

# 论文阅读

上周阅读论文时没看liujian师兄的gitee，以至于以乱序阅读论文（这导致我先阅读了拼好model llava 和 gpt2 后，返回阅读 bert 和 gpt1 两个 encoder-only 和 decoder-only。

在再一次反复理解 transformers 的 encoder-decoder 架构以及与 gemini 讨论翻译任务的前世今生后，感觉自己的大脑再一次学会了 transformers 架构。（究竟要学会几次（

另外粗读了 gitee 上之后的几篇多模态架构论文，感性地理解了一下 blip 和 llava。（？

经朋友推荐阅读了一篇模型结构相关的论文Atlas: Learning to Optimally Memorize the Context at Test Time

学习到了一些奇技淫巧：

Newton-Schulz 5次矩阵乘法来逼近矩阵的逆或正交形式？（

一些小事：

在跟朋友争论专家模型时，再次阅读了 lora 源论文，一边读一边想到了能不能用 lora 加的那个模块做不同专家，想清楚后一问 gemini 说是 MoE-LoRA。QwQ

有一种从 transformers/17 时代的老人到达了 22/23 年的近现代的感觉，现在看来，大抵此前是没入门罢。\*cry

# 其他

上周想要三天干完的JAVAEE计划爆炸了，vibe coding导致了成堆的问题，吃亿嵌长一智了www。