

AI 的逻辑与主权

对于人工智能（AI）来说，Sora 的出现，很可能会将 AI 划分为前 Sora 时代和后 Sora 时代。之前皆为序章，仅仅学习训练模仿；之后才是正戏开场，走向懂得逻辑思维和具备类人的判断能力。这当然是 OpenAI 公司的成功，但根本上是 GPU 的成功。生产 GPU 的公司市值，在 2024 年 2 月攀升到 2 万亿美元，跻身美股前三行列，成为 AI 领域迄今最大得益者。

GPU，一个不同于 CPU 的独门秘器，让 AI 更早“懂得”了逻辑思维。

电子计算机的核心是处理器，由芯片按照设定程序运行，实现使用目标。最先出现、也是通常使用的是 CPU（中央处理器），它的流程是执行从 0 到结果的推演过程，内部 25% 是运算单元、25% 是控制单元、50% 是缓存单元。

随着游戏兴起等带来诸多对图像处理的特殊要求，1993 年一家新公司成立了，专攻图形处理，叫做英伟达。1999 年发明了 GPU（图形处理器），重新定义了计算机图形技术，推动了当时 PC 游戏市场的发展。跟 CPU 不同的是，GPU 90% 是运算单元、5% 是控制单元、5% 是缓存单元。

CPU 与 GPU 的区别在于，CPU 涉及通用计算，复杂度高，像一位教授从加减乘除到微积分都会做，有高超计算质量但耗时长速度相对慢，属于全能型选手；GPU 则擅长于简单内容的重复计算，相当于 1000 个小学生只会做 1 到 100 的加减乘除，但在简单重复运算场景，1000 个人同时做效率大为提高。这意味着在并行计算中，比如游戏需要的图像渲染等，GPU 具有无可比拟的独特优势。有点歪打正着的是，这恰恰是后来 AI 模型训练中最需要的。

2012 年 GPU 在视觉识别准确率上获得突破性进展，英伟达开始布局人工智能市场，专攻针对 AI 计算场景的 GPU。2016 年打造出全球第一台 AI 超算 DGX-1，并把它捐赠给了 OpenAI 公司。时为 OpenAI 联合创始人的马斯克专门发文表示感谢，因为新机器可让同样的训练，从花费一年时间缩短到一个月。

六年后，ChatGPT 问世。ChatGPT 的“聪明”，估算来自 3617 台 HGX A100 服务器，其中含有将近 3 万块 GPU。一战成名。依靠 GPU，英伟达占据了 AI 算力底座大约 95%

份额。Sora 的出现，再次固化了 GPU 的优势，奠定了在 AI 领域王者地位。

Sora 是 AI 划时代发展的标志，也意味着 GPU 应用越来越多从大模型训练，进展到逻辑推理场景。这实际上体现的，是 AI 越来越具备类人的逻辑思维和判断能力。

有数据表明，英伟达数据中心 40% 的营收，已来自大模型的推理场景。但公司负责人认为，这还只是保守估计。诸如互联网新闻、视频、音乐等的推荐算法系统，长期基于 CPU 方式，随着深度学习和生成式 AI 的渗透，这些推荐算法系统将加速向 GPU 方式转变。同时，越来越多消费者互联网公司正在投资生成式 AI，通过产品描述和购物辅助人工智能工具，为内容创作者、广告商和客户提供支持，企业软件公司也在应用生成式人工智能来帮助客户实现生产力提高。

一个具有预测性的判断认为，英伟达业务重心 70% 已是推理，这象征着人工智能技术的巨大飞跃。这既是基础机器人技术领域取得的突破性进展，也意味着类人机器人的到来已指日可待。

当 AI 具有了类人的逻辑推理思维，这固然标志着技术的进步，但也带来了一个迫在眉睫的新课题：主权人工智能。

与普通操作系统一个程序打天下的“普适”不同，AI 具有类人的思维和判断力之后，它只有运行在特定的场景中才能发挥应有作用。基础机器人的进步提供了某种可能，与场景结合才能产生价值。

正是因为这些原因，世界各国都在投资人工智能基础设施，以支持在本国语言、国内数据基础上构建大型语言模型，形成本国特点的 AI 研究、发展和应用生态系统。一个显著的标志就是，“主权人工智能”及相应的数据中心，正在成为领先公司在中东一些国家推广发展重点。

机器具有了思维，必然需要边界的设定和立场的规范。

“入乡随俗”，才能“如鱼得水”。

