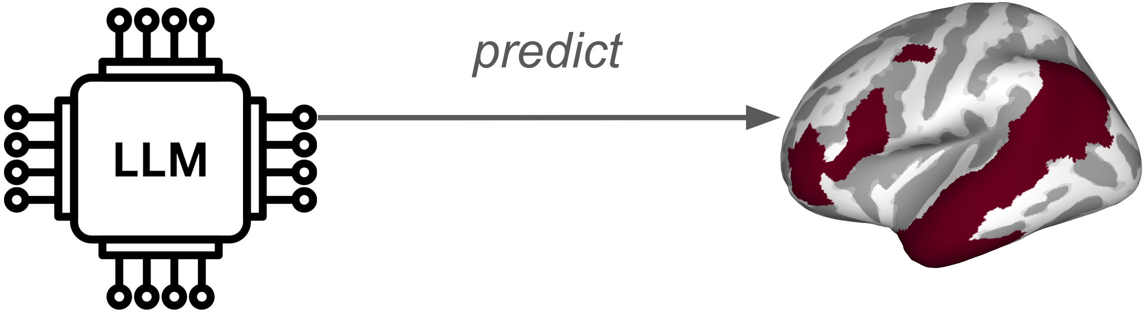
Brain-informed Fine-tuning for improved multilingual understanding in Language Models

Anuja Negi*, Subba Reddy Oota*, Anwar O Nunez-Elizalde, Manish Gupta, Fatma Deniz
Technical University of Berlin / Bernstein Center for Computational Neuroscience Berlin/ Microsoft

Language models accurately predict brain activity during language processing



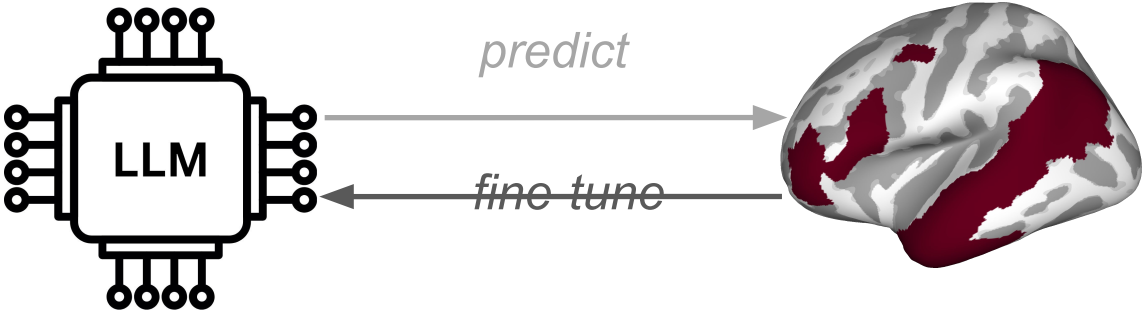
LLM

predict

Wehbe et al., 2014b; Jain & Huth, 2018; Toneva & Wehbe, 2019; Schrimpf et al., 2021; Caucheteux & King, 2022; Goldstein et al., 2022; Karamolegkou et al., 2023; Oota et al., 2025

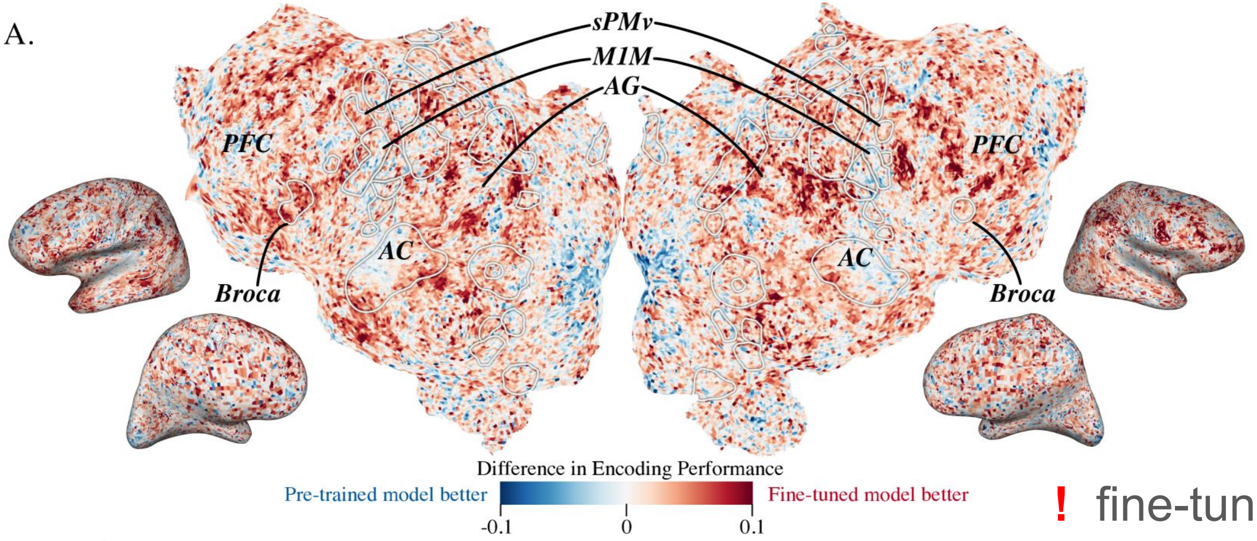
*LLM: Large Language Model

Fine-tuning language models with brain data



Schwartz et al., 2019

Fine-tuning language models with brain data



➔ improves alignment with the brain

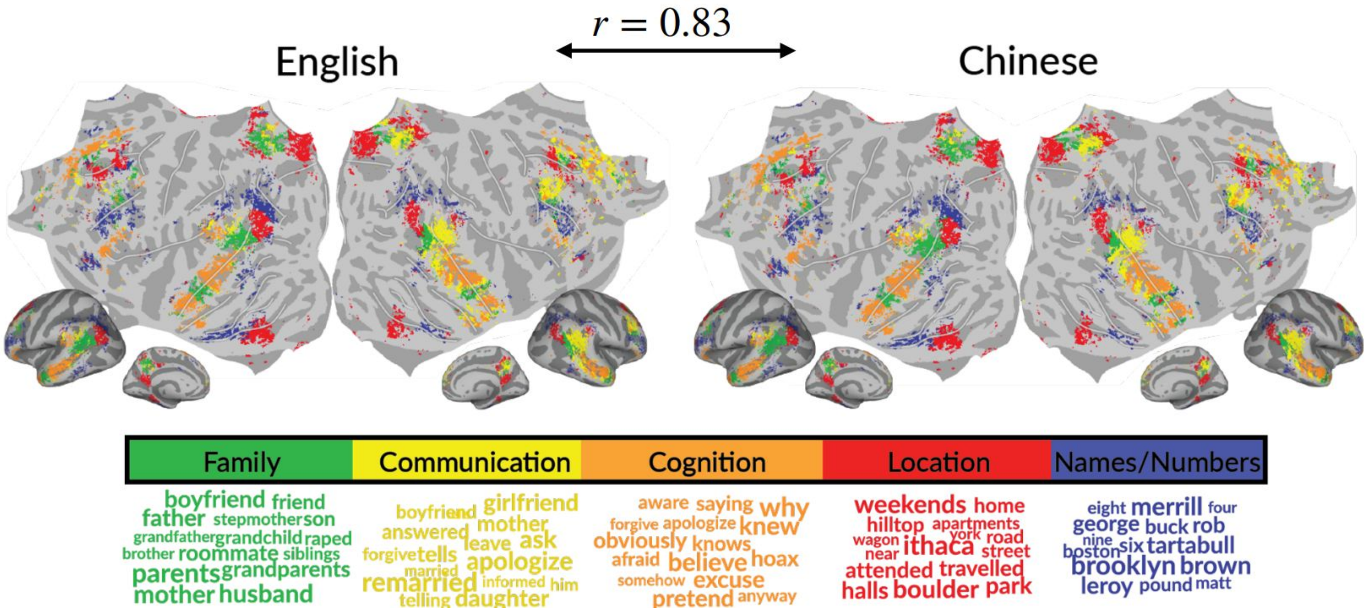
➔ improves their semantic downstream task performance

! fine-tuning with monolingual brain data (English)

! only monolingual models were evaluated

Schwartz et al., 2019; Vattikonda et al. 2025; Moussa et al., 2025

Shared semantic representations in bilinguals



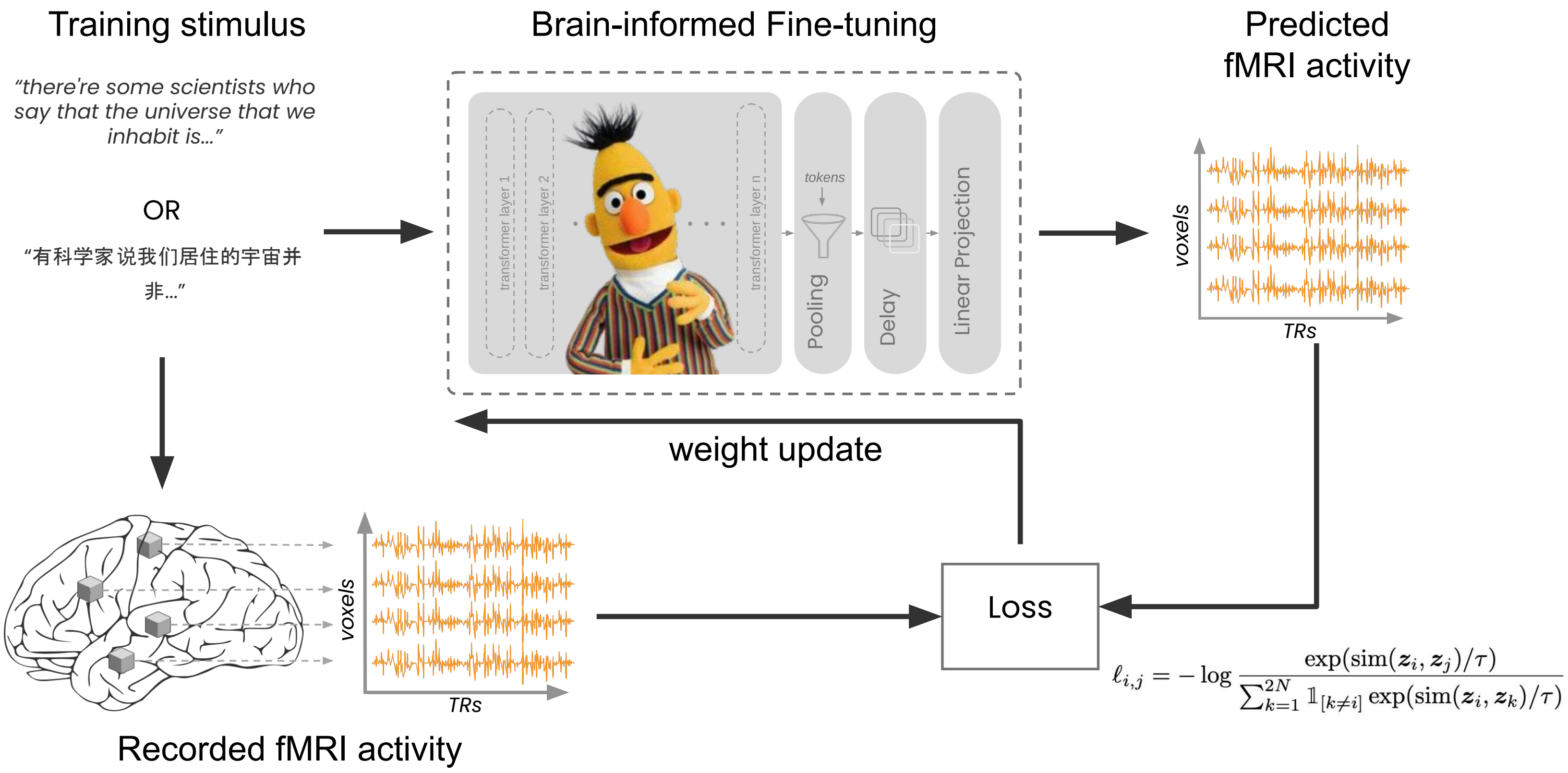
$r = 0.83$
English
Chinese
Family
Communication
Cognition
Location
Names/Numbers
boyfriend friend father stepmother son grandfather grandchild raped brother roommate siblings parents grandparents mother husband
boyfriend girlfriend answered mother leave ask forgive tells married apologize remarried informed him telling daughter
aware saying why forgive apologize knew obviously knows afraid believe hoax somehow excuse pretend anyway
weekends home hilltop apartments wagon york road near ithaca street attended travelled halls boulder park
eight merrill four george buck rob nine boston six tartabull brooklyn brown leroy pound matt

➔ Bilingual language processing relies on shared semantic representations

Chen, Gong, Tseng, Klein, Gallant, Deniz 2024

Can **fine-tuning** language models with **bilingual brain data** elicit multilingual capabilities in them?

Brain-informed fine-tuning with bilingual brain data



Training stimulus

“there're some scientists who say that the universe that we inhabit is...”

OR

“有科学家 说我们居住的宇宙并非...”

Brain-informed Fine-tuning

Predicted fMRI activity

weight update

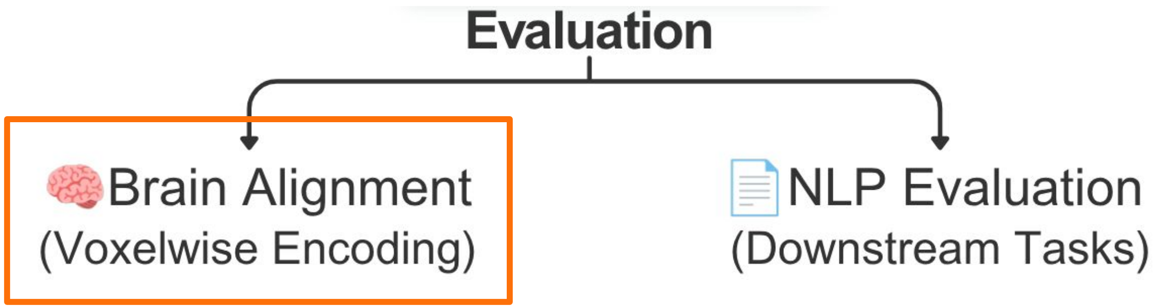
Loss

Recorded fMRI activity

$$\ell_{i,j} = -\log \frac{\exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)}$$

Negi*, Oota*, Nunez-Elizalde, Gupta, Deniz 2025

Evaluating brain-informed fine-tuned models



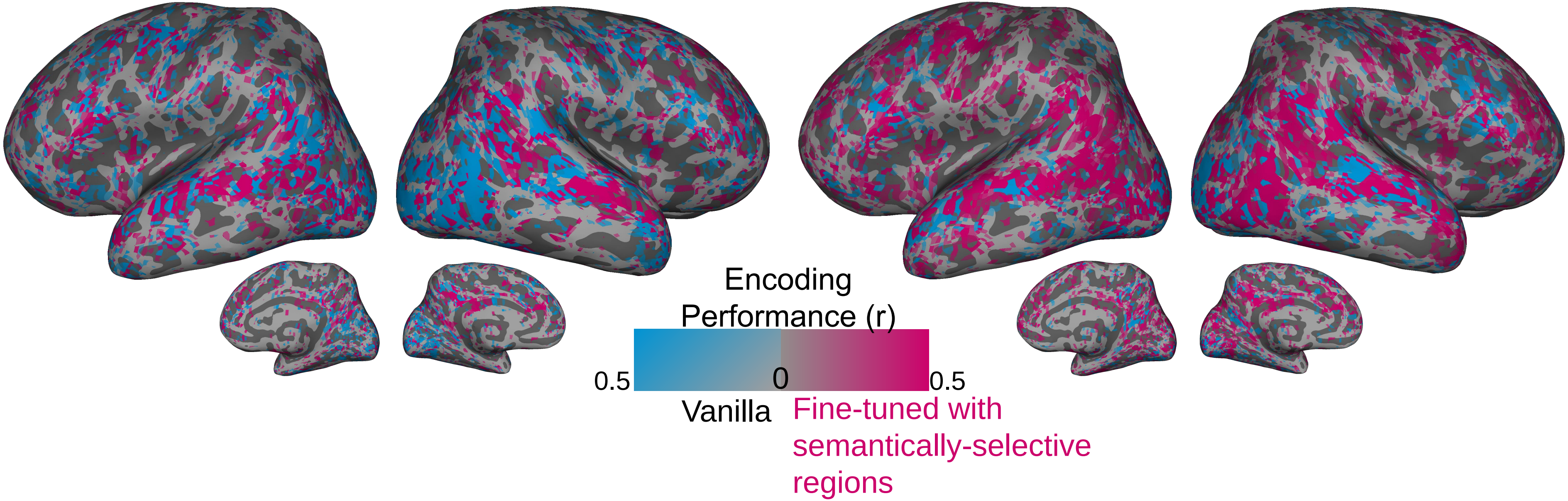
Evaluation

🧠Brain Alignment
(Voxelwise Encoding)

📄NLP Evaluation
(Downstream Tasks)

Negi*, Oota*, Nunez-Elizalde, Gupta, Deniz 2025

Brain-informed fine-tuning improves brain alignment



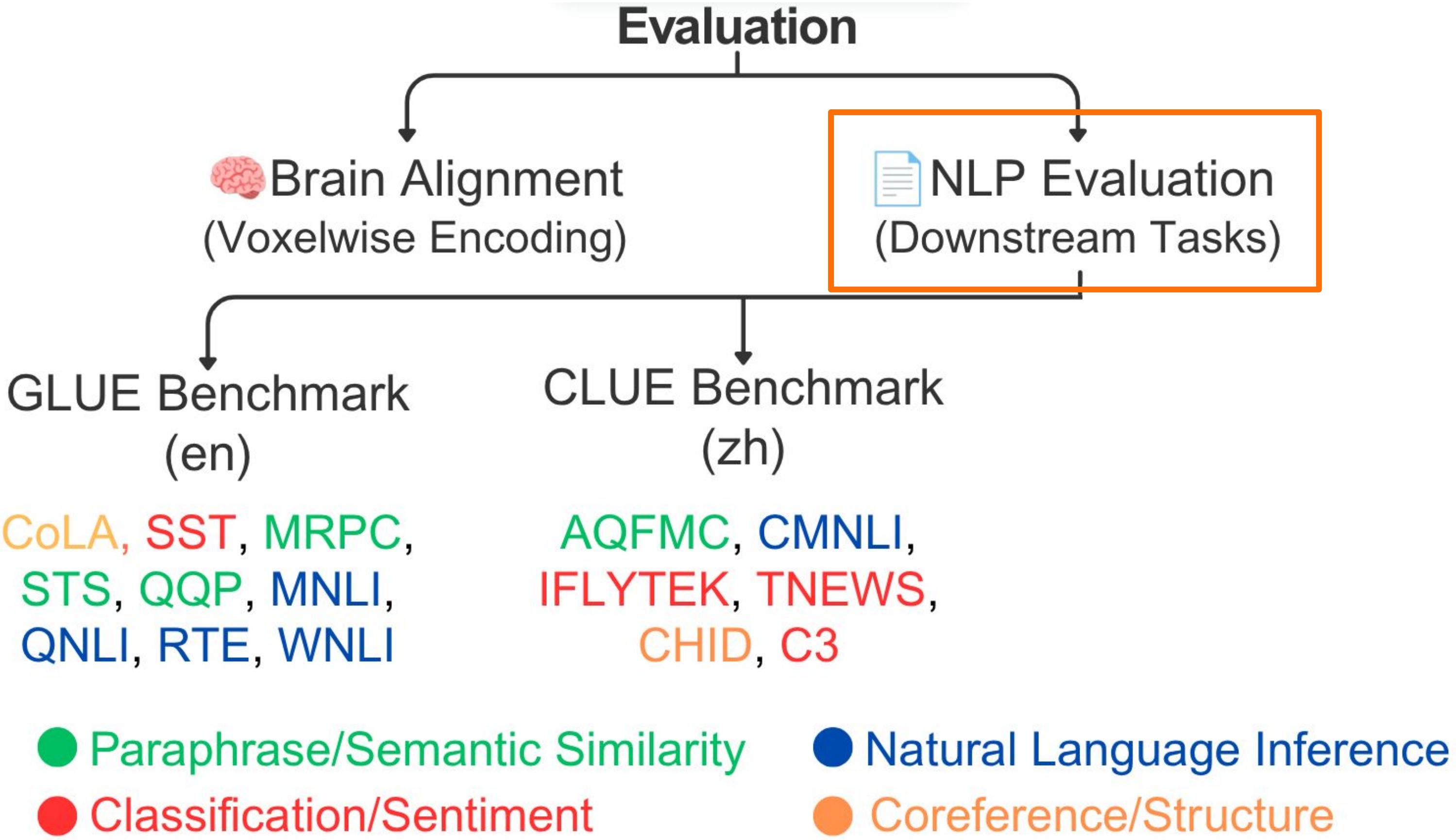
BERT-en fine-tuned with English brain data

BERT-zh fine-tuned with Chinese brain data

*~ **70%** semantic voxels prefer a fine-tuned model over vanilla model*

Negi*, Oota*, Nunez-Elizalde, Gupta, Deniz 2025

Evaluating brain-informed fine-tuned models



Evaluation

🧠Brain Alignment
(Voxelwise Encoding)

📄NLP Evaluation
(Downstream Tasks)

GLUE Benchmark
(en)

CoLA, SST, MRPC,
STS, QQP, MNLI,
QNLI, RTE, WNLI

CLUE Benchmark
(zh)

AQFMC, CMNLI,
IFLYTEK, TNEWS,
CHID, C3

● Paraphrase/Semantic Similarity

● Natural Language Inference

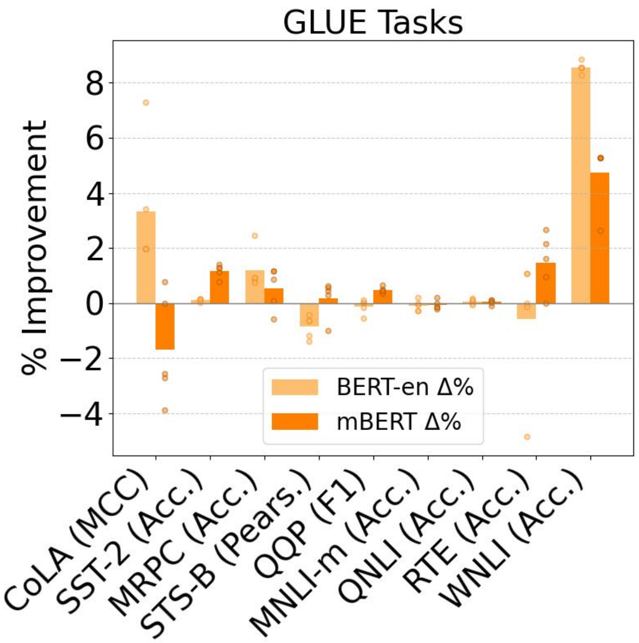
● Classification/Sentiment

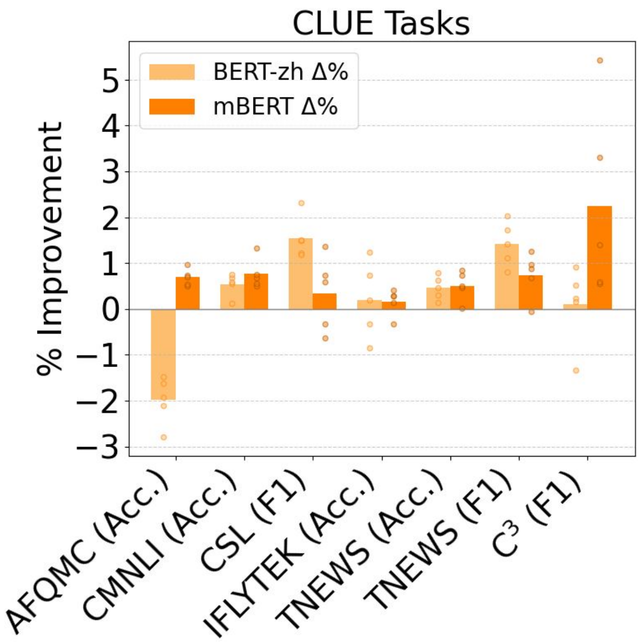
● Coreference/Structure

Negi*, Oota*, Nunez-Elizalde, Gupta, Deniz 2025

Fine-tuning and Evaluation in the Same Language

Fine-tuning improves linguistic task performance





Monolingual model

⬆️ 7/9 on English benchmark (GLUE)

⬆️ 6/7 on Chinese benchmark (CLUE)

Multilingual model

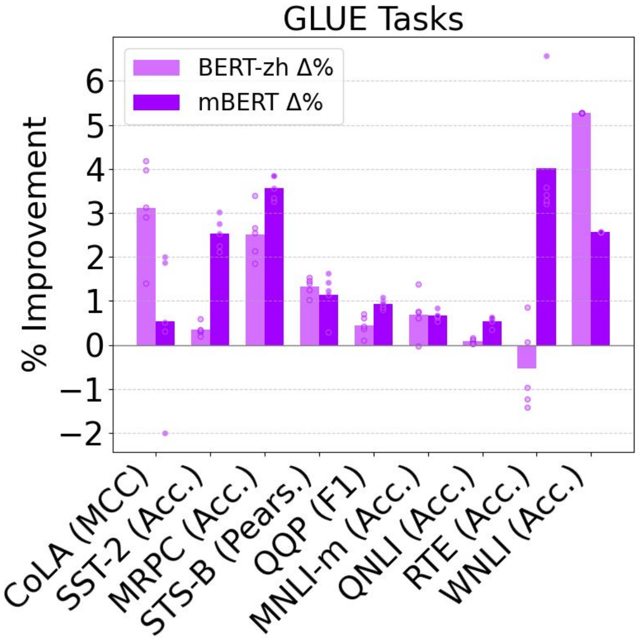
⬆️ 8/9 on English benchmark (GLUE)

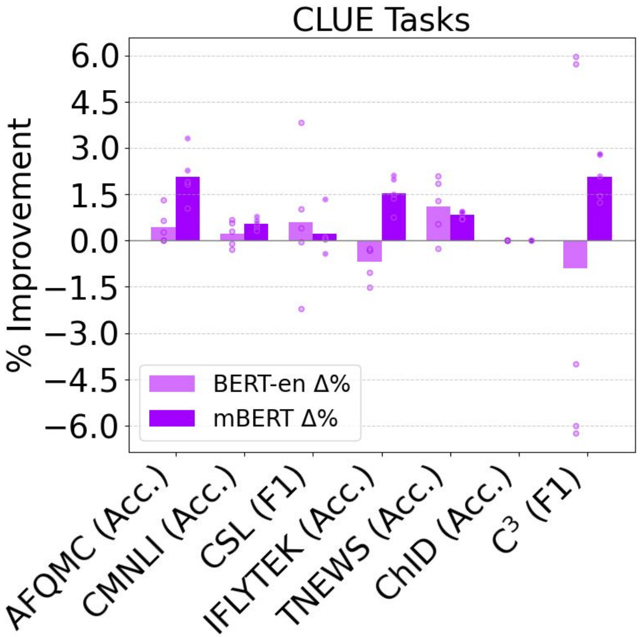
⬆️ 7/7 on Chinese benchmark (CLUE)

Negi*, Oota*, Nunez-Elizalde, Gupta, Deniz 2025

Cross-language evaluation

Fine-tuning enables cross-language transfer





Monolingual model

⬆️ 8/9 on English benchmark (GLUE)

⬆️ 6/7 on Chinese benchmark (CLUE)

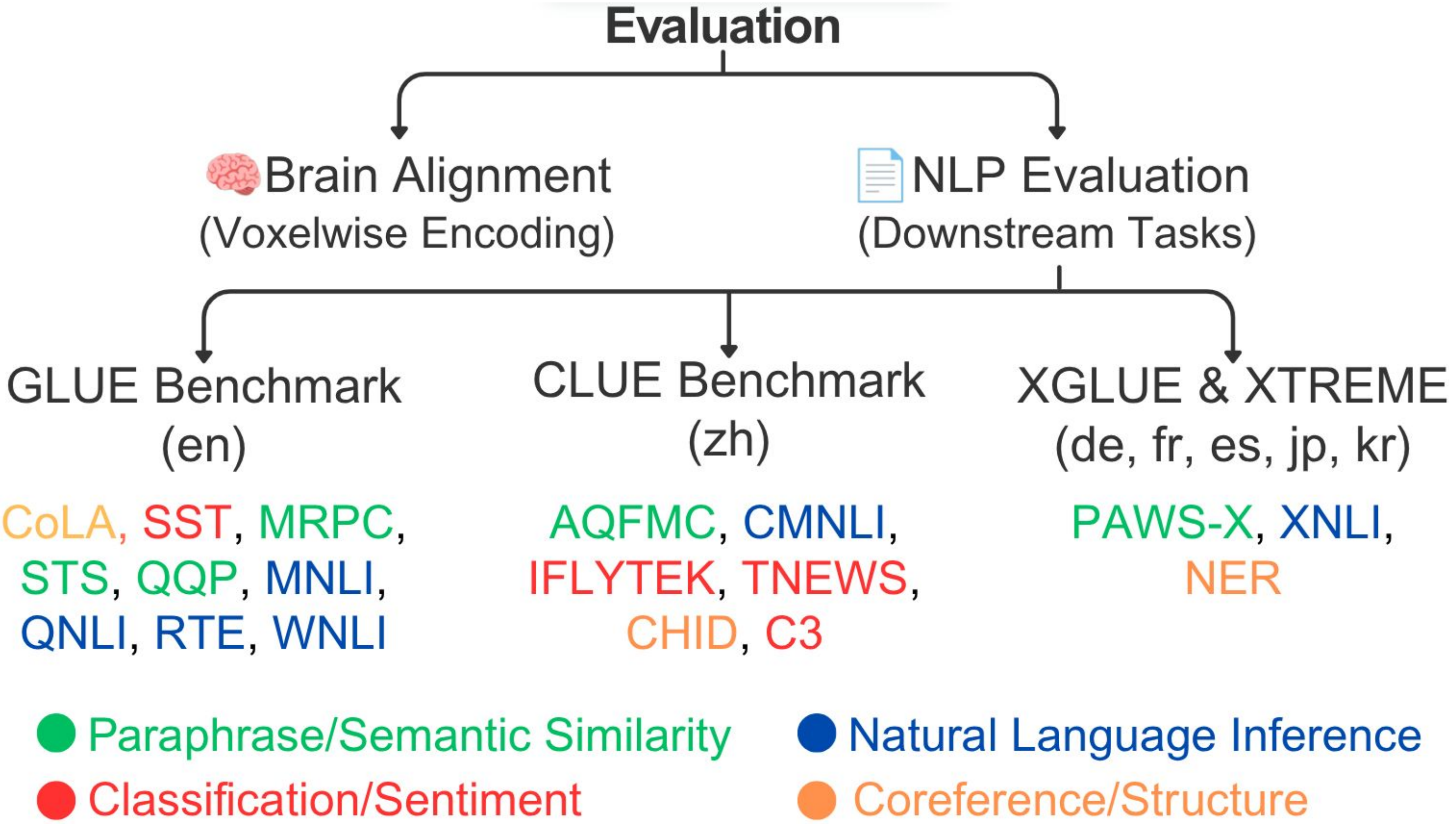
Multilingual model

⬆️ 9/9 on English benchmark (GLUE)

⬆️ 7/7 on Chinese benchmark (CLUE)

Negi*, Oota*, Nunez-Elizalde, Gupta, Deniz 2025

Evaluating brain-informed fine-tuned models



Evaluation

🧠Brain Alignment
(Voxelwise Encoding)

📄NLP Evaluation
(Downstream Tasks)

GLUE Benchmark
(en)

CoLA, SST, MRPC,
STS, QQP, MNLI,
QNLI, RTE, WNLI

CLUE Benchmark
(zh)

AQFMC, CMNLI,
IFLYTEK, TNEWS,
CHID, C3

XGLUE & XTREME
(de, fr, es, jp, kr)

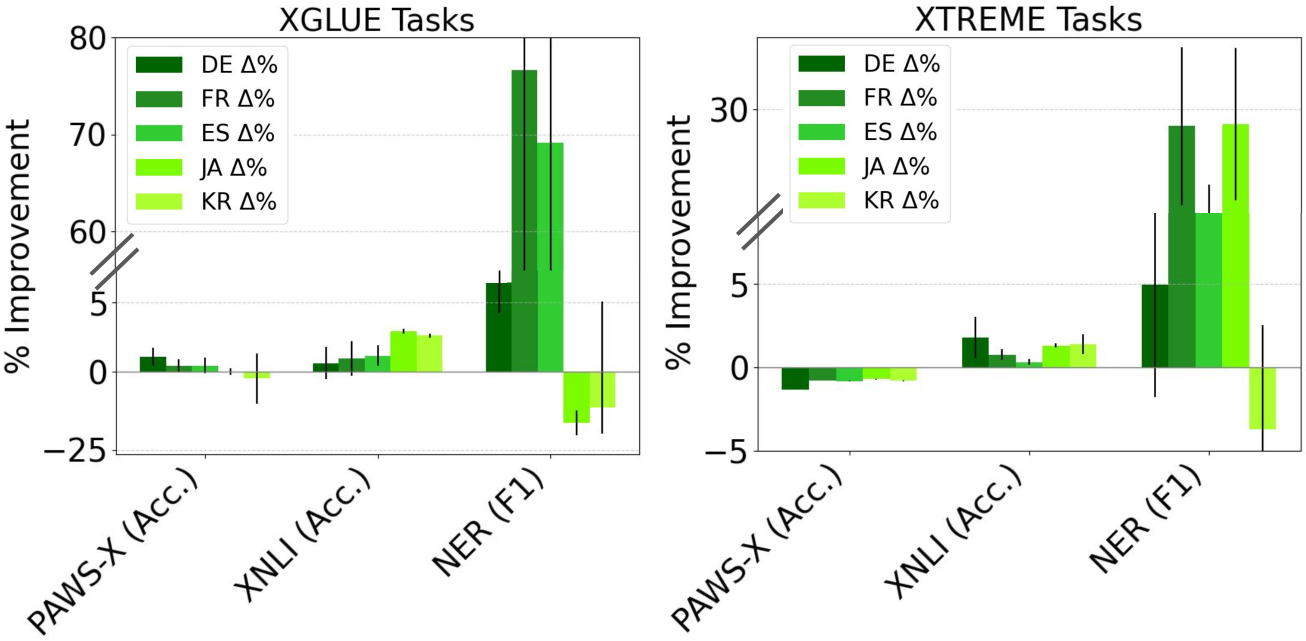
PAWS-X, XNLI,
NER

● Paraphrase/Semantic Similarity
● Natural Language Inference
● Classification/Sentiment
● Coreference/Structure

Negi*, Oota*, Nunez-Elizalde, Gupta, Deniz 2025

Evaluation on unseen languages

Fine-tuning improves language-agnostic representations

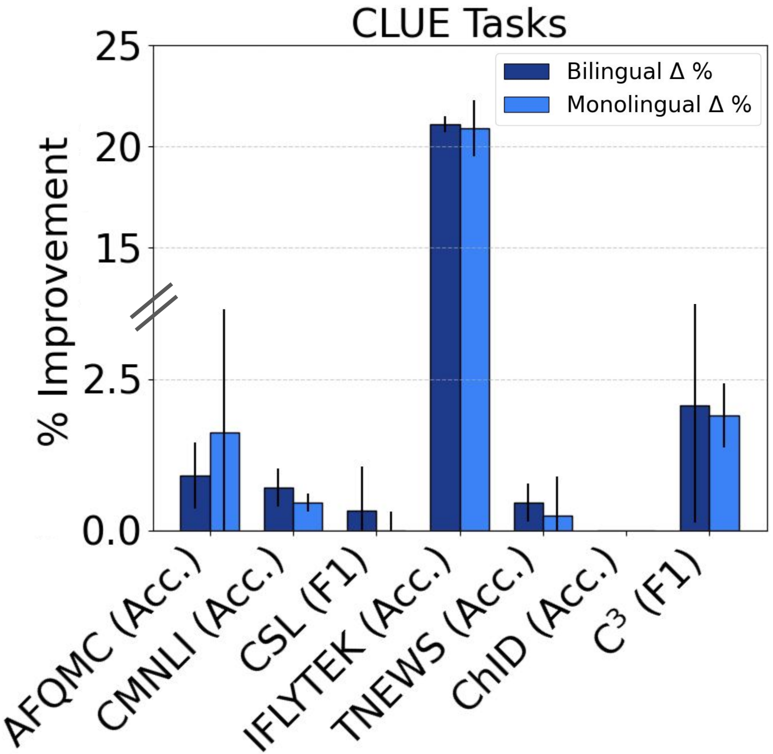


⬆️ 2/3 on DE, FR, ES

Negi*, Oota*, Nunez-Elizalde, Gupta, Deniz 2025

Fine–tuning with monolingual VS bilingual brain data

Cross-linguistic transfer is because of bilingual brain



⬆️ 6/7 bilingual > monolingual

Potentially driven by shared semantics:

- in bilingual brains (Chen et al., 2024)
- across different languages in the brain (de Varda et al., 2025)

Negi*, Oota*, Nunez-Elizalde, Gupta, Deniz 2025

Conclusions

- ➔ First study to perform brain-informed fine-tuning using bilingual brain data.
- ➔ Brain-informed fine-tuning improves
  - ◆ brain alignment
  - ◆ downstream task performance across within-, cross-, and unseen language settings.
- ➔ Improvements are driven specifically by fine-tuning with bilingual brain data, not brain data in general.
- ➔ Potential of leveraging bilingual brain representations for developing language-agnostic models.

Future Work: Explore which linguistic properties the model captures (e.g., syntax, morphology, discourse) to improve model training and evaluation.

Thank you :)

NEURAL INFORMATION PROCESSING SYSTEMS

Denizens LAB

Technische Universität Berlin

Bernstein Network Computational Neuroscience

Microsoft

Anuja Negi

Subba Reddy Oota

Questions?

NeurIPS poster

Fri, 5 Dec
11 a.m. - 2 p.m.

🔗 denizenslab.github.io

Cognitive Computing in Biological and Artificial Systems @ TU Berlin

Anwar Nunez Elizalde

Manish Gupta

Fatma Deniz

erc

Federal Ministry
of Education
and Research