

# 大型模型时代下的实验

朱梦

初稿于 2025-08-17，修改于 2025-08-17

（一）大模型架构趋于同质化。随着大型模型（Large Language Model, LLM）时代的到来，各种大模型架构趋于同质化。令人反思到：99% 在小模型有效的技巧，迁移到大模型上几乎都失效。在小模型架构百花齐放的时代，实验通常包含不同数据集上的实验、不同架构上的实验。现在，就一定要补上在大模型上的实验。

（二）多大才属于大模型？那么，多大数量级的模型被定义为大模型呢？中国《人工智能大模型第 1 部分：通用要求》（GB/T 45288.1-2025）规定：“大模型参数量一般不低于 1 亿”（即 100 million）。而行业通常认为大模型的门槛是 1B（即 1,000 million）参数。

（三）显存需求。如果 FP32（全精度），那么 1 B 参数需占用 4 GB 显存（每个参数占 4 字节）。加上梯度（4 GB）和优化器状态（4 GB~8 GB），总显存占用约 12 GB~16 GB。因此，显存至少 12 GB。