

## 中国发展高层论坛 2026 年年会 “十五五”的中国：高质量发展与共创新机遇”

### 人工智能治理：挑战与合作专题研讨会 (小组讨论一)

由国务院发展研究中心主办、中国发展研究基金会承办的中国发展高层论坛 2026 年年会，于 2026 年 3 月 22-23 日在北京钓鱼台国宾馆举行。3 月 23 日下午举行了“人工智能治理：挑战与合作专题研讨会”，其中，小组讨论（一）的发言嘉宾为中国可持续发展研究会理事长、科学技术部原副部长李萌，伦敦政治经济学院教授、2010 年诺贝尔经济学奖获得者克里斯多夫·皮萨里德斯，索奈有限责任公司董事长、首席执行官雷默，清华大学苏世民书院院长、教授薛澜。本场会议由国务院发展研究中心公共管理与人力资源研究所所长李佐军主持。

李萌在发言中提出，人工智能要走包容性发展之路，实现四个目标：即速度（技术创新、促进经济发展）、温度（惠及民生、促进社会公平）、绿度（绿色低碳、促进可持续发展）、善度（智能向善、安全可控）。围绕如何实现包容性发展，李萌分享了五条路径。一是由微观主体增效向宏观经济增长延展，发展全要素生产率友好型人工智能；二是由人力替代向就业创造延展，发展就业友好型人工智能；三是由高能耗向低能耗延展，发展绿色低碳友好型人工智能；四是由 AGI 焦虑向人机协同进化延展，发展人机友好型人工智能；五是由一国一企独享向全球公共产品延展，发展全球友好型人工智能。展望未来，李萌认为，在人工智能的发展中，人类必须确立主体地位，发挥主导作用。人类要定义人工智能的未来，而不是让人工智能发展出定义人类未来的能力，坚决杜绝人工智能形成脱离人类掌控的能力。这

需要国际合作、敏捷治理，推动构建人类社会命运共同体框架下的智能社会新生态。

克里斯多夫·皮萨里德斯认为，人工智能治理模式的选择既取决于政治考量，也受到经济与社会因素影响。他从国家、企业、劳动者等方面阐释了对人工智能治理的看法。对国家来说，现阶段劳动力市场虽由人类主导，但滞后治理人工智能风险极易造成不可逆的损害。同时，也要抵制过度监管，监管若扼杀创新，则本质上则是自我挫败，而非有效保护。国家层面的人工智能治理可借鉴金融科技监管的沙盒机制，如规定高风险领域人工智能应用在可控环境内完成压力测试、试点验证，经监管机构审批后发布。此外，针对新型应用，由技术、社会、经济等领域专家组成咨询小组开展综合评估，低风险应用无需进入沙盒。欧盟的《人工智能法》就构建了这样的风险分级监管体系。

克里斯多夫·皮萨里德斯提出，国家监管是人工智能治理底线，企业内部治理是实践关键。人工智能虽会淘汰部分岗位，也会创造新就业机会，此转型过程的核心是保障劳动者就业能力。企业需转变理念，摒弃将劳动力视为成本的观念，将劳动者视为合作伙伴，加大终身教育与再培训投入。同时，通过政府与企业共同努力保障劳动者对人工智能技术部署的话语权，避免侵入式监控，防范人工智能偏见，消除劳动者对技能过时造成利益受损的担忧。他表示，人工智能具备变革潜力，可赋能民生改善、疾病攻克、气候应对等领域，但其正向作用实现离不开良好治理。人工智能治理模式可能因地域、政治偏好、社会价值观与经济优先事项等差异呈现不同形态，但需坚守前瞻性、包容性等原则，基于技术服务公共利益的目标，平衡创新与安全，实现多方共赢。

雷默从竞争与合作的角度，分享了对人工智能治理模式的看法。不同技术有不同治理模式，人工智能治理模式的选择尤为重要。目前技术界将人工智能当做一场赢者通吃的赛跑。但这衍生出两个没有明显答案的问题，即人工智能赛跑的终点在哪里、如何定义输赢。雷默认为，美国技术巨头如微软、Facebook 等虽已形成赢家通吃的格局，但手机操作系统如 iOS 与 Android 却是良性竞争、并存发展。因此，赢家通吃并非唯一路径。最后，雷默提出应摒弃竞赛的思维，形成花

园共治思维。他将人工智能的发展比喻成培育多元独立的花园，各自自主生长又彼此互联互通，彼此不再是对抗，而是双向兼顾，以此维护人工智能发展的生态稳定。

薛澜指出，当前以 OpenClaw 为代表的智能体工具热度攀升，相较于传统人工智能大模型，智能体可自主完成复杂任务。但智能体发展带来新的治理挑战：一是恶意使用风险加剧，以适配于 OpenClaw 的“skills”（技能）为例，研究人员对 3984 个“skills”评估后，得出其中 36.8% 都存在安全问题的结论；二是智能体“幻觉”风险凸显，存在推理、执行、感知、记忆等“幻觉”问题；三是智能体对就业岗位的冲击进一步扩大。他认为，智能体治理需坚持“以人为本、智能向善”原则，强调“敏捷治理、多元共治”，实现从内容安全向行为安全、从静态管控向动态边界管理、从单一风险防控向可信交互三方面转变。从应对挑战的治理举措来看，一是技术层面通过权限隔离、全程监控、搭建可信执行环境，实现智能体行为可验证、可干预、可追溯；二是监管层面明晰权责边界，规范用户授权与平台权责，明确创新发展秩序，建立分级包容审慎监管制度；三是社会层面重点关注劳动者技能重塑、就业缓冲、社会保障以及公共教育。

（中国发展出版社 撰稿：赵海娟、吴佳、徐乾；审稿：杨良敏；基金会审核：夏天、王路）

—背景信息—

中国发展高层论坛由国务院发展研究中心主办，中国发展研究基金会承办。自2000年首次举办以来，论坛坚持“与世界对话，谋共同发展”的宗旨，形成了专业化、高层次的鲜明特色，已成为中国政府高层、全球商界领袖、国际组织和中外学者之间重要的交流对话平台。



中国发展研究基金会  
China Development Research  
Foundation

中国北京东城区安定门外大街138号  
皇城国际中心A座15层

邮编: 100011  
电话: 86-10-64255855  
传真: 86-10-64255855-8100  
网址: [www.cdrf.org.cn](http://www.cdrf.org.cn)  
电邮: [cdrf@cdrf.org.cn](mailto:cdrf@cdrf.org.cn)

Floor15, Tower a, Imperial International Center, No.138  
Andingmen Wai Avenue, Dongcheng District, Beijing, 100011, China

Tel: 86-10-64255855  
Fax: 86-10-64255855-8100  
Website: [www.cdrf.org.cn](http://www.cdrf.org.cn)  
E-mail: [cdrf@cdrf.org.cn](mailto:cdrf@cdrf.org.cn)