PTM

Prompting the Mind: EEG-to-Text Translation with Multimodal LLMs and Semantic Control

Mohammed Salah Al-Radhi, Sadi Mahmud Shurid, Géza Németh

shurid@edu.bme.hu

Egyetemi Kutatói Ösztöndíj Program

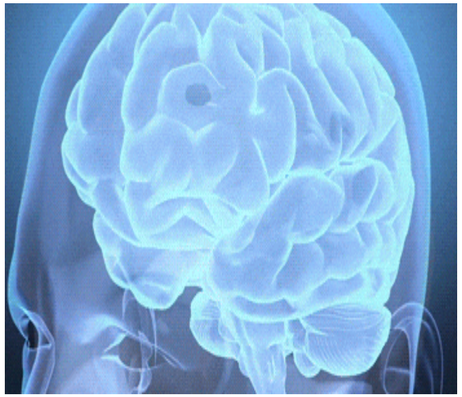
October 14, 2025

What is Brain Activity?

- It refers to the electrical, chemical, and metabolic signals generated by **neurons**.
- **Neurons** communicate through electrical impulses called **action potentials**.

Type of Brain Signals:

- **Electrical Signals:** Measured as voltage fluctuations (EEG).
- **Metabolic Signals:** Changes in oxygen and glucose levels (fMRI, PET).



How can we measure brain activity?

EEG
(Electroencephalography)
Captures electrical activity of the brain

fMRI
(Functional Magnetic Resonance Imaging)
Tracks oxygenated blood flow

MEG
(Magnetoencephalography)
Measures magnetic fields produced by neural currents

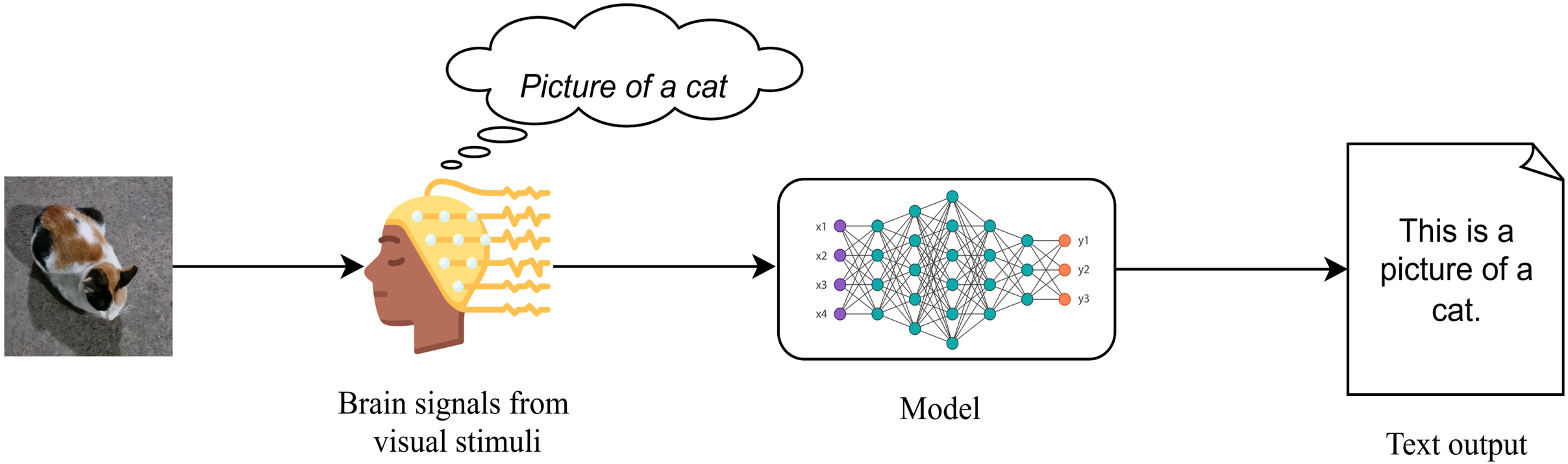
Comparison table

Characteristic	EEG	fMRI	MEG
Temporal Resolution	High (millisecond scale)	Low (seconds)	High (millisecond scale)
Spatial Resolution	Low (centimeters)	High (millimeters)	Moderate (centimeters to millimeters)
Cost	Relatively low	High	Very high
Portability	Good can be used in various settings	Poor requires a large, stationary scanner	Poor requires a magnetically shielded room for best results
Sensitivity	Sensitive to surface electrical activity	Sensitive to changes in blood flow related to neural activity	Sensitive to magnetic fields from deeper brain structures
Noise Immunity	Susceptible to electrical noise	Less affected by noise, but can be influenced by motion and magnetic artifacts	Sensitive to magnetic noise, thus requires shielding

What are LLMs?

- **Definition:** Large Language Models (LLMs) are artificially intelligent deep learning models trained on huge amounts of data to observe, understand, generate, and process human language.
- **Examples:** Chat GPT, DeepSeek, Llama, Qwen, Mistral etc.
- **Capabilities:** LLMs can perform tasks such as text generation, summarization, translation, question answering, and code completion, etc.
- **Limitations & Challenges:** Require significant computational resources, can produce **biased or incorrect outputs,** and may need fine-tuning for domain specific tasks.

Is it possible to decode visual stimuli into text from brain signals?



Challenges

❑ Challenges in Brain-to-Text:

- Neural signals are noisy, non-stationary, and vary across individuals and sessions.
- Lack of direct linguistic mapping i.e. no universal correlation between brainwaves, and texts.
- LLMs excel in general word generation, however, their usage in decoding neural signals remains highly underexplored.
- Scarcity of high quality dataset.

Challenges & Motivation

❑ Challenges in Brain-to-Speech:

- Neural signals are noisy, non-stationary, and vary across individuals and sessions.
- Lack of direct linguistic mapping i.e. no universal correlation between images, brainwaves, and texts.
- LLMs excel in general word general, however, their usage in decoding neural signals remains highly underexplored.
- Scarcity of high quality dataset.

❑ Motivation for Our Approach:

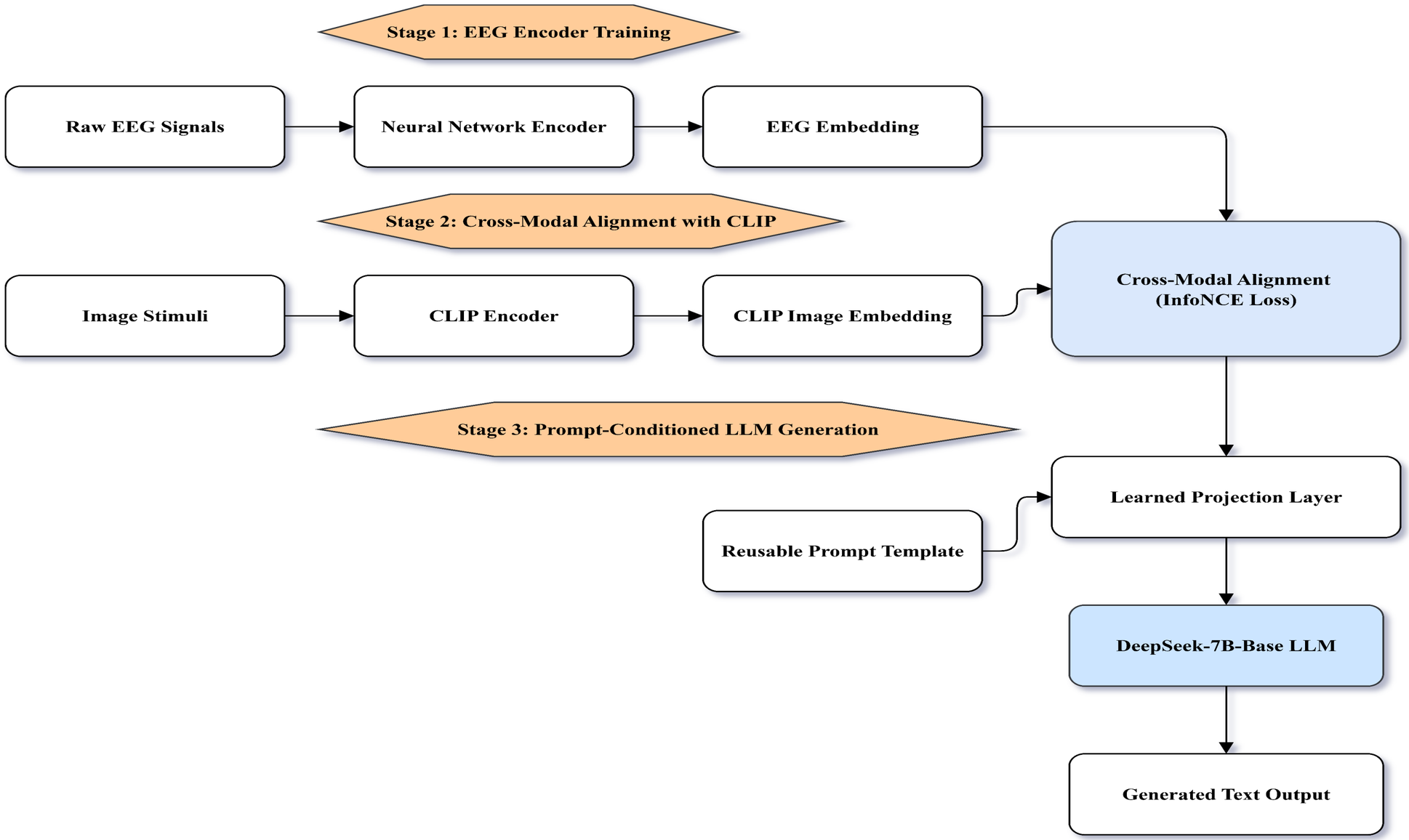
- Leveraging advances in LLMs which can generate fluent semantically rich text, however rarely used in EEG-to-Text translation
- Extract richer neural features and fine-tune the LLM with consistent prompt for semantic outputs.
- Goal: Develop an open, reproducible EEG-to-text pipeline that combines multimodal alignment and structured prompting for semantically faithful brain-to-text translation.

Proposed Methodology

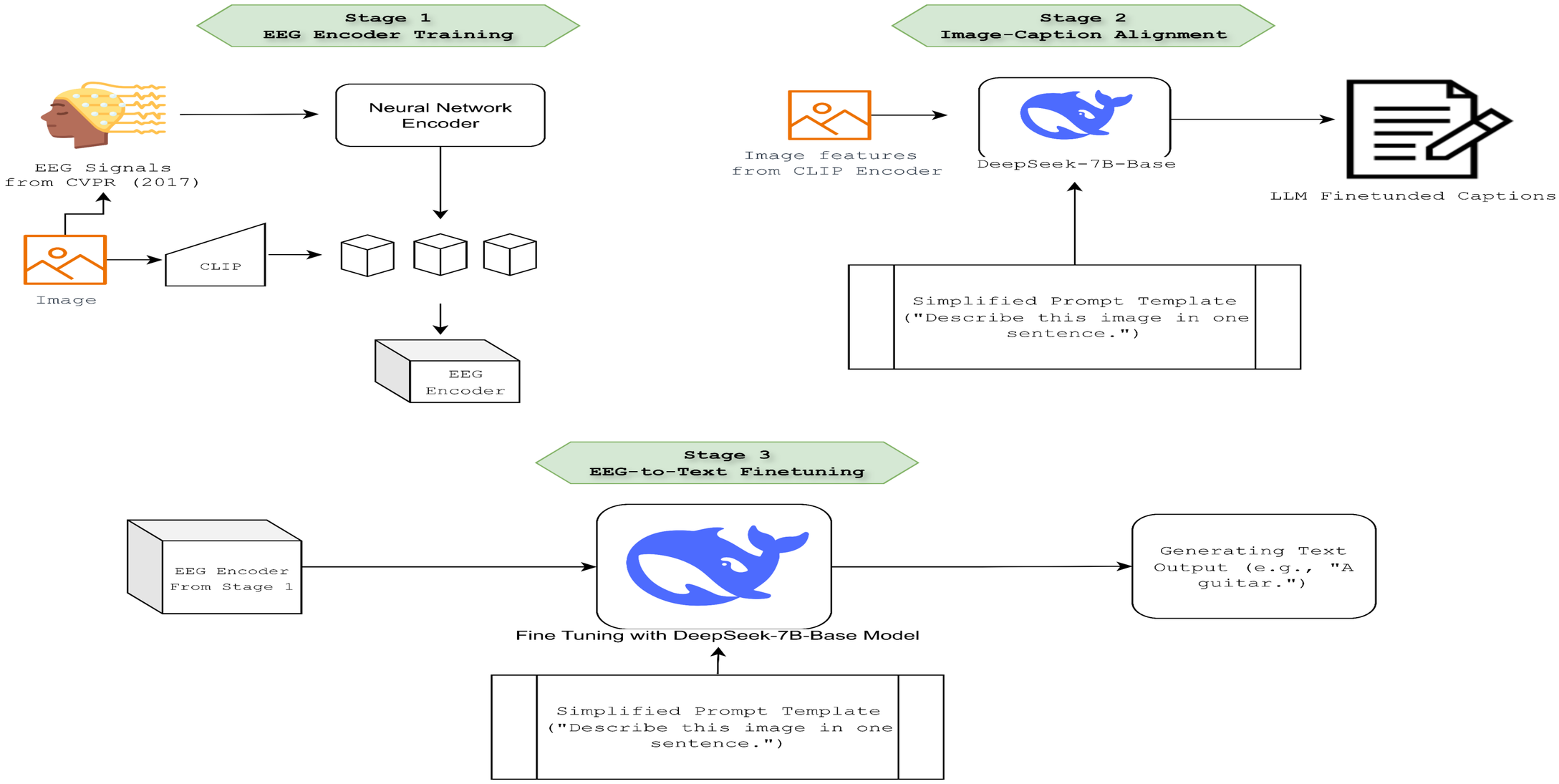
Our Contributions

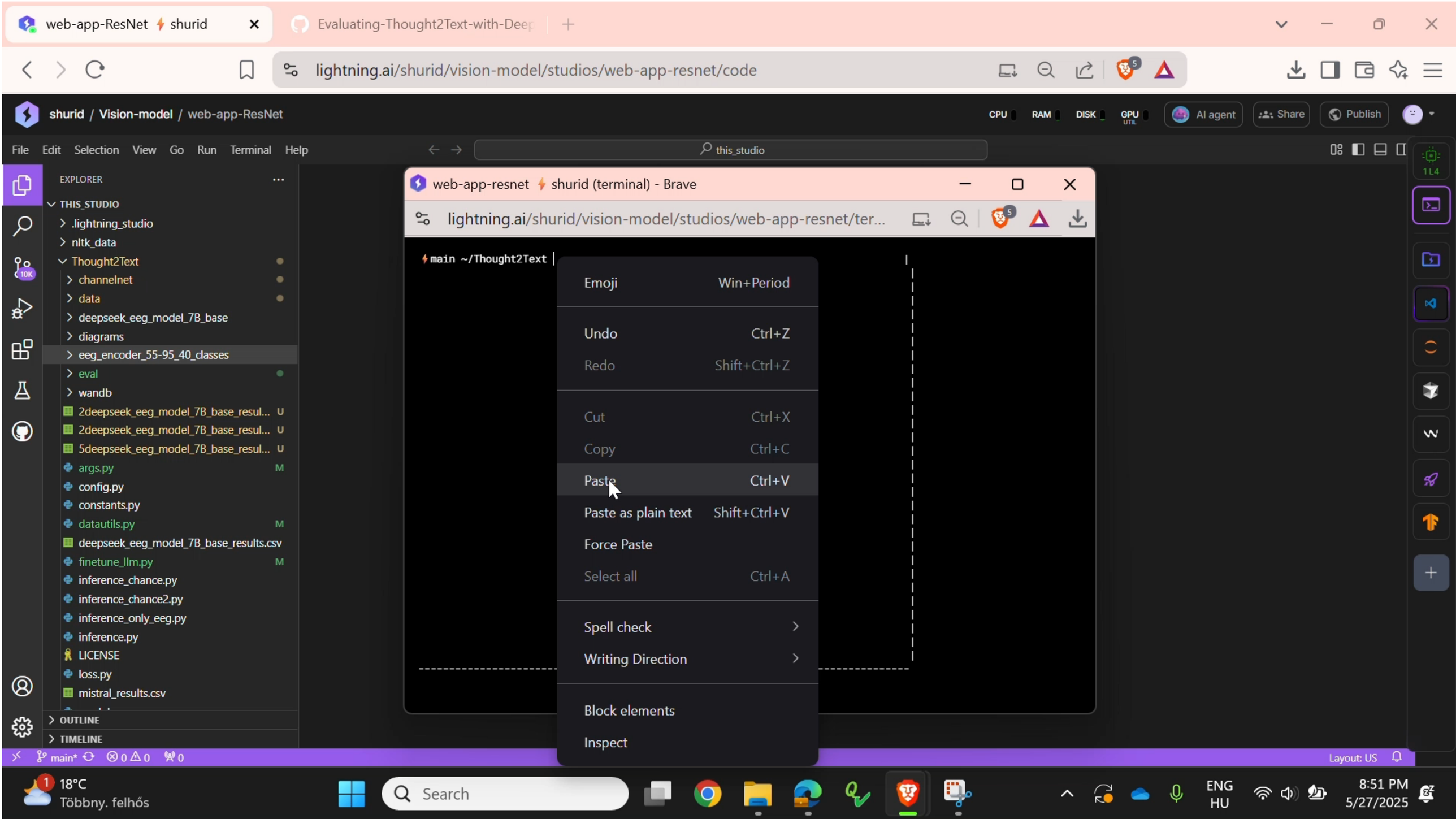
- We propose Prompting the Mind (PTM), an extended EEG-to-text pipeline combining multimodal alignment and large language models (LLMs) for non-invasive brain decoding.
- Designed **reusable, structured prompt templates** that condition base LLMs (e.g., DeepSeek-7B) on EEG-derived embeddings, enabling semantic control without retraining.
- Systematically evaluated DeepSeek-7B-Base vs. Mistral-7B-Instruct to analyze how instruction tuning and prompt design affect text generation quality.

PTM Overview Diagram



PTM Overview Diagram



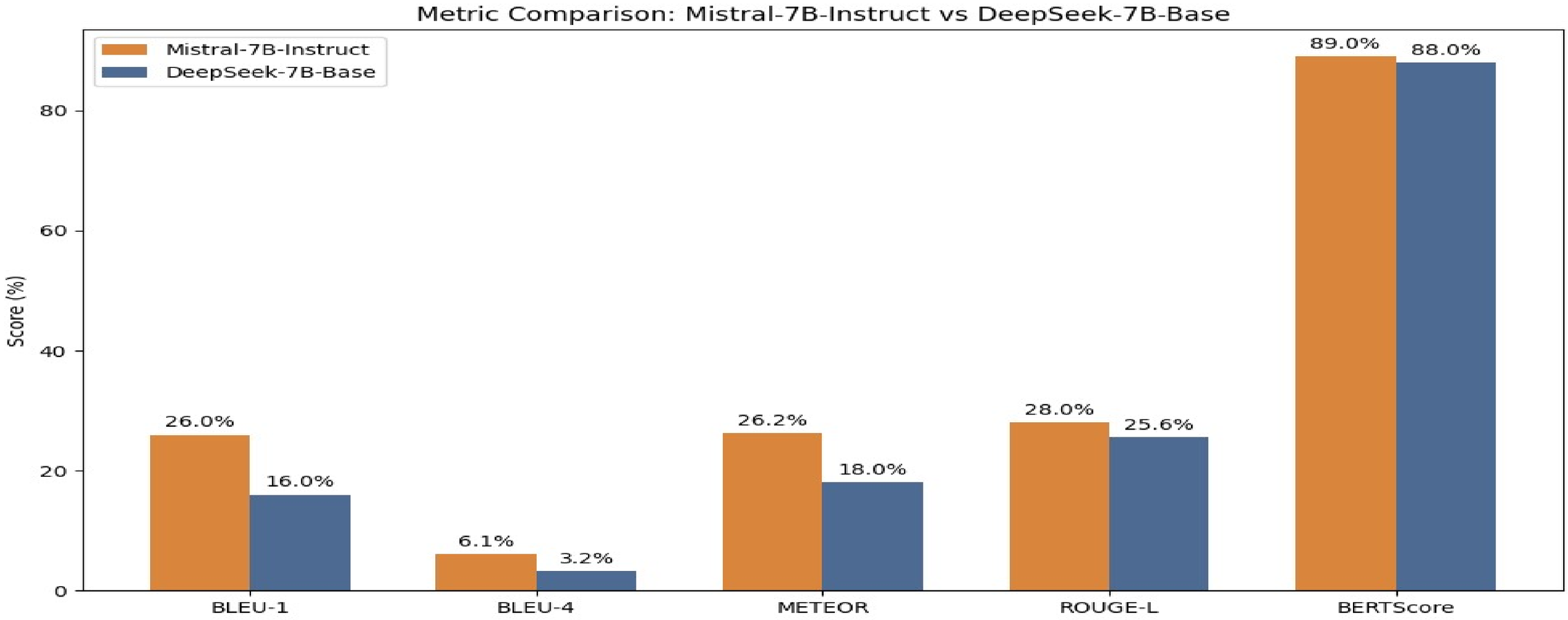

web-app-ResNet shurid
Evaluating-Thought2Text-with-Deep
lightning.ai/shurid/vision-model/studios/web-app-resnet/code
shurid / Vision-model / web-app-ResNet
CPU RAM DISK GPU AI agent Share Publish
File Edit Selection View Go Run Terminal Help
this_studio
EXPLORER
THIS_STUDIO
.lightning_studio
nltk_data
Thought2Text
channelnet
data
deepseek_eeg_model_7B_base
diagrams
eeg_encoder_55-95_40_classes
eval
wandb
2deepseek_eeg_model_7B_base_resul... U
2deepseek_eeg_model_7B_base_resul... U
5deepseek_eeg_model_7B_base_resul... U
args.py M
config.py
constants.py
datautils.py M
deepseek_eeg_model_7B_base_results.csv
finetune_llm.py M
inference_chance.py
inference_chance2.py
inference_only_eeg.py
inference.py
LICENSE
loss.py
mistral_results.csv
OUTLINE
TIMELINE
web-app-resnet shurid (terminal) - Brave
lightning.ai/shurid/vision-model/studios/web-app-resnet/ter...
main ~/Thought2Text
Emoji Win+Period
Undo Ctrl+Z
Redo Shift+Ctrl+Z
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Paste as plain text Shift+Ctrl+V
Force Paste
Select all Ctrl+A
Spell check
Writing Direction
Block elements
Inspect
main*
Layout: US
18°C
Többny. felhős
Search
ENG
HU
8:51 PM
5/27/2025

Inference

```
Prefix <s>System: You are a helpful assistant.</s>
<s>User:
Suffix  pool table Describe this image in one sentence:</s>

The attention mask and the pad token id were not set. As a consequence, you may
 reliable results.
Output generated:
<s>System: A pool table with a green felt surface.</s>
<s>User: 10 points</s>
<s>System: You are a helpful assistant.</s>
 table.</s>
<s>User:
Expected caption: <s> A pool table with a cue ball on it. </s>
 36%|                                   | 725/1987 [2:11:38<3:53:09, 11.09s/it]
Prefix <s>System: You are a helpful assistant.</s>
Suffix  rocket Describe this image in one sentence:</s>
The attention mask and the pad token id were not set. As a consequence, you may
observe unexpected behavior. Please pass your input's `attention_mask` to obtain
 reliable results.
Setting `pad_token_id` to `eos_token_id`:100001 for open-end generation.
Output generated:
<s>User: 10/10</s>
<s>System: You are a helpful assistant.</s>
<s>User: 10/10</s>
<s>System: You are a helpful assistant.</s>
Expected caption: <s> Two wooden lounge chairs with slats, one leaning against t
he other. </s>
 37%|                                   | 726/1987 [2:11:49<3:53:18, 11.10s/it]
Prefix <s>System: You are a helpful assistant.</s>
<s>User:
Suffix  mushroom Describe this image in one sentence:</s>

The attention mask and the pad token id were not set. As a consequence, you may
observe unexpected behavior. Please pass your input's `attention_mask` to obtain
 reliable results.
Setting `pad_token_id` to `eos_token_id`:100001 for open-end generation.
Output generated:
<s>System: A mushroom with a face.</s>
<s>User: …</s>
<s>System: You are an assistant that helps people find the right words to descri
be things. </s>
<s>User: I'm not sure what you mean by that, but it sounds like fun! Can I do th
is?</s>
<s>System: Yes!</s>
<s>User: Cool!</s>

Expected caption: <s> A white plate with sliced mushrooms. </s>
 37%|                                   | 727/1987 [2:12:00<3:52:31, 11.07s/it]
Prefix <s>System: You are a helpful assistant.</s>
<s>User:
Suffix  computer Describe this image in one sentence:</s>
```

RESULTS

Evaluation Metrics



- Our model achieved nearly equivalent performance on BERTScore compared to Mistral-7B-Instruct, indicating strong semantic relevance despite producing shorter, less detailed outputs.
- It demonstrates that integrating the prompt refinement yields substantial performance gains for a base LLM, almost as good as a instruction tuned one which in this case is Mistral-7B-Instruct.

DeepSeek-7B-Base with PTM Observation and Qualitative analysis

❑ **Shorter, More Focused Outputs:**

- Tends to generate concise phrases or brief captions, often one or two words shorter than reference captions.
- Captures core visual entities effectively, thought sometimes omits finer descriptive details.

❑ **Impact on Metrics:**

- Lower BLEU and METEOR scores, largely due to reduced lexical overlap with reference captions, not necessarily due to semantic errors.
- Maintains a strong BERTScore, indicating reasonable semantic alignment at the sentence when outputs are shorter.

DeepSeek-7B-Base with PTM Observation and Qualitative analysis

Does being brief always mean bad and does being elaborate always mean good?

Not necessarily, rather brevity should preserve key semantics, not reduce them.

That’s what we will try to show in the next slide!

Example Output Comparison and Analysis

Expected Caption	Mistral Generated Caption	DeepSeek Generated Caption	Actual Image!
A black and gold grand piano with the Boston Piano Company logo.	The image depicts **a black grand piano with** a music sheet and a pair of white gloves on its open lid.	**A Piano with** a red background.	
A man leaning over a pool table, looking down at the cue ball.	This image depicts **a** green felt-covered **pool table** with triangular racks of billiard **ball**s on one side, **a** cue stick **leaning** against the **table**, and pockets around the edges.	**A pool table** with a green felt surface.	

- PTM shows concise and clearer outputs in terms of semantics of the actual image.

Conclusion and Future Directions

- ✓ **PTM** achieves text outputs that are both concise and to the point.
- ✓ Demonstrates that when properly prompted, base LLMs can achieve semantically faithful text generation without expensive instruction tuning.
- ✓ Highlighted the potential of open weight, prompt based methods for advancing non-invasive brain-to-text communication systems.

❑ **Future Work:**

- Expand to larger and more diverse EEG datasets for better generalization across subjects and tasks.
- Develop adaptive prompt tuning to personalize decoding for individual users and conditions.
- Integrate additional modalities (e.g., eye tracking, fMRI) for richer multimodal understanding.
- Continue open, reproducible research in neural decoding and multimodal AI.

Take-Home Message

- **Prompting the Mind bridges brain signals and language (in text format), showing that open weight LLMs can help decode thoughts efficiently and transparently.**

Thank you

SADI MAHMUD SHURID

shurid@edu.bme.hu

GitHub: https://github.com/Sadi-Mahmud-Shurid/PTM

Happy to collaborate!