



· 主办单位 ·

清华大学方塘研究院

· 协办单位 ·

清华大学人文与社会科学高等研究所

清华大学人工智能国际治理研究院

人工智能治理研究中心（清华大学）

清华大学艺术与科学研究中心



会议手册

2025年6月23日
JUNE 23, 2025

北京·清华大学
TSINGHUA UNIVERSITY, BEIJING



方塘论坛

“方塘论坛”于2025年首次举办，以“科学、技术与文明的未来”为主题，关注世界前沿议题、积极推动人文与科技的跨学科对话、助推本源性问题的研究和基础理论的发展。

2025 首届方塘论坛将围绕“AI时代的伦理奇点”展开讨论。

人工智能的飞速发展，将人类推入充满不确定性的时代。千百年来，人们的祖先从经验中提炼伦理准则，然后在实践中逐渐校正，形成与时代圆融一体的价值体系。然而，AI时代骤临，伦理实践的速度显而易见地落后于技术迭代的速度，迫使人们用前瞻的思维，去思考未来的问题。当人们为技术奇点何时降临争论不休之时，与之相应的伦理奇点也应得到足够的讨论和准备。

AI伦理不仅是技术问题，更是文明问题，它不仅改变了技术应用，更将深刻影响人类自我认知和社会关系等层面，引发新的哲学思考：如果现有工作被大规模取代，人类如何体面地、有意义地生活？如果智能体和机器人深入人类社会，成为人类的照护者、朋友和伴侣，现有的社会关系和伦理结构会出现怎样的变化？如果将机器作为新坐标，人的定义是否会要变化？人的哪些属性变得格外重要，哪些又变得不再重要？

对人工智能相关伦理问题的回应，既需要人文社会科学的研究与思考，也需要对人工智能系统的运行逻辑、技术发展的可能路径及其物理现实的客观认识。因此，方塘论坛将汇聚全球各地的哲学家、人工智能专家、科幻作家和其他各领域关注人工智能问题的有识之士，打破学科与地缘的藩篱，展开本质性前沿性的对话与交锋，共同探讨事关人类文明未来的大问题，孕育走向未来的学术理论。

为此，2025年6月23日，清华大学方塘研究院携手清华大学人文与社会科学高等研究所、清华大学人工智能国际治理研究院、人工智能治理研究中心（清华大学），清华大学艺术与科学研究中心共同举办“科学、技术与文明的未来——AI时代的伦理奇点”国际论坛，邀请图灵奖得主姚期智、中国科学院学部委员赵汀阳、纽伦堡大学人工智能哲学研究中心主任 Vincent Muller、剑桥大学莱芙赫姆智能未来中心主任 Stephen Cave、清华大学哲学系主任唐文明、浙江大学人文高等研究院院长赵鼎新等18位学者参与方塘论坛的讨论，共同回应人工智能时代人类面临的核心问题。

Tsinghua Fangtang International Forum

The Tsinghua Fangtang International Forum will make its debut in 2025 under the theme “The Future of Science, Technology, and Civilization.” Focusing on frontier global issues, the forum actively promotes interdisciplinary dialogue between the humanities and technology and advances the study of fundamental questions and the basic theory.

Its first session in 2025 will explore “The Ethical Singularity in the AI Era.”

The rapid advancement of AI is propelling humanity into an era marked by uncertainty. For centuries, ethical norms were distilled from our ancestors’ experience and refined through ongoing practice, and the value systems in harmony with their times gradually came into shape. However, the sudden arrival of the AI era has significantly hindered ethical practice from keeping up with the rapid technological iteration. This situation pressures humans to grapple with challenges that lie ahead in a forward-looking manner. While debate over the time when the technological singularity might occur continues, we must also explore and prepare ourselves well for the corresponding ethical singularity.

AI ethics is not merely a technical issue but also a civilizational challenge. AI has transformed technology applications and will further reshape our self-understanding and social relations, thus provoking profound philosophical questions. If existing jobs are widely replaced by AI, how will human beings live with dignity and purpose? If intelligent agents and robots are deeply integrated into society as caregivers, companions, and partners, what will be their impact on our current social relations and ethical structure? If machines rise as a new reference point, will human beings be defined in another way? Which human attributes will grow important, and which may decline to being irrelevant?

The ethical dimension of AI requires rigorous research in and reflections from the humanities and social sciences, as well as a grounded understanding of how AI systems operate, the possible trajectories of technological development, and their physical reality. This forum will convene philosophers, AI experts, science fiction writers, and leading scholars in other fields from around the world who pay close attention to AI-related issues. By breaking down disciplinary and geographic boundaries, the forum will foster essential and frontier discussions, intellectual engagement on the future of civilization, and new forward-looking academic theories.

To this end, the international forum “The Future of Science, Technology, and Civilization: The Ethical Singularity in the AI Era,” hosted by the Fangtang Institute at Tsinghua University, co-hosted by the Tsinghua Institute for Advanced Study in Humanities and Social Sciences, the Institute for AI International Governance of Tsinghua University, the Center for AI Governance (Tsinghua University), and the Art and Science Research Centre of Tsinghua University, will be held on June 23, 2025. 18 leading scholars have been invited to attend the forum. Together, they will explore the core challenges humanity faces in the age of AI.

清华方塘研究院

清华大学方塘研究院成立于 2024 年，立足清华大学多学科优势和人文社科学科特色，着力建设具有全球影响力的原始创新策源地，打造全球领先的科学与人文思想交叉研究高地。

“方塘”二字取自朱熹《观书有感》“半亩方塘一鉴开，天光云影共徘徊。问渠那得清如许？为有源头活水来。”方塘研究院以“坚持探索和传播培育世界一流创新人才，诞生重大原始创新成果的思想前提、文化机理和教育规律”为使命和前进方向，希望通过持续的努力，为学术研究引入更多“源头活水”，

清华大学方塘研究院自成立之始，即以打造思想的探索之地、新思想的产生之地为己任，以改革精神与开放意识，积极培育创新文化，厚植创新土壤，促进跨领域的深度思想交流。研究院将以“对人、问题与方法的学术审美品味”为纽带，凝聚秉持纯粹学术理想、不懈追求一流理论成果的全球顶尖学者，构建学术新生态，推动人文与科学前沿领域的突破性进展，为人类步入人工智能时代提供思想指引。

Fangtang Institute at Tsinghua University

The Fangtang Institute at Tsinghua University was established in 2024. Leveraging the university’s strengths in multidisciplinary research and its distinctive humanities and social sciences foundation, the institute is dedicated to building itself into a source for original innovation with global influence and a world-leading hub of scientific and humanistic thoughts.

“Fangtang” derives from the poem Thoughts from Reading by Zhu Xi, a renowned Neo-Confucian scholar of the Southern Song Dynasty: “There lies a glassy oblong pool (Fangtang), Where light and shade pursue their course. How can it remain so clear? For fresh water flows in from the source.” The institute’s mission is to “explore and spread the ideological premise, cultural mechanism, and educational law of cultivating world-class innovative talents and producing major original, innovative achievements.” Through sustained effort, it hopes to introduce more “fresh water from the source” into the pool of academic research.

Since its inception, the institute has positioned itself as the site for bold intellectual exploration and the cradle of new ideas. With the spirit of reform and an open outlook, it actively fosters a culture of innovation, accumulates fertile soil for creative thoughts, and encourages deep, cross-disciplinary intellectual exchange. With “aesthetic appreciation for people, problems, and methods in scholarship” as the anchor, it gathers global top scholars who uphold the ideals of pure academic inquiries and pursue excellence in theoretical innovation, builds a new academic ecosystem, facilitates breakthroughs at the frontier domains of humanities and science, and provides intellectual guidance for humanity as it steps into the AI era.

清华大学人文与社会科学高等研究所

TSINGHUA INSTITUTE FOR ADVANCED STUDY IN HUMANITIES AND SOCIAL SCIENCES

清华大学人文与社会科学高等研究所（简称“文科高研所”）是一个跨学科、国际化的学术平台，汇聚全球学者开展人文社科领域的理论研究和前沿课题探索。依托清华多学科环境，本所促进不同文化、领域和背景的学者对话，并与国内外研究机构开展形式多样的合作。自2009年成立以来，文科高研所举办多次国际性学术会议，以及“人文与社会”系列讲座百余场，邀请了来自亚洲、非洲、拉丁美洲、欧洲和北美的学者至清华进行学术演讲、参加会议或工作坊，这些活动也吸引了清华及其他在京高校和研究机构的学者和学生前来对话、交流。文科高研所现有常驻所学者包括：汪晖教授（文科高研所所长）、崔之元教授、冯象教授、沈卫荣教授、宋念申教授、王中忱教授、严海蓉教授、冯乃希副教授和袁先欣副教授。由文科高研所编辑的人文学术集刊《区域》，迄今已出版十一辑。本刊向跨学科、跨文化、跨区域的研究开放，鼓励人文与社会科学领域的理论探索和突破。

Tsinghua Institute for Advanced Study in Humanities and Social Sciences (TIAS) is an interdisciplinary and international academic platform dedicated to advancing theoretical research and cutting-edge scholarship in the humanities and social sciences. Building on Tsinghua University's multidisciplinary strengths, the institute facilitates dialogue among scholars across cultures, disciplines, and backgrounds while fostering diverse collaborations with research institutions worldwide.

Since its founding in 2009, TIAS has organized numerous international conferences and hosted over a hundred lectures in its "Humanities and Societies" series. These events have drawn scholars from Asia, Africa, Latin America, Europe, and North America to present research, participate in workshops, and engage in academic exchange at Tsinghua. The institute's activities also attract faculty and students from Tsinghua and other Beijing-based universities and research institutions, creating a vibrant intellectual community.

Current Permanent Faculty:

- WANG Hui (Director of TIAS)
- CUI Zhiyuan
- FENG Xiang
- SHEN Weirong
- SONG Nianshen
- WANG Zhongchen
- YAN Hairong
- FENG Naixi (Associate Professor)
- YUAN Xianxin (Associate Professor)

TIAS edits the academic series Remapping, which has published eleven volumes to date. The series welcomes interdisciplinary, cross-cultural, and transregional research, promoting theoretical innovation and scholarly breakthroughs in the humanities and social sciences.



清华大学人文与社会科学高等研究所

Tsinghua Institute for Advanced Study in Humanities and Social Sciences



清华大学人工智能国际治理研究院

INSTITUTE FOR AI INTERNATIONAL GOVERNANCE TSINGHUA UNIVERSITY

清华大学人工智能国际治理研究院是2020年4月由清华大学批准成立的校级科研机构，由知名科技政策专家、清华大学文科资深教授薛澜担任院长，图灵奖得主，中国科学院院士姚期智先生担任学术委员会主席，拥有一支由知名专家学者、中青年学术骨干和业界资深人士组成的专兼职研究队伍。研究院依托清华大学在人工智能与国际治理方面的已有积累和跨学科优势，面向人工智能国际治理重大理论问题及政策需求开展研究，致力于提升清华在该领域的全球学术影响力和政策引领作用，为中国积极参与人工智能国际治理提供智力支撑。

研究院成立以来承担了包括国家科技创新2030重大项目，科技部科技创新战略研究专项等一系列重大项目，发布了一系列研究报告，向科技部、外交部、工信部、发改委、网信办等部门提交了一批政策建议，研究院主办的人工智能合作与治理国际论坛已成为国内该领域重要会议，AI促进可持续发展青年创造营（Bootcamp）吸引了来自世界各地的青年学子，推出的人工智能治理大讲堂、AI治理对话录、《人工智能国际治理观察》和人工智能热点话题在线直播等引起了社会各界的广泛关注。研究院积极拓展对外合作，已与慕尼黑工大、剑桥大学、卡内基梅隆大学等国外高校，联合国开发计划署、国际电信联盟、世界工程组织联合会等国际组织，以及产业界建立了密切交流与合作。

Approved to be established by Tsinghua University in April 2020, the Institute for AI International Governance (I-AIIG) is a university-level research institution headed by Xue Lan, a renowned science and technology (S&T) policy expert and senior professor of liberal arts at Tsinghua University. Andrew Chi-Chih Yao, a Turing Award winner and academician at the Chinese Academy of Sciences, serves as the chair of the academic committee of I-AIIG. The Institute takes pride in its strong resident and nonresident research forces made up of leading scholars, young academics, and industry heavyweights. Leveraging Tsinghua's existing resources and interdisciplinary strengths in AI and international governance, I-AIIG conducts research on major theoretical issues and policy needs regarding global AI governance, as it aims to enhance Tsinghua's global academic influence and policy leadership and to provide intellectual support for China's active engagement in this arena.

Since its inception, I-AIIG has undertaken major projects including China's Major S&T projects (known as 'Innovation 2030') and Strategic Research Projects on S&T Innovation commissioned by the Ministry of Science and Technology, published a series of research reports, and proposed policy recommendations to the Ministry of Science and Technology, Ministry of Foreign Affairs, Ministry of Industry and Information Technology, National Development and Reform Commission, and Cyberspace Administration of China, among other departments. The International AI Cooperation and Governance Forum hosted by I-AIIG has grown to be an important gathering in this field in China, and its U'nAI Youth AI4SDG Bootcamp has convened young people from around the world. In addition, the Institute has launched widely discussed programs and publications including the AI Governance Lectures, AI Governance Dialogue, International AI Governance Newsletters, AI On Air, etc. Actively forging international cooperation, I-AIIG has established close partnerships with prestigious institutions, organizations and the industry including the Technical University of Munich, Cambridge University, Carnegie Mellon University, United Nations Development Programme, International Telecommunication Union, World Federation of Engineering Organizations, etc.



清华大学人工智能国际治理研究院

INSTITUTE FOR AI INTERNATIONAL GOVERNANCE
TSINGHUA UNIVERSITY

人工智能治理研究中心

CENTER FOR ARTIFICIAL INTELLIGENCE GOVERNANCE

人工智能治理研究中心是北京市教委批复成立的北京人文社会科学研究中心，中心依托清华大学公共管理学院，联合清华大学社会科学学院、人文学院、法学院、教育研究院、人工智能研究院、交叉信息学院等，利用清华大学在人工智能技术、公共政策、科技治理、科技社会学、科技哲学、法学方面的多学科优势和积累，面向人工智能文化、伦理、教育、社会认知与调查开展研究，形成人工智能治理“北京学派”，致力成为人工智能治理领域国内领先、世界一流的知名学术高地；搭建人工智能治理学术研究与人文思想共享平台，提出人工智能治理“北京共识”；培育人工智能治理理论创新型人才，打造人工智能治理一流研究团队，建设全国具有引领示范效应的人工智能治理交叉学科；传播智能文化与数字人文，打造“智能文化与数字人文”高地，为数字时代国家和首都治理现代化提供理论支撑和对策建议。

The Center for AI Governance is among the Beijing Humanities and Social Science Research Centers approved by the Beijing Municipal Education Commission. The Center is based at the School of Public Policy & Management, Tsinghua University, and supported by the School of Social Sciences, the School of Humanities, the School of Law, the Institute of Education, the Institute for Artificial Intelligence, and the Institute for Interdisciplinary Information Sciences of Tsinghua University. At the same time, the Center makes best of Tsinghua's multidisciplinary advantages and achievements in artificial intelligence (AI) technology, public policy, science and technology governance, technology sociology, technology philosophy, and law. It studies the AI culture, ethics, education, social cognition and investigation, and forms the "Beijing School" of AI governance. The Center is committed to becoming a world-class, leading academic highland on AI governance in China. By building a platform for sharing academic research and humanistic ideas of AI governance, it proposes the Beijing AI Principles. Furthermore, it aims to cultivate innovative talent in AI governance theory, build a first-class research team on AI governance, and establish a national interdisciplinary AI governance with leading demonstration effects. The Center also works to popularize and create a highland of "intelligent culture and digital humanities," to provide theoretical support and suggestions for the modernization of national and capital governance in the digital era.



人工智能治理研究中心
CENTER FOR ARTIFICIAL INTELLIGENCE GOVERNANCE



清华大学艺术与科学研究中心

ART & SCIENCE RESEARCH CENTER TSINGHUA UNIVERSITY

2001年5月，在著名艺术家吴冠中先生和诺贝尔物理学奖得主李政道先生的共同倡议下，清华大学艺术与科学研究中心正式成立。吴冠中先生认为，科学揭示宇宙的奥妙，艺术揭示情感的奥妙；李政道先生提出，科学和艺术的基础是人类的创造力，它们追求的目标都是真理的普遍性。

中心由清华大学美术学院与清华大学计算机系、医学院和新闻与传播学院共建，以组织和开展跨学科的艺术与科学研究学术活动为任务，使艺术与科学融合并升华到新的高度。中心宗旨是为了促进艺术与科学的结合，探索和建设适应新世纪发展需求的新兴学科，完善现代高等教育体系，继承清华大学优秀文化传统，为创造世界一流大学和一流学科奠定基础。

In May 2001, under the joint initiative of Mr. Wu Guanzhong, a renowned artist, and Mr. Tsung-Dao Lee, a Nobel laureate in Physics, the Art and Science Research Center, Tsinghua University, was officially established. Mr. Wu Guanzhong believed that science reveals the mysteries of the universe, while art reveals the mysteries of emotion; Mr. Tsung-Dao Lee proposed that the common foundation of science and art lies in human creativity, and both pursue the universality of truth as their goal.

Co-established by the Academy of Arts and Design, Tsinghua University, along with the Department of Computer Science and Technology, School of Medicine, and School of Journalism and Communication of Tsinghua University, the center undertakes the mission of organizing and conducting interdisciplinary academic activities in arts and science research, aiming to integrate and elevate arts and science to new heights. Its purpose is to promote the integration of arts and science, explore and develop emerging disciplines that meet the development needs of the new century, improve the modern higher education system, inherit the excellent cultural traditions of Tsinghua University, and lay the foundation for building a world-class university and first-class disciplines.



清华大学艺术与科学研究中心
Art & Science Research Center Tsinghua University



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”
文森特·穆勒 Vincent C. Müller 埃朗根-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
- 10:30–11:00 • 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
- 11:00–11:30 • 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

- 13:30–16:00 • 专题讨论
1 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？
2 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？
3 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

AGENDA OF TSINGHUA FANGTANG INTERNATIONAL FORUM 2025 JUNE 23

Morning Session The A301 Lecture Hall of the Academy of Fine Arts

- 08:00–09:00 • Sign in
- 09:00–09:10 • Remarks Wang Hui Chair of the Tsinghua Fangtang International Forum
- 09:10–09:20 • Qiu Yong Secretary of the CPC Tsinghua University Committee
- 09:20–09:30 • Yang Jinfang Donor and Member of the Management Committee of the Fangtang Research Fund
- 09:30–10:00 • Keynote speech Yao Qizhi AI Safety and Governance
- 10:00–10:30 • Vincent C. Müller Universal AI ethics: Scope and method
- 10:30–11:00 • Zhao Tingyang Ready for a world of bi-subjectivities?
- 11:00–11:30 • Zeng Yi Redefining Superalignment for Human–Living AI Symbiosis

Lunch Break

- 11:30–13:30 • Lunch and informal discussions

Afternoon Session Meng Minwei Humanistic Building

- 13:30–16:00 • Panel discussions
Panel 1: The Human–Superintelligence Relationship in 2100: Metaphors, Analogies, and Re-invention
The year 2025 marks the 75th anniversary of Alan Turing’s proposed test for machine intelligence. As we look ahead to the next 75 years, can we envision a hopeful future of coexistence between humans and superintelligent AI? What metaphors, narratives, or analogies can help us comprehend our future relationship with these pervasive intelligent agents? What are the inherent risks involved in realizing such optimistic visions, and how might we effectively manage them? How will the existence of superintelligent AI influence the dignity and meaning of human life?
Participating Scholars: Stephen Cave、Chen Qiufan、Tang Wenming、Atoosa Kasirzadeh
Panel 2: Understanding Machines: Consciousness, Morality, and Risk
Is it genuinely possible for humans to create machines with morality? In the human–machine relationship, are humans destined to be the sole moral agents? By what criteria could we determine whether a machine is truly conscious? And if consciousness is attainable for machines, might it exist in varying stages or levels? Among the array of hypothetical risks posed by superintelligence, which are more pressing, long-term threats or more immediate, near-term concerns? Could superintelligence introduce unforeseen variables into this equation, extending beyond the simple dichotomy of just humans and machines?
Participating Scholars: Shimon Edelman、Vincent C. Müller、Wang Yue、Zhang Weite
Panel 3: On the Eve of the Singularity: How Should Humanity Find Its Place and Purpose?
As AI continues its rapid advance, existing human jobs are poised to be replaced on a massive scale. Intelligent agents and robots are becoming deeply integrated into human society as new caregivers, companions, and partners. The emergence of AI-enhanced humans and digital immortality is forcing philosophical, religious, and ethical issues back into the spotlight. AI is not only transforming the application of technology but is also deeply reshaping human cognition, ethics, and social relationships, raising new philosophical questions and demanding fresh reflections. How will AI redefine every aspect of human life? What will be its impact on our current social orders and moral traditions? And how should we navigate this new reality?
Participating Scholars: Baoshu、Zhao Dingxin、Francesco Verso、Sun Zhouxing
- 16:00–16:30 • Tea break
- 16:30–18:30 • Roundtable
Moderator: Wang Hui
Participating Scholars: Tang Wenming、Wu Yan、Stephen Cave、Zhao Tingyang、Zhao Dingxin、Liang Zheng、Sun Zhouxing



汪 晖 | Wang Hui

清华大学文科资深教授，人文学院教授，清华大学人文与社会科学高等研究所所长，欧洲科学院院士，领域为中国思想史、中国现代文学和社会理论等。主要著作包括《世纪的诞生》、《现代中国思想的兴起》（四卷本）、《去政治化的政治》《东西之间的“西藏问题”》《亚洲视野：中国历史的叙述》、《短二十世纪》等近 20 种，大量作品被翻译为英文、日文、韩文、意大利文、德文、西班牙文、斯洛文尼亚文等各种文字，约 20 余种。

WANG Hui is the Distinguished Professor in the School of Humanities at Tsinghua University and Director of the Tsinghua Institute for Advanced Study in the Humanities and Social Sciences. He has been elected as a foreign member of Academia Europaea (MAE) since 2024. His research interests include Chinese intellectual history, Chinese literature and social and political theory. He has received numerous awards for his scholarship, including the Luca Pacioli Award (2013) and Anneliese Maier Research Award (2018). His recent English publications include *The Rise of Modern Chinese Thought* (2022), *China's Twentieth Century* (2016), *China from Empire to Nation-State* (2015), and *The Politics of Imagining Asia* (2011).



姚期智 | Andrew Chi-Chih Yao

世界著名计算机学家，2000 年图灵奖得主，中国科学院院士，清华大学交叉信息研究院及人工智能学院院长。研究涵盖计算理论、密码学和量子计算等方面，在通信复杂性、密码学基础和分布式量子计算等领域做出了基础性贡献。

World-renowned computer scientist, 2000 A.M. Turing Award Laureate, Academician of the Chinese Academy of Sciences, Director of the Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University. His research spans computational theory, cryptography and quantum computing, with fundamental contributions to communication complexity, cryptographic foundations and distributed quantum computing.



梁 正 | Liang Zheng

清华大学公共管理学院教授，清华大学人工智能国际治理研究院副院长，清华大学中国科技政策研究中心副主任，清华大学科技发展与治理研究中心学术委员会秘书长，兼任中国科学学与科技政策研究会理事、副秘书长，公共管理专业委员会副主任。主要研究方向为科技创新政策、研发全球化、标准与知识产权、新兴技术治理。

Professor at the School of Public Policy and Management, Tsinghua University, Deputy Director of Tsinghua University's Institute for AI International Governance (I-AIIG), Deputy Director of Tsinghua University's China Institute for Science and Technology Policy (CISTP), Secretary-General of the Academic Committee of Tsinghua University's Center for Science & Technology Development and Governance, Council Member and Deputy Secretary-General of the China Association for Science of Science and S&T Policy, Deputy Director of the Public Management Specialized Committee. Liang's research focuses on science and technology innovation policy, globalization of research and development, standards and intellectual property, and governance of emerging technologies.



唐文明 | Tang Wenming

清华大学人文学院哲学系教授、系主任，清华大学学术委员会委员，人文学院学术委员会主任，清华大学道德与宗教研究院副院长，中华孔子学会秘书长。主要研究领域为伦理学、政治哲学与中国哲学。主要著作有《与命与仁》《近忧》《隐秘的颠覆》《敷教在宽》《彝伦攸斁》《极高明与道中庸》等。

He is currently serves as Professor and Chair of the Department of Philosophy at Tsinghua University Member of Tsinghua University Academic Committee, Chair of the School of Humanities Academic Committee. He is also Vice-Dean of Institute for Ethics and Religions at Tsinghua University, and Secretary general of Chinese Confucius Society. His main interests are Chinese philosophy, Ethics, Political philosophy, and Religious Studies. He has published seven books and more than one-hundred academic papers. Some of his books and papers have been translated into English, French, German and Korean.



吴 岩 | Wu Yan

南方科技大学教授，科幻作家，学者。主持国家首个科幻社科项目。获全国优秀儿童文学奖和 Thomas D. Clareson 奖等国内外奖项。

Professor at Southern University of Science and Technology, science fiction writer, scholar. He presided over China's first social science project on science fiction and has won awards such as the National Outstanding Children's Literature Award and the Thomas D. Clareson Award.



薛 澜 | Xue Lan

清华大学文科资深教授，现任清华大学苏世民书院院长。同时兼任清华大学人工智能国际治理研究院院长，清华大学中国科技政策研究中心主任，清华大学可持续发展研究院联席院长。薛澜教授的研究领域包括公共政策与公共管理，科技创新政策，危机管理及全球治理，并在这些领域中多有著述。2000-2018 年，他曾先后担任清华大学公共管理学院副院长，常务副院长，院长。

Tsinghua University Distinguished Professor of Arts, Humanities, and Social Sciences, Dean of Schwarzman College, Tsinghua University, Director of Tsinghua University's Institute for AI International Governance (I-AIIG), Director of Tsinghua University's Center for China Innovation and Development Strategy, Co-Director of Tsinghua University's Institute for Sustainable Development Goals. Xue's research spans public policy and public administration, science and technology innovation policy, crisis management, and global governance, with extensive publications in each of the fields. From 2000 to 2018, he successively served as Deputy Dean, Executive Deputy Dean, and Dean of the School of Public Policy and Management, Tsinghua University.



赵鼎新 | Zhao Dingxin

浙江大学人文高等研究院院长、芝加哥大学 Max Palevsky 荣休讲席教授。研究领域包括历史社会学、政治社会学和社会科学方法论。中文专著有《社会与政治运动讲义》、《东周战争和儒家国家的形成》、《民主的限制》、《合法性的政治》、《什么是社会学》、《社会科学发问方法》、以及《国家、战争与历史发展：前现代中西模式的比较》

Director of the Institute for Advanced Studies in Humanities, Zhejiang University; and Max Palevsky Professor Emeritus at the University of Chicago. His research encompasses historical sociology, political sociology, and social science methodology. His works include *The Confucian-Legalist Stages: A New Theory of Chinese History and Political Movements*, *Wars in Eastern Zhou Dynasty and the Birth of Confucian-Legalist States and Nation*, *War and Historical Development*.



曾毅 | Zeng Yi

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，北京前瞻人工智能安全与治理研究院院长，联合国人工智能高层顾问机构专家组专家，世界卫生组织人工智能伦理与治理专家组专家，牛津大学人工智能伦理研究所顾问委员会委员。主要研究领域是类脑人工智能、人工智能哲学与伦理、人工智能安全与治理、人工智能与可持续发展。

Research Fellow at the Institute of Automation, Chinese Academy of Sciences, Director of the CAS Center for AI Ethics and Governance, President of the Beijing Institute for Forward-looking AI Security and Governance, Member of the UN High-Level Advisory Body on Artificial Intelligence, Member of the WHO Expert Group on AI Ethics and Governance, Member of the Advisory Board of the Institute for Ethics in AI at the University of Oxford. His research focuses on brain-inspired artificial intelligence, AI philosophy and ethics, AI security and governance, and AI for sustainable development.



宝树 | Baoshu

中国新生代科幻代表作家，北大哲学系本科，比利时鲁汶大学硕士。代表作《时间之墟》《美食三品》等，屡获银河奖、星云奖等，作品被译为多国语言。

one of China's leading new-generation sci-fi writers. He holds a bachelor's degree from the Department of Philosophy, Peking University; and a master's degree from the University of Leuven in Belgium. His works include *Ruins of Time*, *Three Tastings of the Delicacy*. He has won multiple Galaxy and Nebula Awards, with his works translated into many languages.



陈楸帆 | Chen Qiufan

中国科幻作家，代表作《荒潮》《AI 未来进行式》，以“科幻现实主义”风格闻名，作品被译为 20 余种语言，获星云奖、银河奖等奖项。

Chinese sci-fi writer known for *Waste Tide* and *AI 2041*, pioneer of "sci-fi realism". His works have been translated into 20+ languages and have won Galaxy & Nebula Awards.



孙周兴 | Sun Zhouxing

浙江大学敦和讲席教授、图书馆馆长，长江学者。专攻德国哲学与技术艺术哲学，主持翻译《海德格尔文集》等经典，著有《未来哲学序曲》《人类世的哲学》，倡导技术时代的人文创造。

Chair Professor at Zhejiang University and library director, specializes in *German philosophy (Heidegger, Nietzsche)* and techno-art philosophy. He edited *Heidegger's Collected Works*, advocating humanistic innovation in the technological era.



王 珏 | Wang Jue

西安交通大学哲学系教授，研究方向为外国哲学、比较哲学、应用伦理学。近期研究重点为身体现象学、人工智能哲学、技术伦理。

Professor of Philosophy at Xi'an Jiaotong University. Her research interests include foreign philosophy, comparative philosophy and applied ethics. Her recent research focuses on body phenomenology, philosophy of artificial intelligence, and ethics of technology.



张伟特 | Zhang Weite

清华大学新雅书院副院长，清华大学哲学系助理教授，主攻西方近代哲学、元哲学及人工智能哲学。

He serves as Associate Dean of Xinya College at Tsinghua University, Assistant Professor of Philosophy at Tsinghua University, specializing in modern Western philosophy, metaphilosophy, and philosophy of artificial intelligence.



赵汀阳 | Zhao Tingyang

国家文科一级教授，中国社会科学院学部委员，国家文史馆馆员。赵汀阳教授研究存在论、政治哲学和历史哲学，代表理论包括“天下体系”，“漩涡模式”，“动词哲学”等。著有《一个或所有问题》、《第一哲学的支点》、《坏世界研究》、《天下体系》、《天下的当代性》、《惠此中国》、《历史，山水，渔樵》、《寻找动词的形而上学》等。

Member of the Academic Divisions of the Chinese Academy of Social Sciences, and Research Fellow of The Central Research Institute of Culture and History." Zhao has long engaged in philosophical research on ontology, political philosophy, and philosophy of history. His representative theories include "Tianxia System", "Whirlpool Model", "Verb Philosophy", etc. His works include *Alles Unter dem Himmel ; Tianxia tout sous un Meme Ciel ; All-under-heaven ; The Tianxia System for a Possible World Order ; Redefining a Philosophy for World Governance ; Tianxia: Una filosofia para la gobernanza global; Tutto Sotto il Cielo ; Nowa Filozofia ladu Swiatowego.*



Stephen Cave

剑桥大学莱弗赫尔姆智能未来中心主任，同时担任技术与人文研究所联席主任。他出版专著包括《人工智能叙事》、《女性主义人工智能》、《永生》和《你是否选择永生》。他的研究成果被全球超过 100 家主流纸媒报道，并经常受邀在媒体上评论人工智能技术与哲学的相关问题。

Director of the Institute for Technology and Humanity, and of the Leverhulme Centre for the Future of Intelligence, University of Cambridge. His works include *AI Narratives*, *Feminist AI*, *Immortality*, and *Should You Choose to Live Forever?*. His research has been featured in over 100 major global print media outlets, and he regularly comments on AI technology and philosophy-related issues.



Shimon Edelman

康奈尔大学心理学教授。其百余篇学术论文与数十部专著章节涵盖研究主题甚广，涉及运动控制、视觉感知、语言演化与习得，延展至人工智能、心智哲学及意识计算理论领域。近著包括《生与死及其他令人不安的真相》与《意识革命》。

Professor of Psychology at Cornell University. His 100+ academic papers and scores of book chapters span a broad range of research topics, including motor control, visual perception, language evolution and acquisition, and extend to artificial intelligence, philosophy of mind, and computational theories of consciousness. His recent books include *Life, Death, and Other Inconvenient Truths* and *The Consciousness Revolutions*.



Francesco Verso

意大利科幻作家，《未来科幻》丛书编辑。发表科幻作品《Nexhuman》《The Roamers》等，一直探索人类、科技与环境之间的关系。

Francesco Verso, Italian science fiction writer, Editor of *Future Fiction*. His published works include *Nexhuman* and *The Roamers*, which consistently explore the relationship between humanity, technology, and the environment.



Atoosa Kasirzadeh

卡内基梅隆大学哲学系与软件及社会系统系双聘助理教授。她作为国际计算机协会（ACM）公平、问责与透明会议指导委员会成员推动算法治理的全球对话。2025–2027 年，她出任世界经济论坛“全球未来理事会 – 人工智能分委会”理事，深度参与全球 AGI（通用人工智能）治理框架的设计。

Assistant Professor jointly appointed in the Department of Philosophy and the Software and Societal Systems Department at Carnegie Mellon University. She serves on the Steering Committee of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT), driving global conversations on algorithmic governance. From 2025 to 2027, she will serve as a member of the Global Future Council on AI at the World Economic Forum, deeply involved in designing the global AGI (Artificial General Intelligence) governance framework.



Vincent C. Müller

埃尔朗根-纽伦堡大学(FAU)亚历山大·冯·洪堡教席教授, 并担任该校人工智能哲学研究中心(PAIR)主任; 埃因霍温理工大学访问教授; 欧洲认知系统协会主席; 人工智能哲学学会主席; 欧洲机器人技术联盟“伦理、法律与社会经济议题”专家组主席。

Alexander von Humboldt Professor at University of Erlangen-Nuremberg (FAU), Director of FAU's Centre for Philosophy and AI Research (PAIR), Visiting Professor at Eindhoven University of Technology, President of the European Society for Cognitive Systems, Chair of the Society for the Philosophy of Artificial Intelligence, and Chair of the euRobotics topics group on 'ethical, legal and socio-economic issues.'

联系我们



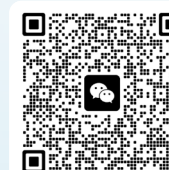
扫描二维码关注公众号
“清华方塘研究院”
获得更多学术活动信息

扫描二维码关注视频号
“方塘思想试验室”
回味方塘学术活动的精彩瞬间



打开小红书扫描二维码
关注“方塘研究院”
加入方塘研究院的红薯地分舵

扫描二维码加入社群
“2025方塘论坛的朋友圈”
获得论坛相关内容的
第一手资料和素材分享



如有大会合作与咨询请邮件联系: fangtang@tsinghua.edu.cn