Sam Altman 萨姆·奥尔特曼

« Back to blog

« 返回博客

Sam Altman 萨姆·奥尔特曼

Posted 7 days ago

发布 7 天前

June 11, 2025 at 5:12 AM

2025年6月11日上午5:12

The Gentle Singularity 温柔的奇点

We are past the event horizon; the takeoff has started. Humanity is close to building digital superintelligence, and at least so far it's much less weird than it seems like it should be.

Robots are not yet walking the streets, nor are most of us talking to AI all day. People still die of disease, we still can't easily go to space, and there is a lot about the universe we don't understand.

机器人还没有在街上行走，我们大多数人也还没有整天和人工智能对话。人们仍然会死于疾病，我们仍然无法轻易进入太空，宇宙中还有很多我们未解之谜。

And yet, we have recently built systems that are smarter than people in many ways, and are able to significantly amplify the output of people using them. The least-likely part of the work is behind us; the scientific insights that got us to systems like GPT-4 and o3 were hard-won, but

will take us very far.

然而，我们最近构建的系统在很多方面都比人类更聪明，能够显著提升使用者的产出。这项工作中最不可能的部分已经过去；成就 GPT-4 和 o3 等系统的科学洞见来之不易，但将引领我们走得更远。

AI will contribute to the world in many ways, but the gains to quality of life from AI driving faster scientific progress and increased productivity will be enormous; the future can be vastly better than the present. Scientific progress is the biggest driver of overall progress; it's hugely exciting to think about how much more we could have.

In some big sense, ChatGPT is already more powerful than any human who has ever lived. Hundreds of millions of people rely on it every day and for increasingly important tasks; a small new capability can create a hugely positive impact; a small misalignment multiplied by hundreds of millions of people can cause a great deal of negative impact.

从某种意义上来说，ChatGPT 已经比任何人类都强大。数亿人每天都依赖它，并完成日益重要的任务；一个小小的新功能就能带来巨大的积极影响；一个小小的偏差乘以数亿人，就能造成巨大的负面影响。

2025 has seen the arrival of agents that can do real cognitive work; writing computer code will never be the same. 2026 will likely see the arrival of systems that can figure out novel insights. 2027 may see the arrival of robots that

can do tasks in the real world.

2025 年，我们会看到能够进行真正认知工作的代理的出现；编写计算机代码将不再是以前的样子。2026 年，我们可能会看到能够得出新颖见解的系统的出现。2027 年，我们可能会看到能够在现实世界中执行任务的机器人的出现。

A lot more people will be able to create software, and art. But the world wants a lot more of both, and experts will probably still be much better than novices, as long as they embrace the new tools. Generally speaking, the ability for one person to get much more done in 2030 than they could in 2020 will be a striking change, and one many people will figure out how to benefit from.

将会有更多的人能够创造软件和艺术。但世界对软件和艺术的需求远超于此，只要专家能够接受新工具，他们可能仍然比新手更优秀。总的来说，到2030年，一个人能够完成的工作量将远超2020年，这将是一个显著的变化，许多人将会找到从中受益的方法。

In the most important ways, the 2030s may not be wildly different. People will still love their families, express their creativity, play games, and swim in lakes.

从最重要的方面来看，2030年代或许不会有太大的不同。人们仍然会爱家人，发挥创造力，玩游戏，在湖里游泳。

But in still-very-important-ways, the 2030s are likely going to be wildly different from any time that has come before. We do not know how far beyond human-level intelligence we can go, but

we are about to find out.

但在一些仍然非常重要的方面，2030年代很可能与以往任何时期都截然不同。我们尚不清楚人类智能究竟能超越到何种程度，但我们即将找到答案。

In the 2030s, intelligence and energy—ideas, and the ability to make ideas happen—are going to become wildly abundant. These two have been the fundamental limiters on human progress for a long time; with abundant intelligence and energy (and good governance), we can theoretically have anything else.

2030年代，智力和能源——创意以及将创意付诸实践的能力——将变得异常丰富。长期以来，这两者一直是人类进步的根本制约因素；有了丰富的智力和能源（以及良好的治理），理论上我们什么都可以拥有。

Already we live with incredible digital intelligence, and after some initial shock, most of us are pretty used to it. Very quickly we go from being amazed that AI can generate a beautifully-written paragraph to wondering when it can generate a beautifully-written novel; or from being amazed that it can make life-saving medical diagnoses to wondering when it can develop the cures; or from being amazed it can create a small computer program to wondering when it can create an entire new company. This is how the singularity goes: wonders become routine, and then table

stakes.

我们已经生活在令人难以置信的数字智能之中，在最初的震惊之后，大多数人都已经习以为常。很快，我们就会从惊叹人工智能能够写出优美的文字，转变为好奇它何时也能写出一部优美的小说；或者，从惊叹它能够做出救命的医疗诊断，转变为好奇它何时能够研发出治愈方法；或者，从惊叹它能够编写一个小型计算机程序，转变为好奇它何时能够创建一家全新的公司。奇点就是这样：奇迹变成常规，然后成为筹码。

We already hear from scientists that they are two or three times more productive than they were before AI. Advanced AI is interesting for many reasons, but perhaps nothing is quite as significant as the fact that we can use it to do faster AI research. We may be able to discover new computing substrates, better algorithms, and who knows what else. If we can do a decade's worth of research in a year, or a month, then the rate of progress will obviously be quite different.

我们已经听到科学家说，他们的生产力比人工智能出现之前提高了两到三倍。高级人工智能之所以引人注目，原因有很多，但或许没有什么比我们能够利用它更快地开展人工智能研究更重要。我们或许能够发现新的计算基础、更好的算法，以及其他未知的成果。如果我们能在一年或一个月内完成十年的研究成果，那么进步的速度显然会截然不同。

From here on, the tools we have already built will help us find further scientific insights and aid us in creating better AI systems. Of course this isn't the same thing as an AI system completely autonomously updating its own code, but nevertheless this is a larval version of

recursive self-improvement.

从现在开始，我们已经构建的工具将帮助我们获得进一步的科学洞见，并帮助我们创建更好的人工智能系统。当然，这与完全自主更新自身代码的人工智能系统不同，但无论如何，这都是递归式自我改进的雏形。

There are other self-reinforcing loops at play. The economic value creation has started a flywheel of compounding infrastructure buildout to run these increasingly-powerful AI systems. And robots that can build other robots (and in some sense, datacenters that can build other datacenters) aren't that far off.

还有其他自我强化的循环在起作用。经济价值的创造已经启动了一个飞轮，不断增加基础设施的建设，以运行这些日益强大的人工智能系统。而能够构建其他机器人的机器人（从某种意义上说，能够构建其他数据中心的数据中心）的实现也并非遥不可及。

If we have to make the first million humanoid robots the old-fashioned way, but then they can operate the entire supply chain—digging and refining minerals, driving trucks, running factories, etc.—to build more robots, which can build more chip fabrication facilities, data centers, etc, then the rate of progress will obviously be quite different.

如果我们必须以老式的方式制造出第一批一百万个人形机器人，但随后它们可以操作整个供应链——挖掘和提炼矿物、驾驶卡车、运营工厂等——来建造更多的机器人，从而可以建造更多的芯片制造设施、数据中心等，那么进步的速度显然会大不相同。

As datacenter production gets automated, the cost of intelligence should eventually converge

to near the cost of electricity. (People are often curious about how much energy a ChatGPT query uses; the average query uses about 0.34 watt-hours, about what an oven would use in a little over one second, or a high-efficiency lightbulb would use in a couple of minutes. It also uses about 0.000085 gallons of water; roughly one fifteenth of a teaspoon.)

随着数据中心生产自动化，智能成本最终应该会趋近于电力成本。（人们常常好奇 ChatGPT 查询耗电多少；平均每次查询耗电约 0.34 瓦时，相当于烤箱一秒多一点的耗电量，或高效灯泡几分钟的耗电量。此外，它还会消耗约 0.000085 加仑水；大约相当于十五分之一茶匙的水。）

The rate of technological progress will keep accelerating, and it will continue to be the case that people are capable of adapting to almost anything. There will be very hard parts like whole classes of jobs going away, but on the other hand the world will be getting so much richer so quickly that we'll be able to seriously entertain new policy ideas we never could before. We probably won't adopt a new social contract all at once, but when we look back in a few decades, the gradual changes will have amounted to something big.

技术进步的速度将持续加快，人们几乎能够适应任何事物，这一点将持续下去。虽然会有一些非常艰难的部分，比如整个行业的工作岗位将消失，但另一方面，世界将以如此快的速度变得更加富裕，以至于我们能够认真考虑过去从未考虑过的新政策理念。我们或许不会一下子就采纳一项新的社会契约，但几十年后回首往事，这些渐进式的变革将产生巨大的影响。

If history is any guide, we will figure out new things to do and new things to want, and

assimilate new tools quickly (job change after the industrial revolution is a good recent example). Expectations will go up, but capabilities will go up equally quickly, and we'll all get better stuff. We will build ever-more-wonderful things for each other. People have a long-term important and curious advantage over AI: we are hard-wired to care about other people and what they think and do, and we don't care very much about machines.

如果历史可以借鉴，我们会发现新事物、新需求，并迅速掌握新工具（工业革命后的工作变动就是一个典型的例子）。期望会上升，但能力也会同样快速提升，我们都会拥有更好的产品。我们将为彼此创造更加美好的事物。人类比人工智能拥有一个长期重要且令人好奇的优势：我们天生就关心他人以及他们的想法和行为，而我们不太关心机器。

A subsistence farmer from a thousand years ago would look at what many of us do and say we have fake jobs, and think that we are just playing games to entertain ourselves since we have plenty of food and unimaginable luxuries. I hope we will look at the jobs a thousand years in the future and think they are very fake jobs, and I have no doubt they will feel incredibly important and satisfying to the people doing them.

一千年前，一位自给自足的农民会看着我们许多人所做的工作，说我们所做的工作是虚假的，认为我们只是在玩游戏自娱自乐，因为我们有充足的食物和难以想象的奢侈品。我希望一千年后，我们看待这些工作时，会认为它们非常虚假，而且我毫不怀疑，从事这些工作的人会感到无比重要和满足。

The rate of new wonders being achieved will be immense. It's hard to even imagine today what we will have discovered by 2035; maybe we will go from solving high-energy physics one year to beginning space colonization the next year; or from a major materials science breakthrough one year to true high-bandwidth brain-computer interfaces the next year. Many people will choose to live their lives in much the same way, but at least some people will probably decide to "plug in".

新的奇迹将以惊人的速度不断涌现。今天甚至很难想象到2035年我们会有什么样的发现；或许我们今年解决了高能物理难题，明年就开始了太空殖民；又或许我们今年在材料科学上取得重大突破，明年就实现了真正的高带宽脑机接口。许多人会选择以类似的方式生活，但至少有些人可能会选择"融入生活"。

Looking forward, this sounds hard to wrap our heads around. But probably living through it will feel impressive but manageable. From a relativistic perspective, the singularity happens bit by bit, and the merge happens slowly. We are climbing the long arc of exponential technological progress; it always looks vertical looking forward and flat going backwards, but it's one smooth curve. (Think back to 2020, and what it would have sounded like to have something close to AGI by 2025, versus what

the last 5 years have actually been like.)

展望未来，这听起来难以理解。但经历这一切或许会让人印象深刻，但并非不可避免。从相对论的角度来看，奇点是一点一点出现的，融合也是缓慢发生的。我们正在攀登技术进步的长弧；它总是向前看是垂直的，向后看是平坦的，但它是一条平滑的曲线。（回想一下2020年，如果2025年能实现接近通用人工智能，那听起来会是什么样，而过去五年的实际情况又如何呢？）

There are serious challenges to confront along with the huge upsides. We do need to solve the safety issues, technically and societally, but then it's critically important to widely distribute access to superintelligence given the economic implications. The best path forward might be something like:

在巨大的机遇和挑战面前，我们也面临着严峻的挑战。我们确实需要从技术和社会层面解决安全问题，但考虑到其经济影响，广泛普及超级智能的使用权也至关重要。最佳的前进方向或许是这样的：

1. Solve the alignment problem, meaning that we can robustly guarantee that we get AI systems to learn and act towards what we collectively really want over the long-term (social media feeds are an example of misaligned AI; the algorithms that power those are incredible at getting you to keep scrolling and clearly understand your short-term preferences, but they do so by exploiting something in your brain that overrides your long-term

preference).

解决对齐问题，这意味着我们可以强有力地保证人工智能系统能够学习并采取行动，实现我们长期真正想要的目标（社交媒体信息流就是错位人工智能的一个例子；支持这些算法的算法非常善于让你继续滚动并清楚地了解你的短期偏好，但它们是通过利用你大脑中某种超越你长期偏好的东西来做到这一点的）。

2. Then focus on making superintelligence cheap, widely available, and not too concentrated with any person, company, or country. Society is resilient, creative, and adapts quickly. If we can harness the collective will and wisdom of people, then although we'll make plenty of mistakes and some things will go really wrong, we will learn and adapt quickly and be able to use this technology to get maximum upside and minimal downside. Giving users a lot of freedom, within broad bounds society has to decide on, seems very important. The sooner the world can start a conversation about what these broad bounds are and how we define

collective alignment, the better.

然后，专注于让超级智能变得廉价、广泛可用，并且不会过于集中于任何个人、公司或国家。社会具有韧性、创造力，并且适应能力强。如果我们能够驾驭人们的集体意志和智慧，那么尽管我们会犯很多错误，有些事情会变得非常糟糕，但我们能够快速学习和适应，并能够利用这项技术获得最大的好处和最小的坏处。在社会必须决定的宽泛范围内赋予用户很大的自由度，这似乎非常重要。世界越早开始讨论这些宽泛的界限是什么，以及我们如何定义集体共识，就越好。

We (the whole industry, not just OpenAI) are building a brain for the world. It will be extremely personalized and easy for everyone to use; we will be limited by good ideas. For a long time, technical people in the startup industry have made fun of "the idea guys"; people who had an idea and were looking for a team to build it. It now looks to me like they are about to have their day in the sun.

我们（整个行业，不仅仅是 OpenAI）正在为世界构建一个大脑。它将极其个性化，人人皆可轻松使用；我们将受到好想法的限制。长期以来，初创企业的技术人员一直嘲笑“创意人”；那些有想法却想找团队来实现它的人。现在在我看来，他们即将迎来辉煌的一天。

OpenAI is a lot of things now, but before anything else, we are a superintelligence research company. We have a lot of work in front of us, but most of the path in front of us is now lit, and the dark areas are receding fast. We feel extraordinarily grateful to get to do

what we do.

OpenAI 现在有很多事情要做，但首先，我们是一家超级智能研究公司。我们面前有很多工作要做，但前方的道路大部分已经照亮，黑暗区域正在迅速消退。我们非常感激能够从事我们所做的事情。

Intelligence too cheap to meter is well within grasp. This may sound crazy to say, but if we told you back in 2020 we were going to be where we are today, it probably sounded more crazy than our current predictions about 2030.

廉价到无法计量的情报唾手可得。这听起来或许有些不可思议，但如果我们在2020年告诉你我们将会达到今天的水平，那可能比我们现在对2030年的预测更疯狂。

May we scale smoothly, exponentially and uneventfully through superintelligence.

希望我们能够通过超级智能平稳、指数级、平静地扩展。

A posthaven user upvoted this post.

A posthaven user upvoted this post.

A posthaven user upvoted this post.

2066 visitors upvoted this post.