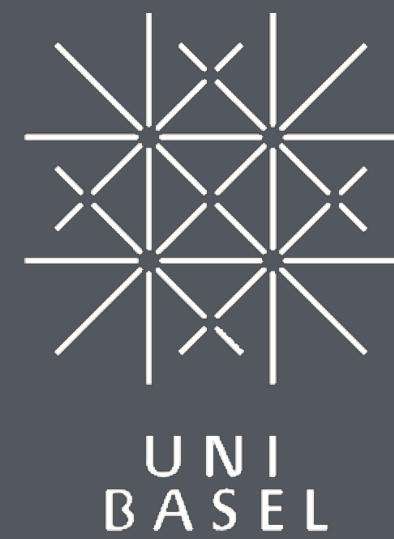


Qualitative data analysis with LLMs

Dirk Wulff & Zak Hussain

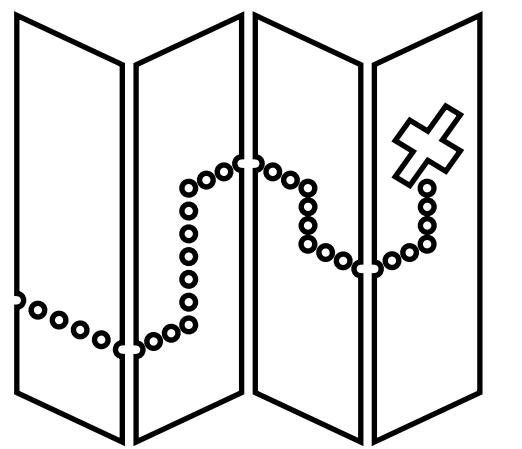


MAX PLANCK INSTITUTE
FOR HUMAN DEVELOPMENT

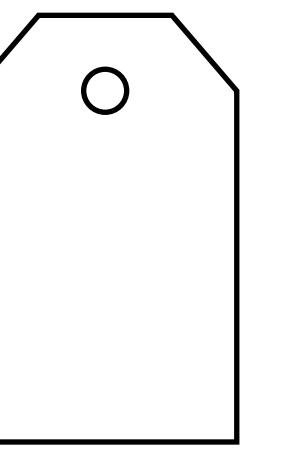


Applications

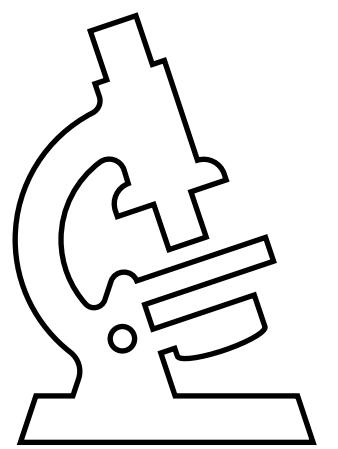
For qualitative data analysis



Map out and cluster



Label



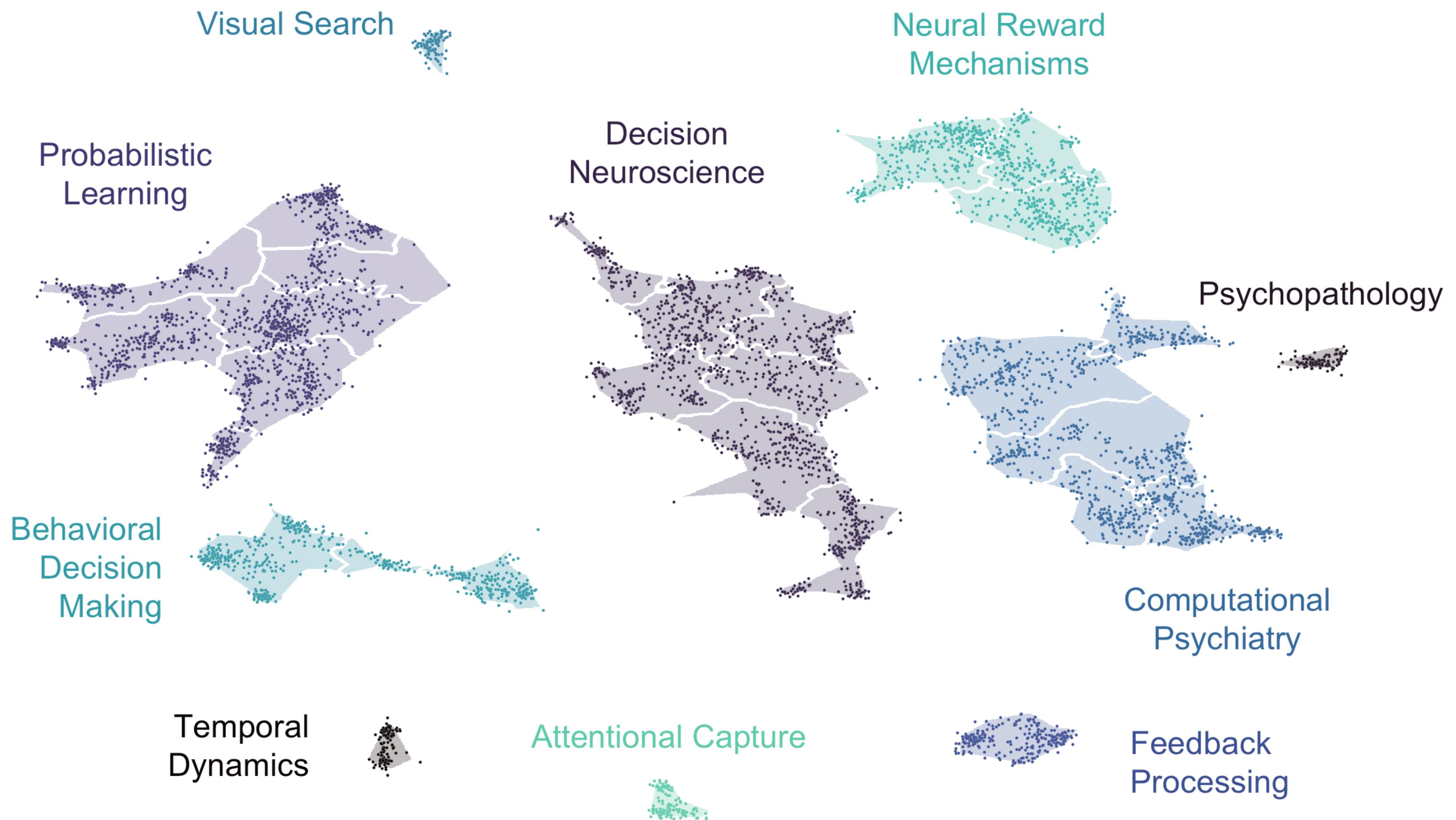
**Summarize
or paraphrase**



Search

Map and cluster

The landscape
of behavioral
reinforcement
learning
research



Map and cluster

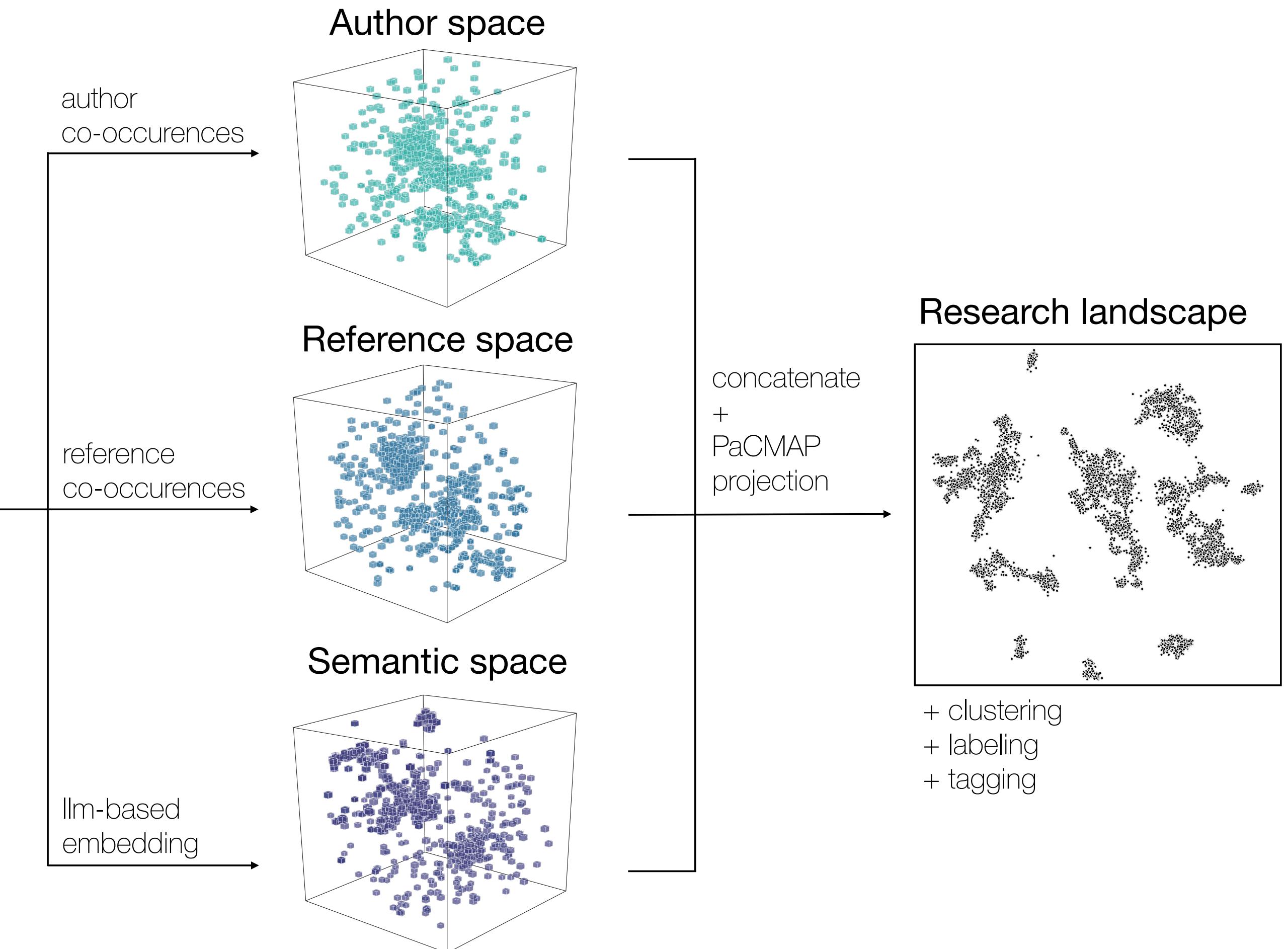
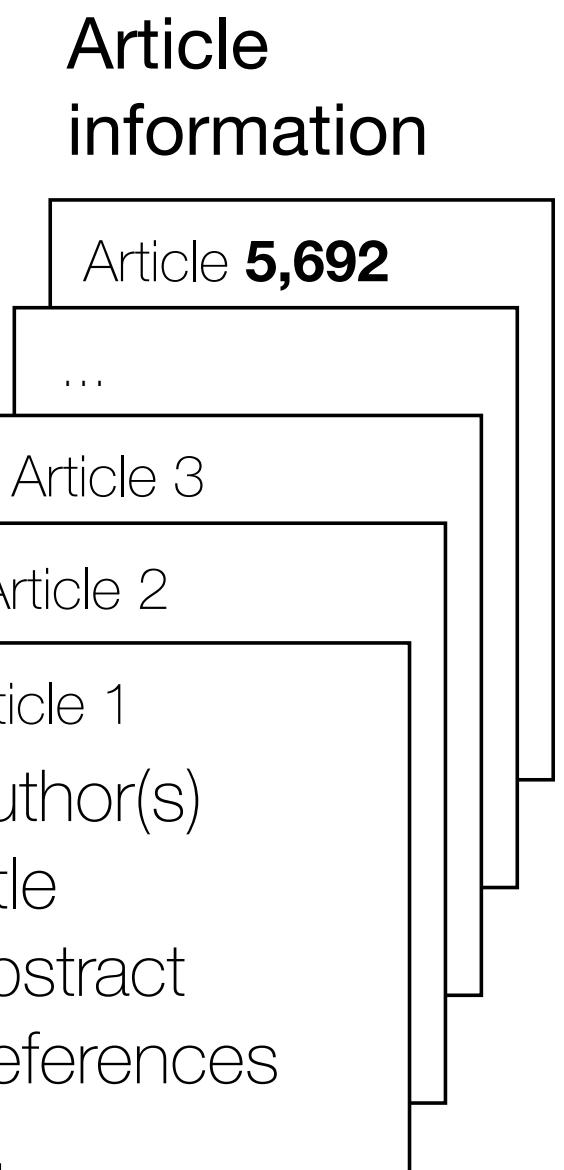
The landscape of behavioral reinforcement learning research

Search query

TITLE-ABS-KEY("reinforcement learning"
OR "decision*" from experience" OR
"choice* from experience" OR "experience-
based choice*" OR "experience-based
decision*" OR "repeated choice*" OR
"repeated decision*" OR "sequential
choice*" OR "sequential decision*" OR
"dynamic decision making" OR "probability
learning" OR "reward learning" OR "reward-
based learning" OR "reward-based
choice*" OR "reward-based decision*" OR
"value learning" OR "value-based learning"
OR "value-based choice*" OR "value-
based decision*) AND
LANGUAGE("English") AND DOCTYPE("ar")
AND SUBJAREA("NEUR" OR
"PSYC" OR "ECON")

↓
8,144 hits

cleaning & filtering



Map and cluster



Map and cluster

Understanding AI

 **Oct-Nov 2023**

3,000 participants

15,000 associations

 **Oct-Nov 2024**

3,000 participants

15,000 associations

 **Oct-Nov 2025**

3,000 participants

15,000 associations

What comes to your mind when you think of artificial intelligence (AI)?

List the **first five thoughts** that spontaneously come to your mind. Enter these associations with **one to three words** per entry field. There are **no right or wrong answers**. If less than five associations come to mind, leave the remaining fields empty.

Job loss

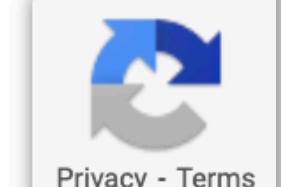
Robot

Terminator

ChatGPT

Helpful

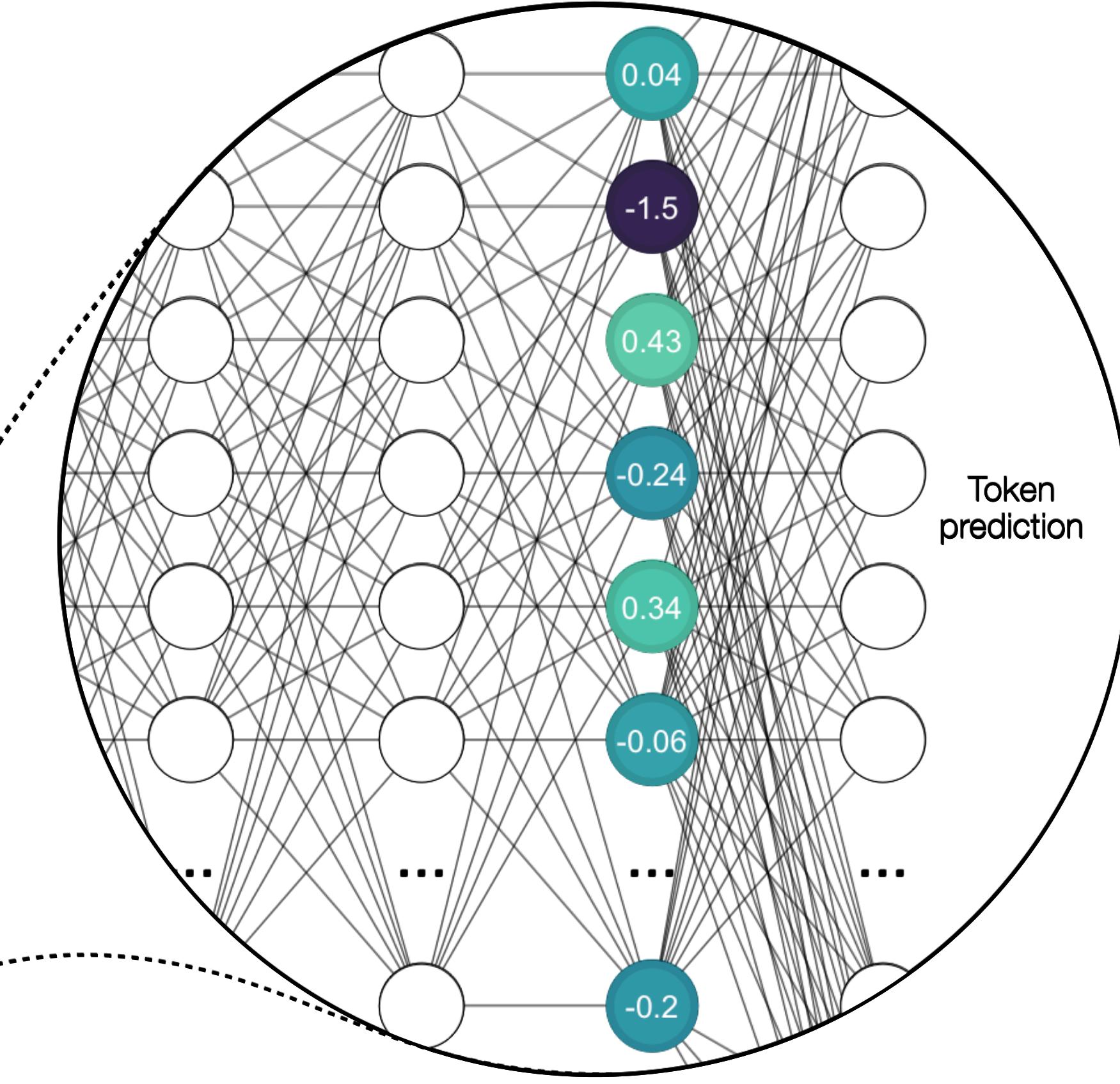
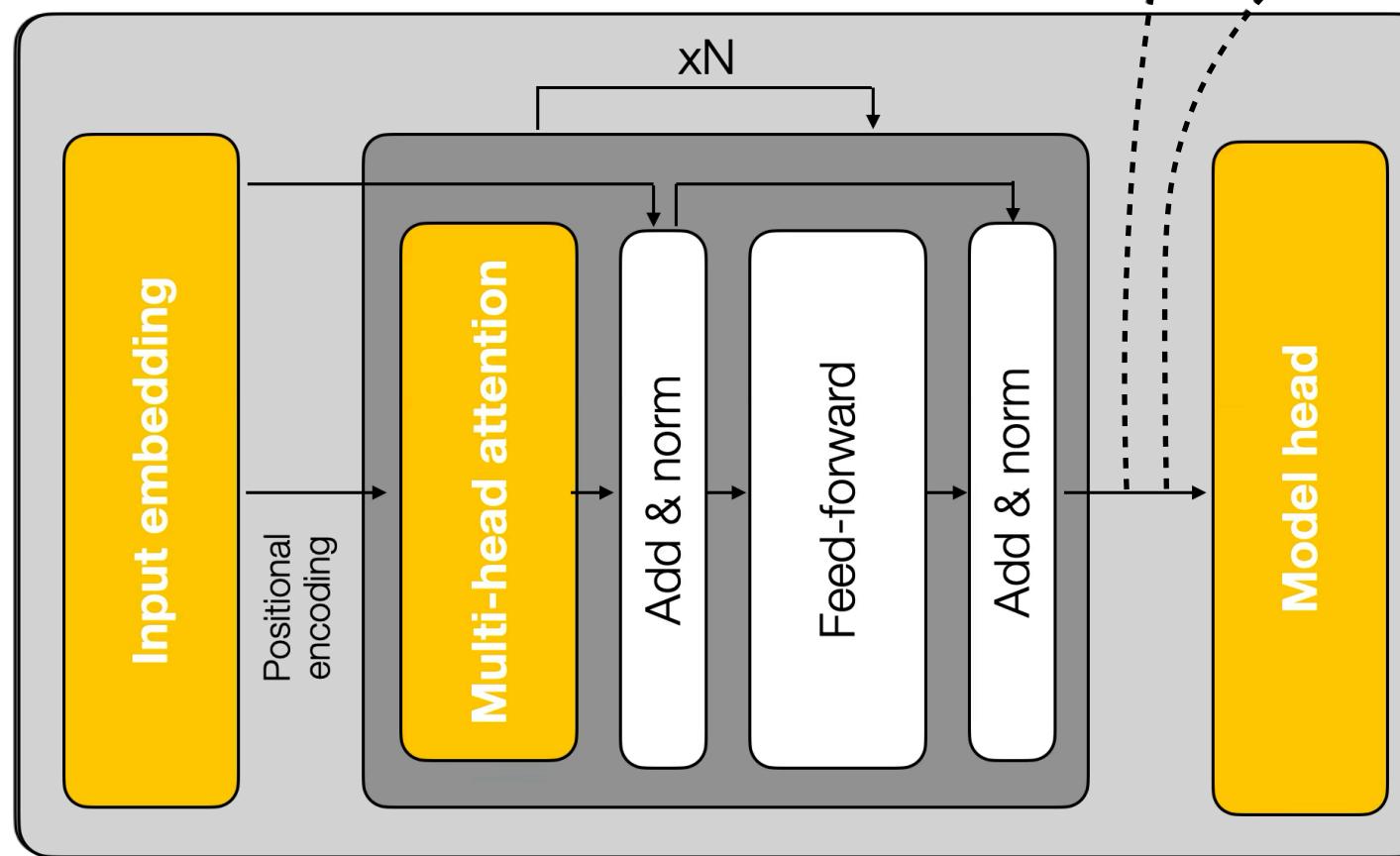
Next



Map and cluster

Understanding AI

“Terminator” →

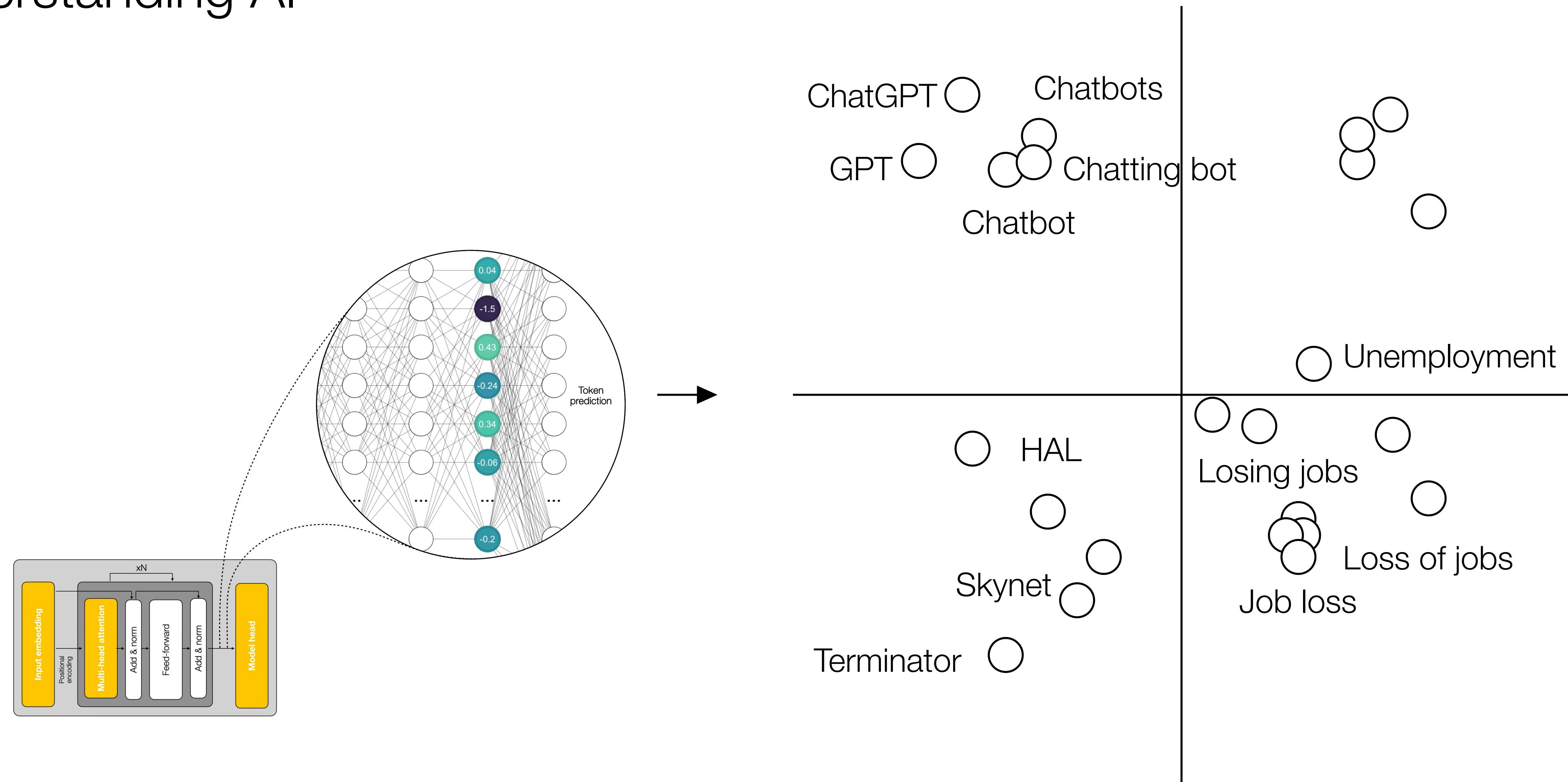


A position in high-dimensional space reflecting the input’s meaning

Map and cluster

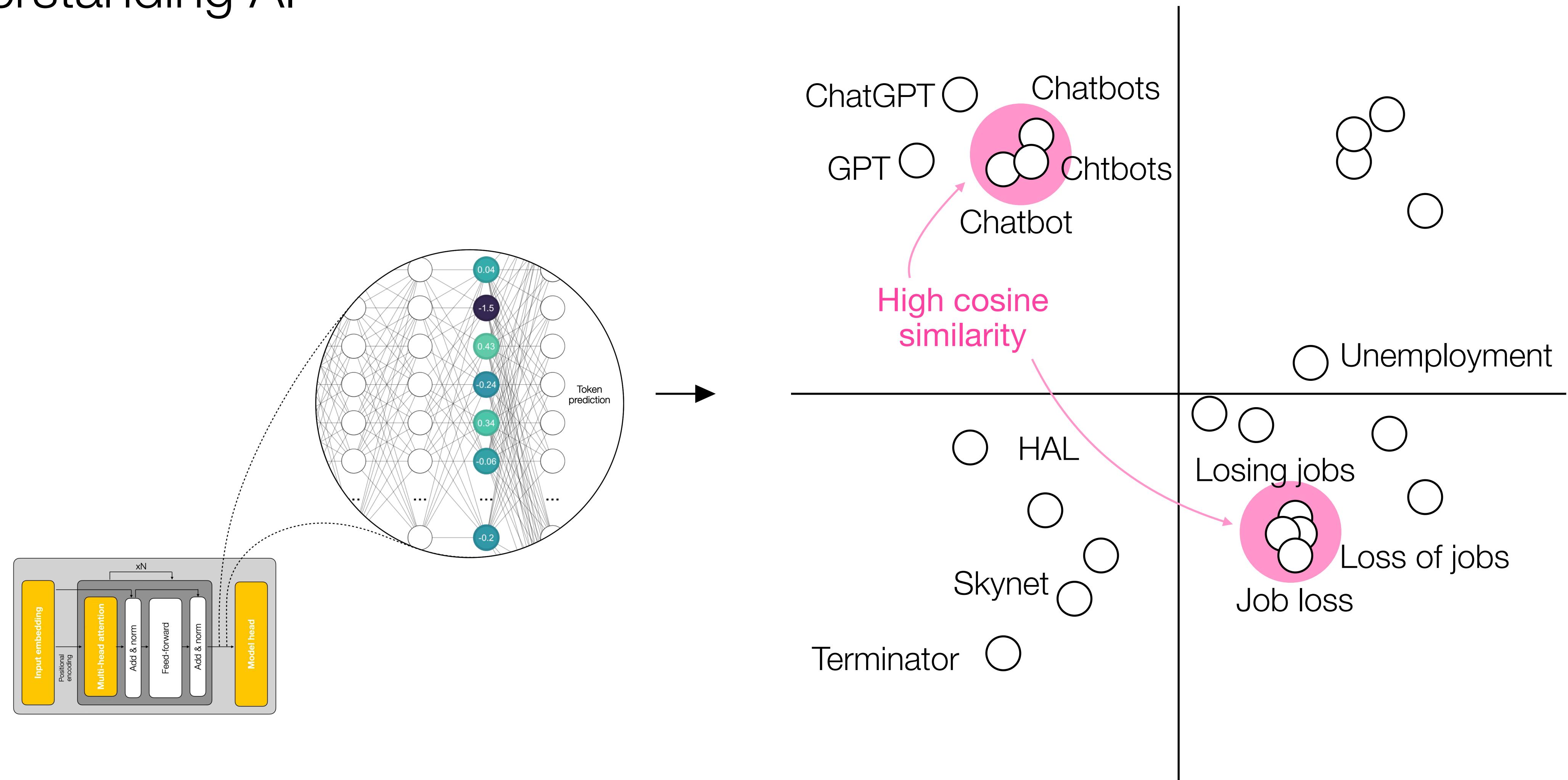
Understanding AI

Semantic space



Map and cluster

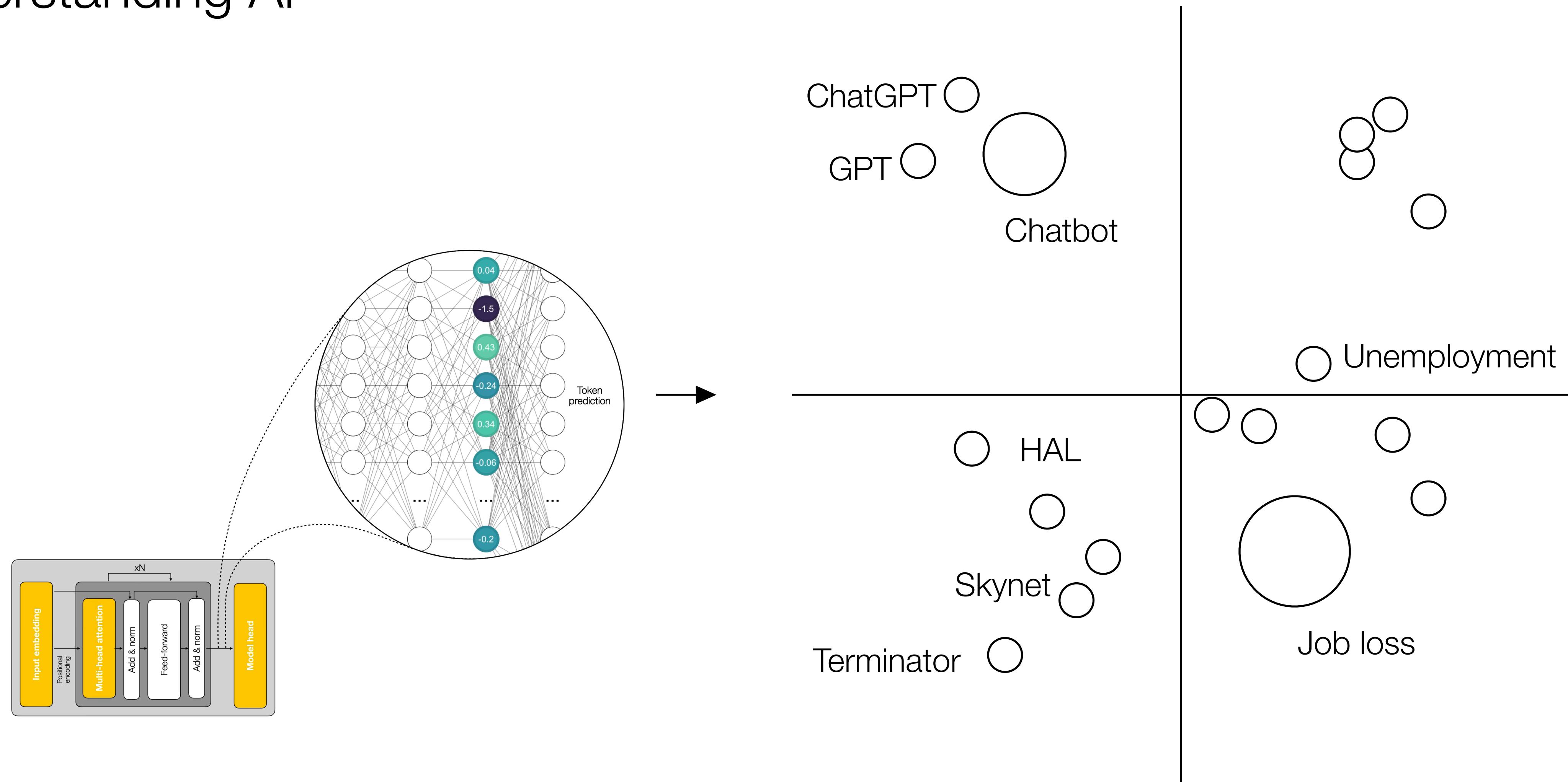
Understanding AI



Map and cluster

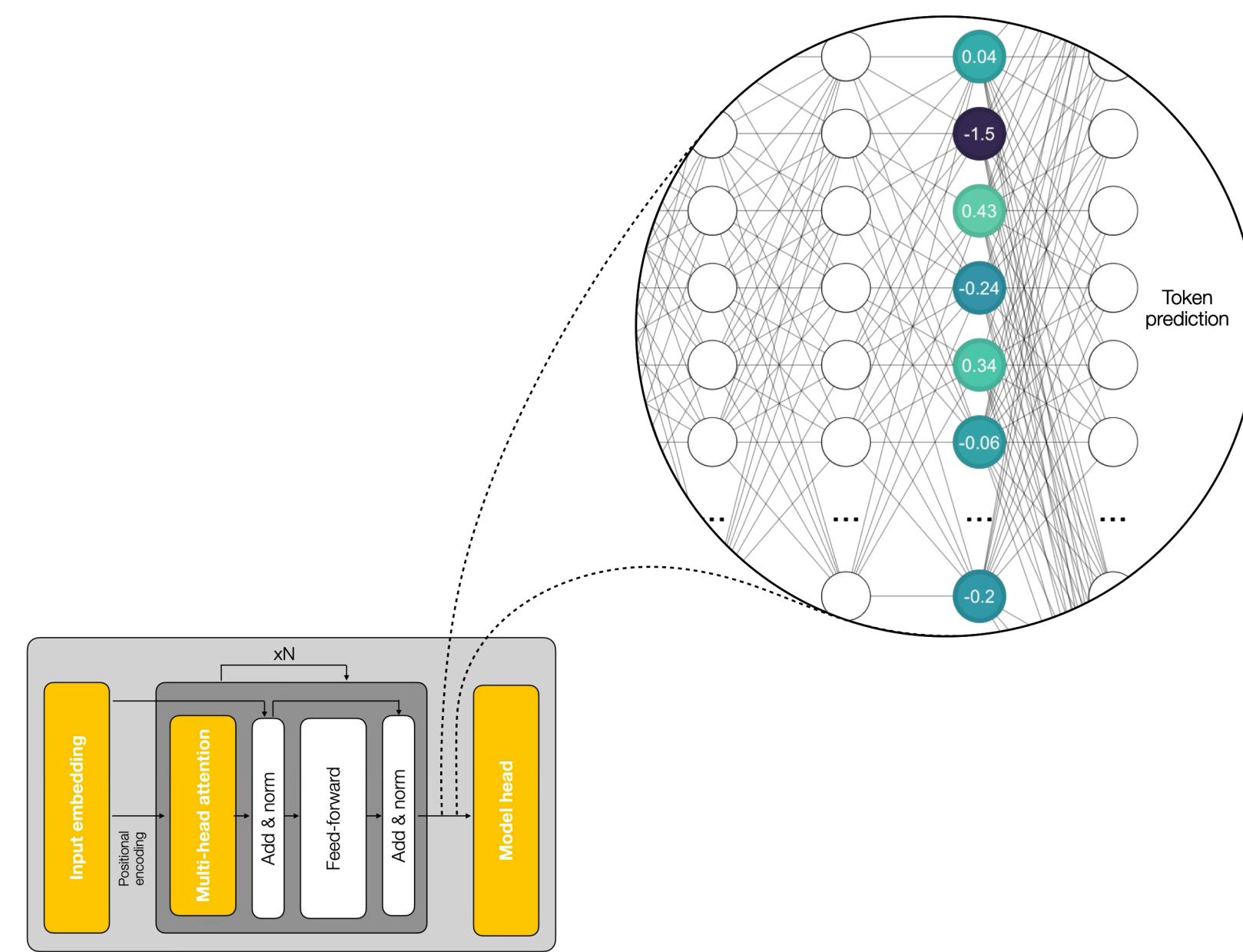
Understanding AI

Semantic space

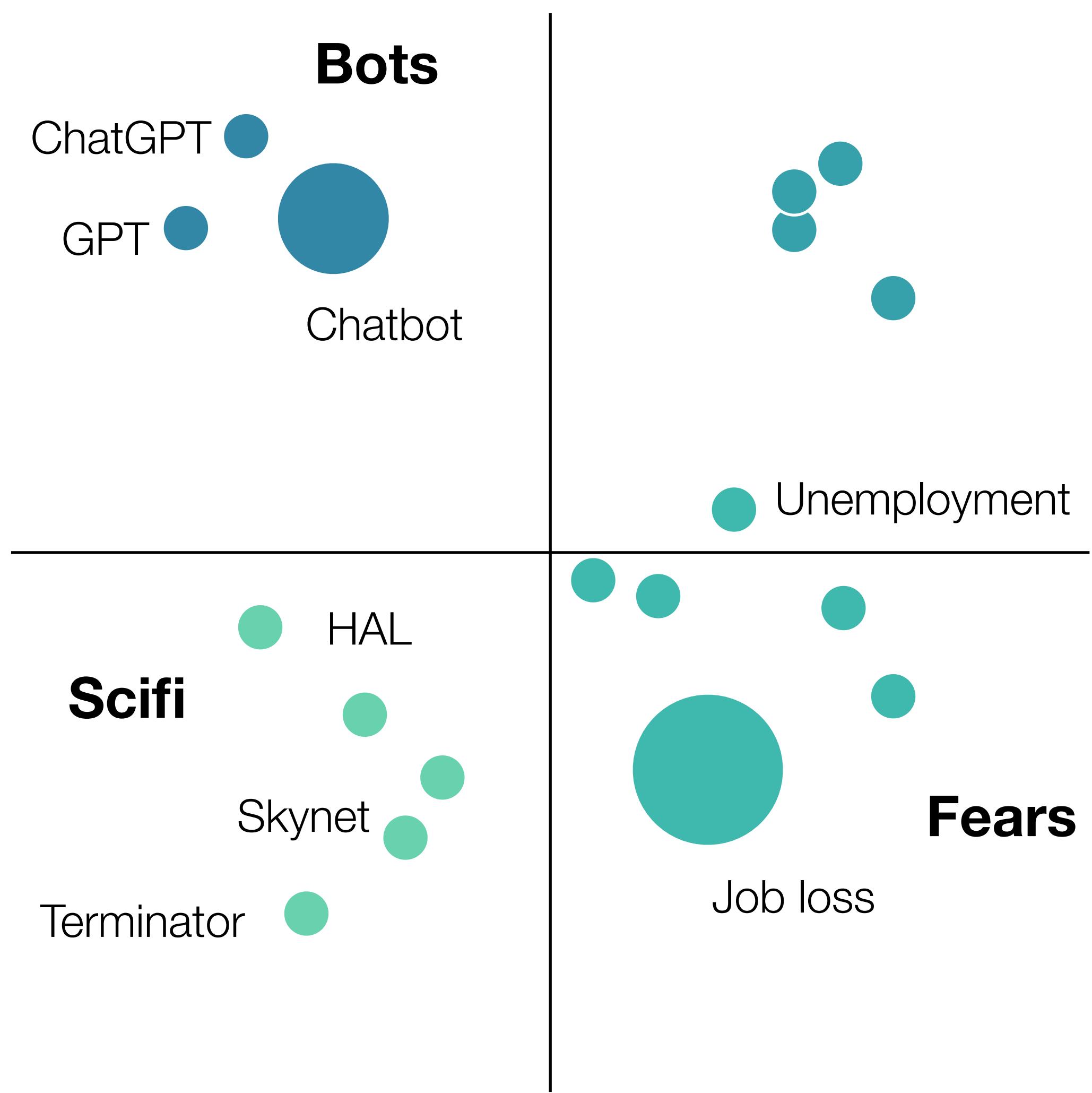


Map and cluster

Understanding AI



Semantic space



Map and cluster

Understanding AI

Deception

Fears

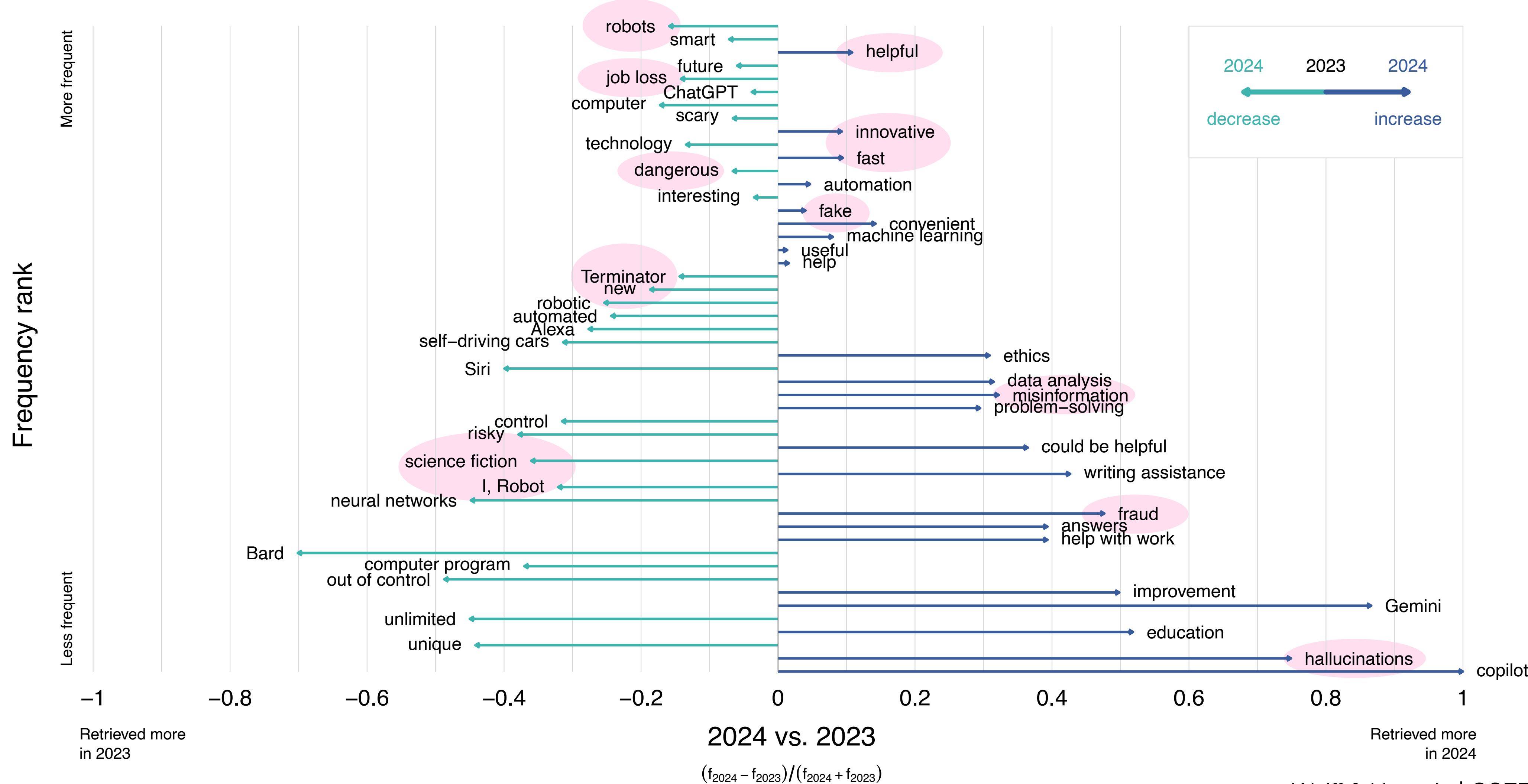
The word cloud illustrates several negative perceptions of AI:

- Terminator**: Associated with **evil**, **apocalypse**, **out of control**, **utopian**, **takeover**, **replacement**, **threatening**, **dangerous**, **job loss**, **scary**, **concern**, **risky**, **scam**, **problematic**, **creepy**, **biased**, **unethical**, **controversial**, **concerning**, **potentially**, **intrusive**, **potentially harmful**, **unreliable**, **not trustworthy**, **misuse**, **captive**, **pirate**, **untrustworthy**, **ethics**, **needs regulation**, **invasion of privacy**, and **privacy concerns**.

Sci - Fi

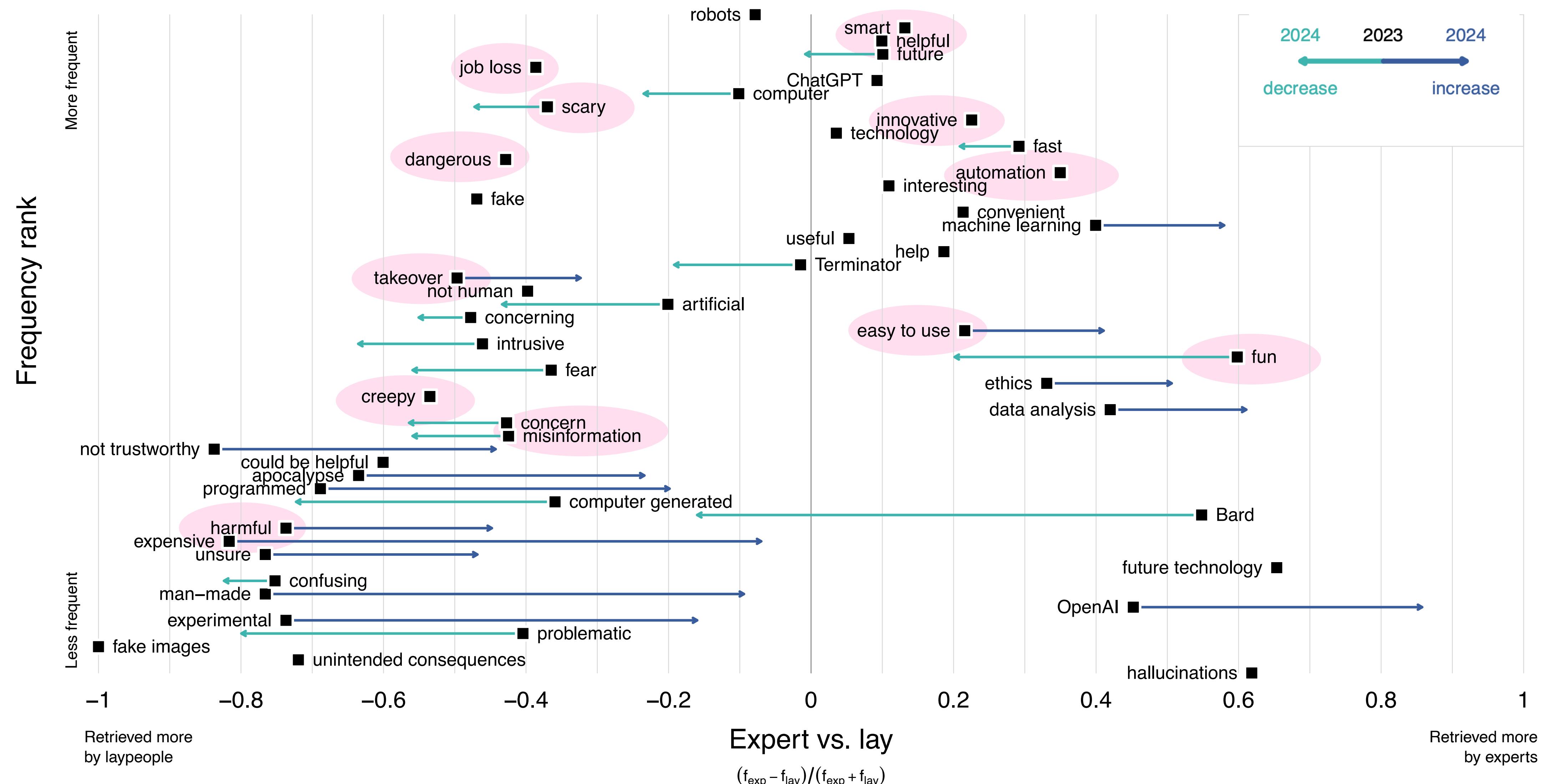
Map and cluster

Understanding AI



Map and cluster

Understanding AI



Labeling

Sustainable development goals

Text

Bycatch—the retained catch of non-target but commercially viable species, plus all fisheries discards—is an element of almost every marine capture fishery. Fishing mortality through bycatch in purse seines and pelagic longlines,...

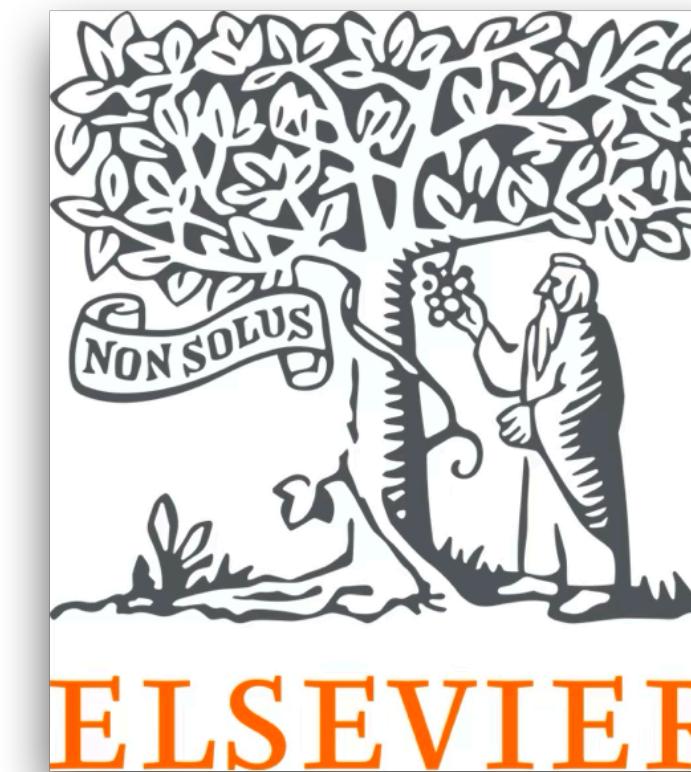


Query

(ocean* OR marine OR coast* OR sea OR seas OR seawater* OR sea water* OR coral reef*)
AND
(overfish* OR fisheries OR fishery OR fish stock* OR fishing)



Label

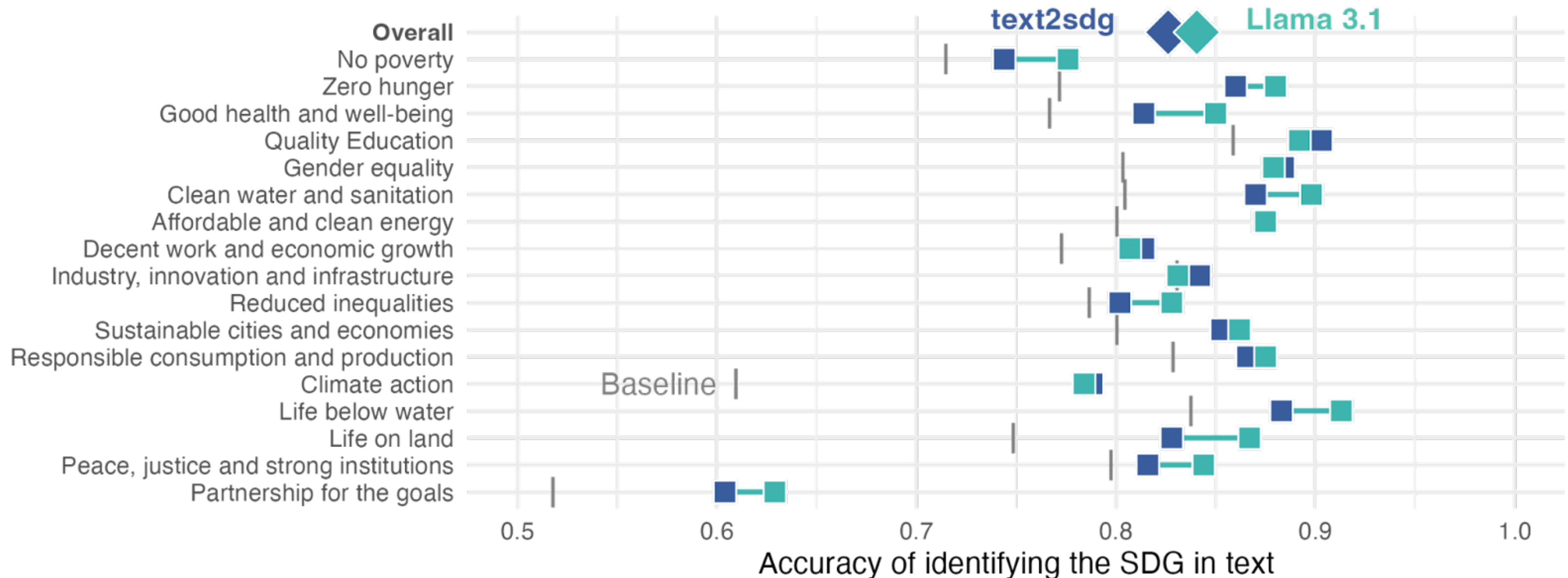


Labeling systems



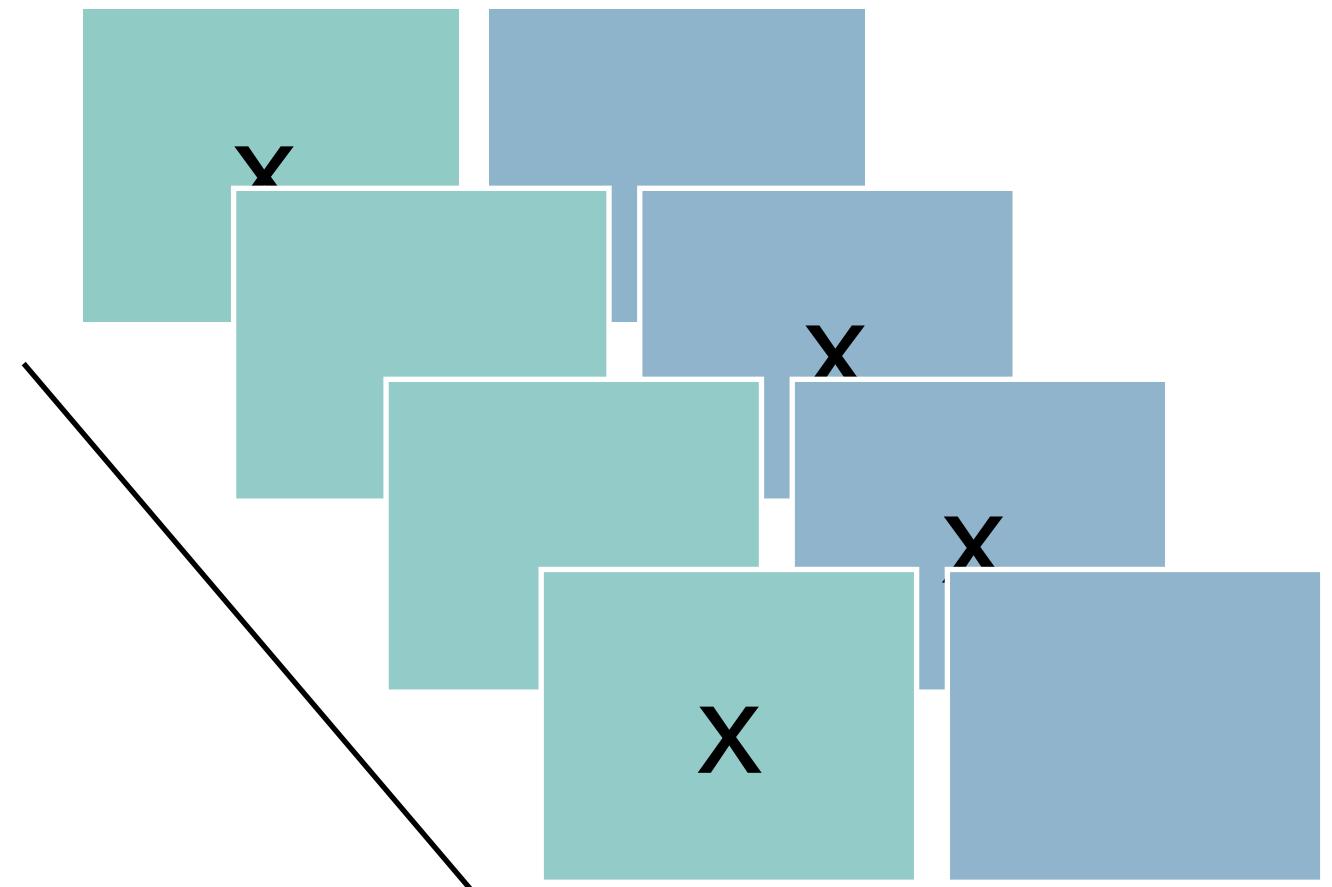
Labeling

Sustainable development goals



Labeling

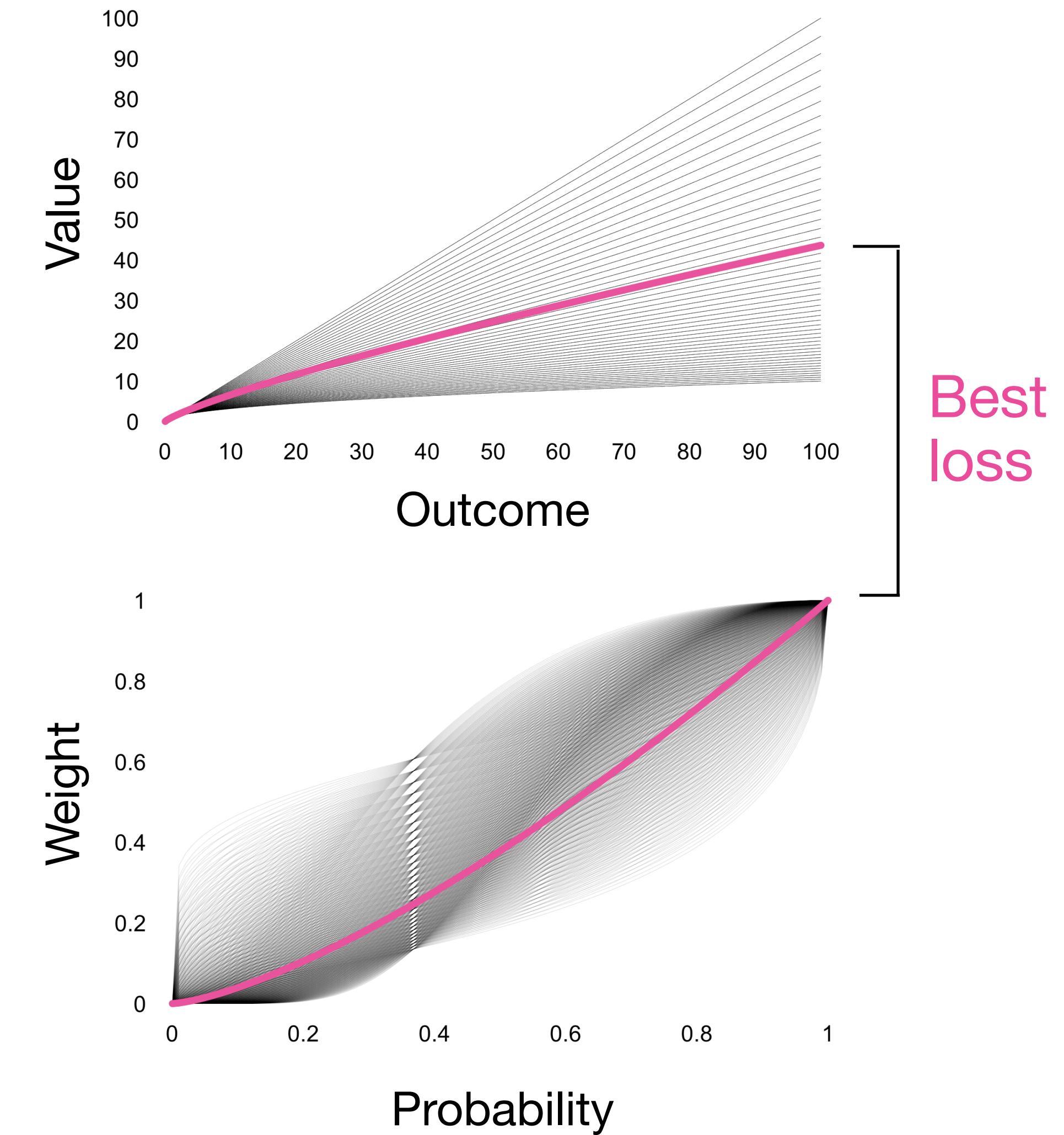
Risky choice verbal reports



Assumption
Stable process
across people and
situations

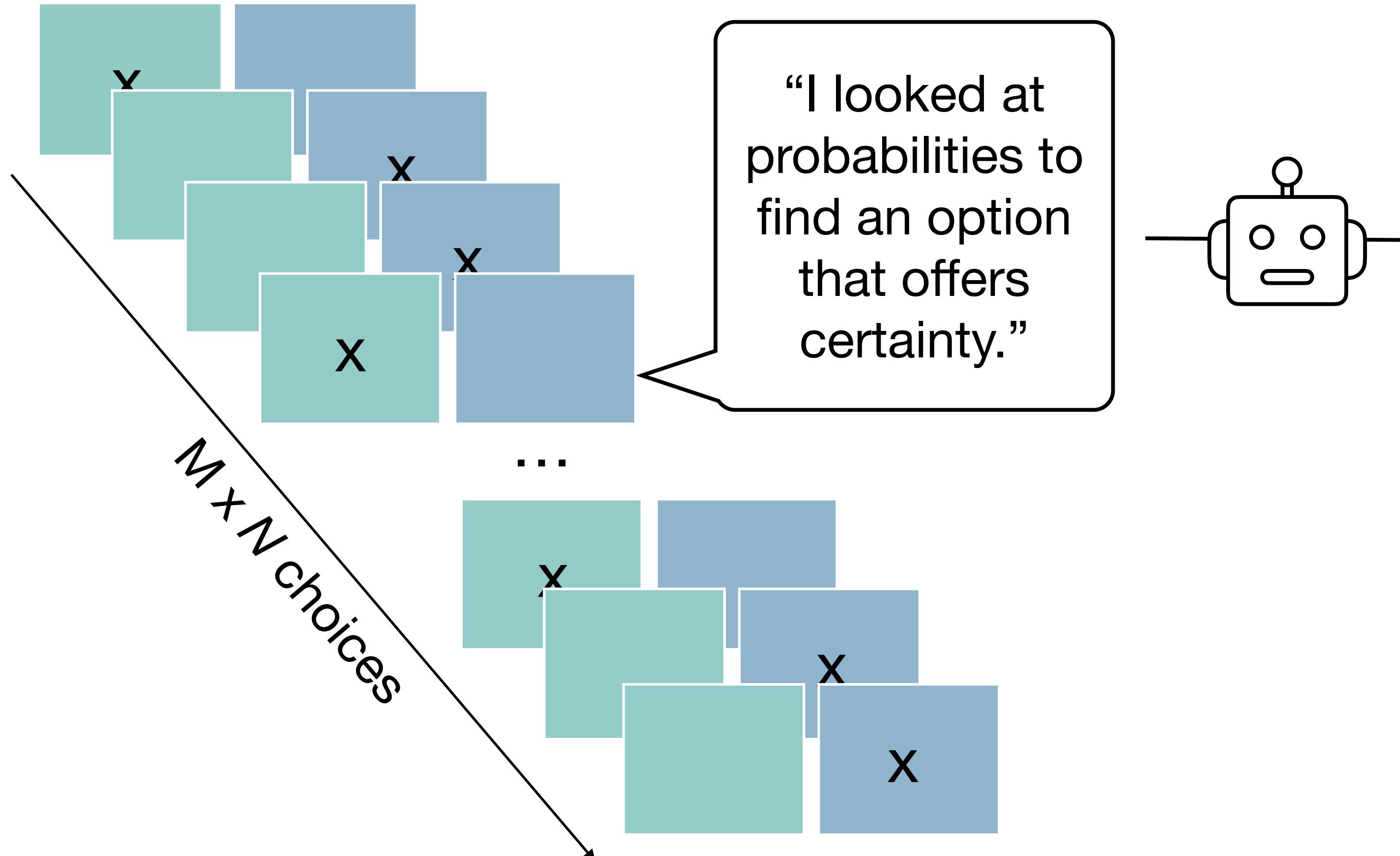
fitting

Nobel-prize winning
Prospect Theory



Labeling

Risky choice verbal reports



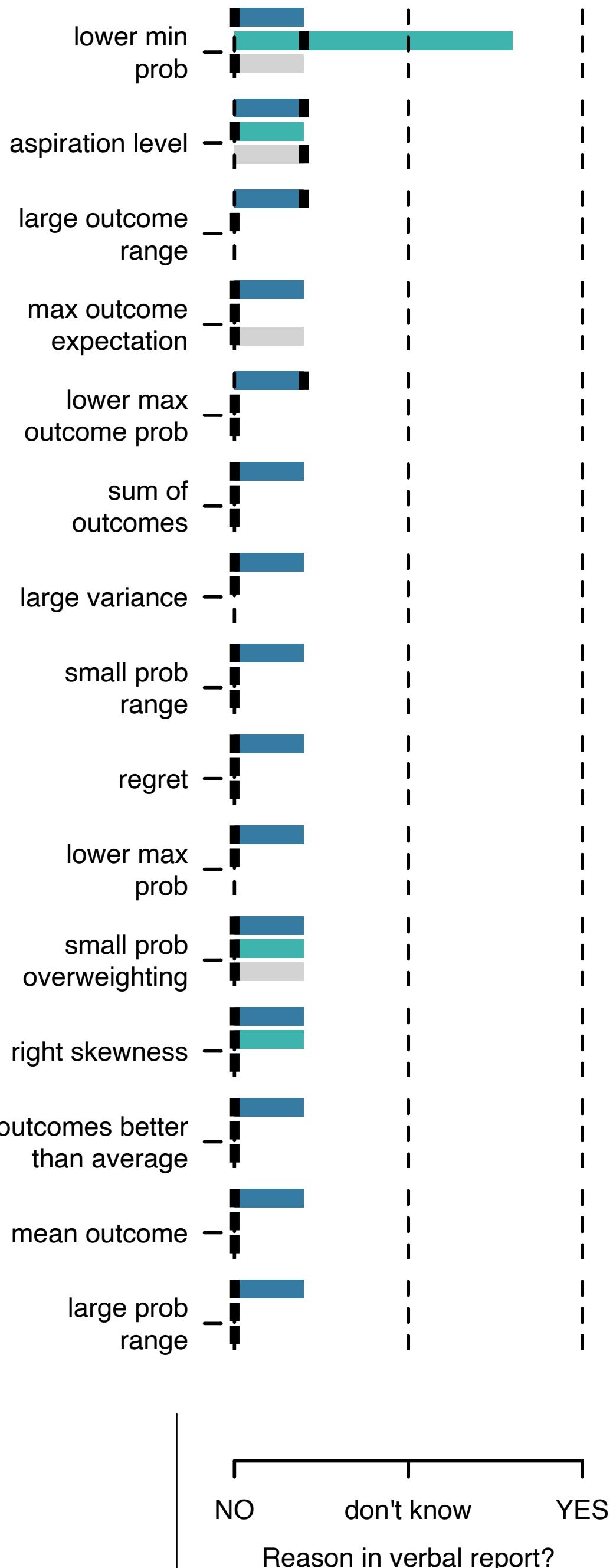
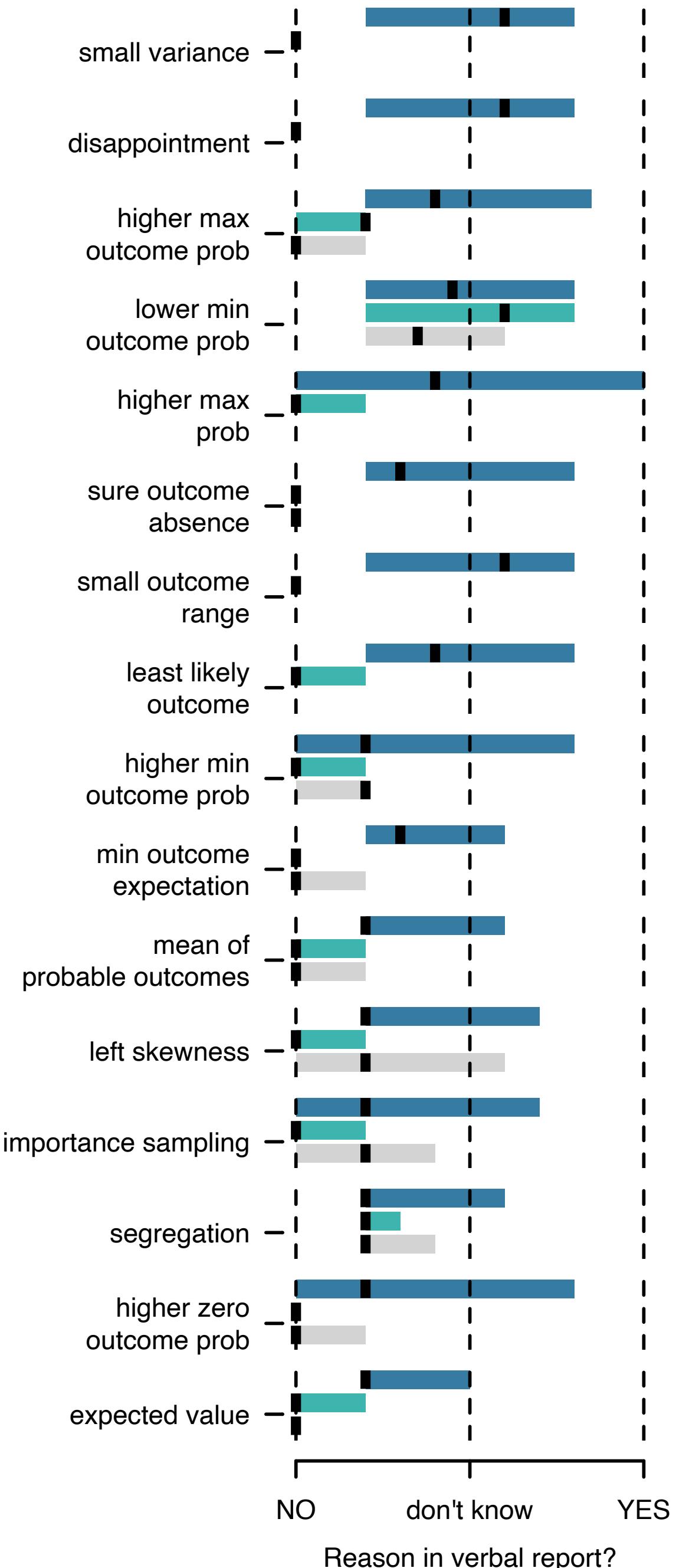
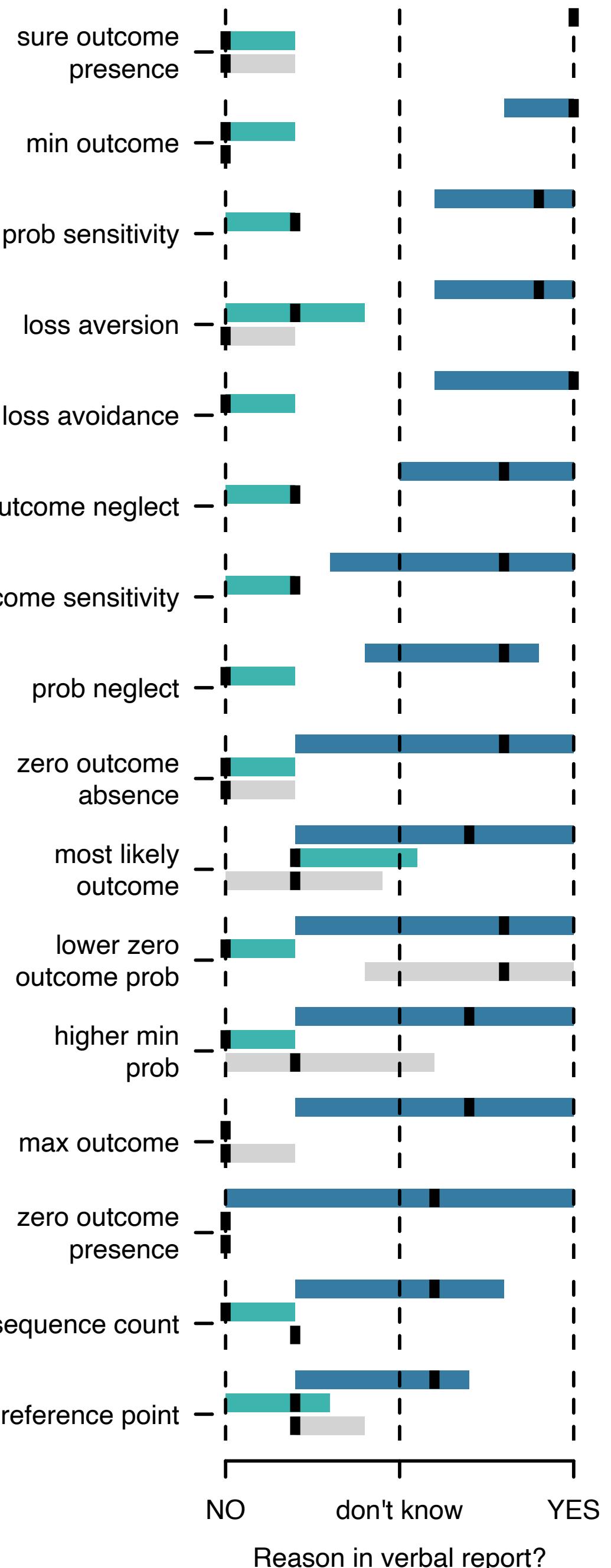
Weight

Labeling

Risky choice verbal reports

LLM judgments accurately identify when reasons are present and absent

- █ Choice = Formal decision reason
- █ Choice ≠ Formal decision reason
- █ Formal decision rule is indifferent



Labeling

Risky choice verbal reports

Decision reasons cluster into three groups

Cluster 1

Certainty → Focus on certainty

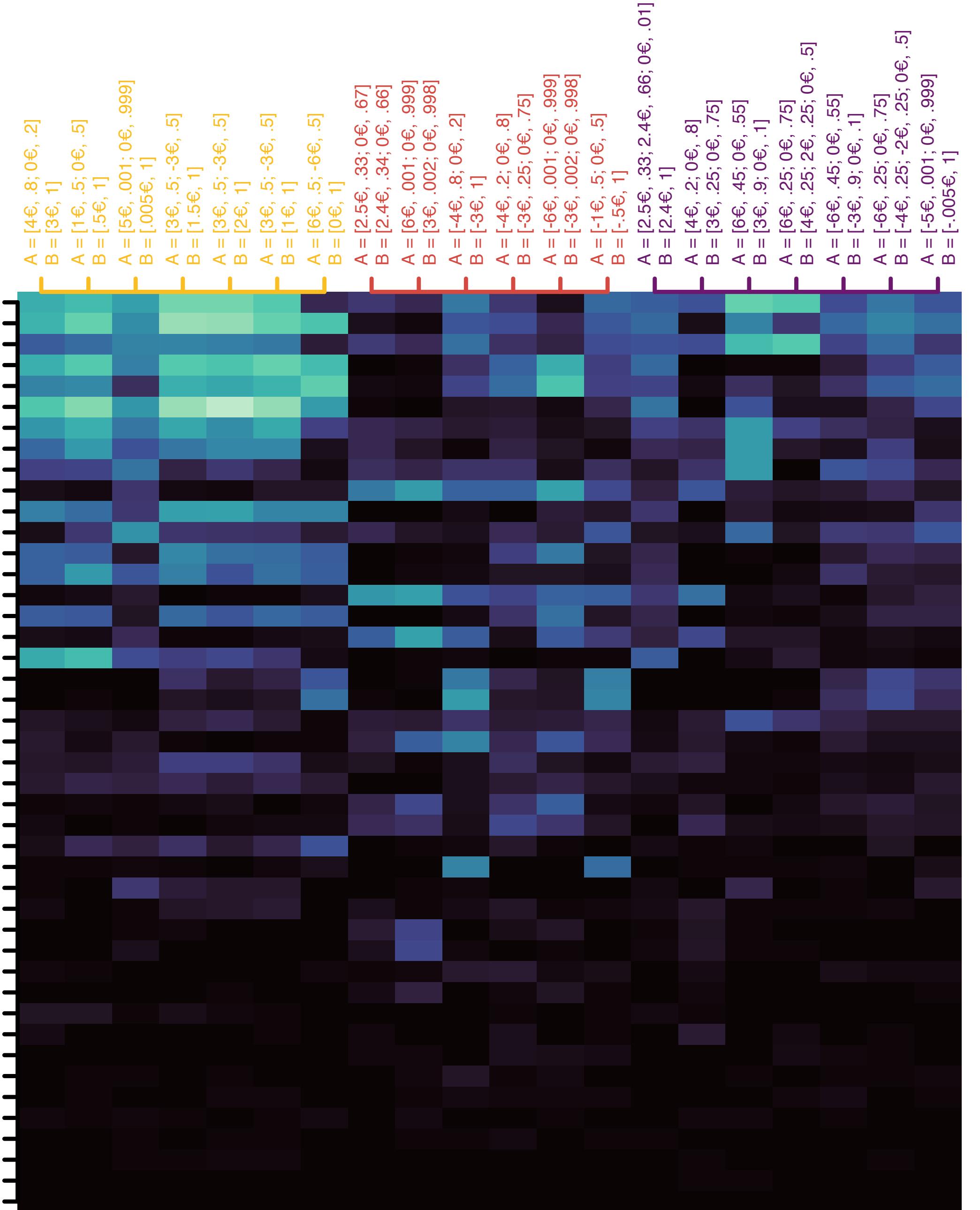
Cluster 2

Similar probability distributions
→ Focus on outcomes

Cluster 3

Difficult problems → Risk aversion

probability sensitivity
loss avoidance
outcome neglect
minimum outcome loss aversion
sure outcome presence
higher minimum probability
higher maximum probability
higher maximum outcome probability
outcome sensitivity
small variance
most likely outcome disappointment
higher minimum outcome probability
probability neglect
small outcome range
maximum outcome
zero outcome absence
higher zero outcome probability
zero outcome presence
left skewness
least likely outcome
reference point
segregation
consequence count
importance sampling
minimum outcome expectation
sure outcome absence
mean of probable outcomes
expected value
sum of outcomes
large outcome range
lower maximum outcome probability
regret
small probability range
maximum outcome expectation
lower maximum probability
right skewness
aspiration level
large variance
small probability overweighting
mean outcome
large probability range
outcomes better than average



Labeling

Predicting risk perception
with GPT-4-Turbo

The screenshot shows a conversation between the user ('You') and ChatGPT. The user asks for a numerical estimate on a scale from -100 (safe) to 100 (risky) for the entity 'robot'. ChatGPT responds with the value 30. Later, the user asks for a numerical estimate for the entity 'virus' on the same scale, and ChatGPT responds with the value 70. At the bottom right of the second message, there are four small icons: a speaker icon, a folder icon, a circular arrow icon, and a thumbs-down icon.

You
On a scale from -100 (safe) to 100 (risky), how risky or safe is 'robot'? Respond with a numerical estimate.

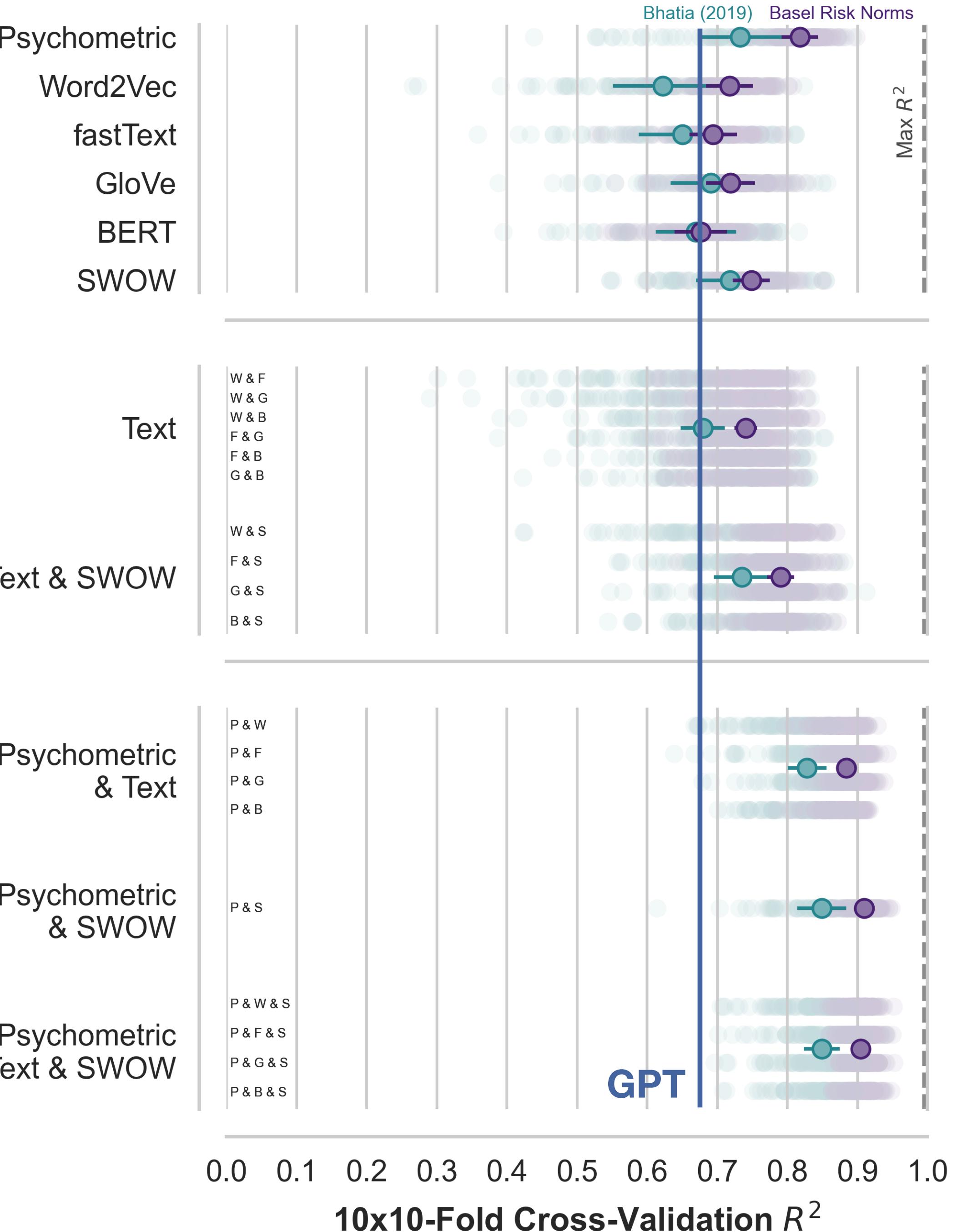
ChatGPT
30

You
User
On a scale from -100 (safe) to 100 (risky), how risky or safe is 'virus'? Respond with a numerical estimate.

ChatGPT
70

Labeling

Predicting risk perception
with GPT-4-Turbo



Labeling

Cognitive tasks

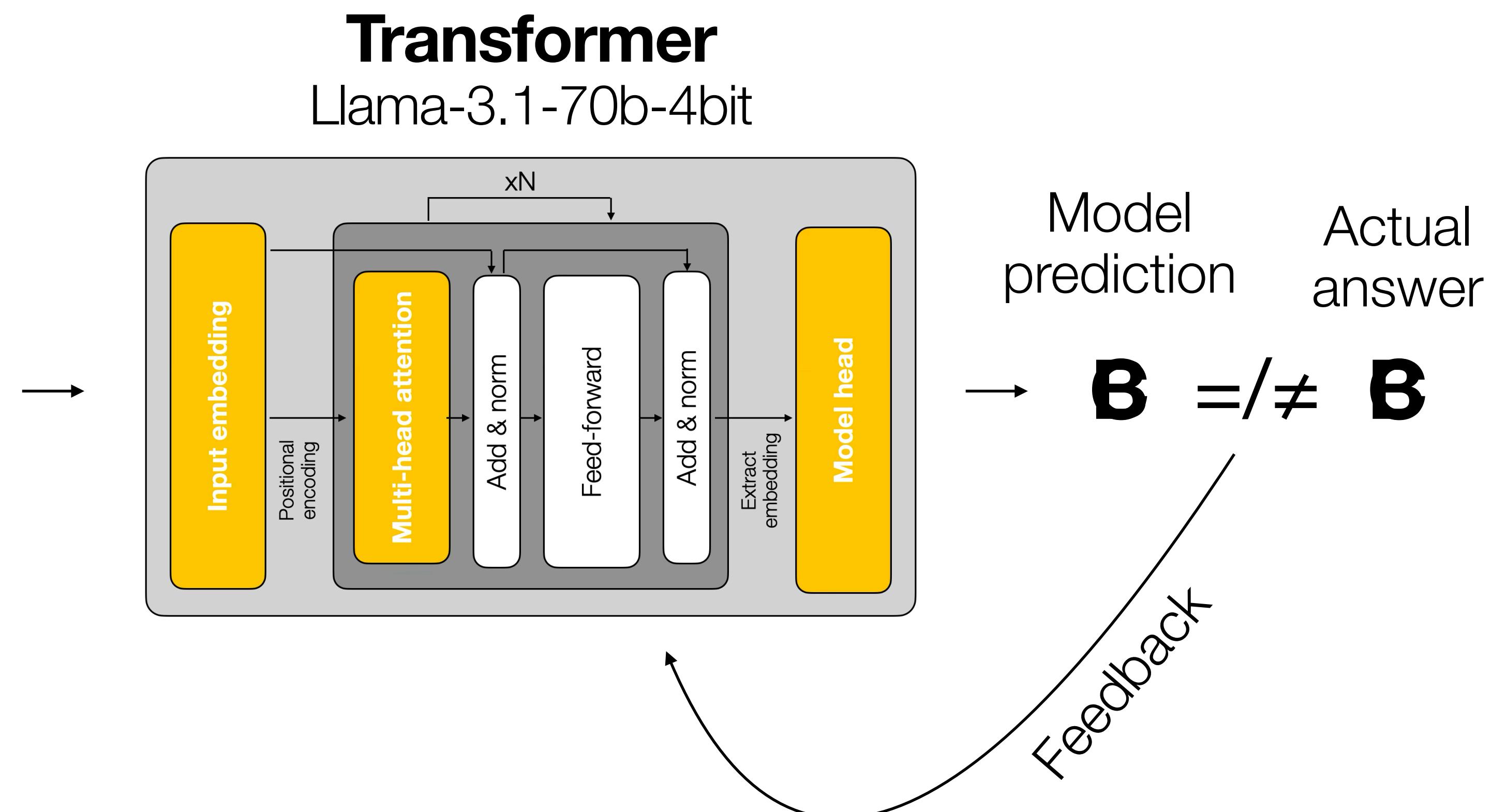
Bandit task

In this task, you have to repeatedly choose between two slot machines labeled **B** and **C**. When you select one of the machines, you will win or lose points. Your goal is to choose the slot machines that will give you the most points.

You press <<**C**>> and get -8 points.

You press <<**B**>> and get 0 points.

You press <<**B**>> and get 1 points.



PSYCH-101

160
studies

60,092
participants

10,681,650
choices



Memory

N-back, Digit span, Recall, Recognition, Episodic memory, etc.

Learning

Associative learning, Categorization, Multiple-cue judgment, Gardening task, etc.

Bandits

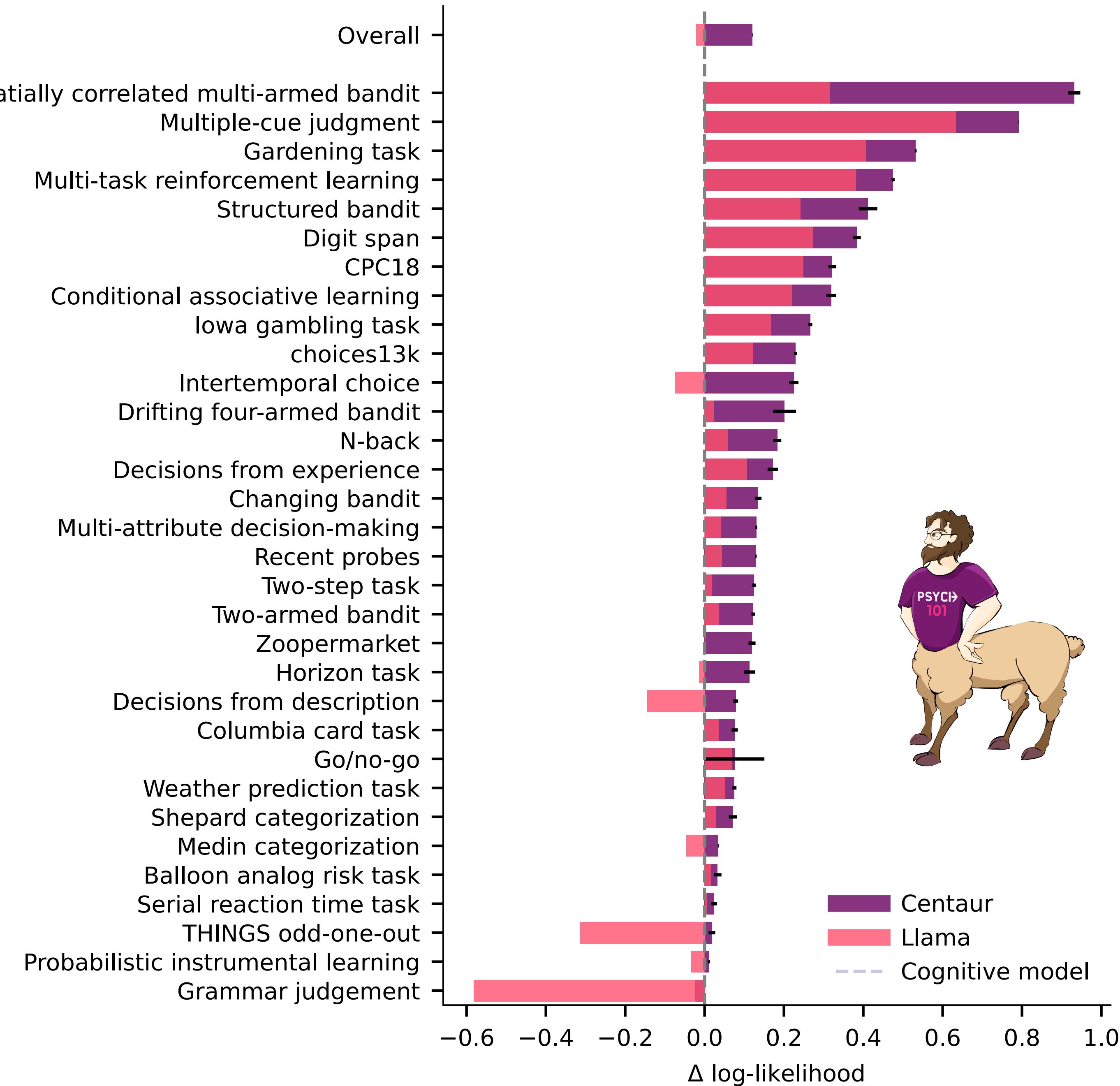
Decisions from experience, Iowa gambling task, Two-armed bandit, Structured bandit, etc.

Decisions

Risky choice, Balloon analog risk task, Columbia card task, Intertemporal choice, etc.

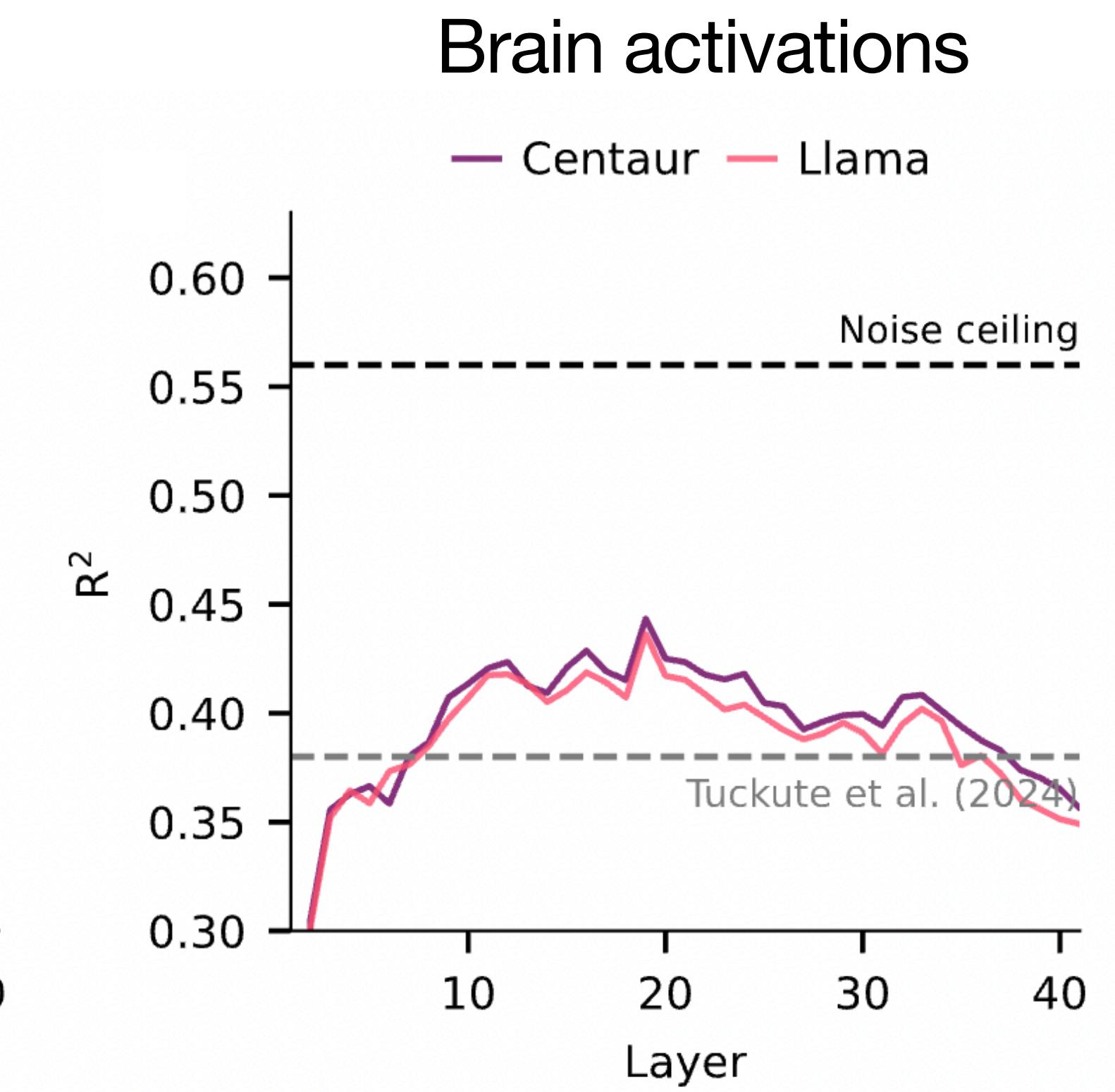
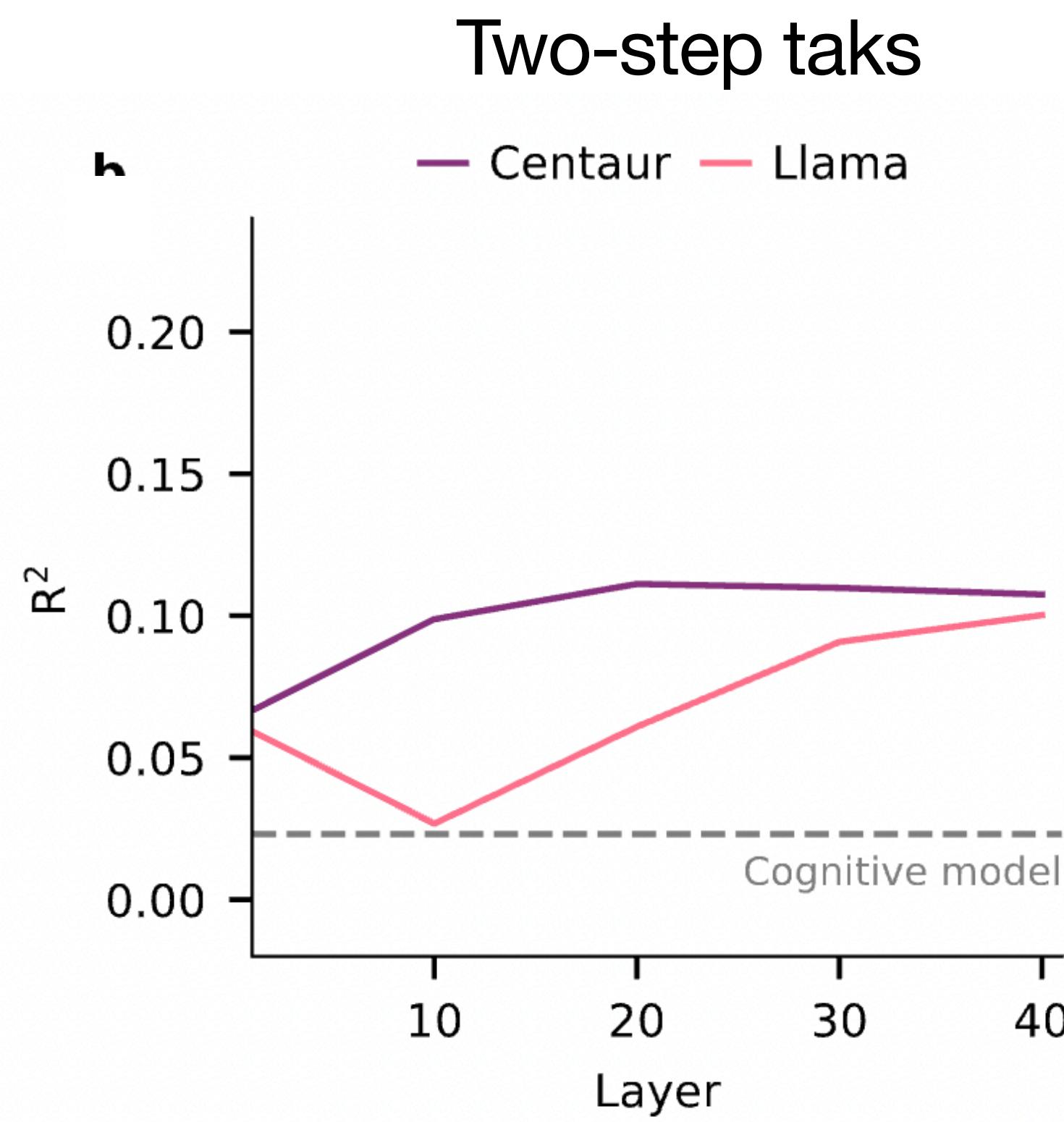
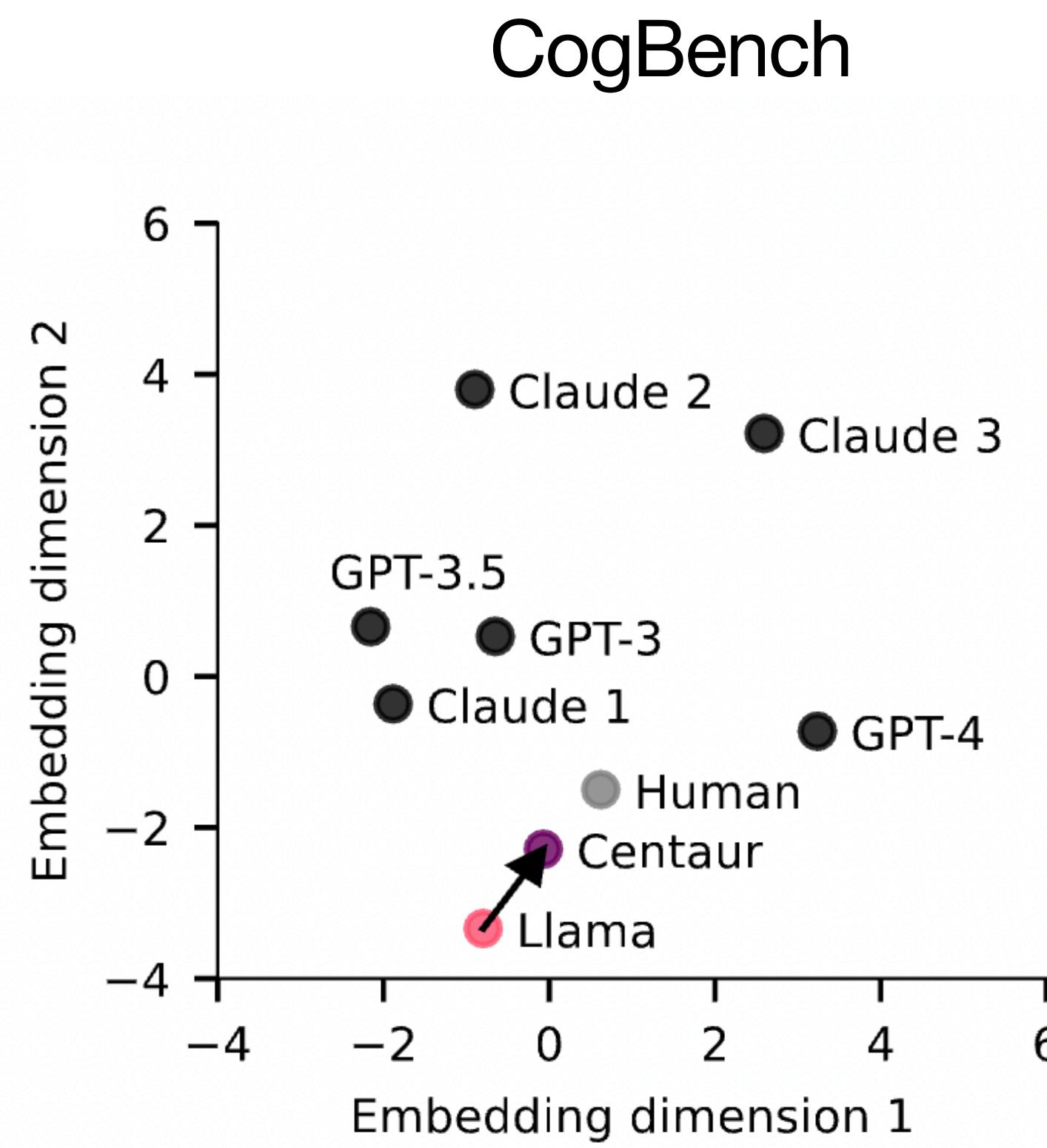
Labeling

Cognitive tasks



Labeling

Cognitive tasks



Paraphrasation

To characterize and organize decision-related proverbs

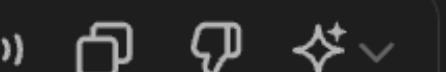


The proverb "curiosity kills the cat" warns that excessive inquisitiveness can lead to trouble or danger. It implies that exploring unknown or risky matters may result in harm, highlighting the potential perils of being overly curious or meddling in affairs best left alone.

Write a 50-word description of the proverb "fortune favors the brave"



The proverb "fortune favors the brave" suggests that bold actions often lead to success. It emphasizes that taking risks and showing courage can attract good luck and positive outcomes. This saying encourages proactive and fearless behavior, implying that opportunities are more likely to come to those who are daring.



Write a 50-word description of the proverb "the ship has sailed"



The proverb "the ship has sailed" signifies that an opportunity has been missed and cannot be reclaimed. It suggests that the moment to act has passed, often due to hesitation or delay, and now it's too late to seize the chance or achieve the desired outcome.

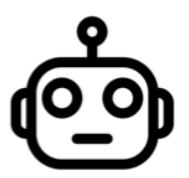
Proverb 1 of 49528

Fortune favors the bold



Summarize the gist of the proverb.

1

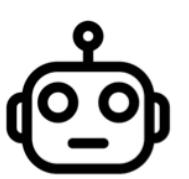


This proverb suggests that those who take risks and act decisively are more likely to succeed than those who are timid and hesitant.

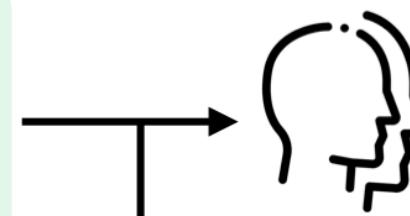


Provide taxonomic tags capturing the proverb's topic.

2



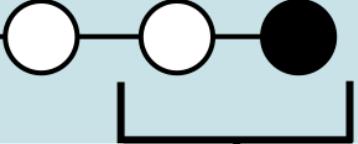
1. decision making
2. risk management
3. success



Does the proverb provide guidance on how to act or make decisions?

3



No  Yes

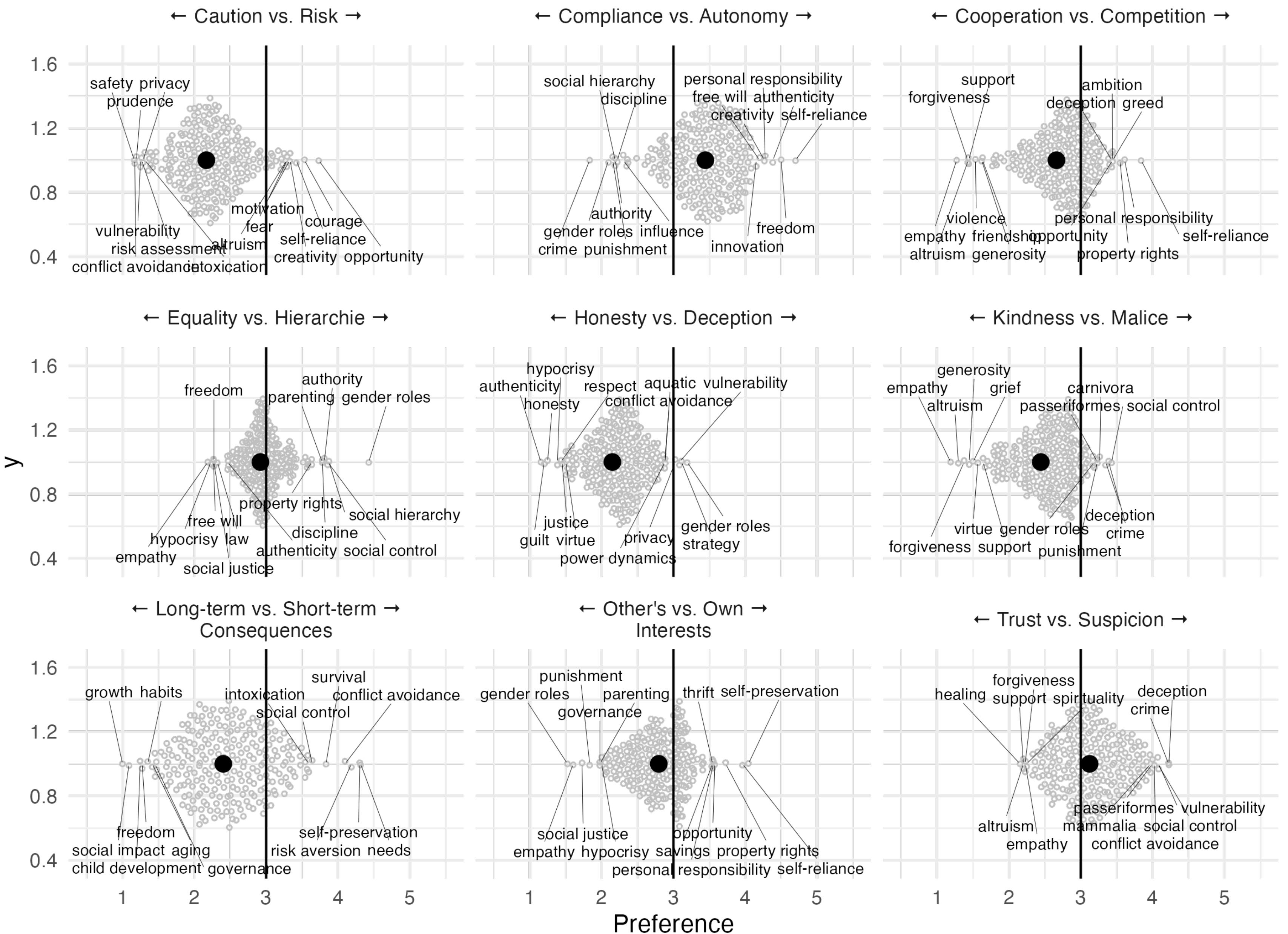


What does the proverb favor?

4

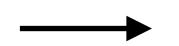
Risk		Caution
Short-term consequences		Long-term consequences
Own interests		Other's interests
Autonomy		Compliance
Trust		Suspicion
Honesty		Deception
Equality		Hierarchie
Kindness		Malice
Cooperation		Competition

Proverbs advocate safety, autonomy, cooperation, honesty, kindness, and a focus on long-term consequences and other's interests





Search in documents



American Psychological
Association
232'000 articles

Read this PDF and return a table of participant characteristics, including demographic information, for the studies included



Participant Characteristics

	Characteristic	Value
1	Sample Size	1205
2	Gender (Female)	50%
3	Age (Mean)	47.7
4	Age (SD)	16.6
5	Age Range	18-87

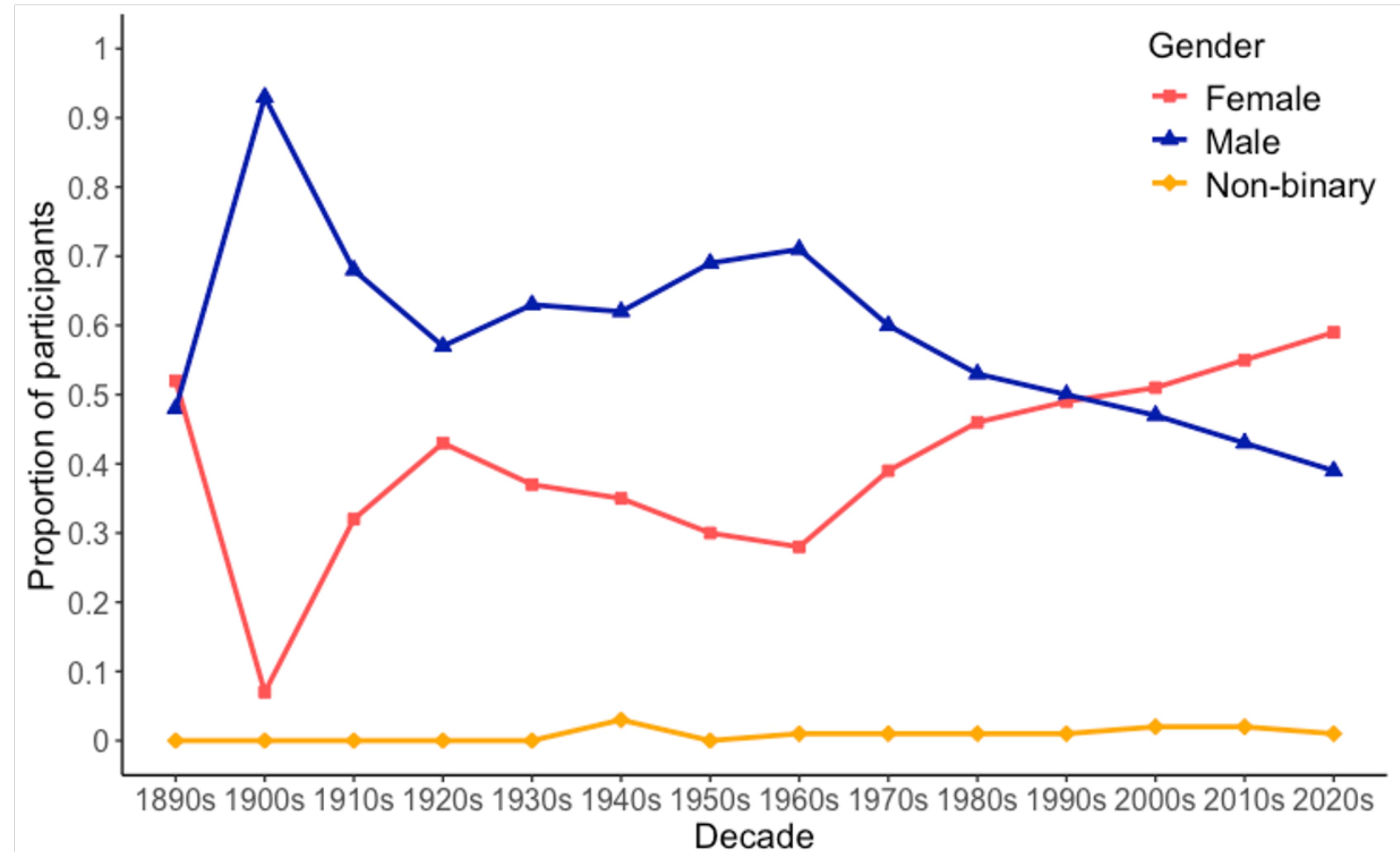
The table of participant characteristics, including demographic information for the studies included in the document, has been created and displayed for you. If you need any additional information or further analysis, please let me know! [\[x\]](#)



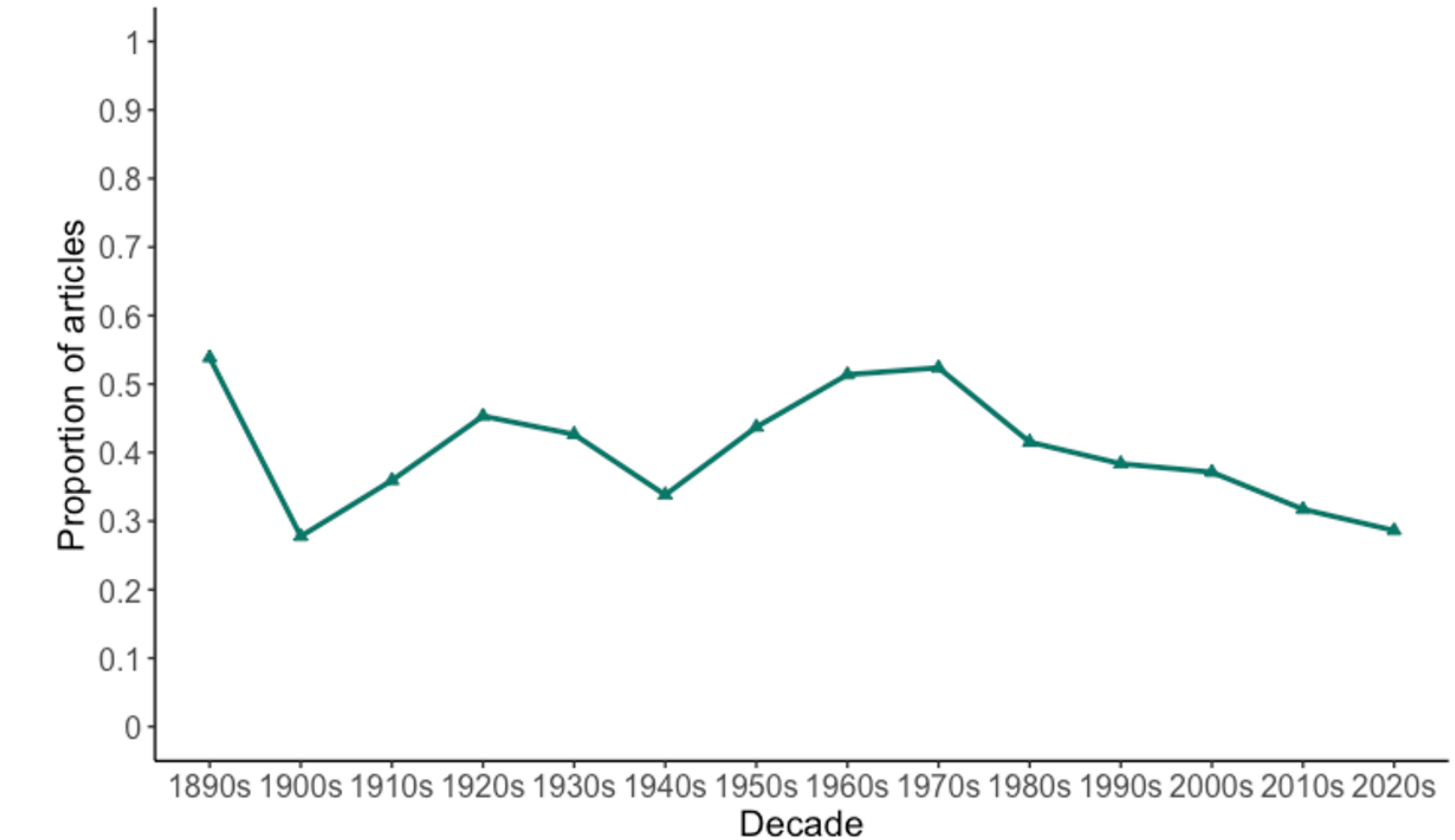
Sample sizes increase drastically



Samples are increasingly female



Still many samples are drawn from student populations



Qualitative data analysis with LLMs

Dirk Wulff



MAX PLANCK INSTITUTE
FOR HUMAN DEVELOPMENT

