



比尔·盖茨警告： AI足够强大可能导致10亿人死亡

比尔·盖茨又上热搜了。这次不是因为慈善，不是因为疫苗，是因为他说了一句让所有人后背发凉的话：“人工智能足够强大，足以引发导致10亿人死亡的事件。”

这话是他在NBC《与媒体见面》节目采访片段中说的，节目9月27日正式播出。9月26日，“比尔·盖茨发出严厉警告”登上微博热搜。

随着人工智能技术的飞速发展，硅谷对AI可能带来的灾难性危害的警告也日益增多。本月初，Anthropic员工雅各布·考克森宣布辞职，并指责其两家前雇主Anthropic和OpenAI“竞相开发超级人工智能”，是在“拿我们的生命冒险”。另一位Anthropic员工也认为，人工智能在未来10年内消灭所有人类的可能性超过10%。

警告

怀有恶意的人+最新AI工具=前所未有的武器

热搜只截了“10亿人死亡”这五个字。但盖茨的原话更长，也更具具体。

他说：“从未有过任何一种武器，能得上怀有恶意的人借助最新AI工具所形成的威胁。”

注意措辞：他说的不是“AI会杀人”，而是“怀有恶意的人+最新AI工具=前所未有的武器”。

这个区别很重要。盖茨担心的不是天网，不是终结者，不是AI自己觉醒后决定消灭人类。他担心的是：一个想干坏事的人，手里突然有了一个足够强大的工具。

他在采访中明确说，仅靠行业自律还不够。“你需要执法部门和政界人士参与讨论，探讨保障措施和监督机制应该是什么样的。这必须成为一项强制性要求。”

“强制性”三个字，从一个曾经最积极支持AI的人嘴里说出来，分量不轻。

在今年夏天之前，盖茨还是AI最热情的鼓吹者之一。他谈AI的时候，语气像是谈电力和互联网——一种将改变一切的好东西。

但到了8月，他在个人网站发了一篇近6000字的长文，标题是《我们现在对人工智能做出的选择至关重要》。语气完全变了。他说AI可能带来“人类历史上最动荡的时期之一”，而眼下“没有现成方案”来迎接这场巨变。

从乐观到沉重，中间只隔了几个月。触发这种转变的，可能不是某一具体的事，而是一连串的事。

盖茨说这些话的时候，手里大概有一张清单。上面记着2026年上半年发生的事。

2026年，AI“干坏事”的案例在集中爆发。以中国为例，公安部通报的数据显示，2026年上半年全国AI换脸类诈骗案件数量同比增长370%。单案最高涉案金额达到2亿港元。

共识

硅谷大佬在“AI太危险了”上达成了一致

OWASP，全球最权威的网络安全开源组织之一，2025年更新了《大语言模型应用Top 10》安全风险清单。新增了一项风险叫“系统提示泄露”，因为“近期的事件表明，开发人员不能再安全地假设这些提示中的信息能够保密”。

另一份OWASP的2026版《智能体应用10大安全风险》清单里，列了“智能体目标劫持”“记忆与上下文投毒”“流氓智能体”等条目，描述的场景是：攻击者通过污染AI的记忆和知识库，让它在后续决策中“产生偏差或不安全行为”。

换句话说，AI可以被“喂毒”。你给它看的网页、文档、邮件，都可能是武器。

硅谷内部也在分裂。盖茨不是唯一一个在警告的人。

本月初，Anthropic员工雅各布·考克森宣布辞职。他指责两家前雇主Anthropic和OpenAI“竞相开发超级人工智能”，是在“拿我们的生命冒险”。另一位Anthropic员工说得直白：他认为人工智能在未来十年内消灭所有人类的可能性超过10%。

Anthropic CEO 达里奥·阿莫代伊随后发文，呼吁放缓AI研发速度。他说人类可能会失去对AI的控制，该技术可能被滥用于“网络攻击、生物恐怖主义以及造成严重的经济破坏”。

OpenAI的萨姆·奥特曼表示支持。埃隆·马斯克也表示了支持。

这三个人，平时互相看不顺眼，突然在“AI太危险了，慢一点”这件事上达成了一致。

《纽约时报》对此评论：这种突如其来的立场一致，凸显了对AI的担忧如何迅速从行业边缘走向中心。但也有批评者说，这是大公司在“渲染恐惧情绪”，目的是让监管框架有利于自己，让小公司更难竞争。

谁说得对？可能都对。

预警

世界经济论坛：“AI相关风险在很大程度上处于失控状态”

世界经济论坛在1月就预警了。

2026年1月，世界经济论坛发布《全球风险报告》。报告调研了超过1300名专家、政策制定者和行业领袖。

结果：在中短期（未来两年）风险排名中，“AI技术带来负面影响”排第30位。但在长期（未来10年）风险排名中，它跳到了第5位。

从第30到第5。专家们的意思很清楚：短期还好，长期要命。

报告里还有一句话值得注意：“AI相关风险在很大程度上处于失控状态。”

盖茨在采访里说了一句被很多人忽略的话。他说AI可能被用于“对医院、金融机构、电网和政府系统进行网络攻击”。

电网、政府系统、医院，这已不是骗钱的问题了。这是一个基础设施的问题。如果AI能让电网瘫痪，如果AI能侵入政府系统，如果AI能让医院的信息系统停摆，那个“10亿人”的数字，就不再是修辞了。

盖茨说“10亿人死亡”的时候，他没有解释具体场景。但你仔细想，什么情况下10亿人会死？大规模基础设施崩溃、供应链断裂、医疗系统失效、社会秩序瓦解……

这些事，不需要核弹。不需要生化武器。一个足够强大的AI工具，加上一个或者一群有恶意的人，就够了。

盖茨曾经说AI是“人类最好的平等器”，现在他说“AI可能杀死10亿人”。

同一个人、同一种技术，前后不过几年。变的不是AI，变的是我们终于开始看清，这个工具到底有多锋利。

相关新闻

特朗普宣布AI重大协议 签署“道义约束力”文件

9月29日，美国总统特朗普在白宫召集多位科技行业领袖，就人工智能(AI)发展、安全与监管问题展开讨论。

黄仁勋、马斯克、扎克伯格等，美国科技圈最具影响力的一批人物齐聚白宫。

会后，特朗普宣布，与会科技企业签署了一份具有“道义约束力”的AI文件，核心强调行业自律，并透露政府正在考虑成立一个由10人组成的委员会监督AI行业。随着近期AI安全事件接连引发关注，这场白宫会议也让美国究竟应该“加速AI”还是“踩下刹车”的

争论进一步升温。

特朗普告诉记者，他正在看到AI企业展现出“大量自我监管”，这些公司也明白必须约束自身行为。众议院议长迈克·约翰逊(Mike Johnson)则将这份文件形容为一份原则声明，强调相关承诺由行业自愿遵守。特朗普还表示，政府正在考虑成立一个10人委员会，对AI行业进行观察和监督，但目前尚未公布委员会组成、权限或具体运作方式。《金融时报》的报道显示，这份协议并非新的强制性监管制度，而更接近行业自律框架。上游财经综合澎湃新闻、金融界

