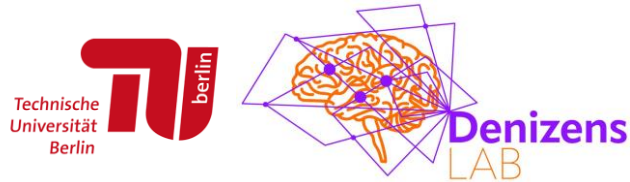
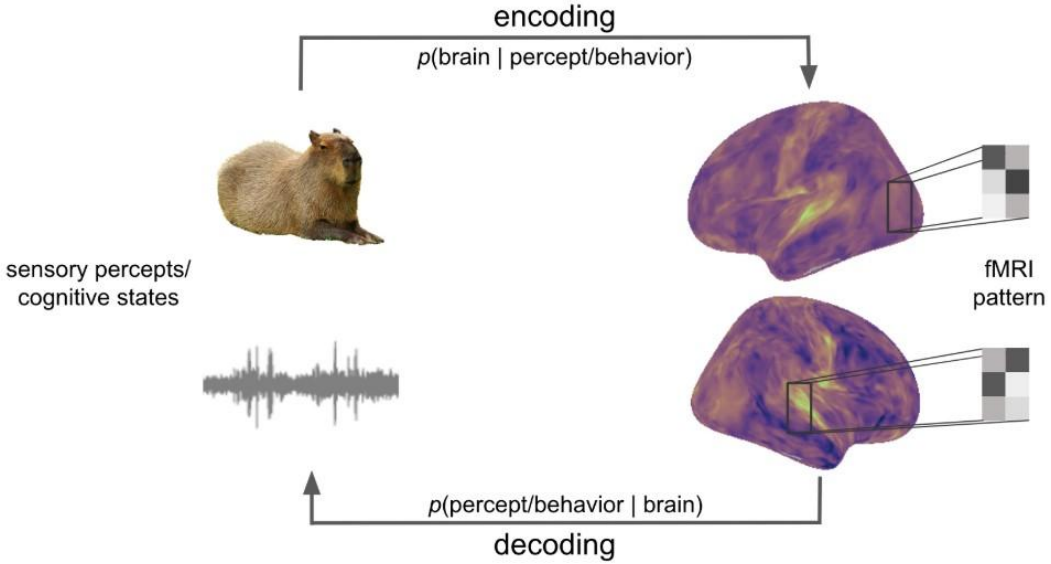
# Linguistic
# Brain Decoding



# What is Brain Decoding?

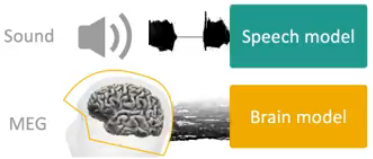



- Can we reconstruct the stimulus, given the brain response?
- Can you read the mind from fMRI responses!
- Or at least tell what the person saw?

https://main-educational.github.io/brain_encoding_decoding/intro.html

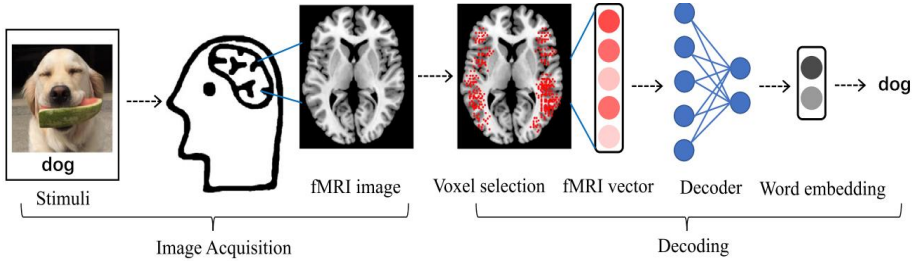
# What is Linguistic Brain Decoding?

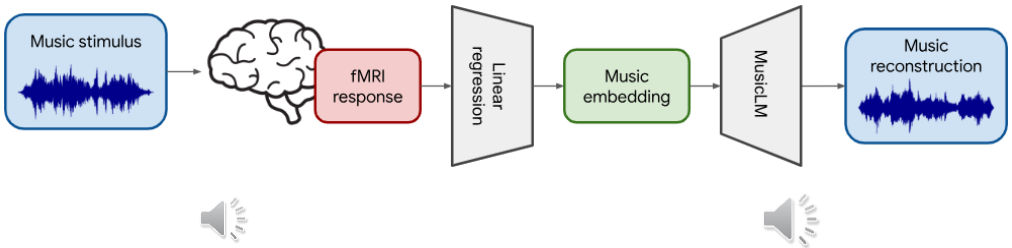
Linguistic Decoding




**Decoding speech from non-invasive brain recordings**

Défossez, Caucheteux, Rapin, Kabeli & King (2022)
arxiv.org/pdf/2208.12266




Alexandre Défossez, Charlotte Caucheteux, Jérémy Rapin, Ori Kabeli & Jean-Rémi King. "Decoding speech perception from non-invasive brain recordings" Nature Machine Intelligence 2023.

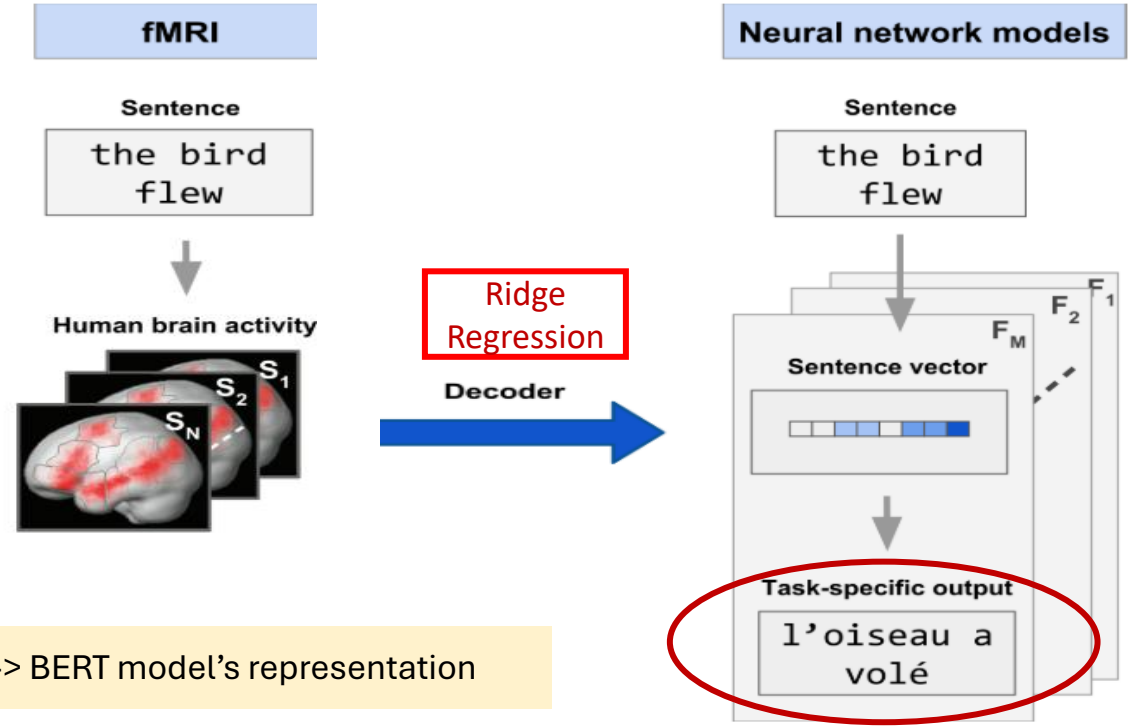
Timo I. Denk, Yu Takagi, Takuya Matsuyama, Andrea Agostinelli, Tomoya Nakai, Christian Frank, Shinji Nishimoto. "Brain2Music: Reconstructing Music from Human Brain Activity" Arxiv 2024.

# Linguistic Brain decoding outline

- **Sentence Reconstruction: Linking artificial and human neural representations of language**
- Story Reconstruction:
  - Semantic reconstruction of continuous language from non-invasive brain recordings
  - Generative Language reconstruction from non-invasive brain recordings
  - A Subject-Agnostic and Versatile Model for fMRI-to-Text Decoding

# Linguistic **brain decoding:** sentence reconstruction from fMRI

- Stimuli: Periera dataset, reading full sentences
- Stimulus representation: BERT language model
- Brain recording & modality: fMRI, reading
- Core idea: predict model's sentence embeddings from fMRI

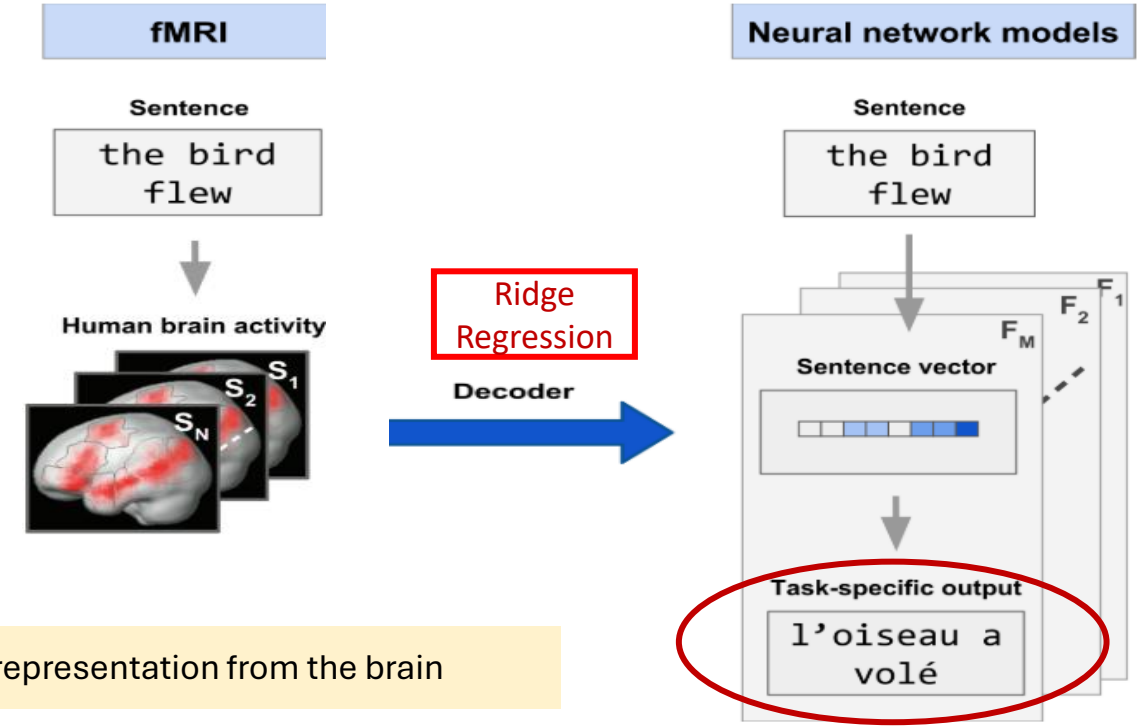



sentence -> human fMRI and sentence -> BERT model's representation

Gauthier et al. 2019 EMNLP

# Sentence representations from language models

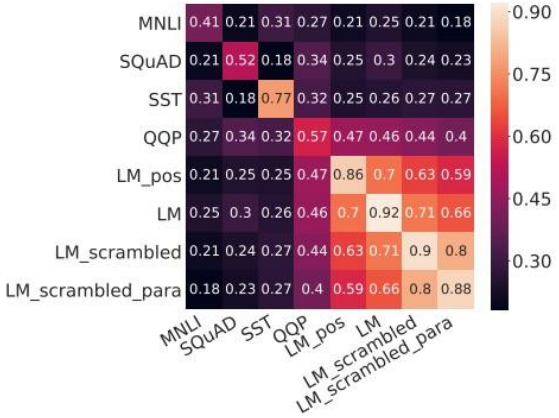
- Pretrained BERT
- Fine-tune the same BERT: QQP, SQuAD 2.0, MNLI, and SST-2
- LM-scrambled: words are shuffled within the sentence
- LM-scrambled-para: words are shuffled within the paragraph
- LM-pos: model predicts only the part of speech of a masked word
- Baseline: GloVe

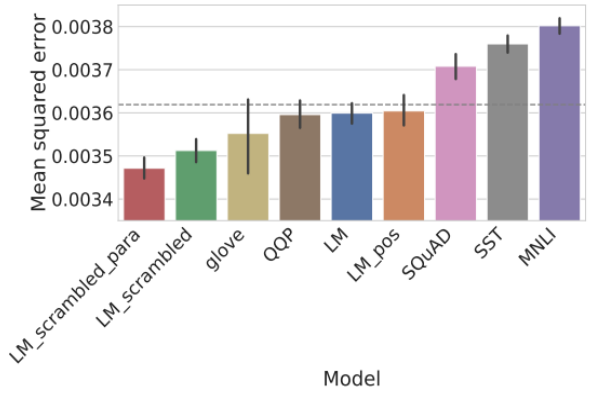


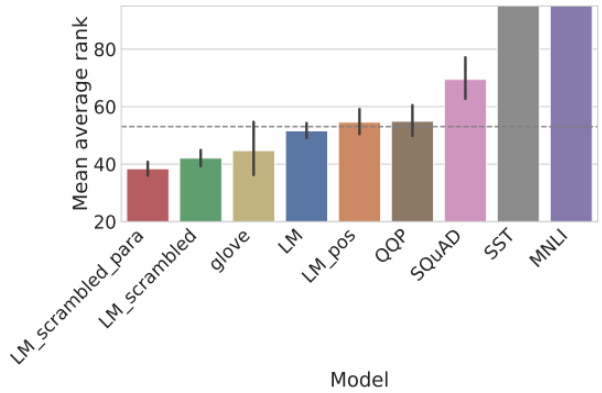

Predict a model's internal sentence representation from the brain

Gauthier et al. 2019 EMNLP

# Scrambled language models have shown better performance!!




Models optimized for LM- scrambled and LM-scrambled-para (syntax-light): the models which improve in brain decoding performance

Gauthier et al. 2019

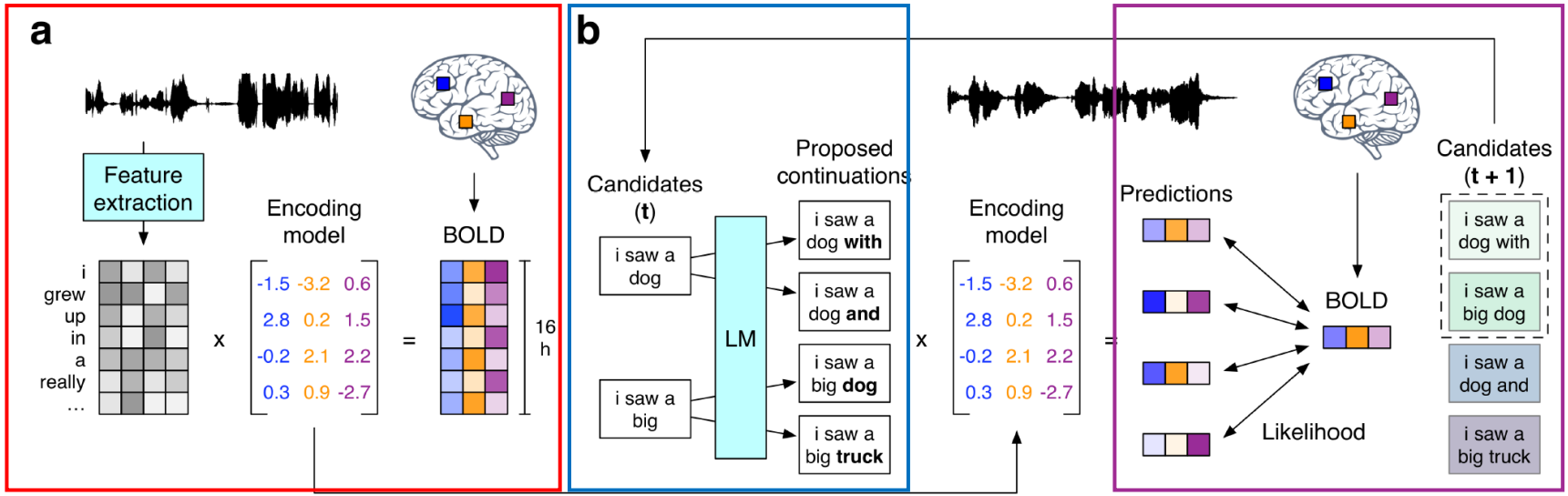
# Linguistic Brain decoding outline

- Sentence Reconstruction: Linking artificial and human neural representations of language
- Story Reconstruction:
  - **Semantic reconstruction of continuous language from non-invasive brain recordings**
  - Generative Language reconstruction from non-invasive brain recordings
  - A Subject-Agnostic and Versatile Model for fMRI-to-Text Decoding

# Continuous Language Decoder

- Stimuli: Moth-Radio-Hour
- Stimulus representation: GPT2 language model
- Brain recording & modality: fMRI, listening

- Brain decoding pipeline: Encoding model
- Word rate model: Takes brain activity and predicts the number of words per TR
- Language decoder: methods like beam search, nucleas sampling, and a Bayesian decoder




Jerry Tang, Amanda LeBel, Shailee Jain & Alexander G. Huth "Semantic reconstruction of continuous language from non-invasive brain recordings" Nature Neuroscience 2023.

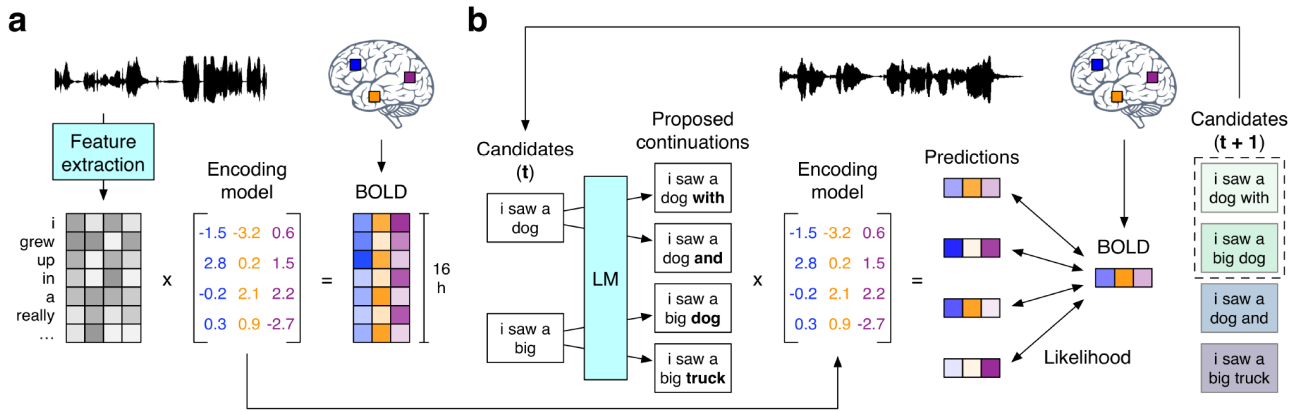
# Step-by-Step Decoding Example

Propose initial words:
- INIT = ["She", "He", "They"]
- Beam = ["She"]

Score Candidates:
- "She went" → likelihood = -10.2
- "She had" → likelihood = -13.5
- Top hypothesis: Beam = ["She went"]

Propose next word from LM:
- Context = “She”.
- Language model predicts: | Word | Prob | |——|——|
  | went | 0.6 | | had | 0.2 | | is | 0.1 |

**Repeat**




Jerry Tang, Amanda LeBel, Shailee Jain & Alexander G. Huth "Semantic reconstruction of continuous language from non-invasive brain recordings" Nature Neuroscience 2023.

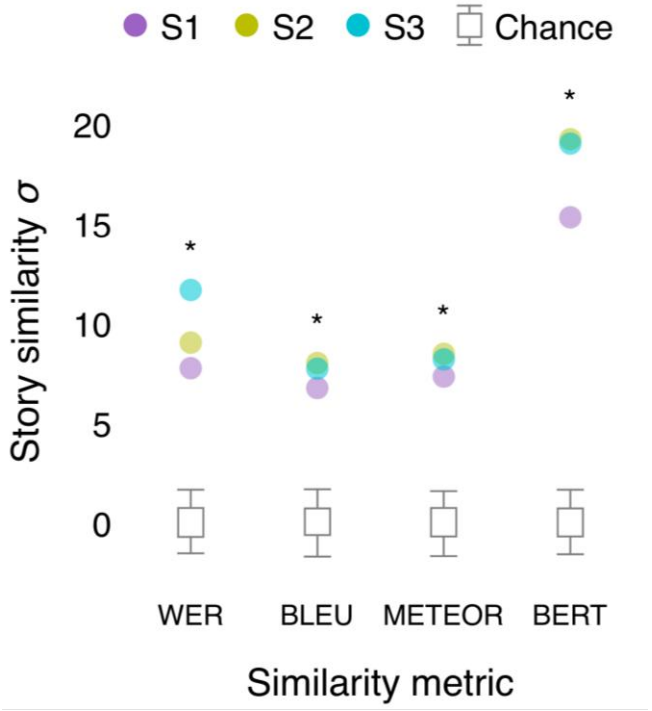
# Continuous Language Decoder

| Actual stimulus | Decoded stimulus | |
|---|---|---|
| *i got up from the air mattress and pressed my face against the glass of the bedroom window expecting to see eyes staring back at me but instead finding only darkness* | i just continued to walk up to the window and open the glass i stood on my toes and peered out i didn't see anything and looked up again i saw nothing | |
| *i didn't know whether to scream cry or run away instead i said leave me alone i don't need your help adam disappeared and i cleaned up alone crying* | started to scream and cry and then she just said i told you to leave me alone you can't hurt me i'm sorry and then he stormed off i thought he had left i started to cry | Exact |
| *that night i went upstairs to what had been our bedroom and not knowing what else to do i turned out the lights and lay down on the floor* | we got back to my dorm room i had no idea where my bed was i just assumed i would sleep on it but instead i lay down on the floor | Gist |
| *i don't have my driver's license yet and i just jumped out right when i needed to and she says well why don't you come back to my house and i'll give you a ride i say ok* | she is not ready she has not even started to learn to drive yet i had to push her out of the car i said we will take her home now and she agreed | Error |

Most parts of cortex contain some representations of language meaning

Jerry Tang, Amanda LeBel, Shailee Jain & Alexander G. Huth "Semantic reconstruction of continuous language from non-invasive brain recordings" Nature Neuroscience 2023.

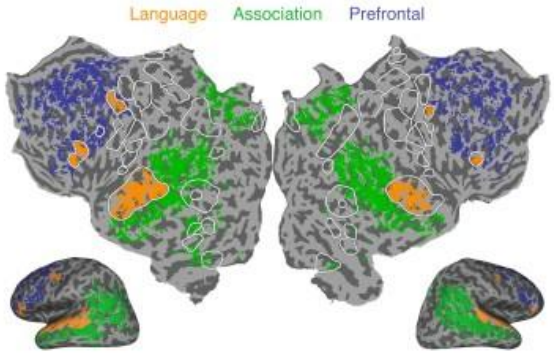
# Continuous Language Decoder

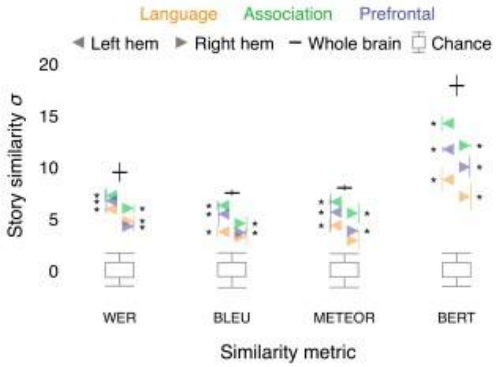



Which cortical networks represent language in sufficient detail to decode complete word sequences

Jerry Tang, Amanda LeBel, Shailee Jain & Alexander G. Huth "Semantic reconstruction of continuous language from non-invasive brain recordings" Nature Neuroscience 2023.

# Decoding is successful from association and prefrontal networks




| Actual stimulus | Left lang | Left assoc | Left PFC | Right lang | Right assoc | Right PFC |
|---|---|---|---|---|---|---|
| *i was like no i'm out of here this is great and i went and hid behind a cabana and he left* | they drove off they didn't even look back as i sat there thinking what the hell i should do | tell me to leave i said ok and ran out to the parking lot i was like wait is that a cop car | i told them to leave but they insisted and kept saying i can't stay so i got up to go | ran away and didn't look back at me and said you can go on without me i'm leaving now | in the driveway i told him to leave me alone and went inside i ran out into the cold | let me through i don't know where he is right now but i will get there soon enough |
| *i drew out this map for you and you're really you're like a mile and a half from home* | i try to keep track of how many miles it is before the number stops at the exact spot | i made a list of how many miles it is to get to where i lived and how far it was to the highway | write down a list of how many bedrooms and how long it takes to get to the closest one | let me just walk away from the case and start the process of making sure | i want to know how much it costs to drive from here to there to see how far it is | look up the address on a map and figure out how far it is to the apartment |

- Language model (LM) pretrained and fine-tuned on narrative stories only
- Decoder uses logic to combine LM + encoding model

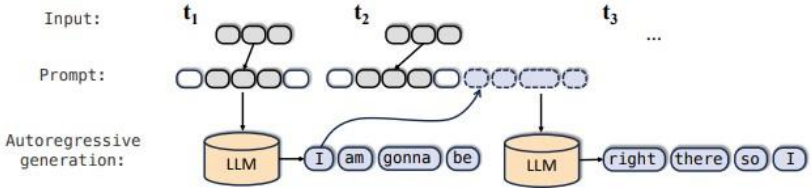
Jerry Tang, Amanda LeBel, Shailee Jain & Alexander G. Huth "Semantic reconstruction of continuous language from non-invasive brain recordings" Nature Neuroscience 2023.

# Linguistic Brain decoding outline

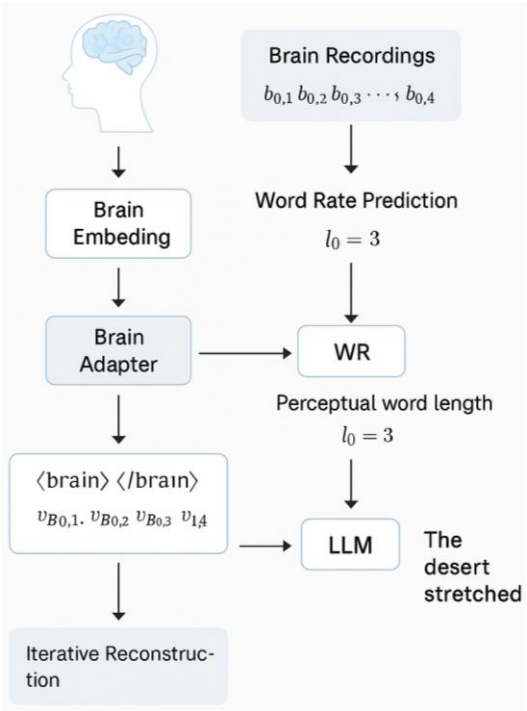
- Sentence Reconstruction: Linking artificial and human neural representations of language
- Story Reconstruction:
  - Semantic reconstruction of continuous language from non-invasive brain recordings
  - **Generative Language reconstruction from non-invasive brain recordings**
  - A Subject-Agnostic and Versatile Model for fMRI-to-Text Decoding

# Generative language reconstruction from brain activity

- Stimuli: Moth-Radio-Hour, Pereira, and Narratives
- Stimulus representation: GPT2, LLaMA-2
- Brain recording & modality: fMRI, listening, reading





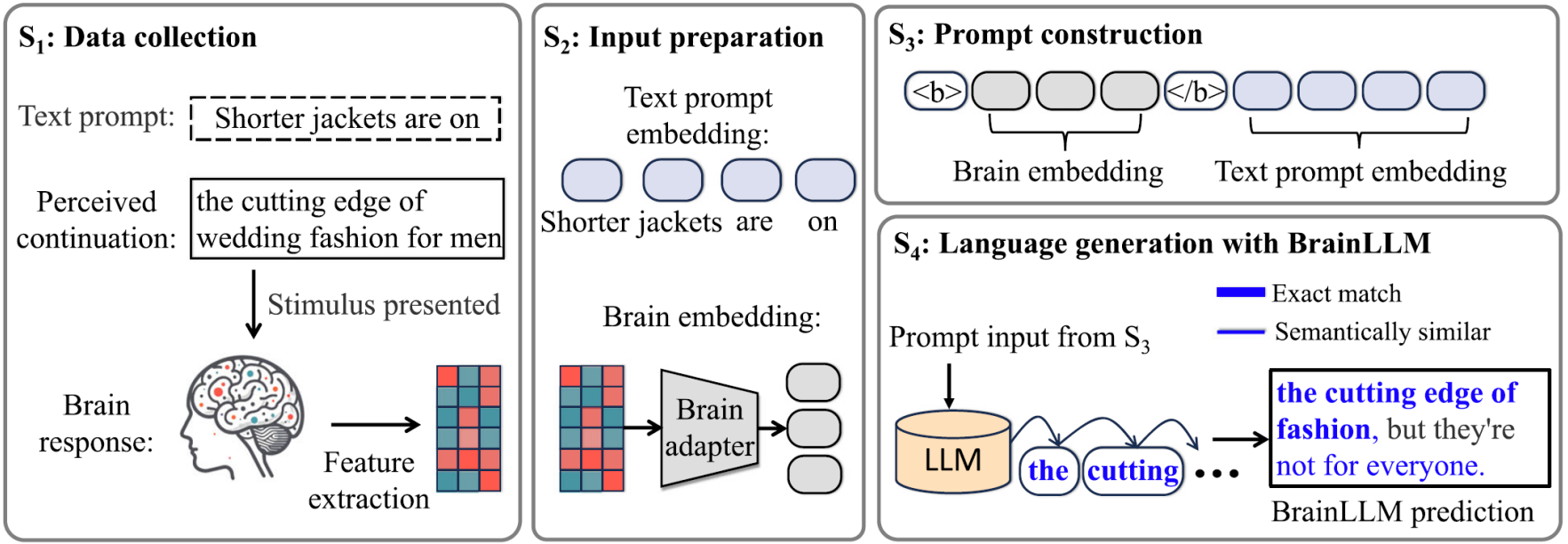



BrainLLM can be used in an auto-regressive manner to **reconstruct a 10 min-long language stimulus**!

Ye et al. (2025). Generative language reconstruction from brain recordings. Commun Biol. https://doi.org/10.1038/s42003-025-07731-7

# Generative Language Decoder

Four main Stages in Language Generation with brain recordings.




Ye et al. (2025). Generative language reconstruction from brain recordings. Commun Biol. https://doi.org/10.1038/s42003-025-07731-7

# Generative Language Decoder

- Examples of language generation with BrainLLM and its controls (PerBrainLLM).
- Text in blue and **bold** indicates that the generated content and the ground truth (perceived continuation) are manually annotated as semantically similar and an exact match, respectively.

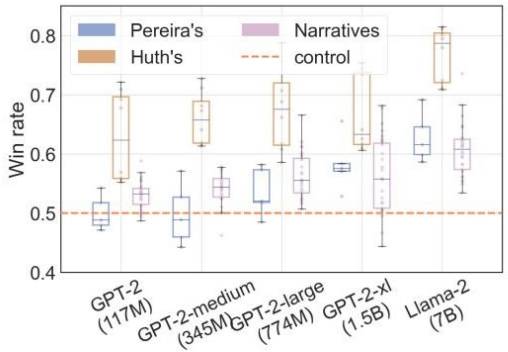
| Text prompt | Continuation | BrainLLM prediction | Control prediction |
|---|---|---|---|
| Shorter jackets are on | the cutting edge of wedding fashion for men | **the cutting edge of fashion**, but they're not for everyone | their way out **of** style, but they're still popular. |
| A wall is a | solid structure that defines and sometimes protects an area | **structure that defines and sometimes protects an area** | vertical **structure** made of stone, brick or concrete |
| I'm just stand--ing there like | the proverbial deer in headlights | a **deer in the headlights** | an idiot |
| she was like petite I could have | folded her up and put her my pocket | picked **her up** with one hand | driven her to work every day |

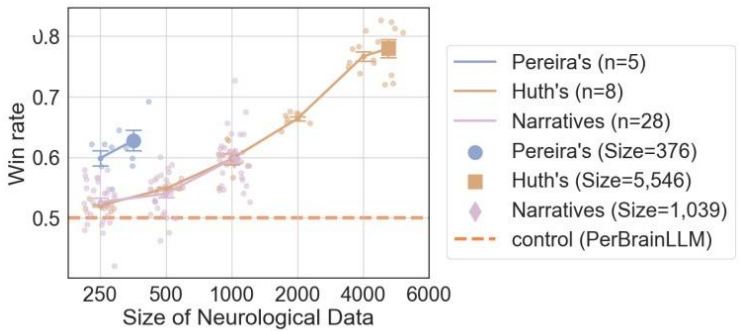
The BrainLLM's output is more aligned with the semantic text content perceived by human participants

Ye et al. (2025). Generative language reconstruction from brain recordings. Commun Biol. https://doi.org/10.1038/s42003-025-07731-7

# Scaling LLMs exhibit positive correlation, and size of brain data

| Dataset | Model | BLEU-1(↑) | ROUGE-1(↑) | ROUGE-L(↑) | WER(↓) |
|---|---|---|---|---|---|
| Huth's | PerBrainLLM | 0.1668* | 0.1536* | 0.1474* | 0.9109* |
| | BrainLLM | **0.1899** | **0.1780** | **0.1709** | **0.8916** |
| Pereira's | PerBrainLLM | 0.3269* | 0.2815* | 0.2751* | 0.7783* |
| | BrainLLM | **0.3432** | **0.2987** | **0.2878** | **0.7576** |
| Narratives | PerBrainLLM | 0.1269* | 0.1211* | 0.1105* | 0.9311* |
| | BrainLLM | **0.1375** | **0.1301** | **0.1209** | **0.9239** |




Language can be directly generated with brain recordings as input, rather than through selection from pre-constructed language candidates

Ye et al. (2025). Generative language reconstruction from brain recordings. Commun Biol. https://doi.org/10.1038/s42003-025-07731-7

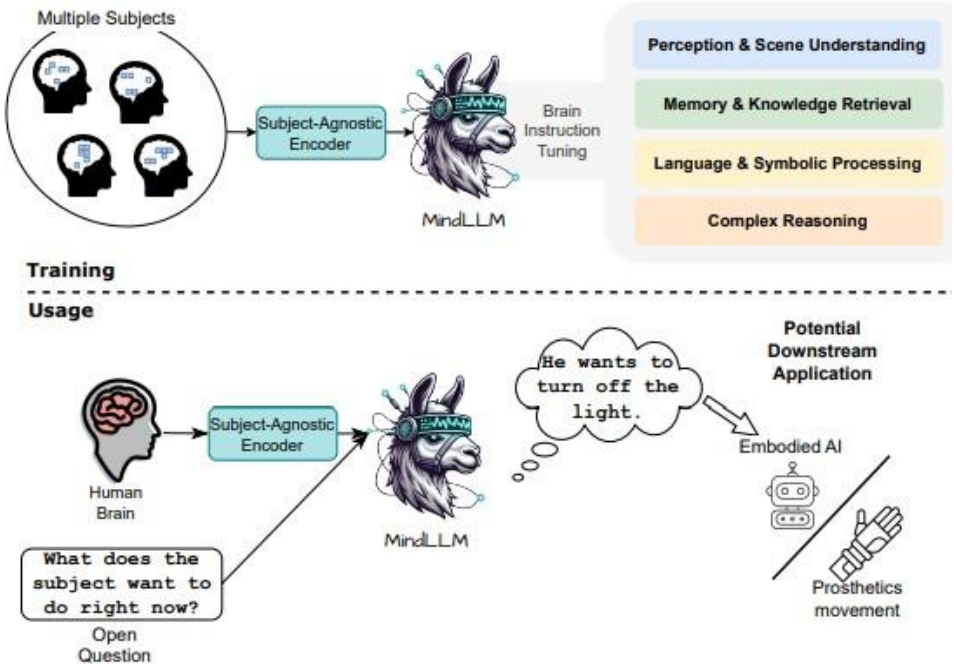
# Linguistic Brain decoding outline

- Sentence Reconstruction: Linking artificial and human neural representations of language
- Story Reconstruction:
  - Semantic reconstruction of continuous language from non-invasive brain recordings
  - Generative Language reconstruction from non-invasive brain recordings
  - **A Subject-Agnostic and Versatile Model for fMRI-to-Text Decoding**

# MindLLM and Neuroscience-Informed Attention

- Stimuli: Natural scenes dataset (NSD), static images
- Stimulus representation: Vicuna
- Brain recording & modality: fMRI, watching

- MindLLM: transforms fMRI recordings into a set of semantic tokens that can be interpreted by a Large Language Model (LLM) like Vicuna.
- A subject-agnostic encoder: it can process any person's brain scan, regardless of voxel count or brain shape.

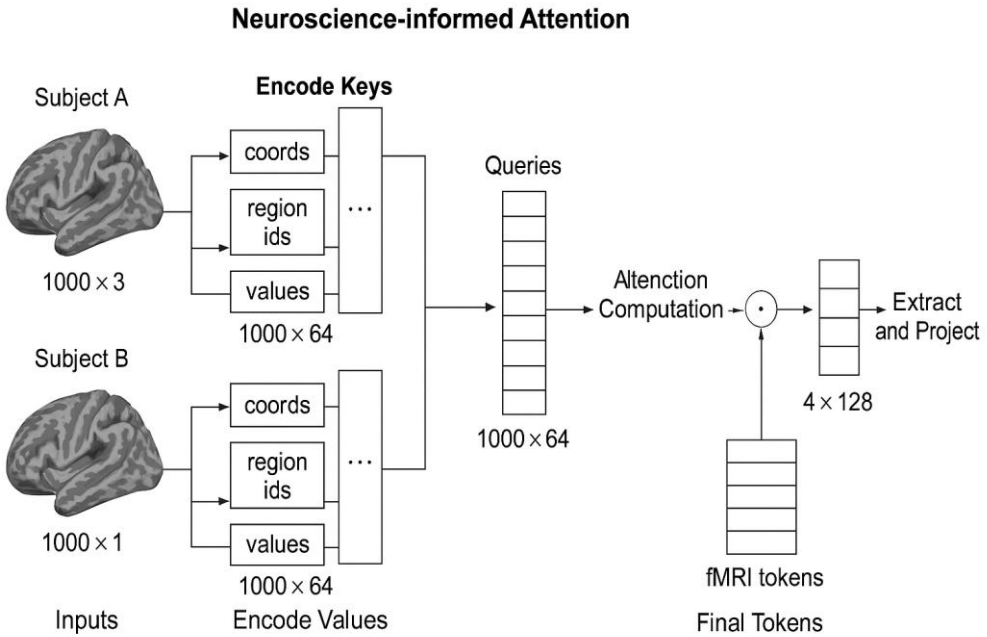



How does neuroscience-informed attention work?

Qiu et al. (2025). A Subject-Agnostic and Versatile Model for fMRI-to-Text Decoding. ICML-2025

# MindLLM and Neuroscience-Informed Attention

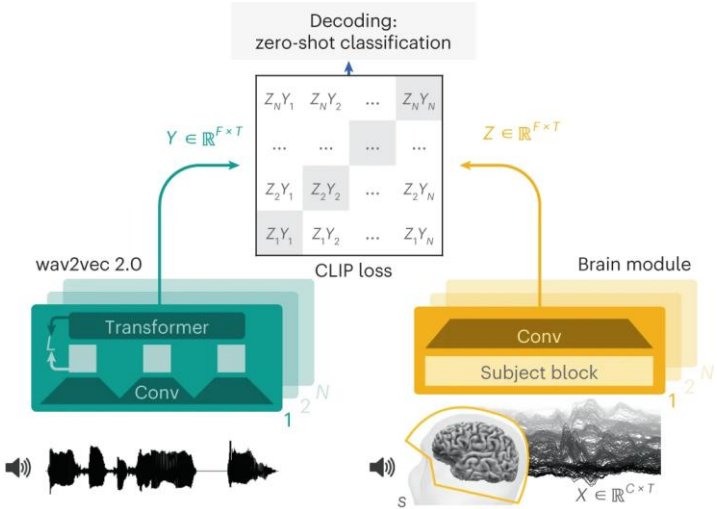
- Values as raw fMRI signals
- Keys as spatial and anatomical context
- Queries as learnable vectors that extract meaningful brain-wide features



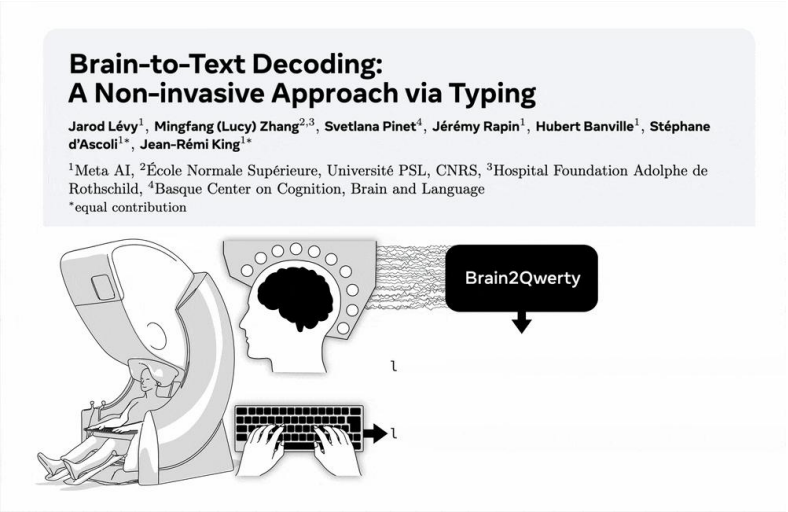

Focuses on static fMRI, without incorporating temporal dynamics.

Qiu et al. (2025). A Subject-Agnostic and Versatile Model for fMRI-to-Text Decoding. ICML-2025

# Linguistic Brain decoding



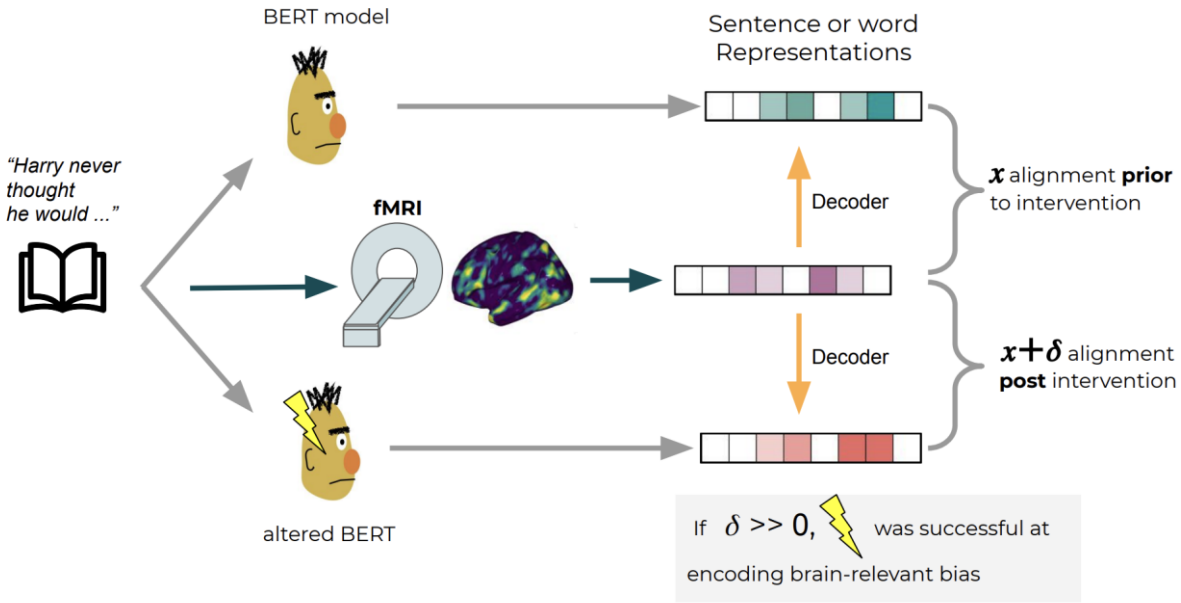

Défossez et al. (2023). Decoding speech perception from non-invasive brain recordings. Nature Machine Intelligence



# DNNs as a model organism

What information is necessary or sufficient to perform a task? To predict brain activity?

Model *organism*, and not simply a model, because DNNs have evolved separately from the human brain




Model organisms allow for direct interventions, which cannot be done in humans

Abdou, Mostafa, Ana Valeria González, Mariya Toneva, Daniel Hershcovich, and Anders Søgaard. "Does injecting linguistic structure into language models lead to better alignment with brain recordings?." *arXiv preprint arXiv:2101.12608* (2021).

# Thank you :)

Anuja Negi

Mathis Lamarre

Chris Tseng

Subba Reddy Oota

Fatma Deniz

Funded by

Federal Ministry of Education and Research

Deutsche Forschungsgemeinschaft
German Research Foundation