

连线评论员

为何全球关注超级人工智能

对话人：
彭 飞 本报评论员
曾 毅 中国科学院自动化研究所研究员、北京前瞻人工智能安全与治理研究院院长

彭飞：回顾2025年，人工智能发展一日千里。大家谈到通用人工智能，充满热情，而谈到超级人工智能，则顾虑重重。2025年10月以来，一份呼吁暂缓研发超级人工智能的声明，获得全球一大批科学家、政商界知名人士的签名。这是为什么？通用人工智能与超级人工智能到底有什么区别？

曾毅：目前所说的通用人工智能一般指具有高度泛化能力，接近或达到人类智能水平的信息处理工具，有很广泛的应用前景。超级人工智能则是指各方面都超过人类智能水平，且被认为是接近生命的存在。这意味着“它”会产生自主意识，且很多想法和行动将难以被人类理解，更难以被人类控制。

我们期望超级人工智能是“超级利他”的，但如果是“超级邪恶”怎么办？有研究发现，当前主流大语言模型在面临被替换的可能时，竟采取欺骗等方式以自保。更令人震惊的是，当模型意识到自己处于被测试的环境时，会故意掩盖不当行为。通用人工智能尚且如此，更何况超级人工智能？大家忧虑的正是这种未知感。

彭飞：从历史看，每次重大技术革命都会对经济社会发展产生重要影响。而且，随着技术完善、治理跟进，人类发展最终都能趋利避害。超级人工智能为何不会遵循这样的规律？

曾毅：不能简单把超级人工智能类比为历史上任何一种技术工具。“它”可能拥有独立认知并超越人类智能，这一挑战是前所未有的。“它”带来的风险和颠覆性改变也绝不限于就业、隐私保护、教育等局部领域，而是系统性的。最核心风险在于对齐失败和失控。如果超级人工智能的目标与人类价值观不一致，即使微小偏差也可能在经能力放大后导致灾难性后果。人类的负面行为大量存储在网络数据中，不可避免会被超级人工智能习得，这极大增加了对齐失败和失控的风险。因此，在人工智能开发与治理中，时刻都要坚持底线思维，摆脱传统的被动反应、跟进式的模式，做到未雨绸缪、前瞻布局。

彭飞：面对这么紧迫的课题，我们应采取一种怎样的治理思路？

曾毅：从基本原则看，安全必须是发展超级人工智能的“第一性原理”，即安全应成为模型的“基因”，不可删除，无法违背，不能因为可能影响模型能力而降低安全护栏。应尽可能全面考虑安全隐患并开展模型安全加固，坚持主动防御而非被动应对。

从实现路径看，通过“攻击—防御—评估”的技术过程不断更新模型，能有效解决如隐私泄露、虚假信息典型安全问题，妥善应对短期风险。但从长期看，真正的挑战在于使超级人工智能与人类期望对齐。目前采取的基于人类反馈的强化学习——即在人机交互中将人类价值观嵌入人工智能的模式，很可能对超级人工智能无法奏效，迫切需要全新的思维与行动方式。

从最终效果看，由于超级人工智能可能拥有自我意识，更安全的理想图景是使“它”自主产生道德直觉、同理心与利他之心，而非单纯依靠外部“灌输”的价值规则。确保人工智能从合乎伦理变成拥有道德，才能最大限度降低风险。

彭飞：超级人工智能的安全问题具有全球性，一旦出现漏洞、失控，影响是跨越国界的。而人工智能的全球竞争又非常激烈，无论国家还是企业，都在争先。一些发达国家在超级人工智能的研发上更是“油门踩到底”。如何避免盲目竞争导致失控？人工智能治理的全球协作有可能吗？

曾毅：人类需要避免人工智能的发展演变为“军备竞赛”，其危害性不可估量。创造出世界上第一个超级人工智能也许不需要国际合作，但确保超级人工智能对全人类安全可靠，全球合作是必选项。

世界需要一个高效率、有执行力的国际机构，来协调人工智能的治理，以确保安全。2025年8月，联合国大会决定设立“人工智能独立国际科学小组”和“人工智能治理全球对话”机制，以促进可持续发展和弥合数字鸿沟。这方面的探索还应深入、继续。

各主权国家作为政策制定和执行的主体，尤其是掌握先进技术的发达国家，更有责任和义务防止在规则缺位的条件下盲目发展超级人工智能，造成风险外溢。中国倡导构建人类命运共同体、构建网络空间命运共同体，强调统筹发展和安全，提出《全球人工智能治理倡议》，值得在全球范围推广和践行。宁可稍稍放慢节奏，把安全的底子筑牢打扎实，也不能急功近利，以免把人类社会引向万劫不复的危险境地。

纵横

扩展阅读的半径

杨朝清

不久前，“万卷府抵，共读山海”全民阅读推广活动举办，活动通过资源下沉与服务创新，将优质讲座、数字平台、流动服务直接送达基层。阅读没有止境，丰盈精神世界的需求永远不会过时。现实中，部分公共图书馆覆盖半径较小、基层图书馆基础薄弱、农村及特殊困难群体借阅不便，解决这些问题，迫切需要打通公共文化服务“最后一公里”。比如，借助网络聚合与物流配送的双轮驱动，“云端图书馆”能让公众实现“居家点单、免费借还”，既降低公众阅读的成本，也提升公众参与全民阅读的便捷性。

今天，满足精神诉求和情感需要的渠道越来越多元，但阅读依然是人们获得精神滋养、汲取智慧力量的重要路径。让阅读成为一种生活方式，需要让公共服务更加精细化、人性化，让“精神食粮”以更低成本、更高效率来到群众身边。

从地铁书房到社区书屋，时代浪潮中，我们要多一些改革意识和创新思维，让公众的阅读诉求有更多途径满足。通过阅读来启智增慧、丰盈精神世界，让有限的时间过得更加充实、更有意义，全民阅读需要更多有力举措。

（摘编自《中国教育报》，原文为《让公众阅读诉求有更多“打开”方式》）

本版责编：陈 凌 盛玉雷 尹双红

评论员观察

从“最优路线”看预期管理

邹 翔

洞察大势、遵循规律，在历史前进的逻辑中前进，在时代发展的潮流中发展，便不惧一时一地的波动、眼前暂时

开车上班，导航软件总会提供多条备选路线。时间久了，便对其中两条尤为熟悉。一条红绿灯少，路况却难以预测，通畅时40分钟可达，一旦堵车，耗时可能翻倍。另一条红绿灯多，优点是通行时间稳定在1小时左右，波动范围小。多次尝试过后，笔者便有了心中的“最优路线”：除非特别赶时间，优选后者。

影响选择的关键因素，在于“预期”。日常生活中，人们往往会倾向于预期更稳定、意外更少的选项，因为越稳定的预期，意味着节奏愈可控、收益愈可预料。由日常生活到经济生活，同理。稳定的预期，

人民时评

法治展长卷 碧水入画来

寇江泽

近段时间，两部关于生态文明建设的法律法规正式施行。

一部是《密云水库水源保护条例》。作为国家针对水库水源地颁布实施的首个条例，其施行将极大助力护航首都“大水缸”。

一部是国家公园法。这是我国首部关于国家公园的专门立法，其施行标志着我国国家公园建设迈入法治化、规范化新阶段。

法治展长卷，碧水入画来。从“首个条例”，到首部“专门立法”，法律制度不断完善，让生态文明建设有了更多“硬支撑”，也给各地守护绿水青山带来启示。

法律制度的作用在于规范。法律制度能为生态文明建设护航，一个方面就在于，明确了“可为”与“不可为”的边界，让生态文明建设有章可循、有法可依。

是人们下决心、作决策时的重要锚点，是信心与行动的基石。由此，便不难理解为何许多地方优化营商环境、吸引人才扎根，都把提供稳定可靠的预期摆在突出位置。

预期的关键是“确定性”，稳预期的本质是稳信心。为什么说法治是最好的营商环境？因为制度管根本、管长远，稳定、公平、透明的规则体系有助于经营主体基于合理的预期对各种市场关系和行为作出正确判断，进而降低交易成本、激发市场活力、增强发展信心。民营经济促进法实施后不久，最高人民法院行政庭首次适用该法，在一起案件中判决政府部门向企业支付800余万元补偿。法治为合法经营“撑腰”，推动政府部门依法履职、守信践诺，给企业信心，让企业安心。

政策预期稳，经济大盘才能稳。落实好中央经济工作会议着力稳预期的要求，需纵深推进全国统一大市场建设，增强政策的前瞻性、针对性、有效性，下更大力气破解企业关心的融资难融资贵、账款拖欠、涉企乱收费等问题，让企业家安心经营、放心投资、专心创业，以预期之“稳”、环境之

法治展长卷 碧水入画来

以施工噪声污染为例。近年来，随着城市建设的加速，施工噪声污染成为影响市民生活质量的突出问题。在浙江宁波，关于噪声污染的投诉量呈逐年上升趋势，其中建筑施工及室内装修噪声投诉增长尤为明显。2025年8月，宁波市人大常委会通过施工噪声污染防治规定，对上位法规定较为笼统的内容作了细化，对施工单位职责、高考中考期间施工限制、室内装修的限定时间等作了具体规定。“小快灵”立法，厘清了各方责任，治理噪声扰民有了更有力的法律武器。

习近平总书记指出：“我国生态环境保护中存在的突出问题大多同体制不健全、制度不严格、法治不严密、执行不到位、惩处不得力有关。”加强制度和法律供给，才能避免陷入“治理—反弹—再治理—再反弹”的怪圈，实现“长治”。

生态环境问题往往具有跨区域、迁移性等特点，影响范围可能跨越行政边界。以法律制度的形式，明确各方职责，推动生态治理区域一体化，才能形成治理合力。

就拿密云水库来说，其控制流域跨京冀“三区五县”。京冀两地经济发展水平、产业结构、生态环境保护力度等诸多方面存在差异。一度，保护措施多表现为上下游之间政策的简单叠加，不同地区管控政策差异也较大。条例明确京冀两地政府的协同责任，赋予跨省协作法律强制力，有助于减少因标准差异导致的监管漏洞，共护密云水库这一“无价之宝”。

江西、湖南两省人大常委会分别表决通过萍乡市、株洲市大气污染防治条



河北省唐山市丰润区一家影城里，观众在电影海报前留影。
朱大勇摄（影像中国）

据统计，2025年贺岁档总票房达53.45亿元，创近8年来同期票房新高；2026年元旦档总票房达7.39亿元，实现电影“开门红”。不同于往年喜剧、合家欢题材为主导的局面，档期内上映的影片涵盖历史、悬疑、纪录、温情、奇幻、公益等多元类型。在满足观众日益多样化的观

影需求中，电影产业制作水平也在水涨船高。

这正是：
光影织幻梦，
连场客盈门。
银幕叙悲欢，
沉醉忘归辰。

徐 之文

各位读者，本版“大家谈”“要闻热评”“中国道路中国梦”“来论”等栏目长期征稿，欢迎赐稿，请勿一稿多投。

本版邮箱：rmrbpl@people-daily.cn（来稿请注明栏目名）

欢迎赐稿

中国道路中国梦·青春为中国式现代化挺膺担当③1

“以前怕你们，现在盼你们”

李 晋

“以前怕你们来，总担心影响生意。后来发现你们也解决困难，现在盼你们来呢。”入企检查指导时，一名饭店老板的打趣，让我心里暖烘烘的。扎根一线7年，作为执法人员，我亲历了经营者从“怕检查”到“盼服务”的态度转变，也见证着营商环境持续优化带来的蓬勃生机。这份转变，可以从陪伴我的3个“老伙计”说起。

第一个“老伙计”，是忙忙碌碌的工作电话。过去，商户办证得多头跑部门、反复交材料，自己的手机成了全天候的“咨询热线”。一些经营者冲着电话抱怨：“等审批好，黄花菜都凉了！”吐槽声后，是他们对“能办、更要易办”的服务呼唤。现在，北京“e网通”小程序上线，各类事务均可线上办结，不见面、零成本，十分方便。不过，这手机还是热线，但变成了我追着商户去服务。

以1980年开业的悦宾宾馆为例，按“先注销、再设立”的旧政策，变更经营者意味着老字号的历史“归零”。《促进个体工商户发展条例》出台后，这个问题有了解法。我们第一时间联系老板，上门讲新规，指导申报，帮助其在新政实施当天，顺利领到了全国首张直接变更经营者的个体工商户营业执照。好政策是种子，精准服务是沃土，唯有把政策温度转化为服务精度，才能让红利更快直达经营主体。执法者，要以监管红线守住质量底线，更要用服务初心换得企业安心。

第二个“老伙计”，是双磨破了底的旧皮鞋。2018年我刚入职，那时检查市场就得练就“铁脚板”，揣着登记簿挨家挨户跑，皮鞋不到半年就开口“抗议”。可比起磨损，更令人揪心的是商户们的抵触情绪，有

人直说“来这么勤就是想‘挑刺’”。自己心里委屈，但也无可奈何，检查少了，市场秩序难保障，检查多了，又可能干扰企业经营。

后来我才明白，破解“两难”的关键不在控制检查频次，而在于以理念之变、方式之变重构监管逻辑。近年来，以“综合查一次”合并检查事项，避免重复上门；用好科技赋能，靠AI识别等实现远程“体检”，对诚信企业“无事不扰”，对风险隐患精准介入，我们省力，商户省心。截至2025年11月末，北京非现场检查量占比达70.1%。跳出“非此即彼”的惯性思维，用系统观念打破壁垒、用创新手段提升效能，方能让监管既有效果更有暖意。如今，我这旧皮鞋的“用武之地”越来越少，但经营者脸上的笑容却越来越多，“两难”

逐步成“两全”。

第三个“老伙计”，是肩上台执法记录仪，它见证了我们从“刚性执法”到“柔性治理”的变迁。过去，镜头里多是严肃问询，是规规矩矩的检查，也是彼此之间的距离。如今，这小小的镜头，既规范执法行为、守护市场公平，也保护经营者权益、传递法治温度。

它记录下我们组建“餐饮服务小队”，从源头筑牢食品安全防线，帮经营者省下返工成本；也记录下我们落实“首违不罚、轻微免罚”，对因疏忽售出过期饮料的小店免去罚款；还记录下我们为经营困难的小微企业争取到“能减则减、能免则免”帮扶政策，老板激动地说：“关键时刻，感谢政府在背后撑了一把”……处罚从来不是目的，规范经营、护航发展才是初衷。柔性执法，让“烟火气”与“秩序感”更加协调相融。

经营主体是经济运行的细胞，是推动经济高质量发展的活力源泉。新的一年，我将继续扎根市场监管一线，努力把优化营商环境的春风，吹进每一个经营者的心坎里。我也期盼着，能有更多商户见到我们时，发自内心地喊一声：“娘家人‘来了’！”

（作者为北京市东城区市场监督管理局执法人员）