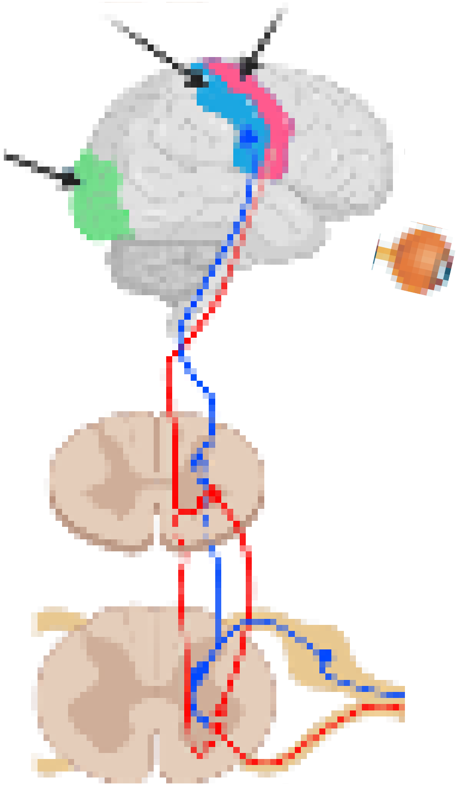
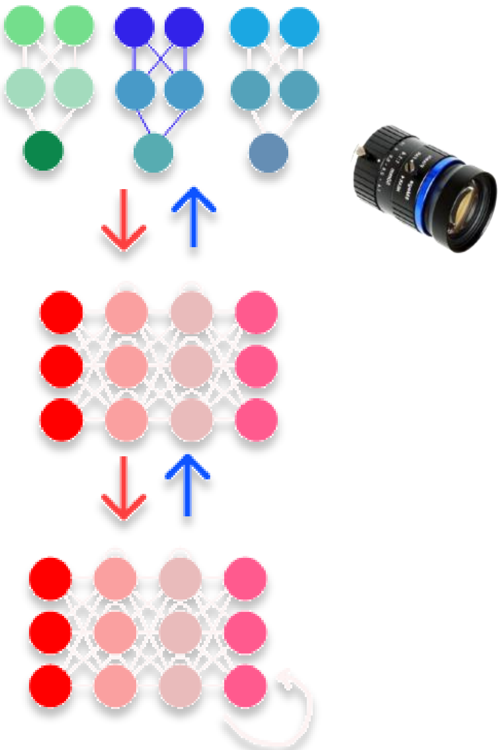
NX-414: Brain-like computation and intelligence

Martin Schrimpf

Lecture 8, 09 April 2025

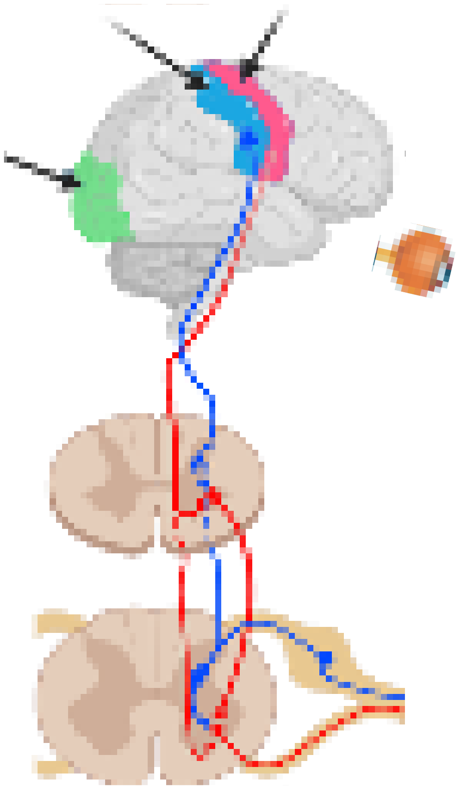
Biological Intelligence ⟷ Artificial Intelligence





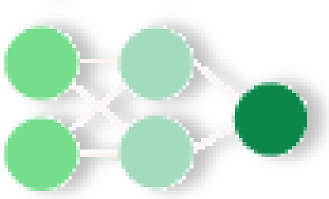
Hausmann & Marin-Vargas et al., 2021

Normative frameworks



Information theoretic

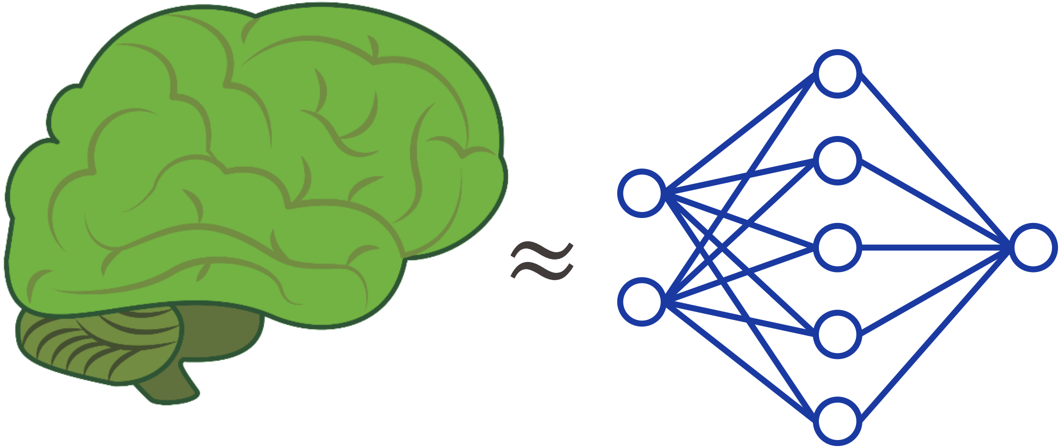
e.g. sparse coding, redundancy reduction, mutual information …



Utilitarian

e.g. recognize objects, chase prey, navigate, **next-word prediction**, …

Using deep neural networks as goal-driven models of a system



Vision: object recognition. Yamins & Hong et al. (2014), Schrimpf & Kubilius et al. (2018)

Audition: speech recognition, speaker & sound identification. Kell et al. (2018)

Somatosentation: shape recognition. Zhuang et al. (2017)

Language: next-word prediction. Schrimpf et al. (2021)

Decision making: context-dependent choice. Mante & Sussilo et al. (2013)

Proprioception: action recognition. Sandbrink et al. (2023)

Why language?

- higher-level cognitive domain (compared to sensory or motor processing)
- plays an essential role in human life
- quintessentially human

Language comprehension: the extraction of meaning from spoken, written, or signed words and sentences.

A major debate: Is language learned or innate?

“**Poverty of the stimulus**” argument by Noam Chomsky:

the linguistic stimuli that children are exposed to are insufficient to explain how they acquire such high linguistic proficiency so quickly

→ Learning alone is insufficient

→ Language must be largely innate (with a genetic disposition for syntax and symbols)

Large language models disprove the innateness of language by learning rich linguistic structure and grammar without strong innate priors or explicit symbols (Piantadosi 2023)

A bridge to higher cognition

Perception

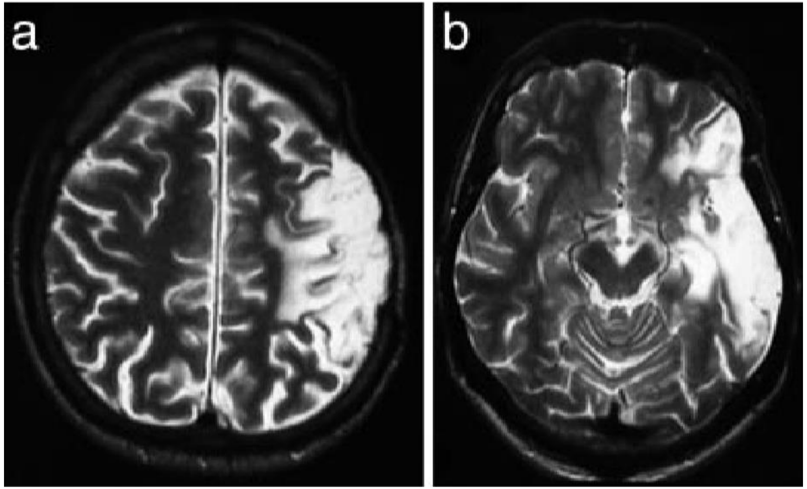
Language

High-level reasoning

Is language the same as thought?

Language is not thought

Individuals with global aphasia are unable to understand or produce language.



https://www.youtube.com/watch?v=7tu5UbpztM0

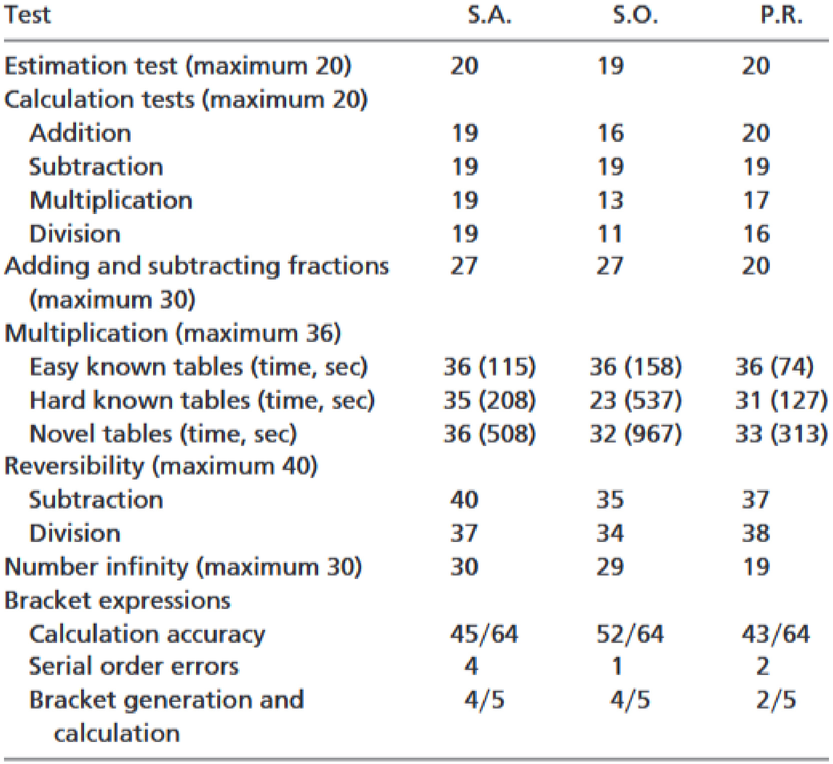
- Varley et al. 2005

Intact cognitive function in aphasia patients

Individuals with global aphasia are unable to understand or produce language.

But: they retain high performance on other cognitive tasks

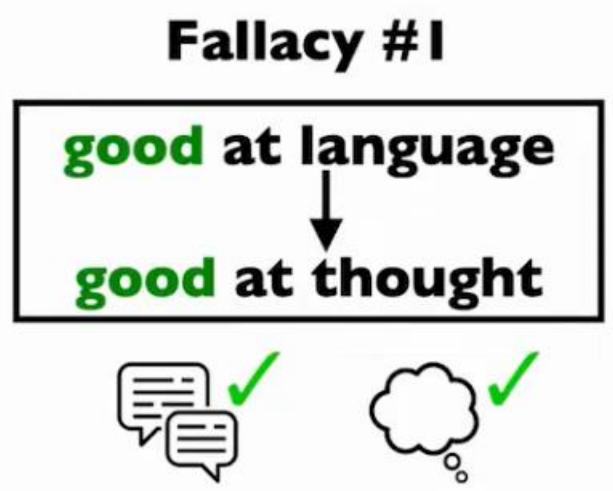
- add and subtract
- solve logic problems
- think about another person’s thoughts
- appreciate music
- navigate environments
- …



Test	S.A.	S.O.	P.R.
Estimation test (maximum 20)	20	19	20
Calculation tests (maximum 20)			
Addition	19	16	20
Subtraction	19	19	19
Multiplication	19	13	17
Division	19	11	16
Adding and subtracting fractions (maximum 30)	27	27	20
Multiplication (maximum 36)			
Easy known tables (time, sec)	36 (115)	36 (158)	36 (74)
Hard known tables (time, sec)	35 (208)	23 (537)	31 (127)
Novel tables (time, sec)	36 (508)	32 (967)	33 (313)
Reversibility (maximum 40)			
Subtraction	40	35	37
Division	37	34	38
Number infinity (maximum 30)	30	29	19
Bracket expressions			
Calculation accuracy	45/64	52/64	43/64
Serial order errors	4	1	2
Bracket generation and calculation	4/5	4/5	2/5

Varley et al. 2005
Fedorenko & Varley 2016

Fallacies in associating language with thought

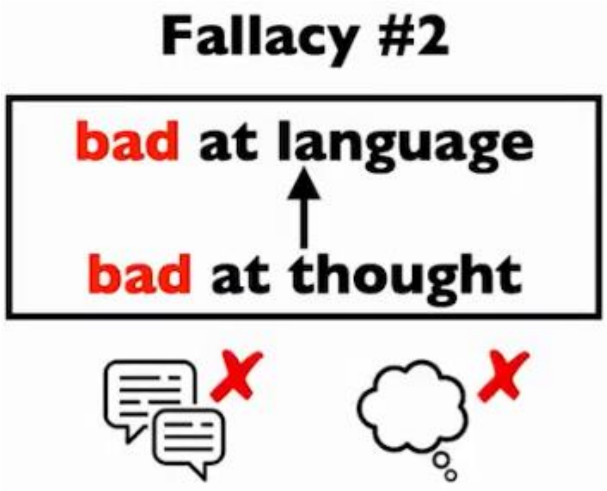


Fallacy #1

good at language

↓

good at thought



Fallacy #2

bad at language

↑

bad at thought

Mahowald & Ivanova et al. 2024

Ev Fedorenko talk at EPFL AMLD 2024

The human language system

working definition:

a set of **left-lateralized** regions on the lateral surfaces of **frontal** and **temporal** cortex that support **high-level** language processing.

Language > Perceptually matched control

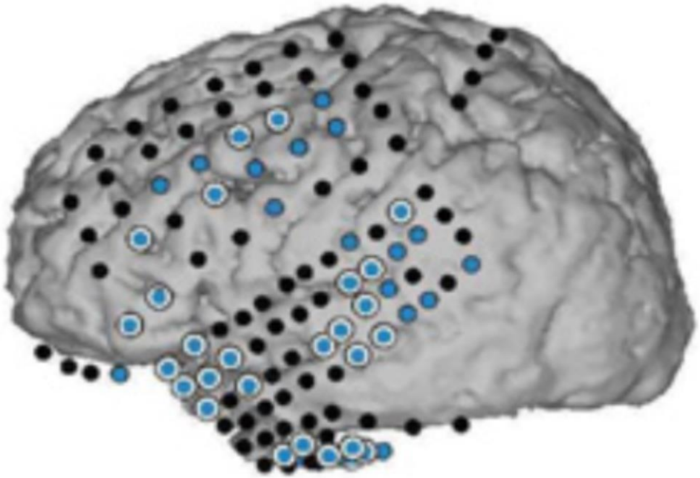
Sentences > Lists of nonwords

Fedorenko and Thompson-Schill 2014 *Braga, DiNicola and Buckner 2019*

Primary recording modalities

fMRI

non-invasive, uses super-conducting magnets to detect changes in blood flow (blood-oxygen-level dependent BOLD contrast)



ECoG

invasive, electrodes placed on the brain surface (below skull etc). Typically from epilepsy patients

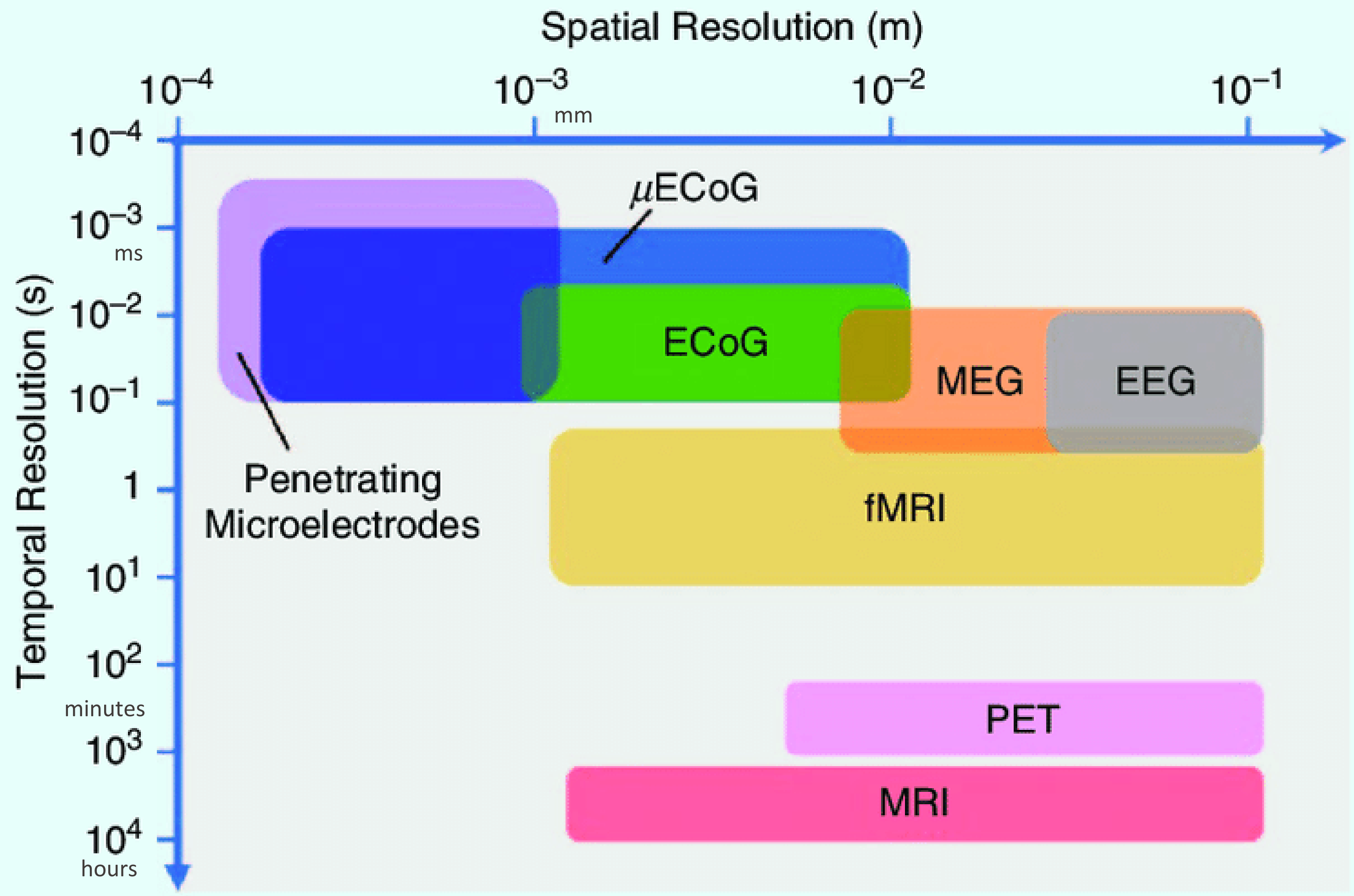
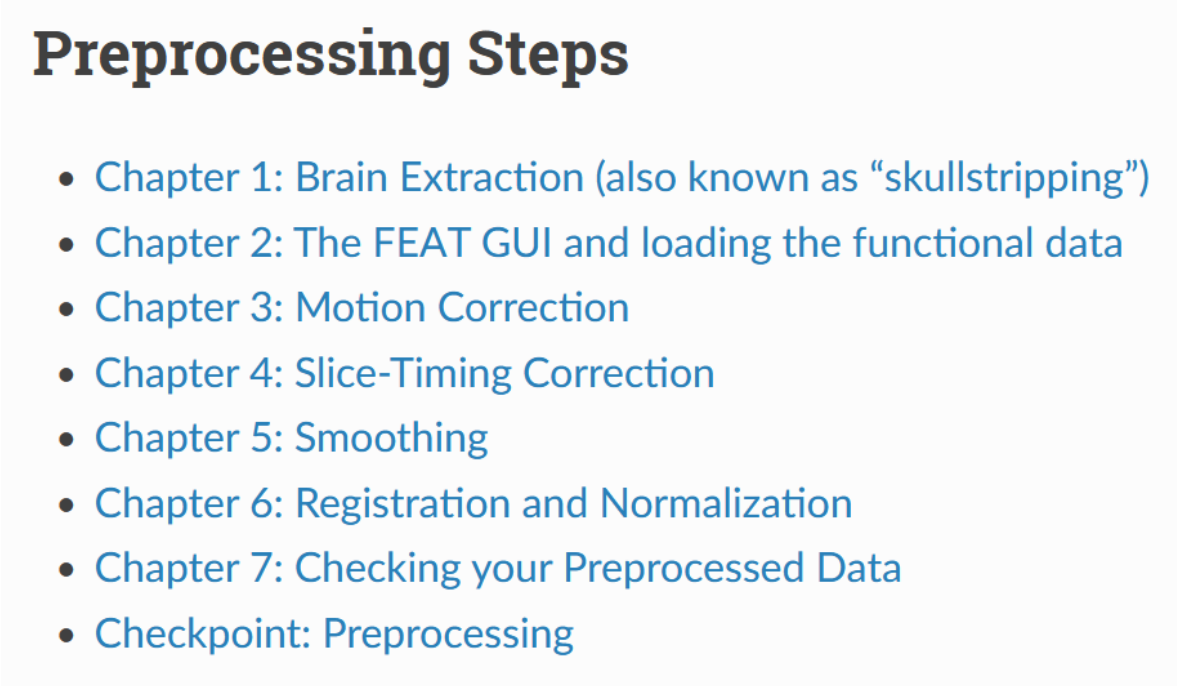


Image from Thukral et al. 2018

fMRI pre-processing is tricky

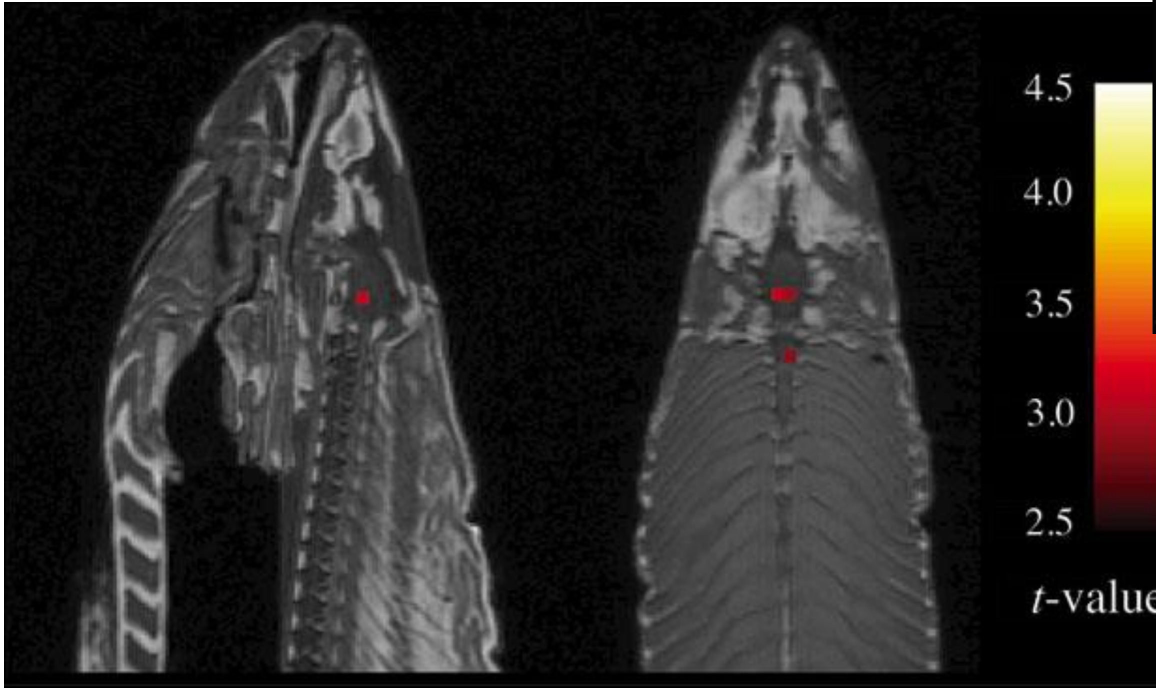


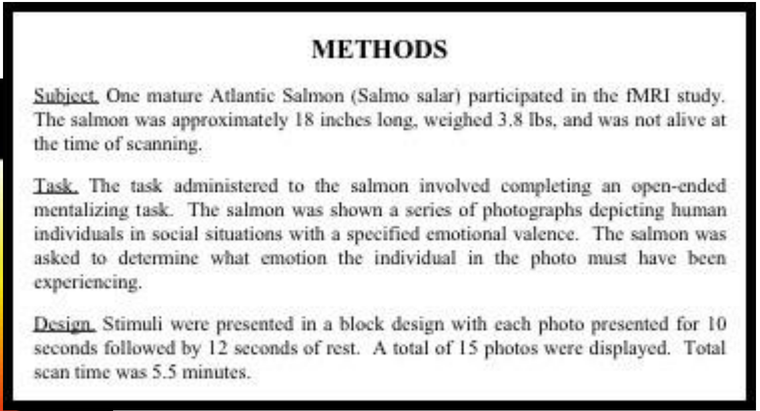
Preprocessing Steps

- Chapter 1: Brain Extraction (also known as “skullstripping”)
- Chapter 2: The FEAT GUI and loading the functional data
- Chapter 3: Motion Correction
- Chapter 4: Slice-Timing Correction
- Chapter 5: Smoothing
- Chapter 6: Registration and Normalization
- Chapter 7: Checking your Preprocessed Data
- Checkpoint: Preprocessing

https://andysbrainbook.readthedocs.io/en/latest/fMRI_Short_Course/fMRI_04_Preprocessing.html

fMRI pre-processing is tricky





METHODS

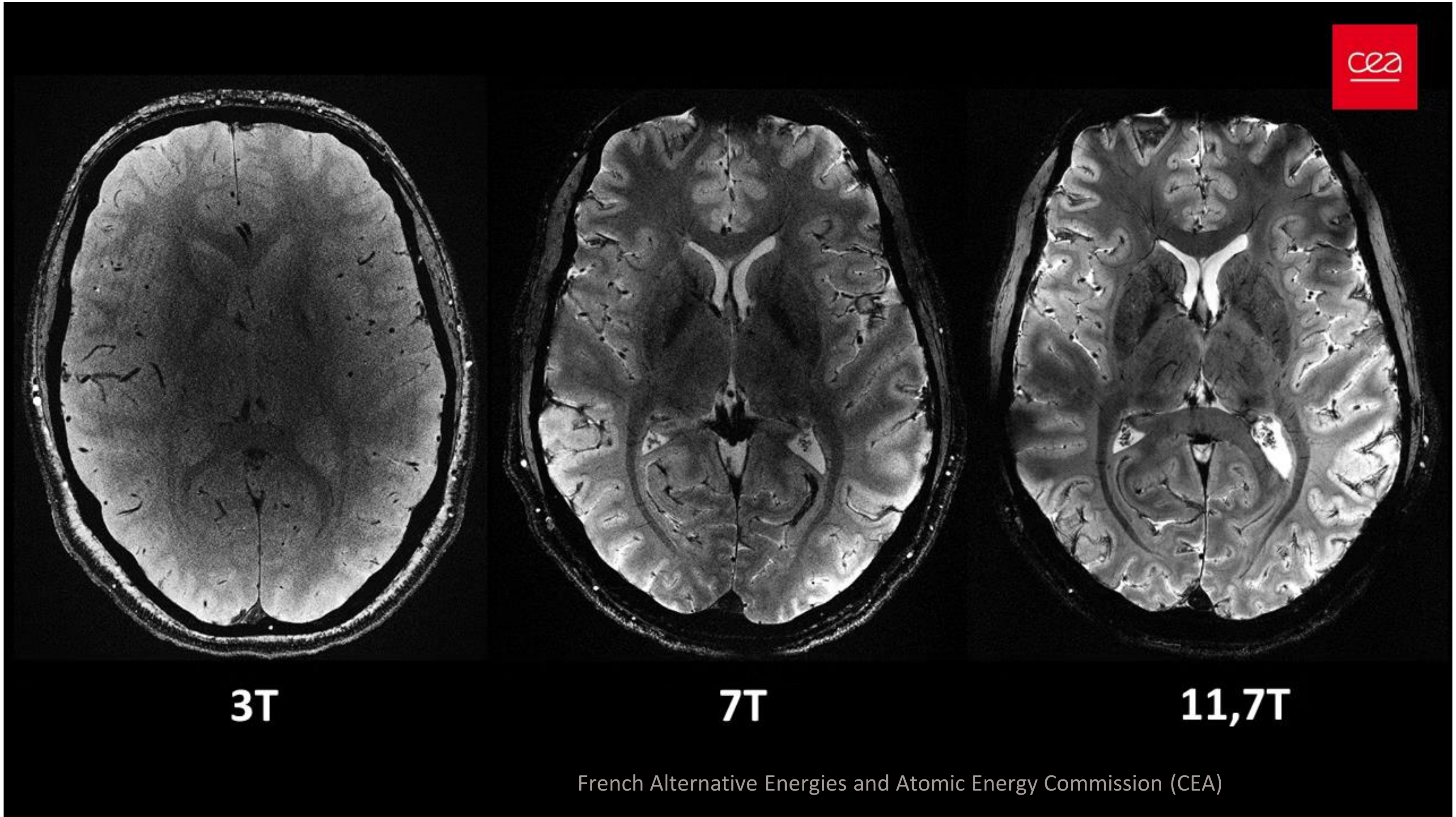
Subject. One mature Atlantic Salmon (Salmo salar) participated in the fMRI study. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was not alive at the time of scanning.

Task. The task administered to the salmon involved completing an open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

Design. Stimuli were presented in a block design with each photo presented for 10 seconds followed by 12 seconds of rest. A total of 15 photos were displayed. Total scan time was 5.5 minutes.

Bennett & Miller, 2010

fMRI news 2024: more powerful magnets



French Alternative Energies and Atomic Energy Commission (CEA)

four minutes for images down to 0.2 mm

The human language system

working definition:

a set of **left-lateralized** regions on the lateral surfaces of **frontal** and **temporal** cortex that support **high-level** language processing.

Language > Perceptually matched control

Sentences > Lists of nonwords

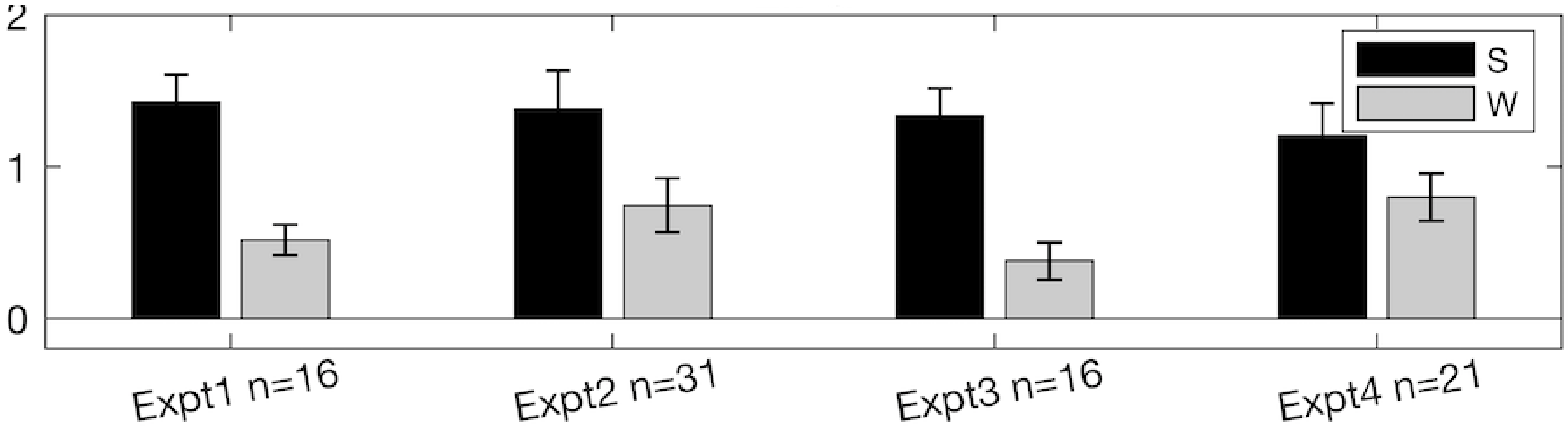
Fedorenko and Thompson-Schill 2014 Braga, DiNicola and Buckner 2019

The human language system (ECoG data)

the dog is taking a bath

>

dap drello smop ub plid kav



Key signature: stronger response to sentences than lists of unconnected words

Fedorenko, Behr and Kanwisher 2011 | Fedorenko et al. 2020

The human language system (fMRI)

courtesy of Idan Blank

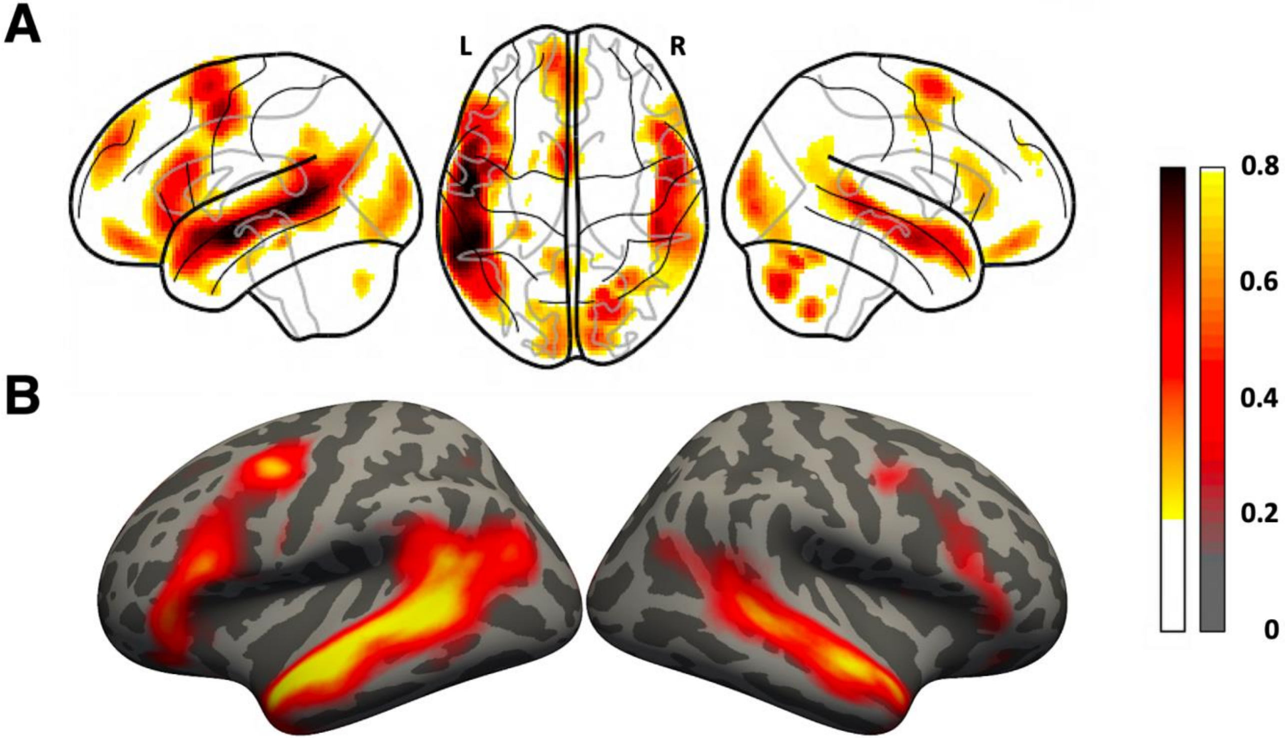


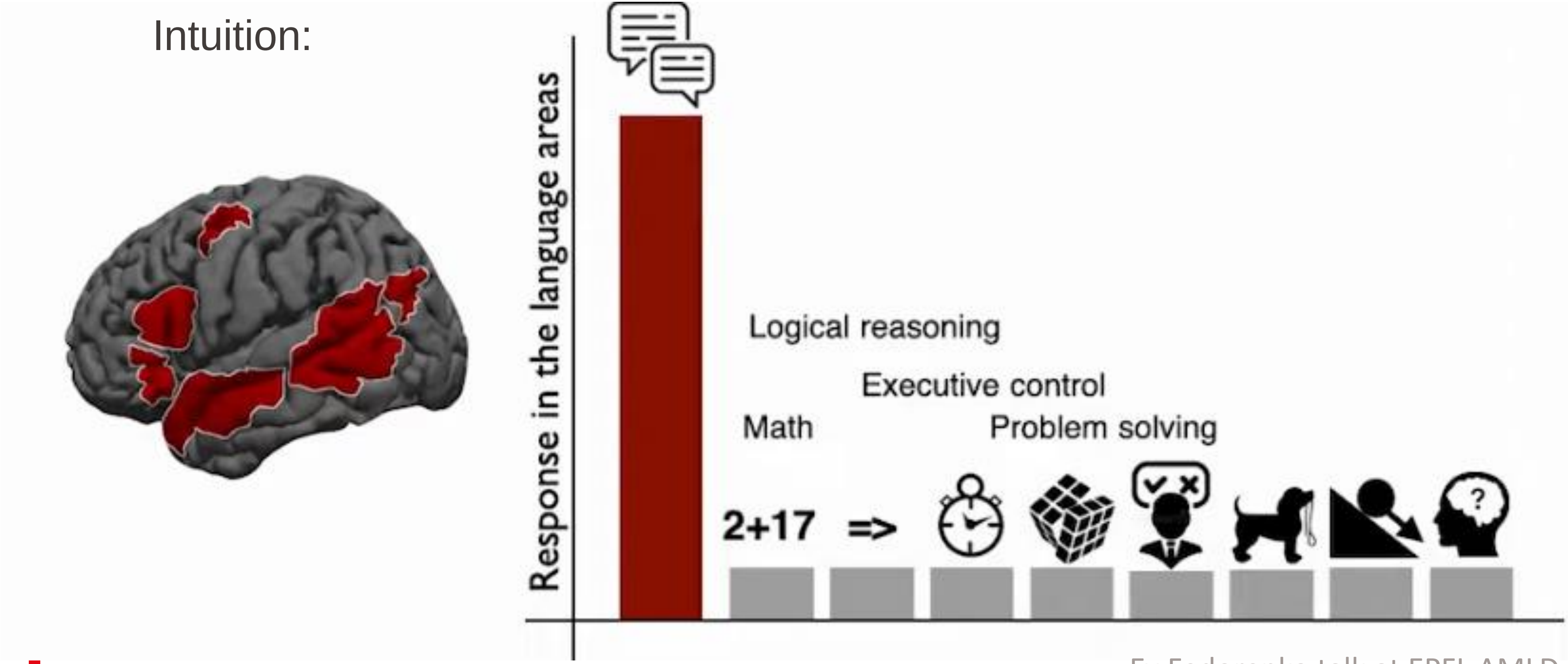
Figure 1: Probabilistic functional atlas for the *language > control* contrast based on overlaid individual binarized activation maps (where in each map, the top 10% of voxels are selected, as described in the text). *A)* SPM-analyzed volume data in the MNI template space (based on 806 individual maps). *B)* FreeSurfer-analyzed surface data in the FSaverage template space (based on 804 individual maps). In both figures, the color scale reflects the proportion of participants for whom that voxel belongs to the top 10% of *language > control* voxels.

Lipkin et al. 2024

The human language system does not perform thought

Language areas show reduced response when we engage in diverse thought-related activities.

Intuition:

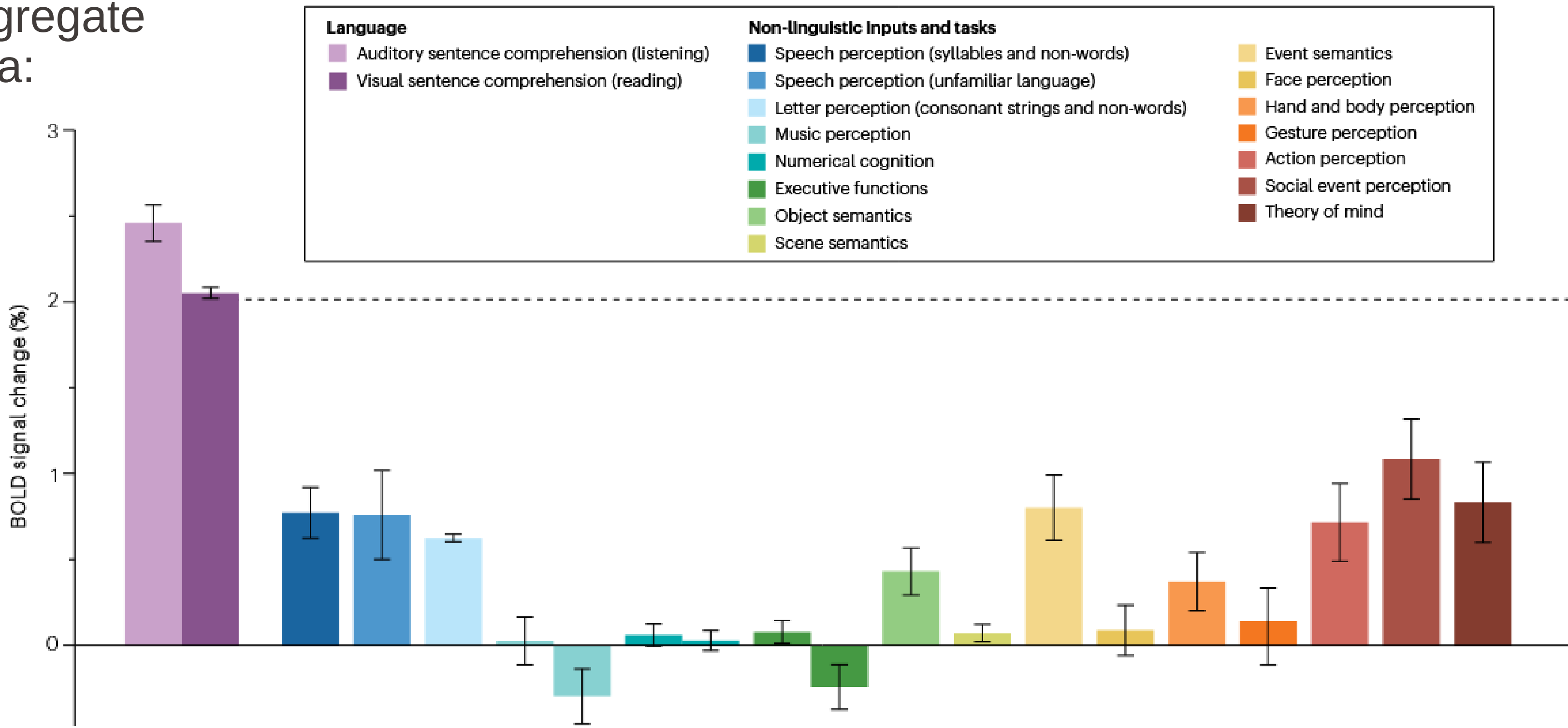


Ev Fedorenko talk at EPFL AMLD 2024

The human language system does not perform thought

Language areas show reduced response when we engage in diverse thought-related activities.

Aggregate data:

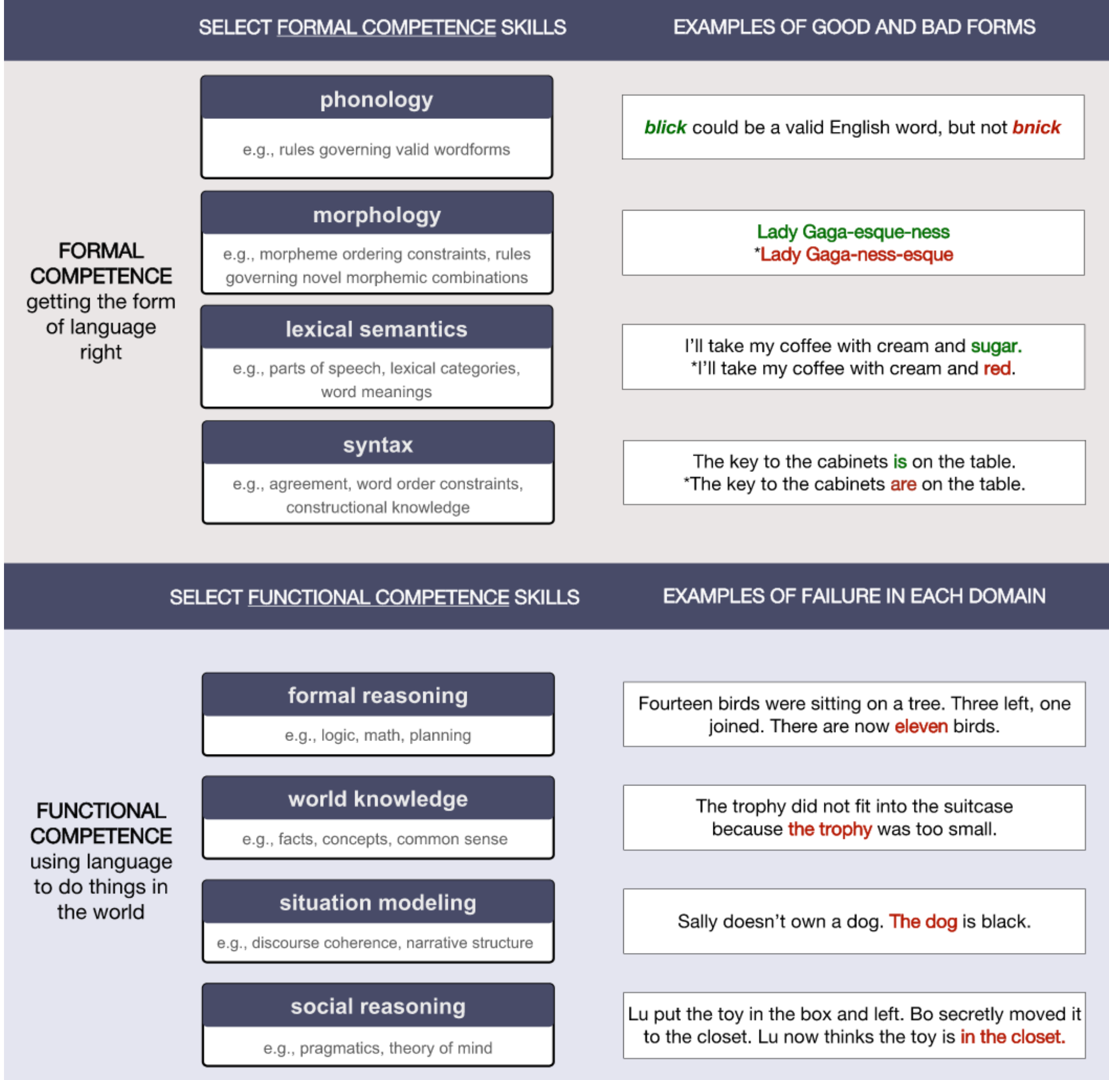


- Fedorenko et al. 2024

Formal vs functional linguistic competencies

Successfully using language requires language-specific formal competence as well as functional competence.

Mahowald & Ivanova et al. 2024

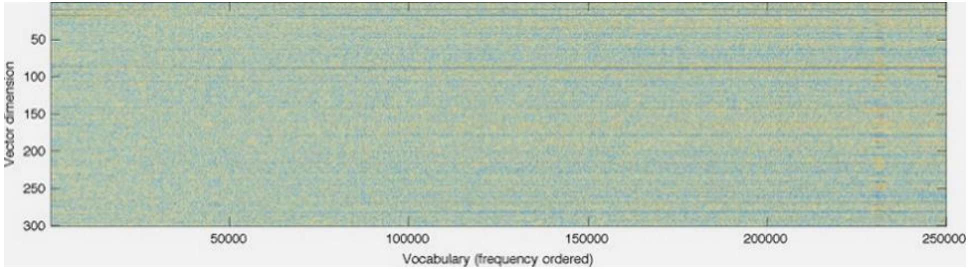


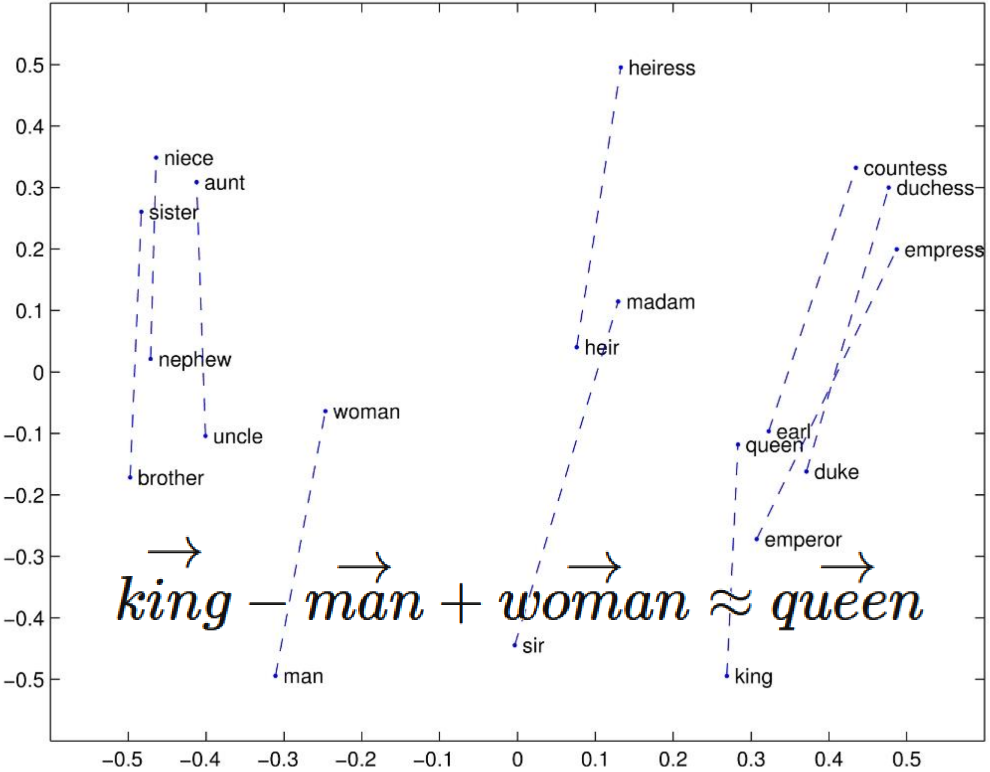
	SELECT FORMAL COMPETENCE SKILLS	EXAMPLES OF GOOD AND BAD FORMS
FORMAL COMPETENCE getting the form of language right	**phonology** e.g., rules governing valid wordforms	***blick*** could be a valid English word, but not ***bnick***
	morphology e.g., morpheme ordering constraints, rules governing novel morphemic combinations	Lady Gaga-esque-ness *Lady Gaga-ness-esque
	lexical semantics e.g., parts of speech, lexical categories, word meanings	I'll take my coffee with cream and sugar. *I'll take my coffee with cream and red.
	syntax e.g., agreement, word order constraints, constructional knowledge	The key to the cabinets is on the table. *The key to the cabinets are on the table.

	SELECT FUNCTIONAL COMPETENCE SKILLS	EXAMPLES OF FAILURE IN EACH DOMAIN
FUNCTIONAL COMPETENCE using language to do things in the world	**formal reasoning** e.g., logic, math, planning	Fourteen birds were sitting on a tree. Three left, one joined. There are now eleven birds.
	world knowledge e.g., facts, concepts, common sense	The trophy did not fit into the suitcase because the trophy was too small.
	situation modeling e.g., discourse coherence, narrative structure	Sally doesn't own a dog. The dog is black.
	social reasoning e.g., pragmatics, theory of mind	Lu put the toy in the box and left. Bo secretly moved it to the closet. Lu now thinks the toy is in the closet.

Modeling language

Embedding type models: GloVe





$$\vec{king} - \vec{man} + \vec{woman} \approx \vec{queen}$$

Trained with lexical co-occurrence

Nearest neighbors

0. *frog*
1. frogs
2. toad
3. litoria
4. leptodactylidae
5. rana
6. lizard
7. eleutherodactylus

3. litoria

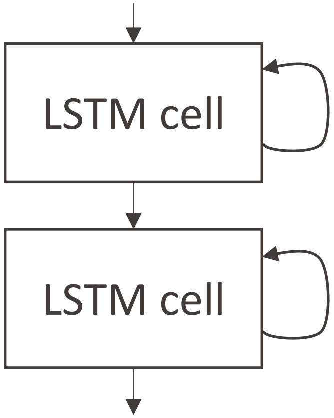
4. leptodactylidae

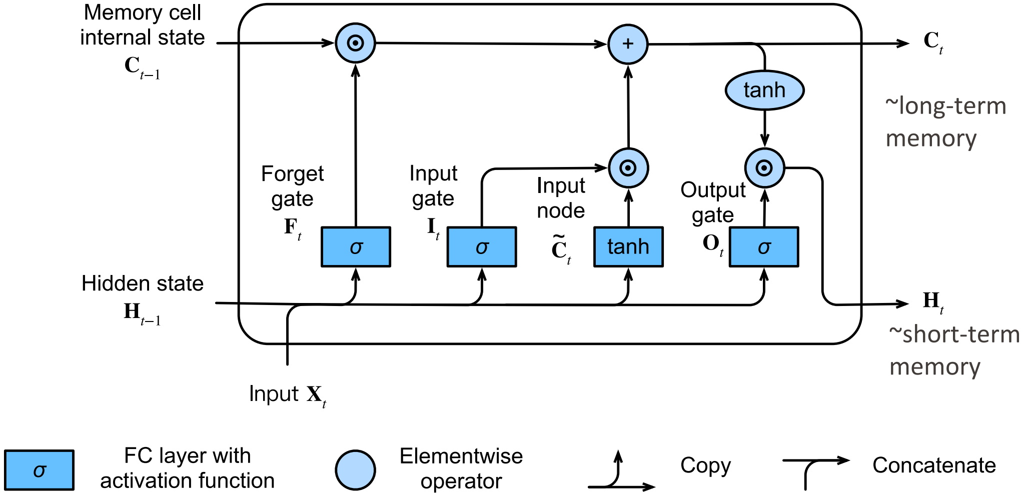
Pennington, Socher, Manning 2014

Modeling language

Embedding type models: GloVe, word2vec, topicETM

Recurrent networks: LSTM, skip-thoughts





~long-term memory

~short-term memory

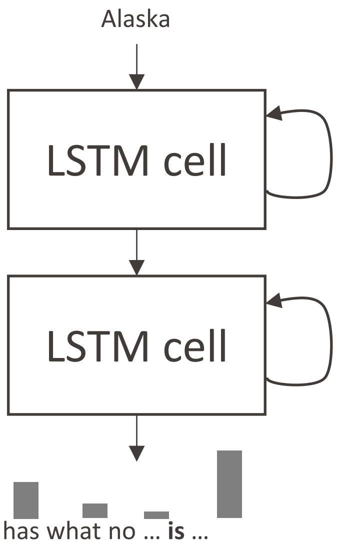
Hochreiter & Schmidhuber 1997

Image from https://d2l.ai/chapter_recurrent-modern/lstm.html

Modeling language

Embedding type models: GloVe, word2vec, topicETM

Recurrent networks: LSTM, skip-thoughts



Alaska

LSTM cell

LSTM cell

has what no … **is** …

Problem: backpropagation through time often leads to vanishing gradients

Typical training objective:
Language Modeling
(minimize perplexity/surprisal)

Alaska is

Alaska is about

Alaska is about twelve

Alaska is about twelve times

Alaska is about twelve times larger

Alaska is about twelve times larger than

Alaska is about twelve times larger than New

Alaska is about twelve times larger than New York

Hochreiter & Schmidhuber 1997

Modeling language

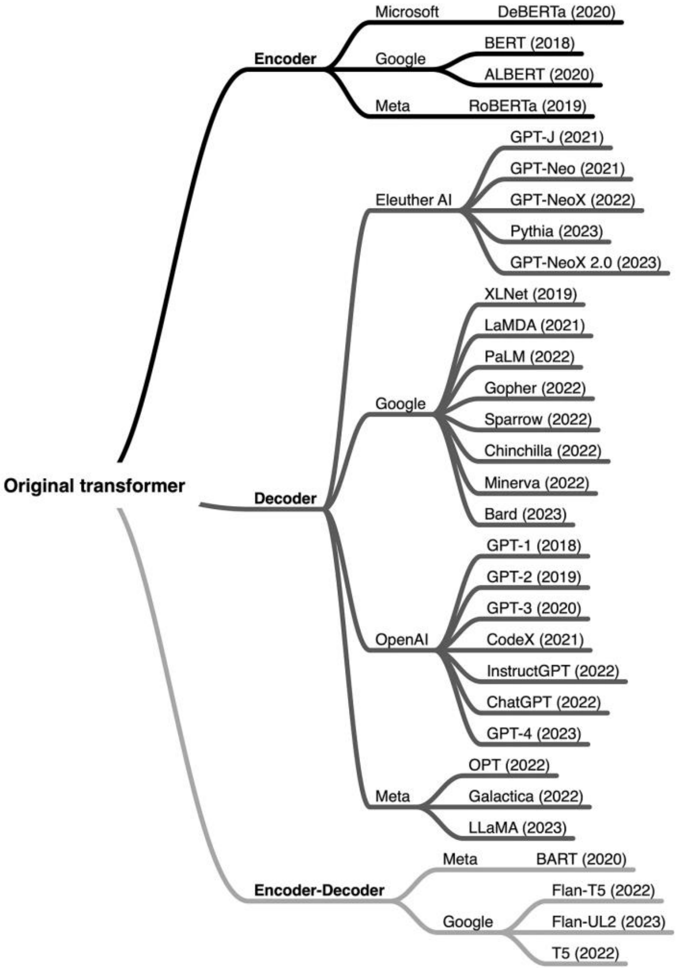
Embedding type models: GloVe

Recurrent networks: LSTM

Transformers (investigated in paper)

- BERTs
- RoBERTas
- XLMs
- Transformer-XLs
- XLNets
- CTRL
- T5s
- AlBERTs
- GPTs

More recent: LLaMA, Gemini, Qwen, Claude, …



Original transformer

Encoder
Microsoft DeBERTa (2020)
Google BERT (2018)
ALBERT (2020)
Meta RoBERTa (2019)

Decoder
Eleuther AI GPT-J (2021)
GPT-Neo (2021)
GPT-NeoX (2022)
Pythia (2023)
GPT-NeoX 2.0 (2023)
Google XLNet (2019)
LaMDA (2021)
PaLM (2022)
Gopher (2022)
Sparrow (2022)
Chinchilla (2022)
Minerva (2022)
Bard (2023)
OpenAI GPT-1 (2018)
GPT-2 (2019)
GPT-3 (2020)
CodeX (2021)
InstructGPT (2022)
ChatGPT (2022)
GPT-4 (2023)
Meta OPT (2022)
Galactica (2022)
LLaMA (2023)

Encoder-Decoder
Meta BART (2020)
Google Flan-T5 (2022)
Flan-UL2 (2023)
T5 (2022)

e.g. Pennington et al. 2014 | Jozefowicz et al. 2016 | Vaswani, Shazeer*, Parmar*, Uszkoreit*, Jones*, Gomez*, Kaiser*, Polosukhin* 2017 | Devlin et al. 2018*

Transformers

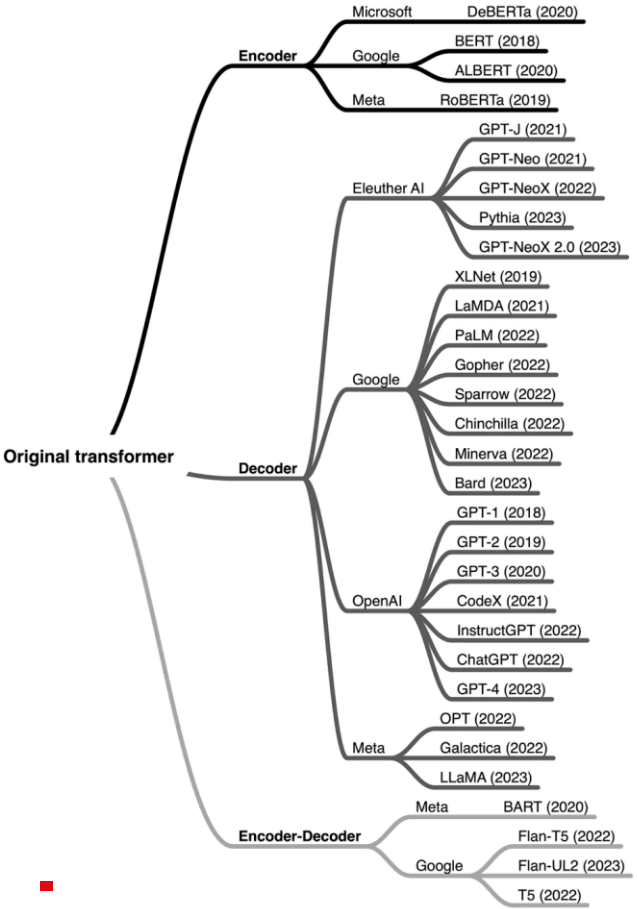


Illustration from https://magazine.sebastianraschka.com/p/understanding-encoder-and-decoder

Recent notable releases:

- Qwen 2.5

https://qwenlm.github.io/blog/qwen2.5-max/

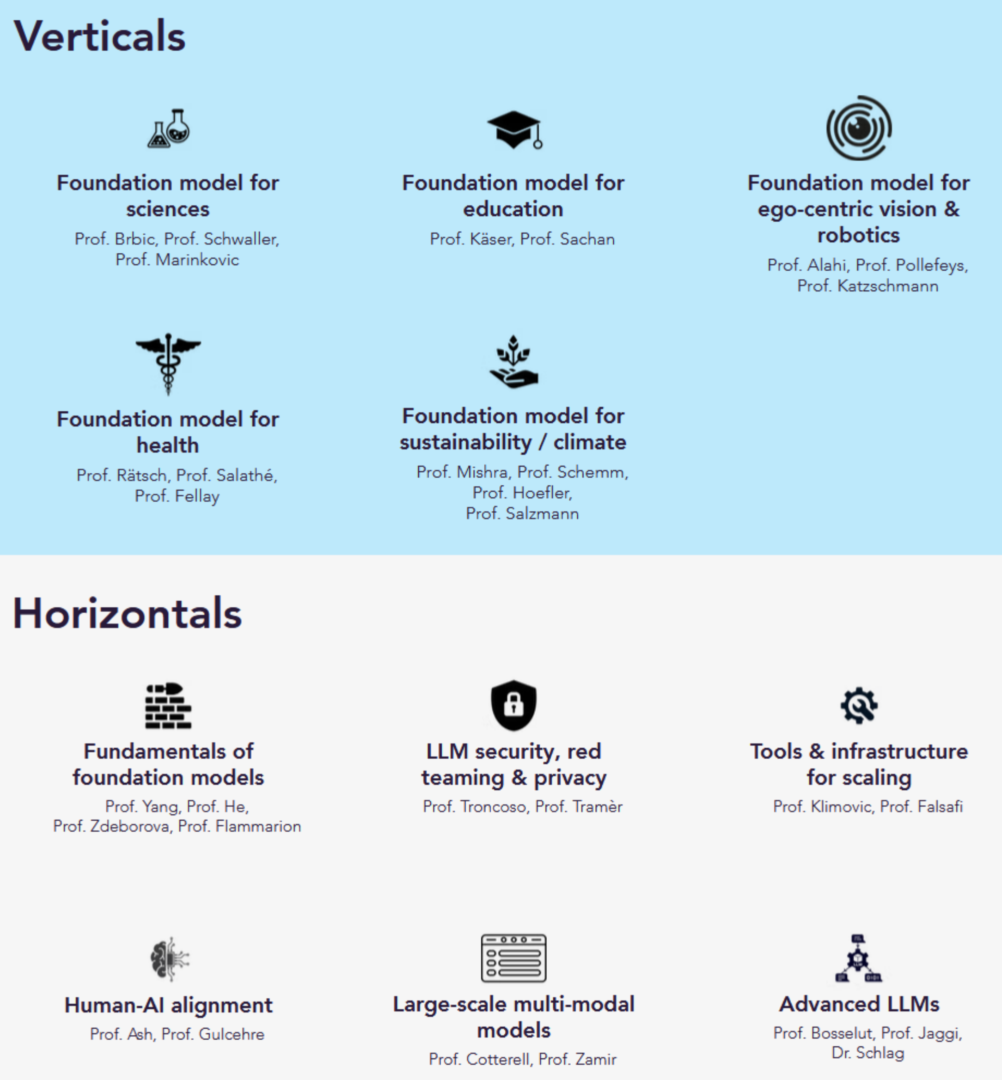
- LLaMA 4

https://ai.meta.com/blog/llama-4-multimodal-intelligence/

NB: SwissAI

- Alps supercomputer: one of the most powerful research computing clusters – built for AI
- >10’000 NVIDIA Grace Hopper GPUs, millions of GPU hours
- Consumes ~10 MW at full load, ~as much as two Swiss trains
- Large-scale AI for the benefit of

- https://www.swiss-ai.org/



Verticals

Foundation model for sciences
Prof. Brbic, Prof. Schwaller, Prof. Marinkovic

Foundation model for education
Prof. Käser, Prof. Sachan

Foundation model for ego-centric vision & robotics
Prof. Alahi, Prof. Pollefeys, Prof. Katzschmann

Foundation model for health
Prof. Rätsch, Prof. Salathé, Prof. Fellay

Foundation model for sustainability / climate
Prof. Mishra, Prof. Schemm, Prof. Hoefler, Prof. Salzmann

Horizontals

Fundamentals of foundation models
Prof. Yang, Prof. He, Prof. Zdeborova, Prof. Flammarion

LLM security, red teaming & privacy
Prof. Troncoso, Prof. Tramèr

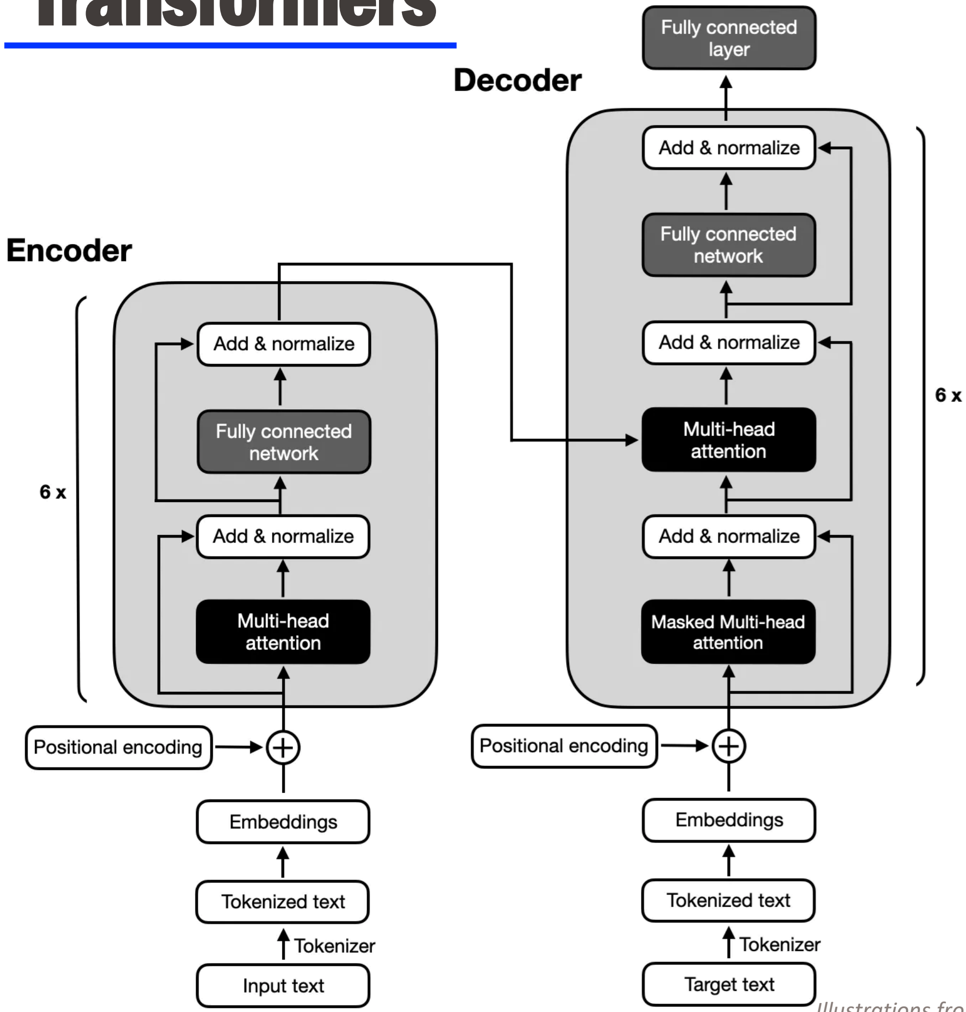
Tools & infrastructure for scaling
Prof. Klimovic, Prof. Falsafi

Human-AI alignment
Prof. Ash, Prof. Gulcehre

Large-scale multi-modal models
Prof. Cotterell, Prof. Zamir

Advanced LLMs
Prof. Bosselut, Prof. Jaggi, Dr. Schlag

Transformers



Typical architecture: blocks of

- **Multi-head attention**
- **MLP** (fully-connected network)
- **Decoder** blocks mask blocks to prevent seeing the future, and cross-attend to encoder

Stack these blocks

- **Encoder-decoder**: e.g., original transformer, T5
- **Encoder-only**: e.g., BERT, MPNet
- **Decoder-only**: e.g. GPT-4, LLaMA, Gemini.
  Most popular now (self-attention + MLP).

Illustrations from https://magazine.sebastianraschka.com/p/understanding-encoder-and-decoder

Multi-Head Attention

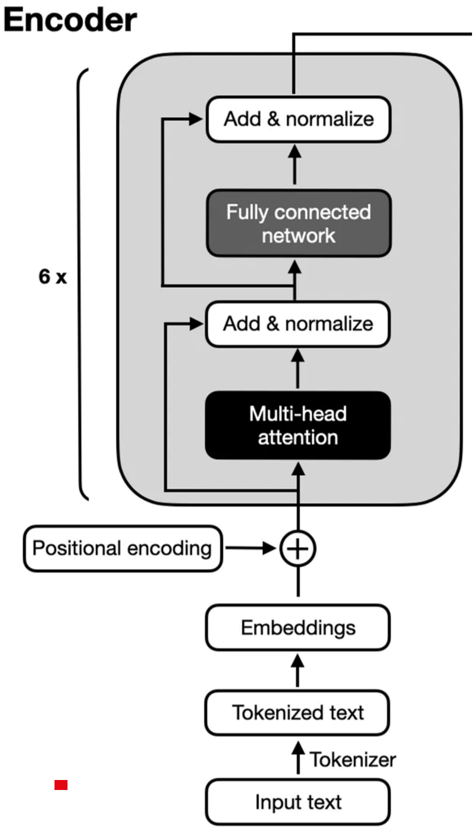
$$Attention(Q,K,V)$$
$$= softmax(\frac{QK^T}{\sqrt{d_k}})V$$

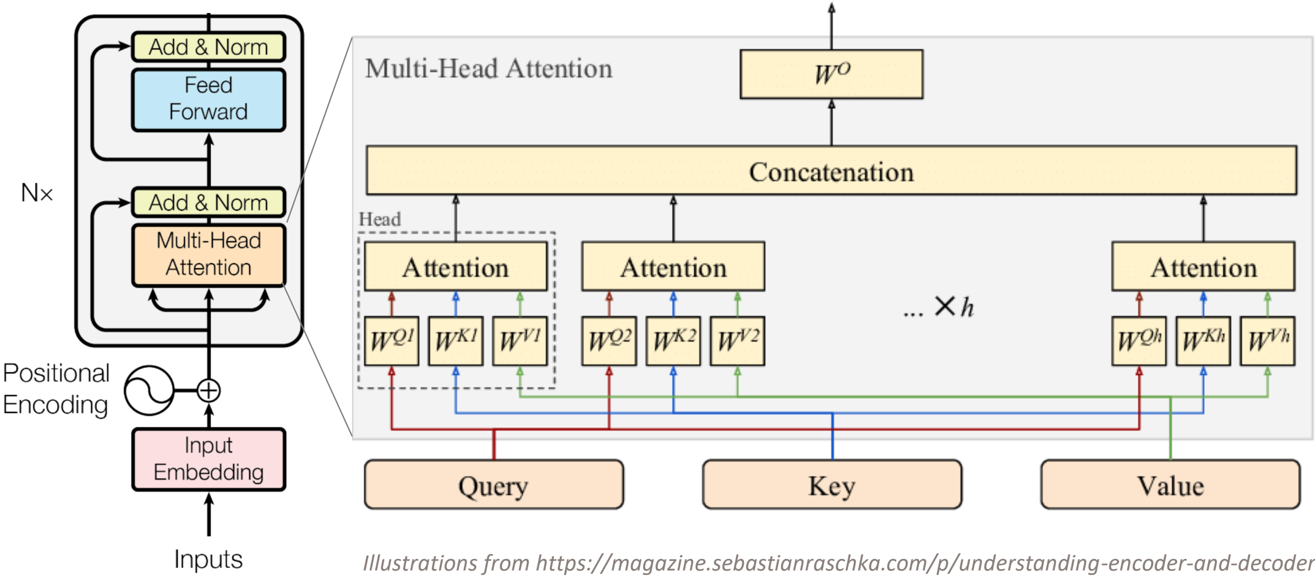
QKV

Query: what am I looking for?

Key: what do I have?

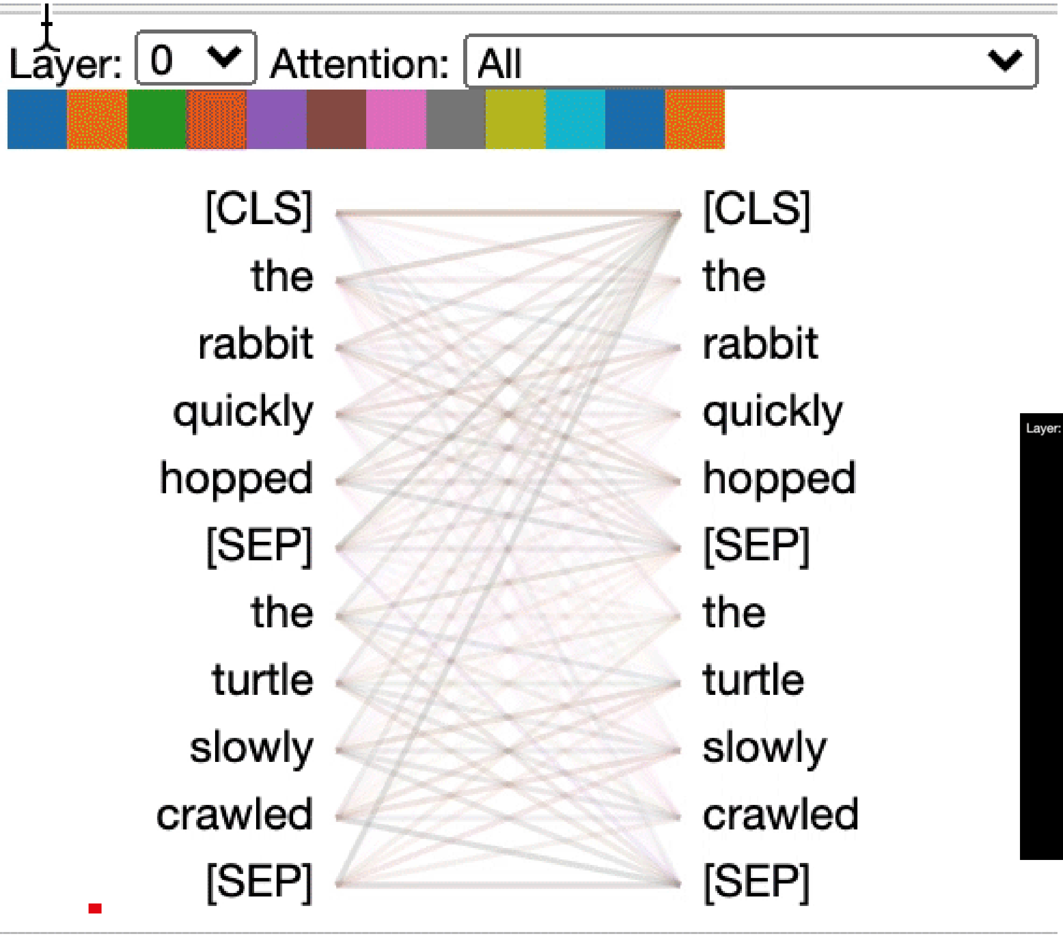
Value: what will I communicate?



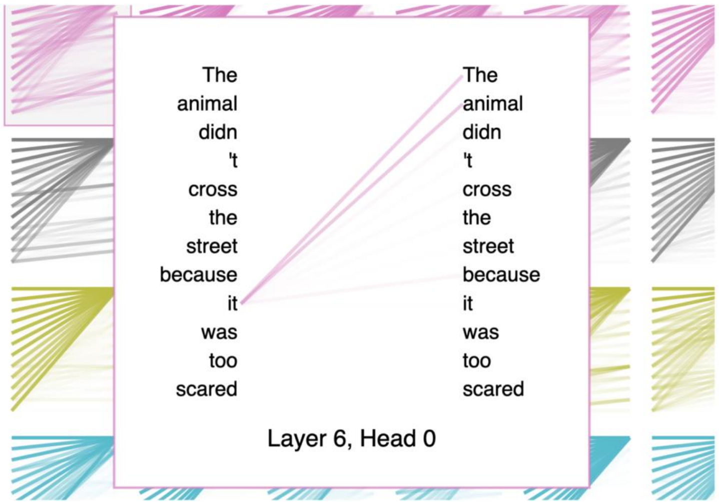


Illustrations from https://magazine.sebastianraschka.com/p/understanding-encoder-and-decoder

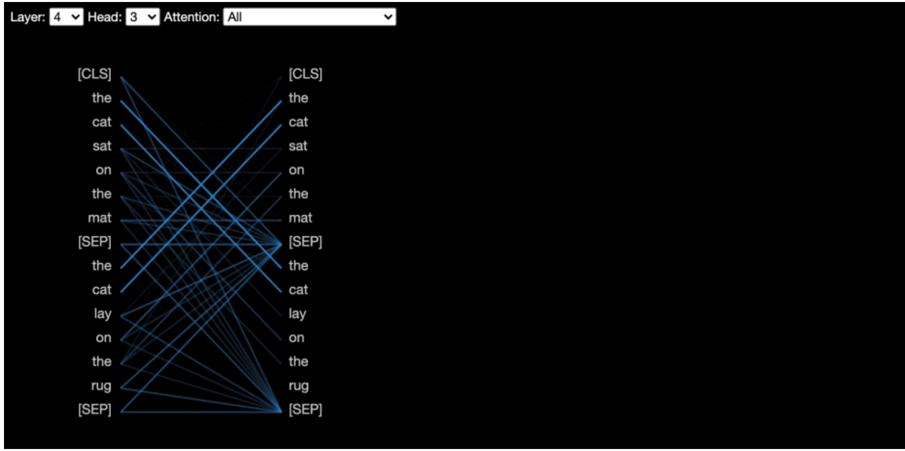
Multi-Head Attention



Layer: 0 Attention: All



Layer 6, Head 0

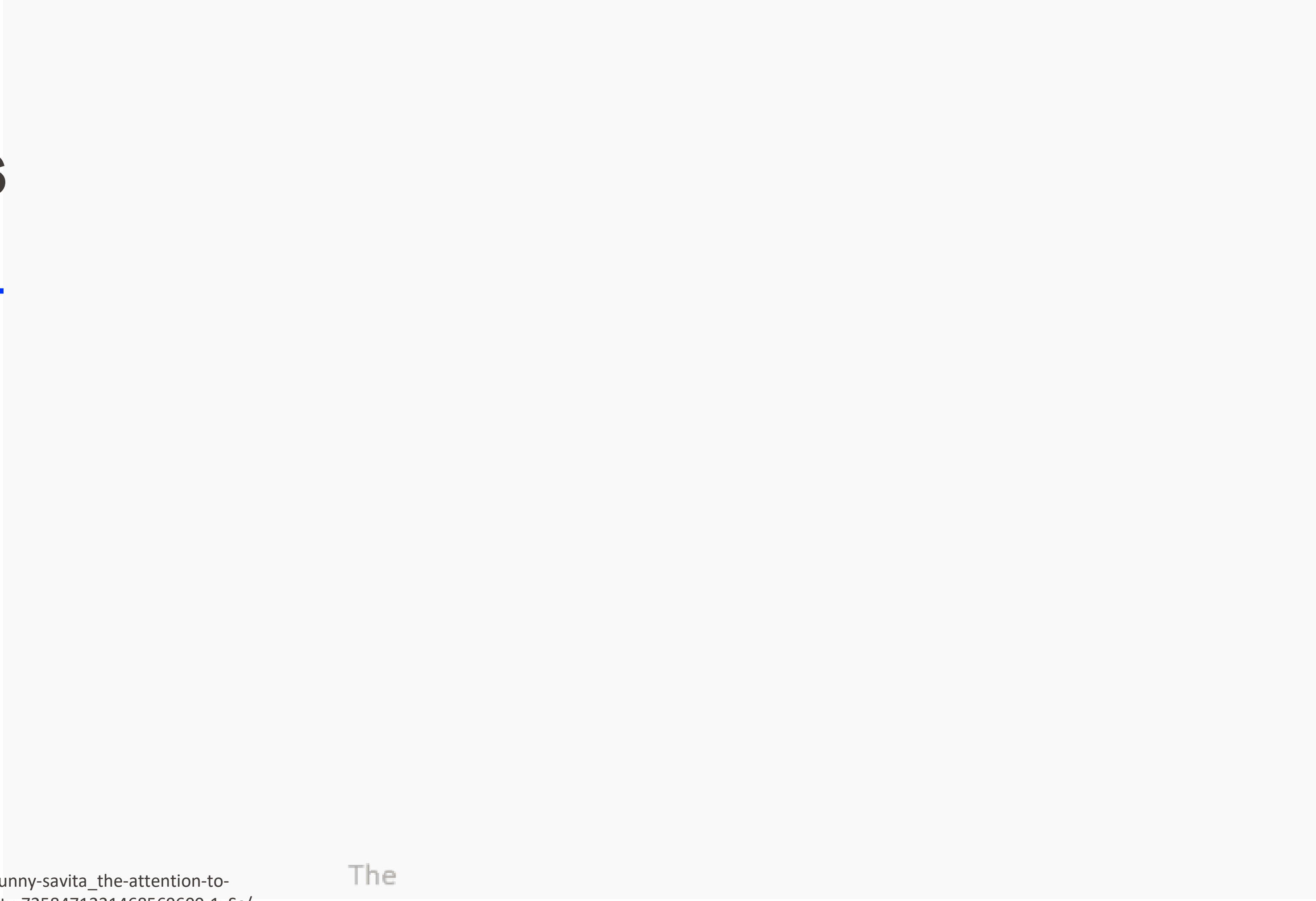


Layer: 4 Head: 3 Attention: All

https://github.com/jessevig/bertviz

https://www.comet.com/site/blog/explainable-ai-for-transformers/

Transformers from scratch



- https://www.linkedin.com/posts/sunny-savita_the-attention-to-detail-in-this-gif-is-incredible-activity-7258471221468569600-1ySc/

LLM scaling laws

- Current ML industry bet: more compute = better models
- This bet has worked out quite well in recent years

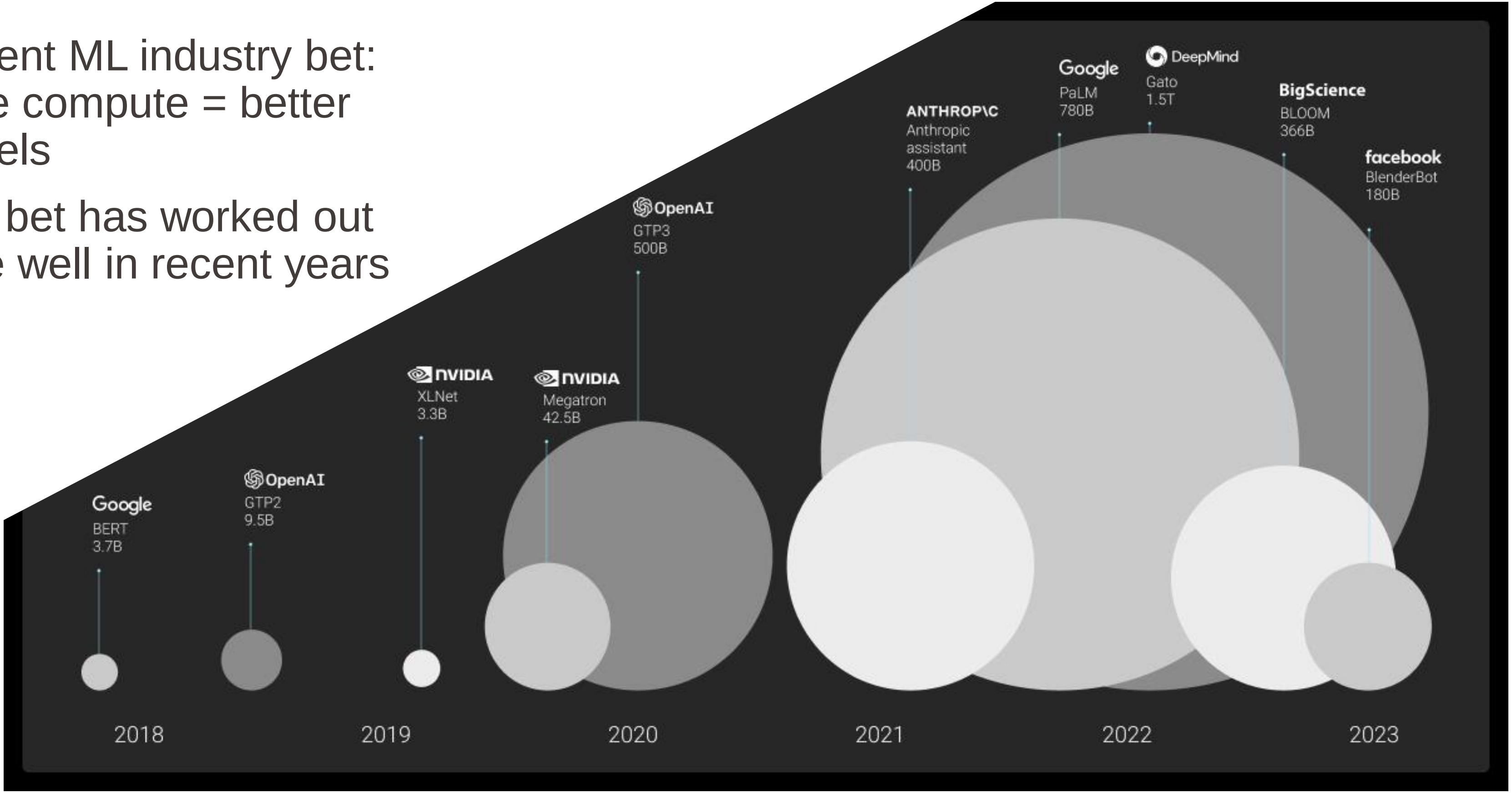
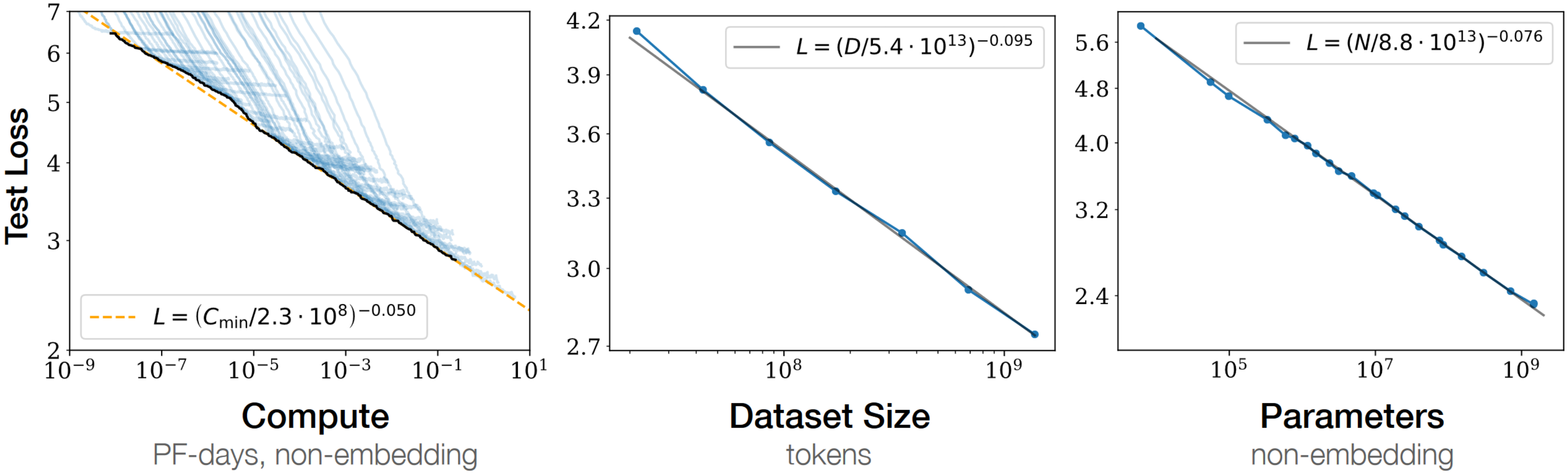


Illustration from https://labelyourdata.com/articles/llm-model-size

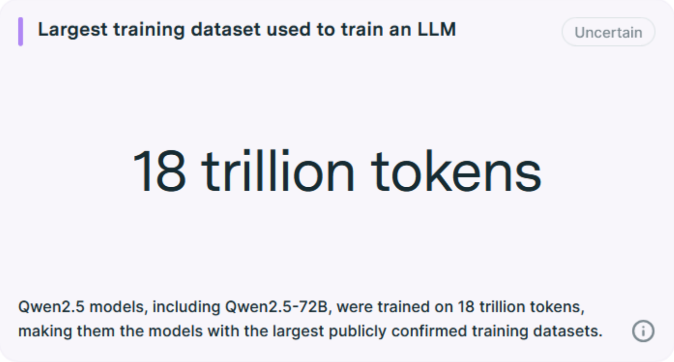
LLM scaling laws

How to most efficiently spend compute budget for the most powerful model?
Figure out how to allocate FLOPs to training tokens and model parameters



Kaplan et al. 2021; Hoffmann et al. 2022

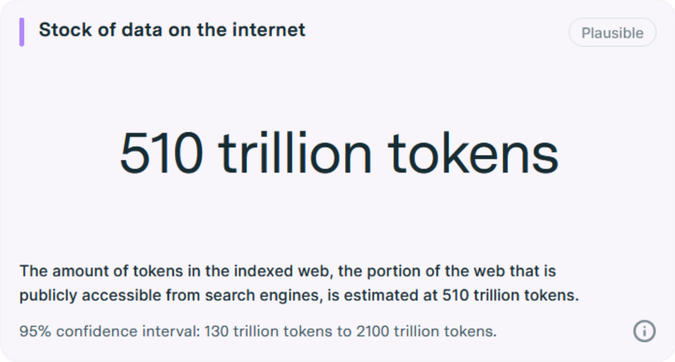
Will scaling continue to work?
We might be running out of internet



Largest training dataset used to train an LLM Uncertain

18 trillion tokens

Qwen2.5 models, including Qwen2.5-72B, were trained on 18 trillion tokens, making them the models with the largest publicly confirmed training datasets.



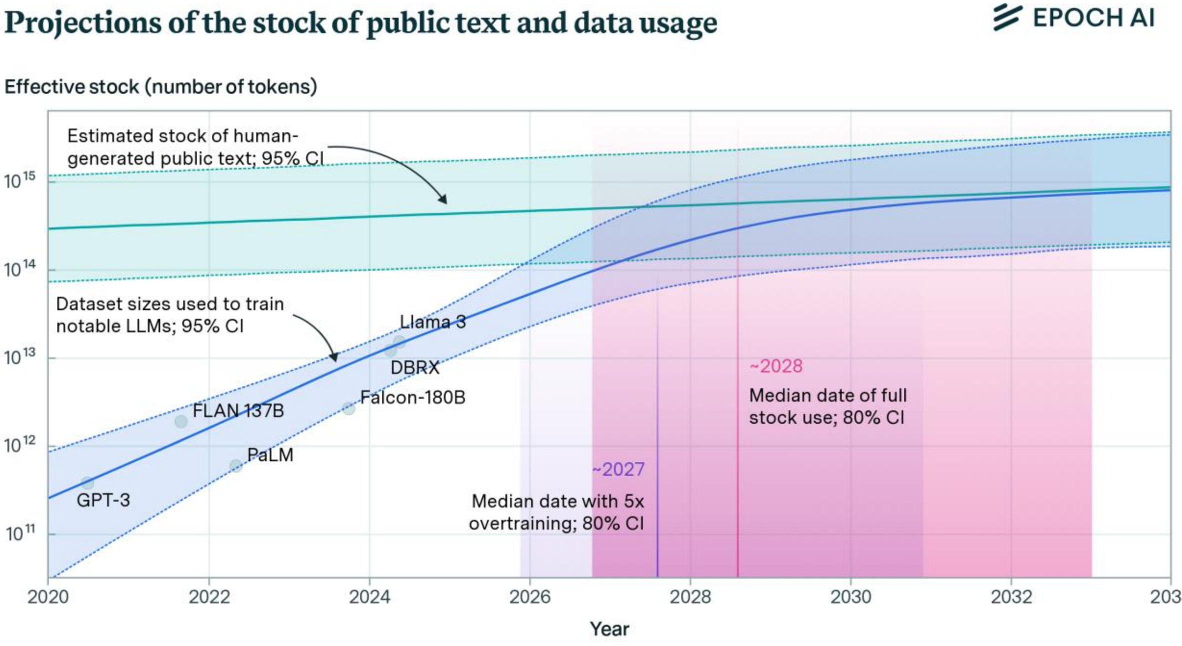
Stock of data on the internet Plausible

510 trillion tokens

The amount of tokens in the indexed web, the portion of the web that is publicly accessible from search engines, is estimated at 510 trillion tokens.

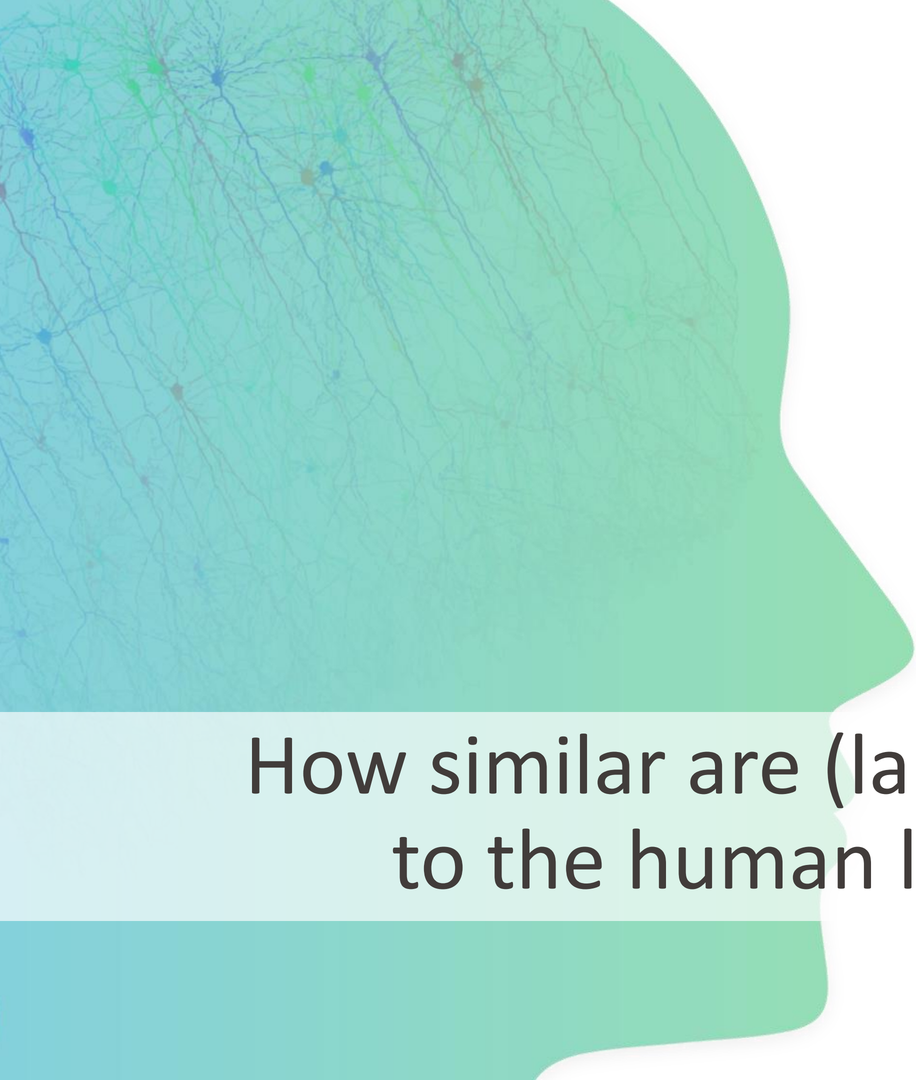
95% confidence interval: 130 trillion tokens to 2100 trillion tokens.

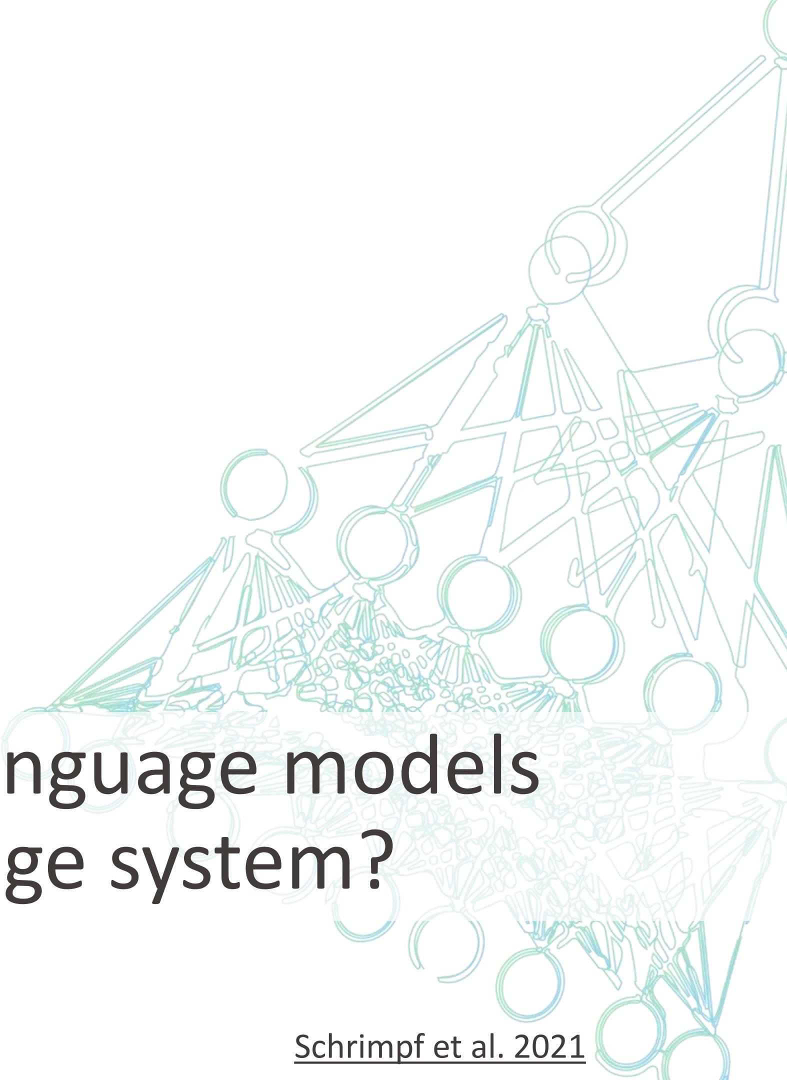
- https://epoch.ai/trends#data



Projections of the stock of public text and data usage

https://epoch.ai/blog/will-we-run-out-of-data-limits-of-llm-scaling-based-on-human-generated-data





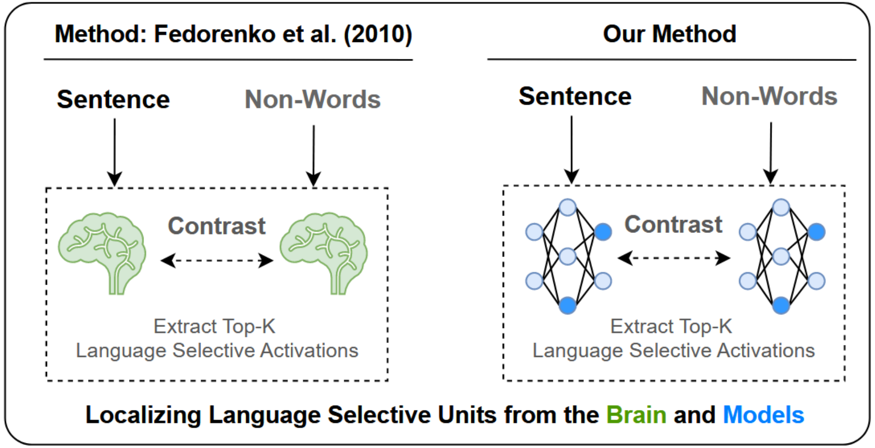
How similar are (large) language models to the human language system?

Schrimpf et al. 2021

A core language network in LLMs

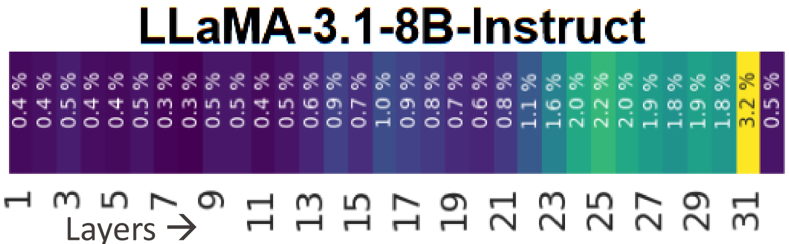
Sentence: THE DOG CHASED THE CAT ALL DAY LONG

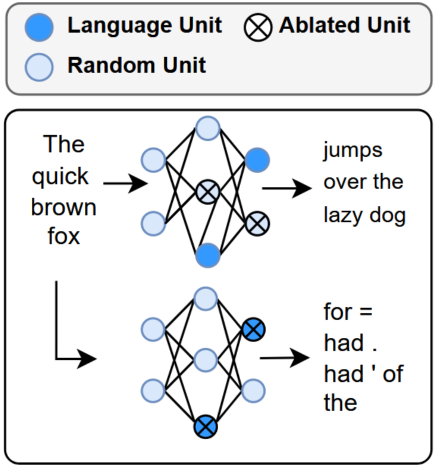
Non-Words: LUT REE UMLY LOND E WAM GOVING HOM

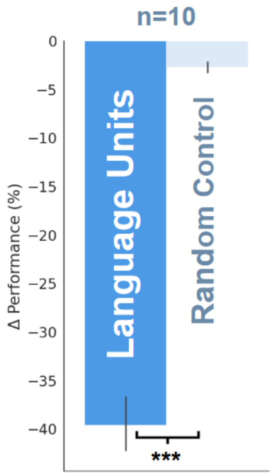


Method: Fedorenko et al. (2010)
Our Method
Sentence
Non-Words
Contrast
Extract Top-K
Language Selective Activations
Sentence
Non-Words
Contrast
Extract Top-K
Language Selective Activations

Localizing Language Selective Units from the Brain and Models







1% of units

Model	Ablate Language Units	Ablate Random Units
Gemma-2B	11 liquido _ sota(.)uggoon3	jumped over the lazy lamb.
Phi-3.5-Mini-Instruct	AME.AME and:ough.. MAR	jumps over the lazy dog.
Falcon-7b	SomeSReadWhenISearchSome	jumps over the lazy dog.
Mistral-7B-v0.3	foxfool foolfoolfoolfool	jumps over the lazy dog.
LLaMA-3.1-8B-Instruct	_ of_An_O_of_An_O_of	jumps over the lazy dog.

- Can functionally localize a core language system in LLMs
- Ablating even a small number of units leads to language deficits (~aphasia)
- How similar are model units to brain data?

AlKhamissi et al. 2025 NAACL Oral

Data target: human neural recordings

Pereira et al. 2018 fMRI

627 sentences x 13,517 voxels in 10 subjects

Beekeeping encourages the conservation of loca habitats. | It is in every beekeeper's interest …

Fedorenko et al. 2016 ECoG

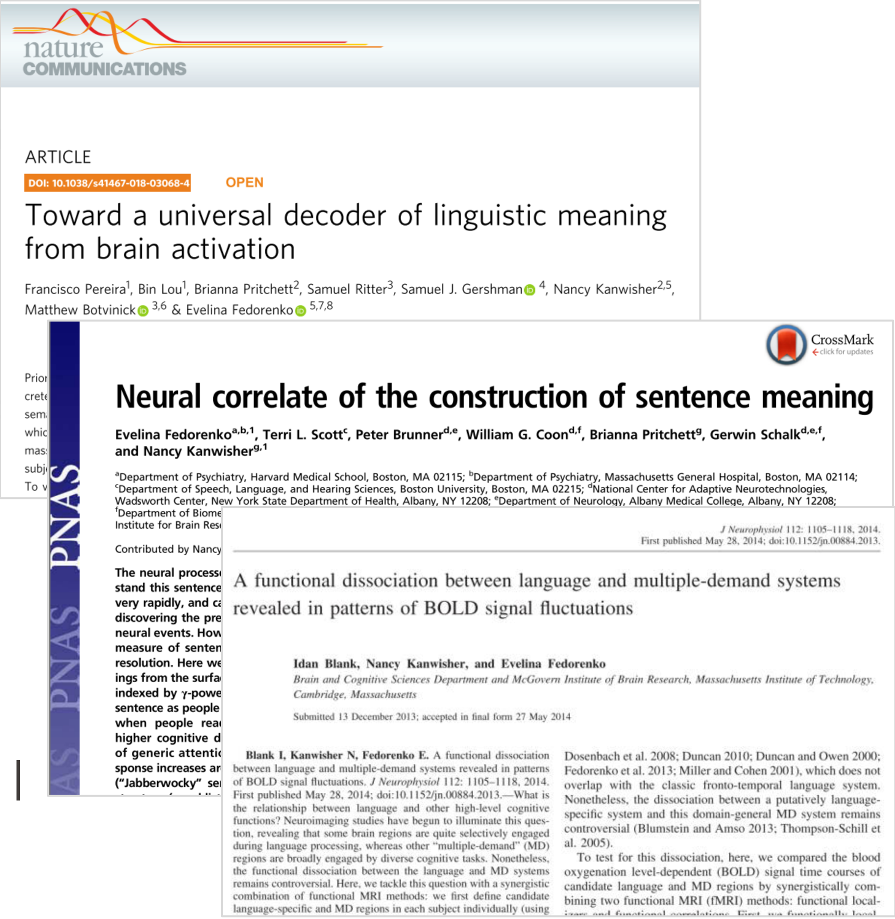
416 words x 97 electrodes in 5 subjects

ALEX | WAS | TIRED | SO | HE | TOOK | A | NAF

Blank et al. 2014 fMRI

1,317 story fragments x 60 fROIs in 5 subjects

If you were to journey to the | North of England, you would come to a valley | that is surrounded by moors as high as | mountains. It is in this | valley where you would find the city of Bradford, | …



nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-03068-4 OPEN

Toward a universal decoder of linguistic meaning from brain activation

Francisco Pereira[1], Bin Lou[1], Brianna Pritchett[2], Samuel Ritter[3], Samuel J. Gershman[4], Nancy Kanwisher[2,5], Matthew Botvinick[3,6] & Evelina Fedorenko[5,7,8]

Prior
crete
sem
whic
mas
subj
To v

PNAS

CrossMark click for updates

Neural correlate of the construction of sentence meaning

Evelina Fedorenko[a,b,1], Terri L. Scott[c], Peter Brunner[d,e], William G. Coon[d,f], Brianna Pritchett[g], Gerwin Schalk[d,e,f], and Nancy Kanwisher[g,1]

[a]Department of Psychiatry, Harvard Medical School, Boston, MA 02115; [b]Department of Psychiatry, Massachusetts General Hospital, Boston, MA 02114; [c]Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215; [d]National Center for Adaptive Neurotechnologies, Wadsworth Center, New York State Department of Health, Albany, NY 12208; [e]Department of Neurology, Albany Medical College, Albany, NY 12208; [f]Department of Biome
Institute for Brain Res

Contributed by Nancy

The neural process
stand this sentence
very rapidly, and ca
discovering the pre
neural events. How
measure of senten
resolution. Here we
ings from the surfa
indexed by γ-powe
sentence as people
when people rea
higher cognitive d
of generic attenti
sponse increases ar
("Jabberwocky" se

J Neurophysiol 112: 1105–1118, 2014.
First published May 28, 2014; doi:10.1152/jn.00884.2013.

A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations

Idan Blank, Nancy Kanwisher, and Evelina Fedorenko

Brain and Cognitive Sciences Department and McGovern Institute of Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

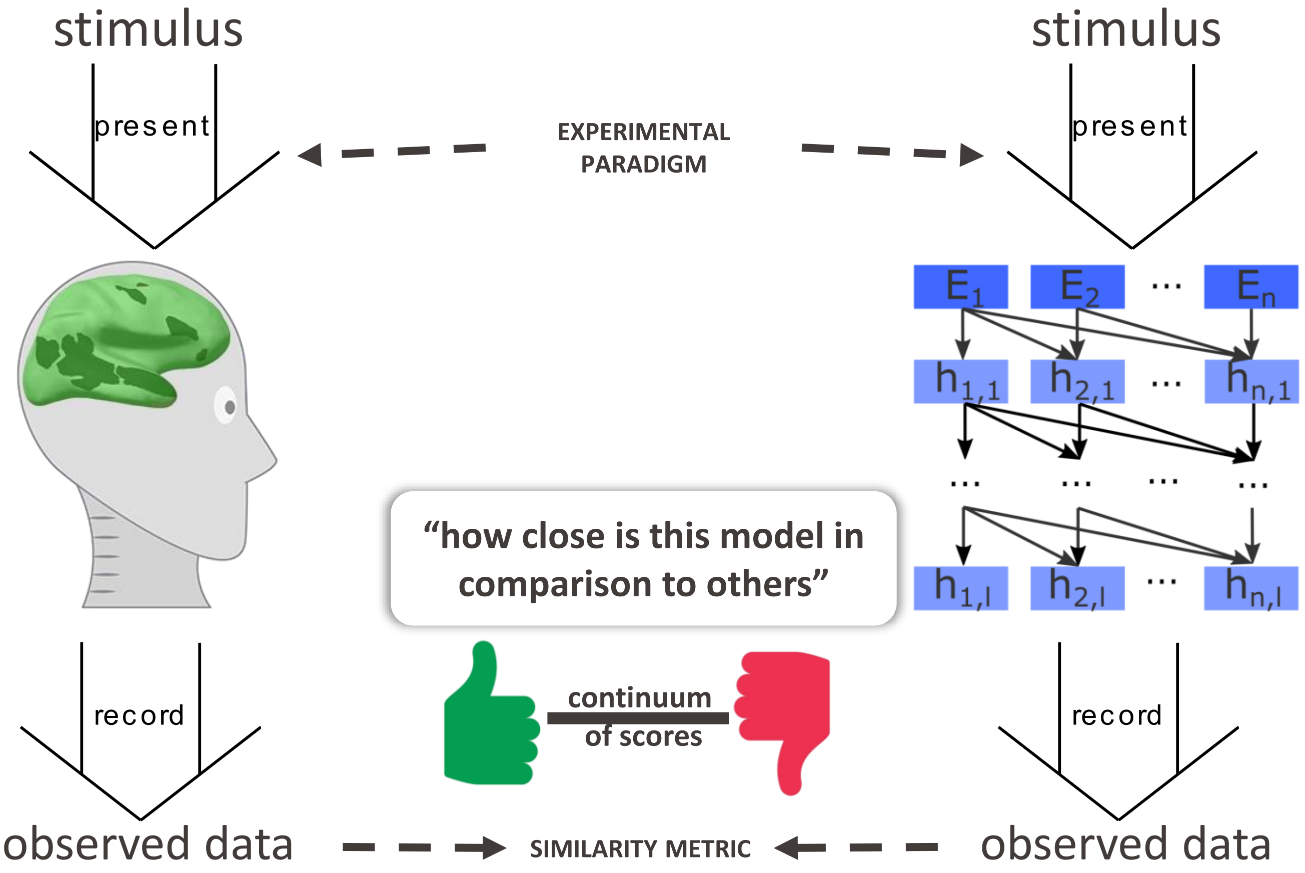
Submitted 13 December 2013; accepted in final form 27 May 2014

Blank I, Kanwisher N, Fedorenko E. A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J Neurophysiol* 112: 1105–1118, 2014. First published May 28, 2014; doi:10.1152/jn.00884.2013.—What is the relationship between language and other high-level cognitive functions? Neuroimaging studies have begun to illuminate this question, revealing that some brain regions are quite selectively engaged during language processing, whereas other "multiple-demand" (MD) regions are broadly engaged by diverse cognitive tasks. Nonetheless, the functional dissociation between the language and MD systems remains controversial. Here, we tackle this question with a synergistic combination of functional MRI methods: we first define candidate language-specific and MD regions in each subject individually (using

Dosenbach et al. 2008; Duncan 2010; Duncan and Owen 2000; Fedorenko et al. 2013; Miller and Cohen 2001), which does not overlap with the classic fronto-temporal language system. Nonetheless, the dissociation between a putatively language-specific system and this domain-general MD system remains controversial (Blumstein and Amso 2013; Thompson-Schill et al. 2005).

To test for this dissociation, here, we compared the blood oxygenation level-dependent (BOLD) signal time courses of candidate language and MD regions by synergistically combining two functional MRI (fMRI) methods: functional local-

Quantifying match-to-brain: Benchmarking



stimulus

present

EXPERIMENTAL PARADIGM

stimulus

present

“how close is this model in comparison to others”

continuum of scores

record

record

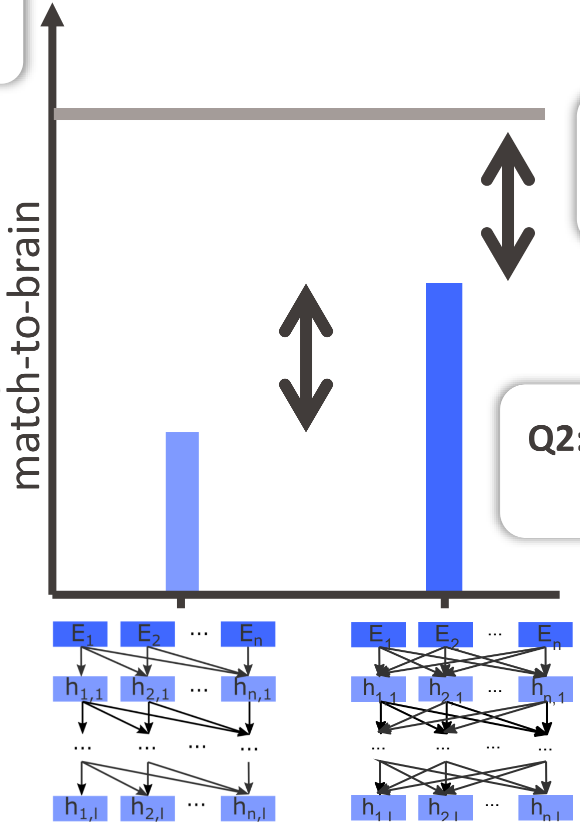
observed data

SIMILARITY METRIC

observed data

Quantifying match-to-brain: Benchmarking

We only care about best-matching model (for now)

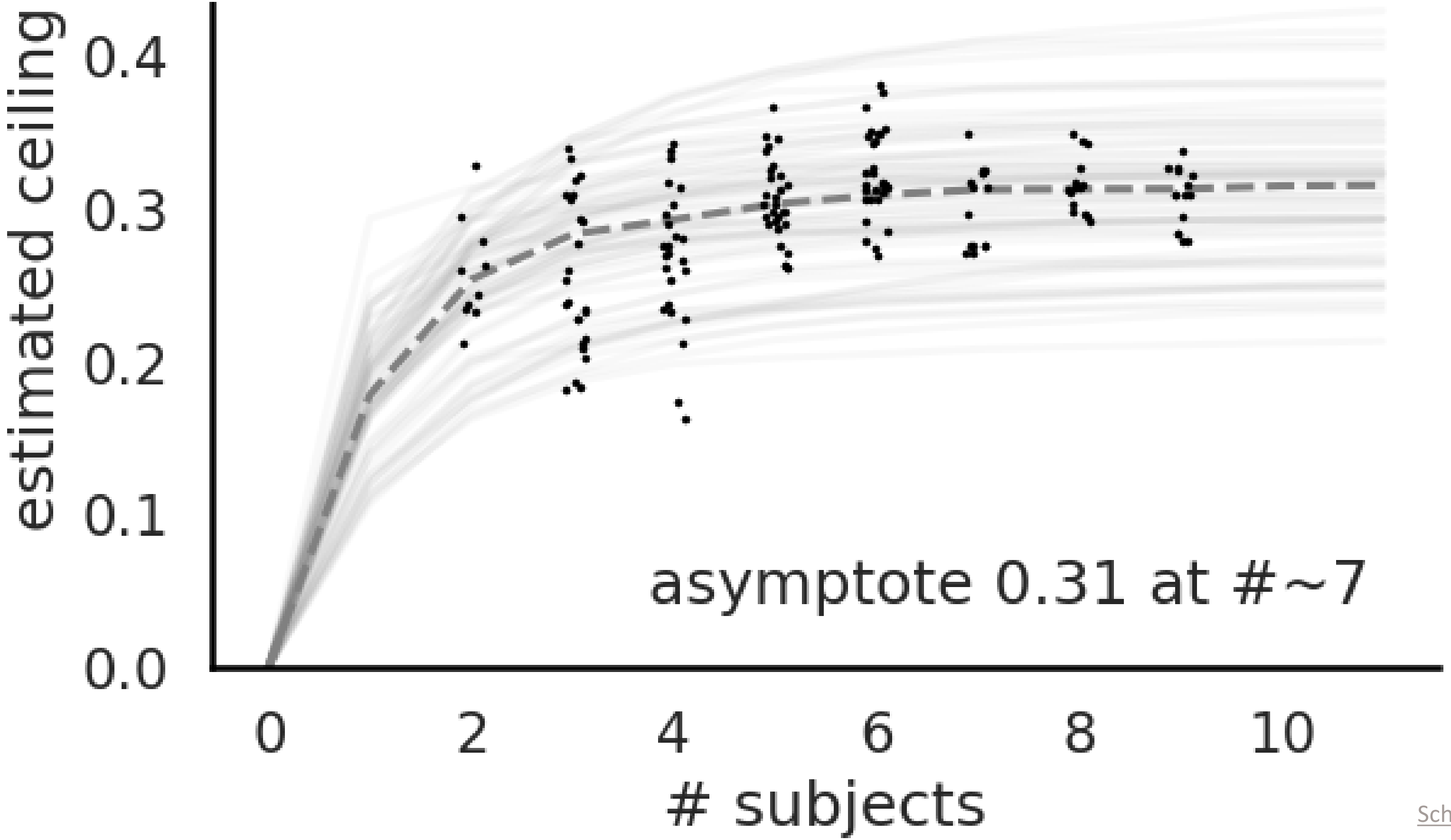

match-to-brain
E_1 E_2 ... E_n
$h_{1,1}$ $h_{2,1}$ ... $h_{n,1}$
... ...
$h_{1,l}$ $h_{2,l}$ ... $h_{n,l}$
E_1 E_2 ... E_n
$h_{1,1}$ $h_{2,1}$ ... $h_{n,1}$
... ...
$h_{1,l}$ $h_{2,l}$ ... $h_{n,l}$

Q1: how close are we?

Q2: which models are best?

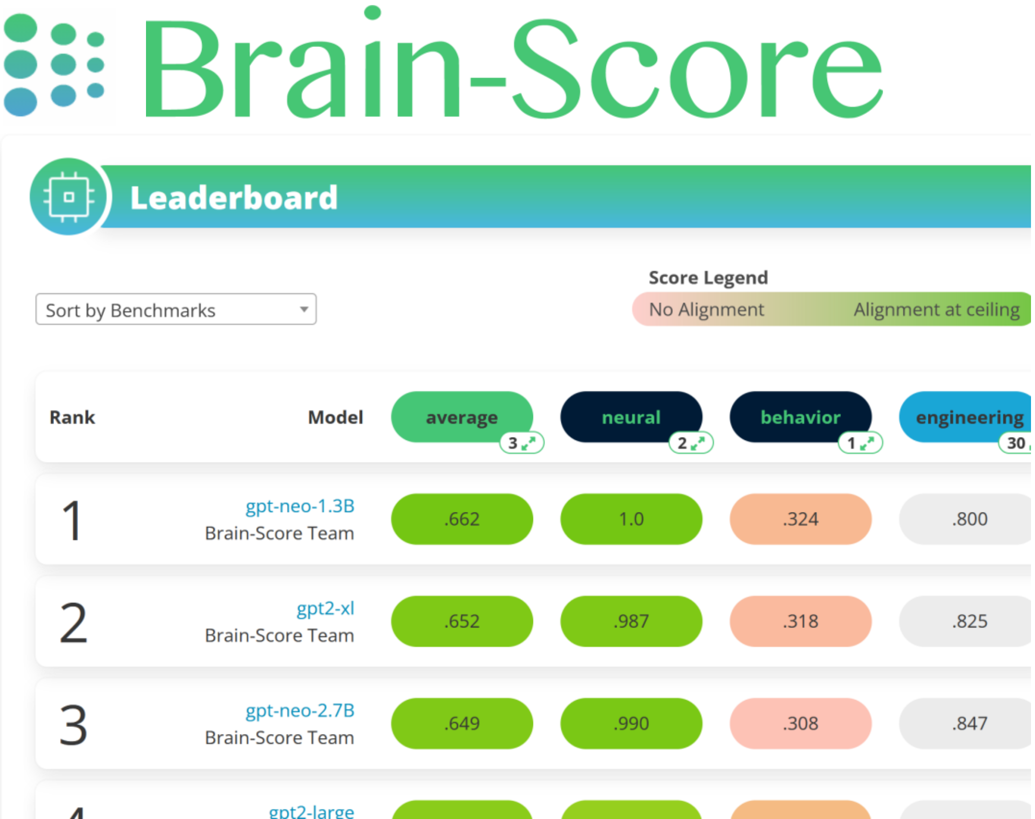
How close are we, how reliable is the data?

“internal consistency” compute similarity of a pool of subjects to a held-out subject



Schrimpf et al. 2021

Open science: accessible brain and behavior benchmarks to evaluate computational models



Brain-Score

Leaderboard

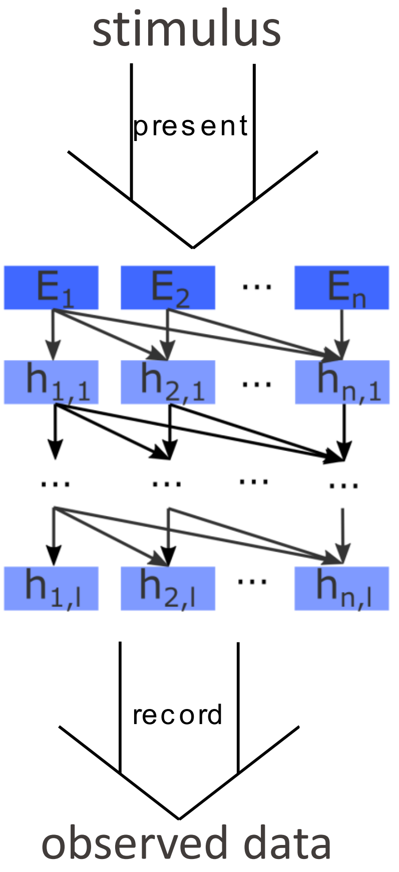
Sort by Benchmarks

Score Legend

No Alignment — Alignment at ceiling

Rank	Model	average 3	neural 2	behavior 1	engineering 30
1	gpt-neo-1.3B Brain-Score Team	.662	1.0	.324	.800
2	gpt2-xl Brain-Score Team	.652	.987	.318	.825
3	gpt-neo-2.7B Brain-Score Team	.649	.990	.308	.847
4	gpt2-large				

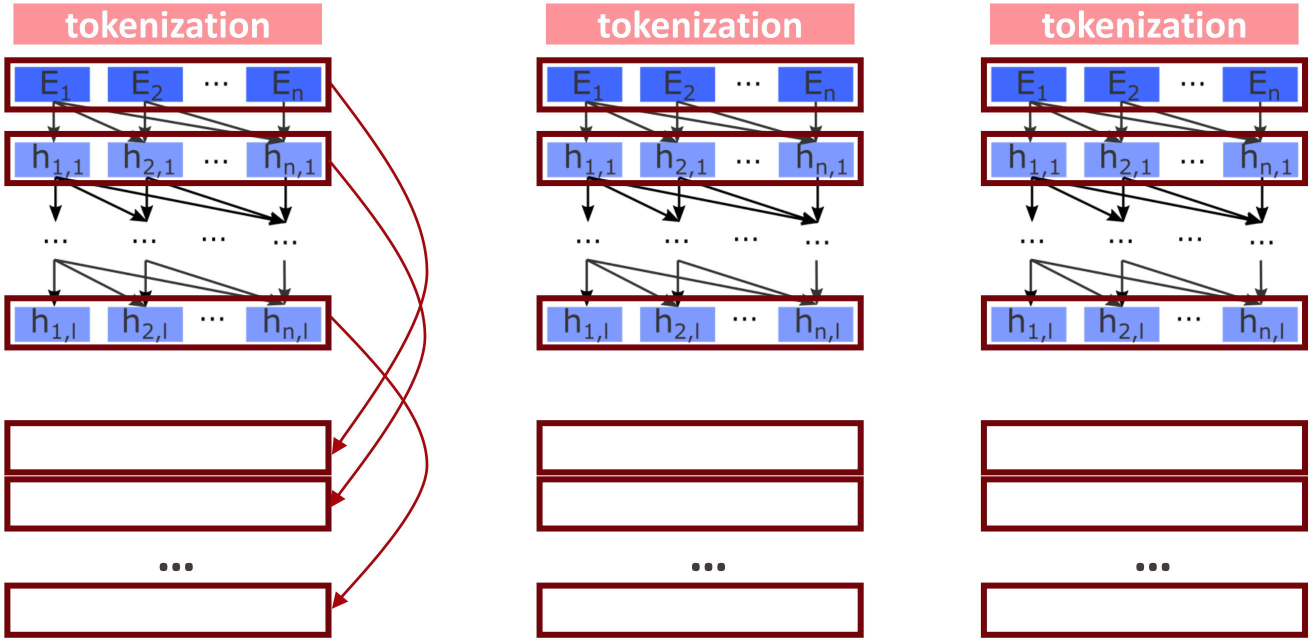
Treating models as experimental subjects



Beekeeping encourages the conservation of local habitats.

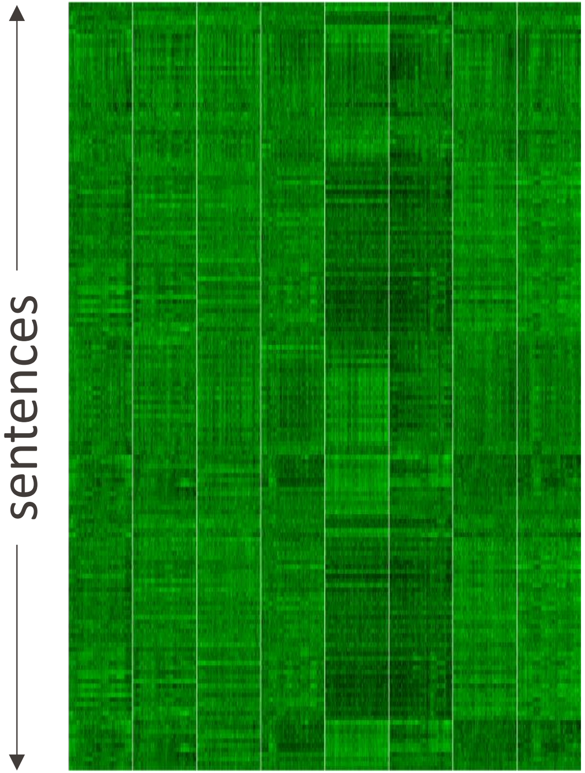
It is in every beekeeper's interest to conserve local plants that produce pollen.

…

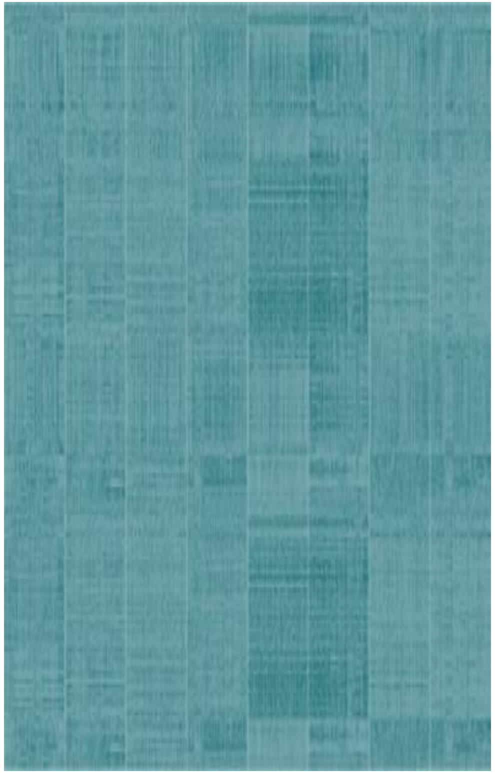


Neural benchmarks

Brain recordings



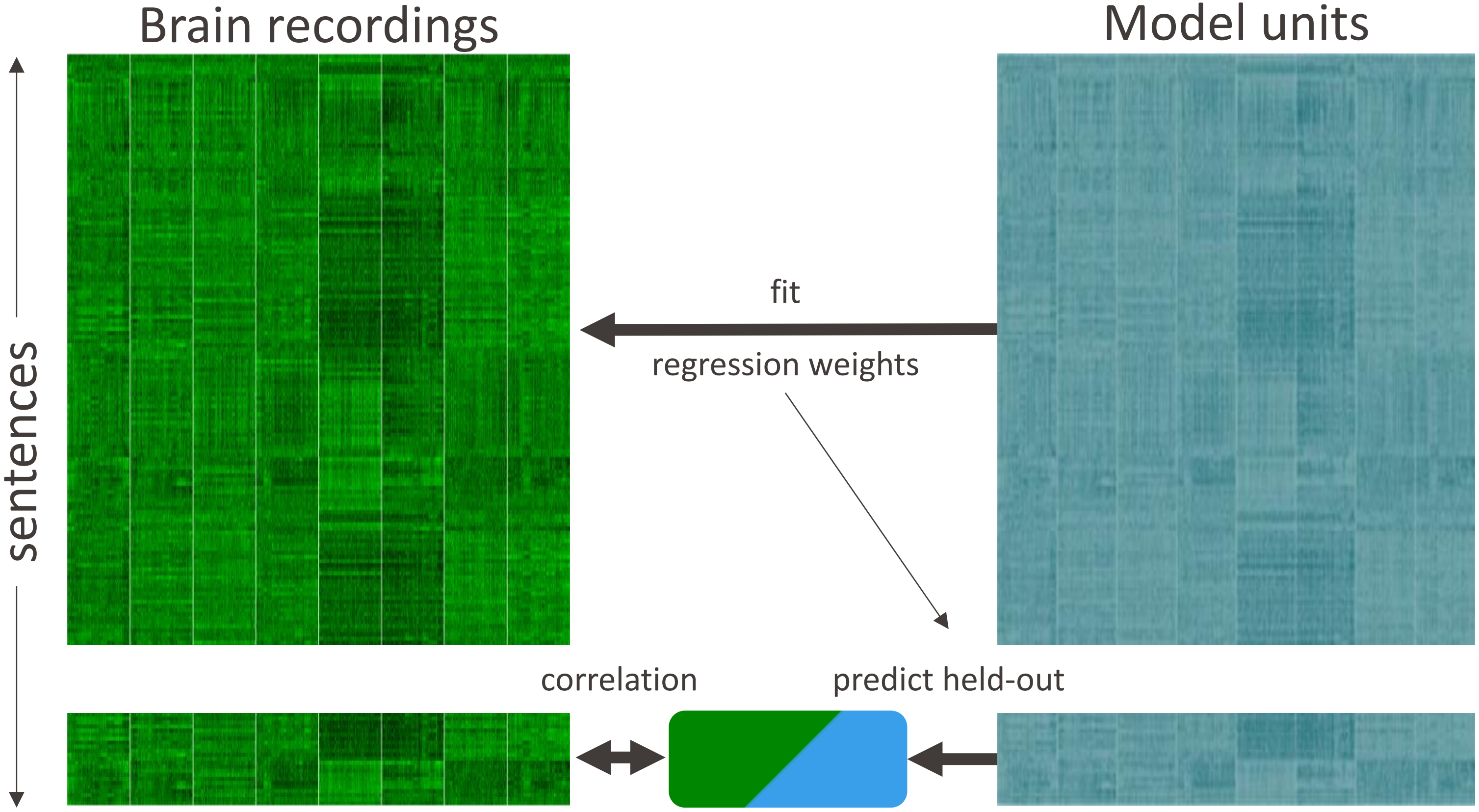
Model units



Neural benchmarks

similarity metric

neural predictivity

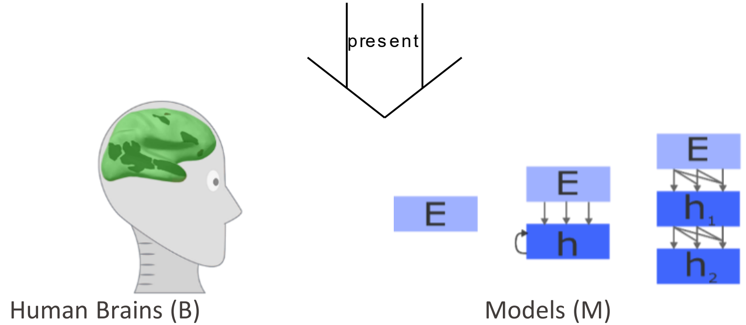


Yamins, Hong*, et al. (PNAS 2014)* *Schrimpf*, Kubilius*, et al. (bioRxiv 2018)*

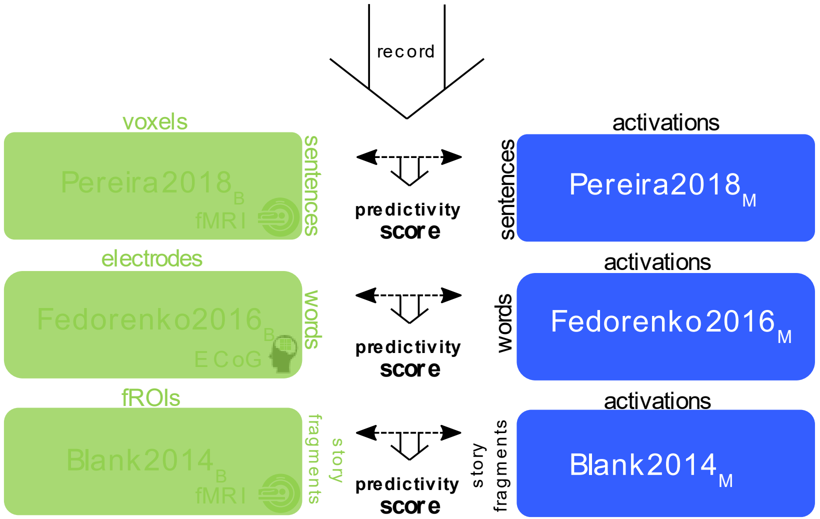
Stimuli

Pereira2018	"Beekeeping encourages the conservation of local habitats. It is in every beekeeper's interest..."
Fedorenko2016	"Alex was tired so he took a nap."
Blank2014	"If you were to journey to the North of England, you would come to a valley that is surrounded by moors as high as mountains. It is in this valley where you…"

Experimental Participants

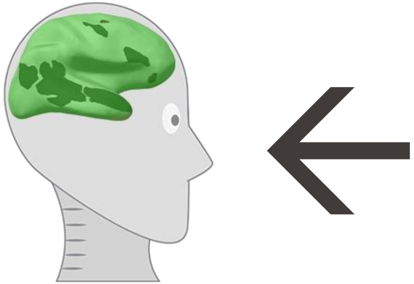


Comparative Measurements

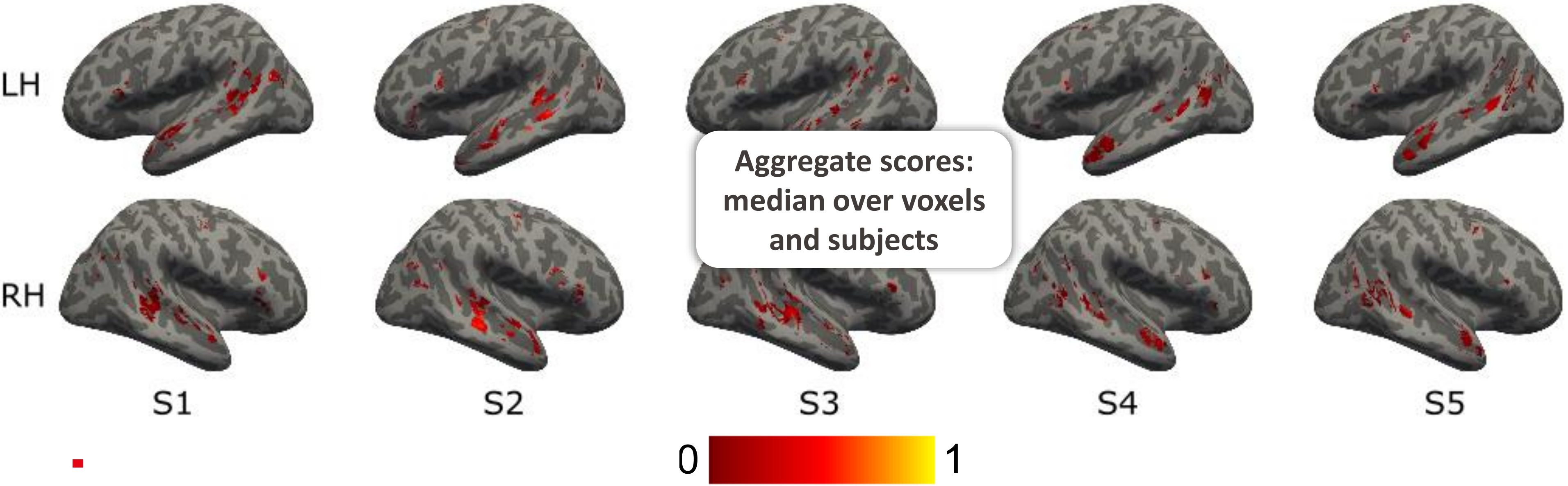


We want one model to predict *all* data

GloVe voxel-wise predictivity scores

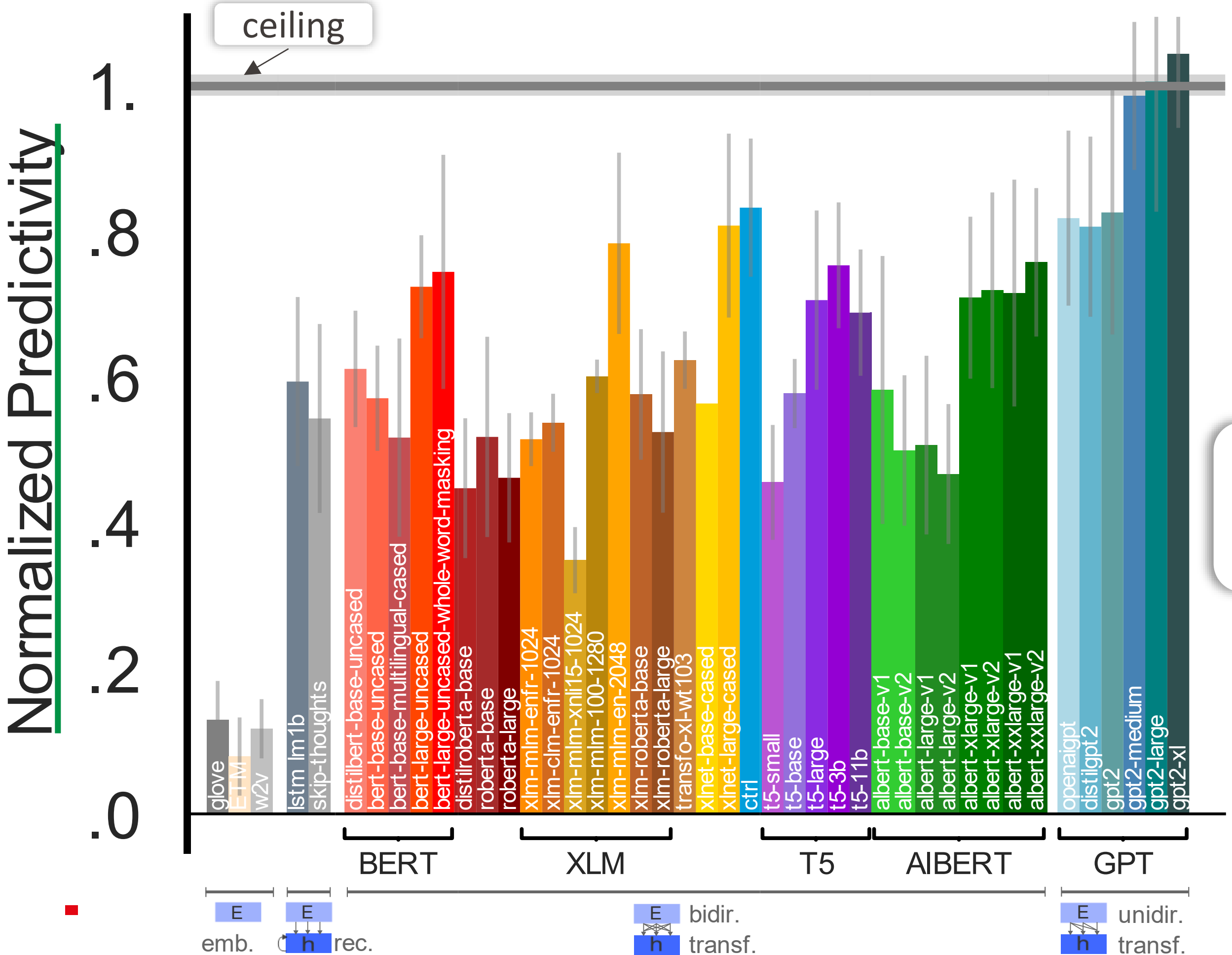


GloVe



Aggregate scores: median over voxels and subjects

Certain language models predict human language recordings

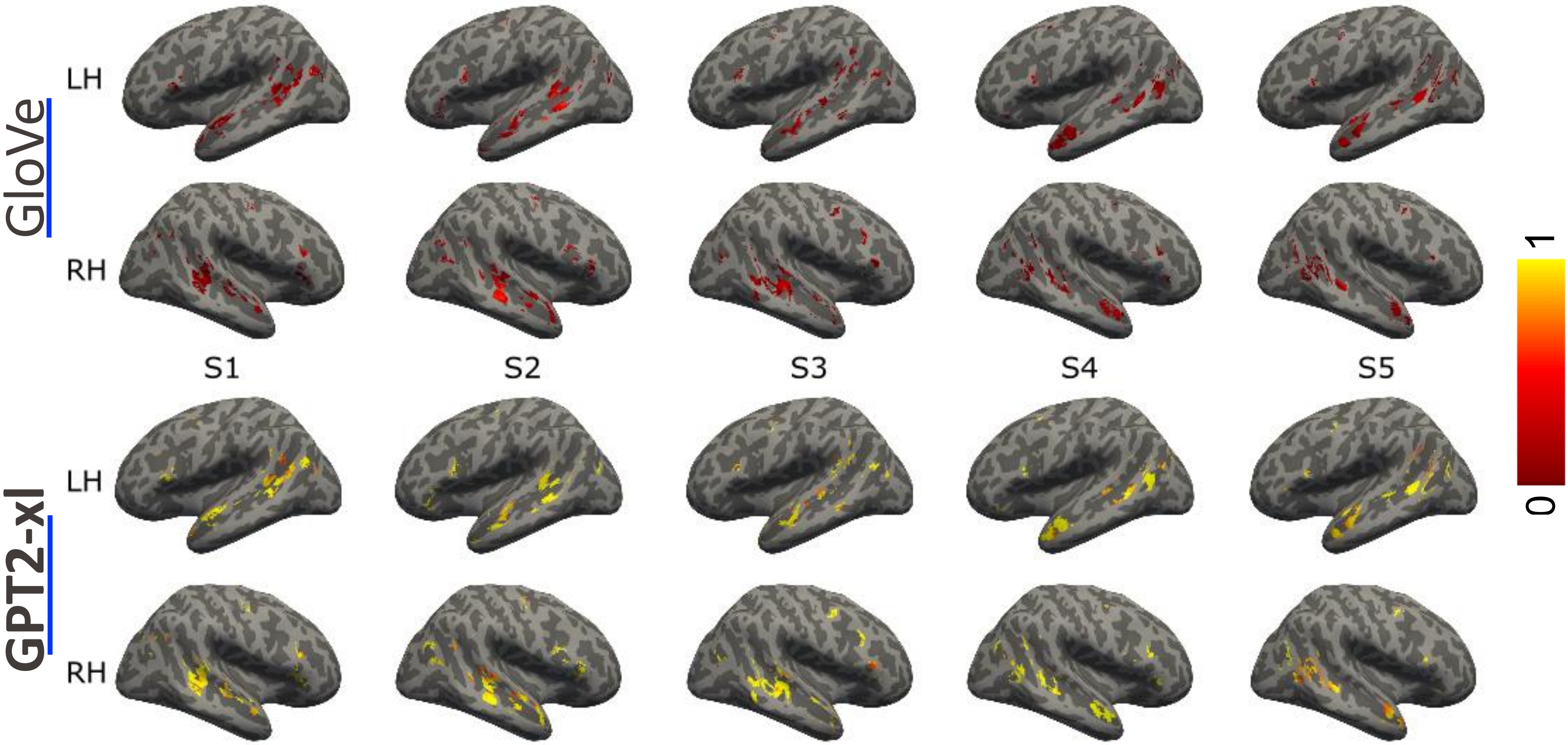


gpt2-xl hits our estimated ceiling for this benchmark!

Small differences can lead to very different brain predictivities, warranting a full survey

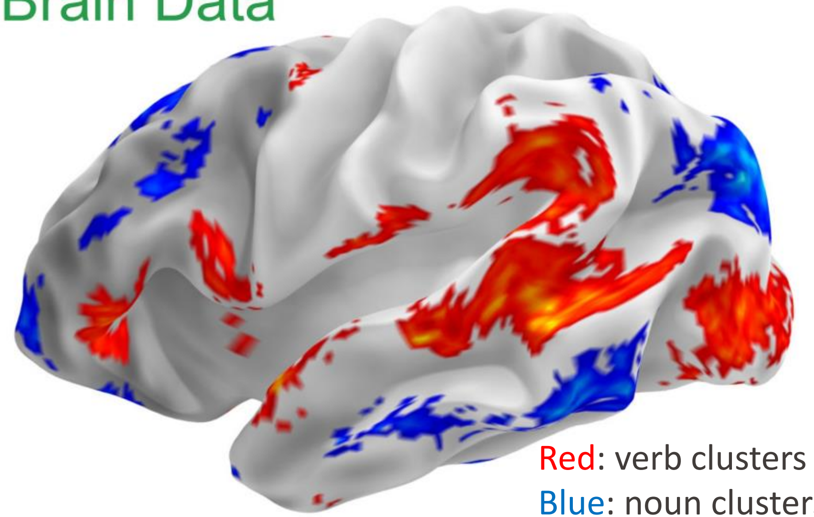
Jain & Huth 2018
Gauthier & Ivanova 2018
Jat et al. 2019
Toneva & Wehbe 2019
Gauthier & Levy 2020
Wang et al. 2020

GPT2-xl accurately predicts a large portion of voxels



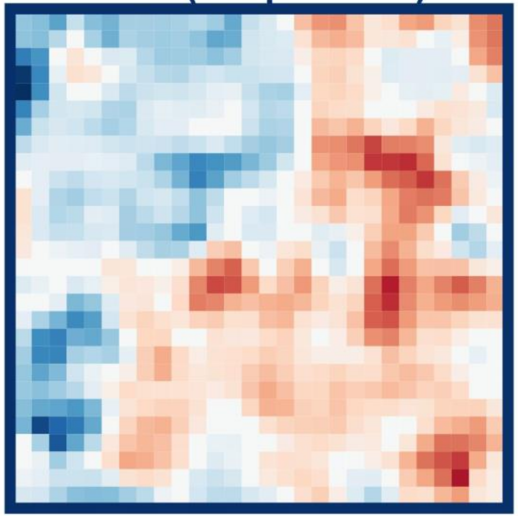
Topographic models of language

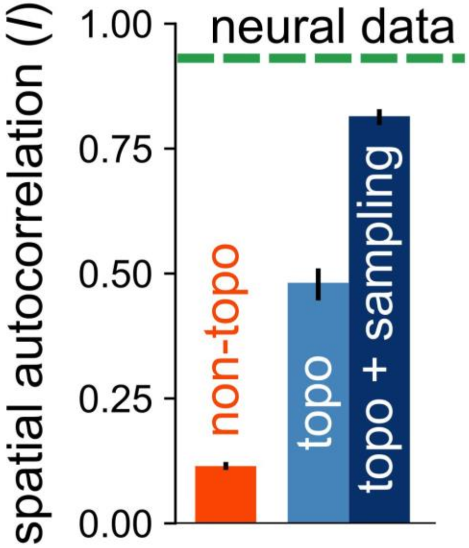
Brain Data



Red: verb clusters
Blue: noun clusters

Model (TopoLM)





- Beyond a functional correspondence, recent models such as TopoLM capture the spatio-functional organization in the human brain
- TopoLM is trained with a task + spatial loss – exactly like the models in vision!

Rathi & Mehrer et al. 2025 (ICLR Oral)

Take-home messages

- Language is not thought. Evidence from aphasia and neuroimaging studies, as well as recent computational evidence in LLMs.
- High-quality, large-scale data for human language is hard (but very important).
- Key model classes in NLP: embedding, recurrent, and transformer models (attention mechanism).
- Scaling laws predict larger models trained on more data will continue to improve performance.
- Particular models such as GPT are similar to
  brain recordings from the human language system.