

论坛议程

上午 美院A301报告厅

- 08:00-09:00 • 签到
- 09:00-09:10 • 开场 **汪晖** 清华大学文科资深教授、清华大学人文与社会科学高等研究所所长、2025方塘论坛主席
- 09:10-09:20 • 校领导致辞
- 09:20-09:30 • 致辞 **杨锦方** 方塘研究基金捐赠人、方塘研究基金管委会委员
- 09:30-10:00 • 主题发言 “人工智能的安全治理”
姚期智 图灵奖获得者、中国科学院院士、清华大学人工智能学院院长
- 10:00-10:30 • 主题发言 “普适性人工智能伦理：研究范围与方法”
Vincent C. Müller 埃尔朗根-纽伦堡大学（FAU）亚历山大·冯·洪堡教席教授、FAU人工智能哲学研究中心主任、人工智能哲学学会主席
- 10:30-11:10 • 主题发言 “我们是否准备好迎接二元主体性的未来？”
赵汀阳 中国社会科学院学部委员、国家文史馆馆员
- 11:10-11:40 • 主题发言 “为人类与人工智能的和谐共生重构超级对齐”
曾毅 中国科学院自动化研究所研究员、人工智能伦理与治理研究中心主任

中午

- 11:30-13:30 • 午餐和自由讨论

下午 蒙民伟人文楼 B201、B202、B203

- 13:30-16:00 • 专题讨论
- 2100年人与超级智能的关系：隐喻、类比与重塑**
2025年是图灵测试提出的第75年，如果站在今天去勾画下一个75年，我们能否构想出人类与超级人工智能共存的积极愿景？哪些隐喻、故事、类比能帮助我们理解未来人与无处不在的智能体之间的关系？实现这些积极愿景的风险有哪些，又该如何有效管理？超级人工智能的存在会如何影响人类生命的尊严与意义？
 - 认识机器：意识、道德与风险**
人类是否能制造出“有道德的”机器？人类是否是人机关系中唯一的道德主体？如何界定机器是否有意识？机器的意识又是否存在阶段和分级？关于超级智能带给人类的风险假设中，远虑和近忧哪个更关键？超级智能是否会带来人、机之外的其他变量？
 - 奇点前夜：人类如何安身立命？**
随着人工智能的高速发展，人类现有工作将被大规模取代，智能体和机器人深入人类社会成为人类新的照护者、朋友和伴侣等现象的普及，AI增强人、数字永生等现象的涌现，让哲学、宗教、伦理等人文问题再次成为热点。人工智能不仅改变了技术应用，更深刻影响了人类认知、伦理、社会关系等多个层面，引发了新的哲学问题和思考。AI会如何重塑人类生活的方方面面？会对现有的社会秩序和伦理传统产生怎样的影响？又该如何应对？
- 16:00-16:30 • 茶歇
- 16:30-18:30 • 圆桌讨论 **主持人：汪晖**