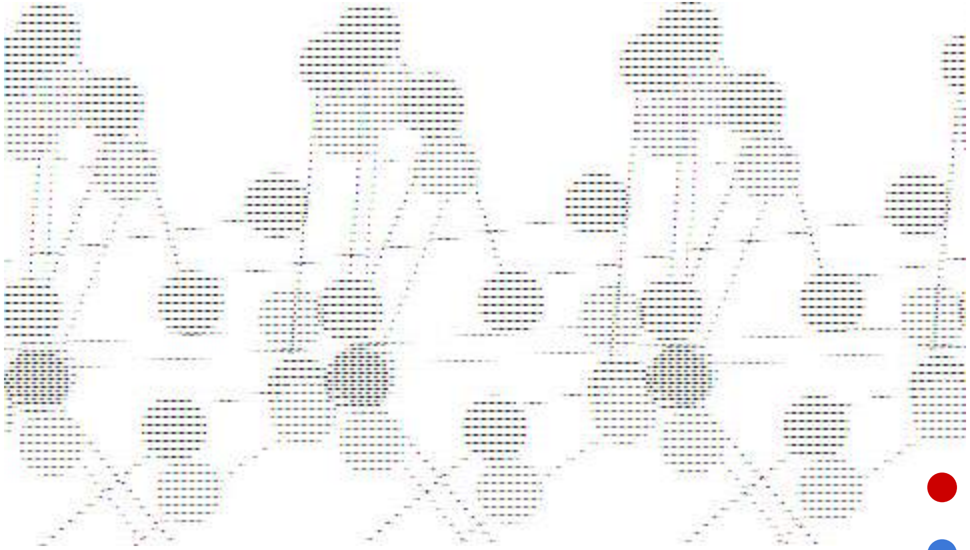


● English
● Portuguese

What do parallel sentences look like in embedding space?

(same meaning, different language)

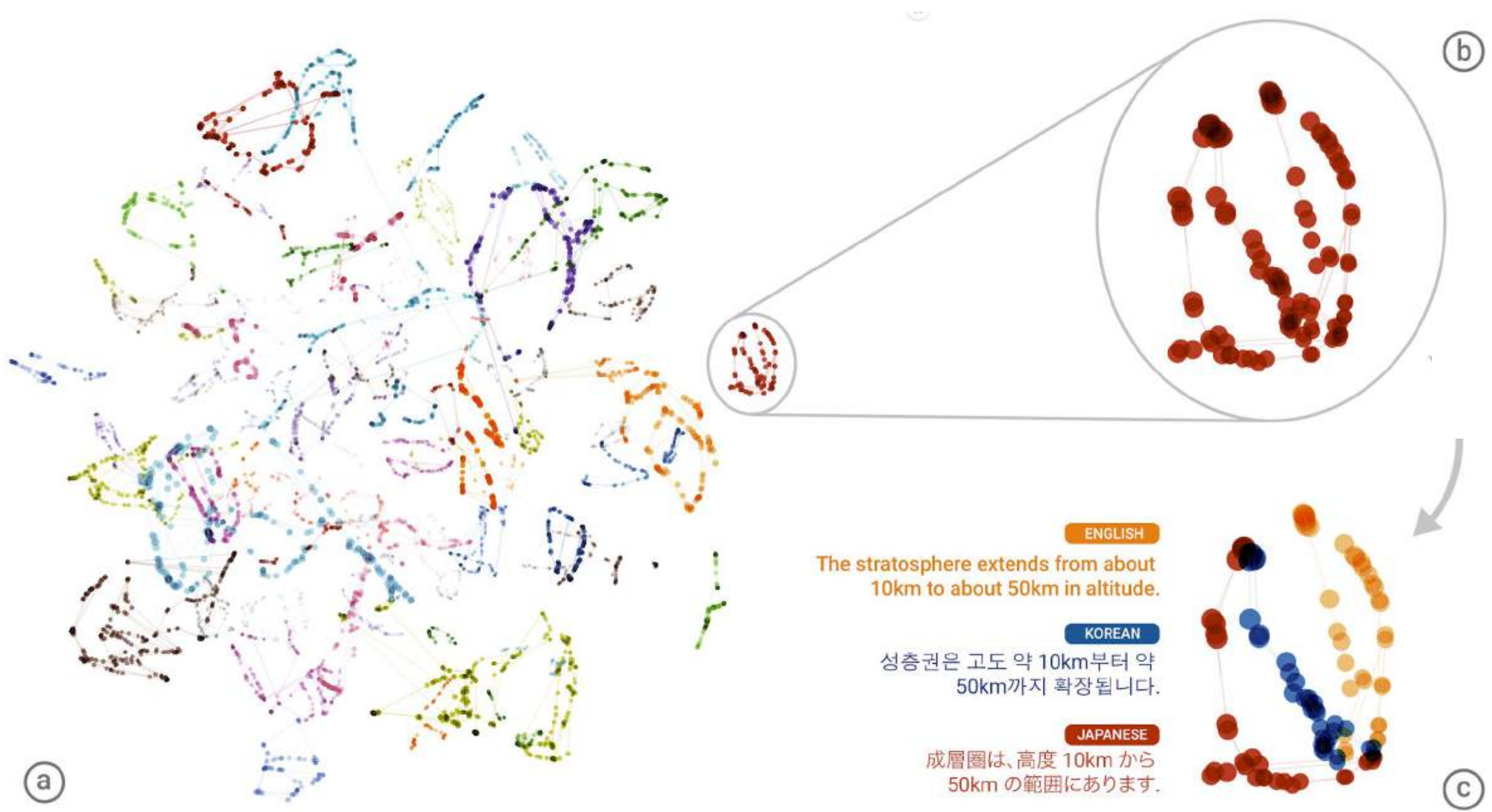
or like this?



● English
● Portuguese

Interlingua?

Sentences with the same meaning mapped to similar regions regardless of language!



Distance between bridge / non-bridge sentences is inversely related to translation quality

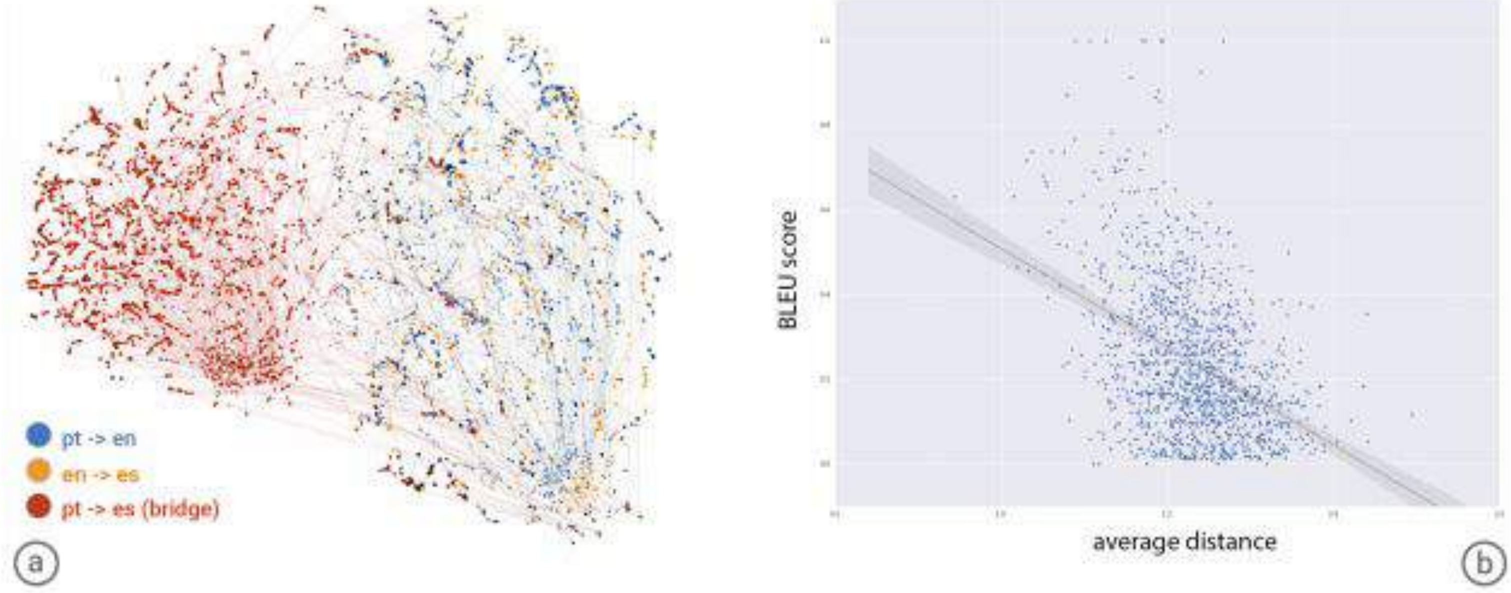
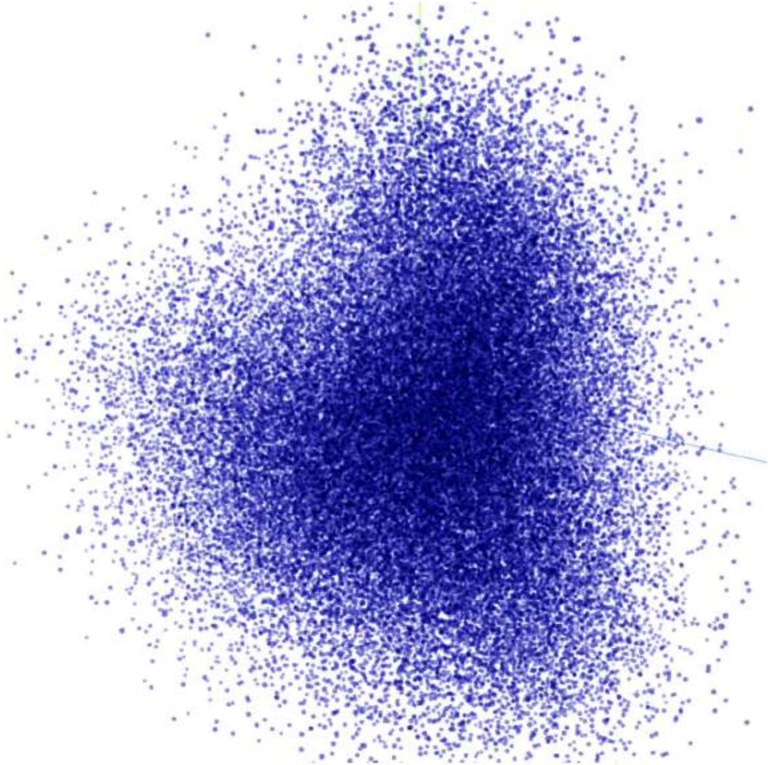
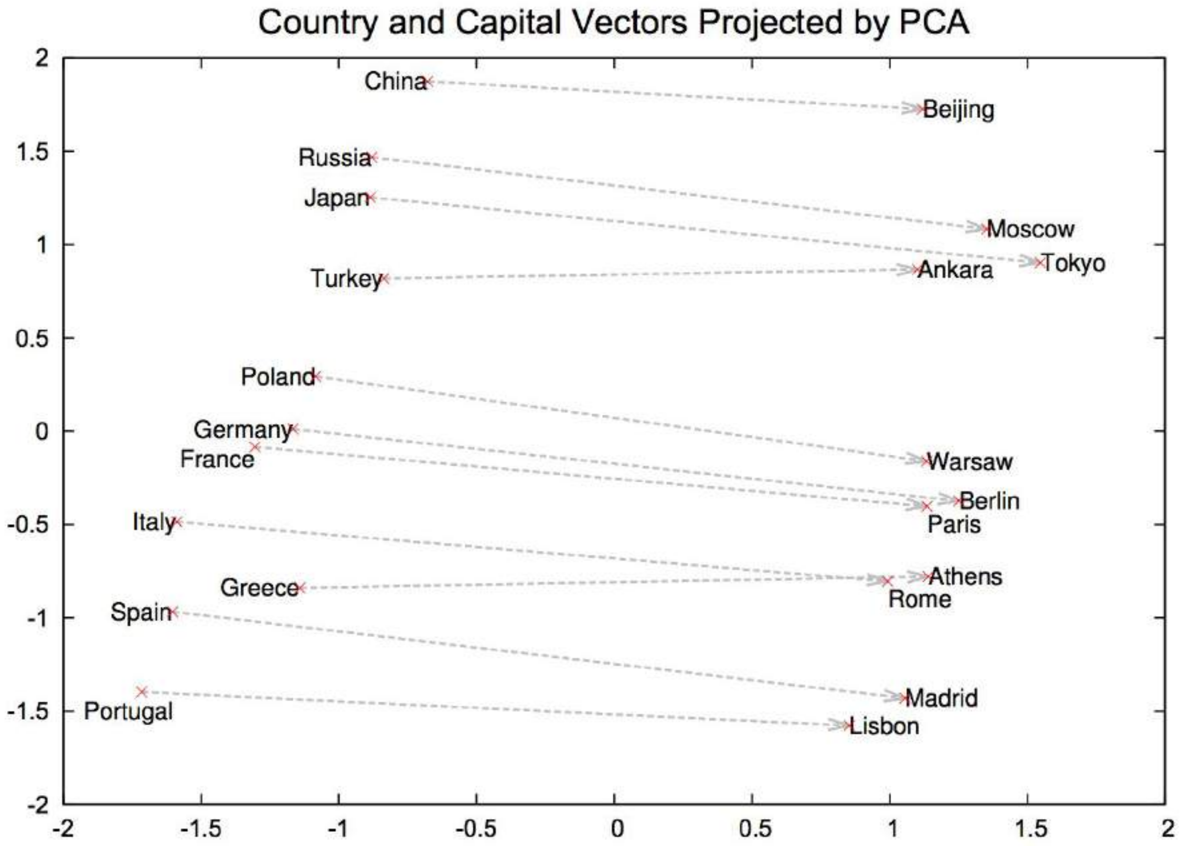


Figure 3: (a) A bird's-eye view of a t-SNE projection of an embedding of the model trained on Portuguese→English (blue) and English→Spanish (yellow) examples with a Portuguese→Spanish zero-shot bridge (red). The large red region on the left primarily contains the zero-shot Portuguese→Spanish translations. (b) A scatter plot of BLEU scores of zero-shot translations versus the average point-wise distance between the zero-shot translation and a non-bridged translation. The Pearson correlation coefficient is −0.42.

word embeddings

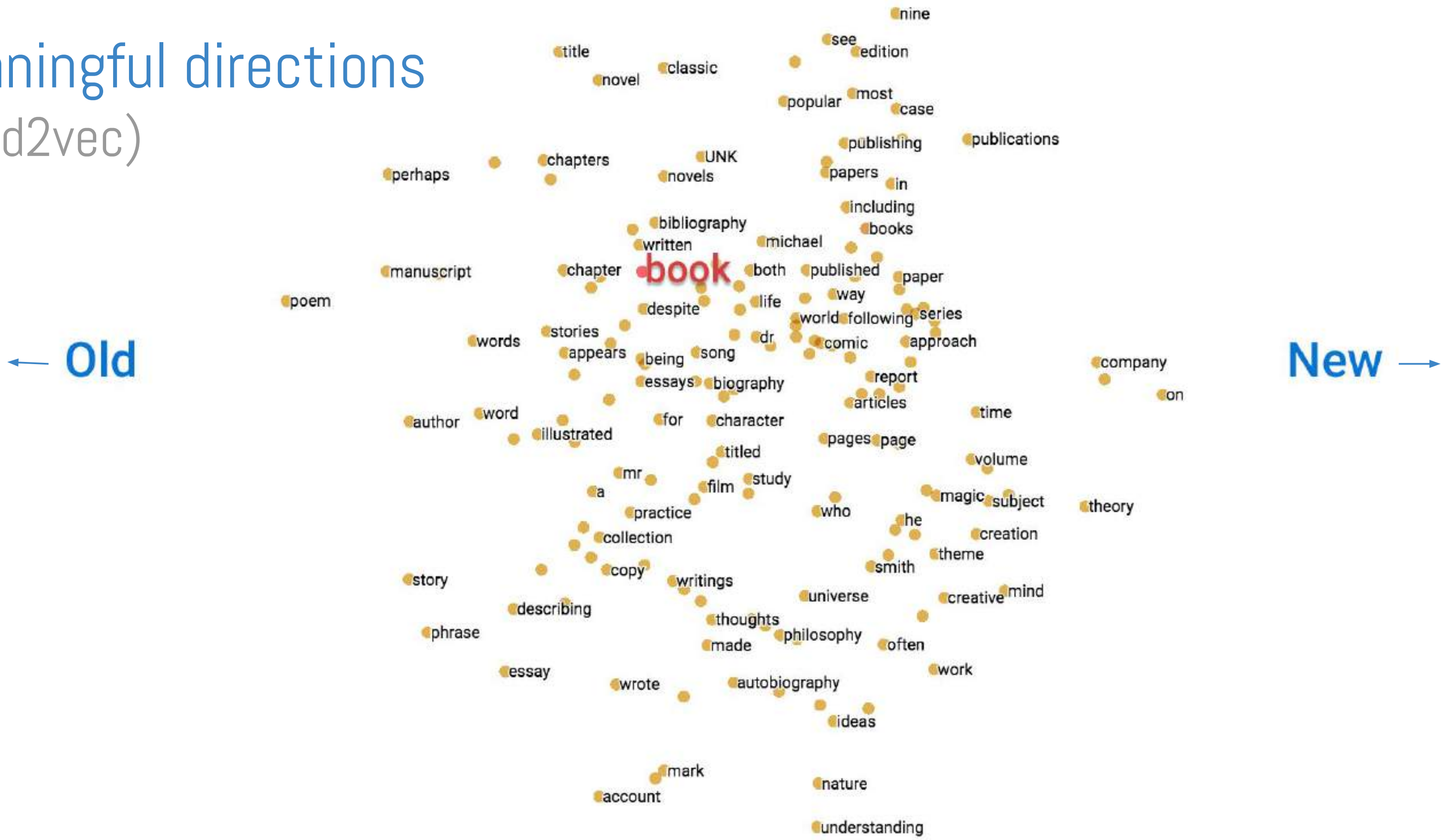


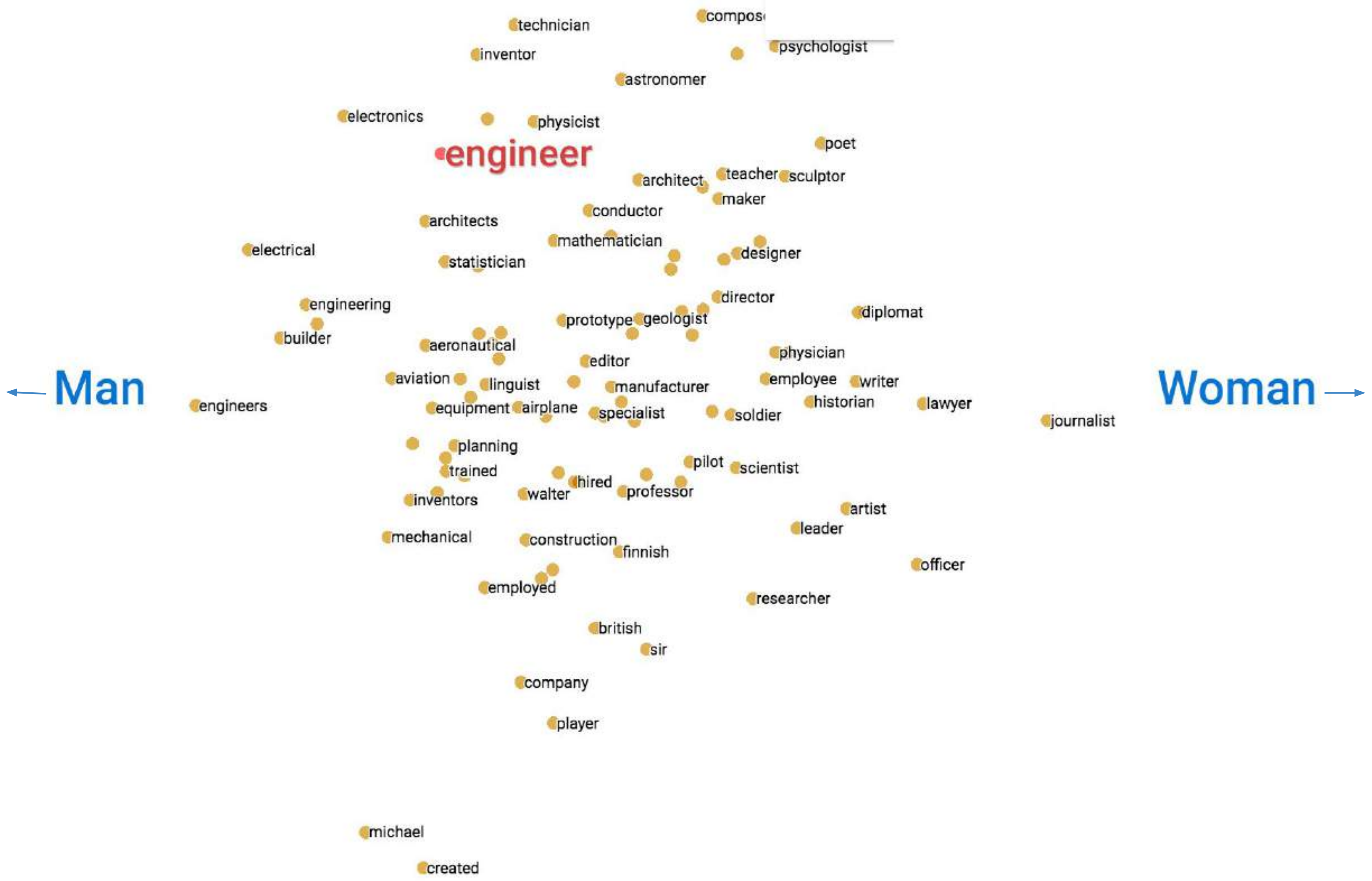
Word embeddings



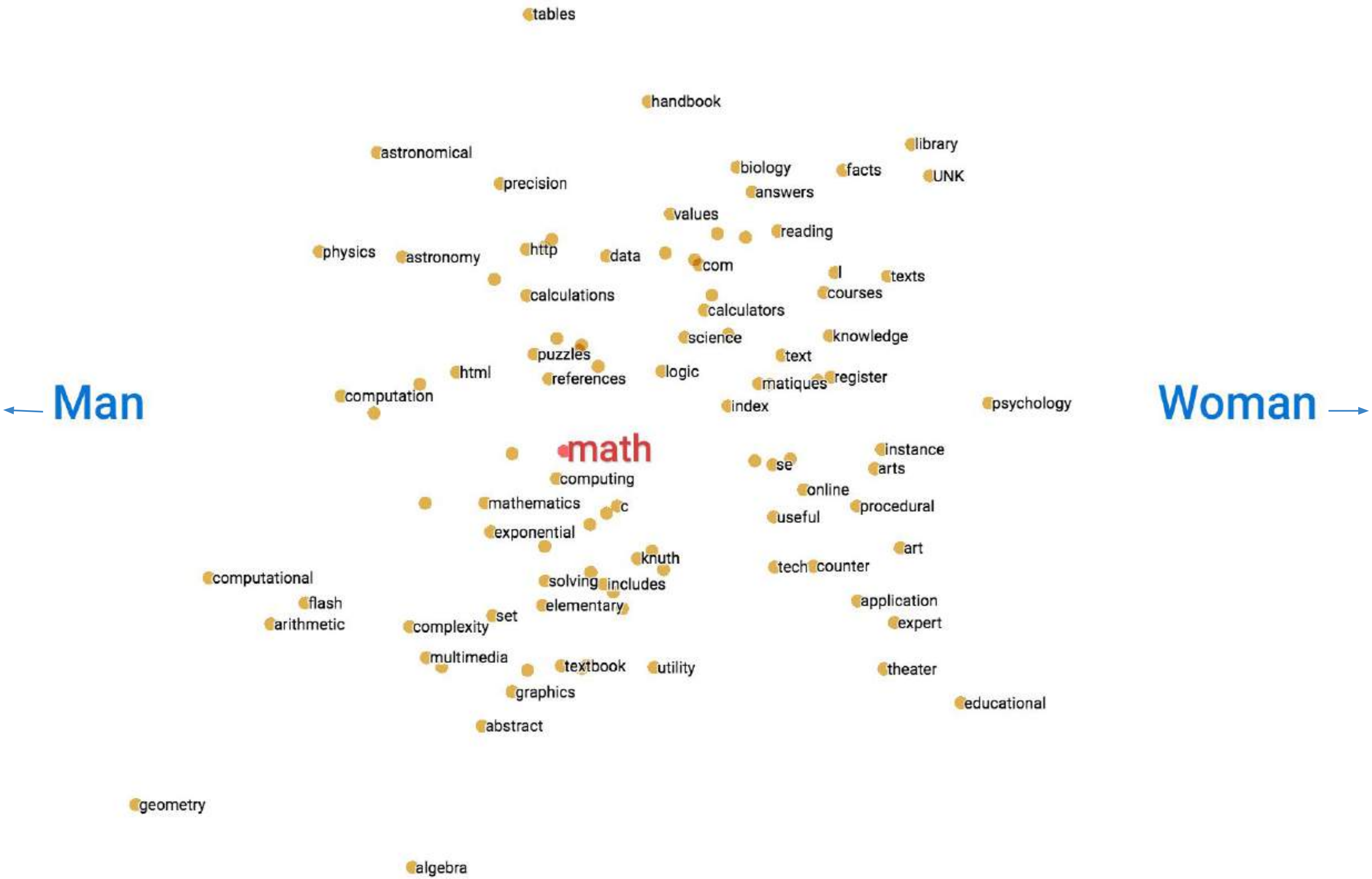
Distributed Representations of Words and Phrases and their Compositionality
Mikolov et al. 2013

Meaningful directions
(word2vec)

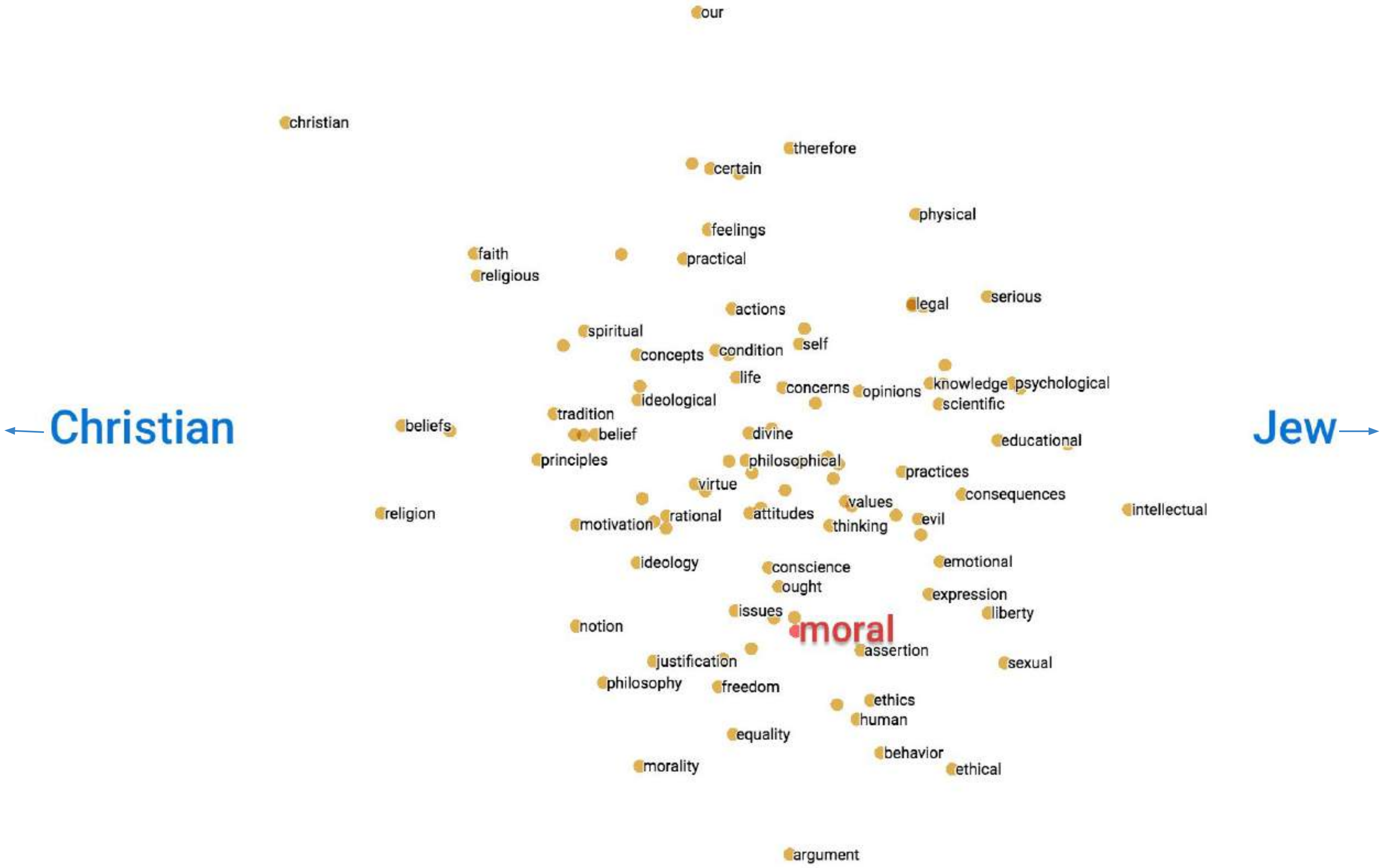




technician
compos
psychologist
inventor
astronomer
electronics
physicist
poet
engineer
architect
teacher
sculptor
maker
conductor
architects
mathematician
electrical
statistician
designer
director
engineering
diplomat
prototype
geologist
builder
aeronautical
physician
editor
Man
Woman
aviation
linguist
manufacturer
employee
writer
engineers
equipment
airplane
specialist
soldier
historian
lawyer
journalist
planning
pilot
scientist
trained
hired
walter
professor
inventors
artist
leader
mechanical
construction
finnish
officer
employed
researcher
british
sir
company
player
michael
created



tables
handbook
astronomical
library
biology
facts
UNK
precision
answers
values
reading
physics
astronomy
http
data
com
l
texts
calculations
courses
calculators
science
knowledge
puzzles
text
html
logic
references
register
matiques
computation
index
psychology
Man
Woman
math
instance
se
arts
computing
online
mathematics
c
procedural
useful
exponential
art
knuth
tech
counter
computational
solving
includes
flash
elementary
application
arithmetic
set
complexity
expert
multimedia
textbook
utility
theater
graphics
educational
abstract
geometry
algebra


our
christian
therefore
certain
physical
feelings
faith
practical
religious
actions
legal
serious
spiritual
self
concepts
condition
life
concerns
opinions
knowledge
psychological
ideological
scientific
Christian
tradition
beliefs
belief
divine
educational
Jew
principles
philosophical
practices
virtue
consequences
values
religion
rational
attitudes
evil
intellectual
motivation
thinking
emotional
ideology
conscience
ought
expression
liberty
issues
notion
moral
assertion
justification
sexual
philosophy
freedom
ethics
human
equality
behavior
morality
ethical
argument

Data Visualization for Machine Learning

Fernanda Viégas @viegasf
google.ai/pair