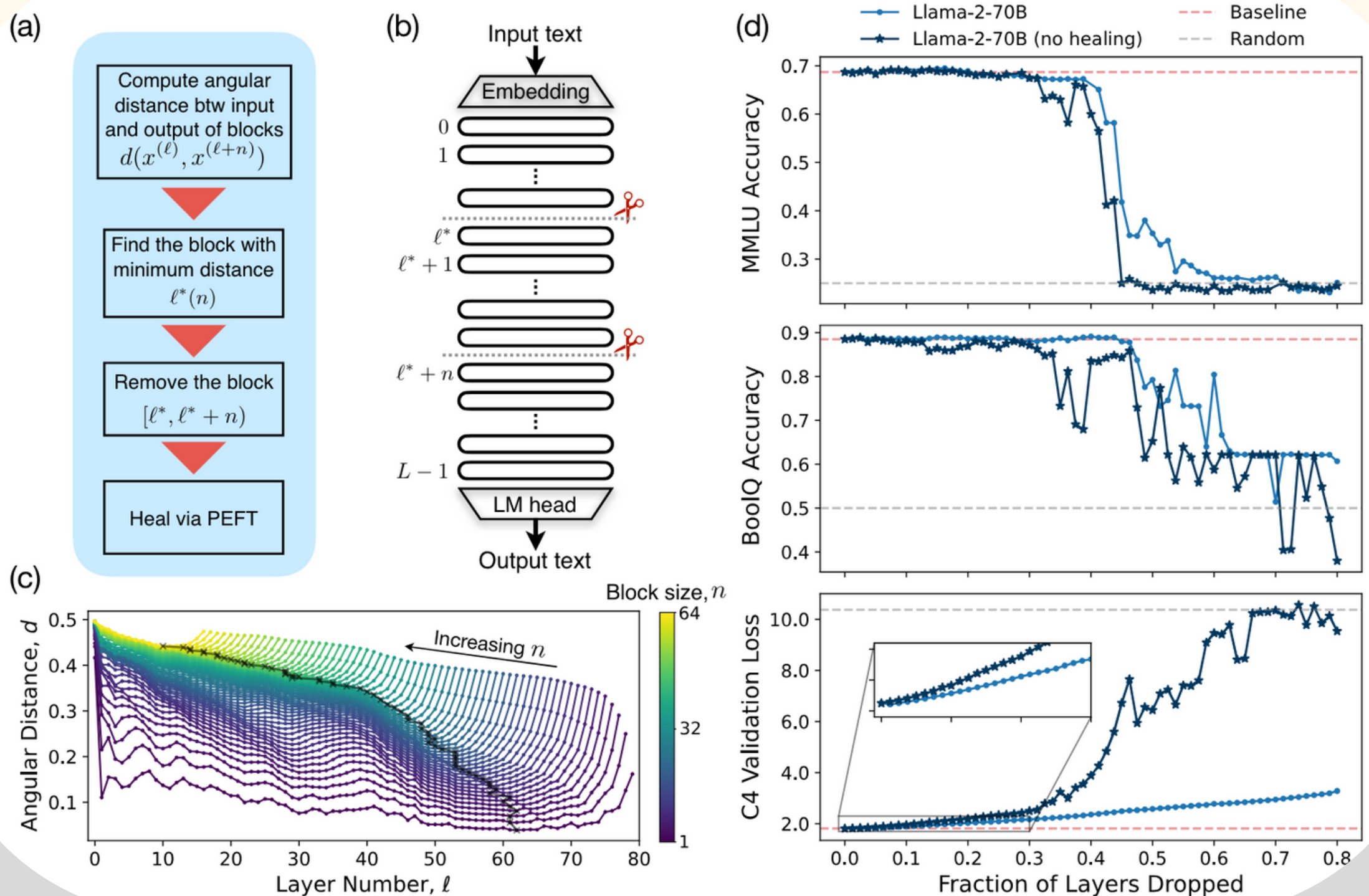


Inefficiency in LLMs

3 Papers Addressing LLM Overparameterisation

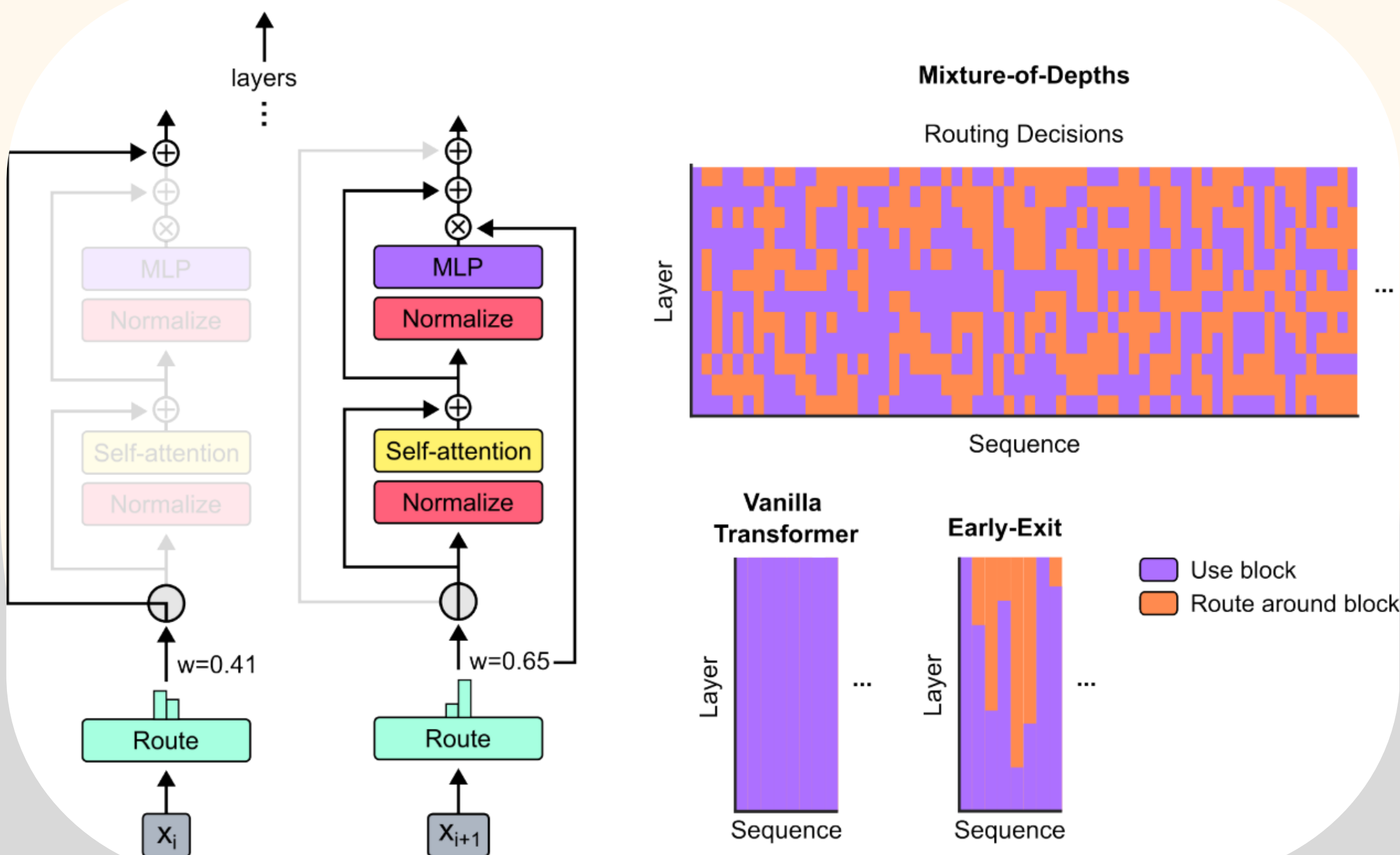
1. “The Unreasonable Ineffectiveness of the Deeper Layers” (*Meta*)



Inefficiency in LLMs

3 Papers Addressing LLM Overparameterisation

2. “Mixture-of-Depths” (Deepmind)



Inefficiency in LLMs

3 Papers Addressing LLM Overparameterisation

3. “Simplified Transformer Blocks” (ETH Zurich)

