Sapienza University of Rome
Master’s Degree in
Artificial Intelligence and Robotics

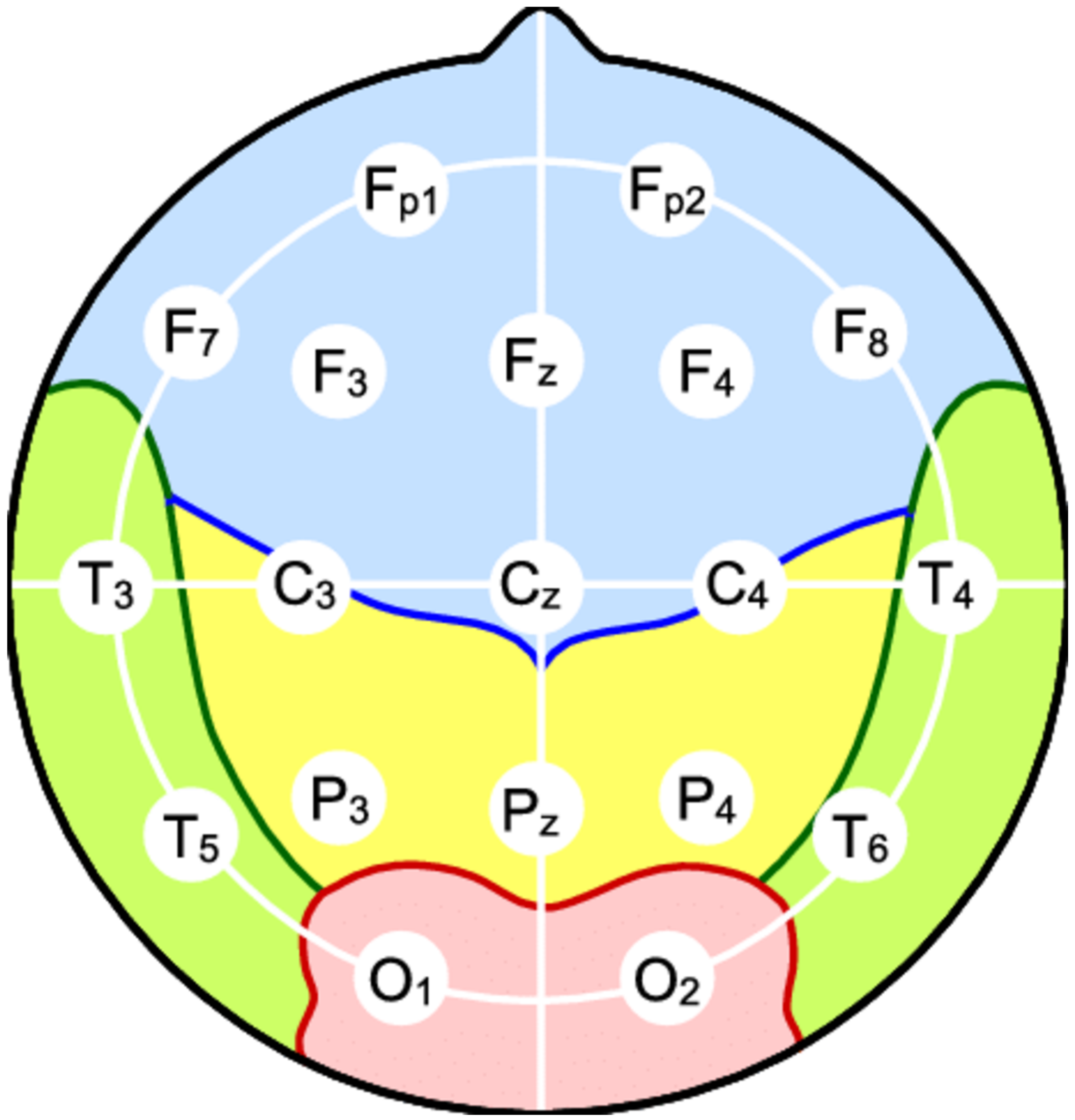
LLM-POWERED EMOTION RECOGNITION FROM MUSIC-EVOKED EEG SIGNALS

Antonello Giorgio, 1836529
Advisor: Prof. Danilo Comminiello

July 22, 2025 — Rome, Italy

What is EEG and Why It Matters

- **EEG** (Electroencephalography) records electrical **brain activity via scalp electrodes**
- Based on the **international 10–20 system** for electrode placement
- Offers high temporal resolution, **ideal for monitoring rapid neural responses**
- **Non-invasive, portable**, and widely used in affective computing
- **Preprocessing** is crucial: filtering, segmentation, artifact removal

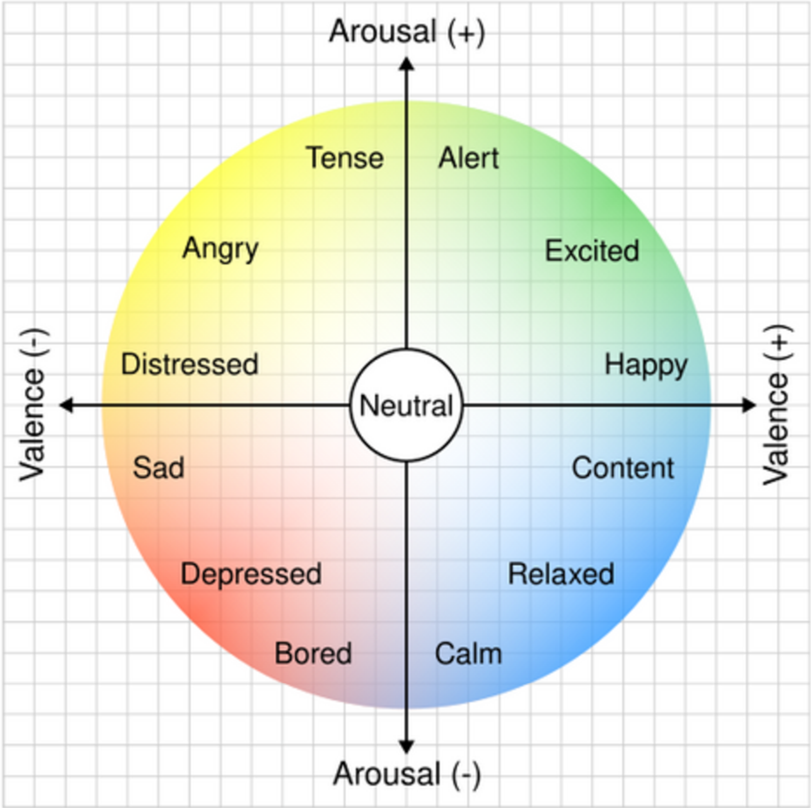


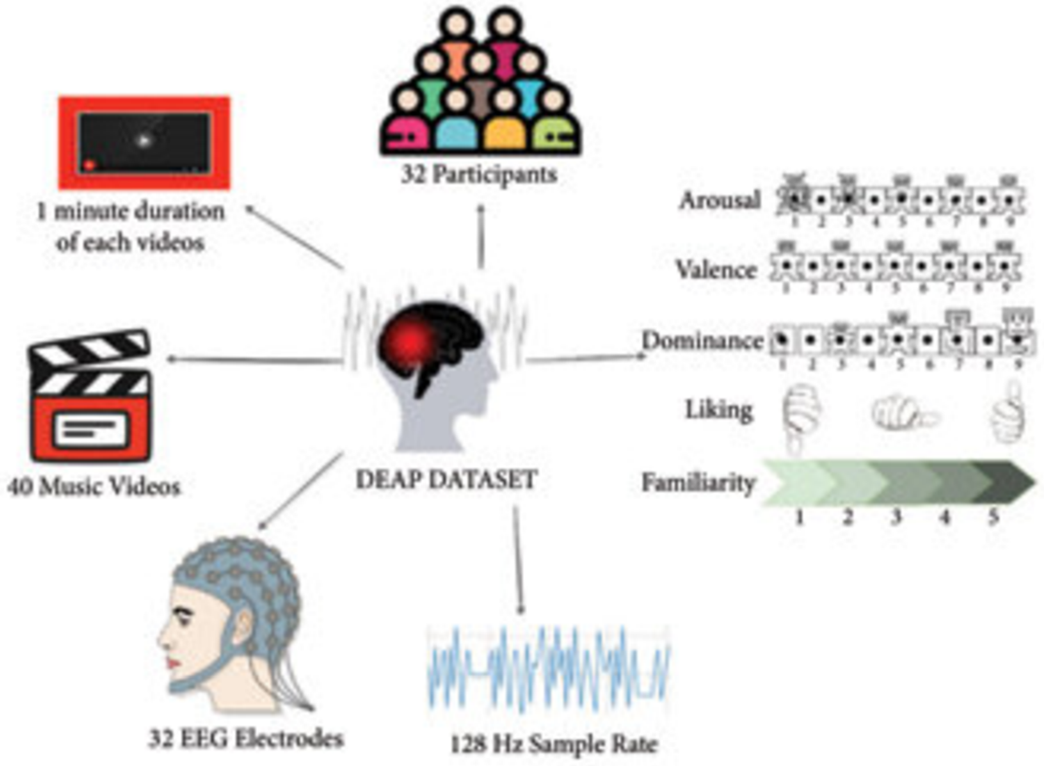
State of the Art

- Early methods use **hand-crafted features + shallow classifiers (e.g., SVM, KNN)**
- **Deep learning** introduced **CNNs and RNNs** to capture **spatial and temporal features separately**
- **GNN-based models** (e.g., LGGNet) learn **dynamic electrode interactions**
- Some models (e.g., TSCeption) use **multi-scale convolutions for temporal-spectral encoding**
- However, **no** model leverages **pretrained Large Language Models (LLMs) to process EEG sequences**

DEAP Dataset and Preprocessing

- **32 participants** watched **40 music videos** (1 min each)
- EEG recorded from **32 channels at 128 Hz** (10–20 system)
- **Self-assessment ratings** for valence and arousal (scale 1–9)
- **Labels binarized**: ≥5 → High, <5 → Low
- **Preprocessing**:
  - Bandpass filtering (4–45 Hz)
  - Artifact rejection (energy and correlation)
  - Z-score normalization (per-channel)
- **Segmented** into **overlapping windows** (1s / 4s, 50% overlap)

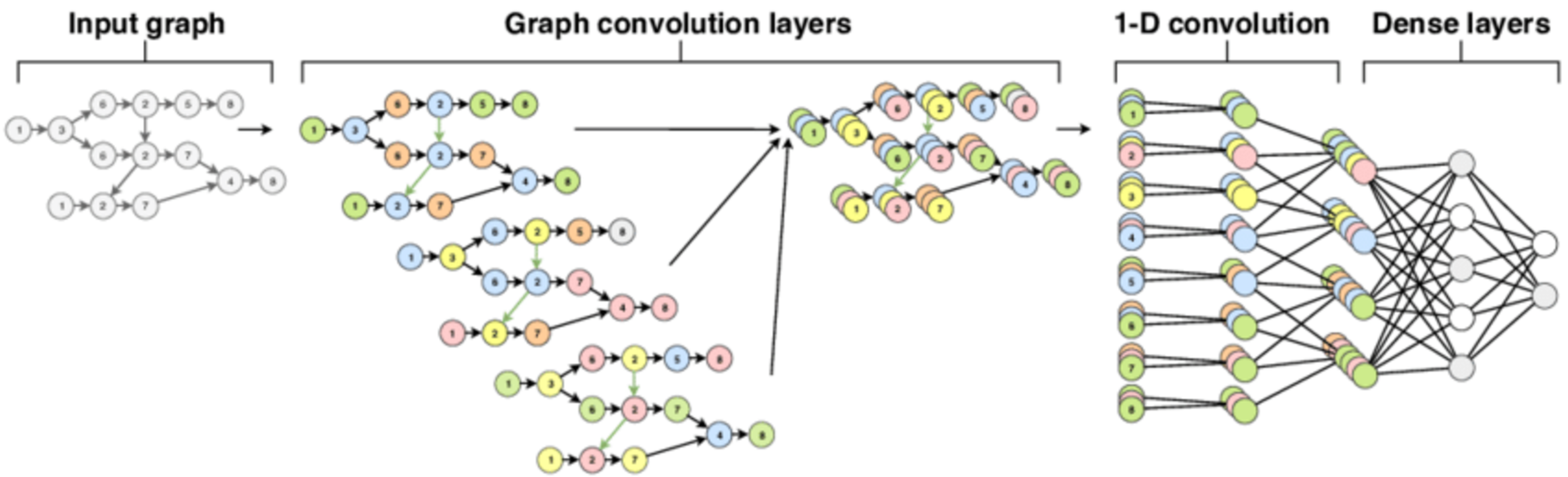




Spatial Branch – Dynamic Graph Convolutions over EEG Channels

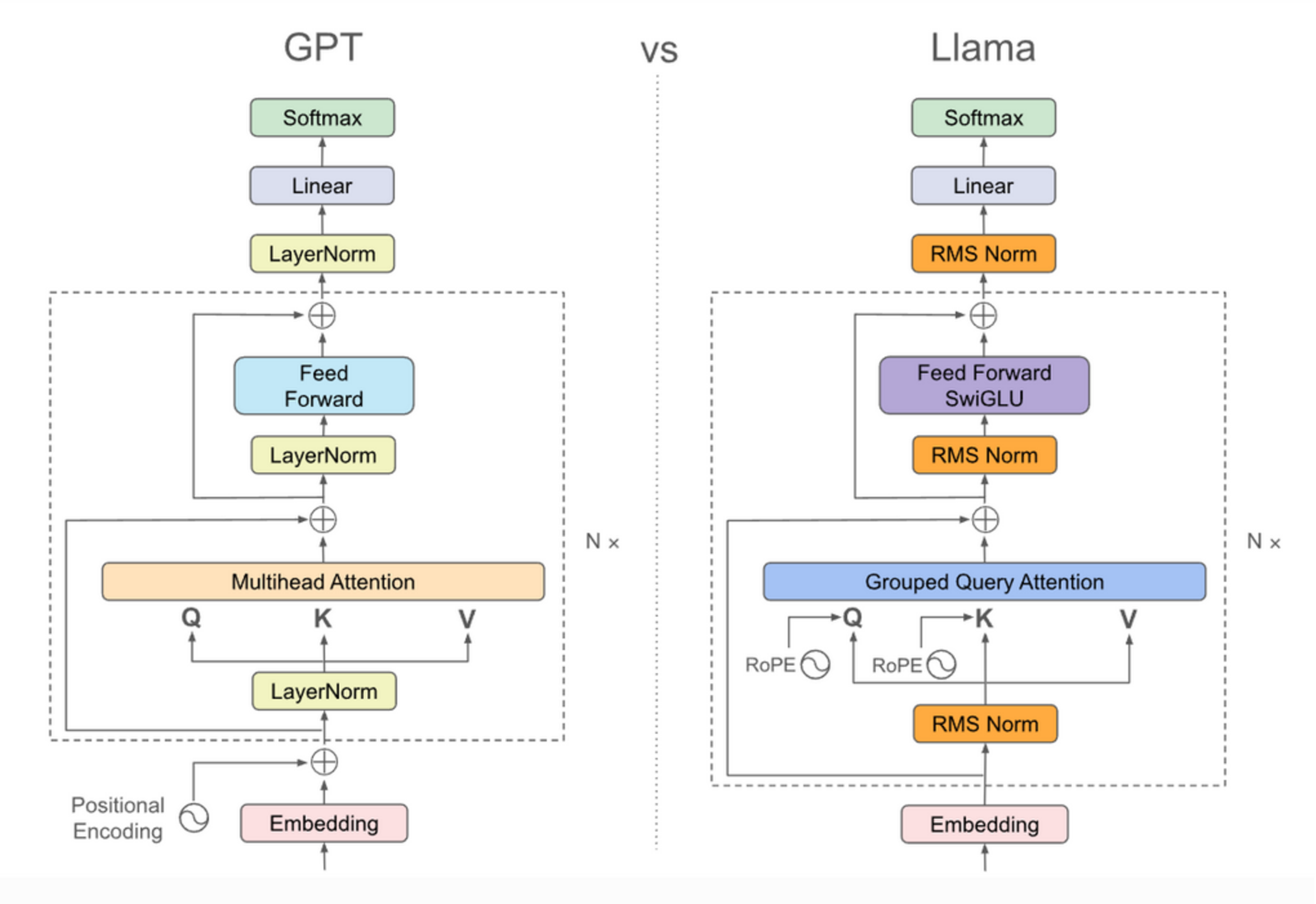
- Addresses the **spatial complexity of EEG signals** with a **learnable graph structure**
- EEG **electrodes** represented **as graph nodes**
- **Inter-electrode relationships** modeled via **learned edge weights**

- Produces a **spatial embedding from each EEG segment**
- Enables **modeling of topological and physiological brain structure**

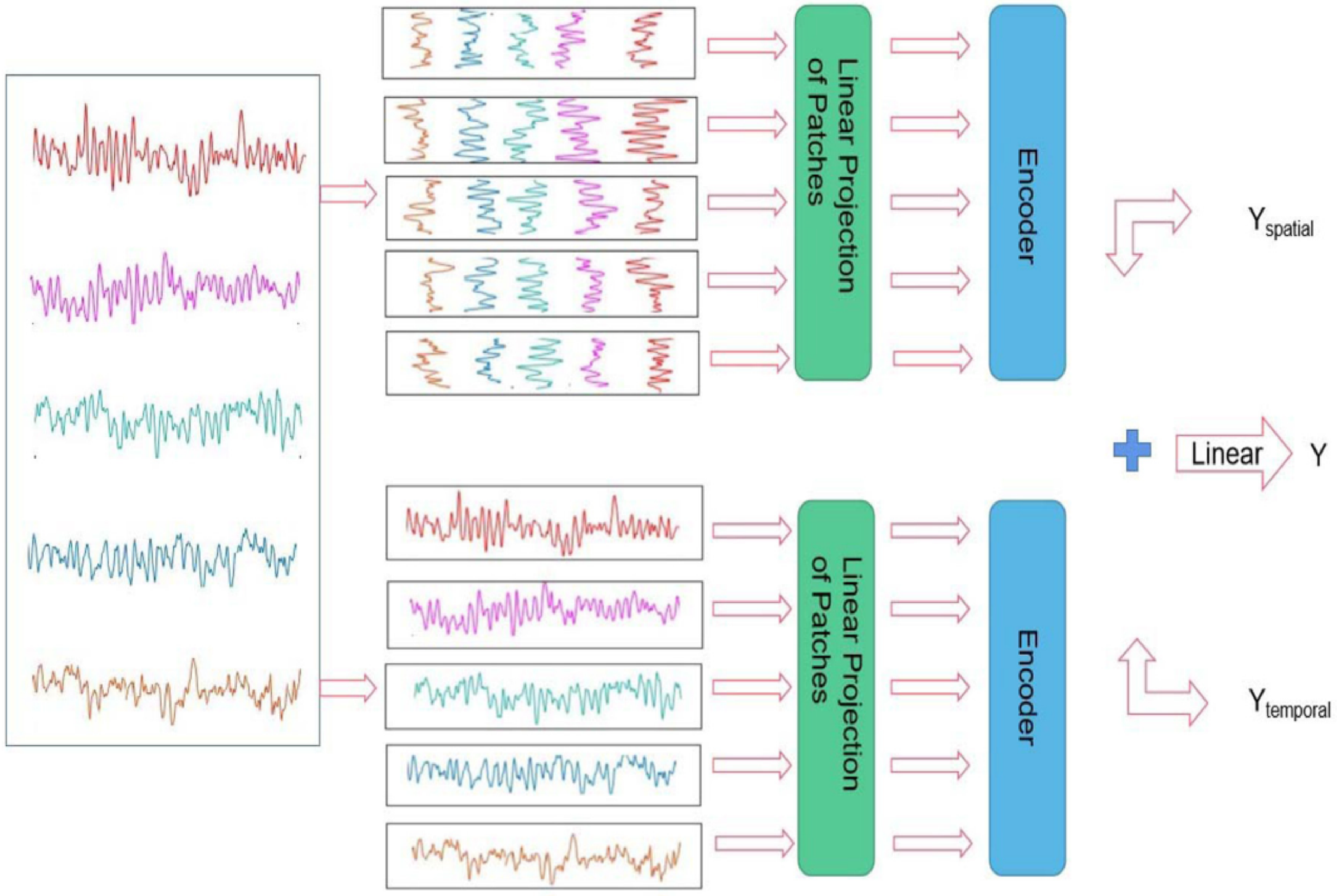


Temporal Branch: EEGTransformer + LLM Decoder

- EEG segments encoded using a **Transformer encoder**
- Encoded **embeddings** are partially **masked and passed** to a pretrained **LLM decoder** (GPT-2 / LLaMA)
- LLM is **fine-tuned with LoRA** on EEG via **masked embedding reconstruction**
- **Simultaneous classification** and **self-supervised decoding**
- Enables **robust temporal representation learning** over EEG sequences



Fusion Strategy and Multitask Training



- Temporal and spatial branches produce **separate logits**
- **Fusion** via:
  - **average**
  - **confidence-based gating**

$$\hat{\mathbf{y}} = \frac{1}{2}\left(\mathrm{softmax}(\hat{\mathbf{y}}_{\mathrm{T}}) + \mathrm{softmax}(\hat{\mathbf{y}}_{\mathrm{S}})\right)$$

- **Shared classification head for valence/arousal prediction**
- Multitask loss:
  - **Cross-entropy (emotion classification)**
  - **MSE Reconstruction loss (LLM decoder)**

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{N-1}\sum_{i=2}^{N} \|\hat{e}_i - e_i\|^2.$$

- Multitask learning **improves generalization on small, noisy datasets**

Experiments: Effect of LLM Fine-Tuning and Reconstruction Loss

Effect of LLM Fine-Tuning

Fine-tuning the LLM with LoRA leads to a clear performance boost compared to both the frozen decoder and the EEG-only baseline. This confirms that **fine-tuning is beneficial to fully exploit the LLM's potential on EEG data.**

Model Variant	ACC	F1
No LLM (EEGTransformer only)	57.48%	53.24%
Frozen LLM (no fine-tuning)	63.94%	61.55%
Fine-tuned LLM (w/ LoRA)	**66.99%**	**65.78%**

Model Configuration	ACC	F1
No reconstruction	64.02%	63.34%
LLM (fine-tuned), with $\mathcal{L}_{\text{rec}}$	66.99%	65.78%

Impact of Reconstruction Loss

Removing the reconstruction loss causes a drop in accuracy and F1 score. Its presence **helps the model learn richer temporal patterns and improves generalization**, especially in low-data regimes like EEG.

Experiments:
Segment Length and LLM-Backbone Comparison

Segment Length

Using **4-second EEG segments consistently outperforms 1-second segments**. Longer windows provide **more temporal context**, allowing both the Transformer and the LLM decoder to extract more informative features.

Segment Configuration	ACC	F1
1.0 s	65.09%	64.38%
4.0 s	**66.99%**	**65.79%**

LLM Backbone	ACC	F1
GPT-2 Small	66.99%	65.78%
LLaMA 3.2 1B	64.02%	63.59%
LLaMA 3.1 8B AWQ INT4	64.50%	64.28%

LLM-Backbone Comparison

Fine-tuned GPT-2 performs better than LLaMA, despite having fewer parameters. This suggests that, for EEG tasks with limited data, **smaller and less complex models offer better optimization and generalization** than larger ones.

Results

- Evaluation performed on DEAP: **subject-dependent classification**
- Our model achieves **66.99% accuracy on arousal prediction,** outperforming state-of-the-art models on DEAP

Comparison with SOTA models:

- **TSCeption** (61.57%) – CNN with temporal-spectral blocks
- **LGGNet** (61.81%) – GNN with local-global graph aggregation
- **MT-LGSGCN** (63.59%) – Multitask GCN with attention and global-local modules

Method	Arousal		Valence	
	ACC	F1	ACC	F1
SVM	60.37%	57.33%	55.19%	57.87%
KNN	59.48%	57.49%	53.03%	55.12%
ShallowConvNet	61.19%	61.19%	59.42%	62.26%
DeepConvNet	61.03%	62.58%	59.92%	62.04%
TSCeption	61.57%	63.24%	59.14%	62.33%
DGCNN	60.80%	60.34%	53.97%	56.27%
LGGNet-Fro	61.19%	63.96%	58.95%	63.89%
LGGNet-Hem	61.52%	63.79%	59.18%	64.34%
LGGNet-Gen	61.81%	64.49%	59.14%	64.58%
MT-LGSGCN	63.59%	65.11%	61.69%	65.23%
Ours (LLM-Powered)	**66.99%**	**65.78%**	**64.69%**	64.53%

Conclusions

- Proposed the **first direct fine-tuning (with LoRA) of an LLM on EEG data** using masked embedding reconstruction **on the DEAP dataset**
- Employed a **masked reconstruction objective** to **enhance temporal representation learning** and **regularize the decoder**
- Demonstrated that **smaller LLMs** (e.g., GPT-2) **can outperform larger LLMs** (e.g., LLaMA) in EEG tasks
- Achieved **state-of-the-art performance on DEAP for subject-dependent valence and arousal** classification
- Limited to the subject-dependent setting and single-modality EEG. Future research could **explore cross-subject generalization and multimodal fusion strategies**.

Sapienza University of Rome
Master’s Degree in
Artificial Intelligence and Robotics

THANK YOU

Antonello Giorgio, 1836529

July 22, 2025 — Rome, Italy