

EEG foundation models: do they actually beat classical decoders?

Brainhack Montréal · Autumn 2026

2 October 2026

Hamza Abdelhedi

PhD student · Université de Montréal · Mila

Before we start

Hi, I'm Hamza

🇨🇦 Telecom engineering

🇨🇦 MSc in AI

🇨🇦 PhD in biomedical engineering, Université de Montréal

Neuro-AI research: vision, decision making, brain disorders

Mila • CoCo Lab • Abundant Intelligences

hamza.abdelhedi@umontreal.ca • hamzaabdelhedi.com

github.com/BabaSanfour • linkedin.com/in/hamza-abdelhedi

Notebooks



github.com/BabaSanfour/meeg-fm-workshop

Plan

TALK

Why foundation models for M/EEG

How they are built

Does it work?

Our experiment

HANDS-ON

Data: BCI IV 2a

01_data

Classical baselines

02_classical

Frozen REVE

03_fm_frozen

Fine-tuning REVE

04_fm_finetune

Who won?

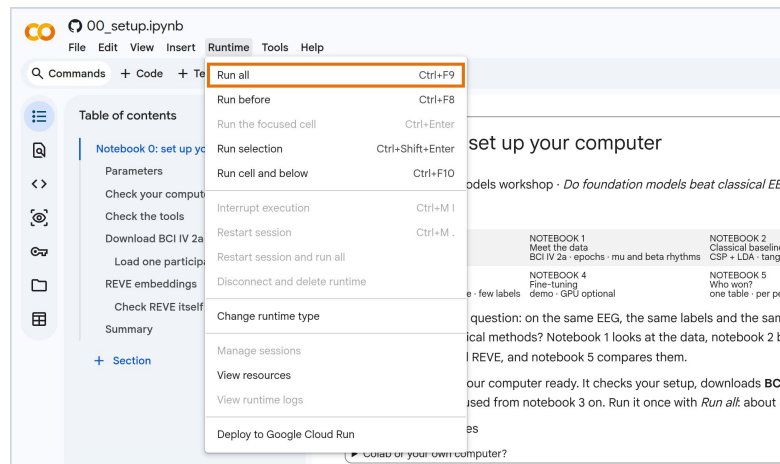
05_compare

Optional: sleep staging on Sleep-EDF

Running 00_setup

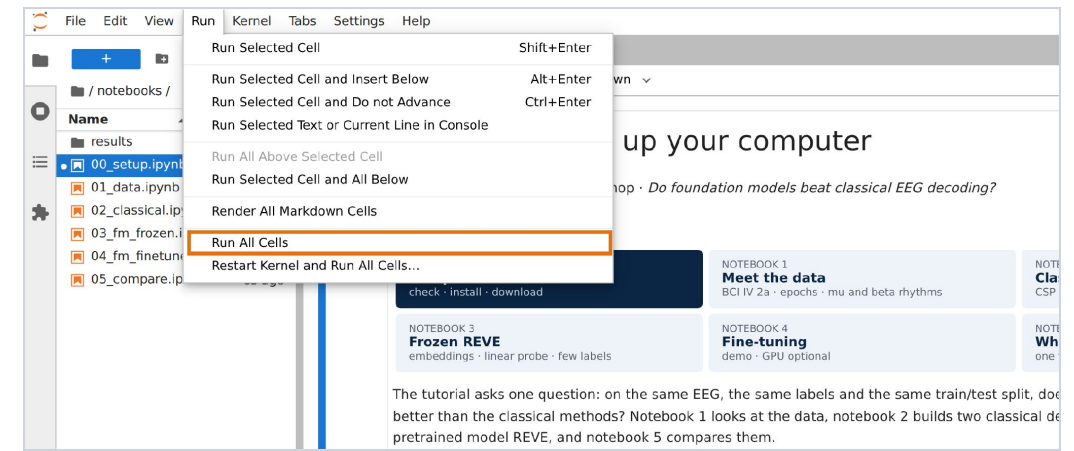
Colab

- 1 Open the repo (QR code on slide 2)
github.com/BabaSanfour/meeg-fm-workshop
- 2 Click **Open in Colab** next to 00_setup
- 3 Runtime > Run all



Local

- 1 `git clone` the repo, then `cd`
github.com/BabaSanfour/meeg-fm-workshop
- 2 `pip install -e "[jupyter]"` or `uv`
- 3 `jupyter lab`, then Run > Run All Cells



Open EEG and MEG data are fragmented

 Search Submit Donate Log In

Quantitative Biology > Neurons and Cognition

[Submitted on 14 Jun 2026]

EEGDash: An open-source platform for machine learning on public neurophysiological data

[Bruno Aristimunha](#), [Aviv Dotan](#), [Pierre Guetschel](#), [Aman Jaiswal](#), [Gal Ashkenazi](#), [Dung Truong](#), [Kuntal Kokate](#), [Amitrava Majumdar](#), [Oren Shriki](#), [Arnaud Delorme](#)

Public neurophysiological datasets are increasingly accessible but remain hard to reuse: turning one into a trained model still takes thousands of lines of code for download, loading, format repair, windowing, and evaluation, and a dataset that meets metadata standards can still fail to load. EEG-Dash is a software resource that catalogues 791 publicly archived recordings (39,778 participants, over 86,051 hours) spanning electroencephalography (EEG), magnetoencephalography (MEG), intracranial EEG (iEEG), electromyography (EMG), and functional near-infrared spectroscopy (fNIRS) from the OpenNeuro and NEMAR archives. It exposes each dataset as an importable, queryable class that preserves signal attributes and loads into machine-learning workflows without custom code, delegating signal handling to MNE-Python, windowing to Braindecode, and format compliance to the official Brain Imaging Data Structure (BIDS) validator. A metadata-first registry adds semantic search, a format-repair layer, automatic dataset-level tags drawn from each source publication, and a feature-extraction framework. The catalogue, with per-record loadability and compliance metadata, supports benchmarking, model development, and cross-dataset analysis.

Access Paper:
[View PDF](#)
[HTML \(experimental\)](#)
[TeX Source](#)
 [view license](#)

Ancillary files (details):

- [dataset_references.bib](#)
- [figures/treemap.pdf](#)
- [references.bib](#)
- [sn-basic.bst](#)
- [supplementary.bbl](#)

(2 additional files not shown)

Current browse context:
q-bio.NC
[< prev](#) | [next >](#)
[new](#) | [recent](#) | [2026-06](#)
Change to browse by:
[q-bio](#)

References & Citations

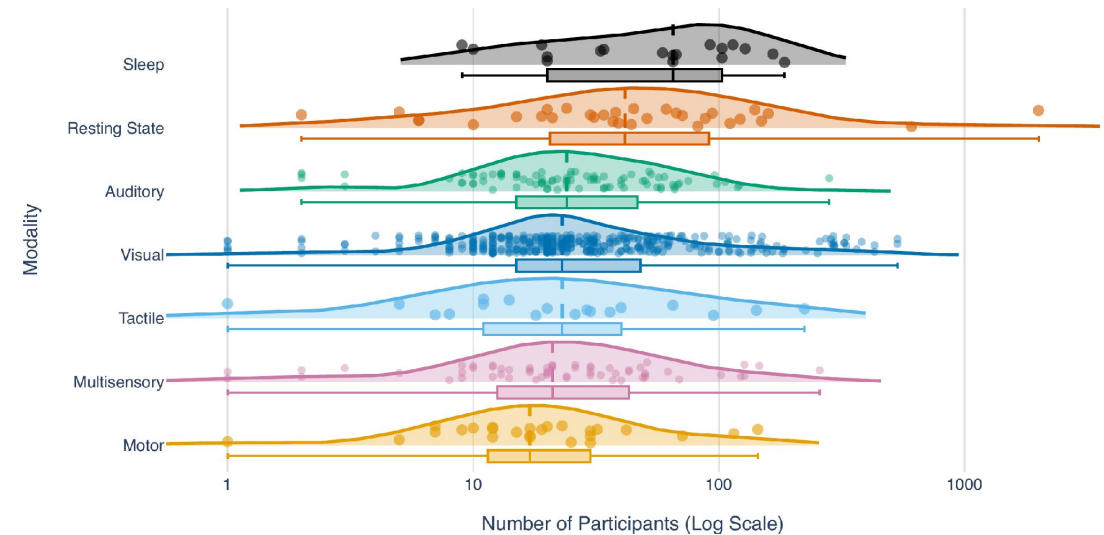
Open EEG and MEG data are fragmented

Open EEG: **659 datasets**, 83,000 h

Open EEG and MEG data are fragmented

Open EEG: **659 datasets**, 83,000 h

Tens of participants



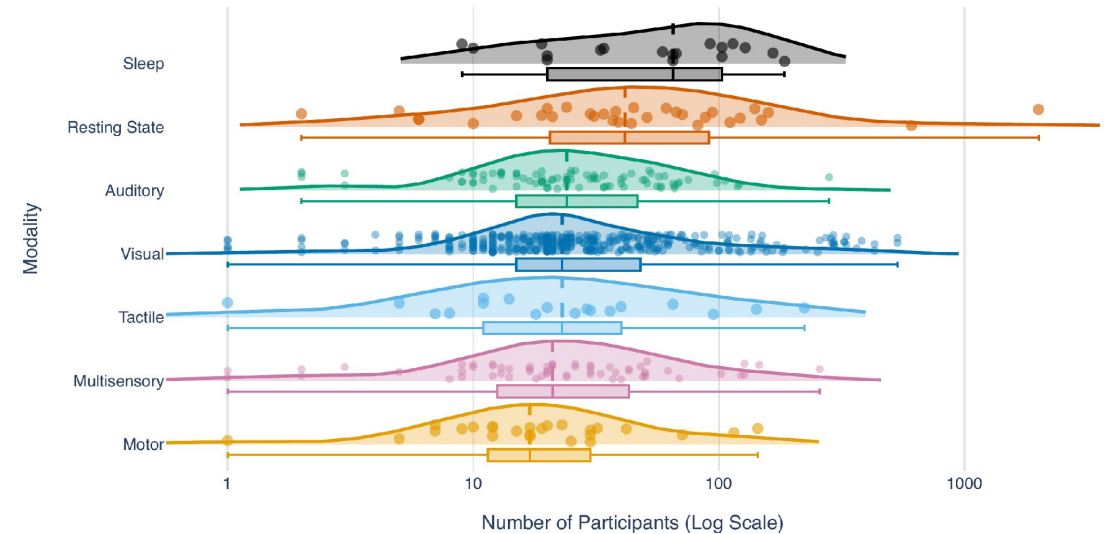
EEG: participants per open dataset, by paradigm (OpenNeuro and NEMAR)

Open EEG and MEG data are fragmented

Open EEG: **659 datasets**, 83,000 h

Tens of participants

Different participants, caps, amplifiers



EEG: participants per open dataset, by paradigm (OpenNeuro and NEMAR)

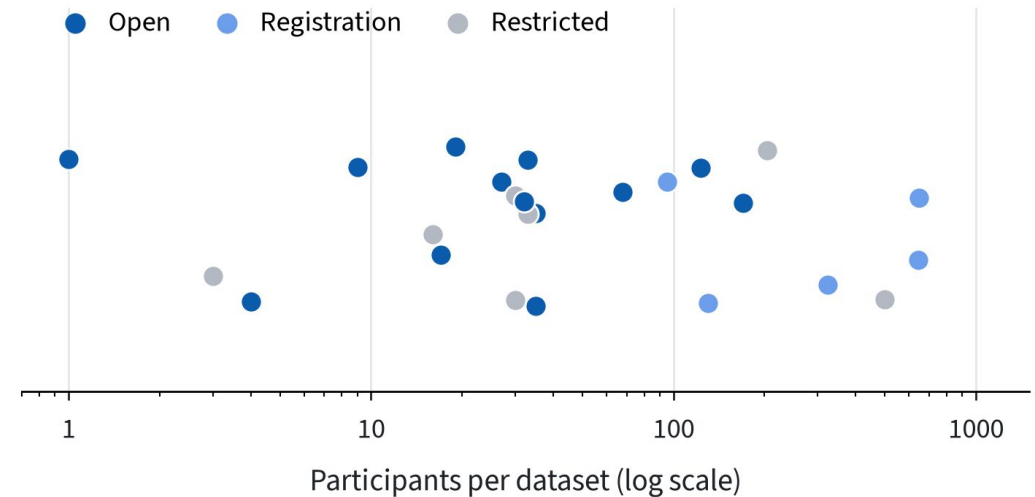
Open EEG and MEG data are fragmented

Open EEG: **659 datasets**, 83,000 h

Tens of participants

Different participants, caps, amplifiers

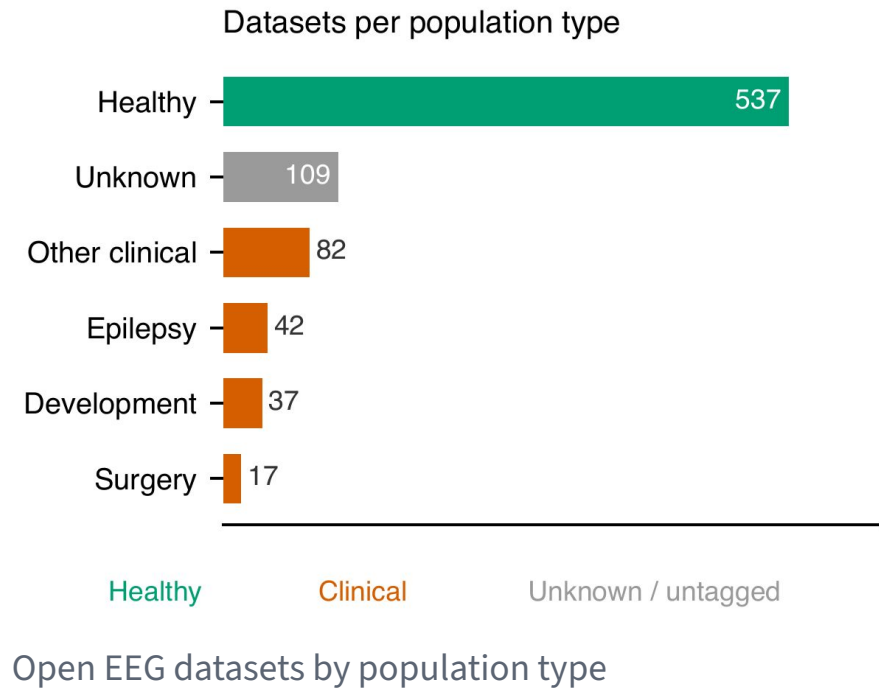
MEG: 25 datasets, ~3,200 participants



MEG: participants per dataset, 25 datasets usable for foundation models

Most EEG data is not open

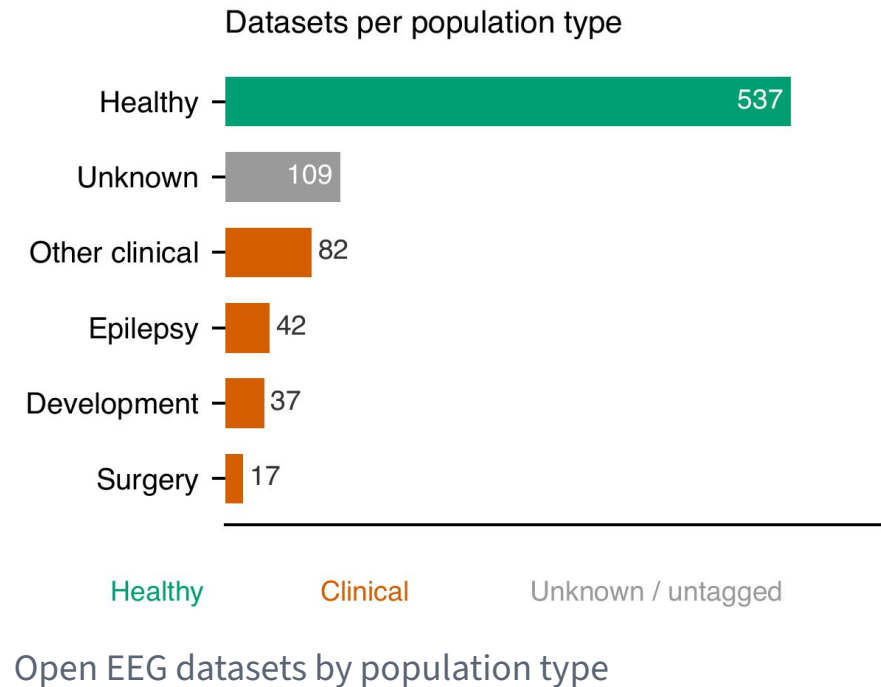
Open EEG: mostly **healthy participants**



Most EEG data is not open

Open EEG: mostly **healthy participants**

Hospitals record hundreds of thousands of EEGs a year, rarely shared

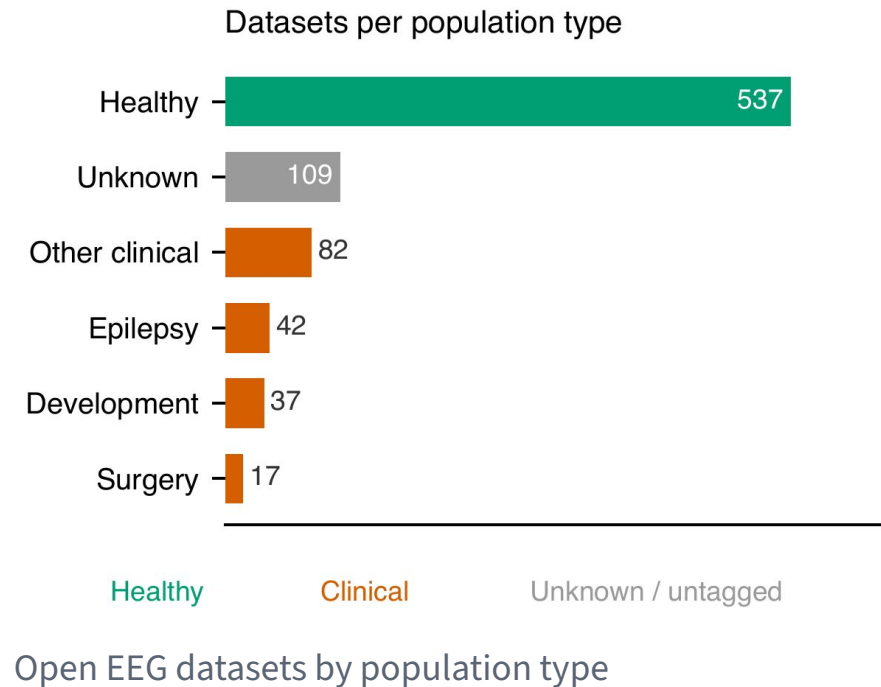


Most EEG data is not open

Open EEG: mostly **healthy participants**

Hospitals record hundreds of thousands of EEGs a year, rarely shared

Rare conditions: few cases per hospital, spread across sites



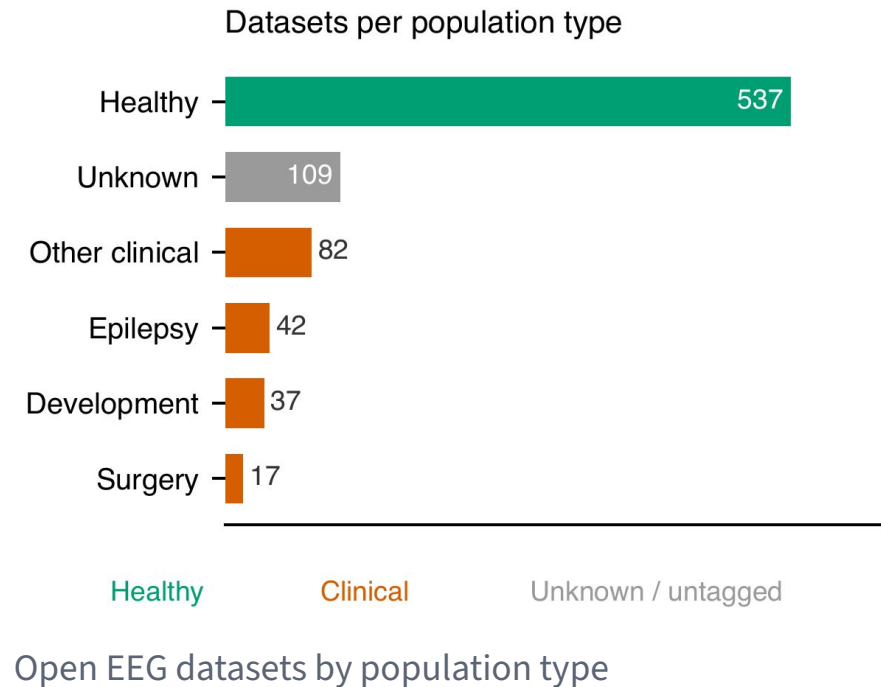
Most EEG data is not open

Open EEG: mostly **healthy participants**

Hospitals record hundreds of thousands of EEGs a year, rarely shared

Rare conditions: few cases per hospital, spread across sites

Lab data: small studies, often never shared



Most EEG data is not open

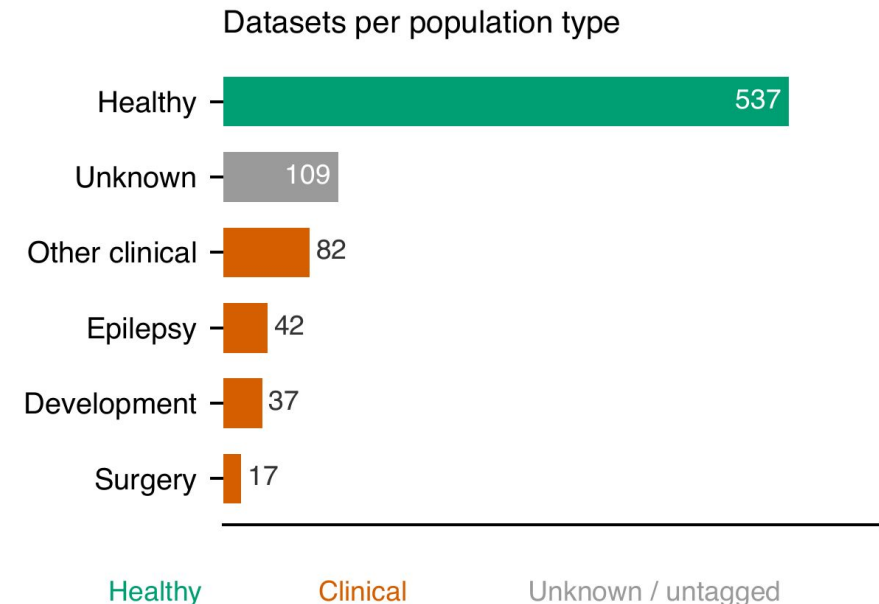
Open EEG: mostly **healthy participants**

Hospitals record hundreds of thousands of EEGs a year, rarely shared

Rare conditions: few cases per hospital, spread across sites

Lab data: small studies, often never shared

Sharing is limited by **patient privacy and consent**



Open EEG datasets by population type

EEG data is noisy and heterogeneous

arXiv Electrical Engineering and Systems Science > Signal Processing [Submitted on 15 Jul 2025 (v1), last revised 24 Dec 2025 (this version, v3)] EEG Foundation Models: A Critical Review of Current Progress and Future Directions Gayal Kuruppu, Neeraj Wagh, Vaclav Kremen, Sandipan Pati, Gregory Worrell, Yogatheesan Varatharajah Premise. Patterns of electrical brain activity recorded via electroencephalography (EEG) offer immense value for scientific and clinical investigations. The inability of supervised EEG encoders to learn robust EEG patterns and their over-reliance on expensive signal annotations have sparked a transition towards general-purpose self-supervised EEG encoders, i.e., EEG foundation models (EEG-FMs), for robust and scalable EEG feature extraction. However, the real-world readiness of early EEG-FMs and the rubrics for long-term research progress remain unclear. Objective. In this work, we conduct a review of ten early EEG-FMs to capture common trends and identify key directions for future development of EEG-FMs. Methods. We comparatively analyze each EEG-FM using three fundamental pillars of foundation modeling, namely the representation of input data, self-supervised modeling, and the evaluation strategy. Based on this analysis, we present a critical synthesis of EEG-FM methodology and find that most EEG-FMs adopt a sequence-based modeling scheme that relies on transformed EEG sequences for self-supervision. However, model evaluations remain heterogeneous, off-the-shelf utility. In addition to adopting standardized and realistic evaluations, future work should make principled and trustworthy choices throughout the EEG representation learning pipeline, benchmarks, software tools, technical methodologies, and applications in collaboration with the world adoption of EEG-FMs.	arXiv Computer Science > Machine Learning [Submitted on 24 Apr 2025 (v1), last revised 21 Sep 2025 (this version, v2)] A Simple Review of EEG Foundation Models: Datasets, Advancements and Future Perspectives Junhong Lai, Jiyu Wei, Lin Yao, Yueming Wang Electroencephalogram (EEG) signals play a crucial role in understanding brain activity and diagnosing neurological diseases. Because supervised EEG encoders are unable to learn robust EEG patterns and rely too heavily on expensive signal annotation, research has turned to general-purpose self-supervised EEG encoders, known as EEG-based models (EEG-FMs), to achieve robust and scalable EEG feature extraction. However, the readiness of early EEG-FMs for practical applications and the standards for long-term research progress remain unclear. Therefore, a systematic and comprehensive review of first-generation EEG-FMs is necessary to understand their current state-of-the-art and identify key directions for future EEG-FMs. To this end, this study reviews 14 early EEG-FMs and provides a critical comprehensive analysis of their methodologies, empirical findings, and unaddressed research gaps. This review focuses on the latest developments in EEG-based models (EEG-FMs), which have shown great potential for processing and analyzing EEG data. We discuss various EEG-FMs, including their architectures, pretraining strategies, pretraining and downstream datasets, and other details. This review also highlights challenges and future directions in the field, aiming to provide a comprehensive overview for researchers and practitioners interested in EEG analysis and related EEG-FM.	arXiv Access Paper: View PDF HTML (experimental) TeX Source CC BY-NC-SA view license Current browse context: eess.SP < prev next > new recent 2025-07 Change to browse by: cs cs.AI cs.LG	arXiv Access Paper: View PDF HTML (experimental) TeX Source view license Current browse context: cs.LG < prev next > new recent 2025-04 Change to browse by: cs cs.AI eess eess.SP References & Citations NASA ADS Google Scholar Semantic Scholar Export BibTeX Citation
--	---	--	---

EEG data is noisy and heterogeneous

Low signal-to-noise: skull and scalp weaken and blur the signal

EEG data is noisy and heterogeneous

Low signal-to-noise: skull and scalp weaken and blur the signal

Artifacts: eye blinks, muscle, heart, movement, power-line noise

EEG data is noisy and heterogeneous

Low signal-to-noise: skull and scalp weaken and blur the signal

Artifacts: eye blinks, muscle, heart, movement, power-line noise

Variability: across people, sessions, age, medication and sleep state

EEG data is noisy and heterogeneous

Low signal-to-noise: skull and scalp weaken and blur the signal

Artifacts: eye blinks, muscle, heart, movement, power-line noise

Variability: across people, sessions, age, medication and sleep state

Recording setups: hardware, montage, reference, sampling rate, protocol

EEG data is noisy and heterogeneous

Low signal-to-noise: skull and scalp weaken and blur the signal

Artifacts: eye blinks, muscle, heart, movement, power-line noise

Variability: across people, sessions, age, medication and sleep state

Recording setups: hardware, montage, reference, sampling rate, protocol

Labels: expert annotation is expensive, subjective and error-prone

EEG data is noisy and heterogeneous

Low signal-to-noise: skull and scalp weaken and blur the signal

Artifacts: eye blinks, muscle, heart, movement, power-line noise

Variability: across people, sessions, age, medication and sleep state

Recording setups: hardware, montage, reference, sampling rate, protocol

Labels: expert annotation is expensive, subjective and error-prone

Formats: one in three OpenNeuro EEG datasets passes BIDS validation

How can we learn meaningful representations from EEG data?

What is a (neural) foundation model?



Search Submit Donate Log In

Quantitative Biology > Neurons and Cognition

[Submitted on 3 Sep 2026 (v1), last revised 21 Sep 2026 (this version, v2)]

A Roadmap for MEG Foundation Models

Philipp Thölke, Hamza Abdelhedi, Yorguin Mantilla-Ramos, Fouad Lbakali, Oumayma Gharbi, Catherine Duclos, Annalisa Pascarella, Vanessa Hadid, Oiwi Parker Jones, Karim Jerbi

Foundation models are beginning to reshape brain-signal analysis by moving the field beyond task-specific decoding pipelines toward reusable models pretrained on broad neural datasets. Magnetoencephalography (MEG) is a compelling but still underdeveloped target for this shift: it captures human cortical dynamics at millisecond resolution while offering stronger spatial interpretability than EEG, making it especially valuable for source-resolved studies of perception, language, cognition, and clinical brain function. Yet MEG foundation models remain at an early stage, with only a small number of MEG-specific and MEG-inclusive multi-modal models, modest pretraining corpora, and emerging but still limited benchmarks. This perspective lays down the basic concepts needed to understand MEG foundation models and provides a didactic overview of the field's key design choices, including tokenization, sensor-versus source-space representations, sensor-geometry encoding, backbone architectures, self-supervised objectives, and pretraining data. We then offer a roadmap for future development, organized around native MEG pretraining, adaptation of EEG foundation models, transfer from generic time-series models, and multi-modal integration with EEG, fMRI, MRI, behaviour, and stimulus features. We highlight the need for coordinated infrastructure, including diverse and reusable MEG datasets, rigorous evaluation across subjects, sites, tasks, and clinical settings, and responsible data-sharing practices that address consent, privacy, access, and governance.

Access Paper:

[View PDF](#)
[HTML \(experimental\)](#)
[TeX Source](#)

 [view license](#)

Current browse context:

q-bio.NC
< [prev](#) | [next](#) >
[new](#) | [recent](#) | [2026-09](#)

Change to browse by:

[CS](#)
[cs.AI](#)
[q-bio](#)

References & Citations

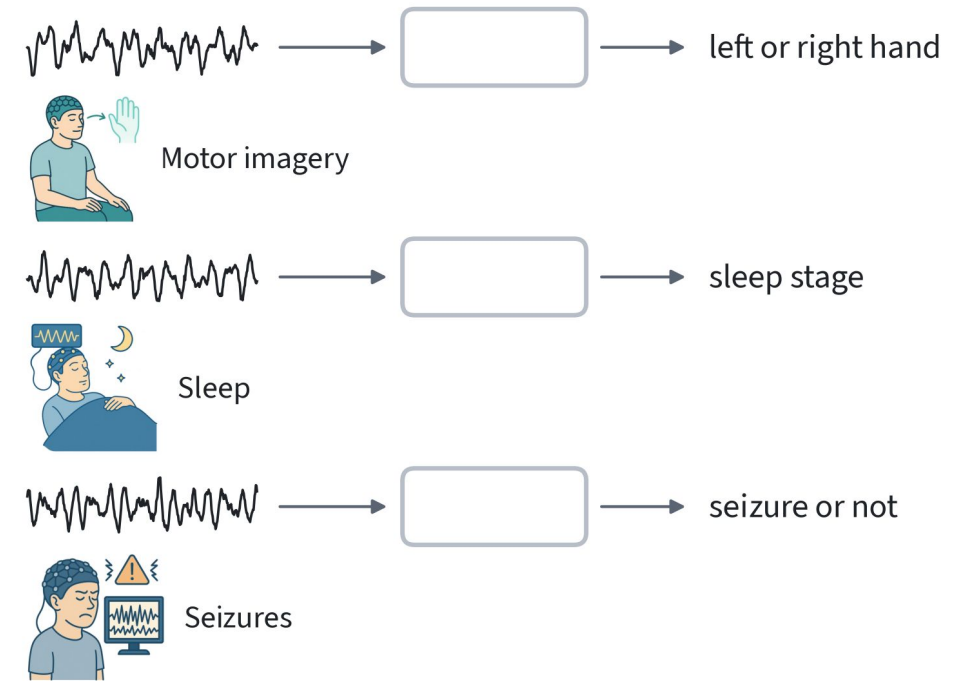
[NASA ADS](#)
[Google Scholar](#)
[Semantic Scholar](#)

[Export BibTeX Citation](#)

Bookmark

What is a (neural) foundation model?

Standard approach: **one model per task**

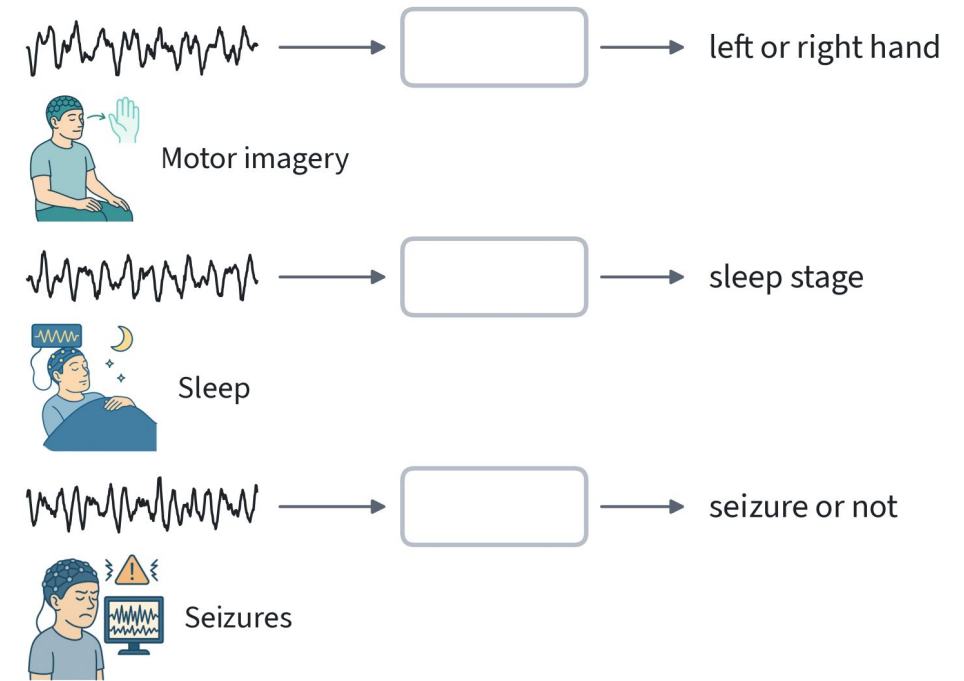


Example: EEG

What is a (neural) foundation model?

Standard approach: **one model per task**

Each model is trained **from scratch**



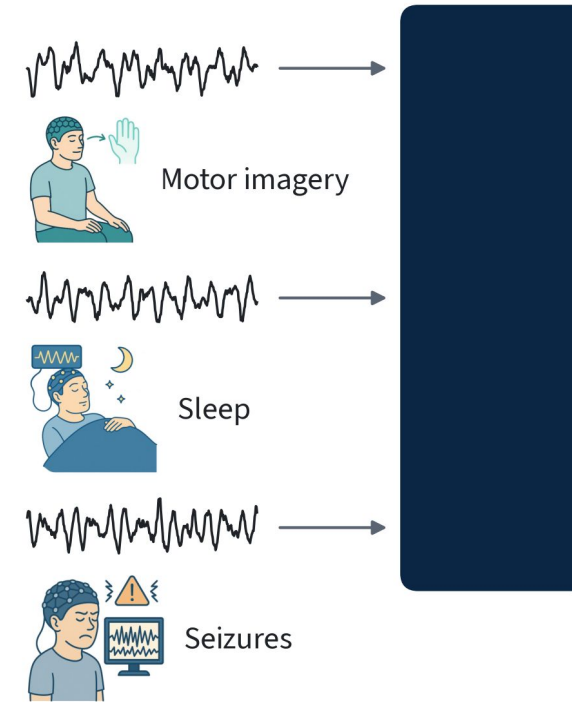
Example: EEG

What is a (neural) foundation model?

Standard approach: **one model per task**

Each model is trained **from scratch**

Foundation model: **pretrained once** on large brain datasets



Example: EEG

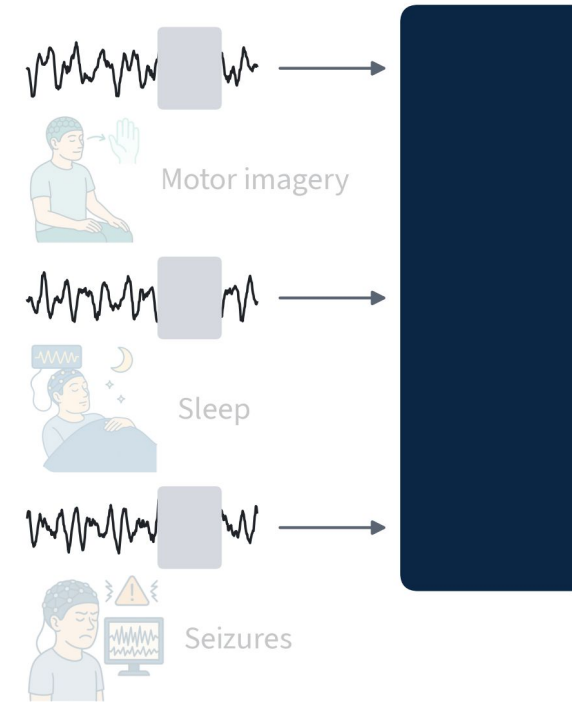
What is a (neural) foundation model?

Standard approach: **one model per task**

Each model is trained **from scratch**

Foundation model: **pretrained once** on large brain datasets

Pretraining uses **no labels**



Example: EEG

What is a (neural) foundation model?

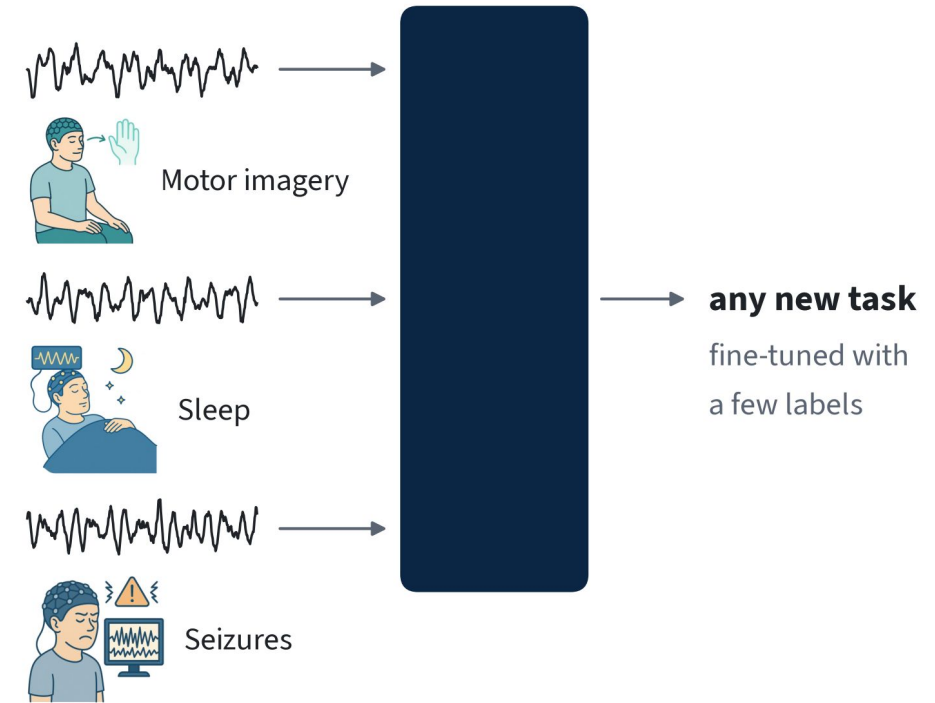
Standard approach: **one model per task**

Each model is trained **from scratch**

Foundation model: **pretrained once** on large brain datasets

Pretraining uses **no labels**

Then **fine-tuned** for each new task



Example: EEG

Ingredients of a foundation model

Data: the recordings used for pretraining

Tokens: how the signal is split into units

Positions: where and when each token was recorded

Backbone: the network architecture

Self-supervised learning: training without labels

Ingredient 1: data

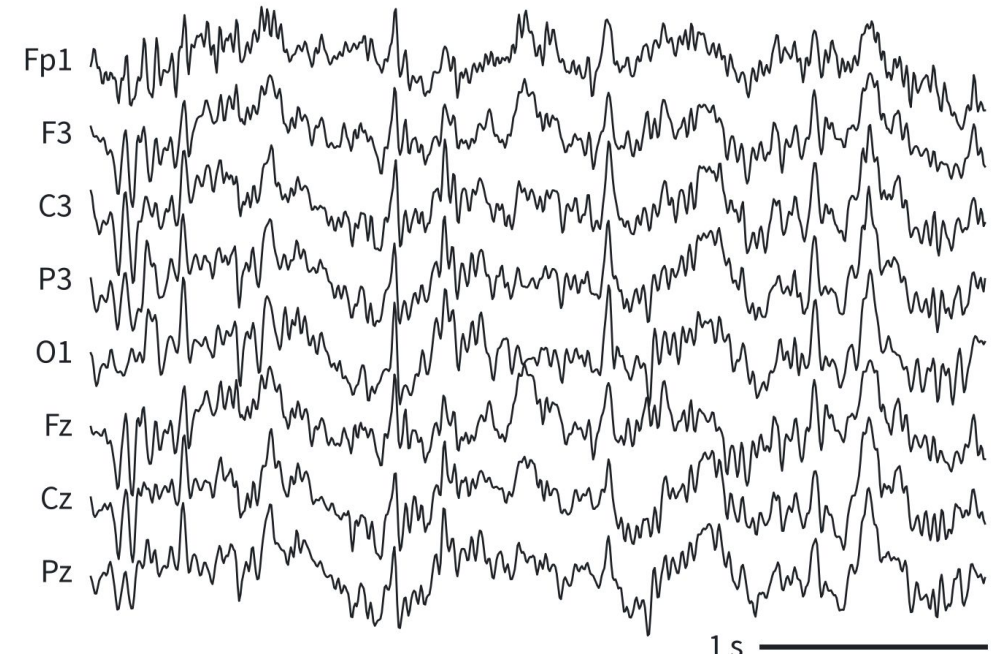
EEG: scalp electrodes measure the brain's electrical activity



Ingredient 1: data

EEG: scalp electrodes measure the brain's electrical activity

Each electrode records one **channel**

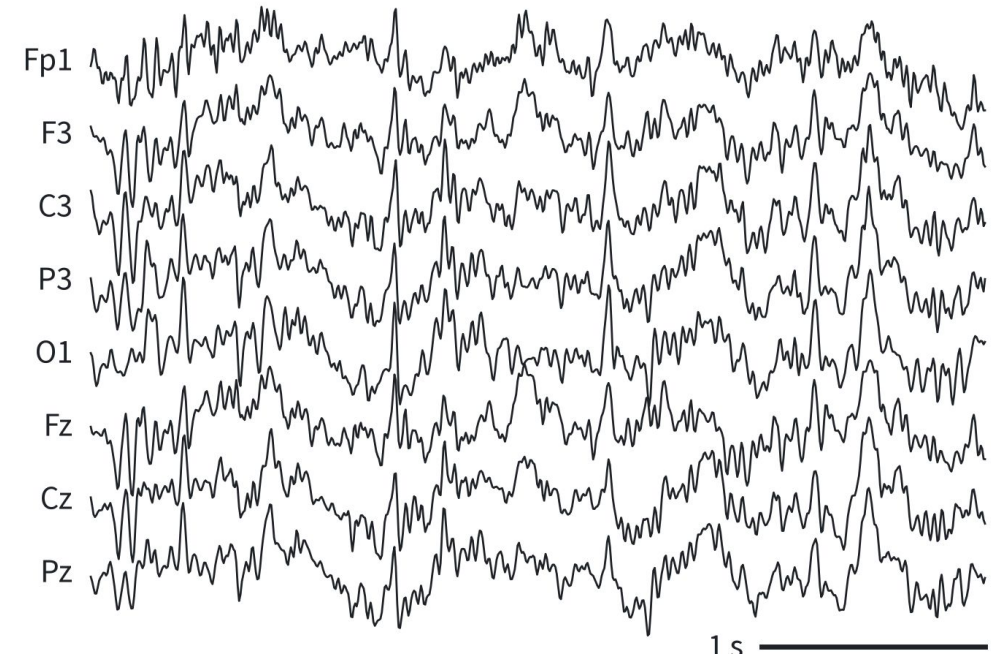


Ingredient 1: data

EEG: scalp electrodes measure the brain's electrical activity

Each electrode records one **channel**

Sampling rate: 128 to 1000 Hz



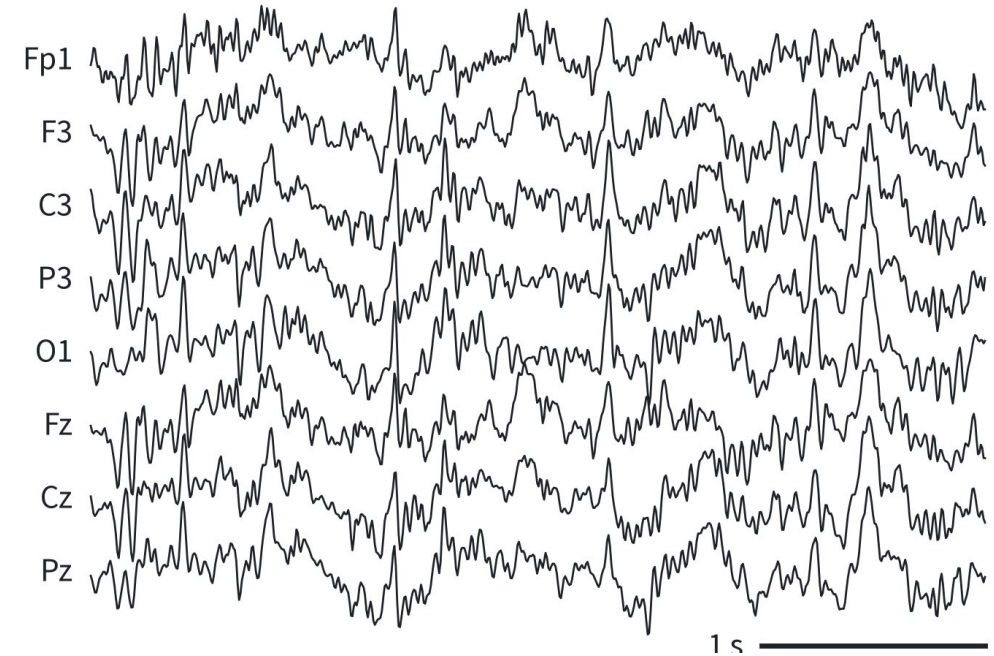
Ingredient 1: data

EEG: scalp electrodes measure the brain's electrical activity

Each electrode records one **channel**

Sampling rate: 128 to 1000 Hz

19 to 256 electrodes per system



Ingredient 1: data

EEG: scalp electrodes measure the brain's electrical activity

Each electrode records one **channel**

Sampling rate: 128 to 1000 Hz

19 to 256 electrodes per system

Pretraining data: thousands of hours, many datasets, **unlabelled**



Motor imagery



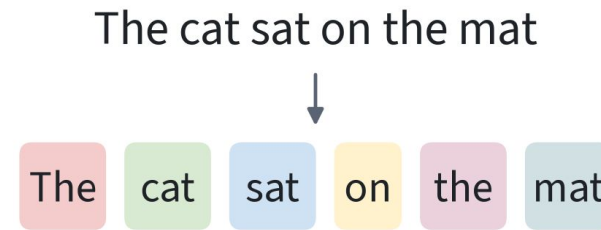
Sleep



Seizures

Ingredient 2: tokens

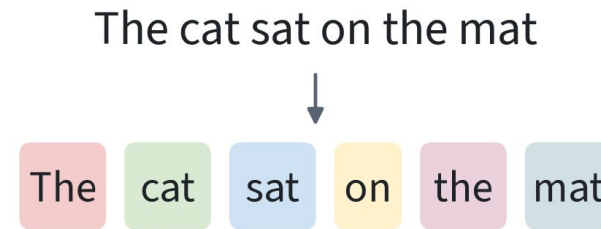
Text: a sentence is split into units called **tokens**



Ingredient 2: tokens

Text: a sentence is split into units called **tokens**

The model processes the token sequence

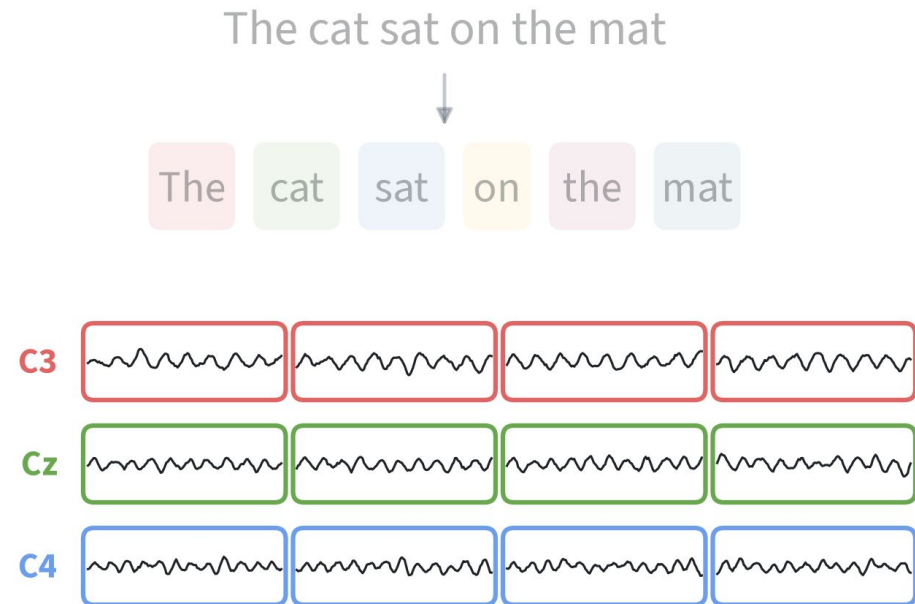


Ingredient 2: tokens

Text: a sentence is split into units called **tokens**

The model processes the token sequence

EEG: each channel is split into short time windows



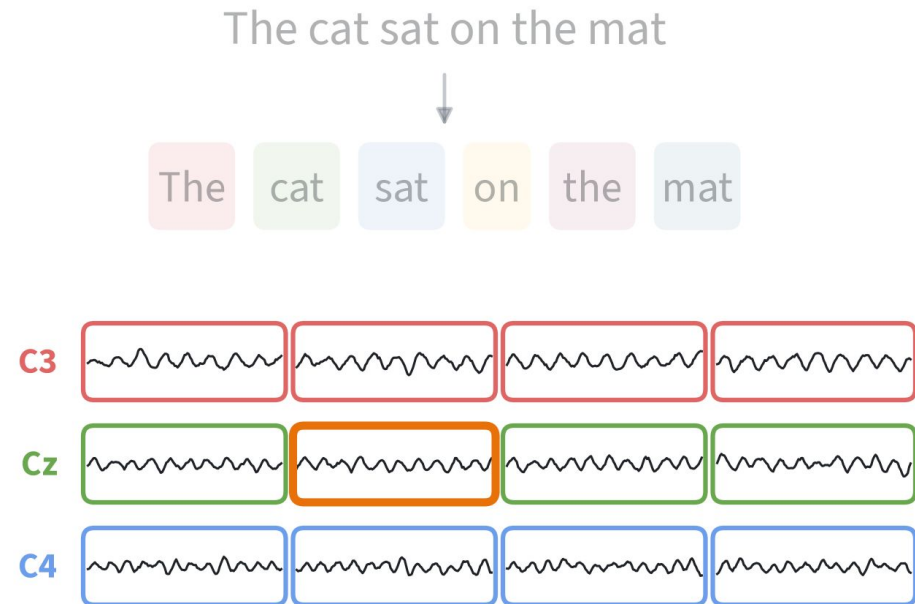
Ingredient 2: tokens

Text: a sentence is split into units called **tokens**

The model processes the token sequence

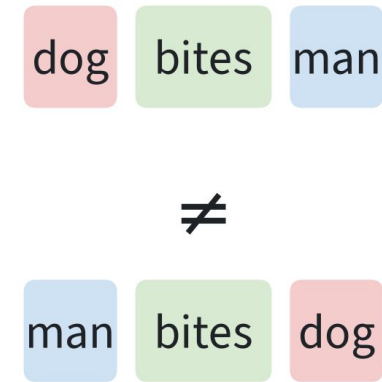
EEG: each channel is split into short time windows

Each window (about 1 s) is **one token**



Ingredient 3: positions

Text: word order changes the meaning



Ingredient 3: positions

Text: word order changes the meaning

Each token is given its **position** in the sequence

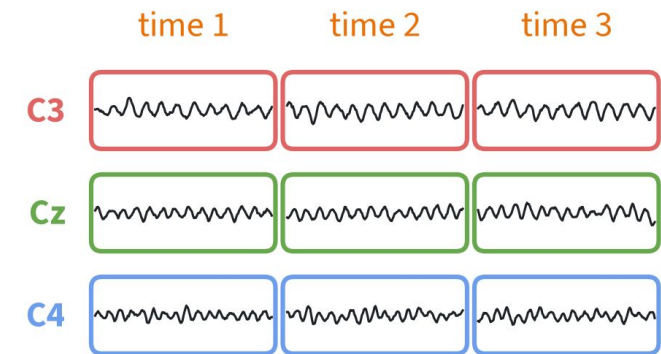
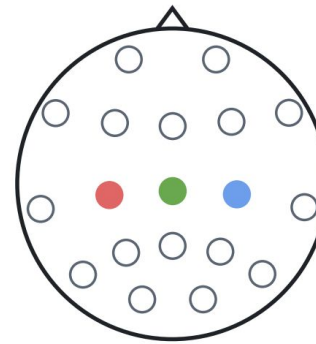


Ingredient 3: positions

Text: word order changes the meaning

Each token is given its **position** in the sequence

EEG: each token is given its **time** and **electrode location**



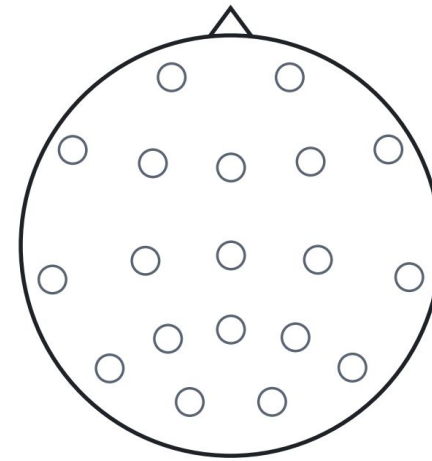
Ingredient 3: positions

Text: word order changes the meaning

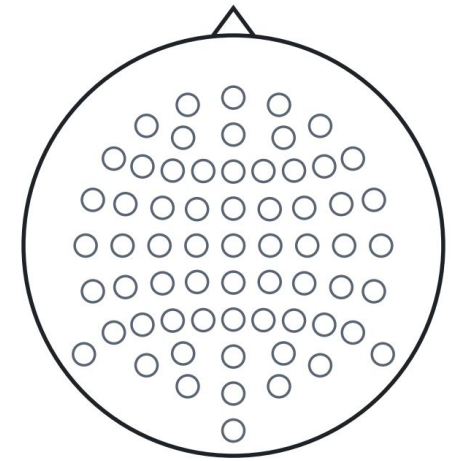
Each token is given its **position** in the sequence

EEG: each token is given its **time** and **electrode location**

Electrode locations let one model handle **different caps**



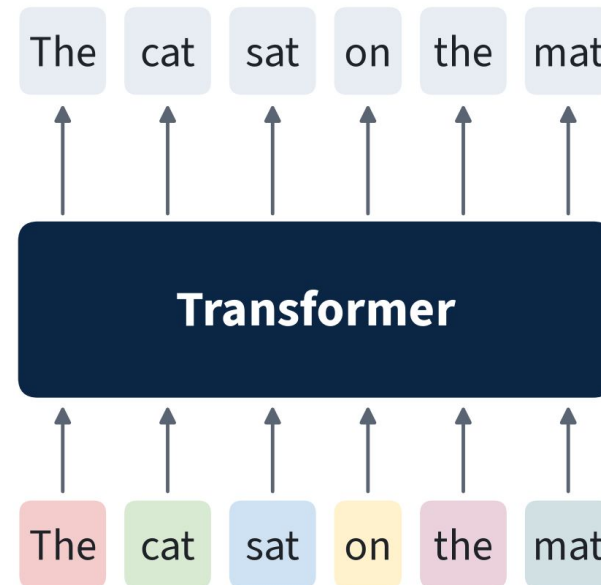
19 electrodes



64 electrodes

Ingredient 4: backbone

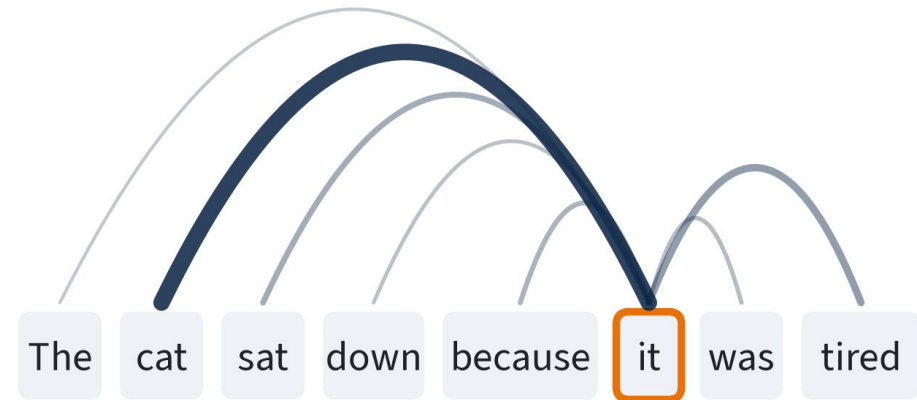
Standard architecture: the **transformer**



Ingredient 4: backbone

Standard architecture: the **transformer**

Attention: each token is updated with a weighted sum of all tokens

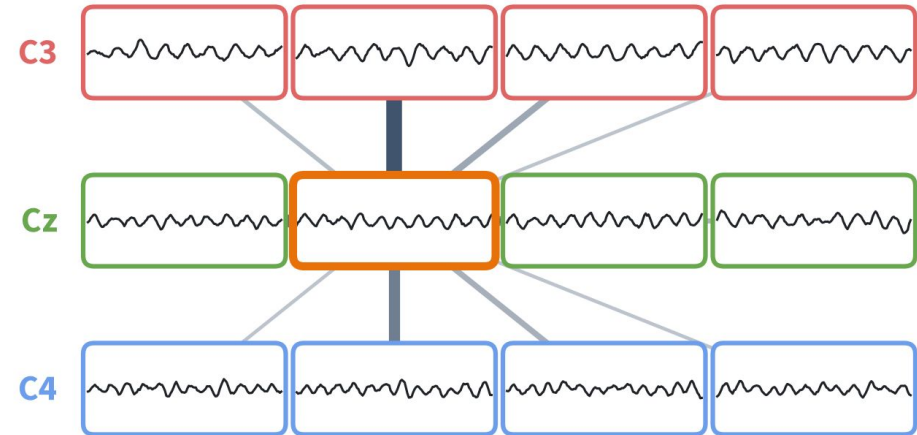


Ingredient 4: backbone

Standard architecture: the **transformer**

Attention: each token is updated with a weighted sum of all tokens

EEG: attention across channels and across time



Ingredient 5: self-supervised learning

The training target comes from **the input itself**, not from labels

Masked word

The cat ? on the mat

Next word

The cat sat on the ?

Ingredient 5: self-supervised learning

The training target comes from **the input itself**, not from labels

Text: predict a **masked** word or the **next** word

Masked word

The cat ? on the mat

Next word

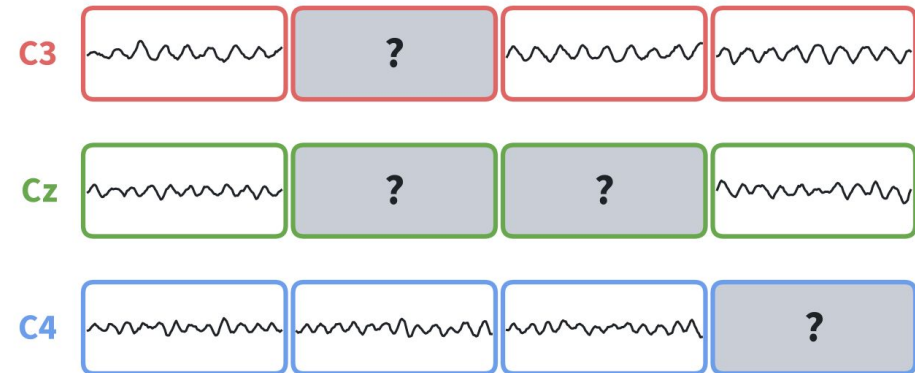
The cat sat on the ?

Ingredient 5: self-supervised learning

The training target comes from **the input itself**, not from labels

Text: predict a **masked** word or the **next** word

EEG: mask some windows



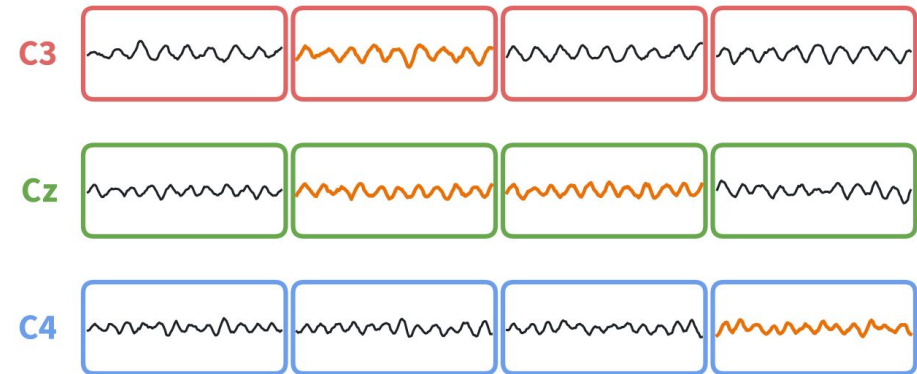
Ingredient 5: self-supervised learning

The training target comes from **the input itself**, not from labels

Text: predict a **masked** word or the **next** word

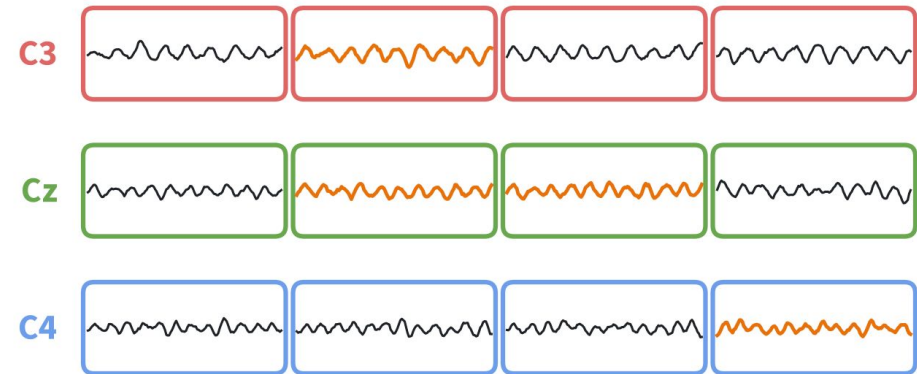
EEG: mask some windows

and **reconstruct** them from the visible ones



Four self-supervised objectives

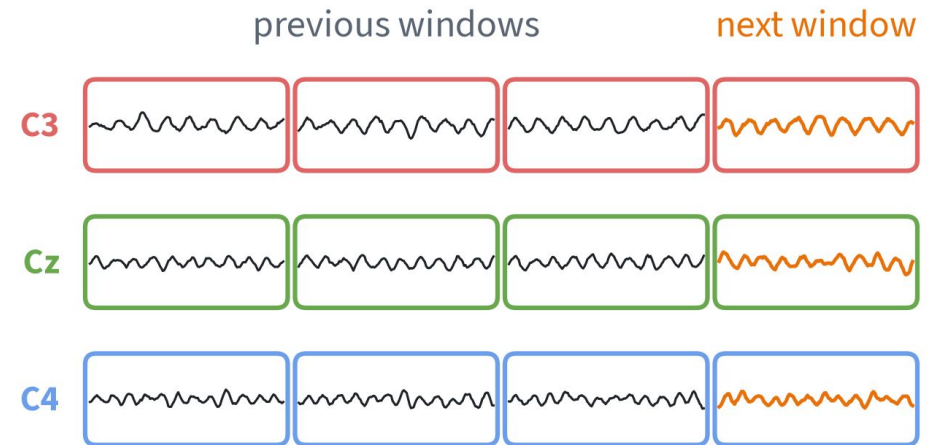
Masked reconstruction: reconstruct hidden windows (REVE)



Four self-supervised objectives

Masked reconstruction: reconstruct hidden windows (REVE)

Next-window prediction: predict the next window (MEG-GPT)

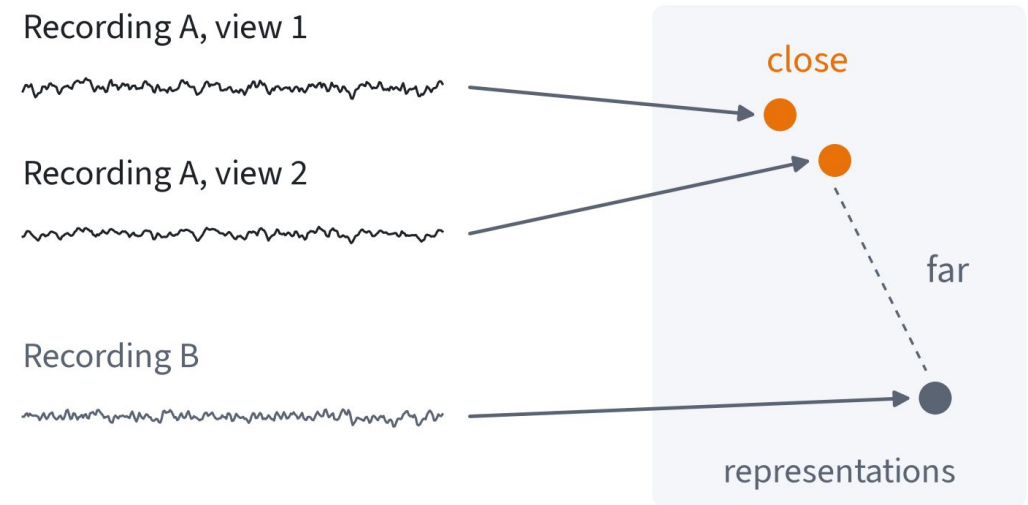


Four self-supervised objectives

Masked reconstruction: reconstruct hidden windows (REVE)

Next-window prediction: predict the next window (MEG-GPT)

Contrastive: two views of one recording get close representations (BIOT)



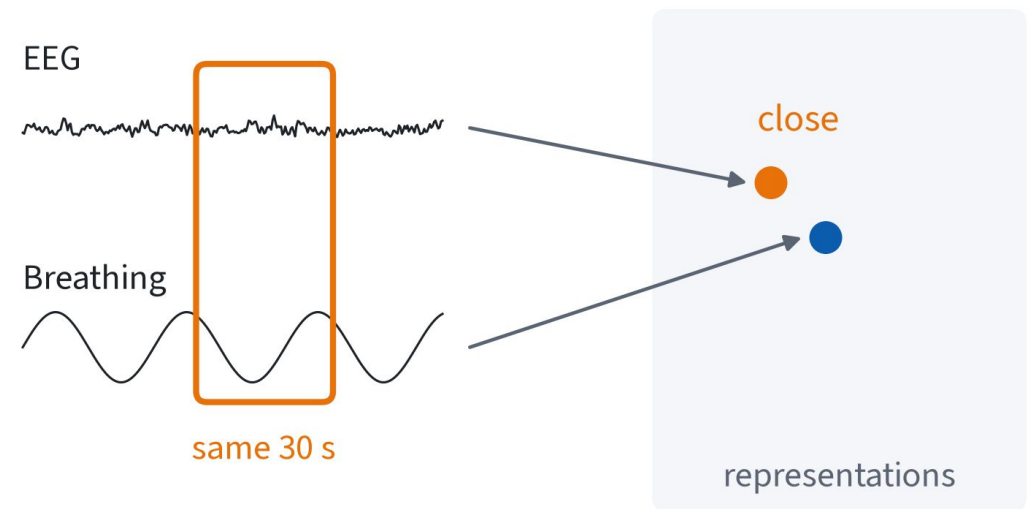
Four self-supervised objectives

Masked reconstruction: reconstruct hidden windows (REVE)

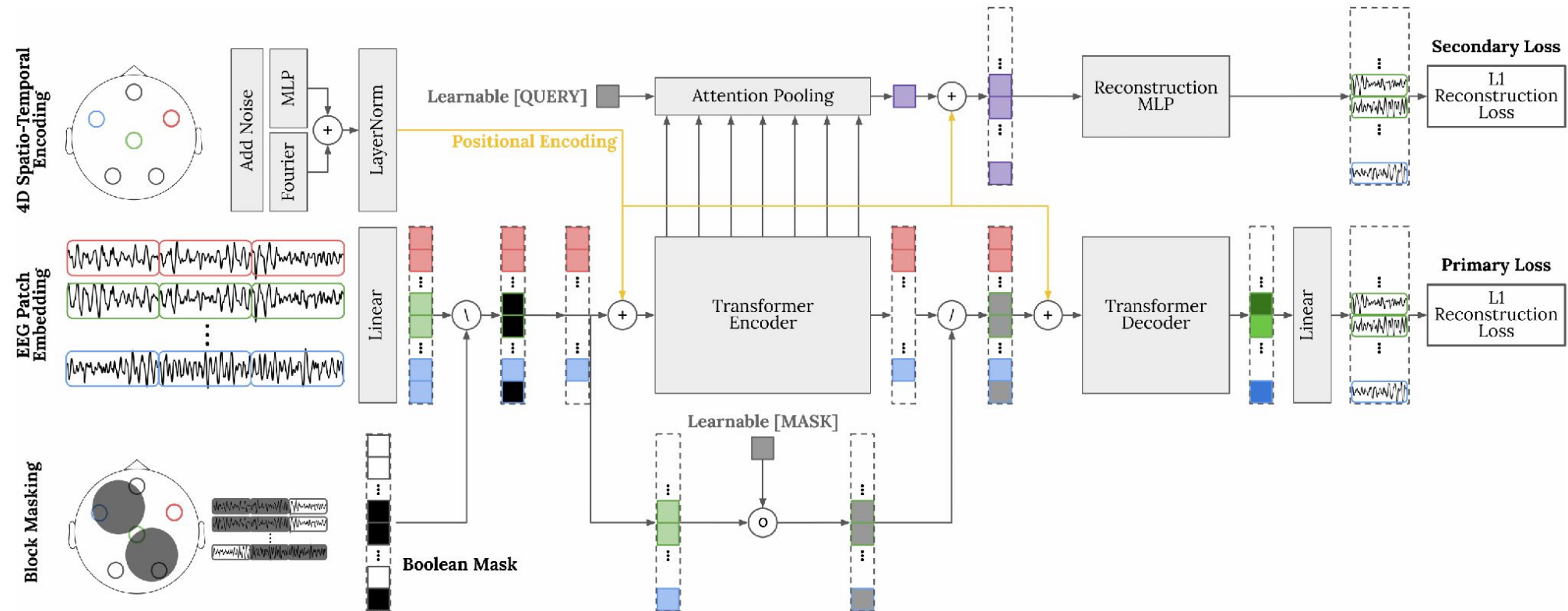
Next-window prediction: predict the next window (MEG-GPT)

Contrastive: two views of one recording get close representations (BIOT)

Cross-modal alignment: match EEG to a signal recorded at the same time (SleepFM)



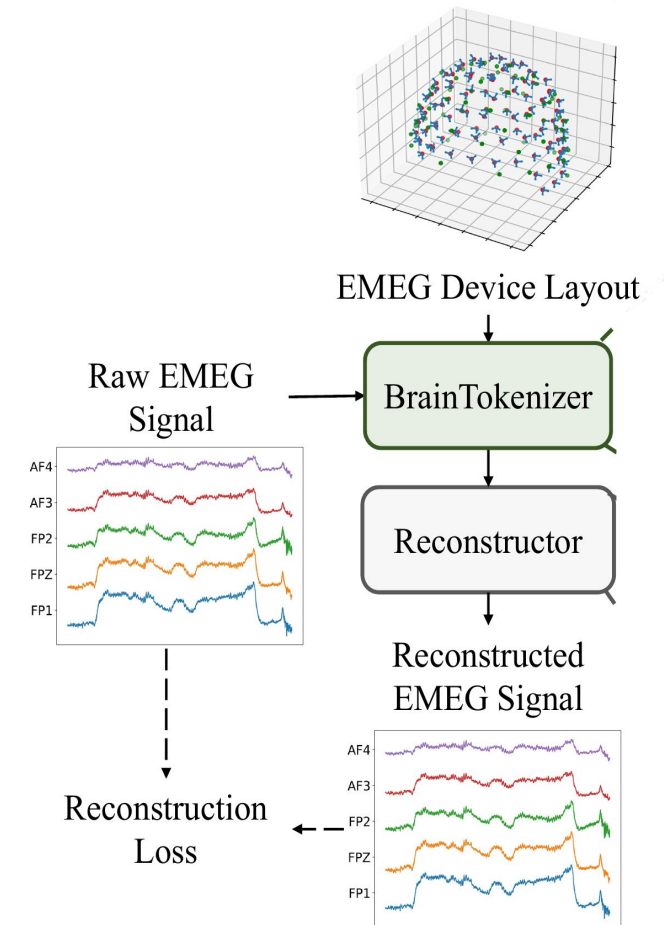
Example: REVE



REVE: pretrained on 60,000 hours of EEG from 92 datasets

Example: BrainOmni

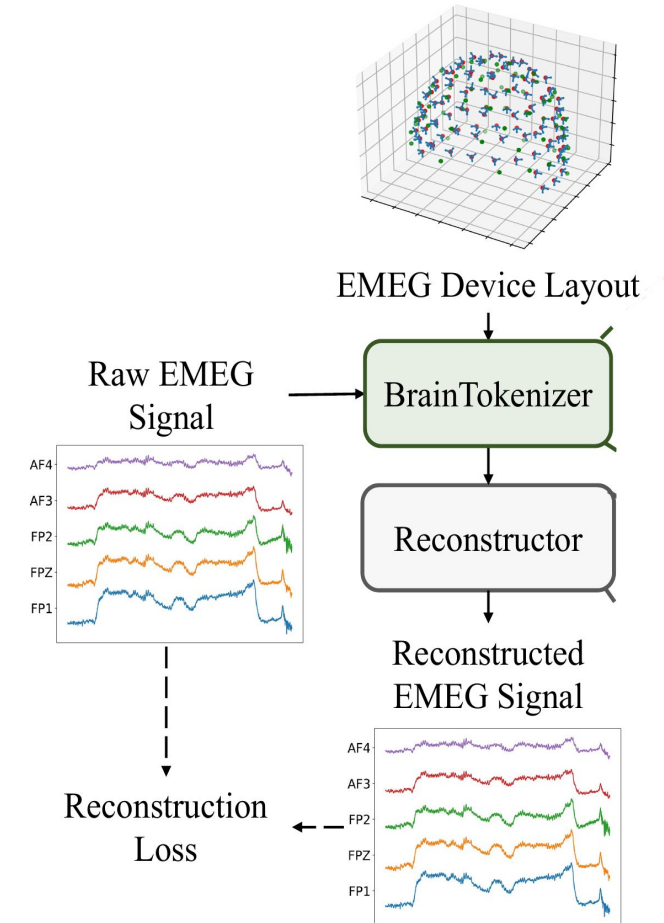
One model for EEG and MEG



Example: BrainOmni

One model for EEG and MEG

Each sensor is described by its **position, orientation and type**

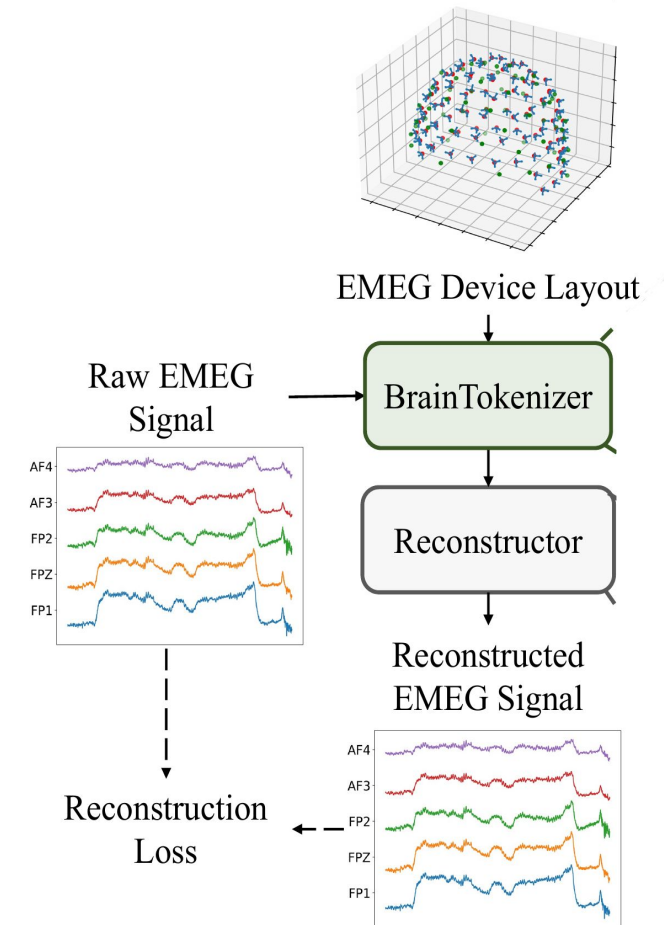


Example: BrainOmni

One model for EEG and MEG

Each sensor is described by its **position, orientation and type**

Pretrained on about **2,000 h of EEG** and **650 h of MEG**



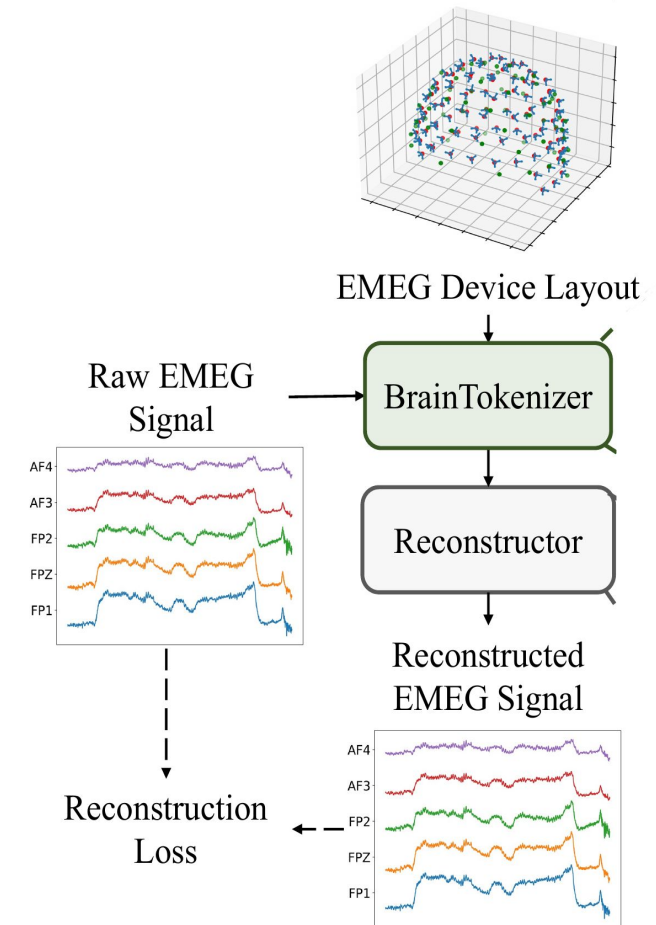
Example: BrainOmni

One model for EEG and MEG

Each sensor is described by its **position, orientation and type**

Pretrained on about **2,000 h of EEG** and **650 h of MEG**

Objective: **mask** tokens and **predict** them



EEG and MEG foundation models since 2021

2021 · BENDR: contrastive pretraining on one clinical EEG corpus

EEG and MEG foundation models since 2021

2021 · BENDR: contrastive pretraining on one clinical EEG corpus

2023 · BIOT: contrastive, inputs with any number of channels

EEG and MEG foundation models since 2021

2021 • BENDR: contrastive pretraining on one clinical EEG corpus

2023 • BIOT: contrastive, inputs with any number of channels

2024 • LaBraM, EEGPT: masked reconstruction across many datasets

EEG and MEG foundation models since 2021

2021 • BENDR: contrastive pretraining on one clinical EEG corpus

2023 • BIOT: contrastive, inputs with any number of channels

2024 • LaBraM, EEGPT: masked reconstruction across many datasets

2025 • CBraMod, REVE: masked reconstruction; REVE adds 3D electrode positions

EEG and MEG foundation models since 2021

2021 • BENDR: contrastive pretraining on one clinical EEG corpus

2023 • BIOT: contrastive, inputs with any number of channels

2024 • LaBraM, EEGPT: masked reconstruction across many datasets

2025 • CBraMod, REVE: masked reconstruction; REVE adds 3D electrode positions

2025 • BrainOmni: one model for EEG and MEG

EEG and MEG foundation models since 2021

2021 • BENDR: contrastive pretraining on one clinical EEG corpus

2023 • BIOT: contrastive, inputs with any number of channels

2024 • LaBraM, EEGPT: masked reconstruction across many datasets

2025 • CBraMod, REVE: masked reconstruction; REVE adds 3D electrode positions

2025 • BrainOmni: one model for EEG and MEG

2025 • MEG-GPT: next-window prediction on MEG

EEG and MEG foundation models since 2021

2021 • BENDR: contrastive pretraining on one clinical EEG corpus

2023 • BIOT: contrastive, inputs with any number of channels

2024 • LaBraM, EEGPT: masked reconstruction across many datasets

2025 • CBraMod, REVE: masked reconstruction; REVE adds 3D electrode positions

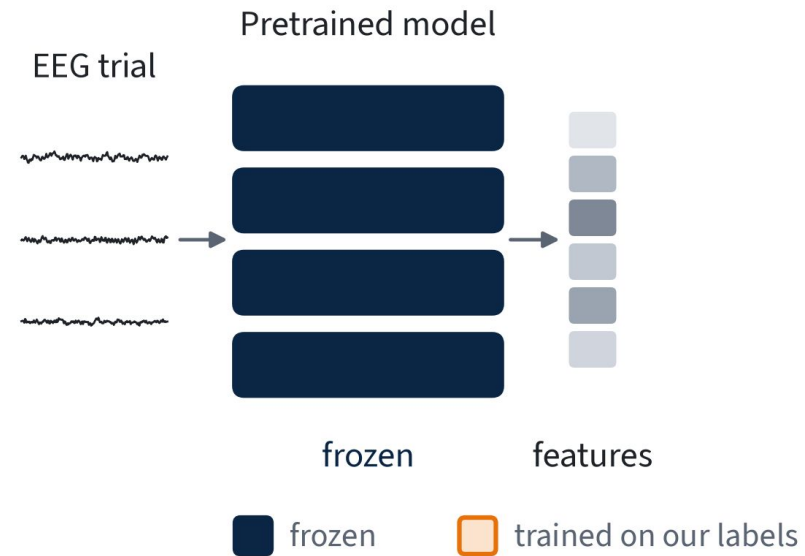
2025 • BrainOmni: one model for EEG and MEG

2025 • MEG-GPT: next-window prediction on MEG

EEG foundation models: **about 10 in 2024, 55 in 2026**

Using a pretrained model

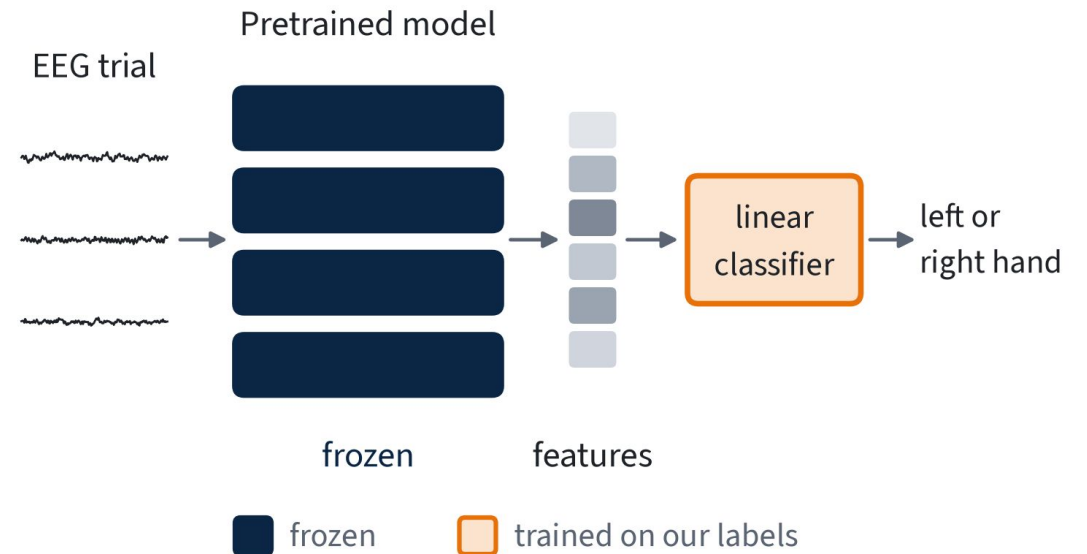
Frozen: the model is not changed; it turns each trial into **features**



Using a pretrained model

Frozen: the model is not changed; it turns each trial into **features**

Linear probe: train only a linear classifier on the frozen features

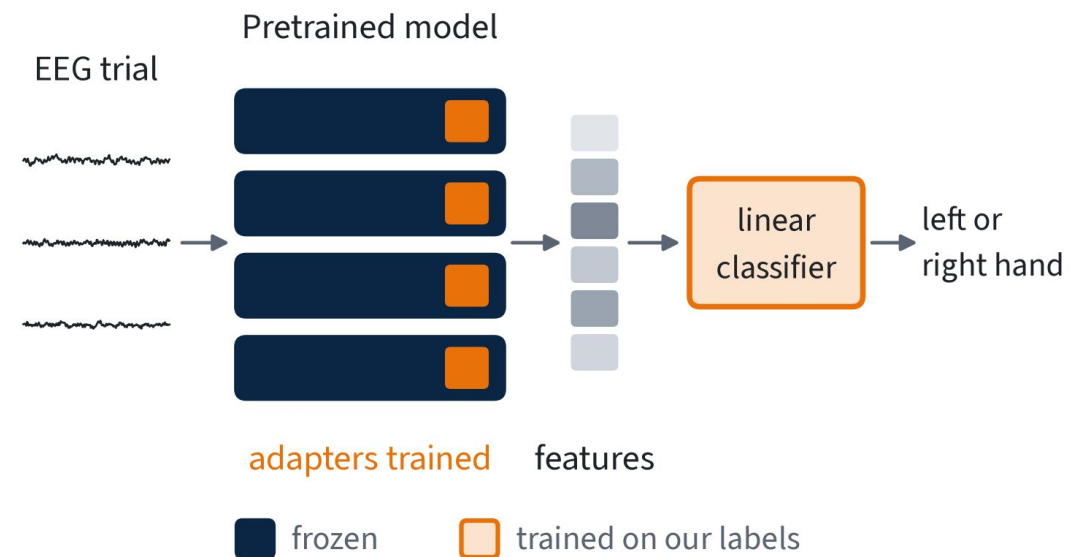


Using a pretrained model

Frozen: the model is not changed; it turns each trial into **features**

Linear probe: train only a linear classifier on the frozen features

Adapters: train small modules added inside the frozen model



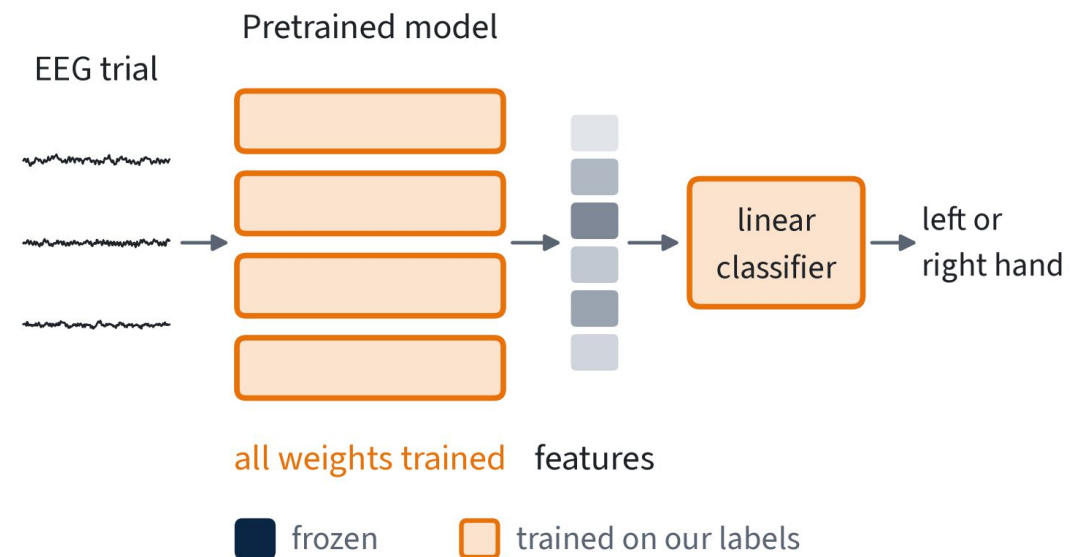
Using a pretrained model

Frozen: the model is not changed; it turns each trial into **features**

Linear probe: train only a linear classifier on the frozen features

Adapters: train small modules added inside the frozen model

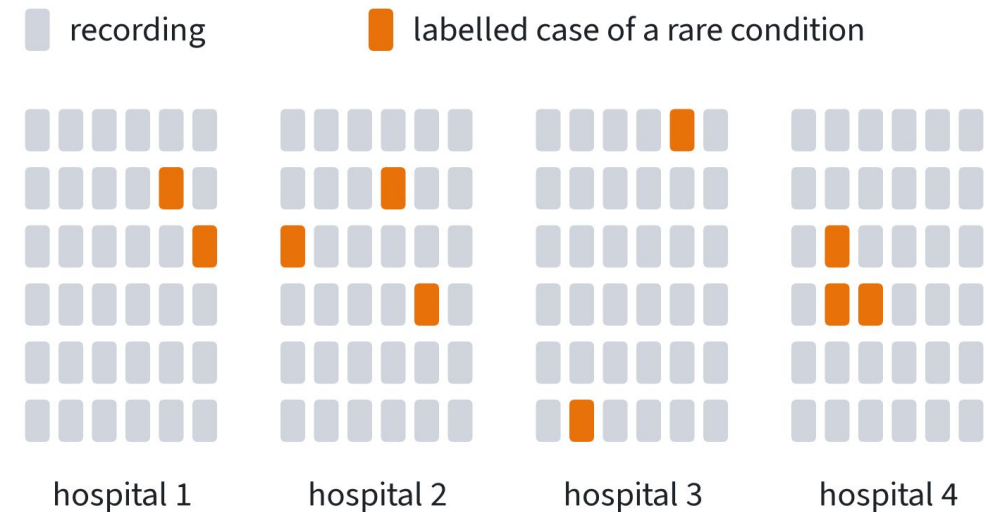
Fine-tuning: update **all** the weights of the model on the new task



What can a pretrained model do that a task-specific model cannot?

Where foundation models could make a difference

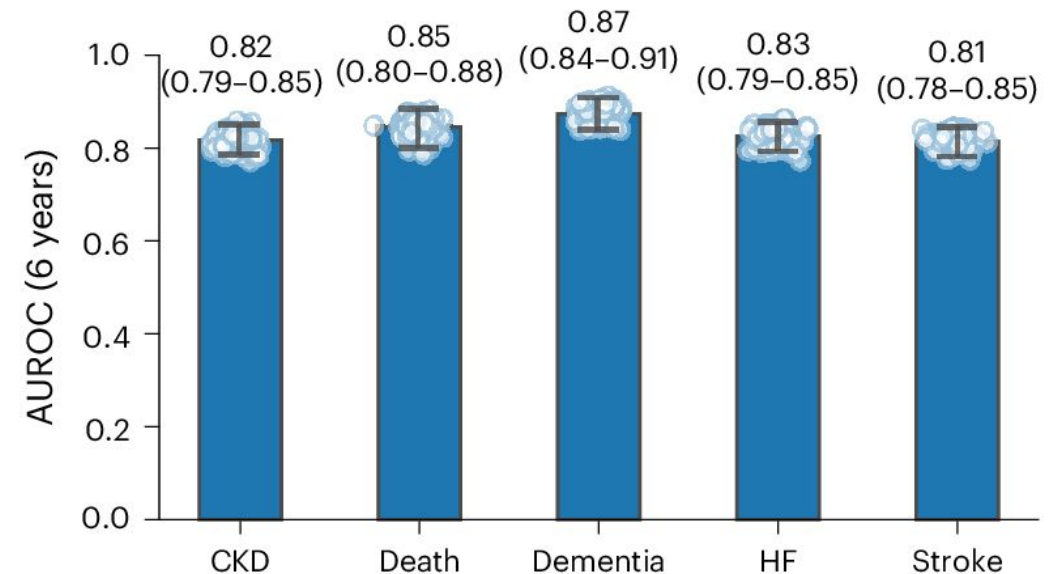
Rare conditions: a few labelled cases per hospital



Where foundation models could make a difference

Rare conditions: a few labelled cases per hospital

Routine recordings: sleep data predicts disease

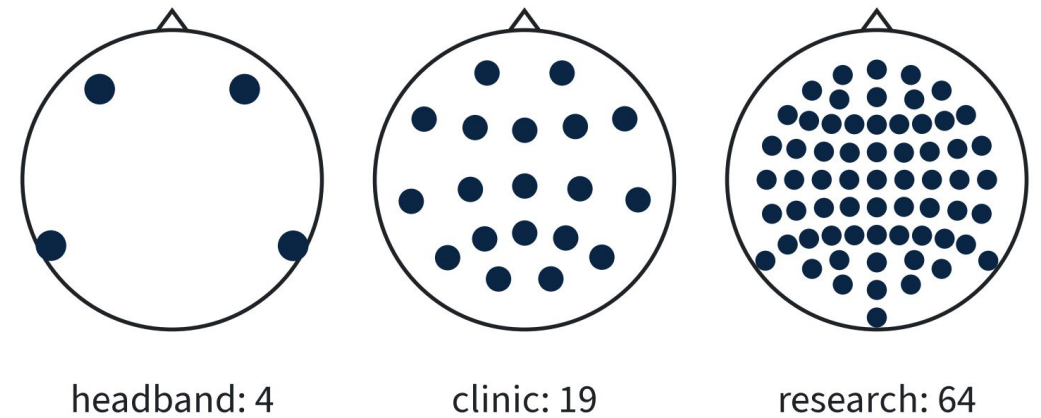


Where foundation models could make a difference

Rare conditions: a few labelled cases per hospital

Routine recordings: sleep data predicts disease

Any device: one model for 4 to 64 electrodes



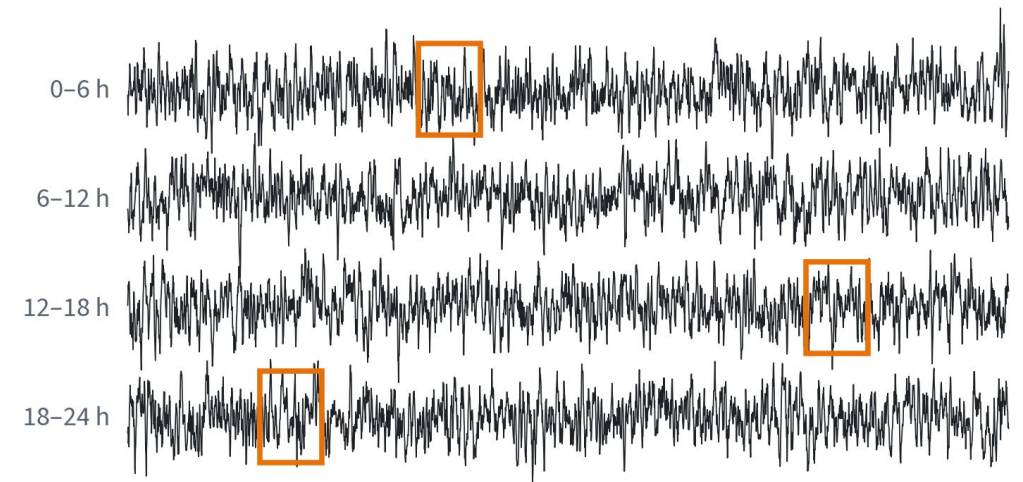
Where foundation models could make a difference

Rare conditions: a few labelled cases per hospital

Routine recordings: sleep data predicts disease

Any device: one model for 4 to 64 electrodes

Long recordings: flag the segments to review



24 hours of EEG, three segments flagged for review

Where foundation models could make a difference

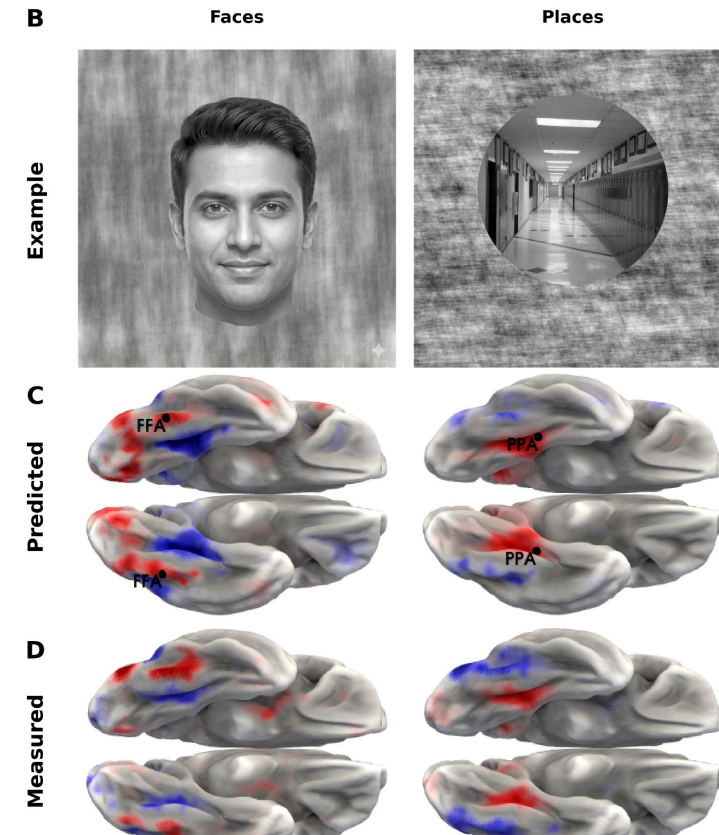
Rare conditions: a few labelled cases per hospital

Routine recordings: sleep data predicts disease

Any device: one model for 4 to 64 electrodes

Long recordings: flag the segments to review

In silico: test a hypothesis before a study



Where foundation models could make a difference

Rare conditions: a few labelled cases per hospital

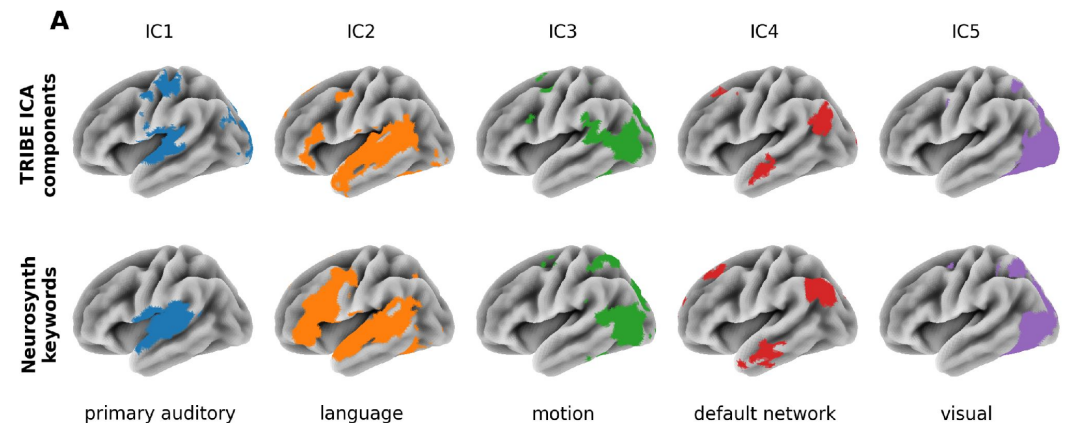
Routine recordings: sleep data predicts disease

Any device: one model for 4 to 64 electrodes

Long recordings: flag the segments to review

In silico: test a hypothesis before a study

Discovery: components match brain networks



Where foundation models could make a difference

Rare conditions: a few labelled cases per hospital

Routine recordings: sleep data predicts disease

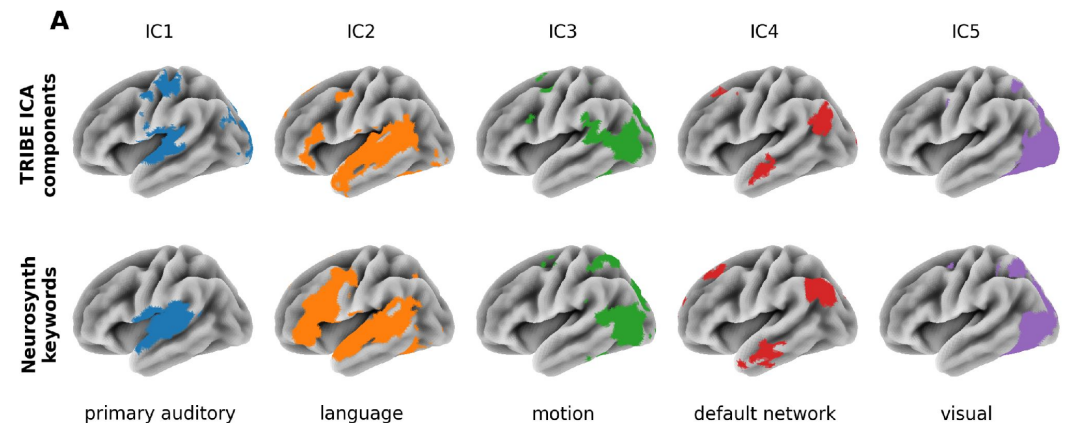
Any device: one model for 4 to 64 electrodes

Long recordings: flag the segments to review

In silico: test a hypothesis before a study

Discovery: components match brain networks

Most results are retrospective

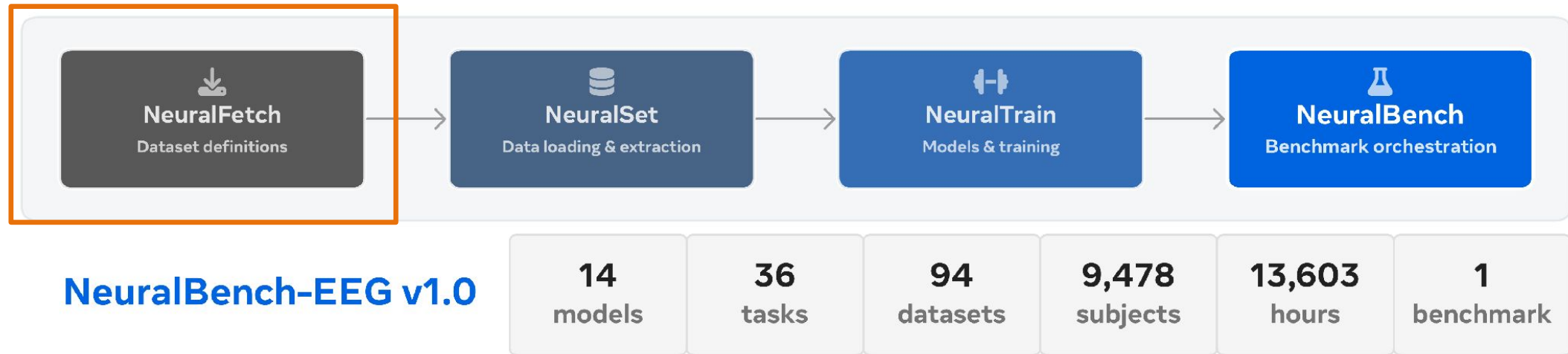


How do we test whether foundation models work?

Twelve EEG benchmarks since July 2025

[Submitted on 14 Jul 2025 (v1), last revised 5 Aug 2025 (this version, v2)] AdaBrain-Bench: Benchmarking Brain Foundation Models for Brain-Computer Interface Applications Jiamin Wu, Zichen Ren, Junyu Wang, Pengyu Zhu, Yonghao Song, Mianxin Liu, Qihao Zheng, Lei Bai, Wanli Ouyang, Chunfeng Song	[Submitted on 25 Aug 2025 (v1), last revised 15 Jun 2026 (this version, v3)] EEG-FM-Bench: A Comprehensive Benchmark for the Systematic Evaluation and Diagnostic Analyses of EEG Foundation Models Wei Xiong, Jiangtong Li, Jie Li, Kun Zhu, Changjun Jiang	[Submitted on 28 Nov 2025] EEG-Bench: A Benchmark for EEG Foundation Models in Clinical Applications Ard Kastrati, Josua Bürki, Jonas Lauer, Cheng Xuan, Raffaele Iaquinto, Roger Wattenhofer
[Submitted on 25 Jan 2026 (v1), last revised 3 Aug 2026 (this version, v3)] EEG-FM-Compass: Progress, Benchmarking, and Future Directions for EEG Foundation Models Dingkun Liu, Yuheng Chen, Zhu Chen, Zhenyao Cui, Yaozhi Wen, Jiayu An, Jingwei Luo, Dongrui Wu	[Submitted on 12 Feb 2026 (v1), last revised 4 Sep 2026 (this version, v2)] Brain4FMs: A Benchmark of Foundation Models for Electrical Brain Signal Fanqi Shen, Enhong Yang, Jiahe Li, Junru Hong, Xiaoran Pan, Zhizhang Yuan, Meng Li, Yang Yang	POSTER Sat, Apr 25, 2026 - 10:30 AM - 1:00 PM -03 Are EEG Foundation Models Worth It? Comparative Evaluation with Traditional Decoders in Diverse BCI Tasks Liuyin Yang · Qiang Sun · Ang Li · Marc Van Hulle
[Submitted on 25 Apr 2026] Channel Adaptation for EEG Foundation Models: A Systematic Benchmark Across Architectures, Tasks, and Training Regimes Kuntal Kokate, Bruno Aristimunha, Dung Truong, Arnaud Delorme	[Submitted on 8 May 2026] NeuralBench: A Unifying Framework to Benchmark NeuroAI Models Hubert Banville, Stéphane d'Ascoli, Simon Dahan, Jérémy Rapin, Marlène Careil, Yann Bencherit, Jarod Lévy, Saarang Panchavati, Antoine Ratouchniak, Mingfang (Lucy)Zhang, Elisa Cascardi, Katelyn Begany, Teon Brooks, Jean-Rémi King	[Submitted on 14 May 2026 (v1), last revised 30 Sep 2026 (this version, v2)] NeuroAtlas: Benchmarking Foundation Models for Clinical EEG and Brain-Computer Interfaces Konstantinos Kontras, Trui Osselaer, Stylianos G. Mouslech, Angeliki-Ilektra Karaïskou, Guido Gagliardi, Thomas Strypsteen, Mohammad Hossein Badiel, Anku Rani, Maarten Vanmarcke, Miguel Bhagubai, Chanakya Ekbote, Jaedong Hwang, Christos Chatzichristos, Paul Pu Liang, Maarten De Vos
[Submitted on 30 May 2026] OmniEEG-Bench: A Standardized Evaluation Benchmark for EEG Foundation Models Ziling Lu, Zongsheng Li, Xinke Shen, Kexin Lou, Yingyue Xin, Xiaoqi Chen, Shinan Wang, Xiang Chen, Jiahao Fan, Chenyu Huang, Xin Xu, Zhoujie Hou, Chen Wei, Quanying Liu	[Submitted on 4 Aug 2026 (v1), last revised 19 Aug 2026 (this version, v2)] BrainBench: Benchmarking Large Language Models for Comprehensive EEG Understanding Yangxuan Zhou, Yuning Chen, Chen Wu, Jiquan Wang, Shijian Li, Gang Pan, Sha Zhao	[Submitted on 26 Sep 2026] Benchmarking EEG Foundation Models at Scale: Lessons from 20,000 Evaluations Zhige Chen, Shu Peng, Chengxuan Qin, Rui Liu, Rui Yang, Kay Chen Tan, Jibin Wu

Meta's open tools for brain data



NeuralFetch: 130+ public datasets (EEG, MEG, fMRI) with downloaders

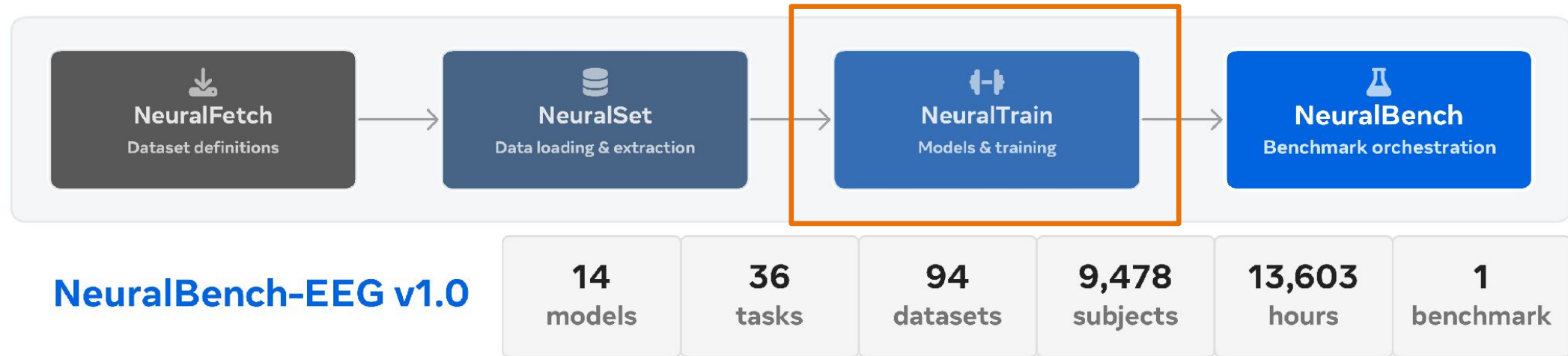
Meta's open tools for brain data



NeuralFetch: 130+ public datasets (EEG, MEG, fMRI) with downloaders

NeuralSet: turns recordings and events into model-ready segments

Meta's open tools for brain data



NeuralFetch: 130+ public datasets (EEG, MEG, fMRI) with downloaders

NeuralSet: turns recordings and events into model-ready segments

NeuralTrain: model zoo, losses and metrics

Meta's open tools for brain data



NeuralFetch: 130+ public datasets (EEG, MEG, fMRI) with downloaders

NeuralSet: turns recordings and events into model-ready segments

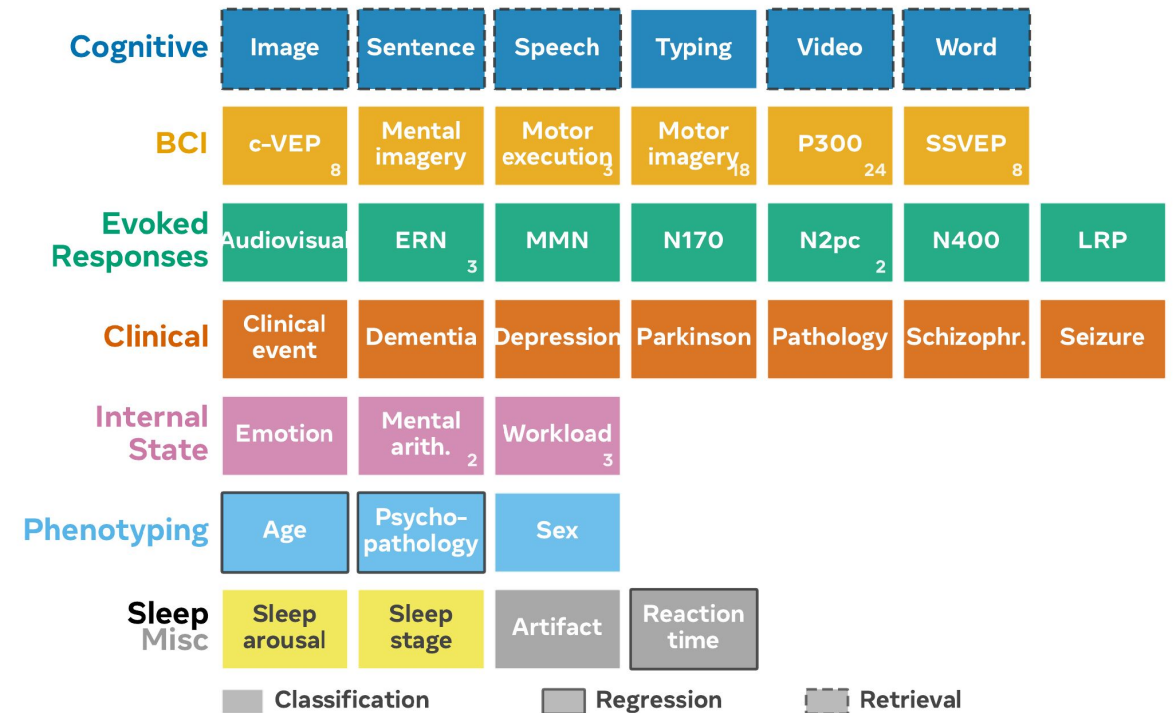
NeuralTrain: model zoo, losses and metrics

NeuralBench: runs every model on every task with one command

NeuralBench-EEG: 36 tasks, one recipe

8 categories: cognitive, BCI, evoked responses, clinical, sleep, ...

A Downstream tasks

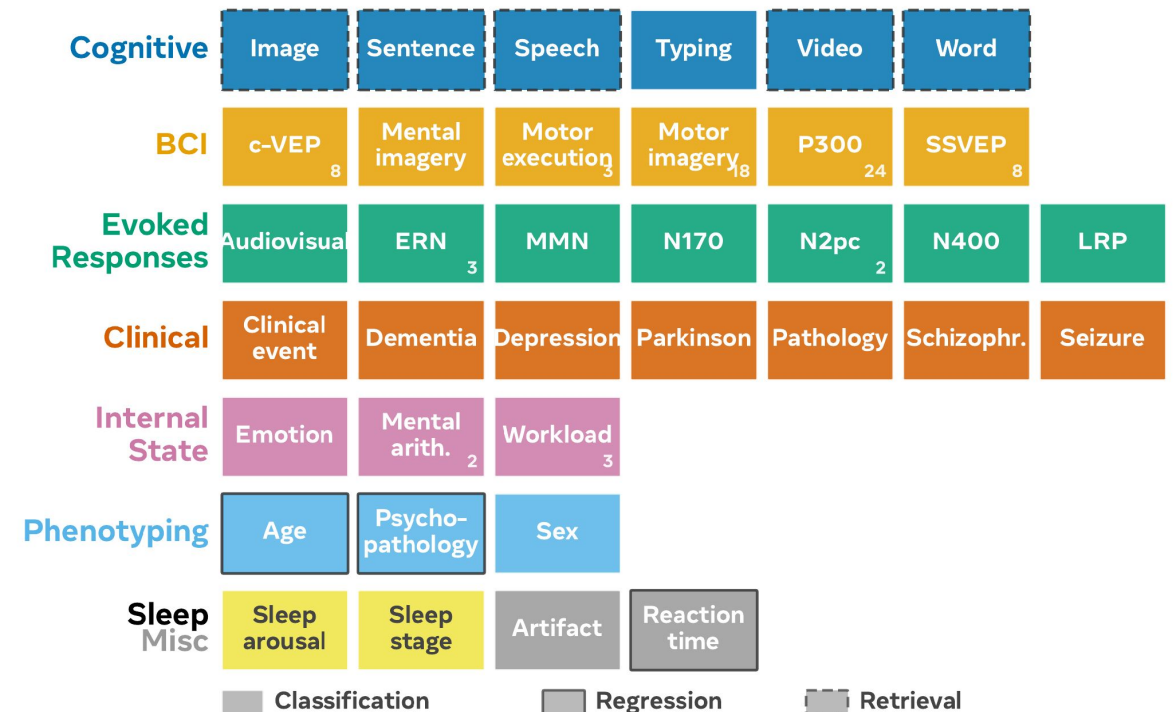


NeuralBench-EEG: 36 tasks, one recipe

8 categories: cognitive, BCI, evoked responses, clinical, sleep, ...

Same recipe for every model: fixed training settings, linear head

A Downstream tasks



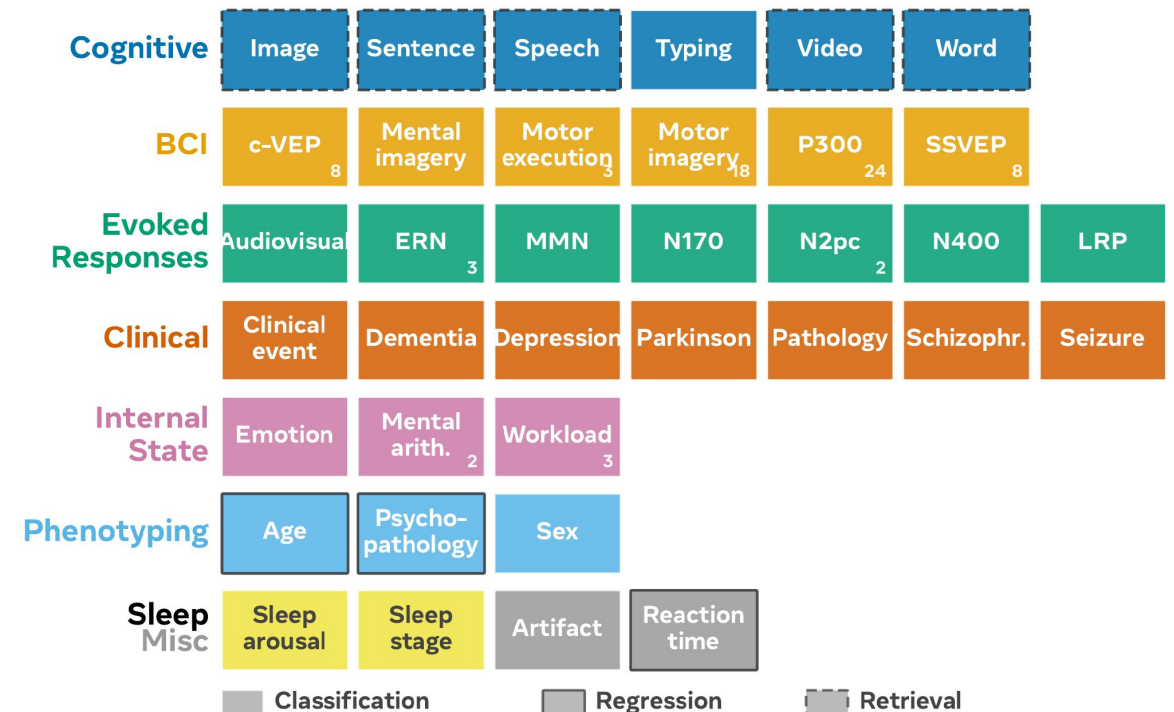
NeuralBench-EEG: 36 tasks, one recipe

8 categories: cognitive, BCI, evoked responses, clinical, sleep, ...

Same recipe for every model: fixed training settings, linear head

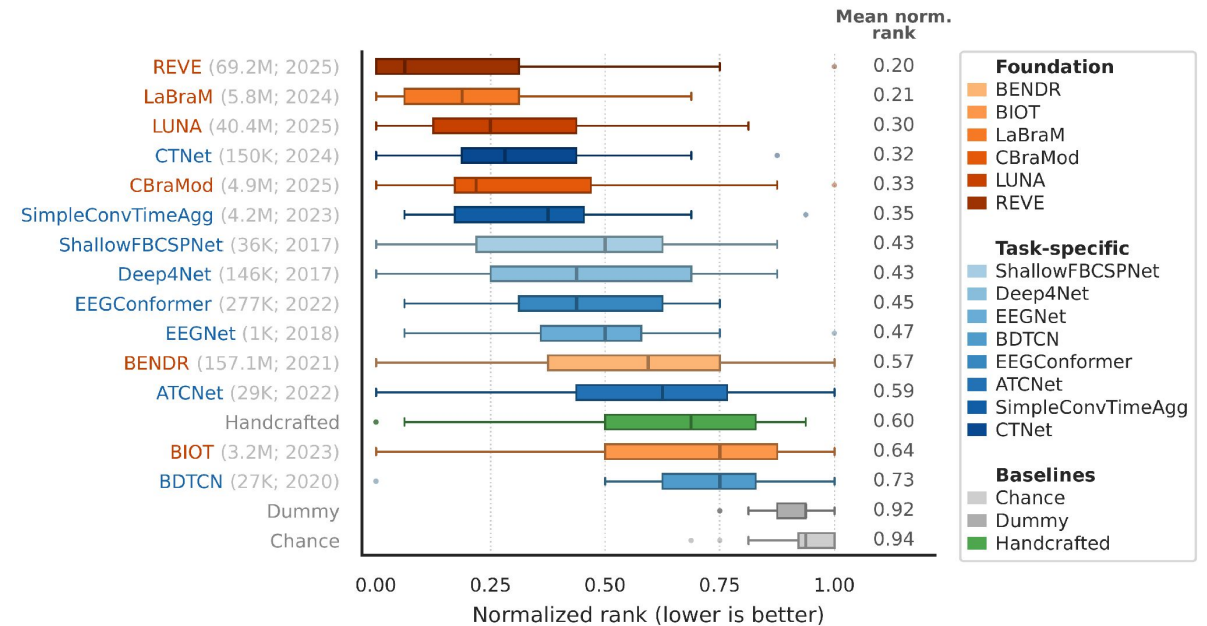
Also **MEG and fMRI** tasks in the same framework

A Downstream tasks



Does it work?

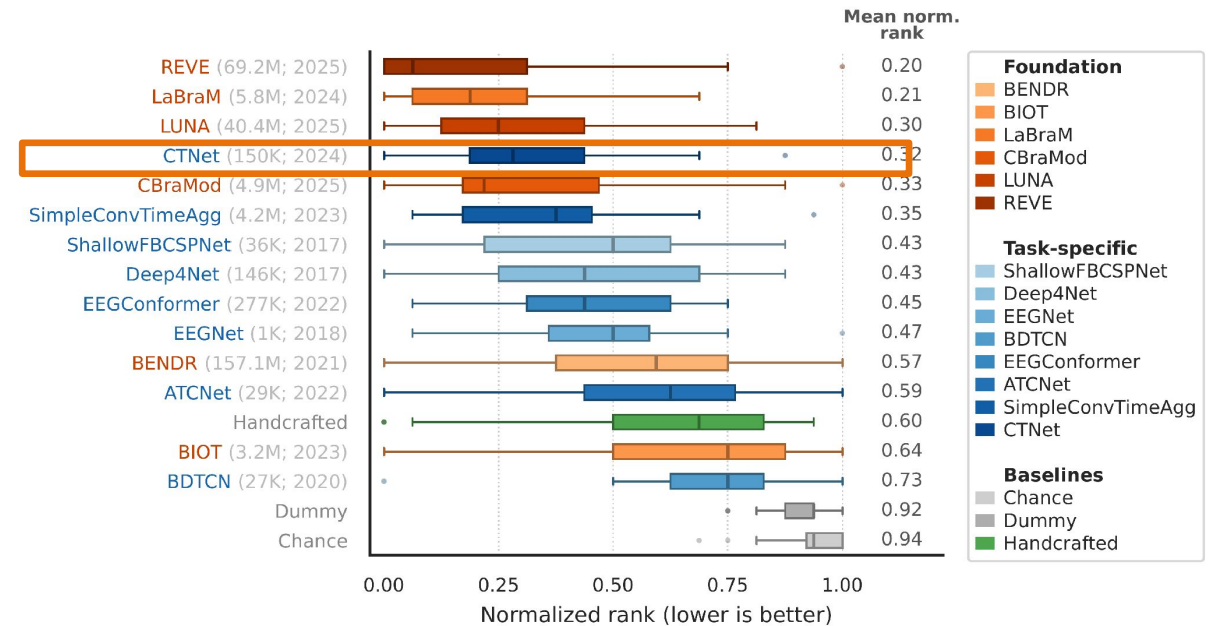
Foundation models rank first: REVE, LaBraM, LUNA



Does it work?

Foundation models rank first: REVE, LaBraM, LUNA

A task-specific network, **CTNet**, is fourth with 0.2% of REVE's parameters

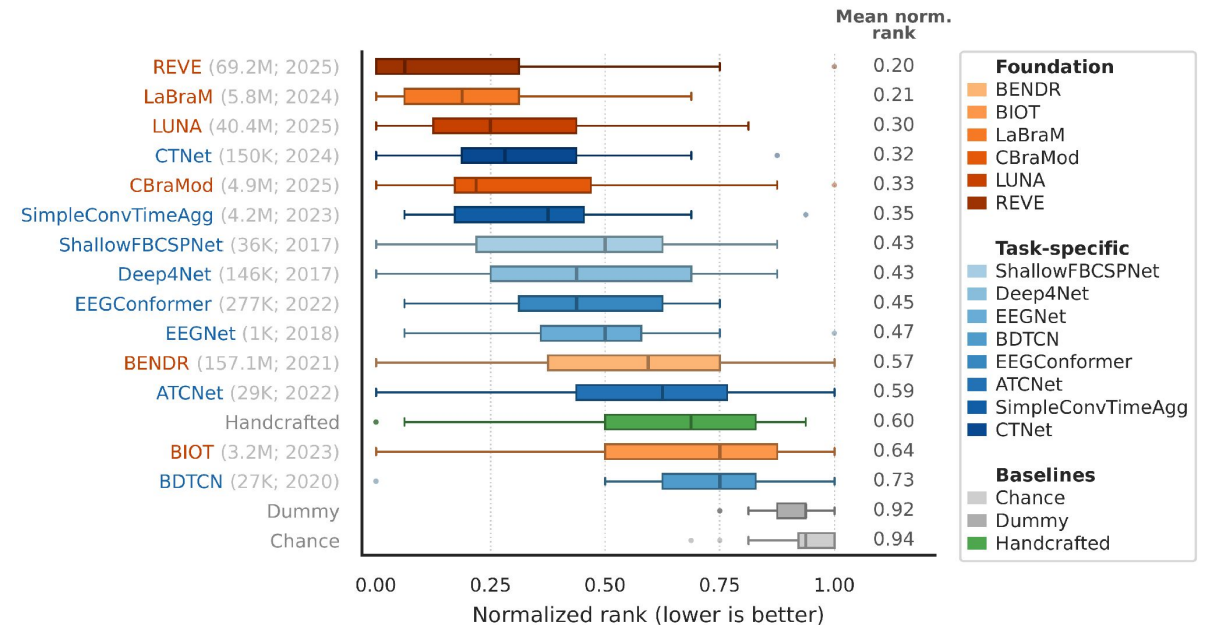


Does it work?

Foundation models rank first: REVE, LaBraM, LUNA

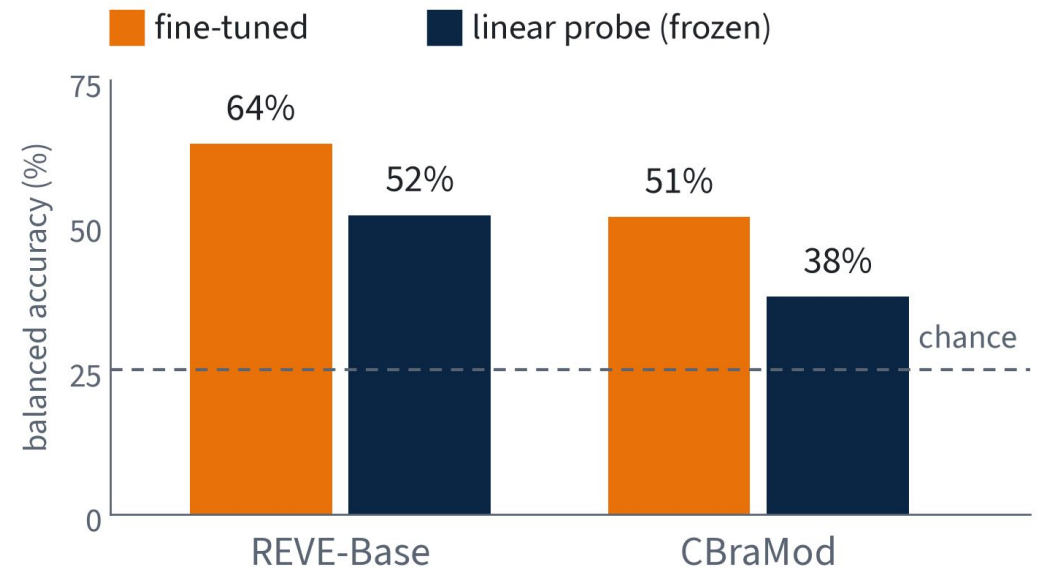
A task-specific network, **CTNet**, is fourth with 0.2% of REVE's parameters

Mixed on evoked-response and clinical tasks; cognitive decoding is hard for all models



The adaptation mode changes the result

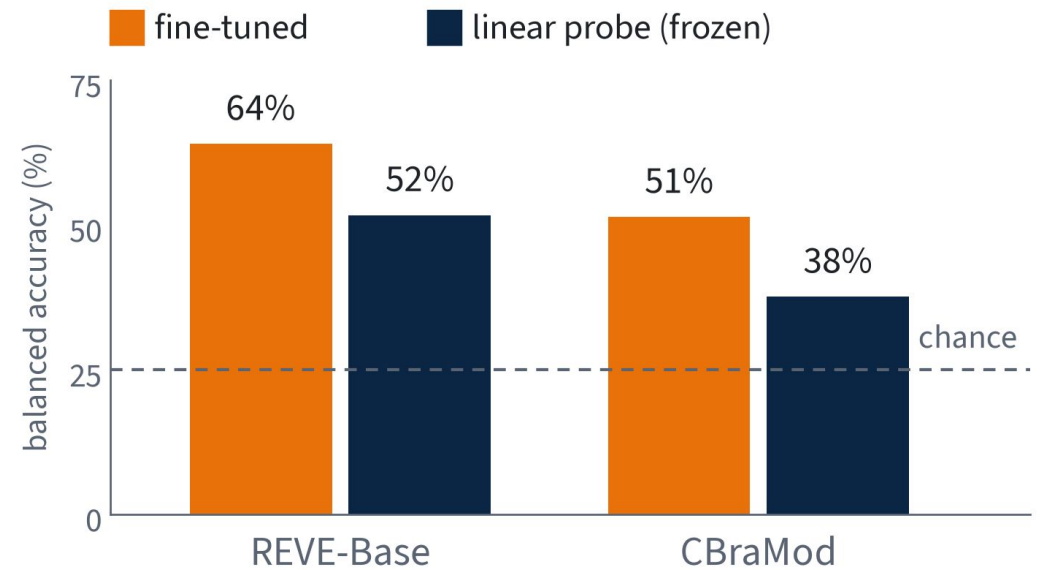
Same models, same dataset: **BCI IV 2a**, four classes



The adaptation mode changes the result

Same models, same dataset: **BCI IV 2a**, four classes

A linear probe can make a model look **much weaker**

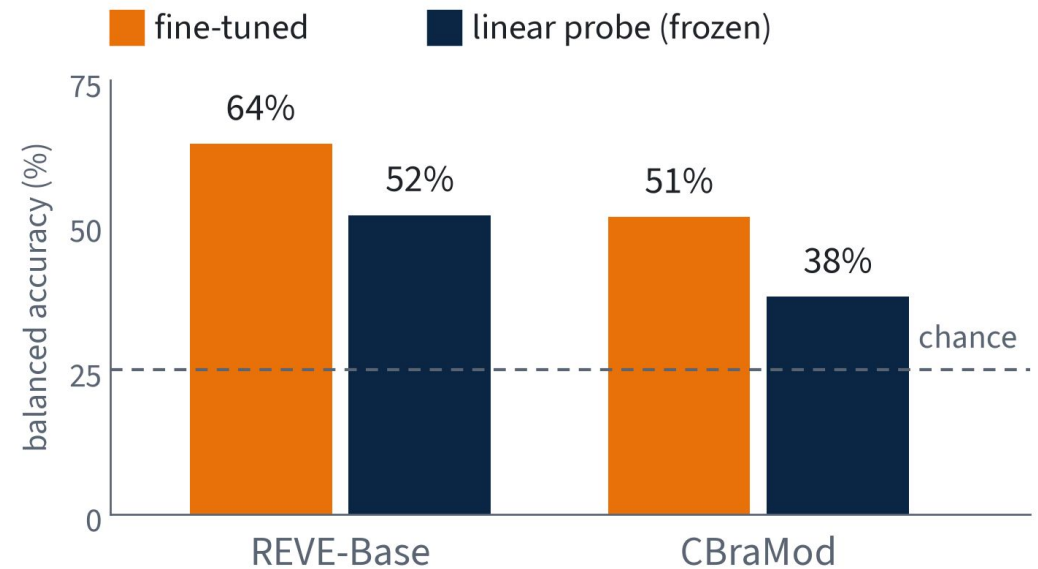


The adaptation mode changes the result

Same models, same dataset: **BCI IV 2a**, four classes

A linear probe can make a model look **much weaker**

Averaging over channels before the probe:
REVE 52% to 43%



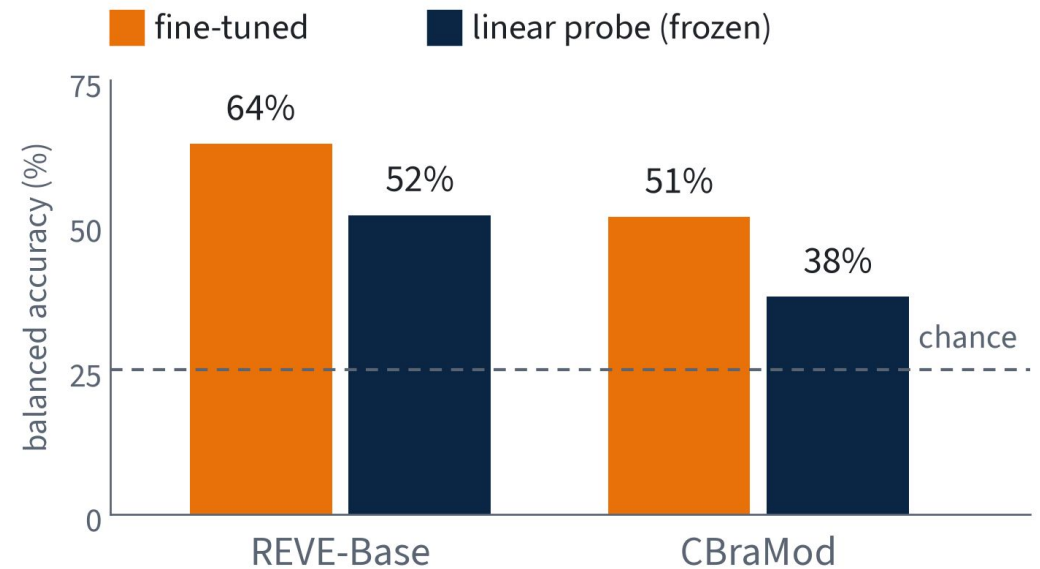
The adaptation mode changes the result

Same models, same dataset: **BCI IV 2a**, four classes

A linear probe can make a model look **much weaker**

Averaging over channels before the probe:
REVE 52% to 43%

We run both: **probe** in 03, **fine-tuning** in 04



Data diversity matters more than model size

 Search Submit Donate Log In

Computer Science > Machine Learning

[Submitted on 25 Jan 2026 (v1), last revised 3 Aug 2026 (this version, v3)]

EEG-FM-Compass: Progress, Benchmarking, and Future Directions for EEG Foundation Models

[Dingkun Liu](#), [Yuheng Chen](#), [Zhu Chen](#), [Zhenyao Cui](#), [Yaozhi Wen](#), [Jiayu An](#), [Jingwei Luo](#), [Dongrui Wu](#)

Electroencephalography (EEG) foundation models (FMs) have recently emerged as a promising paradigm for brain-computer interfaces, aiming to learn transferable neural representations from large-scale heterogeneous recordings. Despite rapid progress, a fair and comprehensive comparison of existing EEG FMs is still lacking, owing to inconsistent pre-training objectives, preprocessing choices, and downstream evaluation protocols. To fill this gap, we present EEG-FM-Compass. We first review 55 representative models and organize their design choices into a unified taxonomic framework including data standardization, model architectures, and self-supervised pre-training strategies. We then evaluate 12 open source FMs and competitive specialist baselines across 13 EEG datasets spanning nine brain-computer interface paradigms. Emphasizing real-world deployments, we consider both cross-subject generalization under a leave-one-subject-out protocol and rapid calibration under a within-subject few-shot setting. We further compare full-parameter fine-tuning with linear probing to assess the transferability of pre-trained representations, and examine the relationship between model scale and downstream performance. Our results indicate that: 1) linear probing is frequently insufficient; 2) specialist models trained from scratch remain competitive across many tasks; and 3) larger FMs do not necessarily yield better generalization performance under current data regimes and training practices.

Access Paper:
[View PDF](#)
[HTML \(experimental\)](#)
[TeX Source](#)
[view license](#)

Current browse context:
cs.LG
< [prev](#) | [next](#) >
[new](#) | [recent](#) | [2026-01](#)
Change to browse by:
[cs](#)
[cs.CV](#)

References & Citations
[NASA ADS](#)
[Google Scholar](#)
[Semantic Scholar](#)

[Export BibTeX Citation](#)

Bookmark

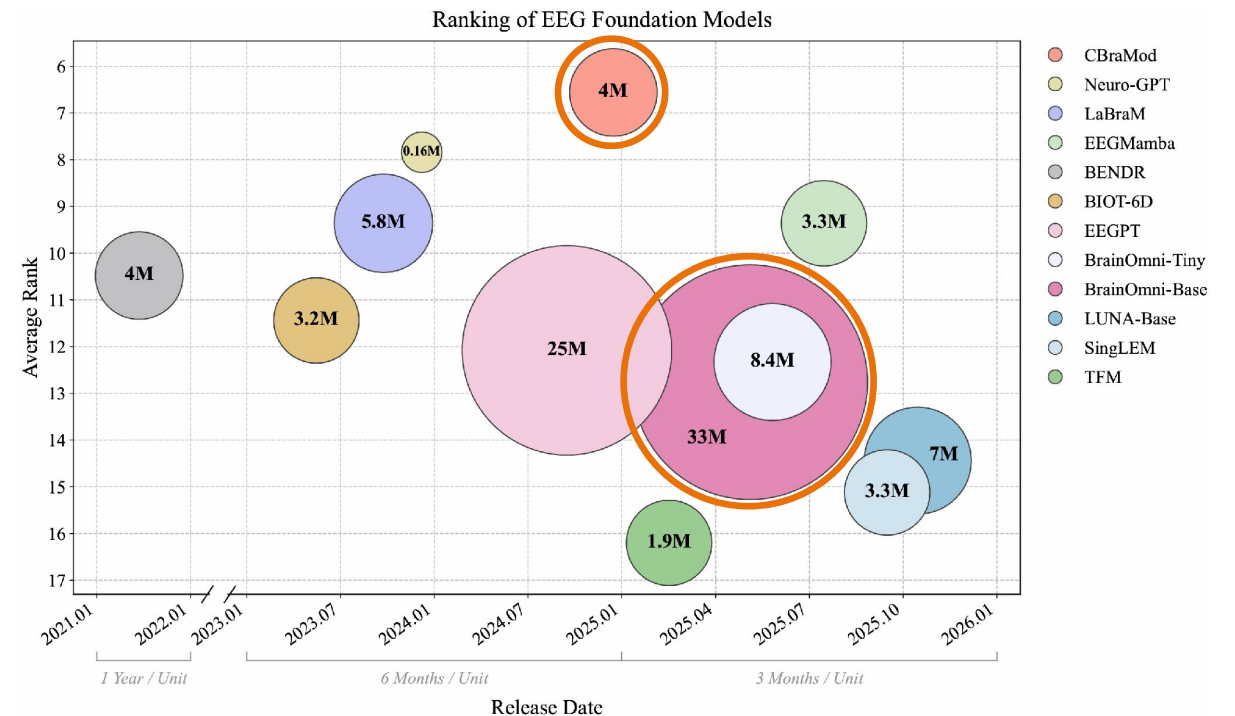
Data diversity matters more than model size

Transfer improves with more
participants, tasks and devices

Data diversity matters more than model size

Transfer improves with more
participants, tasks and devices

Larger EEG models have not consistently
performed better



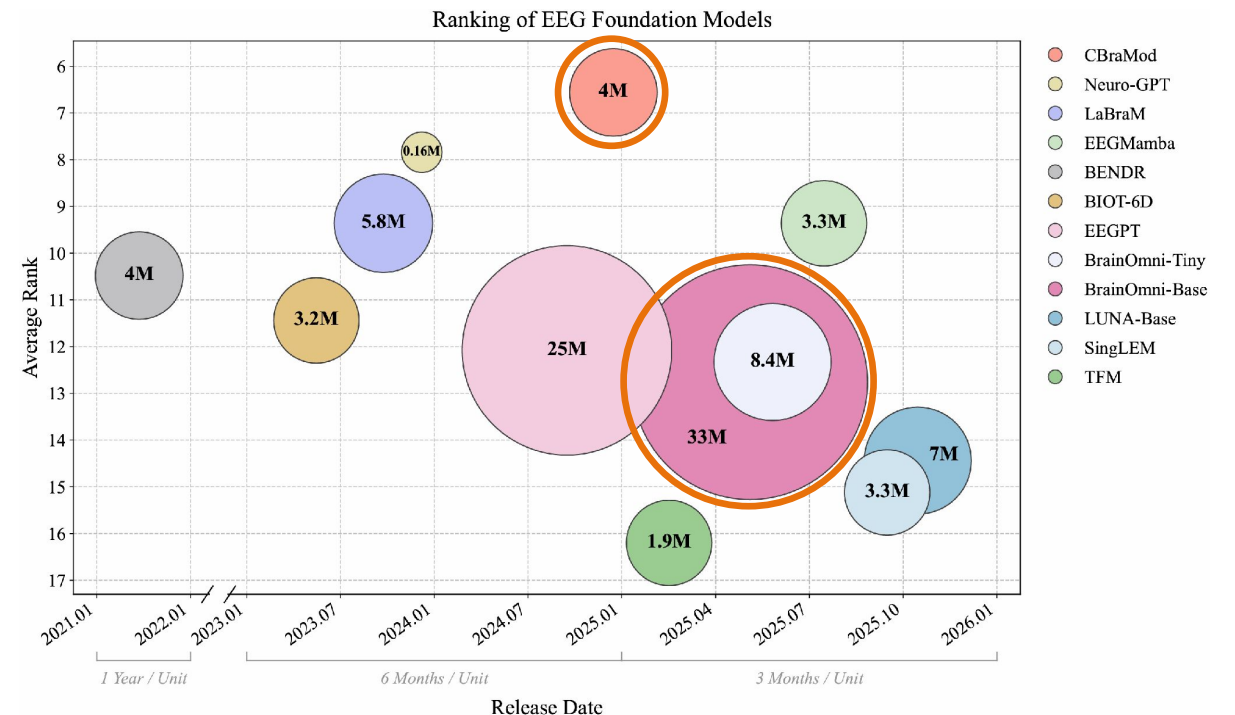
Higher = better average rank. Bubble size = parameters.

Data diversity matters more than model size

Transfer improves with more
participants, tasks and devices

Larger EEG models have not consistently
performed better

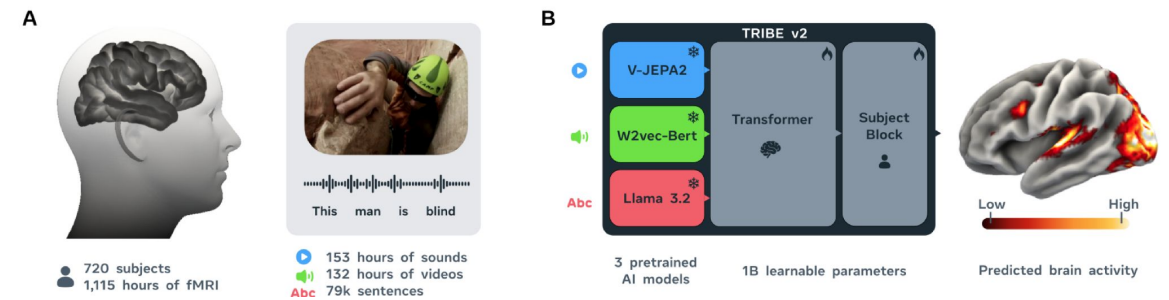
In MEG, **dataset selection** outperformed
pooling all available data



Higher = better average rank. Bubble size = parameters.

Current scene: large brain-data projects

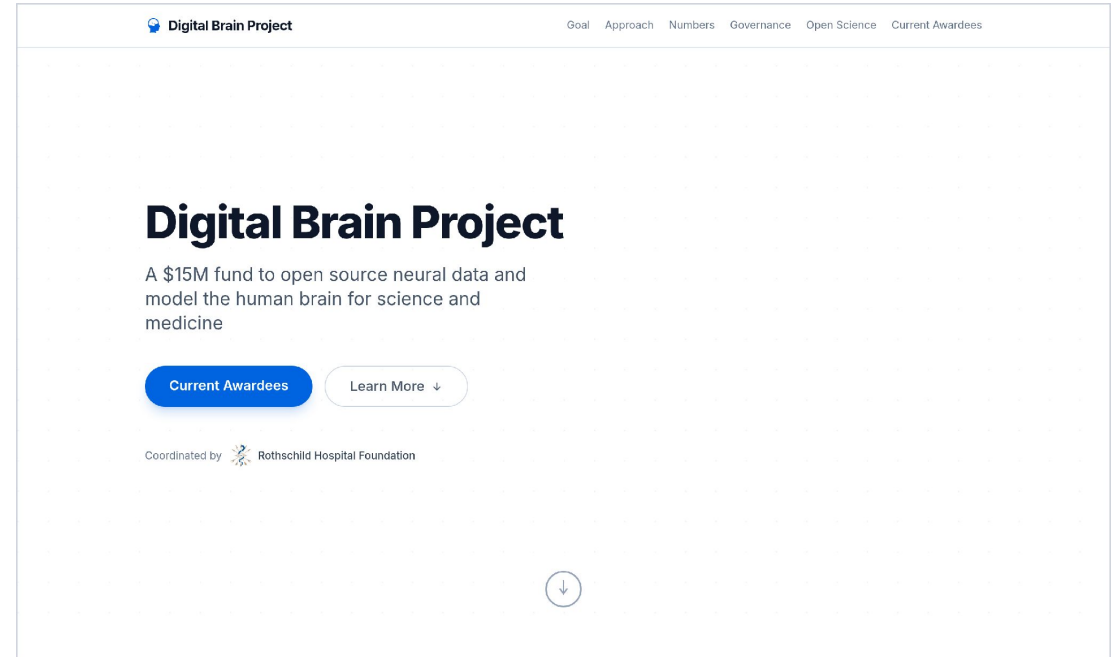
Meta, Brain & AI: TRIBE v2 predicts fMRI responses to video, sound and text



Current scene: large brain-data projects

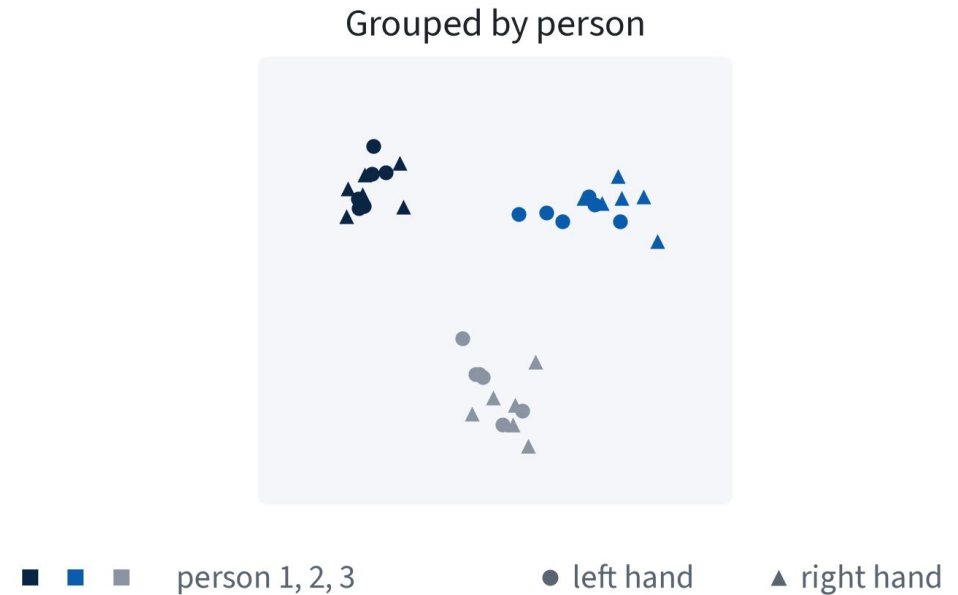
Meta, Brain & AI: TRIBE v2 predicts fMRI responses to video, sound and text

Digital Brain Project: \$15M to record and share brain activity during games



Pitfall: learning who, not what

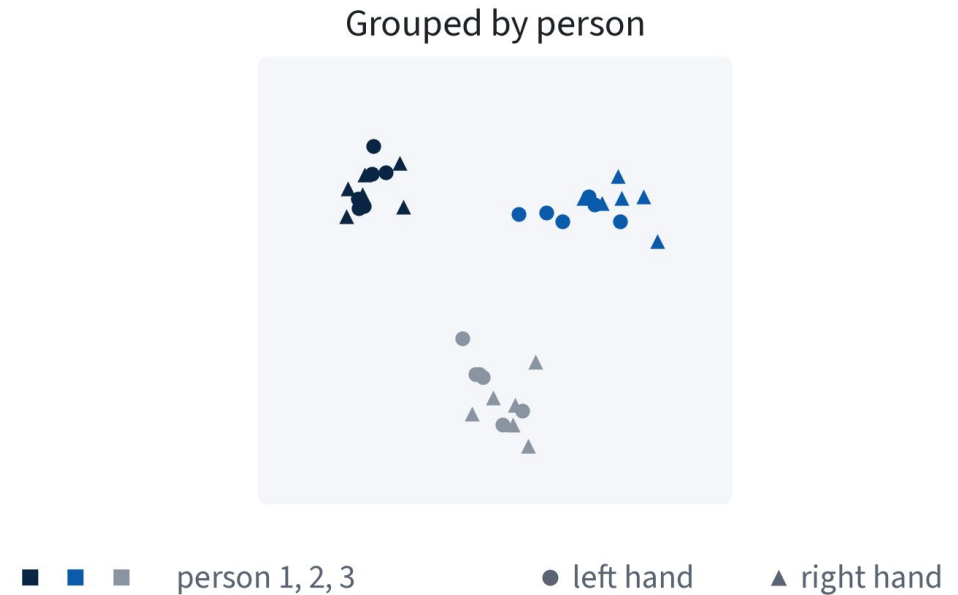
Each person's EEG has an **individual signature**



Pitfall: learning who, not what

Each person's EEG has an **individual signature**

A model can score well by **recognizing the person** or the site

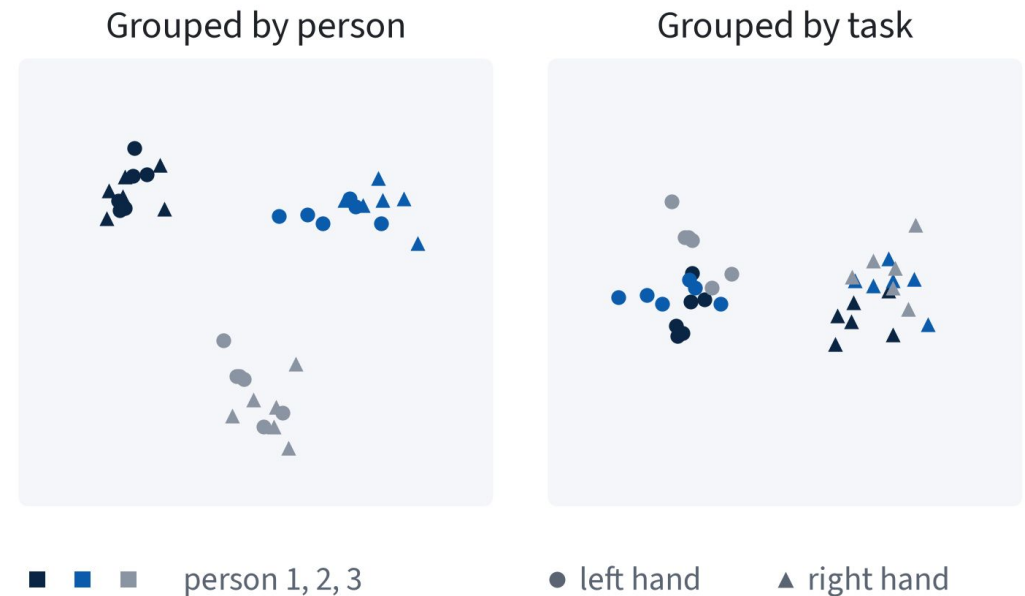


Pitfall: learning who, not what

Each person's EEG has an **individual signature**

A model can score well by **recognizing the person** or the site

Same people in training and test: **inflated scores**



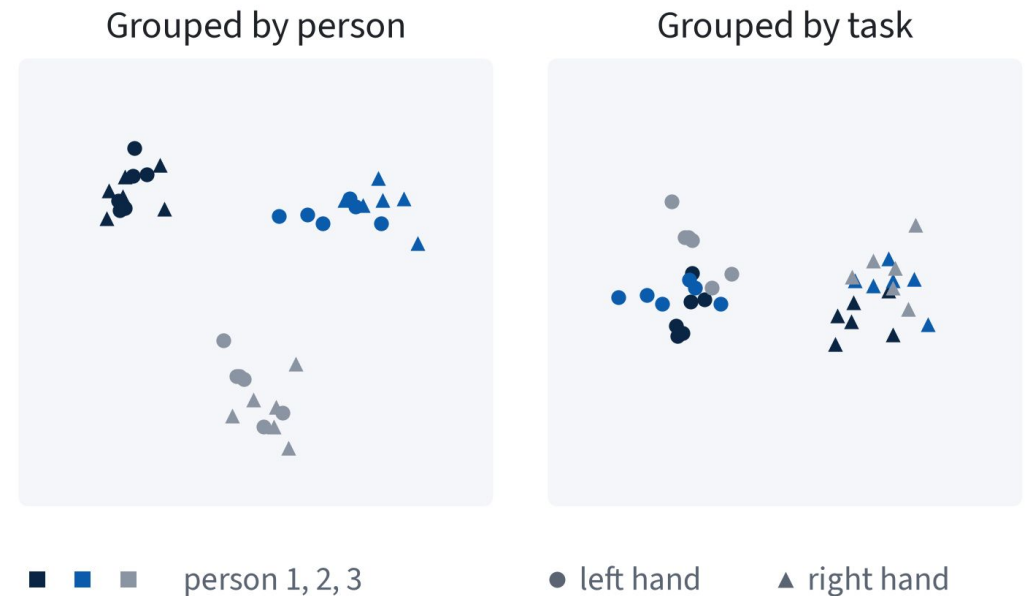
Pitfall: learning who, not what

Each person's EEG has an **individual signature**

A model can score well by **recognizing the person** or the site

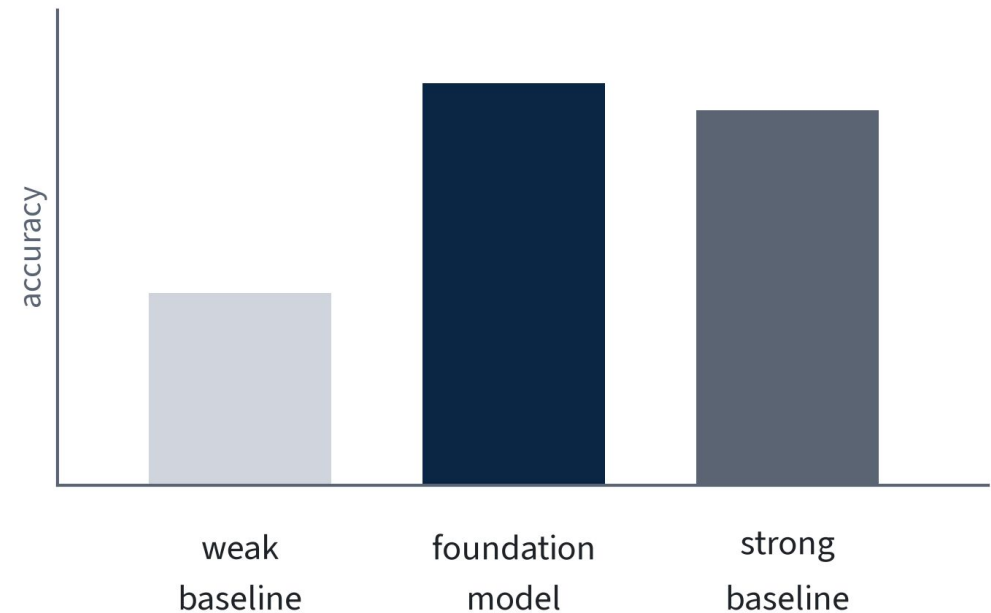
Same people in training and test: **inflated scores**

Test on **new people or new sessions**; we test on a new session



More ways results get inflated

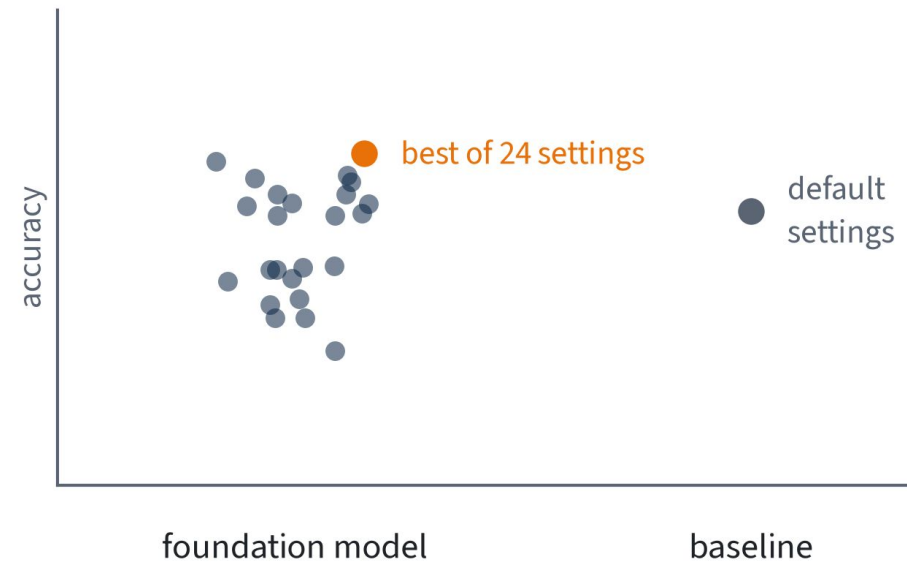
A **weak baseline** makes any model look strong



More ways results get inflated

A **weak baseline** makes any model look strong

One model **tuned carefully**, the other left at its defaults

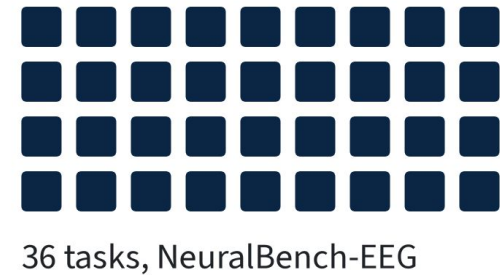


More ways results get inflated

A **weak baseline** makes any model look strong

One model **tuned carefully**, the other left at its defaults

Few or nearly solved test tasks



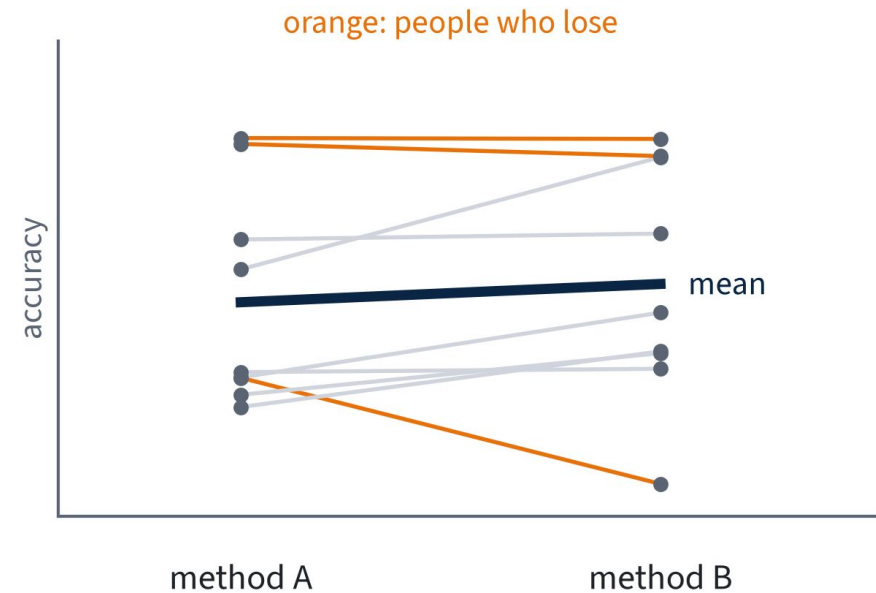
More ways results get inflated

A **weak baseline** makes any model look strong

One model **tuned carefully**, the other left at its defaults

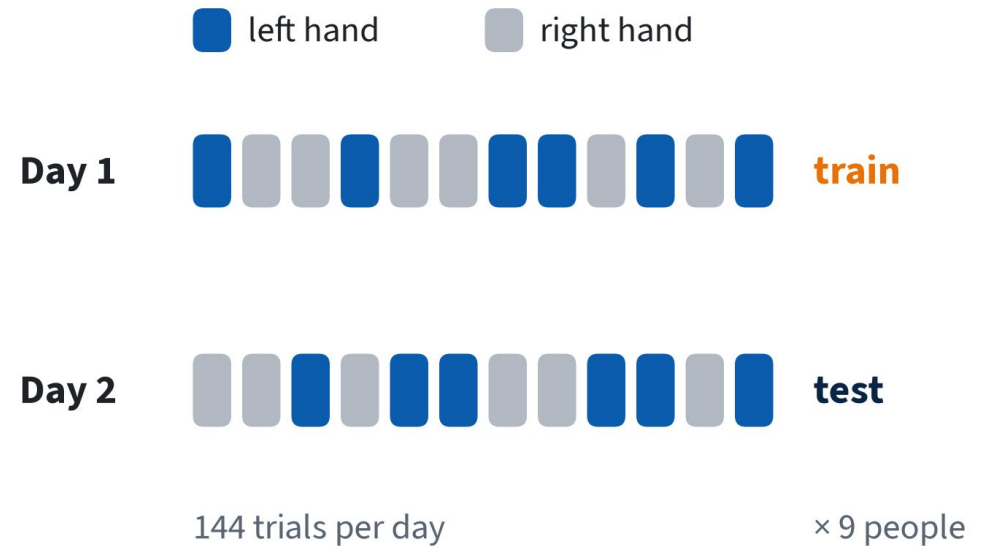
Few or nearly solved test tasks

An **average gain** can hide losses for some people



Classical pipelines and REVE on BCI IV 2a

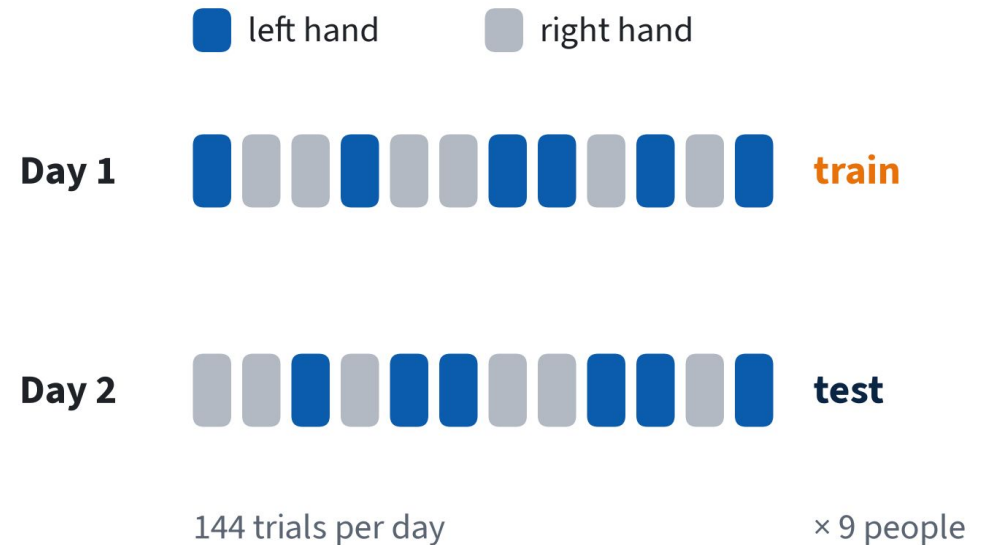
9 people imagine moving the left or right hand



Classical pipelines and REVE on BCI IV 2a

9 people imagine moving the left or right hand

Training on **day 1**, testing on **day 2**

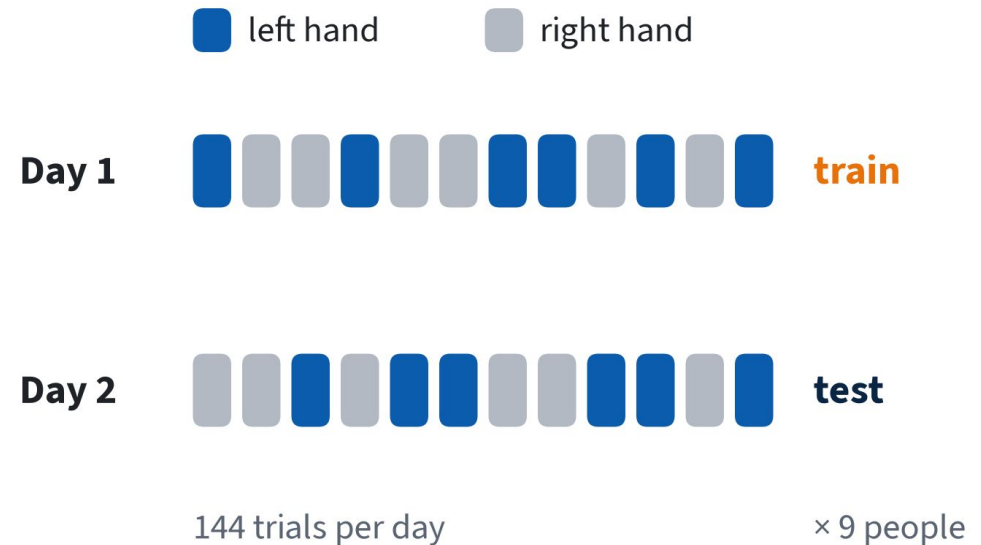


Classical pipelines and REVE on BCI IV 2a

9 people imagine moving the left or right hand

Training on **day 1**, testing on **day 2**

Two classical pipelines: **CSP + LDA**, **tangent space** + logistic regression



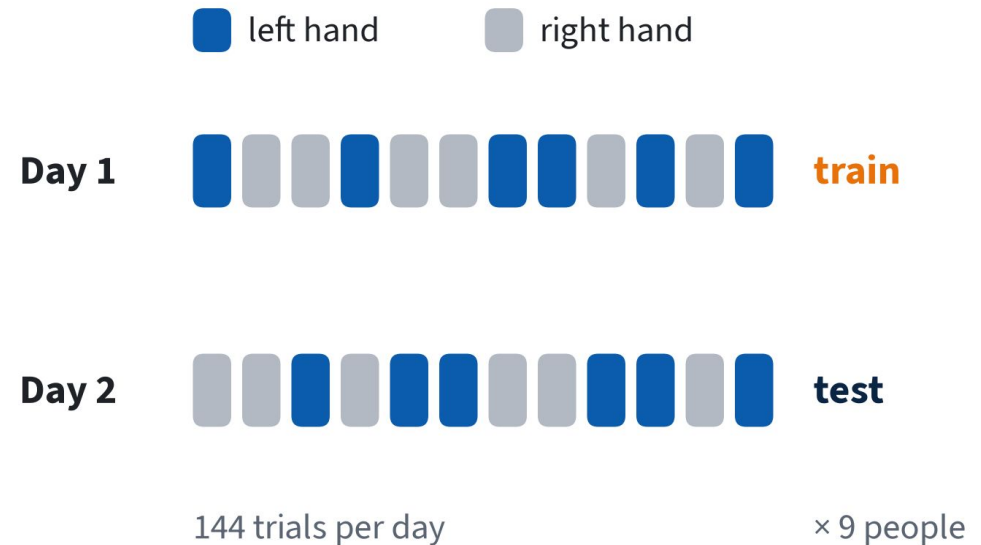
Classical pipelines and REVE on BCI IV 2a

9 people imagine moving the left or right hand

Training on **day 1**, testing on **day 2**

Two classical pipelines: **CSP + LDA**, **tangent space** + logistic regression

REVE frozen with a linear classifier, and
REVE fine-tuned



Classical pipelines and REVE on BCI IV 2a

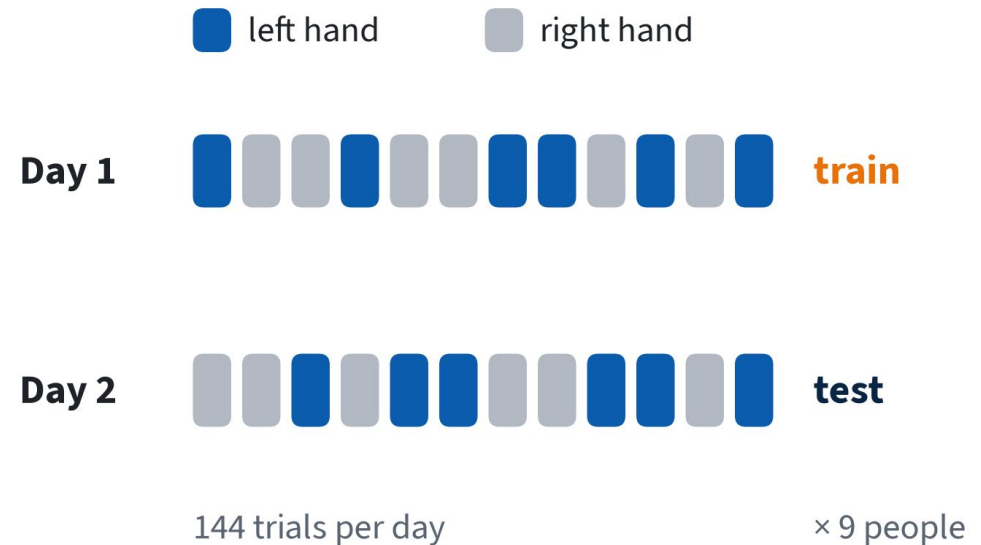
9 people imagine moving the left or right hand

Training on **day 1**, testing on **day 2**

Two classical pipelines: **CSP + LDA**, **tangent space** + logistic regression

REVE frozen with a linear classifier, and **REVE fine-tuned**

All four use the same labelled trials

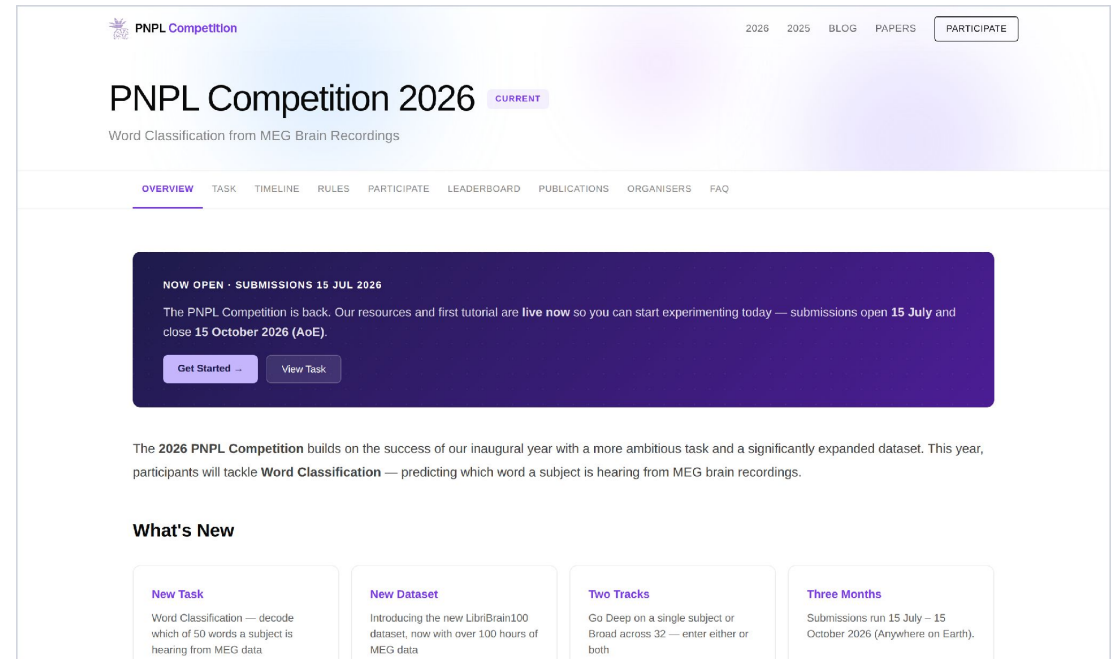


Which pipeline will score best on BCI IV 2a?

- A** CSP + LDA
- B** Riemannian tangent space + logistic regression
- C** REVE, frozen, with a linear classifier
- D** REVE, fine-tuned

Get involved: MEG

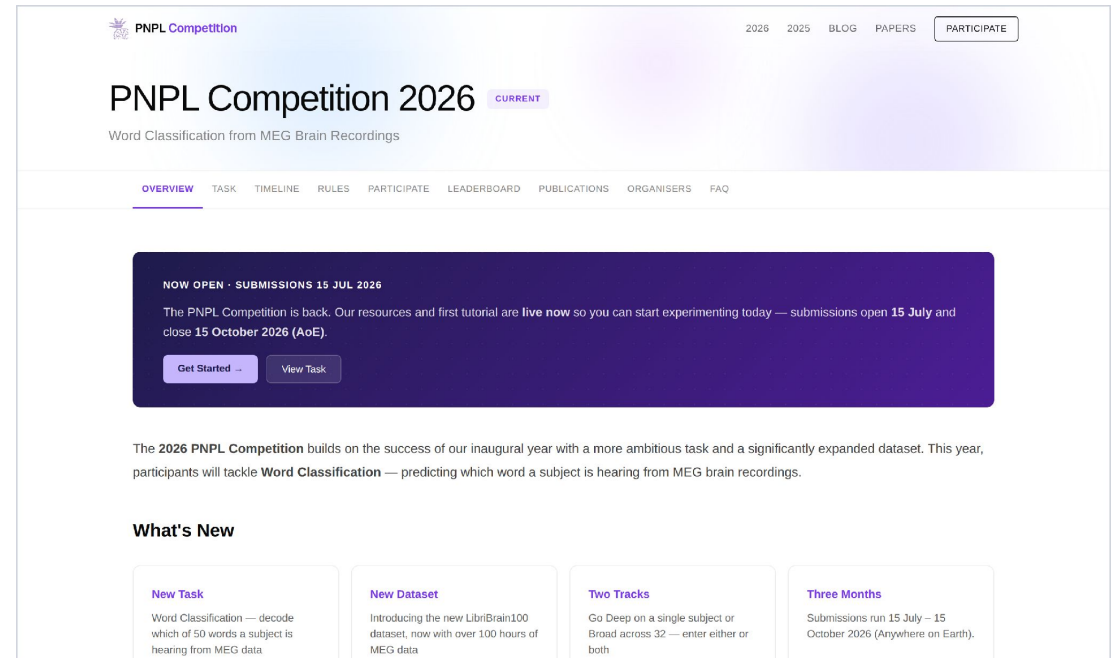
LibriBrain: 50+ h of MEG from one person listening to audiobooks



Get involved: MEG

LibriBrain: 50+ h of MEG from one person listening to audiobooks

PNPL Competition 2026: submissions until 15 October

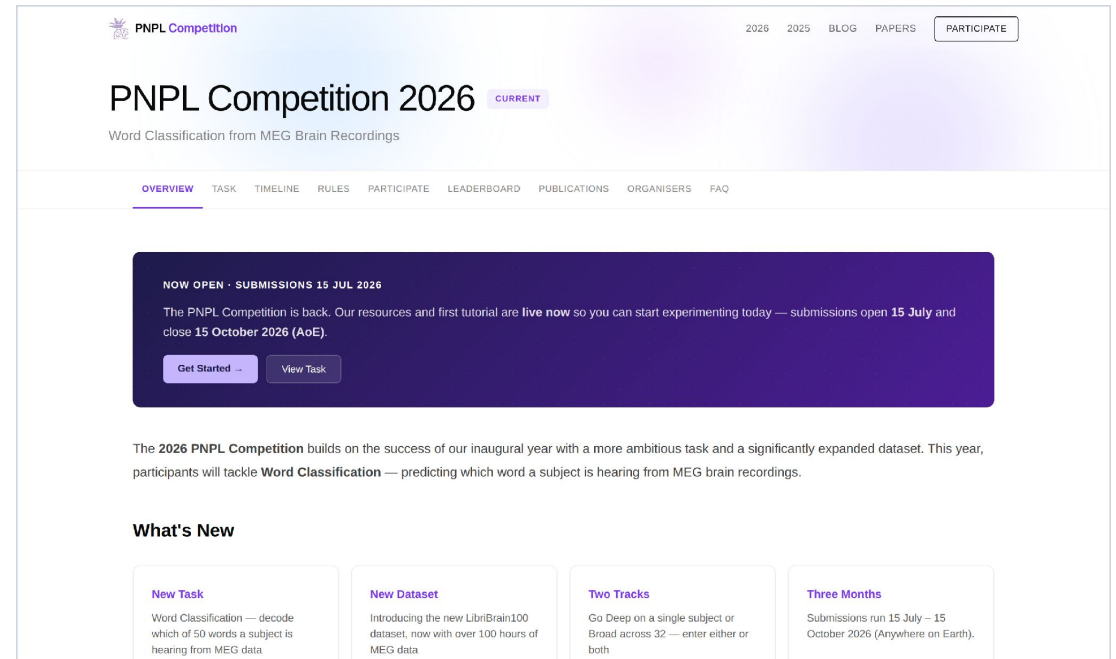


Get involved: MEG

LibriBrain: 50+ h of MEG from one person listening to audiobooks

PNPL Competition 2026: submissions until 15 October

Task: which of **50 words** a person is hearing



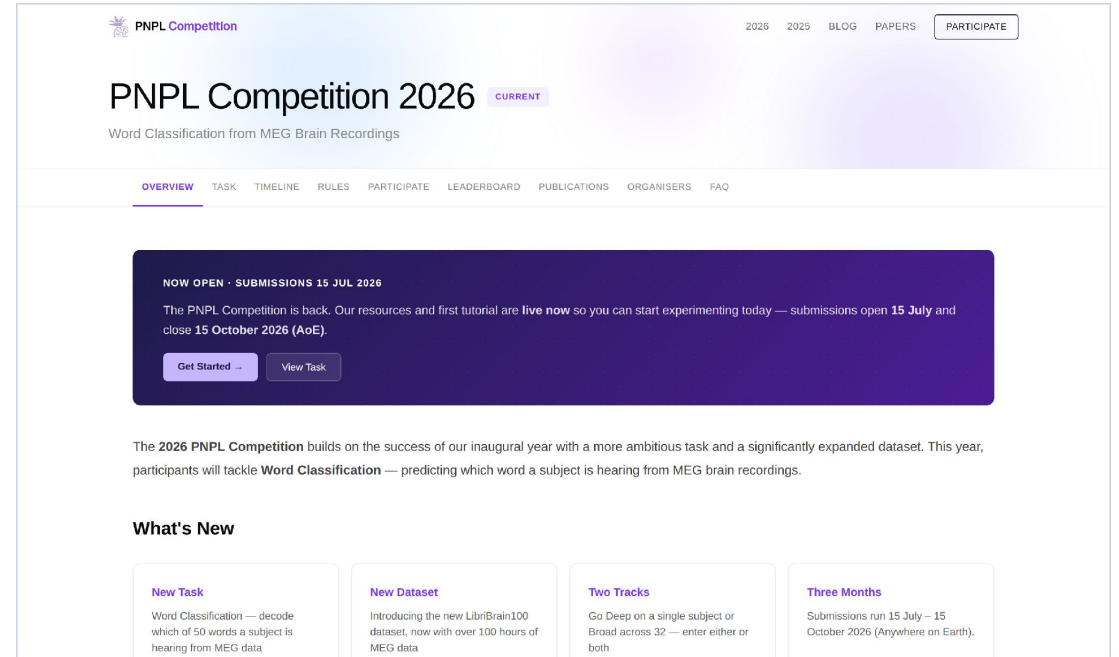
Get involved: MEG

LibriBrain: 50+ h of MEG from one person listening to audiobooks

PNPL Competition 2026: submissions until 15 October

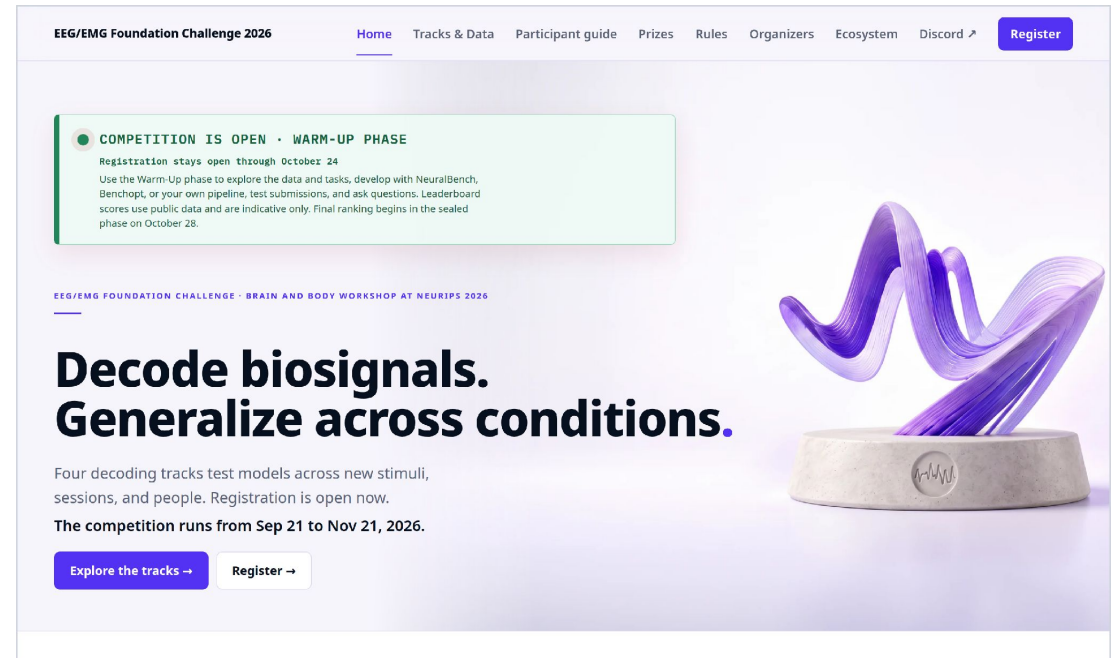
Task: which of **50 words** a person is hearing

LibriBrain100: 100+ h; one person or 32 people



Get involved: EEG

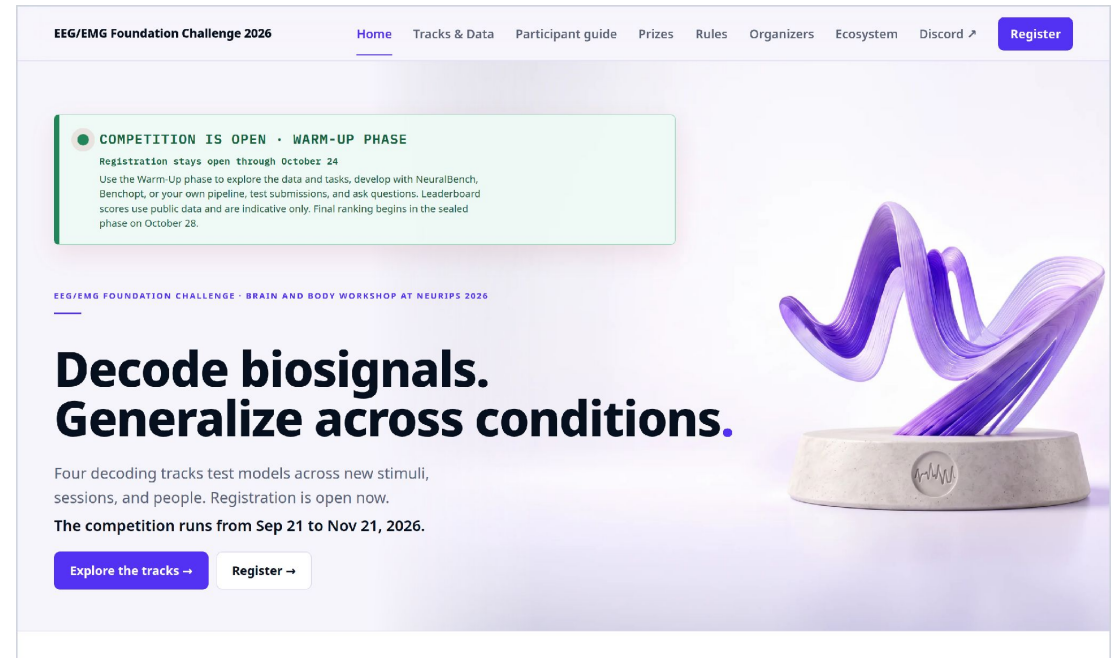
EEG/EMG Foundation Challenge 2026: registration until 24 October



Get involved: EEG

EEG/EMG Foundation Challenge 2026: registration until 24 October

Tracks: EEG to image, BCI across sessions,
sleep onset, EMG to hand pose

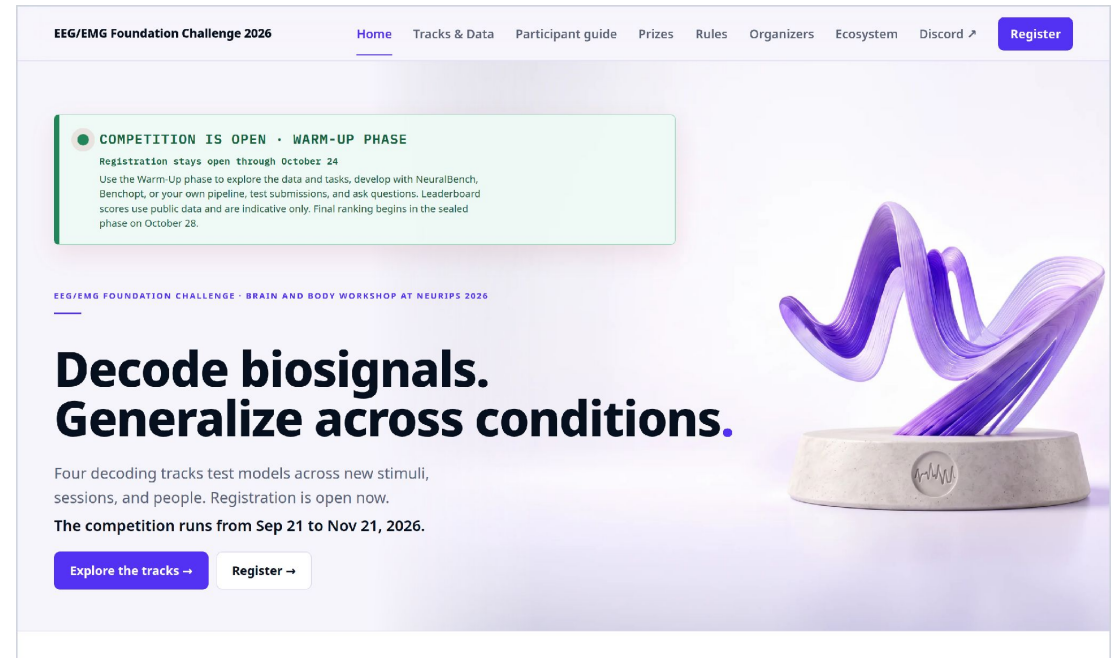


Get involved: EEG

EEG/EMG Foundation Challenge 2026: registration until 24 October

Tracks: EEG to image, BCI across sessions,
sleep onset, EMG to hand pose

Starter kits: NeuralBench, Braindecode,
MOABB, EEGDash



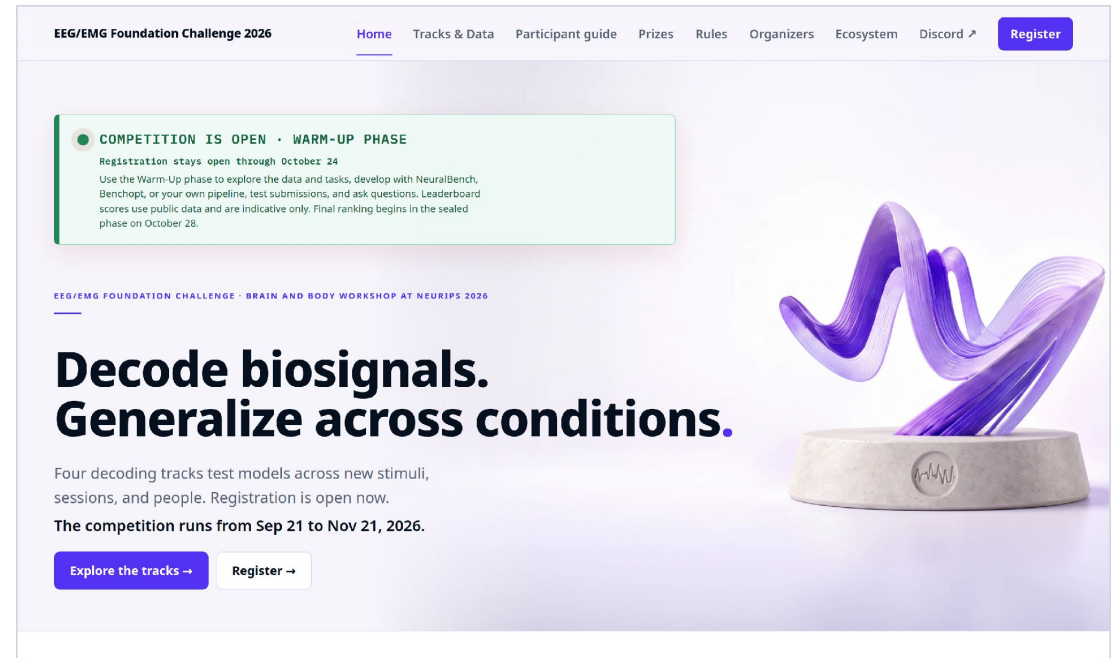
Get involved: EEG

EEG/EMG Foundation Challenge 2026: registration until 24 October

Tracks: EEG to image, BCI across sessions,
sleep onset, EMG to hand pose

Starter kits: NeuralBench, Braindecode,
MOABB, EEGDash

NeurIPS 2026 Brain & Body workshop: Sydney, 11–12 December



Takeaways

Foundation models are **pretrained once** on broad unlabelled data, then adapted

Takeaways

Foundation models are **pretrained once** on broad unlabelled data, then adapted

On benchmarks they have **a slight edge**; strong classical pipelines stay close

Takeaways

Foundation models are **pretrained once** on broad unlabelled data, then adapted

On benchmarks they have **a slight edge**; strong classical pipelines stay close

The result depends on **the dataset, the adaptation mode and the evaluation**

Takeaways

Foundation models are **pretrained once** on broad unlabelled data, then adapted

On benchmarks they have **a slight edge**; strong classical pipelines stay close

The result depends on **the dataset, the adaptation mode and the evaluation**

Check **the split and the baselines**

Takeaways

Foundation models are **pretrained once** on broad unlabelled data, then adapted

On benchmarks they have **a slight edge**; strong classical pipelines stay close

The result depends on **the dataset, the adaptation mode and the evaluation**

Check **the split and the baselines**

On BCI IV 2a today: [\[result from 05_compare\]](#)

Going further

Optional notebook: sleep staging on Sleep-EDF, band power vs REVE vs SleepFM

Going further

Optional notebook: sleep staging on Sleep-EDF, band power vs REVE vs SleepFM

Reviews: MEG roadmap (Thölke et al., 2026); critical EEG-FM review (Kuruppu et al., 2025)

Going further

Optional notebook: sleep staging on Sleep-EDF, band power vs REVE vs SleepFM

Reviews: MEG roadmap (Thölke et al., 2026); critical EEG-FM review (Kuruppu et al., 2025)

Benchmarks: NeuralBench, EEG-FM-Compass, NeuroAtlas

Going further

Optional notebook: sleep staging on Sleep-EDF, band power vs REVE vs SleepFM

Reviews: MEG roadmap (Thölke et al., 2026); critical EEG-FM review (Kuruppu et al., 2025)

Benchmarks: NeuralBench, EEG-FM-Compass, NeuroAtlas

Data: EEGDash, OpenNeuro, NEMAR

Going further

Optional notebook: sleep staging on Sleep-EDF, band power vs REVE vs SleepFM

Reviews: MEG roadmap (Thölke et al., 2026); critical EEG-FM review (Kuruppu et al., 2025)

Benchmarks: NeuralBench, EEG-FM-Compass, NeuroAtlas

Data: EEGDash, OpenNeuro, NEMAR

Code: Braindecode loads 10+ EEG foundation models

Thank you

Questions?

hamza.abdelhedi@umontreal.ca

hamzaabdelhedi.com

github.com/BabaSanfour

linkedin.com/in/hamza-abdelhedi