

## 中国发展高层论坛 2024 年年会 “持续发展的中国”

### 人工智能发展与治理专题研讨会 (小组讨论二)

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2024年年会，于2024年3月24-25日在北京钓鱼台国宾馆举行。3月24日下午，举行了“人工智能发展与治理专题研讨会”，其中“小组讨论二”环节的发言嘉宾有5位，分别为清华大学苏世民书院院长薛澜，福特斯克金属集团董事长安德鲁·福瑞斯特，清华大学智能产业研究院院长、中国工程院院士张亚勤，普华永道全球主席罗浩智，香港大学教授、当代中国与世界研究中心主任李成，美国战略与国际研究中心高级顾问、中国商务经济项目主任甘思德。本场会议由国务院发展研究中心副主任张顺喜主持。

关于人工智能的风险防范和治理，薛澜提出五方面挑战。第一是步调不一致问题。人工智能技术发展非常快，但治理体系的变化相对缓慢，因为每个法律法规的出台都要经过一系列论证、多方考虑。第二是信息不对称问题。调研发现，政府作为治理主体，企业作为被规制主体，双方都存在盲区。首先，企业不知道政府治理最关心的问题；反过来，政府对于技术可能带来的风险也不清楚，双方信息不对称，有时候存在“共同无知”情况。第三是风险规制过程中的成本不对称问题。滥用或误用人工智能的技术成本相当低，但防范人工智能技术风险的成本要远远高于它有可能造成的危害，所以治理的成本很高。

第四是全球治理体系的问题。全球治理往往与机制复合体相联系，任何一个国家、国际组织或企业都无法独立解决这个问题，人工智能就是典型例子。可能存在专业组织、国际组织、企业等多个组织都与某一领域利益相关，且具备一定条件和能力，都希望参与治理。这些组织之间可能没有从属关系，相互之间或有重叠，甚至是矛盾，这使得全球治理体系的形成有很大困难。第五是地缘政治问题。目前，中美在科技领域竞争激烈，存在技术打压问题，而这种地缘政治会带来相应问题。因此在人工智能治理领域，也一定要有中美合作。

对于如何解决，薛澜提到，一是加强安全领域的技术研发，尤其要加强国际合作；二是倡导敏捷治理；三是鼓励企业内部自我规制；四是加强全球治理；五是打破“小院高墙”，加强中美双方在人工智能领域的合作。

安德鲁·福瑞斯特表示，人工智能问题与人类的生存息息相关，可以说是人类的伟大发明之一，但也是具有危险性的创新。无论是从中国、美国的角度，还是从世界的角度看，必须要对人工智能进行监管，人类必须具备控制它的能力。人工智能本质上来源于所有人在网络上分享的思想，它以极快的速度自我完善。如果不加监管，人类的智力水平将会被快速改善迭代的超智慧物种超过，这是全世界共同面临的巨大危险。如果在人工智能的应用中，一旦出现错误，我们需要有方法能够达成共识，让人类来接管，避免产生毁灭性影响。

张亚勤简要总结了人工智能大模型的发展趋势，一是多模态、多尺度、跨模态的新智能；二是未来五年要在整个架构上有大的突破，追随尺度定律（Scaling Law），但未必是transformer模型结构；三是智能逐渐向“人工智能手机”“人工智能PC”边缘侧延伸；四是可以实现定义任务、规划路径、自我升级、自我编码等的自主智能；五是信息式智能逐渐走向物理世界；六是大模型与生物体连结的生物智能。

张亚勤认为，未来五年随着各大领域大规模应用人工智能技术，将带来三方面风险，包括：信息世界的风险，包含错误、虚假信息的风险；已经出现的大模型幻觉问题，随着信息智能大规模应用延伸至物理智能、生物智能时，风险也会规模化；大模型与经济体系、金融体系、军事系统、电力网络连接的风险。

如何防范这些风险问题，张亚勤提出五个长期建议。一是像标识广告一样，对人工智能生产的数字人等智能体进行标识。二是设立映射和注册机制。明确机器人作为从属体，必须映射到主体，这个主体可以是人或公司等法定实体，如果机器人出现问题，就可以追溯主体责任。三是建立分级监管体制机制。对应用在物理世界、生物系统等不同领域的大模型，分级监管。四是加大对大模型风险研究的投入，呼吁政府、科学家、技术从业者和企业家共同参与，一边发展，一边治理。五是制定人工智能发展的红线。目前北京人工智能国际安全对话已经达成了《北京AI安全国际共识》，提出人工智能发展的红线。各国要在全球智能领域共同合作，将人工智能红线作为社会风险共同应对。

罗浩智主要谈了人工智能治理方面的问题。首先是风险问题，有宏观的，有微观的，也有生存的风险。我们应该实验、学习，保持与时俱进，更大地利用人工智能带来的机会，扩大优势，规避风险，实现转型发展。

罗浩智认为，除了政府层面和社会层面对人工智能的监管，公司层面的管理方式也要变化，包括价值观的改变、激励体系的改变、行为的改变和文化的改变，更加全面地考虑到人工智能的影响。因此，要改变企业管理层的制度，加快决策速度，补足短板。为了使优势最大化、劣势最小化，不能仅仅在单个组织的孤岛中进行变革，而要在整个生态系统引导这样的变化，联合多家企业一起变革，使得规划能够顺利实施。连接有多紧密，发展就多有利。

李成表示，人工智能深刻地改变了一切，除了人们的思考方式。人工智能以史无前例的速度、规模、范围和强度影响到我们，但我们其实还没有做好准备，包括治理和就业等方面。我们必须明确，人工智能只是一种工具，它如何被使用？会有什么效果？现在还不确定。但我们要在思维上及时跟上，努力思考，找到应对机制，避免它所产生的负面影响。中美作为超级大国，应该立即行动，好好思考人工智能未来应用的情境，包括可使用的数据、使用的范围等，携手前行。

甘思德分享了人工智能当前面临的三个挑战。第一是人工智能使用安全问题，比如大家都在担心的自动驾驶、大语言模型等的风险。第二是人工智能科技及其应用，国家之间依然存在流通限制，未来一段时间，如何争取人工智能的三个核心元素——数据、算力和算法的主动权，抢占发展先机。第三是人工智能治理，包括三个层面：多边、双边和单边。多边层面，要增强多边治理，来实现安全可信任的人工智能，实现可持续发展；双边层面，需要达成共识并实现合作，形成框架性协议，解决人工智能治理问题；单边层面，可以做的事情很多，以中美为例，美国应该制定全面的法律和框架，管理数据隐私和安全，而且应该取消“防火墙”，加速人工智能领域的发展，方便人们的生活；中国应该更多地参与国际人工智能标准的制定，产生更大的影响、拥有更多的话语权，以更多更好的方式方法向世界展现自己，解决问题，共同发展。

（中国发展出版社 撰稿：刘长杰 雒仁生 徐静；审稿：杨良敏  
中国发展研究基金会 审校：丁孟宇）

--背景信息--

中国发展高层论坛是由国务院发展研究中心主办、中国发展研究基金会承办的国家级论坛。论坛秉承“与世界对话，谋共同发展”的宗旨，着眼于对外释放我国对外开放信号，自2000年创办以来，已发展成为服务中央决策和国家对外经济交往、促进中外企业交流合作的重要平台。

--媒体联络--

中国发展研究基金会

郭丝露 18666028168/64255855-8014

史雅帆 13810361966/64255855-8223

夏天 18801375838/64255855-8086

施婉菁 18801090391/64255855-8090

邱凯娴 18301078627/64255855-8103



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)