DeepSeek r1 闭门学习讨论｜ Best Ideas Vol 3

From：Shixiang

To： Shixiang Friends

「Best Ideas 闭门讨论会 Vol.3」聚焦在引爆全球 AI 社区的 DeepSeek r1，本篇纪要是我们对闭门会上参与讨论的嘉宾成员的观点的总结，不代表任何具体个人及机构观点立场。

I. DeepSeek

1. DeepSeek 有好口碑的原因在于是第一个把复现 MoE、o1 等发出来，胜在做的早，但能不能做的最好，空间还很大，和新挑战在于资源有限，只能把有限的资源放在最亮眼的地方，但后续可能没有精力去做得更好，比如 MoE，这个团队的 research 能力、团队文化还是很好的，如果再给 10、20 万张卡，可能能做出更好的事情。

2. DeekSeek 从 preview 到正式发布这段时间，长上下文能力提升很快。DeepSeek 的 Long context 10K 用非常常规的方法就能够做到。

3. DeepSeek 肯定没有 5 万张卡，公开信息说有 1 万张老的卡，可能有 3 千张禁令之前的 H800，DeepSeek 很注重合规，所以卡应该很少。美国用 GPU 的方式太粗放了。

4. DeepSeek 把所有精力都放在了一个很窄的点，把后续很多东西都放弃了，比如安全、多模态等，不是单纯在服务人，而是做智能本身，可能也是成功的关键因素。

5. DeepSeek 有一个文章是由文生图做图生文做耦合的学习。

6. 量化就是 DeepSeek 的商业模式。幻方就是上一轮 machine learning 的产物。DeepSeek 最重要的事就是 push 智能。钱和商业化的优先级都不高。中国需要有几个领先 lab 来做探索能 beat OpenAI 的东西，智能要走的时间很长，今年又开始分化，肯定要有新东西出来。

7. 单从技术角度，DeepSeek 作为黄埔军校对人才扩散有很大作用。

8. 美国的 AI Lab 商业模式也不好，AI 今天确实没有什么好的商业模式，后面可能需要跑通。梁总是有抱负的，DeepSeek 不在乎形态，往 AGI 走就是了。

9. 梁总是 DeepSeek 最核心的人，和 Sam 不是一类人，梁总是很懂技术的。

10. 读完 DeepSeek 论文的感受是，很多都是节约硬件开销的 tech，在比较大的几个 scaling 方向上，DeepSeek 的技巧可以把成本降下来。

11. 长期不会对算力有影响，但短期大家会想怎么把 AI 做的更加有效率一点。需求还是很强的，各家都是不够用的状态。

12. 做投资，都选择最高级的组合，但现在觉得大家一起磨合好，能力也能慢慢变高级了，挖走一个人是否能打破优势组合是一个问题，现在看对于 DeepSeek 的影响可能不是特别大。

13. 市场上钱还是多，核心是文化组织，DeepSeek 和字节的 research culture 比较像，比较 fundamental，文化好不好的衡量标准在于是否有足够的钱和长期性，有比较重要的商业模式才能有长期性的文化，这两家公司的商业模式都非常好。

14. DeepSeek 为什么能追这么快？

    1) Reasoning model 的需求是更高质量的数据和训练。如果是长文本、多模态，从 0 开始追一个闭源模型会更困难，但纯 reasoning 模型本身的架构没有大动，reasoning 是一个更好追的方向。

    2) r1 能追的快的原因可能在于任务没有特别难，RL 只是让模型选的更准，r1 没有突破 Consensus 32 的效率，同时花了 32 倍效率，相当于把原来并行做探索改成串行了，没有提高智能的边界，只是变得更加容易了。

II. DeepSeek 出圈的影响

1. DeepSeek 的出圈让外界意识到了中国的 AI 很强。以前外界认为中国的 AI 进展落后美国两年，但 DeepSeek 表明其实差距在 3-9 个月，甚至某些方面更强。

2. 有可能导致美国的政策对中国的政策更加不利，但历史上封锁的东西，能被突破的都会很卷，美国的封锁可能给 AI 多三年窗口期。

3. DeepSeek、小红书等公司也受到美国 VC 关注，中国资产的重组值得关注。

4. DeepSeek 做的事大概率是在不利用 H800 或者 A800 算力的前提下，用纯国产来做，如果能做成，会有很大影响。

5. DeepSeek 不是突然爆发的，这次 r1 结果很漂亮，触及到了美国从上到下的核心圈。

6. DeepSeek 是站在巨人的肩膀上，但探索前沿需要的时间和人力成本还是要高很多，r1 并不代表以后的训练成本会同时降低。

7. 中国作为追赶者可以发挥在 engineer 的能力，中美在算力的 gap 会越拉越开的，AI 探索者还是需要更多的算力，中国怎么用较少的算力做出成果，从而有一定的抵御能力甚至做的更好，可能是未来中美 AI 格局的推演。

8. 模型的核心差别在于下一个愿景是什么，而不是技术。

1) 中国今天还是在复现技术方案，reasoning 是 o1 提出的，差距在于谁能提出下一个 reasoning。无限长度的 reason 可能是一个愿景。

2) 如果不了解最大技术的痛点，而选择用蒸馏的技术去避免了解，那在下一代技术提出的时候，可能会掉进坑里，比如千问可能因为蒸馏太多，就掉坑里了，千问就想是不是对过程进行监督，所以尝试了一下过程监督，但直接用结果监督更合适。若直接用结果监督，前一个阶段的 SFT 就不能蒸馏太多 data。

III. SFT

1. DeepSeek 最大的震撼是不需要 SFT 了，但只是在推理层面，推理外可能还是需要的，但需要讨论是不是提出了一个新的范式或架构，使得对数据的利用效率更高了或者模型迭代速度更快。

2. DeepSeek 证明了用 SFT 做蒸馏有很大好处。DeepSeek r1 的第三步骤只做了 SFT，最后一步 alignment 再用了 LHF。r1 本质是 SFT 训练出来的，特殊的是数据是用 LHF 训练出来的模型生成的，说明不需要用特别复杂的方法，只要有足够好的方法，只需要用 SFT 蒸馏就行，GRPO 的本质在于 base model 得足够聪明，一个 prompt 生成用了 16 个 generation，得尝试几次才能大概率有正确的答案。不错的 base model 加上可以 verify，是 r1 提供的思路，math 和 coding 就是比较容易 verify 的。

3. r1 - Zero 没有用 SFT 就出现了 CoT 的过程，CoT 会越来越长，SFT 更像是一个辅助手段，没有 SFT 也能产生，有了 SFT 能很快生成。

4. 现在很多小模型厂商可以用 SFT 去蒸馏大模型，效果会很好，但也没有在 r1 的过程中完全被抛弃。无限长的 CoT 是一台图灵机，是可以解决问题的，但 CoT 本质上只是中间搜索结果，用一种优化的方式去不停 sample potential output，可能会输出正确结果，然后让模型往更可信的方向去推。本质上是模型为了得到这样的结果，必须要做一些 computation，CoT 是 computation 中间必须经过的中间输出。

5. 模型不是真的和人一样在搜索，只是作为模型图灵机，中间会输出，DeepSeek 有做 Long-to-short CoT 的一些提升，CoT generation 的时候也会把超长的 CoT 去掉，猜测最后发布的版本可能是用了更加 clean 的 CoT。

6. SFT 的数据种类有几种：一个是冷启动的数据，更像是给模型一个很好的策略，给一个比较好的初始化，这样能做的探索更好，RL 中有一个优化目标是和原策略更接近；另一种数据是做了 RL 之后，生成很多 data，再加上别的数据，再在 base model SFT，本质上每个 domain 有自己的 data processing pipeline 之类的，这个数据的能力是从 base model 来的，蒸馏是无损的，把多个 domain 放到一起可能会有泛化。

7. 不确定 r1 这个过程的数据效率怎么样。猜测 OpenAI 针对数据效率也做了类似的事情，比如 fine tuning。r1 第三阶段没有用 RL 做出来的模型作为 base 去训练，而是去生成了数据，再去 SFT 得到 r1，数据包含 600K 的 reasoning data 和 200K non-reasoning data。第二阶段的模型可能在 example 的 domain 之外但仍然需要某种 reasoning 的场景下，可能也能展示解题能力，从而得到 reasoning data。而 non reasoning data 是 V3 SFT data 的一部分，是让 V3 脑补出了一个 CoT。800K 的数据还是挺小的，挺有效率的。

IV.数据

1. Scale.AI 不一定会失败，现在需要在各种 domain 上做 RL，比较常用的是 math 和 coding，还是需要 expert 来标注，但数据标注可能会更复杂，但市场会存在。

2. 在 training 上，多模态数据几乎看不出效果，或者说成本太高了，今天还没有任何证据说有用，未来机会可能比较大。

3. DeepSeek 在数据标注上非常重视，特斯拉的标注成本是中国的自动驾驶的 20 倍。特斯拉的机器人的动作是找的小脑非常健康的人做的标注，丝滑程度很好，而中国找的人的丝滑程度很差。

V. 蒸馏

1. 大模型和小模型能力是不匹配的，从大模型往小模型进行蒸馏是真的蒸馏，teacher to student，如果从完全不会中文的模型蒸馏各种中文数据，性能可能会下跌。但实际上蒸馏小模型确实有很明显的性能提升，r1 蒸馏出来后的模型再做 RL 会增长很多，因为是用和模型不匹配的数据做出来的。

2. 蒸馏的坏处是模型 diversity 下降，影响模型上限，无法超越最强的模型。但短期看，蒸馏也是一条路线。

3. 用蒸馏会有一些 hack，早期一般在 instruction 调过的模型做 RL，这个阶段模型会呈现出的特征是，先生成没有用的想法，然后最后突然答对，原因在于很多 RL 的 hack 非常隐晦，模型可能在预训练的时候把很多问题给背了，明面上是在思考，其实只是在靠近背的题。如果不做标注就蒸馏，那现在做 RLVR 的时候，就会导致模型会用更简单的方式解决，而不是去思考这个问题 OpenAI 也没有解决。可能是这一代技术的缺陷。

4. 长期来说，通过走 shortcut 的方式，而没有自己通过愿景去想怎么做技术方案，而是直接复现，中间可能会有不知道的坑。比如在这一代技术 long context 没有质变的前提下，解决问题的上限可能会被限制。r1-zero 可能是一个正确的方向，从头就做 r1-zero 或不通过类 o1 的数据启动可能更好。照着别人的技术方案可能不太好，希望更多探索。

5. 其他模型用蒸馏也能得到较好的结果，未来可能就会区分 teacher、学生，当好学生也是一种可以的商业模式。

6. 在蒸馏和技术路线上，r1 带来的震撼不如 AlphaGo，但在商业上，出圈能力比 AlphaGo 要好很多。

1) 蒸馏分两个阶段，如果只是蒸馏 o1 或者 r1，而没有建立自己的体系和 verifiable reward，会导致大家越来越依赖蒸馏，但通用领域是不可能蒸馏的，因为 reward 无法得到，以及在蒸馏过程中特殊的 CoT 怎么得到。而且第一阶段的蒸馏都有痕迹，用 OpenAI 蒸馏的模型可能遗留了 OpenAI 大量的退火痕迹，为什么 zero 能够在纯 RL 阶段上获得这样的能力，这是和基础模型在退完火之后具有反思能力是有直接关系的。

2) 不太相信纯互联网的数据而不经过退火的模型能做到这样的行为，因为互联网上几乎没有高质量数据。

3) 目前可能只有几个 top Lab 在探索到底需要多少退火阶段的数据和数据配比。蒸馏与否都是 RL 算法的一种，SFT 是行为模仿，是无限的强化学习，但只做 SFT 的上限很低，而且会损害多样性。

7. 一级市场上的创业公司看见 DeepSeek 还是很激动的，如果后续 DeepSeek 还能继续迭代，对于不是大的上市公司来说，使用 AI 上会有非常大的灵活性，DeepSeek 还蒸馏了几个小版本可以在手机上用起来，如果这个方向被证明，对于很多 AI 应用会提高天花板。

8. 蒸馏很重要的是确定目标是什么，OpenAI 是没有数据蒸馏的，要超过 OpenAI 是肯定不能做蒸馏。

9. 未来可能模型需要像人类一样学会跳步回答，在固定 context 长度下，能否提高 performance 上限。

VI.Process Reward

1. Process Reward 不一定不行，但 Process Reward 可能容易被 reward hack，也就是模型没学到什么，但能把 reward 做的很高。如果解决数学问题，用模型生成 1000 个 generation，可能就是没有 1 个能靠近正确答案，那用类似 RLVR 的方式是没有办法训练到任何东西的，如果这时候有个还可以的 process reward，可能能接近正确方向，过程分也是有帮助的。要看解决问题有多难、过程 reward 有多可靠等。

2. 过程分在 PRM 估算中，如果和真实有偏差就很好 hack。过程监督理论上是可能的，问题在于 process 的力度，以及基于 process 力度怎么给到 reward，现在结果监督也是用抽取出来的答案去做匹配，各家也没有很成熟的让模型打分而不 hack 的方案，模型自己迭代是最容易 hack 的。标过程也不难，可以枚举的，只是大家没有做，可能是一个有前途的方向。

3. 过程监督上限是人，人很多是想不到的。结果监督才是模型的上限。

4. AlphaZero 比较有效的原因在于棋局终局的时候是可以做输赢判断的，而且整个 reward 是可以根据胜率计算，但是 LLM 不知道最后不停生成能不能给出答案，有点类似遗传算法，上限可能更高，但也有可能 hack 不到。

5. AlphaGo 到 AlphaZero 的一个优势是围棋的规则是固定的，现在模型从 math 和 coding 开始就是因为比较容易验证，验证的方法是不是足够好会影响最后 RL 的质量。规则得足够完善，不然模型会去 hack，模型能满足规则，但生成的结果不是想要的。

VII. 探索者 VS 追赶者

1. AI 类似阶跃函数，现在做追赶者的算力需求少了 10 倍。追赶者的算力成本一直不太高，但探索者还是要训很多模型，大家对于新算法和架构的探索不会停止。阶跃函数背后其实是有很多人投入了很多，所以算力投入还是会一直往前，还会有很多人投在产品上。除了 reasoning 之外，还有很多方向也很费卡。探索者花费很多卡可能大家看不到，但没有这么多花费，可能不会有下一个阶跃。也有很多人不满足架构、RL 方法，会不断往前推进。

2. 在探索方向的时候，花 1 万张卡的效果不一定比 1 千张卡好，但可能会有一个门槛，即如果只有 100 张卡，那大概率做不出来，因为迭代一次方案的时间太长。

3. 物理学的进步分成学校里的研究者和工业界的实验室，前者需要探索多个方向，不要求回报，后者更关注效率提升。

4. 探索者和追赶者角度，小公司卡很少，就需要考虑效率，而大公司考虑的是怎么更快的得到模型，很多在 2 千卡集群上能提高效率的方法在万卡是不 work 的，大家会更考虑稳定性。

5. CUDA 生态优势在算子的多和全，而华为等国内公司突破的时候是找了一些常用的算子，有后发优势，假如拥有 10 万张卡，在决定资源投入的时候，做领先者的成本很高，做追赶者效率更高，该如何抉择。国内下一个追赶的方向是什么，比如多模态，因为海外 GPT-5 一直迟迟没有出来。

VIII. 其他公司为什么没有用 DeepSeek 的方法？

1. OpenAI 和 Anthropic 之前没有做 DeepSeek 的方向是一个公司聚焦方向的问题，OpenAI 和 Anthropic 可能觉得把现有算力投入其他地方会更有价值。

2. 相比大厂，DeepSeek 可能因为没有在多模态上做事，而是集中在语言，所以能做出成果。大厂的模型能力不弱，但得低调，不能发太多。现在多模态不是很关键，智能来源主要是语言，对于提升智能没有帮助。

IX.25 年 bet

1. 模型在 25 年会发生分化。最诱人的愿景是 push 智能的边界，可能有很多突破的路径，方法可能会发生变化，比如合成数据、别的架构。

2. 25 年首先关注新的架构，除了 transformer 之外能不能找别的，现在已经有了一些探索，可以降低成本，在降低成本的同时也可以探索智能的边界；其次，RL 的全部潜力还没有发挥出来；产品上，大家关心 agent，还没有被大规模应用。

3. 25 年多模态可能会出现能 beat ChatGPT 形态的产品。

X. 模型路线

1. r1 和 V3 带来的低成本、高效果，说明这是一个方向，和另一个扩硬件、涨参数的方向是不冲突的，国内是受到限制只能走前者。

2. 第一，DeepSeek 是从 base model 逼出来的，还是遵循 Scaling Law，第二，从蒸馏角度，DeepSeek 蒸馏还是先大后小，对于越做越大的闭源模型是好事，第三，对技术发展中，还没有出现反规模指标，如果出现，那对于 Scaling Law 可能是一个比较大的打击，而且开源模型的所有东西都可以在闭源模型做一遍，同时还可以降低成本，对于闭源模型也是利好。

3. 在 Meta 复现 DeepSeek 的过程中，目前还没有特别影响 infra 或者长期 roadmap 的地方出现，长期来说除了探索边界之外，也要考虑成本，只有成本更低，才能有更多的玩法。

XI.开发者 & 应用者

1. 开发者是否会从闭源模型迁移至 DeepSeek？目前看还没出现大批迁移，因为领先模型的 coding 指令遵循能力是比较有利的，但不确定这一优势在未来是否会被攻克。

2. 开发者角度来说，Claude-3.5-Sonnet 是做了 tool use 专门训练，对于做 agent 非常有利，但 DeepSeek 之类模型暂时没有提供，但 DeepSeek 带来的空间很大。

3. 对于大模型应用者，DeepSeek V2 就已经满足了所有需求，r1 速度提高了，没有带来特别大的额外价值，但开启深度思考的时候，以前能答对的题目现在反而错了。

1) 应用者选择模型的时候会用工程方法把问题简化，25 年可能是一个应用年，各行各业会使用现有的能力做，可能慢慢会到一个瓶颈了，因为日常可能用不到那么聪明的模型。

2) 现在 RL 是解决了有标准答案的问题，并没有比 AlphaZero 做更多突破，甚至更简单，蒸馏解决了标准答案的问题，有标准答案后用 RL 的方法去训练时可以得到很好的效果，这是为什么现在蒸馏或者 RL 能很快突破的原因。

4. 人类对智能的需求是远远被低估的，比如癌症问题、SpaceX 上的隔热材料都还没有被解决。现有的任务是自动化的问题，还有很多问题，对未来增量的爆发非常乐观，智能是不能停下来的。

XII. 开源 VS 闭源

1. DeepSeek 不仅是中国 VS 美国，而是开源 VS 闭源。

2. 有可能会导致 OpenAI 等把好的模型藏在后面，但 DeepSeek 拿出来之后，其他 AI 公司好的模型可能也藏不住了，但领先的模型都没发布。

3. DeepSeek 成本上做了很多优化，Amazon 等还没有看到因此做出的改变，还是按照既定的计划做，目前是一个共存的状态。开源和闭源模型并不矛盾，高校和小 Lab 应该会优先选择 DeepSeek，不会对云厂商有竞争，因为云厂商对开源、闭源都是支持的，生态不会改变，目前也是共存状态。DeepSeek 在 tool use 等上面还没有像 Anthropic 这么成熟，以及后者已经花了很多时间在 AI 安全上，DeepSeek 如果长期希望得到欧美市场的认可，是需要考虑的。

4. 开源对整个市场的 margin 是有控制的，如果开源能做到闭源的 95%，那如果闭源太贵，那完全就可以用开源来做，如果开源和闭源能力差不多，那对闭源是一个很大的挑战。

XIII. OpenAI Stargate 500B 叙事与算力需求变化

1. DeepSeek 的出现让大家开始质疑 NV 和 OpenAI 最新的 500B 叙事。训练资源问题目前还没有清晰判断，OpenAI 的 500B 叙事是给自己加救命稻草。

2. 对于 OpenAI 500B 基础设施投入的事情是存疑的，因为 OpenAI 是商业公司，如果涉及举债，那可能是有风险的。

3. 500B 是一个很夸张的数字，可能会分 4、5 年去执行。因为 leading 的角色是软银和 OpenAI，前者是资金，后者是技术，软银现在账上的资金没有办法支持 500B，而是用手上的资产去做抵押，而 OpenAI 本身资金也不是很充沛，其他更多是技术参与方，而不是资金提供方，因此要完整实现 500B 是有挑战。

4. OpenAI 500B 的算力是有道理的，在探索阶段，试错成本很高，人力和投资成本都很高，但因为路线是不明确的，从 o1 到 r1 可能也不容易，但至少知道最后是怎么样的一个结果，中间的特征词也可以观察到，可以一开始就对着别人的最终形态去做，比较有方向感。而如果是在前线探索下一代，是最费资源的，而追赶者不需要承担探索，但永远只是追赶。如果 Google、Anthropic 在探索的领域做成功了，可能就会成为最前沿的那家公司

5. Anthropic 把所有的 inference 都换成 TPU 或者 AWS 的事情是既定事实。

6. 国内公司原来受困于算力，现在证明了潜在的技术空间是非常大的。对于更加 efficient 的模型，可能不需要特别大的卡，可以提供相对定制化的芯片，可以在 AMD、ASIC 芯片上提供适配，从投资角度，英伟达壁垒非常高，但 ASIC 也会有更大的机会。

7. DeepSeek 的事情和算力没有太大关系，更多让美国觉得中国比较厉害，比较有效率，英伟达的软肋不在 DeepSeek，只要 AI 还在发展，英伟达就能发展，英伟达的优势在生态，这是靠时间积累的。技术在快速发展的时候，生态就很重要，真正危机在于技术成熟后，类似电力，变成标准品，大家会关注做产品，就会有很多 ASIC 芯片出来做特定场景的优化。

XIV. 二级市场

1. 短期上对股价有影响，pretrain 需求增速放缓，post-train 和 inference scaling 还没有足够快地 scale up，在叙述上会有一个 gap，对于短期交易确实会有影响；

2. DeepSeek 更多是 FP8，美国是 FP16，DeepSeek 所有都是基于有限算力工程能力的提升，对于算力高效的使用是最大亮点。周五 DeepSeek 在北美有巨大的发酵，扎克伯格对 Meta 资本支出给了更高的预期，但英伟达和台积电都是跌，只有博通是涨的，DeepSeek 在短期情绪上对股价、估值有压力，但长期还是看好。二级会担心英伟达从 H 卡到 B 卡的转换上会有一些 air pocket，再加上 DeepSeek 的压力，短期会有股价承压，但可能是长期看更好的机会。

3. 短期在 DeepSeek 在训练上会可能会有体现，比如英伟达的股价，但这是一个增量市场，潜力很大，长期来看，AI 才刚开始，如果 CUDA 还是大家喜欢的选择，那硬件增长空间还是很大的。

4. DeepSeek 短期对美国 AI 圈冲击大，对二级的算力相关公司，甚至能源公司有压力，但长期叙事会继续。