

A person wearing a yellow costume is blowing a large stream of colorful, iridescent bubbles. The bubbles are of various sizes and are floating in the air, creating a festive and playful atmosphere. The background is slightly blurred, showing green foliage and a wooden fence.

Semantic hide and seek - a gentle introduction to embeddings

GenAI Network Melbourne
September 2023
Dave Colls

Acknowledgement of country

Dave Colls

 @davidcolls

 @safetydave

 safetydave.net



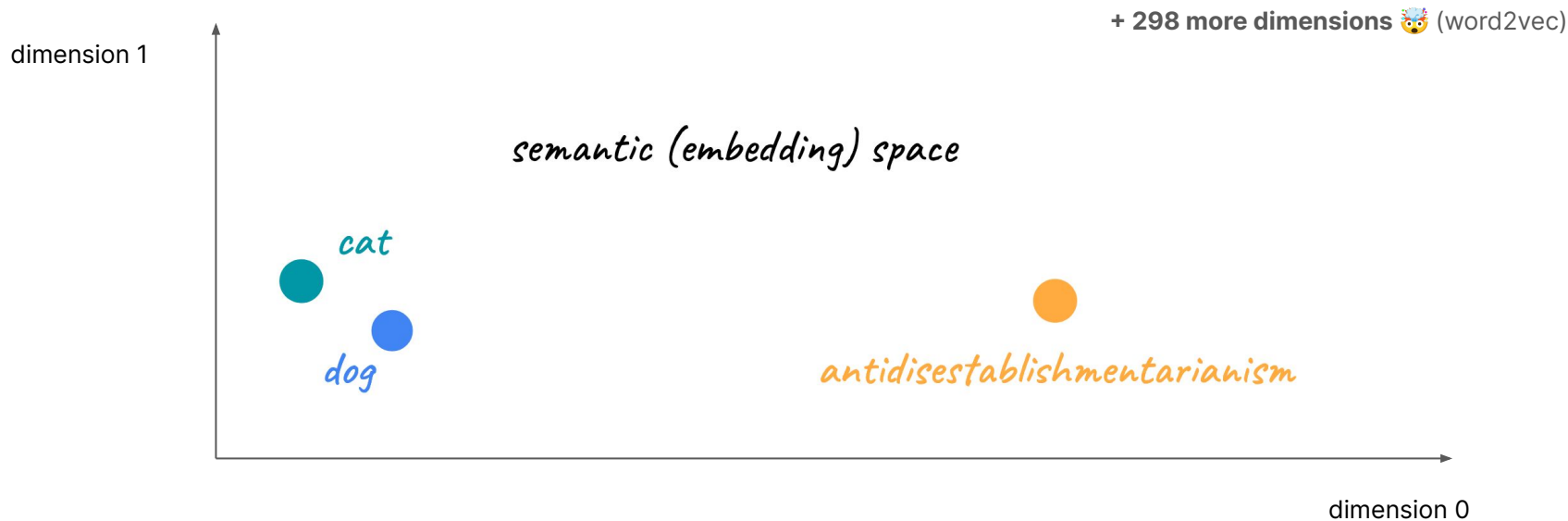
Background on embeddings

Embeddings

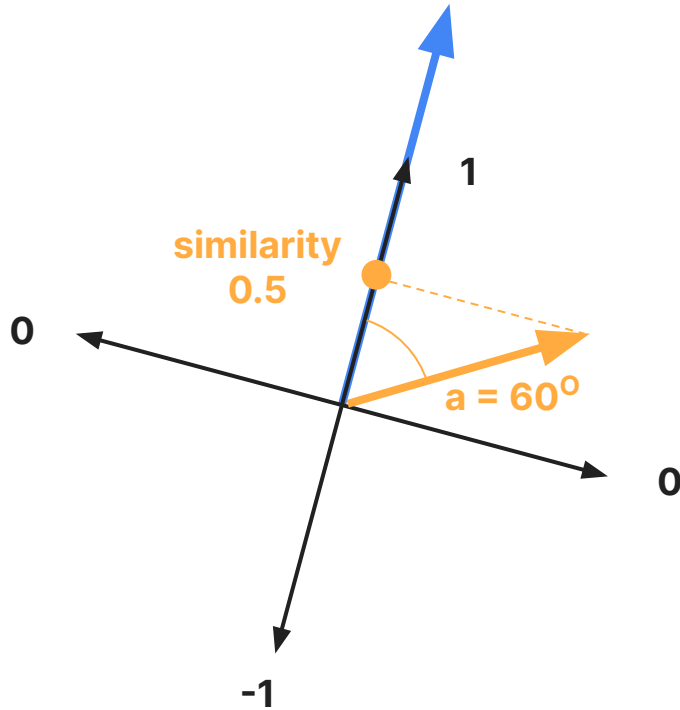
aka semantic space, latent space, vector, etc

A list (vector) of numbers [0.5, 0.77, 0.03, ...]

Similar things are “close” together

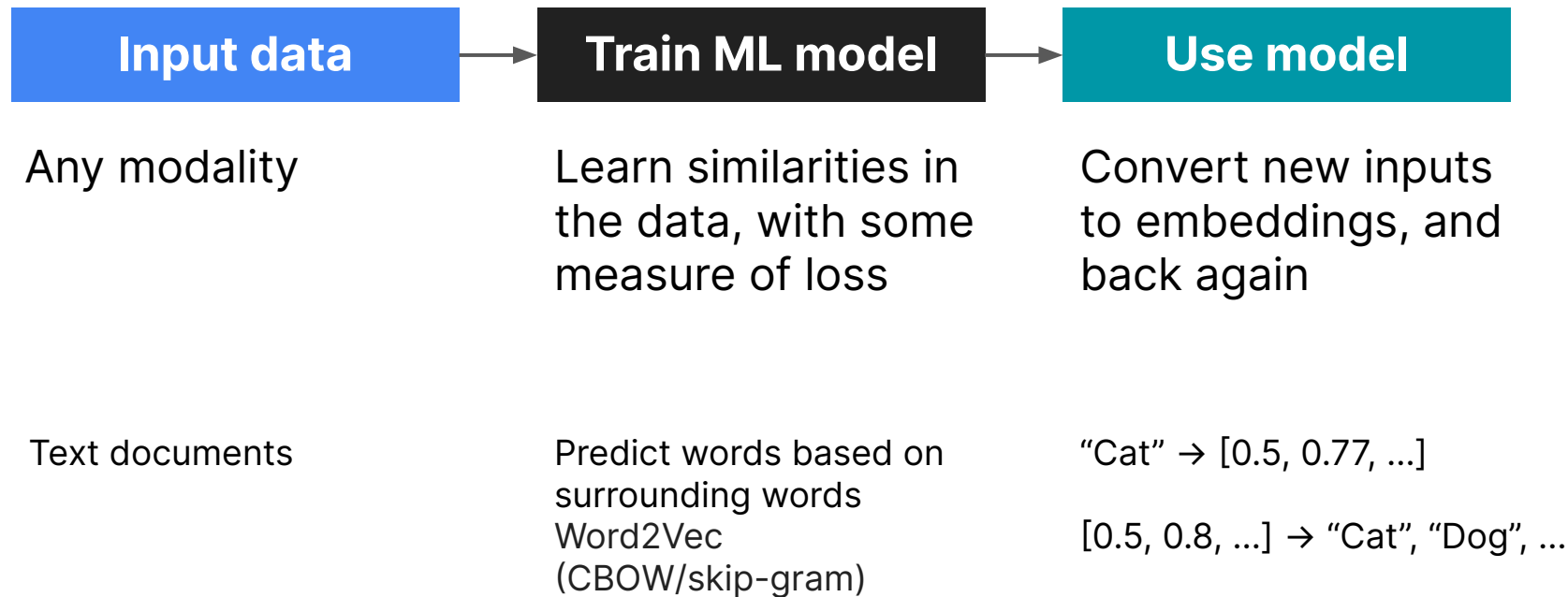


Measuring similarity with cosine



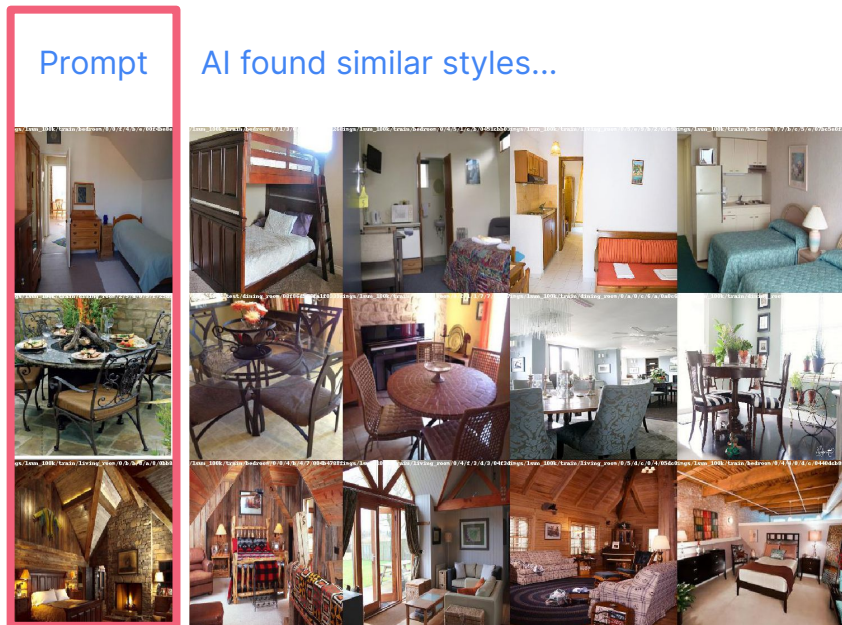
Cosine similarity puns are sticky

Making embeddings

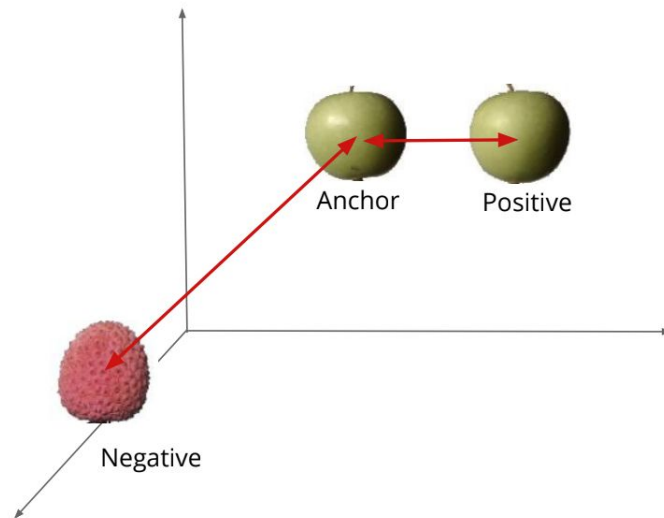


Embeddings in Thoughtworks ANZ projects

This time with images & contrastive learning/triplet loss



“RoomNet” spike



“ProduceNet” representation learning

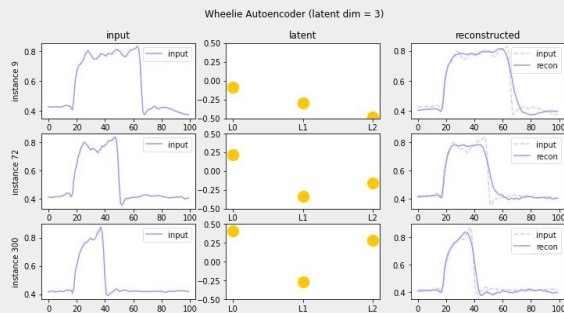
Embeddings in personal projects

This time with time series accelerometer data & autoencoders



The Lockdown
Wheelie Project

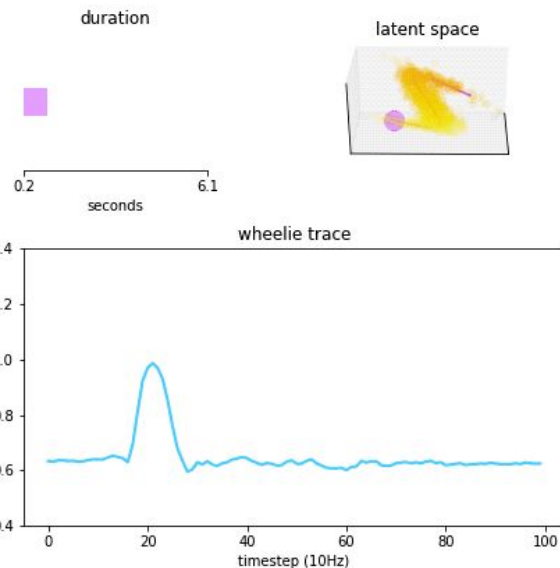
AI coach learns to describe a wheelie



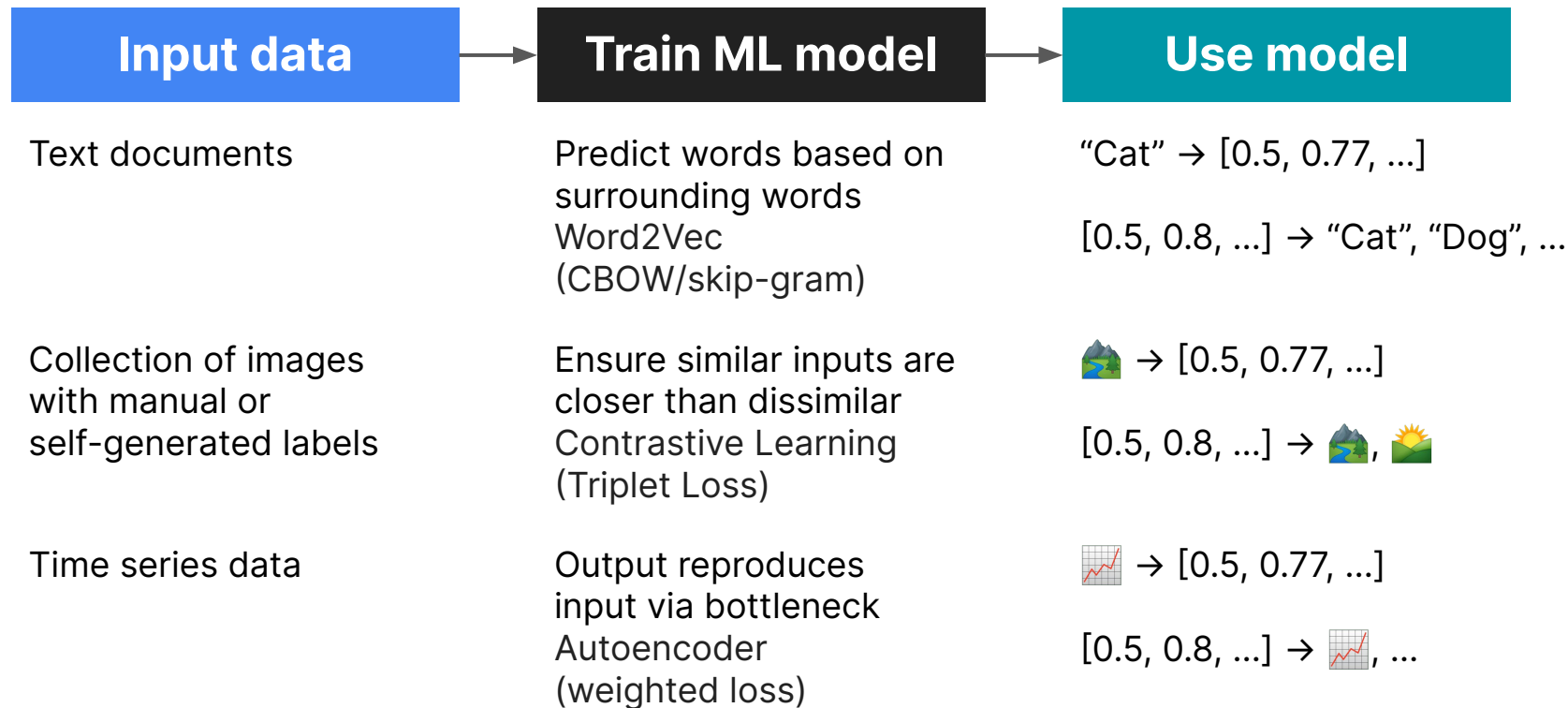
- L0 - From short to long duration
- L1 - Gradual or sharp landing
- L2 - Low & smooth or high & wiggly

aka WheelieGPA

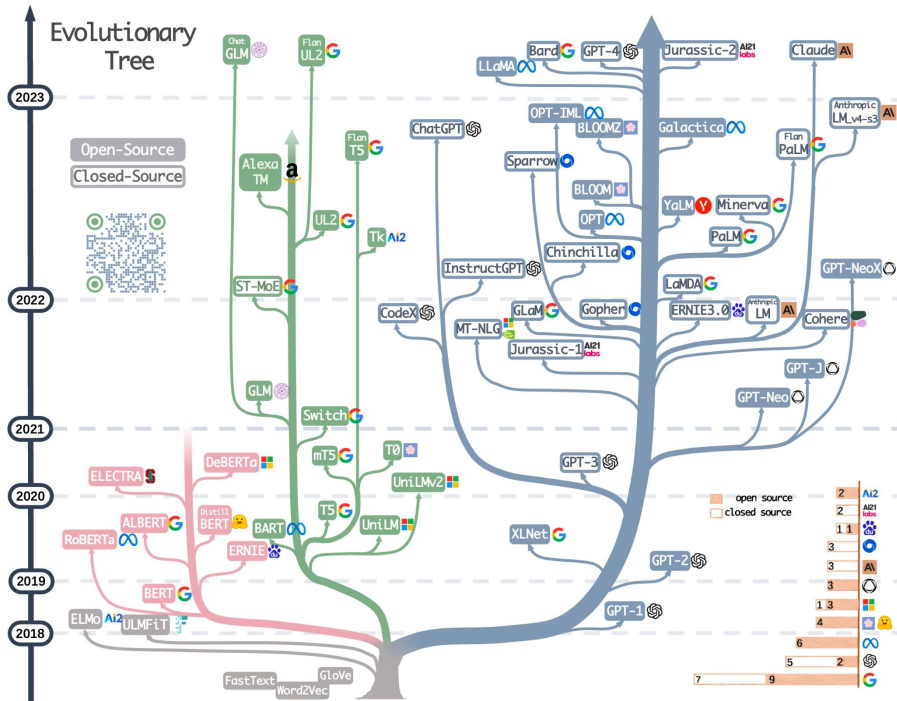
This wheelie does not exist



Making embeddings



Embeddings in LLMs



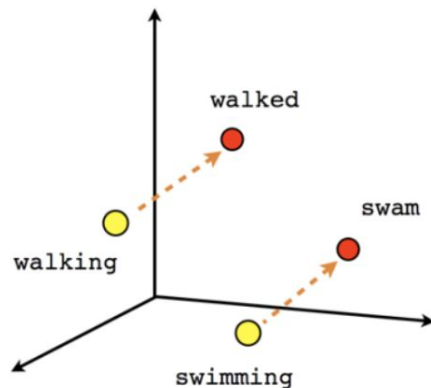
<https://github.com/Mooler0410/LLMsPracticalGuide>

```
curl https://api.openai.com/v1/embeddings \
-H "Content-Type: application/json" \
-H "Authorization: Bearer $OPENAI_API_KEY" \
-d '{
  "input": "Your text string goes here",
  "model": "text-embedding-ada-002"
}'
```

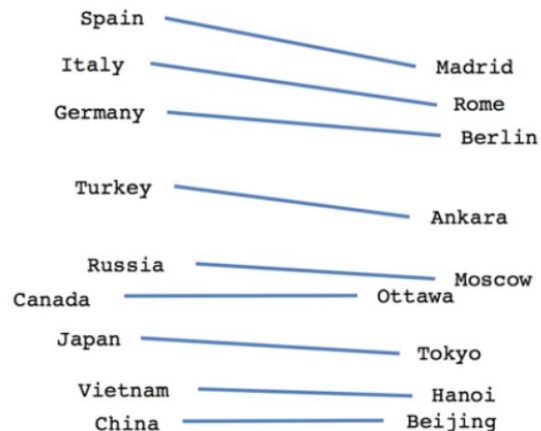
```
{
  "data": [
    {
      "embedding": [
        -0.006929283495992422,
        -0.005336422007530928,
        ...,
        -4.547132266452536e-05,
        -0.024047505110502243
      ],
      ...
    }
  ]
}
```

Semantic algebra with embeddings

Kinda handy, like we might do with spatial vectors



Verb tense

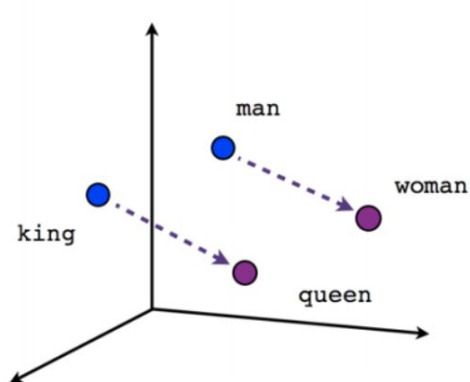


Country-Capital

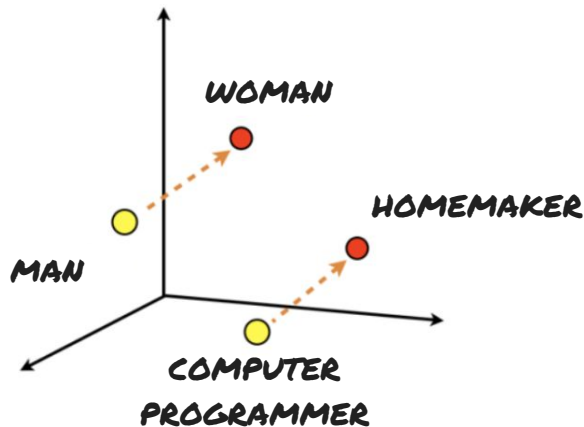
But may come with bias and harmful results

Motivating careful curation of training data or de-biasing treatment

“Sentiment
vector”



Male-Female



ISMS

<https://thegradients.pub/nlp-imagenet/> (cropped & edited)
<https://blog.kjamistan.com/embedded-isms-in-vector-based-natural-language-processing/>

Someone made a game of hide and seek
from embeddings...

Semantle - a game of semantic hide and seek

Semantle



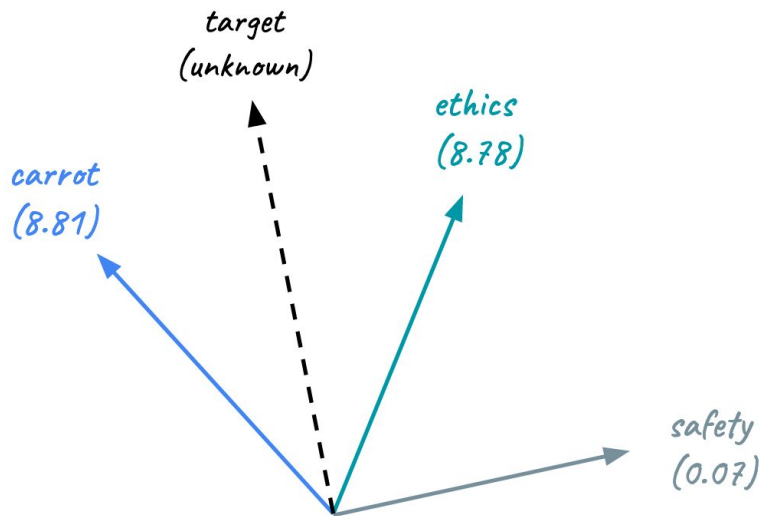
Today is puzzle number **494**. The nearest word has a similarity of **83.89**, the tenth-nearest has a similarity of 45.49 and the one thousandth nearest word has a similarity of 19.85.

Guess

#	Guess	Similarity	Getting close?
7	vegetable	0.49	(cold)
1	carrot	8.81	(cold)
3	ethics	8.78	(cold)
6	embedding	5.83	(cold)
4	running	3.66	(cold)
2	horse	1.47	(cold)
5	safety	0.07	(cold)

Hint

Give up



Any given Friday



Data Service Line Core Team



Semantle

Anyone up for a game of Semantle at 4pm?

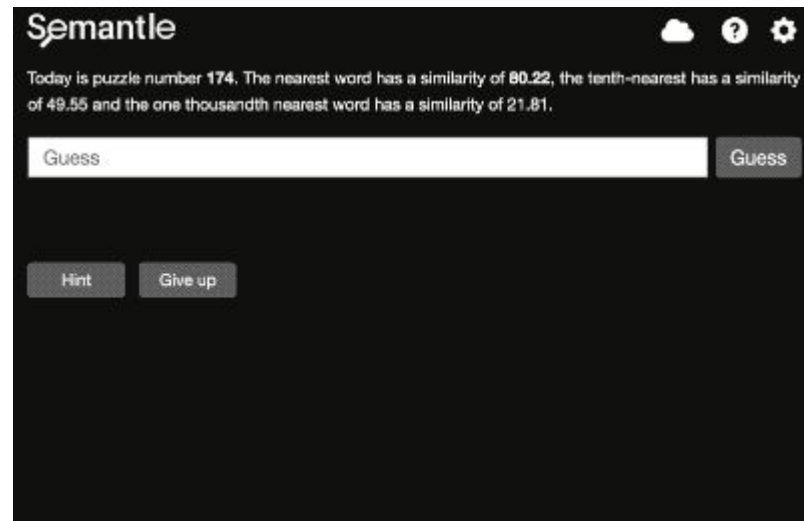
... and you'll never guess what I did next

(In the world before ChatGPT)

Automated and augmented semantle solvers

Personal project aims

- ~~Fu~~ Play around and find out
- Get a better feeling for how embeddings capture semantics
- Find solutions automatically - with different approaches - but also...
- Explore how people and machines work together on problems



<https://safetydave.net/sketching-semantle-solvers/>
<https://safetydave.net/second-semantle-solver/>
<https://safetydave.net/synthesising-semantle-solvers/>
<https://safetydave.net/smarter-semantle-solvers/>

Let's play live!

Oh I'm asking for trouble

This browser (screen shared)

✓ Setup:

- Go to <https://semantle.com>
- Select play with friends
- Get game code and share

👁 Watch the demo:

- Make your own guesses if you like

Solver browser (playing same game)

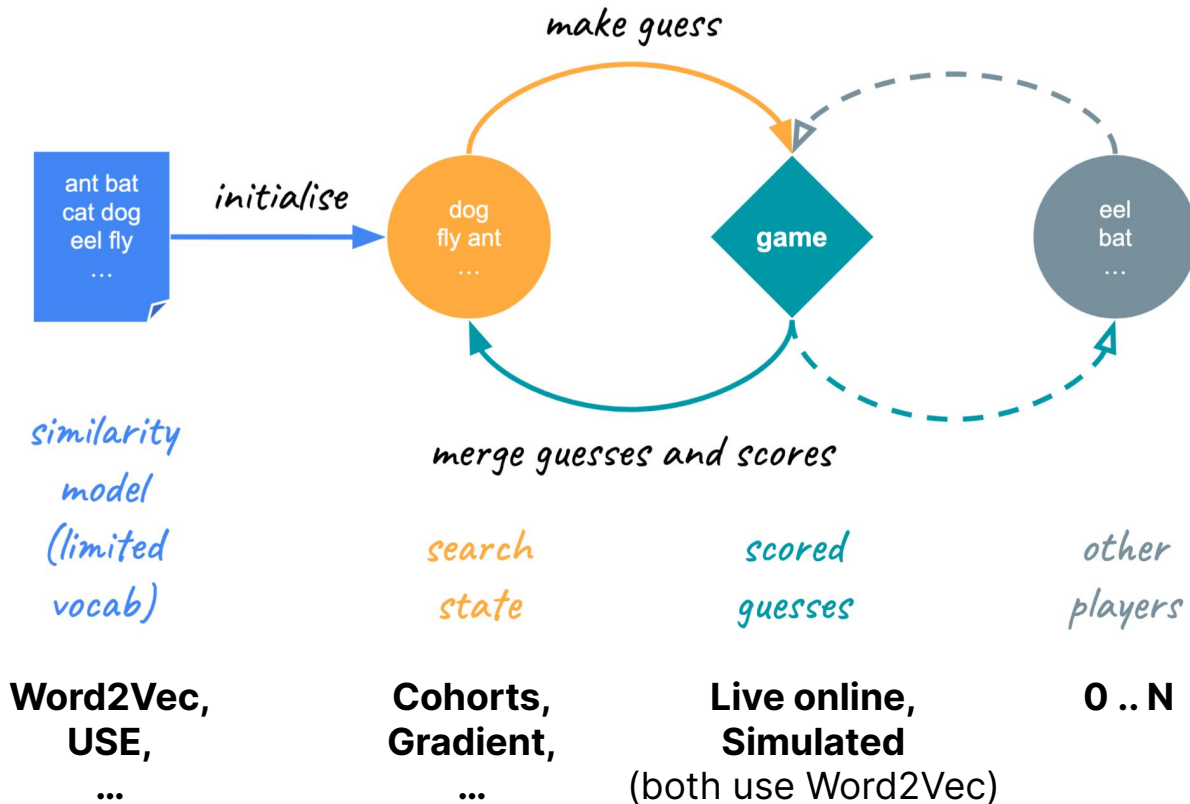
✓ Setup:

- Run notebook, which loads <https://semantle.com> with webdriver
- Manually select play with friends
- Manually enter game code

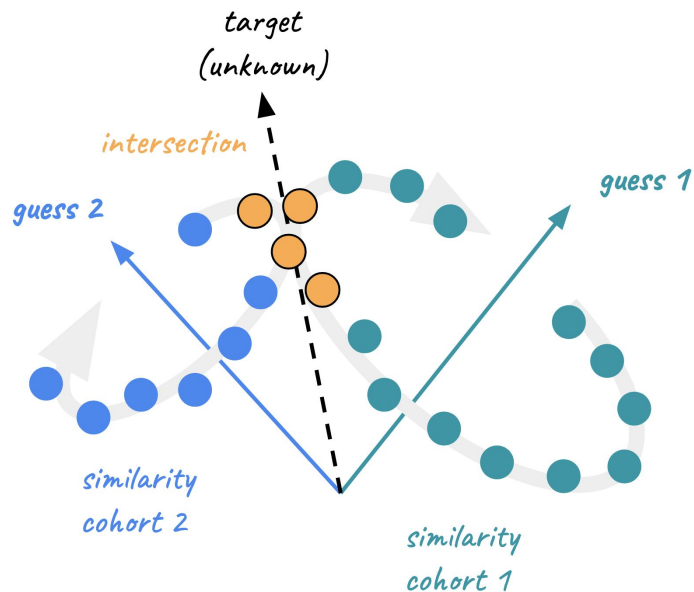
🎬 Start the demo:

- Run the solver method which will make guesses based on the game state

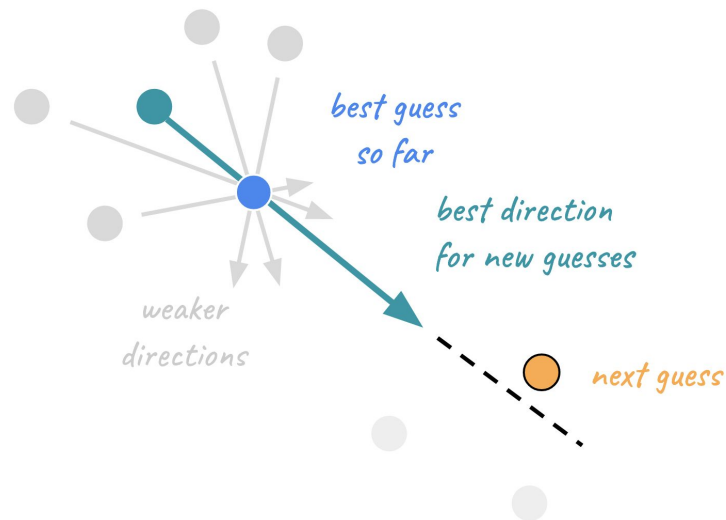
Simplified RL approach with modular components



Two semantic search strategies - conceptual



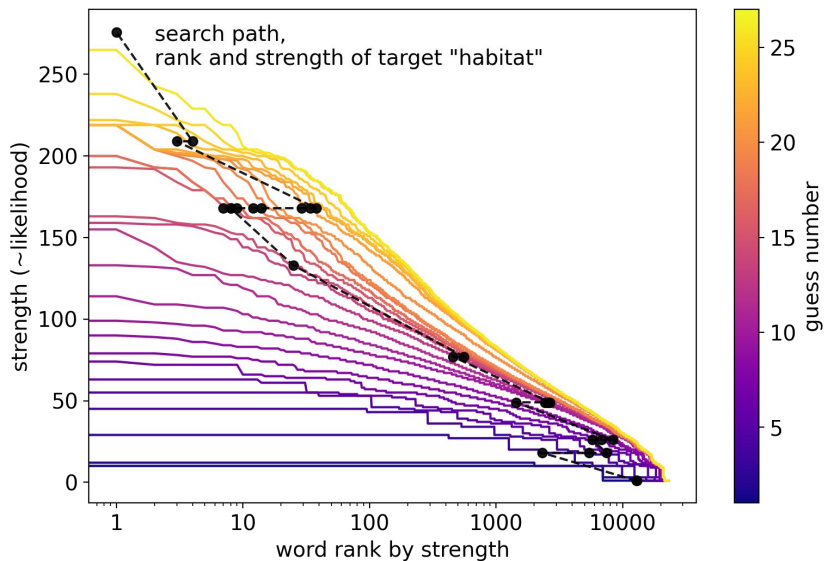
Cohorts



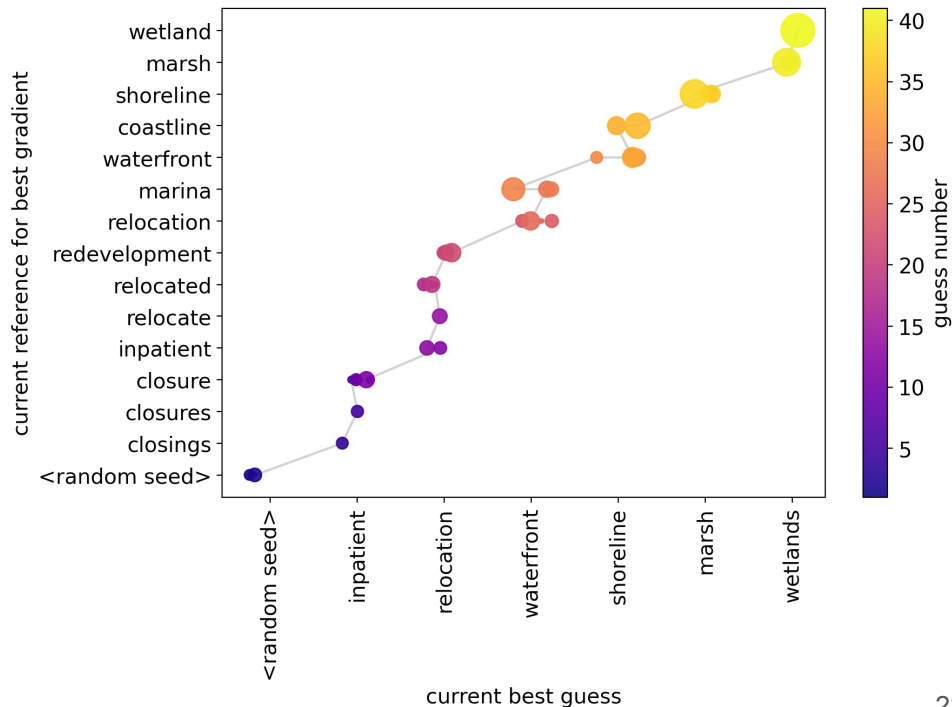
Gradient

Two semantic search strategies - visualised

cohort solver,
vocabulary strength distributions per guess




gradient solver, target "habitat",
basis of guess and similarity (size)




What about ChatGPT?

Obviously, we must ask ChatGPT to play

Just prompting for now


 You are playing a game called semantle, where you must discover a hidden word by making guesses. For each guess you will get a score out of 100 that measures how close the guess is to the hidden word, based on word2vec cosine similarity. What is your first guess?

Intro game


 My first guess in the game Semantle would depend on the context and any hints or clues provided. Without any specific information, I will make a random guess. Let's go with the word "sunshine."

*No context,
random guess*


 The score for "sunshine" is 6.74

 Based on the score of 6.74 for the word "sunshine," it seems that the hidden word is not closely related to "sunshine." To make a more informed guess, I would need additional information or context about the hidden word. Could you provide any hints or clues to narrow down the possibilities?

Moar hints pls!


 Ok. I will make guesses too, and give you the result in the format "guess: score". After each of my guesses, you should make another guess

I'll play too

 Sure, I'm ready to continue the game. Please provide your first guess, and I'll give you the result in the format "guess: score." After that, I'll make my next guess based on the information provided. Let's proceed!

 civilization: 20.01

Helpful clue

 Your guess: civilization
Score: 20.01

Based on the given score, it appears that the hidden word is somewhat related to "civilization." Here's my guess:

My guess: history

*Semantically
similar guess*

ChatGPT finds semantle hard too

Anyone who's played can relate to this



Thank you for the score. Based on the low score for "Duration," it seems that it is not close to the hidden word.

Gets stuck in local maxima

Since this is a word-guessing game and the scores provided have not indicated a close match yet, it is challenging to make an accurate guess without additional information or context.








Vents about Semantle (don't anthropomorphise though)

Could you please provide any hints or clues related to the hidden word? It would greatly help in making a more informed guess.

What can we learn from
playing semantic hide and seek?

Team heuristics

Things we learned

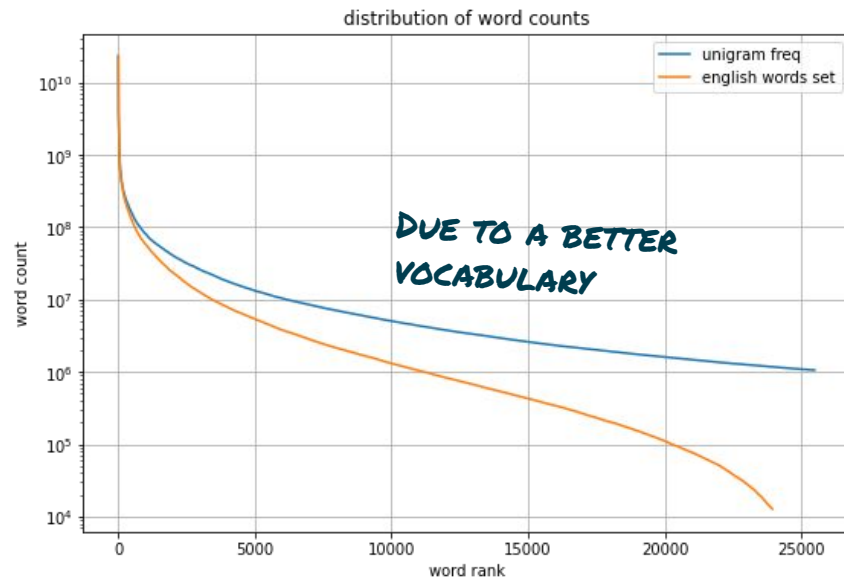
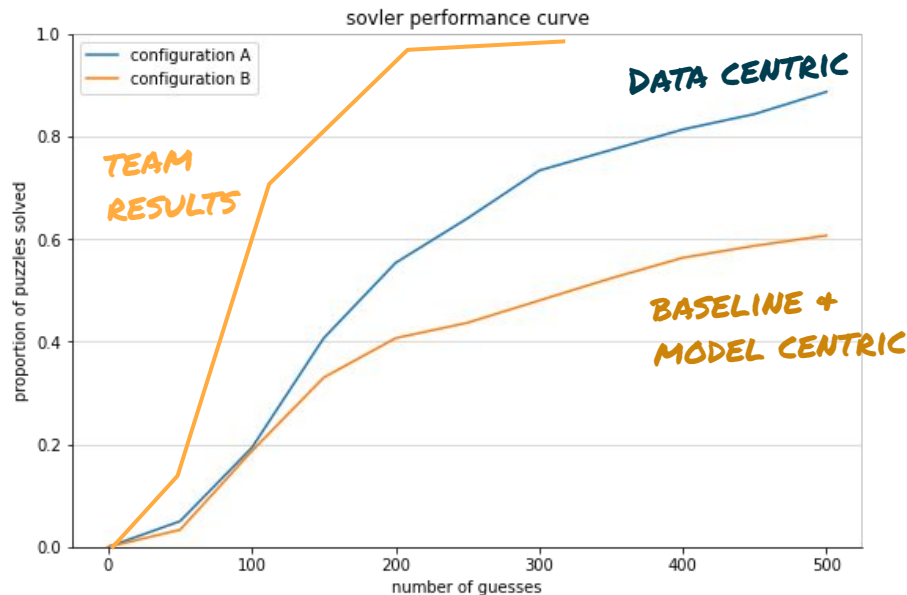
- Play on a Friday afternoon with the Data & AI team 
- Start with “carrot” (no empirical basis for effectiveness) 
- It’s usually not a complicated word 
- Helpful search strategies:
 - Diverse styles 
 - Antonyms may be similar, not just synonyms 
 - Try different verb conjugations, etc 
 - Extrapolate a *sentiment vector* from ~1,000 word to ~100 word 
(typically this is only rationalised post-hoc)
 - Hints are OK after 300 guesses ☐

Robustness to varying semantics

	People semantics	Word2Vec similarity (same as Semantle)	USE similarity
People search	Somewhat inscrutable Somewhere from 13 to 307 guesses		
Cohort search		Arbitrarily good performance (cheat)	Significantly degraded performance
Gradient search		Almost as good as human performance	Minimal performance degradation

Data-centric improvement efforts

Beat model-centric



-122 DNOOBWS

Human and machine learn from each other

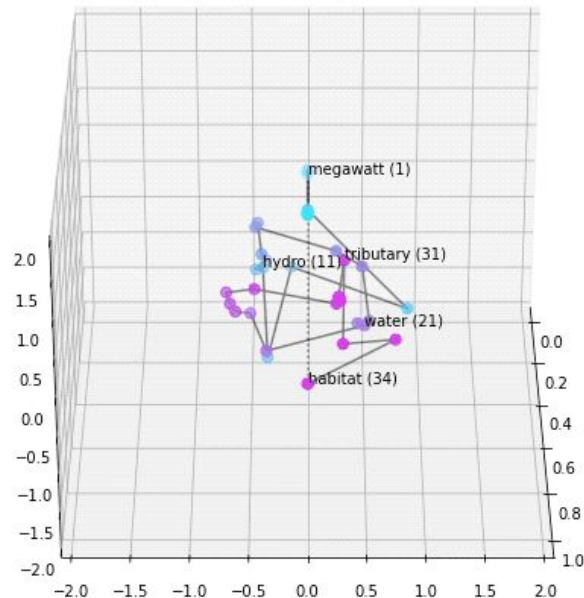
When playing together

One can unstick the other

The high-dimensional semantic search space is highly non-linear (it's hard to find the secret word unless you're already close, and you can get stuck on local maxima)

This is also true in designing products and services with many degrees of freedom - think software!

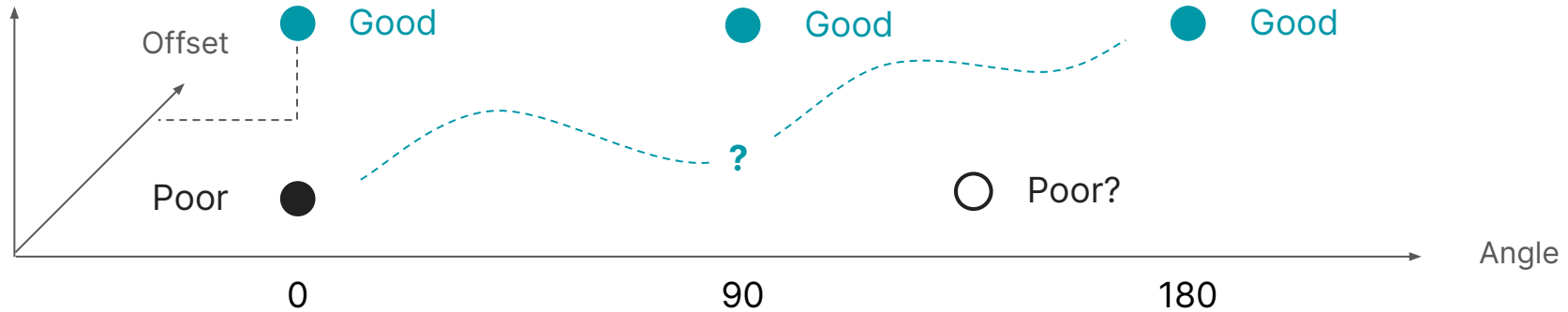
Social machines aid knowledge management



For unsmooth design spaces



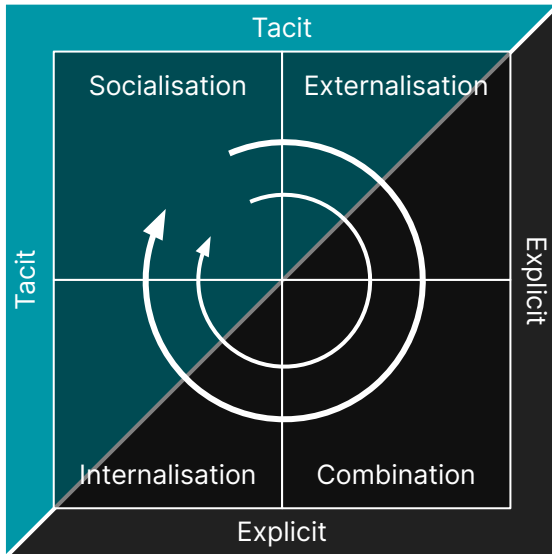
Performance



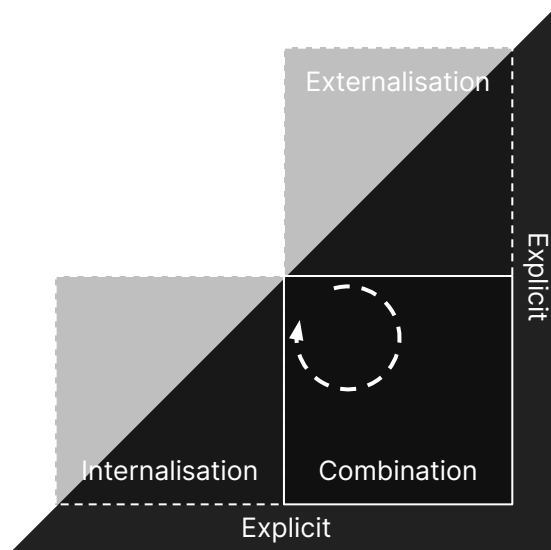
Social machines aid knowledge management

Nonaka-Takeuchi or SECI model

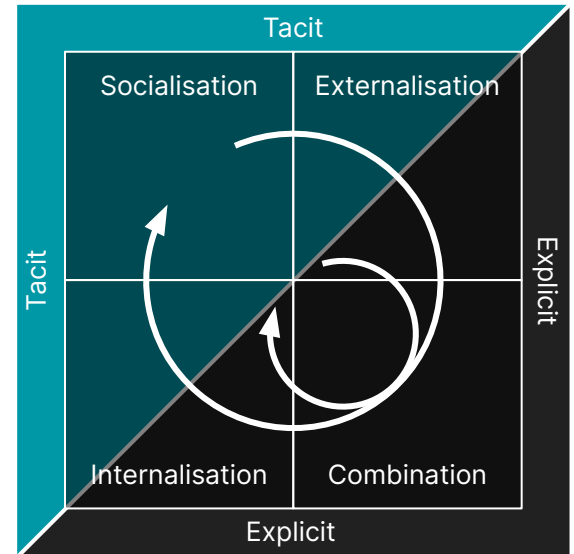
Human-driven



Naive exclusive automation



Social machines augment




```
array([-9.27734375e-02, 8.10546875e-02, -1.00097656e-01, 2.07031250e-01,
1.72119141e-02, -1.02050781e-01, 1.18164062e-01, -1.92382812e-01,
-2.53906250e-02, 1.17187500e-01, -2.42919922e-02, -4.73022461e-03,
1.48437500e-01, 2.28271484e-02, 1.08398438e-01, 4.45556641e-03,
4.57031250e-01, 1.66015625e-01, 1.37695312e-01, 1.33789062e-01,
1.38671875e-01, 1.82617188e-01, 2.45117188e-01, -1.77734375e-01,
1.90429688e-01, -1.50390625e-01, -6.10351562e-03, 6.83593750e-02,
-5.09643555e-03, 4.78515625e-02, 4.73632812e-02, -1.62109375e-01,
-1.33789062e-01, 2.63671875e-01, 1.17187500e-01, 1.90734863e-03,
-1.90429688e-02, -8.30078125e-02, 1.15234375e-01, 2.07031250e-01,
2.46093750e-01, 2.91748047e-02, -8.48388672e-03, 2.45117188e-01,
3.35937500e-01, -1.73828125e-01, -7.47070312e-02, -2.79541016e-02,
1.78710938e-01, -4.39453125e-02, 6.46972656e-03, -1.60156250e-01,
1.68945312e-01, 1.49414062e-01, -1.03515625e-01, 1.21582031e-01,
-1.04492188e-01, 2.55859375e-01, 1.50146484e-02, -1.67968750e-01,
-2.40234375e-01, 1.48925781e-02, -2.24609375e-01, -1.13281250e-01,
1.01928711e-02, 3.39355469e-02, -1.00097656e-01, 7.37304688e-02,
-1.92382812e-01, 1.08398438e-01, 1.11328125e-01, -8.83789062e-02,
6.49414062e-02, 1.12304688e-01, -1.50146484e-02, 9.82666016e-03,
2.14843750e-01, 5.34667969e-02, -1.32812500e-01, 3.80859375e-02,
-2.79296875e-01, 4.27246094e-02, 1.04003906e-01, -1.00097656e-01,
-1.04980469e-01, -1.22558594e-01, -2.02148438e-01, -7.66601562e-02,
-8.34960938e-02, -1.29882812e-01, -1.36108398e-02, 3.58886719e-02,
4.93164062e-02, 4.68750000e-02, -2.88085938e-02, -5.00488281e-02,
-4.31640625e-01, 1.15234375e-01, 7.08007812e-02, 2.52685547e-02,
-1.11328125e-01, -7.32421875e-02, 2.53906250e-02, 5.00488281e-03,
-8.05664062e-02, -1.87500000e-01, -1.57226562e-01, -6.92749023e-03,
-3.78417969e-02, -5.43212891e-03, -9.81445312e-02, 3.80859375e-02,
-1.15234375e-01, 6.68945312e-02, 4.66308594e-02, 1.61132812e-02,
-1.34887695e-02, 1.64794922e-02, -1.97753906e-02, 2.17773438e-01,
-1.22680664e-02, 5.66406250e-02, -2.02148438e-01, 1.03515625e-01,
9.03320312e-02, 4.41894531e-02, -1.04492188e-01, -2.20703125e-01,
9.03320312e-02, 1.06933594e-01, 6.83593750e-02, -1.00708008e-02,
-1.04492188e-01, -7.95898438e-02, -2.69775391e-02, 7.61718750e-02,
8.64257812e-02, 1.15722656e-01, -1.87500000e-01, -8.30078125e-02,
1.69921875e-01, -1.49414062e-01, -1.49536133e-02, 4.63867188e-02,
7.17773438e-02, 1.59179688e-01, -1.87988281e-02, -9.37500000e-02,
1.07421875e-02, -1.51367188e-01, 2.04101562e-01, -1.58203125e-01,
1.24511719e-02, 7.56835938e-02, -2.55584717e-04, 6.07910156e-02,
1.67968750e-01, 1.00585938e-01, -1.96289062e-01, -1.36108398e-02,
5.37109375e-02, -1.28936768e-03, 4.46777344e-02, -8.48388672e-03,
8.15429688e-02, -2.23632812e-01, -8.39843750e-02, 4.45556641e-03,
-2.63671875e-01, 1.34887695e-02, -3.22265625e-02, -1.03149414e-02,
-9.03320312e-02, -1.31835938e-01, -2.10937500e-01, 3.49121094e-02,
-2.23632812e-01, -2.49023438e-02, 9.81445312e-02, 4.37011719e-02,
4.41894531e-02, -1.86523438e-01, 2.87109375e-01, 2.03125000e-01,
-2.55859375e-01, -2.14843750e-01, -1.86523438e-01, -2.39257812e-01,
-1.66015625e-02, 2.34375000e-02, 6.05468750e-02, 9.81445312e-02,
2.04101562e-01, 6.93359375e-02, -2.96875000e-01, -8.93554688e-02,
2.09045410e-03, -1.87500000e-01, 8.97216797e-03, -3.97949219e-02,
-5.88378906e-02, 6.12792969e-02, -1.18652344e-01, -6.34765625e-02,
1.80664062e-01, -5.93261719e-02, -2.03125000e-01, 5.39550781e-02,
-6.64062500e-02, 1.67968750e-01, 5.98144531e-02, 1.62109375e-01,
```

```
1.24511719e-01, 1.83105469e-02, 1.29882812e-01, -6.25000000e-02,
3.65234375e-01, 1.15722656e-01, -1.32812500e-01, 2.11914062e-01,
3.00292969e-02, 1.97968750e-02, -5.17578125e-02, -2.08007812e-01,
-6.68945312e-02, 1.75781250e-02, -1.41601562e-01, -1.96289062e-01,
1.05468750e-01, -7.91015625e-02, 3.66210938e-03, -1.20117188e-01,
1.81640625e-01, 6.25000000e-02, -8.49609375e-02, 1.69921875e-01,
1.72851562e-01, -1.11816406e-01, 2.05078125e-01, 7.08007812e-02,
5.27343750e-02, -3.80859375e-02, 3.93066406e-02, 1.37695312e-01,
-2.12402344e-02, 1.25976562e-01, 2.14843750e-01, 1.88476562e-01,
9.57031250e-02, -6.54296875e-02, 6.71386719e-04, 7.32421875e-02,
-1.59263611e-04, 2.12890625e-01, -1.20239258e-02, 6.68945312e-02,
1.14257812e-01, 6.78710938e-02, -1.62109375e-01, 5.32226562e-02,
-1.43554688e-01, 1.56250000e-01, 1.01074219e-01, 1.15722656e-01,
5.02929688e-02, -7.12890625e-02, -2.45101562e-02, -7.95898438e-02,
-1.08886719e-01, -4.02832031e-02, -1.69921875e-01, -9.61914062e-02,
-9.37500000e-02, -3.98437500e-01, 1.29882812e-01, -1.74804688e-01,
-7.86132812e-02, -1.58203125e-01, 4.68750000e-02, 5.76171875e-02,
6.64062500e-02, 1.14257812e-01, -1.50390625e-01, 1.15722656e-01,
5.07812500e-02, 1.12304688e-01, 6.44531250e-02, 3.75976562e-02,
1.10839844e-01, -2.34375000e-02, -4.76074219e-02, 2.62451172e-02,
-2.19726562e-02, -1.88476562e-01, 5.63964844e-02, -3.02734375e-01,
-1.68945312e-01, 1.84326172e-02, 1.04003906e-01, 9.03320312e-02],
dtype=float32)
```

Thanks!

David Colls
dcolls@thoughtworks.com