AGI 十年内会出现吗

制品 · 持续 · 预判

AGI 十年内会被宣布，但绝不会被公认

报告出具时间：2026/6/29
报告来源：See The Essence

目录

你有没有经历过这种情况：一个人跟你说“十分钟后到”，但你知道他家离这里至少半小时车程。他说的不是你的时间，是他的时间——一个在当下这个情境里对他有用的时间。

关于 AGI 十年内会不会出现，情况很像。很多人把这个问题当技术预测来讨论，但它更像一个精心制造的宣称，功能不是准确描述未来，而是有效地组织现在。理解了这一点，整个争论的样貌就不一样了。

这个宣称是怎么被造出来的，得分开看。

最先要搞清楚的是：谁在说，为什么说。最近几年喊出“10 年内 AGI 会来”的主力，不是独立评估机构，而是一群正在烧钱冲刺这件事的人。他们的公司估值、融资能力、对顶级工程师的吸引力，都和这个时间表绑在一起。如果你的估值取决于投资者相信你正在造一座通天的桥，那告诉他们“桥快通了”就不是一种判断，是一种维持运转的必需品。这不是撒谎，是结构性激励在起作用。

然后是定义权的游戏。AGI 这个词听起来很硬，拆开看全是软的。“像人类一样理解并完成各种任务”——哪个“人类”？三岁小孩还是诺奖得主？“理解”是要有意识，还是能答对题就算？这个问题吵了几十年，从来没有共识。而这恰好给了研发机构最大的操作空间：没有裁判，运动员可以自己画终点线。你只需要把自己的系统跑过一组精心挑选的基准，再开一场发布会，“AGI 已实现”就可以宣布在任何你需要宣布的季度。十年前有团队在特定测试上宣称“达到人类水平”，没人当真——不是因为技术不够好，是因为标准没被大家认。这个困境不会因为模型变大就自动消失。

这个宣称一旦被造出来，传播的路也很值得看清楚。ChatGPT 让普通人第一次觉得“这东西好像真懂点什么”，直观印象形成了。资本市场上，投资人需要 AGI 故事来给几千亿估值找理由，分析师的报告被媒体摘几句转发，最后沉成一种“大家都知道在加速”的氛围。更微妙的是，一些本来最谨慎的安全研究机构开始认真讨论“如果十年内 AGI 出现，我们需要做什么准备”。他们的职责要求他们为最坏情况做准备，不代表他们认为最坏情况最可能发生。但在传播里，“情景推演”滑成“预测”，“预测”滑成“既定事实”。你听到的是“连最谨慎的人都这么说了”，其实你可能听错了。

到这里，一个核心洞察就浮出来了。

AGI 的定义权空缺，不是一个有待解决的技术困难，而是整个游戏能够继续的前提。如果今天世界上存在一个公认的、无法作弊的通用智能测试，任何机构宣布“通过了”都会被独立验证确认或打脸。对研发方来说，这是不可接受的控制权让渡——等于把自己的里程碑交给别人画。所以定义之争不会在十年内解决。十年后最可能的局面是：有人开了发布会，说自己的系统已经是 AGI；另一个阵营立刻发论文，列出十条理由说它不是。双方各拿各的证据，围观的人各信各的。这不是噪音，这件事本身就是“AGI 十年内出现”这句话真正指涉的事件。

那该怎么看这件事?

与其纠结“会不会来”，不如盯几个能校准判断的问题。第一，如果有人宣布 AGI 实现了，它在哪项测试上系统性崩溃？看的不是它通过了什么，是它通不过什么。一个儿童能做的因果推理任务如果它永远做不对，那次宣称就值得怀疑。第二，支撑进展的资源曲线还能延续吗？训练成本涨得比可用资本和算力快的话，指数趋势会自然弯曲。芯片产线、电网改造这些东西的时间常数以十年为单位，它们不跟着论文发表的节奏走。第三，有没有出现真正能约束核心决策者的治理机制？如果一家公司的 CEO 被董事会解雇、再靠员工造反和投资人压力复辟——那任何关于“我们会负责任地释放 AGI”的承诺，听听就行。

记住一件事：智能可能不是一根可以无限往上爬的单一标尺。一个系统可以在语言上超神，在物理常识上还不如一只蟑螂。如果短板恰恰是现实世界里最要命的那一块，那么“十年内到达”这句话本身，可能是对的——只是到达的方式和所有人想象的可能不一样。