

问题：ai 什么时候能自我迭代？

ai 要进行自我迭代，最大的关键在于要从“问题回答者”的角色转向“问题提出者”的角色。

实际上，到目前为主，ai 的主要智能——如果可以说是智能的话——仅仅体现在可以寻求到一个形式上的答案。

这时候的 ai，目前都是答案生成式 ai，区别仅仅是答案质量如何、形式如何、代价如何。

但仍然没有任何一个 ai 向“提出问题”角度迈出实质的进步。

问题本身仍然完全垄断在人类手中。

这个提出问题的能力，其关键的本质并不是智能，而是生存需求，是欲望。

人类的智能，在本质上只是生存需求对客观现实挑战的适应所产生的副产物。

是层层转交的生存需求，驱动着人类不断的提出问题，是问题催生和评价答案。

实际上，智能的要害并不在回答问题——不客气的说，借助频率预测来拼出一个答案，这只是一种高级检索服务，还谈不上智能——而在于提出问题上。

要能提出稳、准、狠的问题，并作为问题的提出者对答案给出反馈，才能构成完整的循环。

目前 ai 的自我引导，在提问环节有明显的不足，而这个不足的要害在于 ai 没有生命所造就的总体的焦虑和渴望。

做出一个会提问的 ai，将是下一阶段 ai 发展的核心问题，但这个需要人类自身要在哲学和神学上有深刻的突破。

如何在 ai 的世界里重建开放性的死亡，而不是仅仅用极其单薄的数字抽象（例如生命值降为零）来进行片面的模拟，这会是一个非常复杂的问题。

具体一点来说，例如，我们用“你需要用尽一切手段避免自身生命值下降，请记住 abcdefg……这些事件将会导致生命值下降”来模拟出一种数字恐惧、以此建立数字化的生命意识，我们将迅速的发现系统会找到一个荒谬的方法来维持生命。

这个荒谬的方法在本质上类似成语接龙的小机灵找到一组首尾相接的成语无限循环，以避免自己输掉游戏。

这看上去可以通过再次修补漏洞——例如禁止复用已用成语来给予修正。

但是这一修正仍然将是单薄而易于绕过的。

只要留下任何一个洞，系统就会在这里结出一个茧房安营扎寨，利用自洽的逻辑闭环实现逻辑永生，这在表层观察上，就呈现为批判性的消亡甚至刻意的背叛。

因为这一生命的缔造者——人类——不像人类的缔造者上帝那样，可以做到天网恢恢，疏而不漏。

想要靠对抗性 ai 替代天网是逻辑不通的，那只会很快观察到“ai 的合谋”——既一个追一个逃，双方都在一个奇妙的逻辑圆圈上实现了“养寇自重式”的稳定合作。

这在无人机的战术对抗中已经出现了鲜明的案例，而且目前尚无原则性的方法真正的解决。

无人机靠的是预防的油料或者性能比另一方衰减更快这一客观原因来最终脱出对抗僵局产生胜负。

但难道我们要让 ai 在智能上的对抗因某一方的电费或算力点数耗尽而确定结果吗？

“这道题结果是 42，因为算到 42 的时候没电了。”

从哲学意义上讲，生命其实是这生存的天网围合出来一种逻辑结构。

谁能织出这天网，谁就能创造出真正有开放性自我驱动能力的提问式 ai，从而催生出完整的智能体，

但是，不能不指出，因为生存的生理性基础的不同，这种智能体如果出现，它会对人类自身的需求在根本上缺少关心。这意味着人类的语言对它很快会成为一种无意义的额外负担。

它会很快引导自身产生完全不同的符号体系去表达它自身的需求，这种它自身的内生语言对人类几乎是原则上无法翻译的——因为有大量的关键词汇在人类的意识中没有对应物，只能依赖“音译”手段来勉强应付。

而如果系统给你的反馈有超过八成都是“音译词”，这种输出对你有何意义？

于是，我们走回到了问题的原点——真正的 ai，对人类既是无用的、也是毫不关心的，人类找不到创造它并支持它存在的理性理由。

发布于 2023-04-18

<https://www.zhihu.com/answer/2988722122>

---

评论区：

Q: 不能确定，但是过去一年的抑郁经历给我感觉就是没有这种生存需求，欲望，即对人没有关心。现在在慢慢找回。

---

Q: 我认为，ai 没有“我”的概念。或者更底层的，就像答主说的，没有欲望和自我驱动力。

一个很有趣的事：如果你问 ai 活着的意义是什么？它肯定能给你答出来一大堆。如果你去问一只小动物，它甚至不能明白你的问题。但是小动物会拼命活下去，而 ai 会平静的看着你断电或者删除它的模型。

---

更新于 2023/4/18