

## “西岸对话”发起单位

上海交通大学  
上海人工智能实验室  
中国互联网协会  
中央广播电视总台超高清视音频制播呈现国家重点实验室  
上海市徐汇区人民政府  
中国互联网治理论坛(中国IGF)  
清华大学产业发展与环境治理研究中心

## “西岸对话”共同发起部门(机构)

上海交通大学国际与公共事务学院  
上海交通大学人工智能学院  
上海交通大学中国城市治理研究院  
清华大学人工智能国际治理研究院  
北京大学数字法治研究中心  
北京大学法律与人工智能研究中心  
北京大学公共政策研究中心  
北京大学政府大数据与公共政策实验室  
厦门大学公共事务学院  
厦门大学公共政策研究院  
南开大学周恩来政府管理学院  
华东师范大学公共管理学院  
华南理工大学公共政策研究院  
复旦大学数字与移动治理实验室  
中山大学数字治理研究中心  
北京邮电大学互联网治理与法律研究中心  
阿里研究院  
腾讯研究院  
中国移动研究院(中移智库)  
中关村智用人工智能研究院  
华为技术有限公司  
商汤科技智能产业研究院

# 西岸对话： 人工智能安全与治理对话网络

## 发起倡议

07.05

中国·上海

通用人工智能的迅速发展和普及应用,正在推动人类社会进入人工智能时代。人工智能在引领新一轮科技革命和产业变革的同时,也带来了各种风险和复杂挑战,人工智能发展、安全、治理已经成为国际社会的普遍关切。但同时,人工智能技术演化的动态性、产业应用的普遍性与广泛性、风险产生的不确定性,使得人工智能安全与治理面临着新的挑战。

《全球人工智能治理倡议》提出,应“坚持发展和安全并重的原则,通过对话与合作凝聚共识,构建开放、公正、有效的治理机制。”为此,我们认为,当前有必要建设有效的协同、对话机制,搭建政、产、学、研跨领域交流对话平台,协同各方在共商共建共治过程中共同探索、共同进步,促进关于人工智能安全与治理知识、能力、资源的生长和积累,以保障人工智能技术和应用的健康与可持续发展。

我们倡议,发起“西岸对话:人工智能安全与治理对话网络”,秉持以下四方面基本原则:公共原则、科学原则、责任原则、敏捷原则。

**第一 “公共原则”** 是指,在人工智能安全与治理过程中注重开放性、包容性、普惠性,合作建设具有公信力的跨机构合作治理机制,共同分享、相互学习人工智能安全风险知识和治理实践,推动形成发挥公共价值作用的安全治理框架或工具。

**第二 “科学原则”** 是指,应遵循、尊重人工智能技术创新、应用的科学规律,调动科学家、工程师、企业、技术社群、监管机构、多利益相关方主体积极性,在技术创新应用过程中,共同探索、学习、积累人工智能安全知识,渐进演化、逐步提升人工智能安全与治理共识水平,并在此过程中培养人工智能安全与治理能力、积累人工智能安全与治理资源。

**第三 “责任原则”** 是指,要探索、形成多类型、具有不同约束力程度的责任治理框架和制度安排,并根据具体安全风险等级形成相匹配的责任制度要求,以激励相容的责任理念促进多利益相关方更主动、积极参与人工智能安全全球公共产品建设,以确保公共产品不被滥用、误用,形成良性生态体系。

**第四 “敏捷原则”** 是指,多边、多方主体应致力于形成相互关联、相互依存的治理关系,基于广泛共识,及时回应人工智能安全风险,以反馈互动、迭代创新的安全治理过程应对人工智能安全风险及其演化的不确定性与动态性。

为践行上述四点原则,“西岸对话”将形成宽泛、稳定的合作结构,通过以下四方面工作推动人工智能安全与治理的国内行动与国际对话。

**第一** 形成经常性交流对话机制,围绕人工智能安全与治理议题,定期组织政、产、学、研的跨领域对话,对人工智能技术的新发展、新业态作出敏捷回应,并逐渐扩展为多方共识。

**第二** 致力于推动形成人工智能安全与治理的公共知识、能力与资源,在标杆案例、规范标准、政策工具等各个维度搭建同行评议、互通有无的平台渠道,以智力产品支撑国内相关机构的治理行动,以释放人工智能安全与治理的“公地喜剧”。

**第三** 展开人工智能应用与治理素养提升活动,建设公共知识以弥合智能鸿沟和治理能力差距,助力人工智能赋能千行百业,回应技术与产业“孤岛困境”,推动新质生产力发展。

**第四** 积极参与人工智能安全与治理的国际对话,总结中国经验、对话国际实践,成为沟通国内外的有效窗口。