

# Pretrained Transformers as Universal Computation Engines

Kevin Lu, Aditya Grover, Pieter Abbeel, Igor Mordatch



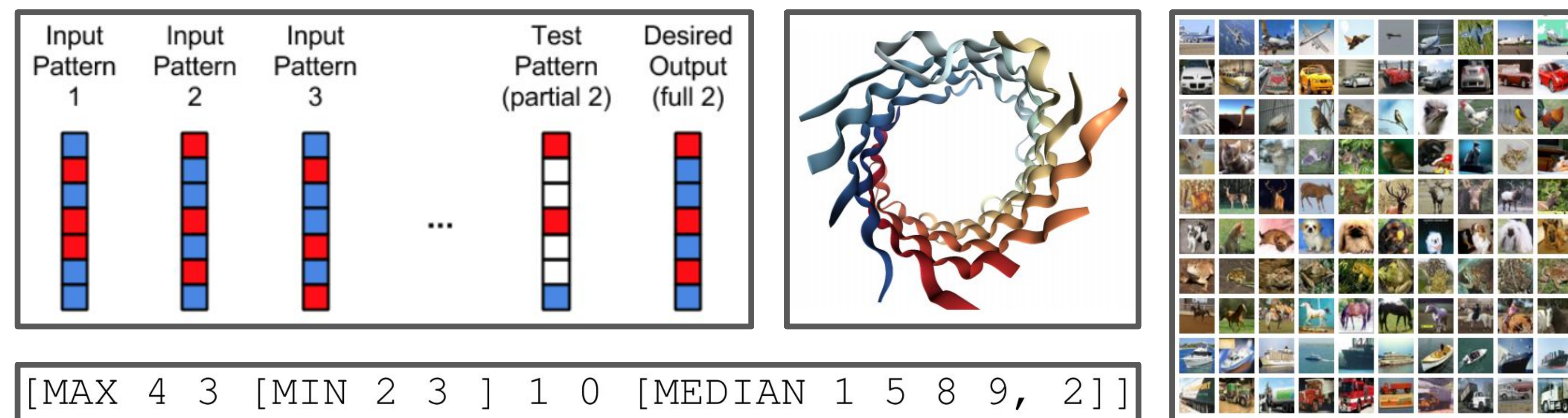
FACEBOOK AI



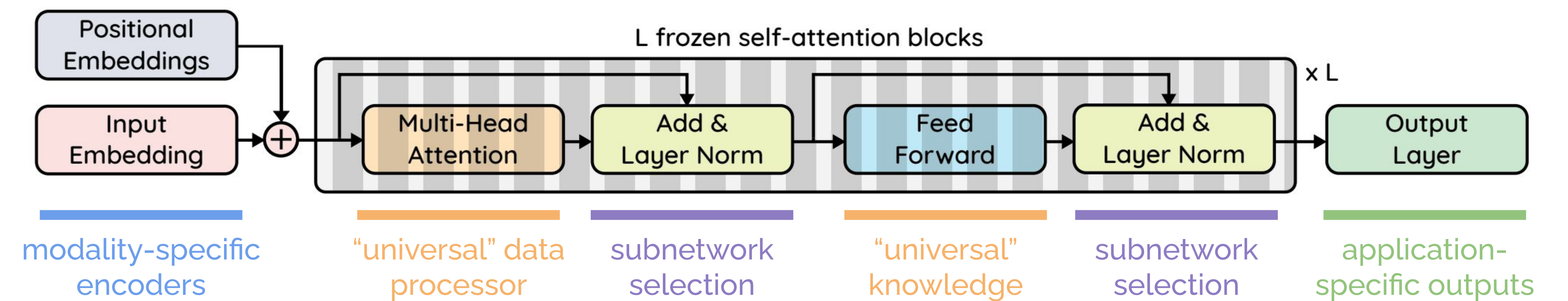
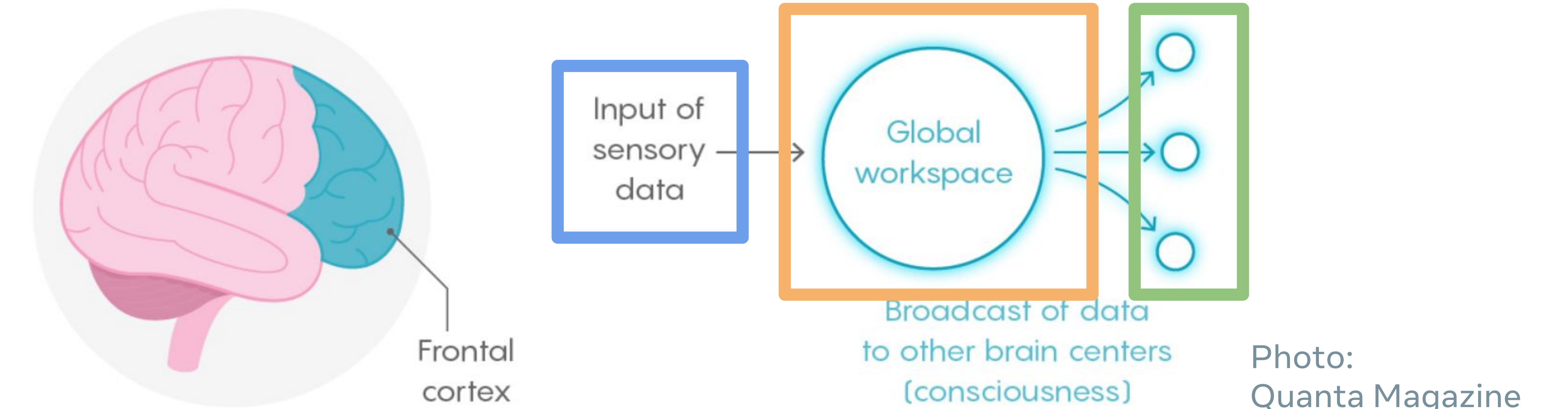
## Can the learned self-attention mechanism generalize across modalities?

We take GPT-2 and freeze the self-attention and feedforward parameters.

Despite tokens coming from new modalities, we find the frozen model exhibits strong classification performance.



## Global Workspace Theory



## Performance on Multimodal Sequence Benchmarks

