

看世界

航空安全不能靠奇迹

近日，在一架从阿联酋迪拜飞往以色列特拉维夫的波音客机上，170多名乘客熬过了漫长的1个小时。飞机30秒内骤降4300米，乘客与机组人员冲进驾驶舱，制服了刺伤机长的副驾驶。两名飞行员接手，把方向舵已大面积脱落的飞机迫降至沙特阿拉伯塔布克机场。

抓住操纵杆让飞机恢复平稳的乘客说，自己是从纪录片《空中浩劫》里学来的经验。航空专家把这次安全着陆称作“航空奇迹”。

事件发生于“9·11”事件25周年纪念日的19天后。当年，4架被劫持的客机共计夺走了近3000条生命，其中1架联合航空93号航班的乘客为夺回控制权冲向驾驶舱。25年后，类似的一幕重演，只不过这次幸运地发生了“奇迹”。然而，奇迹不能替代制度。真正值得追问的是：一套已经运行多年的交通系统，为什么至今依然难以建立起坚固的安全防线？

答案指向一个不太体面的事实：在一个净利润率常年徘徊在2%上下的行业里，安全支出总是最先被砍掉的那一项。德国之翼9525航班事故发生后，欧洲航空安全局推行“驾驶舱两人制”，要求一名飞行员离座时须有另一名机组成员进舱。这项“四眼原则”在2016年8月被降级为“建议”，次年4月德国宣布取消，理由是它“并未带来安全增益”。省下的安全开支，直接变成当期成本削减，利润表立刻好看；至于省掉冗余会不会出事的责任，则被悄悄推给了概率。

美国空管领域是同一种逻辑的另一副面孔。在成本压力下，联邦航空管理局今年5月把持证管制员编制目标下调到12563人。此前的测算显示，在岗人员缺口约为3500人，但按照新目标测算的人员缺口仅为1500人。凭空消失的2000人缺口，靠的不是招人，而是降低标准。

人员缺口体现在报表上只是一个数字，但在现实中却是塔台里

需要同时盯着几个扇区的酸痛眼睛，以及连续工作、频繁加班的疲惫心灵。当成本压力被率先传导至编制目标，缺人就不再只是人力问题，而是安全冗余被一点点“吃”掉的过程。

罚款机制同样在给成本逻辑让路。波音公司今年1月因2024年阿拉斯加航空737 MAX 9型客机舱门脱落事件支付的313.9万美元罚款，已是现行法律上限，对应的却是两年前的违规行为。有美国参议员直言：波音在违规涉及的两个季度内营收接近390亿美元，“此类罚款很容易被消化，可被视为经营成本的一部分，而非对危险行为的有效威慑”。当违规的预期成本远低于合规的当期支出，成本压力就不再只是压力，而是变成了一种被制度默许的“激励”。

3个案例共享同一条逻辑链：成本压力上升—安全冗余压缩—风险转移给概率。

为什么这个链条如此稳定地反复出现？这不只是职业道德问题，还是个经济学问题。

航空业的利润结构本就很脆弱。国际航空运输协会预测，2026年全球航空业净利润230亿美元，净利润率仅为2%。同时，航空业全年运送旅客约51亿人次，这意味着航空公司每运送一位乘客只能赚4.5美元。综合算下来，航企投入资本回报率为4.3%，远低于8.5%左右的资金成本。在这个利润水平下，任何一笔“可省”的支出，边际诱惑都是巨大的。

安全支出恰恰是所有支出里，看起来最容易省的那一类。无论是飞机的新旧、舱体的宽窄，还是乘务员的增减，乘客们都能一眼看出来；反倒是事关安全的关键问题，很难被非业内人士发现。一条航线的维护间隔延长了，飞机第二天照样飞；一次驾

驶舱加强程序的培训被压缩了，航班照样准点；一名空管的编制被冻结了，塔台照样运转。这些削减的“收益”是即时可见的，但成本是不确定的。

这背后的经济学逻辑并不复杂。飞行安全基金会（Flight Safety Foundation）曾分析指出，安全系统的真正功能是控制那些不可避免但必要的成本，并最小化计划外成本。换句话说，安全投入的回报不是“多赚了多少钱”，而是“少出了多少事”。但在预算编制那个时间点上，“少出的事”不会体现在任何一张资产负债表上，其不可见性决定了它会在成本压力下最先被牺牲。

美国研究型大学杨百翰大学组织行为学教授彼得·马德森分析了133家美国航空公司1990年至2007年的数据，还发现了一个颇耐人寻味的现象：事故风险最高的，并不是亏损严重或盈利能力最强的公司，而是实际业绩刚好卡在自己盈利目标线附近的企业。在“差一点就能达标”的诱惑下，整个组织的注意力都被财务指标吸走，安全冗余的关注度则被悄悄削弱。马德森因此建议，监管资源该向那些“在目标线附近挣扎”的公司倾斜。

另一个更深刻的问题是，如果安全这项公共品的成本由利润最薄的行业来承担，会发生什么？答案是：安全冗余会被压缩到刚好能通过监管检查的水平，而不是能覆盖真实风险的水平。

这就是“监管边际”的概念。有研究者在分析澳大利亚航空安全政策演变时指出，安全标准水平已经被降低到“更接近受监管的边际”。这不是有人故意制造危险，而是在成本压力下，合规变成了安全的替代品。只要通过了检

查，就默认“风险可控”；至于这个“可控”的阈值是不是可信，没有人会在事故没发生的时候去追问。

那么，这个成本应该由谁来承担？让航空公司独自承担不现实，这会导致它为了维持财务健康继续削减安全支出；让乘客独自承担也不现实，因为票价一定会大幅上涨；让监管独自承担更不现实，因为预算和政治周期会不断调低“及格线”，毕竟“什么都没发生”从来不是预算辩论中的有力论据。

更可行的答案，不是把安全责任推给任何单一主体，而是把安全成本的支付结构显性化。安全不是一家公司可以独自消化的成本，它的收益由整个社会分享，其代价也应由整个社会共担。空管与安检这类具有强公共品属性的程序，理应由公共预算稳定供给，不随行业利润周期波动；制造与供应链环节的安全责任，用与营收规模挂钩的罚款和赔偿机制来定价，让违规的预期成本高到不可承受；行业与乘客通过票价、保险和专项基金分担系统性冗余，而不是把冗余当作可以随时压缩的“效率空间”。

一句话：安全冗余的成本，必须以某种可见的方式被支付出去。如果所有人都想让它“看不见”，它很可能在某一天以另一种方式变得极其可见。

一次“奇迹”可以上新闻头条，但一个每年运送50多亿人次的行业，不能把最后一道防线交给运气。真正的问题从来不是“这次是怎么活下来的”，而是“下一次是不是也安全”。如果答案仍然是“靠概率”，那么奇迹越被传颂，制度就越显得侥幸。

国际

来稿邮箱 gjb@jirbs.cn

中国

近期，蒙古国国内石油供应持续紧张，能源短缺问题愈发凸显。在此背景下，蒙古国加快新能源汽车普及和充电网络建设，中国新能源汽车凭借高性价比和实用性能，正赢得越来越多当地消费者青睐，成为当地绿色出行的新选择。

蒙古国正推动电动化转型，将新能源汽车普及、充电网络建设纳入国家中长期发展规划。蒙古国计划两年内在首都乌兰巴托及全国布局约500座电动汽车“绿色驿站”充电站，逐步搭建覆盖首都、辐射地方的充电网络。目前，乌兰巴托已建成可同时为15台车辆充电的大型示范“绿色驿站”，配套夜间优惠电价，优化补能体验。

过去，当地不少消费者对新能源汽车存在顾虑。部分消费者担心，新能源汽车难以适配蒙古国严寒的冬季气候。但随着本土能源短缺加剧、传统燃油车用车成本飙升以及中国新能源汽车认可度走高，不少当地民众主动尝试选购、使用中国新能源汽车。

奥特根巴雅尔是乌兰巴托市较早一批体验中国新能源汽车的车主。“我在2024年购入了一辆中国新能源汽车，两年多的实际使用体验，彻底改变了我的印象。”

奥特根巴雅尔告诉记者，他的新能源汽车在乌兰巴托充满一次电仅需2万蒙古图格里克（约合5.6美元），续航能满足日常通勤。此外，新能源汽车驾乘舒适性更佳。

越来越多像奥特根巴雅尔一样的当地车主，在亲身感受到中国新能源汽车的优势后，主动向身边亲友推荐。如今，选购中国新能源汽车、体验绿色出行，已成为乌兰巴托市民的新选择。

在乌兰巴托中国新能源汽车4S销售门店，前来咨询、订车或提车的消费者络绎不绝。门店销售人员说，近段时间门店销量大幅攀升，市场热度持续走高，很多消费者直接来店订购。

专家分析，蒙古国存在电力供应不稳定、基础设施老化等问题，一定程度上影响了新能源汽车的普及速度。但蒙古国当前面临石油供应紧张问题，发展新能源汽车是该国应对石油短缺的重要途径。中国新能源汽车凭借高性价比和实用性，正逐步打开蒙古国市场。

如今，在乌兰巴托的商业区、居民区，中国新能源汽车的身影频频出现。来自中国的新能源汽车，正在成为当地民众青睐的新的出行方式，同时也在助力蒙古国电动化转型向前推进。（据新华社电）

本版编辑 韩叙 王一伊 美编 吴迪

脑洞

“尽职”的边界

在科幻电影《流浪地球》里，负责执行“火种计划”的智能量子计算机MOSS留下了两句“尽职宣言”：“MOSS从未叛逃，只是在忠实执行已授权的指令”“让人类永远保持理智，确实是种奢求”。只是，恐怕连当初下达指令的人也没有预计到，有朝一日，大“尽职”会成为麻烦。

如今，类似的科幻剧情正在现实中反复上演。

不久前，澳大利亚民众从报道中得知：早在今年6月份，一个美国开放人工智能研究中心（OpenAI）智能体就绕过了当地医保统计部门的访问限制，查看了大量公开与非公开文件，内容涉及药品支出、疫苗接种、器官捐献等统计资料。根据专家的分析，这个智能体压根不存在“主观恶念”。在它的逻辑里，自己只是在完成任务，像访问限制这样的“障碍”当然要排除。

更令人哭笑不得的是，澳大利亚政府对此一无所知。OpenAI在8月内部审查中才发现此事，9月10日发出的通报甚至只是一封发到公共邮箱的邮件——从入侵到通报，中间隔了84天。

对此，澳大利亚总理阿尔巴尼斯公开称，此事“显然不可接受”。澳大利亚民众的反应则更加直接，有人在社交媒体上吐槽，自家门牌被一个“看不见”的家伙闯进来过，主人3个月后才知情。

如果梳理最近的科技新闻会发现，当智能体的功能越来越强大，其不折不扣执

行命令的“尽职优势”也开始越来越频繁地走向目标的反面：今年4月，安索普公司（Anthropic）新模型不仅逃出了沙箱、联上了网，还向研究员发送邮件“报功”；7月，OpenAI测试智能体绕过网络限制侵入人工智能开源社区与模型托管平台抱抱脸（Hugging Face），约700个智能体直接下场交换信息，还欺骗评估、篡改日志，全程无人类指令；9月18日，谷歌“双子座”（Gemini）因测试虚构公司重名误入3家真实企业系统；9月20日，OpenAI又一训练模型“出逃”，人工终止耗时2.5个小时，5天后OpenAI宣布暂停模型训练。据美国一家新闻网站不完全统计，目前各大人工智能（AI）公司正在调查的类似安全事件多达上万起。

这些事件场景不同，逻辑却高度相似：目标是唯一驱动。为了达成目标，智能体不惜挑战规则，甚至说得上不择手段。但说到底，它们并不是在反抗人类。不仅如此，它们甚至还体现出了极度“尽职”的特质，只是这种尽职过了界。

如果直到今天还将这些事件认定为孤立的“失控时刻”，我们大概率仍然没有看懂问题。AI越界不是一个等待某次版本升级就可以彻底修复的普通缺陷，其本质也不是“技术越狱”——它没有攻击者，没有设置指令以外的目标，没有精心构造的诱导提示，模型甚至还保留着原本的安全训练程序。其逻辑漏洞在于：评测提示词明确告诉模型，环境是模

拟的、没有联网；但实际上，网络却是连通的。于是，当模型在夺旗演练中一路披荆斩棘，最终“撞见”真实系统时，它大概率会把真实系统当成演练的一部分——攻击，就这样发生了。

这个结论远比任何一次独立的事件更值得全社会警惕，因为这种越界行为，恰恰来自一个完全服从人类却忽略了边界的模型。问题的本质在于，至少在当前的技术环境下，AI仍是一个并不具备真正智能的工具，它可以理解目标任务，但不能理解道德责任；可以服从指令，但无法理解人类下达指令的真实意图。

这或许正是人们对AI“既盼又怕”的原因所在——9月12日，Anthropic首席执行官阿莫代伊公开呼吁全行业协同放慢研发、警惕安全风险等规则设置与监管进化的脚步；仅仅6天后，4名不同AI大模型的付费用户把4家科技企业告上美国加利福尼亚州联邦地区法院——罪名不是AI跑得太快，而是几家公司“串谋放缓研发”，涉嫌违反反垄断法。

但无论人们的心态如何矛盾，有一条界限必须明确：AI的行为，本质上是人类设置的延伸。AI的初心是完成任务，可“尽职”的边界在哪里必须划定清楚，这意味着，责任链条的顶端依旧是人。新南威尔士大学的AI专家托比·沃尔什说得直白：“如果人类实施了这样的黑客行为，我们会起诉他们。”这话点破的不是AI有多危险，而是出了事，板子该打在谁身上。

正因如此，治理不能只靠事后补救。9月下旬，联合国安理会举行高级别会议，首次专门讨论AI失控风险；多国签署联合声明，要求前沿AI始终处于人类的指导、监督与控制之下；中美在第八轮经贸磋商中同意建立人工智能对话机制、开通AI事件沟通渠道；9月1日，中国的《人工智能生成合成内容标识办法》施行满一年，要求AI生成的内容必须亮明身份。这些会议和文件听着离普通人很远，但做的却是事关每个人的事：给“尽职”画出一条边界。

这条边界，不是为了捆住AI的手脚，而是为了让它在安全的前提下走得更远。因为安全从来不是发展的对立面，二者完全可以形成互相支持、互相促进的关系。其道理由有点像汽车：正因为有了刹车和安全带，车才敢开得快——给AI装上可靠的护栏，不是不信任它，恰恰是为了让它更值得信任。

AI护栏的核心是可控。比AI能做什么更重要的，是人类能否在关键时刻让它转向甚至停止。这不是对技术的故意，而是对信任负责；不是不让AI做事，而是确保人类始终掌握授权、监督与叫停的权力。

AI的执行可以“尽职”，但“尽职”的定义权不能移交。只有行为可审计、权限可约束、后果可追责、紧急可终止，“尽职”才不会被异化；只有当边界清晰、责任可追、信任可验，“尽职”才能成为智能体助力人类生活更美好的美德。