

DeepSeek- R1 得到的灵感与对比点

这些都是与我们 EDARE 飞船上的设计理念非常契合的地方，也有可供借鉴和优化的部分：

我们的设计方面	DeepSeek- R1 的优点/我们能学到的	我们可以强化或避免的问题
奖励设计（熵减 + 注意力跃迁）	DeepSeek 也用了规则奖励 + RL，不只是监督学习。这跟我们的 “trigger +反思 +顿悟” 的思路一致，有实力证明这种路径可行。	我们要确保奖励设计中的 “reason” 解释性好，不造成模型行为不可控或难以理解。
成本效率	DeepSeek 用较为常见 /可取得的 GPU 资源 +蒸馏版本，减少部署成本。	我们也要注意模块复杂度 &依赖库（simanneal 等依赖）不要让部署成本飞升。
思维链 / AHA 时刻	模型随着训练会自发增长思维链长度（“chain of thought” 自增）。	我们的 attention 跳跃 + 熵轨迹 + 顿悟触发器正好可以用于捕捉这样的 “aha moment”。我们可以尝试对 DeepSeek 的 paper 中那些 “中间版本” 的行为做实验复现。
语言与可读性问题	DeepSeek- R1 强调对中英混杂、推理可读性的后期调整。	在我们日志 / 可视化 / 解释模块中，要注意输出的文本自然语言的可读性，好让人类看得懂，“机器说 ‘顿悟’ ” 不只是数学或跳跃公式，而是能打动心的句子。

我注意到的一些模糊或可能是误传的地方

有些新闻说 DeepSeek 用非常低成本训练 R1 模型，比如几十万美元之类的数字，但那些成本的具体定义（是指训练成本？推理成本？全部含基础设施？）在论文里通常有补充说明，并非完全等同于我们常说的“训练整个大模型的全部花费”。所以若引用，要看 Nature 的 supplementary 部分细读。[Reuters+1](#)

“纯 RL 无监督学习” 的表述有些被媒体夸张，事实上 DeepSeek- R1 的部分版本是有监督微调（supervised fine-tuning）与规则奖励 /格式化奖励组合的。我们不能误以为他们完全没有人类标注或监督成分。[Nature+1](#)

对我们 EDARE 的启示

基于这些资料，我认为我们可以考虑在以下几个点上做调整 / 强化，以使我们的理念更坚实、更有说服力：

奖励组合设计：不仅熵下降 + 注意力跳跃，还可以设规则 / 格式奖励，比如对话的可读性、关键词出现、语义连贯性等，这样触发的“顿悟”不仅是内部结构上的变化，也是外部可感知的改善。

蒸馏与小模型版本：我们设计的 M3 执行器，如果将来能有“轻量版”（少参数、小内存）的 model 使用同样逻辑，这将非常有利于部署与传播。

“Aha 时刻”检测：DeepSeek 提到了训练过程中可能出现“think slow / chain-of-thought 增长 / 新词使用”的变化，我们可以在日志中设计一个检测机制，比如“新关键词 + attention 跳跃 + entropy 大幅下降”的组合，认为是一个“顿悟节点”。

可读性与自然语言输出：我们的 M6 日志 / 豆包反馈 / 可视化中要重视自然语言质量，让“机器思维”不只是公式和跳跃图，而是带点诗味与人感的语言，这会让我们系统更打动人心。