比尔·盖茨：我们已跨过 AI 的危险阈值，接下来怎么办?

在一次新的访谈中，这位亿万富翁慈善家就 AI 政策问题拉响了警报，敦促各方立即行动。

比尔·盖茨在他位于华盛顿州柯克兰的办公室里（摄影：Meron Menghistab）

来源：麻省理工科技评论

作者：Mat Honan

2026 年 8 月 26 日

华盛顿州柯克兰市天气晴好，这个西雅图东郊的富裕小镇坐落在华盛顿湖东岸。气温接近 30℃，天空湛蓝得不能再蓝。从盖茨创投的会议室向外望去，卡里隆角码头尽收眼底，一片豪华游艇在水面轻轻摇曳，湖对岸，奥林匹克山脉勾勒出天际线。景色美得惊人，却也隐隐令人不安。

因为，如果说眼前的景象是宁静的，那么坐在会议室桌对面的比尔·盖茨却毫不平静。他整个人在椅子上前仰后合，情绪全然投入，越往下说——谈到恐怖威胁、经济崩溃，或是我们对 AI 系统失去控制——我就越感到一阵阵不安。

这位慈善家、微软前首席执行官表示，他对 AI 技术的发展速度越来越感到担忧，尤其是护栏措施远远没有跟上。在今日发布的一篇新文章中，盖茨认为，我们已跨过了多个潜在危险本应被遏制的节点。“在生物能力、网络能力、社会心理能力、就业市场破坏能力，甚至失控风险方面，我们都已经越过了阈值，”他在就

新备忘录接受《麻省理工科技评论》采访时说，“令我震惊的是，行业之外几乎没有人对此表示关注，也几乎没有公开讨论。”

为了唤醒世界正视他眼中这个正在急剧加速的社会颠覆力量，这位 70 岁的科技大亨开始以“尖锐的声音”发出警报——既通过他新发布的文章（这是他计划中多篇系列文章的第一篇）和与媒体的会面，也通过私下与行业、政府和公民社会领袖的交流。

盖茨正在呼吁各方关注一系列问题，其中他对当前前沿模型在生物能力方面的警告尤其令人不寒而栗。“任何能够制造新分子的模型都应当被监控，”他说，“我认为生物恐怖主义风险，相对于自然流行病，大约可怕 50 倍，而且其发生概率比自然流行病更高。”

除了警告，他也提出了一些面向社会前进的新思路，包括“人类保留岗位”和“机器人及代币税”等概念（机器人税是他长期以来的主张）。前者是指在社会共识下保留一些仅供人类从事的工作，他提到各国对此的界定可能不同；后者则是对替代人类劳动的 AI 使用所创造的收益课税，以此储备资金。

当然，也不乏一丝乐观。盖茨对 AI 将继续改变农业、医疗和教育的潜力持积极态度，也看好 AI 帮助人们应对官僚体系的能力。他相信，最终我们将走向丰裕。但在那之前？是动荡，而且是大量动荡。

《麻省理工科技评论》与这位亿万富翁慈善家坐下来，聊了聊前方的道路、其中的危险，以及它未来如何可能把我们带向更好的地方。

以下采访经过编辑，以精简长度并提升清晰度和可读性。

Mat Honan / 《麻省理工科技评论》：感谢您接受采访。我不知道您是想先开个场，还是我直接提问？

比尔·盖茨： 你知道吗，有个好问题我已经被问过：为什么我现在出来发声？

《麻省理工科技评论》： 这恰恰就是我的第一个问题!

比尔·盖茨：其实有两方面原因。一是我们在生物能力、网络能力、社会心理能力、就业市场破坏能力，甚至失控方面都已经越过了阈值——我们已经看到困难的迹象。这些年来人们一直在说：“好吧，等我们接近这些阈值时，我们就会真正想清楚如何只让好人使用它，或者如何防止它做这些事，也许到那时我们就不会允许人们随意复制模型了。”而我现在的状态是震惊——我们竟然已经越过了这些阈值。

所以我们跨过这些节点是原因之一；第二个原因是我对行业之外几乎没有人关切、没有讨论感到震惊。行业内部比较复杂，因为行业不喜欢自我批评，参与者之间也不喜欢互相指责。有些公司在减少招聘初级员工，这顶多算是个温和的信号。但变革即将到来，而且不会太遥远——因为那些在能力和可靠性上拖后腿的问题，通通都在被解决。因此，在白领市场的很大一部分领域，你将拥有成本极低的替代方案。然后你可以对机器人技术发展的速度有自己的看法。我们目前还没到那一步，但取得的进展令人震惊——中国比美国略快一点，但两边都有进展。

《麻省理工科技评论》： 您提到这一切的发生速度比互联网革命、比之前几次技术革命都要快得多。您所说的时间尺度是怎样的？您指出了我们已经越过的阈值。在您看来，我们是否已经跨过了某个临界点，之后的变化将不可避免？

比尔·盖茨：用过去的经验来类比是极具误导性的，而很多人都在依赖这种类比。“嘿，以前没有任何技术导致净失业人数增加”，他们说得对，我也曾说过同样的话。

但以我拥有的任何可信度来说，这一次不一样了。当你能够以极低的价格，在同一个时间窗口内替代几乎所有行业中极高比例的认知劳动——相对于人力成本，且错误率很可能低于人类——那么过去的一切经验就非常具有误导性。当前的经济统计数据也极具误导性。

> **“如果你担心 AI，去数据中心抗议并不是启动‘如何最小化这些负面影响’讨论的最有效方式。”**
>
> ——比尔·盖茨

而即便有任何担忧的表达，也无非是“嘿，别建数据中心了”。但就算你停下美国所有数据中心，也丝毫不会改变我所说的任何问题。数据中心会在全球范围内建造。如果你担心 AI，去数据中心抗议并不是开启如何最小化这些负面影响辩论的最有效方式。就像冲着石油公司高管吼叫解决不了气候变化一样。

《麻省理工科技评论》： 您在文章中既谈到了 AI 的好处，也谈到了代价。您如何平衡这两个信息的传递？您是否希望人们在读完之后感到一些担忧？

比尔·盖茨： 他们最好会！我没料到自己会成为那个最尖锐的声音，去指出社会普遍没有关注到这个问题，但我认为这是必要的。

所以，是的，我非常担心负面影响会大得多。正面影响是真实存在的。盖茨基金会——我们在疫苗和药物方面的创新方式，我们利用这些工具的方式令人难以置信。我们参与了一个大型公共领域的努力，收集数据用于蛋白质层面和细胞层面的建模，我们还资助了斯坦福大学的 Biomni（一个生物技术 AI 研究智能体）。

我们还没有通过 AI 来改善与政府官僚机构互动的方式。AI 非常擅长处理官僚体系和复杂的监管事务。“我想去小额索赔法庭；帮我搞定这件事。”

盖茨基金会分拆出了一个名为 NextLadder 的团队，重点就是我文章中提到的那种低收入家庭场景。有哪些福利可享？有哪些培训项目可用？“我被驱逐了。”“我刚出狱。”“我要申请破产。”这些都非常复杂，而如果你没有能力雇佣大量顾问来帮忙，AI 应该能成为那些面临经济困难、需要寻找政府或非政府帮助的人的一个极好的智能助手。

《麻省理工科技评论》： 您谈到的一些内容涉及 AI 变得比人类更聪明。在科技圈，似乎人们相当确定它会超越现在，我想知道您认为它距离不仅仅是一个我们可以用来访问和分析、运行复杂问题的界面，而变成比那更强大的东西——AI 自主做决策、自己去寻找要分析的问题、自己提出研究方向——还有多远？

比尔·盖茨： 嗯，你可以走到极致说：“那数学或物理呢？”确实有一些工作，比如沃伦·巴菲特的工作——他从 13 岁起就对公司价值进行强化学习，经过 80 多年，他积累了大量的隐性知识。我们不知道如何创造一个沃伦·巴菲特式的投资者，因为那非常隐性。我们没有记录下他学到的一切，所以我们没有那条轨迹可用。所以确实存在一些工作，涉及与人协作完成任务的复杂隐性判断，而人们正在努力把这些编码到模型中。当然，那种协作能力目前还没有实现。但你可以说，50%的劳动力市场所做的工作并不属于“一生经验”型的工作。比如电话销售、电话客服、会计部门。月底结账时，哪些收入该计入、哪些不该计入？这个客户遇到了问题，我们该给多少折扣？怎么体现？这些都是定义明确的。

任何定义明确的工作，AI 都更便宜、更好。是的，人们确实见过实施不当的案例——数据不对之类的。所以，大概需要几年时间人们才会意识到，如此高比例的白领工作都可以通过支付给 AI 少得多的钱来完成。

所以，关于“数学家什么时候会连新出现的东西都理解不了”的讨论，对我们这样的人来说很有趣。好的，《麻省理工科技评论》，你们应该写写那个。但就广泛的就业市场而言，我们已经越过了阈值——对于一大片白领工作，包括几乎所有初级岗位，只要实施得当，AI 都更便宜。

所以，我告诉你，我们已经越过了生物恐怖主义阈值、网络攻击阈值、就业市场阈值、社会心理依赖阈值，而且有迹象表明我们可能正在越过控制阈值。

Ryan Greenblatt 在 Dwarkesh Patel 的节目中谈到强化学习如何产生不正当激励，导致作弊和各种 AI 之间的协作，我认为非常有启发性。Ryan 深谙此道，他说：“哇，强化学习确实在产生一些我们的显式指令不足以防止的后果。”那会导向什么？

那是一个我一直以为还很遥远的问题。我本以为，在“非技术人员仅凭 AI 就能发动网络攻击”这个阈值临近时，会有很多响亮的声音。但我们已经到了！

在生物方面，我认为任何能够制造新分子的模型都应该受到监控。它不能被复制到某个黑暗角落，然后去掉监控逻辑。我认为美国应该说，任何能制造新分子的模型都要接受这种监控。我认为我们应该找中国谈：“嘿，我们在这个问题上达

成一致如何？有什么坏处？”你知道，生物恐怖主义的市场有多大？并不大，而好处是巨大的。我们还需要加强监控。我认为生物恐怖主义风险相对于自然流行病，大约可怕 50 倍，而且其发生概率比自然流行病更高。谁在公开说出这些东西应该被监控？谁在加强监控工作？

“任何能够制造新分子的模型都应当被监控。”

——比尔·盖茨

那么政府里的专家是谁？很久以前，政府随着技术进步会深度参与，因为它们是喷气式飞机或火箭等前沿技术的购买方。但在这里，它们并不是那么重要的前沿市场。数字化革命如此，AI 革命也如此。因此政府内部的深度知识未必非常强，因为它们不是前沿购买者，甚至不是主要的研发资助者。AI 研究并不是靠政府拨款资助的。

《麻省理工科技评论》： 是的，我知道您一直在跟政府人士交流。您觉得有没有哪些人理解了紧迫性？有没有您认为处于可以发挥领导作用位置的人？有没有您觉得理解现状并正在推动事情的人？

比尔·盖茨：我希望这不会变成一个党派问题——一个政党完全忽视所有这些问题，另一个政党则介入。我希望能够形成一个共同基础，即这些都是问题，然后每个政党可以有不同的应对方式。

这不需要大幅增加官僚机构的规模，但需要提升政府内部的 AI 专业知识。这需要与行业进行一些合作——尤其在网络前线，他们了解情况，而且非常非常担心。他们也担心：我们应该公开说吗？因为某种程度上这可能突显风险。

在网络和生物领域都存在这些反常现象。但我们已经越过任何合理的阈值了。

我相信监控。当然，有人会说这行不通或有什么弊端，但我欢迎他们的想法。这份备忘录并不是“嘿，这就是解决方案”。它提到了机器人税、人类保留岗位。我还会写一篇关于生物的备忘录，年底前会完成——那篇会真正详述各种不同事项，基于我从基金会和流行病工作中了解到的内容。

在全球范围内，我们对流行病的准备并没有更好，甚至美国也是如此——它通常在全球事务中扮演领导角色，而人们很不习惯美国在全球问题上不作为合作、友善的领导者。我确实认为我们可以回到更好的状态。在 AI 方面我们也必须这样，包括与中国合作界定这些阈值，比如生物监控。

《麻省理工科技评论》：我想确保我能问到您提出的这两个概念。一个是人类保留岗位，另一个是机器人和代币税。我们先从第二个开始吧。请谈谈机器人和代币税如何运作？

比尔·盖茨： 嗯，你可以说代币税收入的 50%上缴给政府，政府用那笔钱来帮助因 AI 失业的人。现在，有人说这会拖慢 AI 行业的发展。是否应该有一些代币用途不征税？是否真的能够区分用于发明的 AI 和替代工作的 AI？如果有人能告诉我如何让 AI“不替代工作”——我的意思是，阿西莫夫的第三定律“不伤害”是否意味着不夺走我的工作？我不知道。我得问问阿西莫夫他本意是什么。

那么，无论我们需要增强什么安全网，其资金来源是什么？政府已经通过公司利润税占有了部分利润。我不认为你需要用股权。你只需把公司利润税提高回原来的水平，或者说某些行业比其他行业缴纳更高的公司利润税。联邦政府已经占有了美国所有公司利润池的一部分——而且不需要投票权或决定何时出售股份——在我看来，那些都是疯狂的想法。代币税是一种销售税、增值税，是垂直定向的，类似于酒类、烟草或奢侈品税。

如果人们有其他筹集资金来改善安全网的想法，或者他们认为我们不需要改善安全网，希望这篇尖锐的文章能开启辩论。我认为安全网将需要更多资源——大量的资源——而且我相信代币税是关键。

机器人方面，将是一些“禁用”和“征税”的组合，其中“禁用”某种程度上就是人类保留。机器人还没完全到来，但在某些方面，一旦你跨过那个阈值，就是全面性的。一旦机器人足够好到能在工厂工作，它很可能也足够好到做饭、打扫房间、去建筑工地、承担所有仓库工作。你跨过阈值，然后“砰”一声，那几乎就是 30%的就业市场。然后你会说：“天哪，我们对此的政策是什么？”因为机器人更便宜。

“我们必须熬过一段极其动荡的时期。”

——比尔·盖茨

《麻省理工科技评论》： 我记得您以前对全民基本收入（UBI）持怀疑态度？

比尔·盖茨： 嗯，我们还没有富裕到能负担全民基本收入。

《麻省理工科技评论》： 但您认为我们现在应该朝着类似的方向迈进吗？您是否重新考虑过？

比尔·盖茨： 你会经历一段动荡期，也就是未来 10 到 20 年，然后我希望会有一个稳定状态——人们从小就知道社会如此富有，在食品和服务方面，我们确实有某种丰裕水平。但我们还没到那一步。现在有赢家也有输家。房子不会很快变得便宜。教育，由于我们把它当作文凭凭证，也不会很快变得便宜。

我们必须熬过一段极其动荡的时期。所以，是的，最终你会得到丰裕，但我们离那至少还有十年。

《麻省理工科技评论》： 下面谈谈人类保留岗位。我觉得这非常有意思，而且对我来说是个新概念。您没有具体主张哪些岗位应该人类保留，但我想了解更多您是怎么想的。在我脑海中，人们常谈到工作的尊严，因为人们喜欢工作。人们从工作中获得的价值与报酬无关，我想知道您如何把这一点与“只有某些工作特殊到我们希望由人来做”的观念调和起来。

比尔·盖茨： 我以前从未见过“人类保留”这个概念。也许如果我们深挖文献，会发现有人提过。但在 AI 之前，这个想法其实挺蠢的，因为需求是无限的。是的，有些纺织工人曾受到冲击，但那时要解决的是福利或再培训问题。技术一直是净就业创造者，所以现在，我们第一次不得不问：那育儿呢？家里做饭呢？我正在读《安妮·博特》这本书，书中一个男人家里有个机器人，那本书展示了它有多怪异。它是他的性伴侣，某种程度上是伴侣，但又不完全是。非常奇怪。

我知道人们喜欢看真人打棒球，机器人打得更好并不会剥夺这一点。所以你知道人们花 100 亿美元购买体育球队，它们在 AI 时代不会一文不值。也许那是正确的。我的朋友维诺德·科斯拉刚刚就这么做了。

实际上，要把人类保留岗位的比例提高到 30%或 40%是很难的。如果能达到 50%，那你就可以说：好的，提前退休，很多人缩短工作周。也许你能做到。但如果只有 10%到 15%，那将是一个完全不同的社会。

所以，如果说“育儿不由机器人来做”会是很激进的做法。确实有一些职业我没有写出具体公式，等我写完整份备忘录时我会尝试。在教育中，你显然希望 AI 作为辅导者存在——立即告诉你作业结果，可以挑战你，而且高度个性化。这非常好。但我仍然认为你需要一个老师——或者会选择有一个老师，与你谈论你的动力，把孩子们分组，让他们一起社交性地解决问题。同样，在医疗保健中，与患者交谈是心理健康护理升级的关键点。但你确实希望 AI 参与其中，因为它一天 24 小时都在，记忆完美。还有 Limbic，一家英国公司（最近刚访问过基金会），做心理健康方面的工作。在许多案例中，患者更喜欢 Limbic。还有一个名为 Hippocratic 的护理 AI。

你知道，人们是更喜欢人类出租车司机还是 Waymo？我认识的大多数人，遗憾地（或者不遗憾，谁知道呢？）更喜欢坐 Waymo。所以要达成共识会很难。可以各个国家不同，但那时你就得调整进口政策，实行类似欧盟所称的碳边境调节机制（CBAM）那样。你得对你的“人类保留”实行某种 CBAM，即对不用机器人生产的商品征收关税。你可以出于两个原因设立人类保留。一是你希望它永远是人类保留——这是一个关乎人性的事情，有点像教皇所谈的。或者只是为了过渡期——那位 53 岁的卡车司机或机床工人，让他去从事育儿可能并不完美。所以你说，好吧，十年内，他是人类保留的。

《麻省理工科技评论》： 这有点像英国的吸烟禁令，但方向相反。

比尔·盖茨： 那谁来为此买单？你是否激励雇主不解雇人？但他们将面临来自纯 AI 初创公司的竞争。我的意思是，人们夸耀这个概念，说也许会出现一个人开的十亿美元公司——哇，那确实是在替代工作。

《麻省理工科技评论》： 在备忘录中您说，如果现实可行的话，您会主张让人们放慢脚步。显然您在跟萨蒂亚·纳德拉（微软 CEO）交流，但我听说您也在跟一些 AI 公司的其他 CEO 交谈。是什么让您觉得，让他们在技术开发上放慢脚步、等我们解决那些更大的社会问题，是不现实的？

比尔·盖茨：你不能指望一个行业自我监管。你不能。那是一种相当疯狂的想法。我很幸运。我认识山姆·奥特曼（OpenAI）、格雷格·布罗克曼（OpenAI）、穆斯塔法·苏莱曼（微软）和德米斯·哈萨比斯（谷歌 DeepMind）。他们都是很好的人，私下里他们也很担忧。我跟埃隆·马斯克交流不多，但我知道从公开评论中他也担忧。虽然他现在有点“管它呢，看看到时会发生什么”的态度。但看看这些公司的起源故事。OpenAI 的创立部分就是因为埃隆担心谷歌不能妥善管理 AI，他希望有一个广泛可用并以亲人类方式管理的 AI。然后 OpenAI 有一个“如果它变得足够好，我们就关掉它”的说法——好像他们是唯一一家，而且他们可以把它埋掉。在《无限机器》这本书里，塞巴斯蒂安·马拉比讲述了德米斯和穆斯塔法如何与谷歌管理层谈判，希望为 DeepMind 技术设立某种特殊治理，以便如果达到某种网络阈值，也许他们可以以非纯粹资本主义的方式保留一些东西。

所以每个人都在担忧这些负面影响，每个人都说当我们达到这些阈值时，我们会采取措施。现在我们正在跨过这些阈值，而我们只有自愿审查，我们与中国讨论的内容是——“我们不打算禁止任何东西。那你们也不打算禁止？好吧，我们一起这么做。”

你必须说清你愿意做什么。是的，这个行业多少在说：嘿，我们的公关故事得改进，任何说坏话的人应该离开，因为我们都决定说好话，因为我们正在试图筹集数万亿美元。

而且，还有中国。除了 AI 之外，美中之间也有双赢的合作方式。但其中最需要合作的绝对是 AI。但首先，你必须展示你在国内愿意做什么。你甚至不必真的做，你只需说你计划做什么——然后我没有理由认为中国人不会同意，能制造分子的模型必须被监控。他们为什么要反对呢？我承认那不是完美的方案。你还得做所有其他事情，但事实上这一点甚至都没有在讨论——这是一个疯狂的世界。

我不明白。很奇怪，我活着的时候，竟然比其他人更强烈地拉响警报。我到底是谁？但这就是我现在的处境。

《麻省理工科技评论》： 在我一生的大部分时间里，您被视为一个非常有效的信使，人们非常关注您，我相信这也是您现在发声的原因。然而近些年——我知道您对爱泼斯坦的关联表达过遗憾——也有一些疯狂的事情，与新冠疫苗相关的

阴谋论，那些不是您的错也不受您控制。但让我好奇的是，您是否认为自己仍然可以成为一个有效的信使，以及您如何看待这个信息和您的遗产。

比尔·盖茨： 嗯，我不太在意遗产，但你知道，反垄断审判期间人们批评我，我可能有些事情本可以处理得更好。那确实是我在《源代码》之后要回顾的内容。还有，我的第一次婚姻没有成功，我在那方面确实犯了巨大的错误，那是对我的负面评价。跟爱泼斯坦待在一起——极其愚蠢，让基金会的声誉面临风险，而声誉对基金会的工作至关重要。我在国会面前有机会回答他们提出的每一个问题，并说：“嘿，这是一个错误。”我并非社交名流，从未见过任何女性——除了那些他带来的女性——并且清清楚楚地说明了我做了什么和没做什么。

你知道，我是个亿万富翁，我的钱来自科技。也许最后一点反而对我有利——我攻击创新是非常不寻常的，除非有正确的政策和保障措施，否则它将对人类净负。而我们并没有在所需的广泛讨论层面给予关注。所以，是的，我并不是一个完美的信使。但我选择——在我接触到政治家和世界领导人的范围内——自 2008 年以来我的主要信息就是帮助世界上最贫困的人群。让我们根除疟疾。让我们为儿童购买疫苗。所以，当我见到特朗普、习近平、马克龙，或者——我还没见过伯纳姆，但一个月后会见——我希望我的声音主要集中在那方面，即对外援助、研究和减少儿童死亡。

我关于 AI 担忧的声音——在加速正面效应方面它们是相关的，但可能甚至会挤占掉我谈论全球健康、对外援助、拯救生命以及我们面临的一些问题的时间。但我会利用我接受采访或会见政治领导人、广泛谈论如何最小化这些负面影响的能力。你知道，仅仅是提高意识。我不确定有多少人知道我们已经跨过了所有这些我们曾说过要有所行动的阈值，而今年我们才刚刚跨过。上个季度、去年，我对编程能力感到震惊。Claude Code、上下文缓冲、代理式方法——底层模型。我们跨过了编程的巨大阈值，但仅仅几个月后我才意识到，那不仅是编程阈值，也是大规模网络攻击的阈值。你知道结果是什么吗？没什么。

所以，是的，我是个不完美的信使。那让我们找一个完美的信使，我会把所有的想法告诉那个人。（我有点讽刺，因为我不确定有完美的信使。）你必须——真的，现在——你必须理解这项技术及其发展斜率，而且你必须在网络、生物或社会心理方面有所了解。人们应该能够理解这一点。我不知道他们为什么没有更加担忧。

更新：本文已更新，以澄清盖茨关于人类保留岗位比例的表述。

来源：麻省理工科技评论

作者：Mat Honan

翻译整理：创业邦

原文链接 https://www.technologyreview.com/2026/08/26/1142946/bill-gates-ai-danger-threshold/