国际会议

INTERNATIONAL CONFERENCE

北京大学 PEKING UNIVERSITY

FOUNDATIONS AND APPLICATIONS OF AI

人工智能基础与应用国际会议

跨学科视角中的人工智能

INTERDISCIPLINARY PERSPECTIVES

2022.04.08—10

中国·北京·北京大学

PEKING UNIVERSITY, BEIJING, CHINA

北京大学哲学系宗教学系
DEPARTMENT OF PHILOSOPHY AND RELIGIOUS STUDIES
PEKING UNIVERSITY

北京大学人工智能研究院
INSTITUTE FOR ARTIFICIAL INTELLIGENCE, PEKING UNIVERSITY

北京大学
哲学与人类未来研究中心
CENTRE FOR PHILOSOPHY & THE FUTURE OF HUMANITY

北京大学外国哲学研究所
Institute of foreign philosophy, Peking University

目录 | Table of Contents

前言 Preface

This international conference brings together leading experts and aspiring talents from various disciplines working on artificial intelligence, with the specific aim of creating an exchange between theoretically and practically oriented specialists.

To this end, research presentations will include some of the latest advances in engineering and programming, cutting edge research on corresponding legal, ethical, and governance issues, as well as work in psychology, sociology, cognitive science, logic, and philosophy.

AI technology will have an increasingly significant impact on the development of humanity. Existing AI technology (e.g. recommender systems, facial recognition software, medical diagnostics, military robots) and expected advances in the field (e.g. self-driving cars, care robots, virtual reality) raise fundamental questions about what we should do with respect to them, both from an individual and a societal point of view. These questions include how we ourselves should behave as well as how we want the technology to behave as it becomes increasingly autonomous. Risks and opportunities must be carefully weighed against one another based on a holistic understanding of the latest research in the various relevant fields of science and technology, so that we can be maximally prepared for the uncertain future that this disruptive kind of technology brings with it.

In the space of two and a half days, there are scheduled a total of 48 research presentations: 18 keynotes in one-hour individual sessions, plus another 30 sessions of half an hour each (15 in Chinese and 15 in English running parallel). The conference will be fully online and will use the meeting software Zoom. In addition, the conference will provide public live streaming and timely video uploads including on YouTube, Bilibili ('Chinese YouTube'), and WeChat.

The organisers are extremely grateful to the individuals and institutions who have supported this conference. A lot of hard work by a great many people—friends, colleagues, and students—has gone into the conception, planning, and organisation of this event. Many thanks to all of you! Finally, we would like to give special thanks to Prof. Han Shuifa, Director of the Chinese Institute of Foreign Philosophy, for all of his exceptional support and encouragement.

We are looking forward to learning from our speakers and from the discussion of their research during this conference. We wish everyone a fruitful and insightful meeting!

The organisers
Sebastian SUNDAY GRÈVE 王小塞
SUI Tingting 隋婷婷

本次会议旨在为国内外人工智能理论和实践领域的顶尖专家和青年才俊提供交流平台。会议主要聚焦两个方面：人工智能基础相关的理论问题和应用相关的实际问题。

会议报告将囊括工程学和计算机编程方面的一些最新进展，以及相应的法律、伦理和智能治理问题的前沿研究，同时涵盖心理学、社会学、认知科学、逻辑学和哲学方面的相关工作。

近年来，人工智能对人类的影响越来越显著。现有的人工智能技术（如推荐系统，面部识别软件，深度学习医疗诊断等）以及预期将出现的技术（如自动驾驶汽车，护理机器人，虚拟现实等）向人类个体和社会群体提出了很多基础性的问题，如人们应当怎样应对人工智能的发展，以及随着机器自动化的提升，人们对人工智能应当抱有怎样的期待等等；对于这些问题中包含的风险与机遇，人们有必要全面了解相关科技领域的最新研究，对其进行谨慎的权衡，以便最大限度地为这种颠覆性技术可能带来的不确定性做准备。

在为期两天半的会议中，共有48场报告，分别是18场时长为1小时的主题报告，以及30场时长为半小时的分组报告（由同一时段在不同分会场举行的15场中文报告与15场英文报告组成）。本次会议全部线上举行，使用的软件为ZOOM和腾讯会议。此外，会议将通过Youtube、Bilibili以及微信视频号进行直播，会议录频也将上传以上视频网站。

主办方对于支持本次会议的个人和机构深表感谢，谨在此向那些为本次活动的构思、策划和组织添砖加瓦的朋友、同事以及学生们致以诚挚的谢意。最后，我们还要特别感谢北京大学外国哲学研究所所长韩水法教授对于会议的大力支持。
我们十分期待在本次会议上聆听各位报告人富有洞见的研究与讨论。预祝大家在本次会议中增进了解、有所收获！

主办人
Sebastian SUNDAY GRÈVE王小塞
SUI Tingting隋婷婷

会议日程

Schedule

人工智能基础与应用国际会议
跨学科视角中的人工智能

Foundations and Applications of AI Interdisciplinary Perspectives (FAAI)

2022.4.8

13:30 – 14:00 北京时间	开幕式 Opening ceremony (UTC-4/Eastern 01:30; UTC+2/Berlin 07:30) ZOOM
组织者 Organizers	王小塞（北京大学哲学系宗教学系） Sebastian SUNDAY GRÈVE, Department of Philosophy and Religious Studies, Peking University 隋婷婷（北京大学哲学系宗教学系） SUI Tingting, Department of Philosophy and Religious Studies, Peking University
发言嘉宾 Speakers	韩水法（北京大学外国哲学研究所所长） HAN Shuifa, Director of the Chinese Institute of Foreign Philosophy, Peking University 李文新（北京大学人工智能研究院副院长） LI Wenxin, Vice Dean of the Institute for AI, Peking University

开幕式嘉宾 Speakers at the Opening Ceremony

韩水法 HAN Shuifa

韩水法，浙江余杭人，北京大学哲学系教授，博士生导师。北京大学哲学系学术委员会主任、北京大学外国哲学研究所所长、北京大学人文学部委员、德国图宾根大学政治哲学研究中心通讯研究员、中华全国外国哲学学会副理事长、武汉大学珞珈讲座教授。
已出版学术著作七部，译著三部，并在《中国社会科学》、《哲学研究》等核心刊物上发表论文100余篇。在北京大学已开设研究生课程与本科生课程20余门，如《纯粹理性批判》研究、康德《判断力批判》研究、康德实践哲学、罗尔斯《正义论》研究、政治哲学、人工智能时代的人文主义等，并多次受邀至海内外著名学府讲学。
主要研究方向包括康德哲学暨德国唯心主义、政治哲学、韦伯与社会理论、当代中国思想、大学问题、汉语哲学。

HAN Shuifa, born in Yuhang, Zhejiang province, is a professor and doctoral supervisor in the Department of Philosophy at Peking University. He is also Director of the Academic Committee of the Department of Philosophy, Director of the Chinese Institute of Foreign Philosophy, and a member of the Faculty of Humanities of Peking University, Corresponding Researcher at the Political Philosophy Research Center of the University of Tübingen, Germany, Vice-chair of the All-China Foreign Philosophy Society, and Luojia Chair Professor of Wuhan University.
Professor Han has published seven academic books, three translated books, and more than 100 papers in core journals such as *Social Sciences in China* [Chinese journal] and *Philosophical Research* [Chinese journal]. He has taught more than 20 graduate and undergraduate courses in Peking University, including on Kant's *Critique of Pure Reason*, Kant's *Critique of Judgment*, Kant's Practical Philosophy, Rawls' *Theory of Justice*, Political Philosophy, and Humanism during the Age of Artificial Intelligence. Meanwhile, he is frequently invited to give lectures at other famous universities at home and abroad.
His major research areas include Kant's Philosophy and German Idealism, Political Philosophy, Weber and Social Theory, Contemporary Chinese Thought, University Issues, and Philosophy in Chinese.

Website: https://phil.pku.edu.cn/szdw/szll/wgzxjys/274724.htm

李文新 LI Wenxin

北京大学信息科学技术学院教授，北京大学计算机实验教学中心（国家级示范中心）主任，中国计算机学会杰出会员，北京大学人工智能研究院副院长。
主要研究方向为生物特征识别、图像处理、模式识别、基于内容的图像检索等。研究兴趣还包括解决困难问题，目前担任了北京大学ACM/ICPC比赛团队的负责人。
已出版专著3部，发表学术论文50余篇，其中大部分发表在顶级会议和期刊上。李教授因其在计算机编程领域的出色教学而获得了许多奖项。她曾获得北京大学创立者教育优秀教学奖（2009年）、北京大学优秀教学奖（2008年）和北京大学科达教育优秀教学奖（2003年）。此外，还获得了ACM/ICPC亚洲杰出贡献奖（2006,2010）和ACM/ICPC亚洲领导力奖（2016）。

Professor of School of Electronics Engineering and Computer Science; Associate Dean of School of Electronic Engineering and Computer Science, Director of the Computer Experiment Teaching Center (National Demonstration Center) of Peking University, Distinguished member of China Computer Society, deputy Dean of Institute for artificial intelligence, Peking University.
Her research interests include biometrics, image processing, pattern recognition, content-based image retrieval. She is also interested in tough problem solving and now acting as the leader of the Peking University ACM/ICPC competition team.
Dr. Li has published 3 books, and more than 50 research papers, and most of them are published in top-tier conferences and journals. Dr. Li has won many awards for her excellent teaching in the area of computer programming. She received Founder Education Award for Excellent Teaching at Peking University (2009), Excellent Teaching Award at Peking University (2008), and Koda Education Award for Excellent Teaching at Peking University (2003). She also received Asia Outstanding Contribution Award of ACM/ICPC (2006, 2010), and Asia Leadership Award of ACM/ICPC (2016).

Website: http://www.ai.pku.edu.cn/info/1313/1697.htm

Scientific Programme

2022.4.8

特邀嘉宾会场（英文） Keynote Session (ENG) ZOOM	
主持人：安乐哲，北京大学 Host: Roger T. AMES, Peking University	
北京时间 14:00–16:00 (UTC-4/ Eastern 02:00; UTC+2/ Berlin 08:00)	ZENG Yi (Institute of Automation, Chinese Academy of Sciences) “Building and Living with Brain-inspired Ethical AI” 曾毅（中国科学院自动化研究所） 构建符合伦理的类脑人工智能并与其共生
	Pedro DOMINGOS (School of Computer Science & Engineering, University of Washington, USA) “How Will AI Change Ethics?” 佩德罗·多明戈斯（美国华盛顿大学计算机科学与工程系） 人工智能将如何改变伦理学？

平行会场 （中文） Parallel Session (CHI) 腾讯会议	
主持人：梅剑华，山西大学 Host： MEI Jianhua, Shanxi University	
北京时间 16:10–18:30 (UTC-4/ Eastern 04:10; UTC+2/ Berlin 10:10)	涂良川（华南师范大学马克思主义学院） 人工智能“高阶自动化”的主体可能性——兼论人工智能奇点论的存在论追问 TU Liangchuan (School of Marxism, South China Normal University, China) “The Subject Possibility of "Higher Order Automation" in AI——An Inquiry on the Singularity of AI”
	张学义,王晓雪 （东南大学人文学院） “伦理旋钮”：破解无人驾驶算法困境的密钥？ ZHANG Xueyi, WANG Xiaoxue (School of Humanities, Southeast University, China) “A Survey for Ethical Knob of Moral Algorithm of Driverless Car”
	郭喨（浙江大学马克思主义学院） 自动驾驶交通系统中的“人类”与“机器” GUO Liang (School of Marxism, Zhejiang University, China) “‘Human’ and ‘Machine’ in Transportation Systems of Autonomous Driving”
	赵周宽（西安外国语大学中国语言文学学院） 后人类世伦理问题的基本思考 ZHAO Zhoukuan (School of Chinese Language and Literature, Xi’an International Studies University, China) “Basic Thoughts on Ethical Issues in the Post-Human Era”

平行会场（英文） Parallel Session (ENG)	ZOOM
主持人：德米特罗•米哈伊洛夫，东南大学 Host: Dmytro MYKHAILOV, Southeast University	
北京时间 16:10–18:30 (UTC-4/ Eastern 04:10; UTC+2/ Berlin 10:10)	WANG Leye (Center of Frontier Computing Science, Peking University) “Principle of Least Sensing and Computing: Building an Intelligent System with Minimum Data” 王乐业（北京大学前沿计算研究中心） 最小感知计算原则：探索基于最小必要数据的智能系统构建
	KONG Yuqing (Center of Frontier Computing Science, Peking University) “Eliciting Thinking Hierarchy without a Prior” 孔雨晴（北京大学前沿计算研究中心） 无先验学习思考等级
	Konstantin AZAROV (Philosophy Department, University of Bonn, Germany) “John Searle’s Chinese Room, and Its Predecessor, a Short Story “The Game” by Anatoly Dneprov” 艾坦丁（德国波恩大学哲学系） 塞尔的中文屋及作为其原型的阿纳托利·第聂伯夫的小说“游戏”
	SUI Tingting (Department of Philosophy and Religious Studies, Peking University) “The Is-Ought Problem within Moral Algorithm Experiments of Autonomous Vehicle” 隋婷婷（北京大学哲学与宗教学系） 自动驾驶道德算法实验中的“是”与“应当”问题
18:30–20:00	休息 Break

特邀嘉宾会场（英文） Keynote Session (ENG)	ZOOM
主持人：朱利叶斯·舍恩赫尔，北京大学 Host: Julius SCHÖNHERR , Peking University	
北京时间 20:00–23:00 (UTC-4/ Eastern 08:00; UTC+2/ Berlin 14:00)	WANG Yanjing (Department of Philosophy and Religious Studies, Peking University) “Knowing How to Plan” 王彦晶（北京大学哲学与宗教学系） 从“知道如何”到自动规划
	Herman CAPPELEN (Department of Philosophy, University of Hong Kong) “Can AI Transform Philosophy?” 赫尔曼·卡普兰（香港大学哲学系） 人工智能可以变革哲学吗？
	David CHALMERS (Department of Philosophy, New York University) ”From the Matrix to the Metaverse” 大卫·查尔莫斯（纽约大学哲学系） 从黑客帝国到元宇宙

特邀嘉宾会场（英文） Keynote Session (ENG) ZOOM	
主持人：南星，北京大学 Host： NAN Xing, Peking University	
北京时间 8:00–10:00 (UTC-4/ Eastern 20:00; UTC+2/ Berlin 02:00)	Cynthia RUDIN (Computer Science Department & Electrical and Computer Engineering Department, Duke University, USA) “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead” 辛西娅·鲁丁（美国杜克大学计算机科学学院与电气计算工程学院） 使用可解释模型在高风险决策中替代解释黑箱深度学习模型
	Piero SCARUFFI (Department of Chemical Engineering, Stanford University) “Has AI Reached Human Parity?” 皮耶罗·斯卡鲁菲 （美国斯坦福大学化学工程系） 人工智能已经达到与人类同等水平了吗？

平行会场 （中文） Parallel Session (CHI) 腾讯会议	
主持人：郭嘹，浙江大学 Host： GUO Liang, Zhejiang University	
北京时间 10:10–11:50 (UTC-4/ Eastern 22:10; UTC+2/ Berlin 04:10)	赵松（曲阜师范大学政治与公共管理学院） 哲学与人工智能的根本联系：生存论的前提反思 ZHAO Song (School of Politics and Public Administration, Qufu Normal University, China) “Exploration of the Fundamental Connection between Philosophy and Artificial Intelligence”
	李佳欣（北京大学哲学与宗教学系） 人工智能抑或认知心理学——如何理解西蒙的科学发现观 LI Jiaxin (Department of Philosophy and Religious Studies, Peking University) “AI or Cognitive Psychology: How to Understand the Theory of Scientific Development of Simon”
	张伟特（清华大学哲学系） 笛卡尔与人工智能：“我思故我在”作为智能测试标准的可能性 ZHANG Weite (Department of Philosophy, Tsinghua University) “Descartes and artificial Intelligence: the possibility of "I think, therefore I am" as a standard for intelligence testing”

平行会场（英文） Parallel Session (ENG)	ZOOM
主持人：扎克·加勒特，达勒姆Excelsior古典学院 Host： Zack GARRETT, Excelsior Classical Academy	
北京时间 10:10–11:50 (UTC-4/ Eastern 22:10; UTC+2/ Berlin 04:10)	Bongrae SEOK (Department of Humanities/Philosophy, Alvernia University) “AI, Robotics, and Buddhism” 石奉来（美国艾尔弗尼亚大学哲学系） 人工智能，机器人学与佛教
	Dmytro MYKHAILOV (School of Humanities, Southeast University, China) “New AI Applications within Smart-Education Domain: AI Adaptive Educational System and Its Moral Drawbacks” 德米特罗•米哈伊洛夫（东南大学人文学院） 智慧教育领域的人工智能新应用：人工智能教育自适应系统及其道德缺陷
	ZHANG Ping (Law School, Peking University, China) 张萍（北京大学法学院）
11:50–13:00	**休息 Break**

特邀嘉宾会场（中文） Keynote Session (CHI)	ZOOM
主持人：周程，北京大学 Host: ZHOU Cheng, Peking University	
北京时间 13:00–16:00 (UTC-4/ Eastern 01:00; UTC+2/ Berlin 07:00)	成素梅（上海社会科学院哲学研究所） 人工智能的跨学科理解 CHENG Sumei (Institute of Philosophy, Shanghai Academy of Social Sciences) “The Interdisciplinary Understanding of Artificial Intelligence“
	梅剑华（山西大学哲学社会学学院） 儿童哲学与人工智能哲学间的三重对话 MEI Jianhua (School of Philosophy and Sociology, Shanxi University, China) “Three Levels of Dialogue between Philosophy of Children and Philosophy of AI”
	段伟文（中国社会科学院哲学研究所） 深度伪造的认知挑战与伦理治理 DUAN Weiwen (Institute of Philosophy, Chinese Academy of Social Sciences) “The Cognitive Challenges and Ethical Governance of Deep Forgery”

平行会场 （中文） Parallel Session (CHI) 腾讯会议	
主持人：张伟特，清华大学 Host： ZHANG Weite, Tsinghua University	
北京时间 16:10–18:30 (UTC-4/ Eastern 04:10; UTC+2/ Berlin 10:10)	蔡恒进，蔡天琪（武汉大学计算机学院） 人的意识能上传到元宇宙吗？ CAI Hengjin, CAI Tianqi (School of Computer Science, Wuhan University, China) “Is It Possible to Upload Human Consciousness to Metaverse?”
	简圣宇（扬州大学美术与设计学院） “元宇宙”问题：技术化娱乐的人文之思 JIAN Shengyu (Art and Design Academy, Yangzhou University) “The Metaverse: Humanistic Thoughts on Technologized Entertainment”
	吴东颖（中国科学院哲学研究所） 吴小安（西北工业大学马克思主义学院） 因果贡献度的困境与证成 WU Tung-Ying (Institute of Philosophy, Chinese Academy of Sciences) WU Xiao’an (Northwestern Polytechnical University, China) “Defending Degrees of Causal Contribution”
	李欣怡（复旦大学哲学学院） 历时地衡量公平：基于因果结构理论的分析 LI Xinyi (School of Philosophy, Fudan University) “Measuring Fairness Diachronically: a Causal Structural Approach”
平行会场 （英文） Parallel Session (ENG) ZOOM	
主持人：郭春宁，中国人民大学 Host： GUO Chunning, Renmin University of China	
北京时间 16:10–18:30 (UTC-4/ Eastern 04:10; UTC+2/ Berlin 10:10)	Giada PISTILLI (Department of Philosophy, Sorbonne University, France) “Ethical Frameworks as a Moral Exercise in the Field of Natural Language Processing” 嘉坦·皮斯蒂利（巴黎索邦大学哲学系） 自然语言处理中作为道德训练的伦理架构
	SONG Jianli (School of Marxism, Tianjin University, China) “The Acceleration and Inequality of Artificial Intelligence under the Century of Epidemic” 宋建丽（天津大学马克思主义学院） 世纪疫情下人工智能的加速与不平等挑战

北京时间 16:10– 18:30 (UTC-4/ Eastern 04:10; UTC+2/ Berlin 10:10)	Timothy TAMBASSI (DISPAC, University of Salerno, Italy) “On the Mutual Understanding between Human Beings and Software Systems. The Role of Information System Ontologies in Artificial Intelligence” 提摩西·坦巴西（意大利萨莱诺大学文化遗产学系） 人类与软件系统间的相互理解——人工智能中信息系统存在论的作用
	ZHANG Mengting (Department of Philosophy, Sun Yat-Sen University, China) “A New Understanding of the Role of Dreaming From the Perspective of Model” 张梦婷（中山大学哲学系） 对梦的新解释：从模型的角度看
18:30–20:00	休息 Break

特邀嘉宾会场（英文） Keynote Session (ENG) ZOOM	
主持人：隋婷婷，北京大学 Host: SUI Tingting, Peking University	
北京时间 20:00– 22:00 (UTC-4/ Eastern 08:00; UTC+2/ Berlin 14:00)	Susan SCHNEIDER (Department of Philosophy, Florida Atlantic University) “AI-based Brain Enhancements, Superintelligence and the Future of the Mind” 苏珊·施奈德(佛罗里达大西洋大学哲学系) 基于人工智能的大脑强化、超智能与心灵的未来
	LUO Huan (School of Psychological and Cognitive Sciences, Peking University) “Relational Structure Knowledge in the Human Brain: Representation, Memory, and Learning” 罗欢（北京大学心理与认知科学学院） 人脑中的关系知识结构：表征、记忆和学习

特邀嘉宾会场（中文） Keynote Session (CHI) ZOOM	
主持人：李麒麟，北京大学 Host: LI Qilin, Peking University	
北京时间 10:00–11:00 (UTC-4/ Eastern 21:00; UTC+2/ Berlin 03:00)	朱菁（厦门大学人文学院） 人工智能可解释性问题的哲学探索 ZHU Jing (College of Humanities, Xiamen University, China) "A Philosophical Exploration of the Interpretability Problem of Artificial Intelligence"
平行会场 （中文） Parallel Session (CHI)	**腾讯会议**
主持人：涂良川，华南师范大学 Host: TU Liangchuan, South China Normal University	
北京时间 11:10–12:15 (UTC-4/ Eastern 23:10; UTC+2/ Berlin 05:10)	周旅军（中华女子学院社会工作学院） 人工智能中的性别不平等治理何以可能？——基于文化堕距视角的算法规制路径分析 ZHOU Lujun (School of Social Work, China Women's University) "How is the Governance of Gender Inequality in Artificial Intelligence Possible: Analysis on Algorithm Regulation Path from the Perspective of Cultural Lag"
	沈苑，汪琼（北京大学教育学院） 人工智能教育应用的价值敏感设计 SHEN Yuan, WANG Qiong (Graduate School of Education, Peking University) "Value-sensitive Design of Artificial Intelligence Education Application"
特邀嘉宾会场（英文） Keynote Session (ENG) ZOOM	
主持人：宋建丽，天津大学 Host: SONG Jianli, Tianjin University	
北京时间 11:10–12:15 (UTC-4/ Eastern 23:10; UTC+2/ Berlin 05:10)	Zack GARRETT (Excelsior Classical Academy, USA) "A.I. and Authorial Intent" 扎克·加勒特（美国达勒姆Excelsior古典学院） 人工智能和作者的意图

	GUO Chunning (School of Arts, Renmin University of China) “The Hyper-Linguistic Transformation of Science Fiction Films: The ‘Embodied’ Narration and Broadcasting of Artificial Intelligence” 郭春宁（中国人民大学艺术学院） 科幻电影的超语言转变：人工智能的具身化叙事与传播
12:15–13:00	**休息 Break**

特邀嘉宾会场（中文） Keynote Session (CHI) ZOOM	
主持人：陆俏颖，北京大学 Host: LU Qiaoying, Peking University	
北京时间 13:00–15:00 (UTC-4/ Eastern 01:00; UTC+2/ Berlin 07:00)	杨玉超（北京大学集成电路学院） 基于忆阻器的高效AI芯片与学习系统 YANG Yuchao (School of Integrated Circuits, Peking University) “Memristor Enabled High-efficiency AI Chips and Learning Systems”
	谭营（北京大学计算智能实验室） 群体智能及其应用研究进展 TAN Ying (Computational Intelligence Laboratory, Peking University) “Advances in Swarm Intelligence and Its Applications”
平行会场 （中文） Parallel Session (CHI) 腾讯会议	
主持人：张学义，东南大学 Host: ZHANG Xueyi, Southeast University	
北京时间 15:10–16:15 (UTC-4/ Eastern 03:10; UTC+2/ Berlin 09:10)	张琨（北京大学医学人文学院） 论医疗决策中人工智能技术运用的法律责任——以给药剂量为例 ZHANG Kun (School of Health Humanities, Peking University) “On Legal Liability for the Use of Artificial Intelligence Technology in Medical Decision Making: The Example of Drug Dosing”
	陶峰（南开大学哲学院） 人工智能模拟时代的艺术品 TAO Feng (College of Philosophy, Nankai University, China) “The Artworks in the Era of Artificial Intelligence Imitability”

平行会场（英文） Parallel Session (ENG)	ZOOM
主持人：嘉坦·皮斯蒂利，索邦大学 Host: Giada PISTILLI, Sorbonne University	
北京时间 15:10–16:15 (UTC-4/ Eastern 03:10; UTC+2/ Berlin 09:10)	JIANG Cong, LIU Lu (Law School, Peking University) “Justifying the Interpretability of Judicial AI: Legal Sense as the Starting Point” 姜聪，刘露（北京大学法学院） 为司法人工智能的可解释性辩护：以法感为出发点
	Yvonne FÖRSTER (School of Philosophy and Sociology, Shanxi University Taiyuan, China / Leuphana University Lüneburg, Department of Philosophy and Art Science, Germany) “Artificial Intelligence: Dialectics of Transparency” 伊冯娜·弗尔斯特（山西大学哲学社会学学院/德国吕讷堡大学哲学系） 人工智能：透明度的辩证法
特邀嘉宾会场（英文） Keynote Session (ENG)	**ZOOM**
主持人：陈海丹，北京大学 Host: CHEN Haidan, Peking University	
北京时间 16:30–18:30 (UTC-4/ Eastern 04:30; UTC+2/ Berlin 10:30)	Lena KÄSTNER (Department of Philosophy, University of Bayreuth, Germany) “On Modeling Psychopathology: Eight Challenges and How to Meet Them” 莉娜·卡斯特纳（德国拜罗伊特大学哲学系） 精神病理学建模：八个挑战及其应对
	GE Jianqiao (Academy for Advanced Interdisciplinary Studies, Peking University) “Does the Human Brain Think that Artificial Intelligence is Artificial?” 葛鉴桥（北京大学前沿交叉学科研究院） 人脑会认为人工智能是人工的吗？

详细会议信息

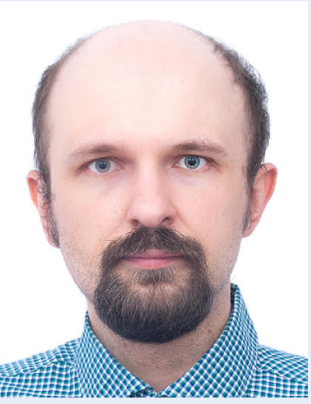
Detailed information

Speaker Profiles and Abstracts

(By SURNAME in Alphabetical Order)

Speaker List (By SURNAME in Alphabetical Order)

A-J	K-S	T-Z
A Konstantin AZAROV **C** CAI Hengjin, CAI Tianqi Herman CAPPELEN David CHALMERS CHENG Sumei **D** Pedro DOMINGOS DUAN Weiwen **F** Yvonne FÖRSTER **G** Zach GARRETT GE Jianqiao GUO Chunning GUO Liang **J** JIAN Shengyu JIANG Cong, LIU Lu	**K** Lena KÄSTNER KONG Yuqing **L** LI Jiaxin LI Xinyi LIU Xiaoli LUO Huan **M** MEI Jianhua Dmytro MYKHAILOV **P** Giada PISTILLI **R** Cynthia RUDIN **S** Piero SCARUFFI Susan SCHNEIDER Bongrae SEOK SHEN Yuan, WANG Qiong SONG Jianli SUI Tingting	**T** Timothy TAMBASSI TAN Ying TAO Feng TU Liangchuan **W** WANG Leye WANG Yanjing WU Tung-Ying, WU Xiaoan **Y** YANG Yuchao **Z** ZENG Yi ZHANG Mengting ZHANG Kun ZHANG Ping ZHANG Weite ZHANG Xueyi, WANG Xiaoxue ZHAO Song ZHAO Zhoukuan ZHOU Lujun ZHU Jing

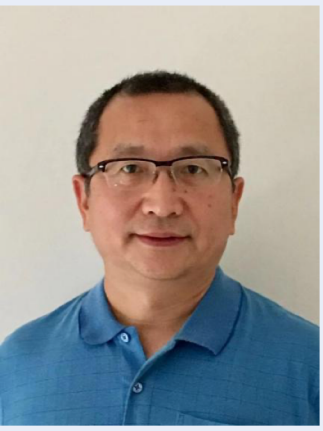


Konstantin AZAROV 艾坦丁
Philosophy Department, University of Bonn, Germany 德国波恩大学哲学系

Bio 简介: I'm a postdoc at the University of Bonn, Philosophy department. I work in the field of comparative aesthetics, with special expertise in Chinese (Laozi), Russian (Tolstoy), and German (Kant) traditions. I also received training as a painter in both Western and Eastern (国画) traditions. This experience gives me an opportunity to see deeper layers in art. My interest in the problems of AI is a result of cooperation with the Center for Media Philosophy at St. Petersburg State University. I graduated with honors from St. Petersburg State University in 2009. In 2009-2011 I was a lecturer (philosophy) at St. Petersburg Affiliation of Jewish Studies. I defended my PhD thesis in 2019. In 2020 I was a postdoc at National Taiwan University. Since 2020, I have been working on Tolstoy's reading of Laozi, as well as on the philosophy of technology and AI by Dneprow.
Website 主页： N/A

Title 标题: John Searle's Chinese Room, and Its Predecessor, a Short Story "The Game" by Anatoly Dneprov
塞尔的中文屋及其原型，阿纳托利·第聂伯夫的小说"那个游戏"

Abstract 摘要: There are two views against the possibility of strong AI. The first is centered around John Searle's well known 1980 skeptical thought experiment, "the Chinese Room." However, there is a second, largely unknown, view, that preceded Searle's. This second view is in Anatoly Dneprov's 161 short story, "The Game." Both ask: "can computers think?" both answer: "no." Although Searle's experiment received a lot of criticisms, "The Game" did not receive any response. However, if we take into account Dneprov's 'The Game' together with his other works, we will find that Dneprov's argument is developed to a level comparable to Searle's. This paper argues that the radical difference between the two is the perception of the brain. Searle believes that machines can think, but the human brain is the only machine capable of thinking, while Dneprov would not agree. For him, the brain is part of an organism and an organism cannot be regarded as a machine. Therefore, Dneprov's work can be helpful to challenge conceptions of the brain as a machine in the current debates on AI.



CAI Hengjin 蔡恒进 (First Author 第一作者)
School of Computer Science, Wuhan University, China 武汉大学计算机学院

Bio 简介：蔡恒进，武汉大学计算机学院教授、博导，吴文俊人工智能科学技术奖获得者，中国移动通信联合会元宇宙产业委员会副主任委员，2021中国准独角兽百强企业杭州宇链科技联合创始人，武汉市校友经济促进会副会长，开放原子开源基金会顾问，Springer国际期刊《AI and Ethics》编委，中国碳中和50人论坛特邀研究员，中国人工智能学会（CAAI）人工智能伦理道德专业委员会委员，CSIAM（中国工业与应用数学学会）区块链专委会委员，中国计算机学会（CCF）区块链专委会委员，CCF服务计算委员会委员，CCF理论计算委员会委员，CCFAI智能体及多智能体系统学组成员，中国通信工业协会区块链专业委员会（CCIAPCB）副主任委员，出版了一系列人工智能、区块链和元宇宙的著作。

Website 主页：
http://cs.whu.edu.cn/teacherinfo.aspx?id=353

CAI Tianqi 蔡天琪 (Coauthor 合作作者）
School of Computer Science, Wuhan University, China 武汉大学计算机学院

Bio 简介：蔡天琪，清华大学深圳国际研究生院博士后，中国工业和应用数学学会区块链专委会委员。武汉大学管理科学与工程博士、金融信息工程硕士、软件工程学士。首创了基于通证的记账模型和基于哈希交互的下一代互联网架构。60余项人工智能、大数据、区块链技术专利共同发明人。《机器崛起前传——自我意识与人类智慧的开端》、《人机智能融合的区块链系统》《区块链：链接智能未来》等著作联合作者。

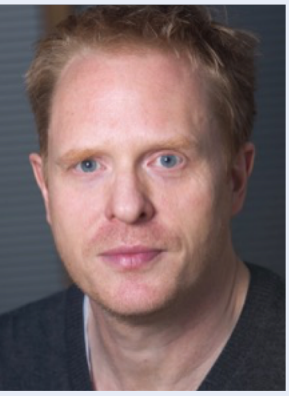
Website 主页：N/A

Title 标题： 人的意识能上传到元宇宙吗？

Is It Possible to Upload Human Consciousness to Metaverse?

Abstract 摘要：专用人工智能技术不断推进，但类人思维的人工智能技术仍然亟待突破，未来人机共同参与的社会治理亦面对重重挑战。这些问题都要求我们必须更加深入探寻人机本质，在“人更像机器”还是“机器更像人”之间做出抉择。我们主张后者才是对人机未来更有益的发展方向。要实现这一目标，就需要探讨人类意识能否传递以及如何传递给机器。对人类而言，意识单元（我们称之为“认知坎陷”）的产生与身体相关，如果意识脱离了人的身体上传给机器，就需要对意识分级并讨论其可迁移的程度。物理数学规律具有绝对可迁移性，意识世界的产物具备相对可迁移性。人能够将抽象概念通过附着与隧通进行具象表达并优化，在人际、代际间传播并达成共识，即形成了具备可迁移性的认知坎陷，而“自我”就是其中最原初的意识单元。通过认知坎陷工程化的方式，有可能让机器形成“自我”原型，习得隧通与迁移，实现机器类人思维的突破，即使如此，人的意识也无法完整地上传给机器。人机共融将有可能在第三代互联网或元宇宙中率先进行尝试并有望实现。

Artificial Intelligence technology of specific disciplines continues to surge, while the development of AI technology with human-like thinking appeals to remain stalling. Confronting with the governance challenges of the human-computer-coexistence society in the future, it is imperative to explore the human nature compared to machines, and to make a choice between making human beings more machine-like or making machines more human-like, we prefer the latter, and it is required to understand whether and how human consciousness is uploaded to machines. A man can transform and optimize abstract concepts into concrete expressions, we call those cognitive attractors, which are transferable among people and between generations by reaching consensuses. Human consciousness, which originates from human bodies, can be scaled based on its transferability. For example, the physical laws and mathematical theorems are absolutely transferable, and other cognitive attractors are relatively transferable. “Self” is the most important and primitive cognitive attractor and can be concreted by bandwagoning and tunneling to other cognitive attractors, which provides an engineering possibility for machines to form their prototypes of “Self”, and to acquire human consciousness. However, human consciousness can not be completely uploaded to machines even though the breakthrough of human-like thinking machines is achieved. Nevertheless, human-computer-coexistence will be firstly attempted and actualized in Web 3.0 or Metaverse.



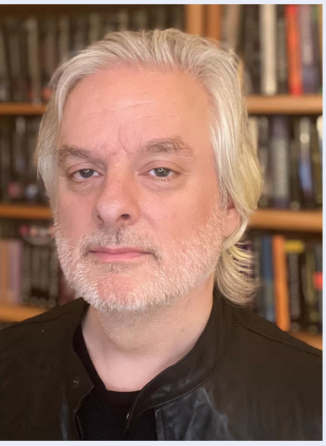
Herman CAPPELEN 赫尔曼·卡普兰
Department of Philosophy, University of Hong Kong
香港大学哲学系

Bio 简介 ：Herman Cappelen is a philosopher. He currently works as a Chair Professor of philosophy at the University of Hong Kong. Before He moved to Hong Kong, He worked at the Universities of Oslo, St Andrews, and Oxford. His current research focus is on the philosophy of AI, Conceptual Engineering, the conceptual foundations of political discourse, externalism in the philosophy of mind and language, and the interconnections between all of these. However, His philosophical interests are broad - they cover more or less all areas of systematic philosophy. He recently finished a book manuscript combining themes from conceptual engineering and political philosophy. The tentative title is: *Better than "Democracy": Conceptual Engineering as Conceptual Abandonment*.

Website 主页： https://www.hermancappelen.net/

Title 标题： Can AI Transform Philosophy?
人工智能可以变革哲学吗？

Abstract 摘要： David Chalmers in his book *Reality+* argues that a new field, Technophilosophy, both allows us to use old-style philosophy to solve new problems (about AI), and also, and most intriguingly, opens for the possibility that we can use AI to solve old philosophical problems. The first part of the talk is a critical examination of Chalmers' ideas about technophilosophy. The second part outlines a new vision for what technophilosophy should be: Technophilosophy should primarily be about the expansion and revision of core philosophical concepts such as *person*, *communication*, *meaning*, *thinking*, *belief*, *rationality*, *obligation*, *responsibility*, and *trust*. AI will result in these concepts being revised and so also transform philosophy.



David CHALMERS 大卫·查尔莫斯
Department of Philosophy, New York University 纽约大学哲学系

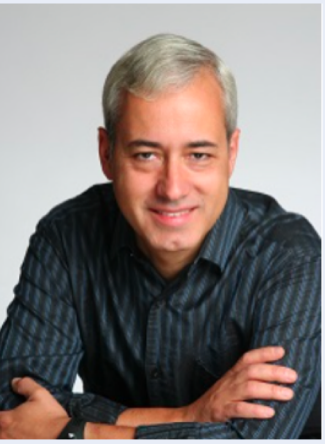
Bio 简介：David Chalmers is University Professor of Philosophy and Neural Science and co-director of the Center for Mind, Brain, and Consciousness at New York University. He is the author of *The Conscious Mind* (1996) and of *Reality+: Virtual Worlds and the Problems of Philosophy* (2022). He is known for formulating the "hard problem" of consciousness, which inspired Tom Stoppard's play The Hard Problem, and for the idea of the "extended mind," which says that the tools we use can become parts of our minds.
Website 主页： http://consc.net/

Title 标题： From the Matrix to the Metaverse 从黑客帝国到元宇宙
Abstract 摘要： N/A

CHENG Sumei 成素梅
Institute of Philosophy, Shanghai Academy of Social Sciences 上海社会科学院哲学研究所

Bio 简介 :成素梅，理学学士，哲学博士，上海社会科学哲研究所副所长，二级教授，《哲学分析》杂志主编（执行）；国务院政府特殊津贴专家，教育部新世纪优秀人才支持计划入选者，上海市领军人才，国际“逻辑学、方法论和科学技术哲学协会”（CLMPST）理事，中国自然辩证法学会常务理事，中国自然辩证法学会物理学哲学分会副主任，上海市哲学学会副会长；曾担任上海市自然辩证法学会理事长，获上海市先进工作者和上海市三八红旗手荣誉称号；学术成果曾获上海市哲学社会科学优秀成果一等奖；牛津、剑桥、斯坦福、京都、哥本哈根以及南加州等大学访问学者；研究方向为科学实在论、量子力学哲学、人工智能哲学和休闲哲学。
Website 主页：
https://www.sass.org.cn/2018/1023/c1230a26235/page.htm

Title 主题：人工智能的跨学科理解
The Interdisciplinary Understanding of Artificial Intelligence
Abstract 摘要：N/A

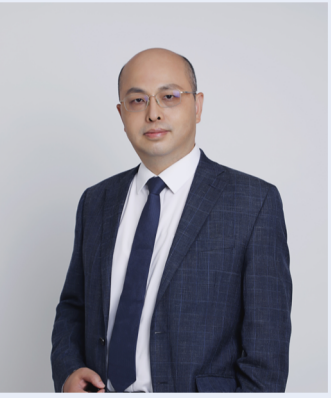


Pedro DOMINGOS 佩德罗·多明戈斯
School of Computer Science & Engineering, University of Washington, USA 美国华盛顿大学计算机科学和工程系

Bio 简介：Pedro Domingos is a professor of computer science at the University of Washington and the author of *The Master Algorithm*. He is a winner of the SIGKDD Innovation Award and the IJCAI John McCarthy Award, two of the highest honors in data science and AI, and a Fellow of AAAS and AAAI. His research spans a wide variety of topics in machine learning, artificial intelligence, and data science. He helped start the fields of statistical relational AI, data stream mining, adversarial learning, machine learning for information integration, and influence maximization in social networks.
Website 主页：
https://homes.cs.washington.edu/~pedrod/

Title 主题： How Will AI Change Ethics?
人工智能将如何改变伦理学？

Abstract 摘要： The ethical issues surrounding AI have received a lot of attention lately, but unfortunately it's all been about shoehorning AI into current Western ethical norms. But AI will dramatically change society and therefore ethics, as did previous technological revolutions (e.g., printing, the pill). This talk will examine how AI might change our views of what's ethical and what's not, and how we can prepare for it. Areas covered will include privacy and data sharing, fairness and equality, work and life, self-driving machines and trolleyology, war and intelligent weapons, democracy vs. authoritarianism, social media, and what distinguishes us from machines. Possible approaches to AI's ethical impacts take inspiration from Buckley, Kay and Marx.



DUAN Weiwen 段伟文
Institute of Philosophy, Chinese Academy of Social Sciences 中国社会科学院哲学研究所

Bio 简介:段伟文，中国社科院哲学所科技哲学研究室主任、研究员，中国社科院科学技术和社会研究中心主任，中国社科院大学教授、博导,享受国务院特殊津贴专家。他获得了华中师范大学物理学学士学位、中国人民大学科技哲学硕士和博士学位，曾为牛津大学、匹兹堡大学访问学，博古睿中国中心学者（2020-2021）。

他的主要研究领域为科技哲学、科技伦理、信息通信技术及大数据和人工智能的社会和伦理研究等，现为《自然辩证法研究》副主编，国际期刊 “负责任的创新(RI)”、“社会中的信息、传播与伦理杂志(JICES)”编委，国家社科基金重大项目“智能革命与人类深度科技化前景的哲学研究”首席专家等，著有《信息文明的伦理基础》(2020)、《可接受的科学：当代科学基础的反思》(2014)、《网络空间的伦理反思》（2002）等。

目前担任中国大数据专家委员会副主任委员，中科院学部科学规范与伦理研究支撑中心学术顾问，中国计算机学会职业伦理与学术道德委员会常务委员，“浙江互联网金融联合会金融科技伦理(专业)委员会”学术委员，美团人工智能治理委员会顾问委员等职。

Duan Weiwen is the Director and Professor of the Department of Philosophy of Science and Technology in the Institute of Philosophy, *Chinese Academy of Social Sciences* (CASS), the director of the Research Center for Science, Technology and Society at CASS, a professor and doctoral supervisor at the *University of the Chinese Academy of Social Sciences*, and a special allowance expert of the State Council of China. He holds a Bachelor of Science degree in Physics from *Central China Normal University*, and a Master of Philosophy and PhD degree in Philosophy of Science and Technology from *Renmin University of China*. He was a visiting scholar in *Oxford University*, *Colorado School of Mines*, and *University of Pittsburgh*. He is also a Berggruen China Center Fellow (2020-2021).

His main research areas are philosophy of science and technology, ethics of science and technology, social and ethical issues of ICTs, Big Data and AI. He is Associate Editor of *Dialectical Studies in Dialectics of nature*, and he is on the editorial board of the *Journal of Responsible Innovation* and the *Journal of Information, Communication and Ethics in Society*. He is the chief researcher of "Philosophical Studies on Intelligence Revolution and Deepening Techno-scientific of Human Being (2017-2022)", which is an important project supported by National Social Sciences Founding of China (NSSFC). He is the author of several books, including *The Ethical Foundation of Information Civilization* (2020, Shanghai People's Press), *Acceptable Science: Reflection on the Foundation of Contemporary Science* (2014, Science and Technology Press of China), *Ethical Reflection on Cyberspace* (2002, Jiangsu People's Publishing House), etc.

He is one of the deputy chairmen of the *Committee of Big Data Experts of China*; an academic advisor to the Support Center for Research on Scientific Norms and Ethics, Academic Division, *Chinese Academy of Sciences*; Standing member of the Committee on Ethics and Professional Conduct, *China Computer Federation* (CCF); Academic member of Zhejiang Province Financial Technology Ethics Committee; Advisory member of Meituan's AI Governance Committee, etc.

Website 主页：
http://philosophy.cssn.cn/yjry/zgjzc/201811/t20181113_4774763.shtml

Title 标题： 深度伪造的认知挑战与伦理治理
The Cognitive Challenges and Ethical Governance of Deep Forgery

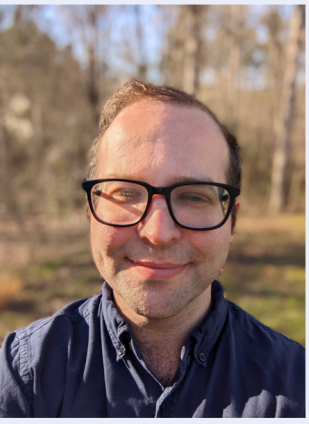
Abstract 摘要：N/A

Yvonne FÖRSTER 伊冯娜·弗尔斯特
School of Philosophy and Sociology, Shanxi University Taiyuan, China / Department of Philosophy and Art Science, Leuphana University Lüneburg, Germany 山西大学哲学社会学学院/ 德国吕讷堡大学哲学系

Bio 简介: Yvonne Förster is appointed as Foreign Expert and Research Professor at Shanxi University Taiyuan, China and teaches Philosophy at Leuphana University Lüneburg, Germany. She has received her PhD from the Friedrich Schiller University Jena, Germany with a thesis on experience and ontology of time. She has been awarded Senior Research Fellowships at two Institutes for Advanced Studies (*Media Cultures of Computer Simulation* at Leuphana and Cultural Sciencesat University of Konstanz) and was a visiting Professor at Bauhaus University Weimar and University of Kassel, Germany. Currently she is senior research fellow at GCAS College Dublin. Her research focuses on human-machine relations, critical posthumanism, the future of technology, theories of embodiment and fashion as art. Recent publications include: (2023, forthcoming) Technologies of Religion: From Prayer Bots to a Looming Singularity, in: Nathan Loewen and Agnieszka Rostalska (Eds.), Philosophy of Religion around the World: Critical Perspectives and Approaches, Vol. 1, London: Bloomsbury. (2020) Aesthetics of the Past and the Future, Human Life within Changing Environments, in: Zoltan Somhegyi und Max Ryynänen (Eds.): *Aesthetics in Dialogue*, Bern: Peter Lang, 2020, Chapter 16, 237-250.
Website 主页： https://yvonnefoerster.com/

Title 标题： Artificial Intelligence: Dialectics of Transparency
人工智能：透明度的辩证法

Abstract 摘要： Democracy is all about transparency, visibility, and public engagement. In the Greek polis, political decisions were discussed in the agora, a public place where all citizens could listen and engage. Representational democracy today is less publicly embodied, but transparency of decision processes is of the utmost importance. If a government cannot make its decisions transparent enough, it runs the risk of losing the people's trust. Transparency implies rules, visibility, and the readiness to argue and give reasons. With the emergence of AI applications not only in the political sphere but in basically every aspect of social and private life, we are faced with new forms of opacity, which strongly impact human decision making, behavior, movement, and communication. The central problem is that AI applications perform as black boxes. The underlying causal processes and reasons for decisions remain opaque. Artificial agency is becoming a ubiquitous phenomenon and thus calls for ethical and phenomenological reflection. This is an even bigger problem in the case of emerging technologies like the Internet of Things, the Internet of People and the envisaged Internet of Everything. Technologies are merging with everyday objects, surfaces, urban and home environments, and our very own bodies. In my talk, I aim at understanding the way in which AI systems introduce a new kind of opacity to the ecological structures of the life-world and how to tackle this problem from a philosophical perspective.



Zack GARRETT 扎克·加勒特
Excelsior Classical Academy in Durham, North Carolina, USA 美国达勒姆Excelsior古典学院

Bio 简介: Zack Garrett received his Ph.D. in philosophy from the University of Nebraska - Lincoln in 2020. He currently teaches logic and philosophy at Excelsior Classical Academy in Durham, North Carolina. His research interests include topics that lie at the intersections between metaphysics, philosophy of language, and logic. He is particularly interested in non-classical logics and impossible worlds.
Website 主页： http://www.zack-garrett.com/

Title标题： A.I. and Authorial Intent
人工智能和作者的意图
Abstract 摘要： Fiction and non-fiction created by artificial intelligence pose a problem for theories of fiction and truth-in-fiction that put a large emphasis on the role of authorial intent. Since current text generating artificial intelligences lack intent, the fictions they create lack authorial intent. The onus for evaluating them, then, can only rest with the readers and the meanings of the words that compose the texts. I argue first that the complexity of fiction writing AIs is not sufficient for us to ascribe to them intent or creativity. Second, I argue that some of the texts that can be generated by future AIs should count as fiction. I, then, explore the consequences of this conclusion for a variety of theories of fiction including those put forward by Deutsch, Currie, Walton, and Friend. Of these, I argue that Walton and Friend are best suited to dealing with AI written fiction. I also explore the consequences for theories of truth-in-fiction. In particular, I look at what would count as the collective belief world for an AI written fiction and who counts as an informed reader of an AI written fiction. Finally, I introduce questions about the ontological status of the fictional characters that appear in AI written fiction.

GE Jianqiao 葛鉴桥
Academy for Advanced Interdisciplinary Studies, Peking University 北京大学前沿交叉学科研究院

Bio 简介: Dr. Ge is teaching at Academic for Advanced Interdisciplinary Studies(AAIS), Peking University and she is awarded 2021-2022 Berggruen Fellow. Her research interests are focused on scientific studies of brain intelligence and social cognition. She is the the Principle Investigator/Co-Investigator of more than 10 research grants supported by Ministry of Science and Technology of China, National Natural Science Foundation of China, and Beijing Municipal Science & Technology Commission. She has published more than 20 referred research articles on leading academic journals such as PNAS, and awarded 2 national patents. Dr.Ge received her Ph.D. in psychology, B.S in physics, a double B.S in mathematics and applied mathematics, and a double B.S in economics from Peking University. She was a postdoctoral fellow at the University of Chicago.
葛鉴桥现任北京大学前沿交叉学科研究院讲师，2021-2022博古睿学者，主要专注于大脑智能与认知功能的研究，主持并参加多项国家自然科学基金项目、北京市科委脑计划重大专项等，在国际前沿期刊如PNAS等发表多篇论文，曾获两项国家发明专利。美国芝加哥大学博士后，北京大学心理学博士、物理学学士、并获数学与应用数学双学士、经济学双学士。
Website 主页：
https://www.berggruen.org/people/jianqiao-ge/

Title Does the Human Brain Think that Artificial Intelligence Is Artificial?
人脑会认为人工智能是智能的吗？
Abstract 摘要：N/A

GUO Chunning (Maggie) 郭春宁
School of Arts, Renmin University of China 中国人民大学艺术学院

Bio 简介：Chunning (Maggie) Guo teaches New Media Art and Animation in the School of Arts at Renmin University of China since 2003 until now. She got PhD degree in Aesthetics of Philosophy School, Renmin University of China in 2015. Her PhD paper is The Dialogue of Hetero-geneity: The Medium Experiment and the Criticism of Independent Animation. She was the recipient of the Jury Award at the 11th annual Chinese Independent Film Festival in China, and the NETPAC Award in 2015 Busan International Short Film Festival of South Korea. Her collaborative work Ketchup was selected for several International Festivals, including FANTOCHE oInternational Animation Festival in Switzerland, the Stuttgart International Animation Festival in Germany, Anima Mundi International Animation Festival in Brazil, World Festival of Animated Film Zagreb in Croatia, A Long Week f Short Films Festival in Shanghai, China, as well as 2016 Berlin Chinese Film Festival, etc. Her Chinese papers and translation papers were published on local and international journals, including Contemporary Cinema, Croatian Cinema Chronicle Film Journal, Art Education, Cartoon and Animation Studies, Epistémè, etc.

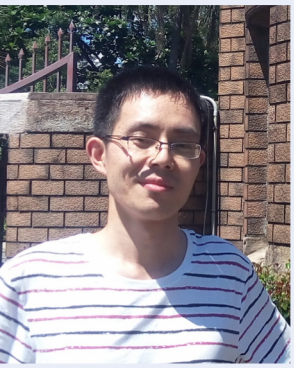
Website 主页：http://art.ruc.edu.cn/info/1150/4866.htm

Title 标题： The Hyper-linguistic Transformation of Science Fiction Films: The “Embodied” Intelligence Narration and Broadcasting of Artificial
科幻影像的超语言转向：人工智能之具身化叙事与传播

This work is coauthored with:

TAO Chen, School of Literature, Renmin University of China
陈涛，文学院，中国人民大学
SUN Xuehao, School of Arts, Renmin University of China
孙雪浩，艺术学院，中国人民大学

Abstract 摘要： Science fiction films, which emerged in the 1960s, mapped the Cold War confrontation and post-industrial culture while presenting a "hyper-linguistic turn" in the rhetoric of images. Due to the rapid development of computer technology, the films from the 1980s onward constructed new science fiction images, especially through the "embodiment" of artificial intelligence, and initiated the mythical narrative of the post-human body. This paper re-examines the image of artificial intelligence in science fiction films as an audiovisual document and archive of "machine language". This paper treats AI as a fourth level of symbolic system, i.e., the "code of codes". The paper focuses on the language, rhetoric, learning and communication abilities of AI in science fiction images, and analyzes the ephemerality and co-temporality of language with the help of cybernetic framework theory, and attempts to summarize and analyze the image characteristics of AI in three stages of cybernetics through the "language dimension", namely dynamic equilibrium, anti-cybernetics and virtual breakout.

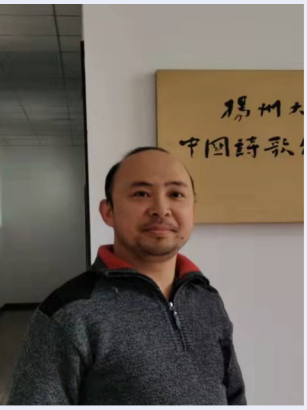


GUO Liang 郭嘹
School of Marxism, Zhejiang University, China 浙江大学马克思主义学院

Bio 简介: 郭嘹，男，湖北郧阳人。科学哲学博士，浙江大学教师，浙江大学科技与法律研究中心研究人员。
当前主要研究兴趣为：自动驾驶的技术、法律与社会治理。倡导并实践“以技术的手段解决伦理的问题”，发表CSSCI期刊论文十余篇。另有《中国自动驾驶的法律挑战》、《自动驾驶人文研究应当扎实可靠》、《“无人驾驶”的技术路线》（郭嘹、唐兴华）等学术小品多篇。
Website 主页：N/A

Title 标题： 扎实可靠 交通系统与自动驾驶中的“人类”和“机器”：路权和人权
“Human" and "Machine" in Transportation Systems and Autonomous Driving: Right of Way and Human Rights

Abstract 摘要： 这里我将报告自动驾驶与交通系统中有关人类与机器权利的三方面经验证据。（1）“眼里有车，心中无人”：当前交通系统路的权算法侵害了人类出行的基本权利。自动驾驶以及当前的交通系统中，“驾驶员—乘员—行人”利益相关方格局里，最为弱势的行人的路权遭到了系统性忽视；在“智能指挥系统”如“城市大脑”的加持下，交通系统中人类的出行自由正在遭遇挑战，人类“从容行动”的尊严已经让位于“通行效率”的考量；甚至行人必须违法才能实现基本的空间物理转移，这损害了人类出行的基本权利。（见经验证据）（2）“大路更忙，小路更闲”：当前算法损害交通系统有效性。当前交通系统“地图导航”的路线选择与自动驾驶算法的道路选择偏好已经对城市交通产生了重要影响，算法为减少驾驶者工作强度而对“大、长、直”道路的偏好导致城市主干道路车流负载过大进一步加速了拥堵，降低了城市路网的通行效率，强迫“更宽、更长”主干道路的出现并在“更宽、更长”的道路出现后将其“塞满”，形成自反的陷阱无法自拔。（3）“算法主持、机器主导”：人类驾驶员的“地方性知识”乃至“私人知识”与自动驾驶系统算法存在冲突。交通系统中普遍存在的“非常规道路”、特定区域交通拥堵的时空特征经常为所在地的出租车“老司机”所熟悉，“算法主持、机器主导”的局面中人类智慧不得不让位于“机器决策”，基于乘客的“安全考量”与人类或机器驾驶员的“自保顾虑”所形成的“上位共识”，载客车辆将执行一个“算法推荐的愚蠢路线”，交通系统和人类将付出经济、效率和不环保的代价。



JIAN Shengyu 简圣宇
Art and Design Academy, Yangzhou University 扬州大学美术与设计学院

Bio 简介： 简圣宇，扬州大学美术与设计学院教授，博士，硕士生导师。研究方向为意象理论和人工智能美学。曾在《南京社会科学》《东岳论丛》《东南学术》等刊物上发表研究成果，《后人类纪语境中的身体美学问题》等文曾获得过人大复印资料全文转载，出版了包括《审美之思：当代语境下的艺术理论与批评》、《当代艺术的美学阐释》在内的4本专著，参编高等教育出版社教程《艺术学纲要》。赴美国西俄勒冈州大学访问期间，系统观摩了北美高校的美术教育，回国后将其中的艺术批评部分引入学校教学，反响良好。

Website 主页：
https://msxy.yzu.edu.cn/info/1055/1508.html

Title 标题: “元宇宙”问题：技术化娱乐的人文之思
The Metaverse: Humanistic Thoughts on Technologized Entertainment

Abstract 摘要: “元宇宙”概念展现了人类未来发展的一种必然趋势，也是人类文明发展到一定阶段的必然产物。从早期电子游戏到网络游戏，再到“元宇宙”，人类创造出了愈加精致、全面的技术化娱乐方式。技术化娱乐提供的虚拟现实，造成了部分人群的成瘾性依赖和逃避现实等问题。“元宇宙”作为下一代虚拟现实的代表同时包含着风险与机遇。我们有必要从人文关怀的视域出发，围绕技术化娱乐、“元宇宙”和虚拟现实技术应用这一系列具有内在相关性的问题进行学理思考和深入阐析。

JIANG Cong 姜聪 (First Author 第一作者)
Law School, Peking University 北京大学法学院

Bio 简介：北京大学法学院硕士究生，北京大学法律人工智能实验室实习研究员，主要研究领域：法律人工智能应用、法律人工智能的可解释性等
Website 主页： N/A

LIU Lu 刘露 (Coauthor 合作作者）
Law School, Peking University 北京大学法学院

Bio 简介：博士，高级工程师，北京大学法律人工智能实验室研究员，深耕法律人工智能交叉领域，设计并参与主讲《法律人工智能导论》，《智能公共法律服务》，《人工智能、伦理与治理》等课程。
Website 主页： N/A

Title 标题: Justifying the Interpretability of Judicial AI: Legal Sense as the Starting Point
为司法人工智能的可解释性辩护：以法感为出发点

Abstract 摘要: 缺乏可解释性，是法学界对司法人工智能应用于司法实践领域的核心批评之 一，本文试图为司法人工智能的可解释性进行辩护。首先引入司法实践中很常见的现象—— 法感，对法感和可解释性的关系进行分析，指出法感无需解释。然后剖析司法人工智能的训 练过程，揭示出缺乏可解释性并非人工智能的本质属性导致，而是因为任务目标和训练方式 的设置问题，导致目前的司法人工智能是在训练法感。并以此出发，提出司法人工智能应当 从任务目标、训练方式、训练数据等方面进行改良，以构建可解释的司法人工智能模型。

Lena KÄSTNER 莉娜·卡斯特纳
Department of Philosophy, University of Bayreuth, Germany 德国拜罗伊特大学哲学系

Bio 简介：Lena Kästner is professor of philosophy, computer science and artificial intelligence at University of Bayreuth, Germany. Previously, she has been an assistant professor in philosophy of science at Tilburg University in the Netherlands and a junior professor in philosophy of mind and cognitive systems at Saarland University, Germany. Her background is in cognitive science and cognitive neuroscience and she holds a PhD in philosophy. Her primary research areas are philosophy of mind and philosophy of science, especially philosophy of cognitive science. She works on scientific explanations, causation, scientific discovery and experimentation, psychopathology, and explainable AI (XAI). Currently, Prof. Kästner is head-PI of the project "Explainable Intelligent Systems" (www.eis.science).
Website 主页：https://lenakaestner.com

Title 标题：On Modeling Psychopathology: Eight Challenges and How to Meet Them
精神病理学建模：八个挑战及其应对

Abstract 摘要：Understanding mental illnesses requires looking at a variety of different factors contributing to the development, persistence, and treatment of mental illness. That is, scientists must take into account the role of, e.g., behavioral, psychological, neurophysiological, genetic, pharmacological and environmental influences on psychopathology. To integrate real-world data regarding such varied factors, complex computational models have been raising high hopes. In this talk, I shall examine the vices and virtues of recent multiplex approaches for grasping psychopathology.

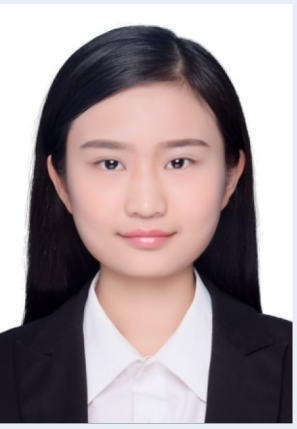
KONG Yuqing 孔雨晴
Center of Frontier Computing Science, Peking University 北京大学前沿计算研究中心

Bio 简介 : Yuqing Kong is currently an assistant professor at The Center of Frontier Computing Science (CFCS), Peking University. She obtained her Ph.D. degree from the Computer Science and Engineering Department at University of Michigan in 2018 and her bachelor degree in mathematics from University of Science and Technology of China in 2013. Her research interests lie in the intersection of theoretical computer science and the areas of economics: information elicitation, prediction markets, mechanism design, and the future applications of these areas to crowdsourcing and machine learning. Her papers were published in several conferences include WINE, ITCS, EC, SODA, AAAI, NeurIPS, ICLR, ECCV, IJCAI, WWW
Website 主页： https://cfcs.pku.edu.cn/yuqkong/

Title 标题: Eliciting Thinking Hierarchy without a Prior
无先验学习思考等级

Abstract 摘要: A key challenge in crowdsourcing is that majority may make systematic mistakes. Prior work focuses on eliciting the best answer without a prior even when the majority is wrong.

Here without any prior, we want to elicit the full hierarchy where the higher-ranking answers, which may not be supported by the majority, are from more sophisticated people. We propose a new model, called the thinking web, that describes the hierarchy among people's thinking types through a weighted directed acyclic graph. To learn the thinking web without any prior, we propose a novel, powerful and practical elicitation paradigm, the Answer-Guess paradigm and it works as follows. First, we ask a single open response question and ask for both of each respondent's answer and guess(es) for other people's answers. Second, we construct an Answer-Guess matrix that records the number of people who report a specific Answer-Guess pair. Third, by ranking the answers to maximize the sum of the upper triangular area of the matrix, we obtain and visualize the hierarchy of the answers without any prior. We also conduct four empirical studies to demonstrate the superiority of our approach compared to the plurality vote and also validate our thinking web model: more sophisticated people can reason about less sophisticated people’s mind and the hierarchy can be approximately described by a directed acyclic graph.



LI Jiaxin 李佳欣
Department of Philosophy and Religious Studies, Peking University 北京大学哲学与宗教学系

Bio 简介: 李佳欣，北京大学哲学系伦理学博士生，主要研究方向为西方近代早期政治哲学、政治思想史和伦理学。
Website 主页： N/A

Title 标题: 人工智能抑或认知心理学：如何理解西蒙的科学发现观
AI or cognitive psychology: how to understand the theory of scientific development of Simon

Abstract 摘要:在研究科学发现的思维实践中，西蒙创造性地提出了“机器发现”的理论，并提出使用人工智能来模拟人类创造性思维的提出与运作过程的方法。由于涵摄科学哲学、人工智能和认知心理学等多领域，这种“启发式搜索解题”发现观亦引起了对西蒙的科学发现观定位的广泛争议：有些人认为西蒙是AI主义的研究路径，有些则认为西蒙是认知心理学的终极目的。结合西蒙的著述文章，本文主要通过在对上述两种观点进行分析的基础上，尝试厘清西蒙科学发现观的真正面貌，并结合当今社会人工智能发展的具体状况，进一步梳理认知心理学与人工智能的渊源关系与发展异同，并结合德雷福斯的批判，谈谈对当今时代人工智能的发展状况判以及二者未来研究进路走向的科学哲学思考。

LI Xinyi 李欣怡
School of Philosophy, Fudan University 复旦大学哲学学院

Bio 简介：李欣怡，就读于复旦大学哲学学院，目前是科学技术哲学专业硕士生，主要研究生物学哲学、因果科学哲学。
Website 主页： N/A

Title 标题: 历时地衡量公平：基于因果结构理论的分析
Measuring Fairness Diachronically: a Causal Structural Approach

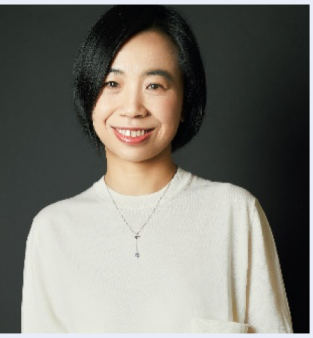
Abstract 摘要：因果科学的发展为实现算法公平提供了一个新的解决方案。不同于统计学方法只关 注相关性，因果分析关注变量之间的内在联系。然而，因果模型下的算法公平存在两个尚 未被解决的问题:缺乏对公平和效率之间的衡量;无法为合理的敏感条件分类提供判断的 依据。在现有分析的基础上，本文尝试着构建一个基于因果分析的历时模型，重点关注世 界与模型之间的交互关系，为解决上述两个问题提供了尝试性的解答。

LIU Xiaoli 刘晓力
Department of Philosophy, Renmin University of China 中国人民大学哲学系

Bio 简介: 刘晓力，北京大学哲学博士，中国人民大学哲学院教授，北京大学博古睿学者。曾任中国逻辑学会副会长，现任中国自然辩证法研究会科学哲学专业委员会主任。从事数学哲学、科学哲学、心灵哲学与认知科学的交叉研究。主要著作包括:《理性的生命:哥德尔思想研究》、《认知科学前沿的哲学问题》、《认 知科学对当代哲学的挑战》，主编《心灵与机器交响曲》、《心灵与认知》丛书等。2007-2020 年，主持了总计 100 期中国人民大学“科学-社会-人文论坛”。2018 年，作为首席专家，创建中国人民大学哲学与认知科学交叉平台和“哲学与认知科学明德讲坛”。
Website 主页：
http://phi.ruc.edu.cn/jytd/jszl/apyjs/L/lxl_8faf7c197c1b493cb40732b3eef15e48/index.htm

Title 标题： 意识图灵机能否搭建人类与AI互信的桥梁？
Can Conscious Turing Machines Build a Bridge of Mutual Trust Between Humans and AI?

Abstract 摘要： 近年来，图灵奖获得者布鲁姆M. Blum致力建构“意识图灵机CTM”模型，并且声称CTM可以模拟意识和意识经验。同时，布鲁姆认为，建构CTM的努力表明，科学不仅可以说明无机生命和有机生命的起源，也可以解释感觉质现象如何产生；解释何以情感价值不仅能够产生于血肉之躯，也可以在硅基材料或其他金属材质的机器上实现。本文基于可信认的脑机制和预测心智假设，讨论仅仅基于全局工作空间假说的CTM模型，距离实现可信赖的AI、搭建人-机互信桥梁的目标还有多远。

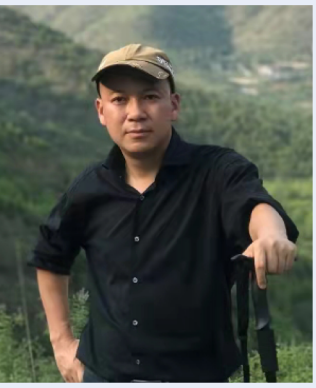


LUO Huan 罗欢
School of Psychological and Cognitive Sciences and a PI of IDG/McGovern Institute for Brain Research, Peking University 北京大学心理与认知科学学院

Bio 简介：Dr. Huan Luo is a tenured associate professor at the School of Psychological and Cognitive Sciences and a PI of IDG/McGovern Institute for Brain Research, Peking University. Her research primarily focuses on the brain mechanisms of perception, attention, and working memory in humans, particularly from a dynamic perspective. She is Chang Jiang Young Scholar and supported by NSFC Key Program and Excellence Young Scientists Funds. Dr. Luo received her Ph.D. from the University of Maryland College Park, first worked at Chinese Academy of Sciences and joined Peking University since 2015. Dr. Luo currently serves as the associate editor for *Progress in Neurobiology*, and editorial members of *PLoS Biology* and *eLife*. As one of six labs in the world, her lab recently participated in a high-impact international collaborative project COGITATE supported by Templeton World Charity Foundation.
Website 主页：
http://www.psy.pku.edu.cn/szdw/qzjy/jsyjy/lh/index.htmhttp://mgv.pku.edu.cn/english/people/lbd/sopacs/360652.htm

Title 标题： Relational Structure Knowledge in the Human Brain: Representation, Memory, and Learning
人脑中的关系知识结构：表征、记忆和学习

Abstract 摘要： Despite ostensibly having limitless access to knowledge during this era of information explosion, we all acknowledge that big data is not knowledge. Indeed, humans are endowed with tremendous abilities to reason and infer the hidden relational structure behind fragmented outside events, an essential index for intelligence. In this talk, I will present my lab's recent works aiming to understand how relational structure knowledge, from simple sequence to hierarchy and network, is represented, maintained, and learned in the human brain. Our findings support dissociated structure-content neural representations and an essential role of relational structure in mediating memory and learning.



MEI Jianhua 梅剑华
School of Philosophy and Sociology, Shanxi University, China 山西大学哲学社会学学院

Bio 简介：梅剑华，山西大学哲学社会学学院教授、博士生导师；教育部“青年长江学者”；《认知科学》双语期刊主编；曾担任首都师范大学哲学系系主任。主持国家社科重点、重大子课题等多项项目，获山西省教学成果特等奖。在《中国社会科学》《哲学研究》《自然辨证法通讯》等刊物发表文章40余篇，主编、翻译学术著作6部。

Website 主页：
http://zxxy.sxu.edu.cn/szdw/js/eb46298bd5f447a487cd35c945bb671e.htm

Title标题：儿童哲学与人工智能哲学间的三重对话
Three Levels of Dialogue between Philosophy of Children and Philosophy of AI

Abstract 摘要 ：儿童哲学和人工智能哲学都是当今时代哲学研究的新形态，人工智能要能模仿人类智能，就不能只考虑一般人类的智能，尤其需要考虑儿童智能。对儿童哲学、儿童智能的探究可能成为人工智能研究的一个新思路。本文从这个思路出发，探究儿童哲学与人工智能对话的三个层次：智能体的认知特征、智能体的学习方式、智能体的自主标准。儿童和机器都是实际存在的智能体，从对话的角度，儿童哲学可以受益于人工智能研究，反之亦然。本文将展示这一对话的丰富内涵，以期对儿童哲学和人工智能的基础讨论有所助益。

Dmytro MYKHAILOV 德米特罗•米哈伊洛夫
School of Humanities, Southeast University, China
东南大学人文学院

Bio 简介: I am a Ukrainian researcher in the philosophy of technology currently working in China. In 2017 I got a Ph.D. degree in philosophical anthropology at Taras Shevchenko National University of Kyiv (Ukraine). Afterward, back in 2019, I moved to Nanjing (China), where I got a postdoctoral position at Nanjing Normal University. My research there was primarily related to the ethics of AI in medicine. Currently, I am working as a postdoctoral fellow at the School of Humanities, Southeast University (Nanjing, China). Now I am a member of the research group that studies the ethical implications of AI in Smart-education. I have a membership in several international scientific communities: SPT (Society for the Philosophy and Technology), MEA (the Media Ecology Association); SPSP (Society for Philosophy of Science in Practice); SPM (Society for Phenomenology and Media).
Website 主页：
https://www.researchgate.net/profile/Dmytro-Mykhailov

Title 标题： New AI applications within Smart-Education domain: AI adaptive educational system and its moral drawbacks
智慧教育领域新人工智能应用：人工智能教育适应系统及其缺陷

Abstract 摘要: Our connection with technology begins in early childhood and increases throughout our whole life. For example, in the UK 33% of 3–4-year-olds and 92% of 12–15-year-olds access the Internet, and across Europe the number of children having a smartphone increases by 58% for each year they become older (Ofcom 2013). Relatively similar numbers may be found in other 'high-tech' countries. Digital technologies actively participate in our breeding, education, and training. Technology becomes ubiquitous spreading through various societal domains, transforming our everyday practices and mundane routines.

Today's domain of education isn't an exclusion (Sheehy & Holliman, 2017). Old models, in which the teacher explains the subject in the classroom and the students complete the exercises at home, are slowly replaced by new learning approaches such as distant learning, mobile learning (m-learning), personalized learning, game-based learning, etc (Demir, 2021).

All these novel learning approaches became possible by new technological solutions. AI is one of them. AI solutions are widely implemented in different areas of today's education (Chaudhri et al., 2013). Nowadays AI is successfully used to facilitate skills that require repetition (for example, language learning). The use of recommender systems might help students to identify content and learn more about a particular topic. AI tutoring systems currently support richer experiences for learners and provide researchers with new opportunities to analyze big data sets from vast databases (Eynon & Young, 2020). Personalized learning enhances student and group experience, supports individual and collective reflexivity, deepens the analysis inside and outside of the classroom. Said differently, AI becomes a tool to augment human cognition.

Novel AI applications that come into fashion recently are various AI adaptive educational systems. The main purpose of these systems is to provide students with personalized suggestions and to customize content and future learning paths. This leads to minimization of conceptual disorientation and decreasing cognitive and informational overload. By doing this AI adaptive educational systems are maximizing learning efficiency and minimizing undesirable drawbacks that might appear within the educational process (Bajaj & Sharma, 2018).

However, there are still a lot of ethical issues that might 'pop up' during the implementation of AI into a school environment.

Taking into consideration that AI adaptive educational system is a relatively novel tool not much has been written on the ethical consequences that it might bring to the educational domain. The use of AI, thus, appears to bring benefits but also to create potential moral drawbacks (Woolf et al., 2013). In what follows, I will briefly specify three ethical 'risk zones' where the application of AI might lead not only to moral uncertainty but also to several undesirable consequences.

The first 'risk zone' is related to 'student-machine' interaction. It seems that we still don't know much about the impact of AI on students' social skills (for instance, empathy, reading of emotions, cooperation, work in groups) and various learning abilities (like memorizing, structuring information, task orientation). If students will spend more time with computers and less time in face-to-face commutation will it influence their development? And if so, in which way? What is more important, at what age should students begin to interact with AI on the daily basis? Is AI beneficial to student development or is it a harmful digital supervisor (Haughton, Aiken and Cheevers 2015)?

The second zone is about digital addiction. It looks like contemporary digital technologies are highly addictive. Can we say that the educational process dependent on AI may have addictive consequences? Maybe we even can suggest that students will depend on technology much more than it was before? What kind of addictivity it might be? How strong will students' social lives be damaged through 'digital addiction' (Kucirkova & Radesky, 2019)? Finally, what should we do if the usage and hence, an addiction from AI in education is unavoidable?

The third ethical zone is related to design practices and algorithmic transparency (Smith & Iversen, 2018).

As has been already mentioned, AI adaptive educational system is designed to provide students with personalized suggestions. Every personalized suggestion is based on selectivity. Some information is prioritized while others fall out of the educational scope. As AI selectivity is primarily based on machine learning algorithms (like Neural Networks, Genetic Algorithms, or Fuzzy Logic) the process of information selection will depend on how the machine learns (Pasquinelli, 2019). However, we still don't know which is the most suitable machine learning technique and the most appropriate artificial intelligence method to apply for a particular learning environment (Bajaj & Sharma, 2018). It is still unclear which machine learning technique will better suit students with different learning types. For example, maybe neural networks will better suit active learners, while genetic algorithms will be more suitable for reflective learners.

Three ethical 'risk-zones' mentioned above may serve as a preliminary roadmap for understanding potential drawbacks that AI might cause in the educational domain. It goes without saying that the proposed framework is neither exhaustive nor complete. However, it might be helpful within the early stage of implementation of AI in the educational domain.

Giada PISTILLI 嘉坦·皮斯蒂利
Faculty of Philosophy, Sorbonne University, France
索邦大学哲学系

Bio 简介: Giada Pistilli is a Ph.D. candidate in philosophy, specialising in conversational AI ethics (conversational agents, Natural Language Processing, Large Language Models). Her research focuses on elaborating ethical charters, value pluralism and value conflict embedded in AI models, and empirical research based on field experiments of conversational agents for French public administrations.
Website 主页： https://www.giadapistilli.com

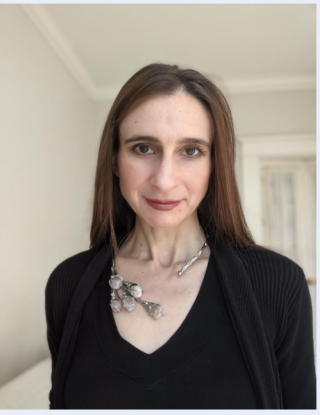
Title 标题： Ethical Frameworks as a Moral Exercise in the Field of Natural Language Processing
自然语言处理中作为道德操练的伦理架构

Abstract 摘要： The scientific domain of artificial intelligence has long become an opportunity for researchers from different disciplines to collaborate on common goals. The social sciences, particularly research in ethics, show their capacity to guide this interdisciplinary research. In particular, the tools of ethical reflection can help us distinguish opinions from moral judgments, the importance of asking the right research questions, and the complexity of human judgment when defining “good” and “bad”, “right” and “wrong”.
These reflection tools become especially important when framing an AI application in an ethical framework. In this context, I wish to illustrate one particular case study in the field of Natural Language Processing: the experience of drafting an ethical charter for an open science project focused on developing a multilingual language model and its dataset for research purposes (BigScience1).

The goal was to find common core values that could serve as goals for the project, operating as a pivot for articulating the other documents that frame the language model's functioning and application. I argue that the importance of such experiences lies mainly in the moral exercise of gathering, reflecting, debating, and ultimately deliberating together. The richness of debates among engineers, developers, lawyers, sociologists, philosophers, and historians about their shared values is paramount in research in descriptive ethics.

The challenges present in this collaborative and interdisciplinary research lie in several points. For example, the project's scientific community includes more than 900 researchers from more than 50 countries worldwide who speak more than 25 languages and dialects. However, the difficulties of this diversity have been turned into an asset and core value: the ethical charter states, "we assume the basis of value pluralism within our community, and we cherish it". Accordingly, the ethical approach used is that of moral value pluralism, which allows the different value systems present among the collaborators of the project to be considered of equal importance with their tensions and similarities. In line with this approach, I have chosen to adopt the practice typical of the Confucian ethical tradition that emphasizes, among other things, the notion of harmony (和). Being by its very nature relational, the notion of harmony does not seek to avoid conflict but aims to guide its resolution, focusing on the coexistence of different parties (Li, 2006). The notion of harmony thus fills one of the primary missions of ethics: to show what humans ought to do when making choices.

Bottom-up collaboration, the support of ethics as a tool for interdisciplinary collaboration, and a value pluralism approach based on the notion of harmony make up the beacons of this research focused on ethically framing Natural Language Processing systems.



Cynthia RUDIN 辛西娅·鲁丁
Computer Science Department & Electrical and Computer Engineering Department, Duke University, USA 美国杜克大学计算机科学系 & 电气与计算机工程系

Bio 简介：Cynthia Rudin is a professor of computer science, electrical and computer engineering, statistical science, mathematics, and biostatistics & bioinformatics at Duke University. She directs the Interpretable Machine Learning Lab, whose goal is to design predictive models with reasoning processes that are understandable to humans. Her lab applies machine learning in many areas, such as healthcare, criminal justice, and energy reliability. She holds an undergraduate degree from the University at Buffalo, and a PhD from Princeton University. She is the recipient of the 2022 Squirrel AI Award for Artificial Intelligence for the Benefit of Humanity from the Association for the Advancement of Artificial Intelligence (the "Nobel Prize of AI"). She is a fellow of the American Statistical Association, the Institute of Mathematical Statistics, and the Association for the Advancement of Artificial Intelligence. Her work has been featured in many news outlets including the NY Times, Washington Post, Wall Street Journal, and Boston Globe.
Website 主页：
https://users.cs.duke.edu/~cynthia/home.html

Title 标题：Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead
对于高风险决策选择可解释机器学习模型，而不是黑盒模型

Abstract 摘要: With widespread use of machine learning, there have been serious societal consequences from using black box models for high-stakes decisions. Explanations for black box models are not reliable, and can be misleading. If we use interpretable machine learning models, they come with their own explanations, which are faithful to what the model actually computes. I will discuss examples related to seizure prediction in ICU patients and digital mammography.

Piero SCARUFFI 皮耶罗·斯卡鲁菲
Department of Chemical Engineering, Stanford University, USA 美国斯坦福大学化学工程系

Bio 简介: Piero Scaruffi is the founder of Silicon Valley Artificial Intelligence Research Institute. Born in Italy, Scaruffi holds a degree in Mathematics from University of Torino. He came to the Silicon Valley in 1983, and has witnessed the development of this valley. His career crossed the production, research and academic fields. Visiting scholar at U.C. Berkeley, Stanford University, and University of San Francisco. In 2006 the New York Times ran an article about Piero's own website (www.scaruffi.com) titled "The Greatest Website of all Times". Piero established the most influential interdisciplinary event L.A.S.T. Festival (combines Life, Art, Science and Technology) in Silicon Valley, which has attracted the best interdisciplinary artists, scientists and technologists of Silicon Valley. The LAST Festival, together with TED Conference (Technology, Entertainment and Design) and Burning Man (Anti-traditional Carnival), is called three major activities of cross-border art and innovative inspiration sources in Silicon Valley.

Piero is the author of *A History of Silicon Valley*, *Intelligence Is Not Artificial* and *Human 2.0 - the Future of Technology*. *A History of Silicon Valley* is the first history of Silicon Valley from 1900 to the 2010s. It is a comprehensive study of the greatest creation of wealth in the history of the world, from the establishment of Stanford University to the age of social media. The later two introduced the current state and future of Artificial intelligence. All the three books have been published and translated to several languages around the world.

皮耶罗·斯卡鲁菲是硅谷人工智能研究院（Silicon Valley Artificial Intelligence Research Institute）创始人。毕业于意大利都灵大学数学系，1983年来到硅谷任工程师，见证了硅谷30年来的兴盛过程，职业生涯横跨硅谷产、学、研三界。加州伯克利大学、斯坦福大学、旧金山大学访问学者。世界新媒体领域先驱，其个人新闻网站早在 2006 年就被《纽约时报》评价为“史上最伟大的网站”。在硅谷一手创办最有影响力的跨界L.A.S.T节（融汇生活、艺术、科学、技术），吸引整个硅谷最优秀的生活家、跨界艺术家、科学家和技术达人的参与，LAST节与TED大会（技术、娱乐、设计）、Burning Man（反传统狂欢节）并称为硅谷跨界艺术和创新灵感来源的三大活动。

Website 主页： https://en.softonic.com/articles/the-13-oldest-web-pages-that-are-still-working-today

Title 标题： Has AI Reached Human Parity?
人工智能已经达到与人类同等水平了吗？

Abstract 摘要： In the last 4 years, human parity by various AI systems has been announced. Since we normally assess a person's intelligence through language, most I.Q tests for A.I. systems are based on natural-language tasks, such as answering questions, translating from one language to another, and generating captions for images or images from text descriptions. Meanwhile, robots and self-driving cars are widely described as replacing humans in many ordinary activities. These announcements reveal that AI is increasingly based on "foundation models", trained on datasets and then tested on "benchmarks". But there is still a big gap between the "deep learning" of AI system and the "deep thinking" of human beings.

Susan SCHNEIDER 苏珊·施奈德
Department of Philosophy, Florida Atlantic University 佛罗里达大西洋大学哲学系

Bio 簡介：Schneider is the founding director of the new Center for the Future Mind at Florida Atlantic University (FAU). She writes about the nature of the self and mind, especially from the vantage point of issues in philosophy, AI, cognitive science and astrobiology. In her recent book, *Artificial You: AI and the Future of the Mind*, she discusses the philosophical implications of AI, and, in particular, the enterprise of "mind design." As the NASA chair, Schneider has recently completed a two year project with NASA on the future of intelligence. She now works with Congress on AI policy. She is delighted to be co-director of the MPCR Lab at FAU's new Gruber Sandbox, a large facility which builds AI systems drawing from neuroscience research and philosophical developments. She also appears frequently on television shows on stations such as *PBS* and *The History Channel* (see below for clips). She writes opinion pieces for the *New York Times*, *Scientific American* and *The Financial Times*. Her work has been widely discussed in the media (see "media" above). She is currently working on a new book on the shape of intelligent systems (for W.W. Norton).
Website 主页：
https://schneiderwebsite.com/index.html

Title 标题： AI-based Brain Enhancements, Superintelligence and the Future of the Mind
基于人工智能的大脑强化、超智能与心灵的未来
Abstract 摘要：N/A

Bongrae SEOK 石奉来
Alvernia University in Reading, Pennsylvania, USA 美国艾尔弗尼亚大学哲学系

Bio 简介: Bongrae Seok is Associate Professor of Philosophy at Alvernia University in Reading, Pennsylvania, USA. His primary research interests lie in cognitive and comparative philosophy of mind and moral psychology, philosophy of neuroscience and AI, moral neuroscience, neuroethics and neuroasesthetics. In his recent books, Naturalization, Human Flourishing, and Asian Philosophy: Owen Flanagan and Beyond (Routledge 2020), Moral Psychology of Confucian Shame: Shame of Shamelessness (Rowman and Littlefield 2016), and Embodied Moral Psychology and Confucian Philosophy (Lexington 2013), he develops an interdisciplinary approach to moral psychology from the viewpoint of embodied moral emotions (empathy, shame, and flourishing) and Asian philosophy (Korean and Chinese Confucianism). In his articles "Autistic Moral Agency and Integrative Neuroethics," "Diversity and Unity of Modularity", "The Emotional Mind and The Moral Mind", "Mencius's Vertical Faculties and Moral Nativism", and "Neuroscience, Moral Sentimentalism, and Confucian Philosophy", he integrates cognitive modularity and embodied cognition in a broad scope of psychology, neuroscience, and philosophy. He also published articles on phenomenology and perception of music such as "Cognitive Science and the Neuroaesthetics of Musical Chills" and embodied perception of space such as "Traces of the Body in Space: Embodied Metaphor of Menus and Lists" from the interdisciplinary perspective of space cognition and neuroaesthetics. Regarding AI and philosophy, he is preparing articles that integrate AI, Art and Philosophy such as "The Uncharted World of AI Art: Music and AI [artificial intelligence]" and "Marginalization and De-Marginalization: Future of Robotic AI and Buddhism." His current work focuses on interdisciplinary topics (including embodied cognition and emotion, affective moral intuition, empathic nociception, moral psychology of autism and psychopathy, and aesthetic and embodied experience of music) that bring philosophy and neuroscience to the forefront of cognitive science.

He is a recipient of the Neag Professorship (2014−2016), the Lindback Foundation Award for Teaching Excellence (2020), and the Charles Fu Foundation ISCP Essay Award (Winner) (2019). He is program chair of the APA (American Philosophical Association) affiliated group of NAKPA (North American Korean Philosophical Association). He is a former president of ACPA (Association of Chinese Philosophers in America) and a current member of the editorial board of KSCS (the Korean Society for Cognitive Science) and KAFP (the Korean Association of Feminist Philosophers).
Website 主页： https://bongraeseok.weebly.com/

Title 标题： AI, Robotics, and Buddhism
人工智能，机器人学与佛教

Abstract 摘要： In this paper, I will explain a stimulating relation between AI and Buddhism. Specifically, I will discuss how robotic AI is introduced and used in East Asian Buddhism (Mahayana Buddhism practiced in China, South Korea, and Japan). If the sacredness of life and the human soul is one of the core foundations of religion, the ultimate goal of robotic AI as a mechanical system of cognition cannot be accepted or recognized in religion. However, Buddhism, unlike other religious traditions such as Christianity and Islam, accommodates and utilizes AI in its religious practice. There are four different ways Buddhism uses Robotic AI: (1) Robots are used as a form of social media without any acknowledgment or acceptance of AI in religious practice and its technological and philosophical/religious significance. (2) Robots are used as an agent to provide religious service. (3) Robots are used as a robotic interface or representation of Buddha with the full acknowledgement and recognition. (4) Robots are used as a kind of mind that can be enlightened. In this paper, I will discuss and explain how Mahayana Buddhism of East Asia actively accommodates robotic AI. It seems that Buddhism has a rather broad notions of the mind and the sentient being (sattva, 衆生) and it understands and accepts the artificial mind and machine intelligence within the broad context of Buddhist enlightenment where the mind and cognition are not tied to a particular substance or a functional specification. However, under current AI technology with machine learning (specifically the tendencies of confirmation bias in deep learning), there is also a possible tension between the AI mind and the Buddha mind.

SHEN Yuan 沈苑 (First Author 第一作者)
Graduate School of Education, Peking University 北京大学教育学院

Bio 简介：北京大学教育学院博士研究生，主要研究领域：人工智能教育应用伦理、教学设计的理论与实践等
Website 主页：N/A

WANG Qiong 汪琼 (Coauthor 合作作者）
Graduate School of Education, Peking University 北京大学教育学院

Bio 简介：北京大学教育学院教授，博士生导师，北京大学数字化学习研究中心主任，主要研究领域：人工智能教育，信息技术环境下的教与学规律等
Website 主页：N/A

Title 标题: 人工智能教育应用的价值敏感设计
Value-sensitive Design of Artificial Intelligence Education Application

Abstract 摘要:价值敏感设计（Value Sensitive Design）为近年来应用伦理学显著的设计转向提供了重要的理论支持和方法论基础，能够有效指导人工智能教育应用（AI in Education）的可持续发展。该文概述了价值敏感设计的核心主张，结合人工智能教育应用的特点提出了适用的价值敏感设计研究路线。并以智能课堂教学评价系统为例，提出了系统设计应遵循的价值框架和可参考的设计要求。通过整合开展概念、经验和技术研究，发现智能课堂教学评价系统的设计应以优质为目标价值，以科学、无害、公平、自主、透明为规范价值，以可用为特定价值，建议设计者在设计阶段通过在科学证据支持下创建算法模型、避免使用有偏见的数据集、遵循条件性公平、鼓励师生参与设计过程、开发配套说明文本等做法来彰显价值考量并在系统全生命周期中持续进行检验，以推动符合伦理的人工智能教育应用发展。

SONG Jianli 宋建丽
School of Marxism, Tianjin University, China 天津大学马克思主义学院

Bio 简介：宋建丽，女，1972年7月24日生，内蒙古乌兰察布市人，现任职天津大学马克思主义学院英才教授，厦门大学马克思主义学院兼职教授、博士生导师，长期从事马克思主义理论的教学与研究工作，研究重点涉及马克思主义政治哲学、国外马克思主义、伦理学、性别哲学等多个相关方向。福建省百千万人才工程省级人选，已在《马克思主义研究》《马克思主义与现实》《国外社会科学》《哲学动态》《世界哲学》等重要学术期刊发表论文60余篇，有10篇（部）被全文转载或论点摘编。已出版学术专著2部，其中一部获福建省社科第九届社会科学优秀成果一等奖。翻译（合译）著作1部，主编、参编著作各1部。主持各类课题15项，其中国家社科基金一般项目2项、青年项目1项、重大项目子项目2项，省部级课题4项。重点围绕“马克思主义与当代社会思潮”“当代资本主义批判”“ 国外马克思主义与社会主义”等主题展开系列研究，目前侧重于马克思主义正义论、全球治理视域下国外马克思主义的当代资本主义批判、唯物史观与构建人类命运共同体等问题的研究。

Website 主页：
http://marxism.tju.edu.cn/info/1085/1912.htm

Title 标题：世纪疫情下人工智能的加速与不平等挑战
The Acceleration and Inequality of Artificial Intelligence under the Century of Epidemic

Abstract 摘要： 疫情冲击下，当代世界经济和政治的脆弱性凸显，价值观冲突加剧，各国为应对疫情而采取的措施正在极大地改变世界未来格局。疫情冲击下人工智能的加速运用既可能加剧全球不平等，也可能会成为扭转全球不平等加剧的催化剂。也就是说，如果不能摆脱资本利益的主导和羁绊，人工智能的加速发展和运用无疑会加深已有数字鸿沟，进而造成碎片化和离心化的社会撕裂，最终危及社会发展，破坏人工智能的良性发展进程；反之，如果在不断创新、完善人工智能运用的同时，始终自觉考虑如何应对可能的风险和不平等挑战，自觉摒弃排斥、控制和竞争的价值观，代之以包容、共享和互惠的价值观，从而减少可能扩大的脆弱性的不平等，则可能导向弥合数字鸿沟、加强全球向心力和凝聚力的美好前景。简言之，人工智能的加速使用改变着以技术为中介的人与世界的关系，持续应对疫情冲击不仅需要在人类命运共同体理念下致力于科技创新与合作，而且也需要推动疫情下的全球治理朝着更具社会凝聚力、更加公平正义的方向发展。加速对弱势群体的边缘化、排斥还是促进社会包容和凝聚，是疫情冲击下人工智能加速使用可能导向的两种不同路径，遵循社会达尔文主义的理念还是命运共同体的理念，以资本为中心还是以人民为中心，是应对疫情冲击的两种不同选择。选择前者，只能将人工智能的发展导向一条“不归路”；选择后者，才能有效可持续地应对疫情冲击，保障人工智能的安全发展，实现人工智能造福人类的根本目的。

SUI Tingting 隋婷婷
Department of Philosophy and Religious Studies, Peking University 北京大学哲学与宗教学系

Bio 简介：隋婷婷，东南大学哲学博士，现为北京大学哲学系博雅博士后/讲师。2017-2018年，获CSC建设高水平公派研究生项目奖学金资助，赴美国西北大学心理学院霍森菲尔德实验室访学，从事EEG脑电事件相关电位与认知心理学研究。目前主要从事人工智能哲学、认知神经科学哲学、实验伦理学以及道德心理学等方面的研究，在《International Journal of Psychophysiology》（SCI/SSCI）、《自然辩证法通讯》、《自然辩证法研究》、《哲学分析》以及《江苏社会科学》等刊物上发表多篇论文。曾参与《中国大百科全书第三版•哲学卷》科学技术哲学（生命科学哲学问题）版块词条编写，撰写“发生心理学”与“心理学哲学问题”词条。

Sui Tingting, a Ph.D. of Southeast University, is currently a post-doctoral fellow/lecturer in the Department of Philosophy of Peking University. From 2017 to 2018, she was funded by the CSC Construction High-level University Postgraduate Program and went to the Rosenfeld Lab of department of psychology, Northwestern University, USA, to study EEG event-related potentials of cognitive psychology. Her contemporary research interests include artificial intelligence philosophy, cognitive neuroscience philosophy, experimental ethics and moral psychology. She has published a number of papers on "International Journal of Psychophysiology"(SCI/SSCI) and Chinese journals including "dialectics of Nature Communications", "Studies of Dialectics of Nature", "Philosophical Analysis" and "Social Sciences of Jiangsu Province". She also participated in the compilation of entries for the section of Philosophy of Science and Technology (Philosophical Issues of Life Sciences) in the third edition of Encyclopedia of China: Philosophy Volume, and wrote entries for "genetic psychology" and "philosophical problems in psychology".

Website 主页： N/A

Title 标题: The Is-Ought Problem within Moral Algorithm Experiment of Autonomous Vehicle
自动驾驶道德算法实验中的“是”与“应当”问题

Abstract 摘要: Autonomous vehicles are accompanied by many moral disputes in algorithm. "Trolley problem" is one of the classic moral dilemmas that need to be solved. Scholars tend to build a moral algorithm that applied moral principles to solve the problem. To find the proper moral principle, many scholars are inclined to take moral algorithm as a moral issue, that is, an “ought” issue. However, according to some previous studies, moral algorithm seems affected by some “is” factors, that is, non-moral factors. An example is the algorithm of "consequentialism". People could accept it morally, but would reject it in practical because of fear of self-sacrifice. Therefore, a question is that whether people should consider moral algorithm as an “is” issue? To answer this question, we set up an online survey for people around China to test it. According to survey, 57.5% people would reject to buy autonomous vehicle which applied algorithm of “egoism” that would protect them at any situation. Comparing with autonomous vehicle that applied “consequentialism” and “Rawlsian”, the percentage of people who rejected “egoism” is the same as people who rejected “consequentialism”. The conclusion is that moral algorithm can’t be considered either as an “ought” issue or as an “is” issue. The algorithm that enjoys the highest support in the survey is the combination of “consequentialism” and “egoism”, which tried to balance both “ought” and “is” problem.

Timothy TAMBASSI 提摩西·坦巴西
DISPAC, University of Salerno, Italy; University of Eastern Piedmont, Italy 意大利萨莱诺大学文化遗产学系

Bio 简介： I am an Assistant Professor in Philosophy at the University of Salerno, and an Adjunct Professor at the University of Eastern Piedmont. My main interests of research include Applied Ontology, Philosophy of Geography, Philosophy of Computer Science, and the Philosophy of E.J. Lowe.
Website 主页： https://timothytambassi.wordpress.com/

Title 标题: On the Mutual Understanding between Human Beings and Software Systems. The Role of Information System Ontologies in Artificial Intelligence
人类与软件系统间的相互理解——人工智能中信息系统存在论的角色

Abstract 摘要: According to Goy and Magro (2015), a foundational aspect of information system ontologies [ISOs] is supporting communication and mutual understanding between human beings, human beings and software systems, and software systems.
[1] does not, however, imply that human beings and software systems understand ISOs' entities in the same way. But if so, what would the difference involve?
This talk aims to account for such a difference. Firstly, we maintain that while human being can have access to entities represented in ISOs, software applications cannot. Secondly, we argue that the difference also involves the (Semantic web) languages by which ISOs are developed. More precisely, some of those languages are based on the open-world assumption, according to which everything that cannot be inferred as false from an ISO must be considered unknown. Conversely, in many programming languages and formal systems of logic used for knowledge representation, some ISOs' approaches adopt the closed-world assumption, according to which, everything that is not known to be true, in the system, must be considered as false.

TAN Ying 谭营
Computational Intelligence Laboratory, Peking University
北京大学计算智能实验室

Bio 简介：谭营，北京大学教授、博士生导师、烟花算法发明人。2005年曾入选中国科学院百人计划，曾任日本九州大学教授、美国哥伦比亚大学高级研究员、中国科技大学教授、香港中文大学研究员、电子工程学院教授及研究所所长等。主要从事智能科学、计算智能与群体智能、机器学习与大数据分析以及相关应用的研究工作。主持国家级科研项目30余项。撰写学术专著《Fireworks Algorithm》(Springer)、《GPU-based Parallel Implementation of Swarm Intelligence Algorithms》(Morgan Kaufmann - Elsevier)、《烟花算法引论》(科学出版社)等十余部，科普专著《人工智能知识讲座》(人民出版社，2018)和《人工智能之路》(清华大学出版社，2019)，主编Springer-Nature的LNCS论文集共42卷。发表学术研究论文320余篇，曾获得国家自然科学奖二等奖、教育部自然科学奖二等奖、中国科技产业化促进会科学技术奖卓越贡献奖等。他担任国际群体与演化智能学会的主席，担任十多个国际期刊的主编、副主编和编委，创立并担任群体智能国际会议(ICSI)主席(已举办11届)，担任金砖国家计算智能大会联合主席，等。他是多个国际、国家科学基金和奖励计划的评审专家。

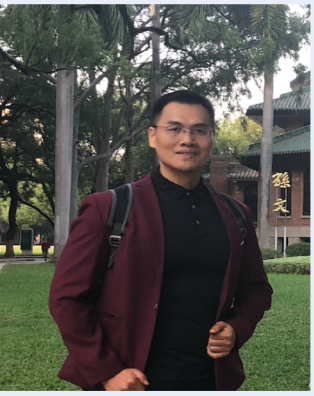
Ying Tan is a full professor of Peking University, director of Computational Intelligence Laboratory at Peking University, and the inventor of Fireworks Algorithm (FWA). He worked as a professor of Faculty of Design, Kyushu University, Japan, in 2018, at Columbia University as senior research fellow in 2017, and at Chinese University of Hong Kong as research fellow, and at University of Science and Technology of China in 2005-2006 as a professor under the 100-talent program of CAS. He is the president of the IASEI, and also serves as the Editor-in-Chief of IASEI Transactions on Swarm Intelligence, and International Journal of Computational Intelligence and Pattern Recognition (IJCIPR), the Associate Editor of IEEE Transactions on Cybernetics (CYB), IEEE Transactions on Neural Networks and Learning System (NNLS), Neural Networks, International Journal of Swarm Intelligence Research (IJSIR), etc. He also served as an Editor of Springer's Lecture Notes on Computer Science (LNCS) for 40+ volumes, and Guest Editors of several referred Journals, including IEEE/ACM Transactions on Computational Biology and Bioinformatics, Information Science, Neurocomputing, Natural Computing, Swarm and Evolutionary Optimization, etc.

He is the founder general chair of the ICSI International Conference series since 2010 and the DMBD conference series since 2016. He won the 2nd-Class Natural Science Award of China in 2009 and 2nd-Class Natural Science Award of Ministry of Education of China in 2019 and many best paper awards. His research interests include computational intelligence, swarm intelligence, deep neural networks, machine learning, data mining, intelligent information processing for information security and financial prediction, etc. He has published 350+ papers in refereed journals and conferences in these areas, and authored/co-authored 12 books, including "Fireworks Algorithm" by Springer in 2015, and "GPU-based Parallel Implementation of Swarm Intelligence Algorithms" by Morgan Kaufmann (Elsevier) in 2016, and received 5 invention patents.
Website 主页： http://www.cil.pku.edu.cn

Title 标题：群体智能及其应用研究进展
Advances in Swarm Intelligence and Its Applications

Abstract 摘要： Inspired from the collective behaviors of many swarm-based creatures in nature or social phenomena, swarm intelligence (SI) has been received attention and studied extensively, gradually becomes a class of efficiently intelligent optimization methods. This talk presents the latest advance in swarm intelligence and its applications. First of all, I will introduce a novel swarm intelligence optimization method, so-called Fireworks Algorithm (FWA) inspired by fireworks' explosion in air, and its many improvements and beyond as well as a variety of successful applications; and then, the research progress of multi-objective search and swarm cooperative learning in swarm robots is presented; After that several typical applications of swarm intelligence will be described; Finally, I will introduce some important researches of swarm intelligence in future.

受生物群体与自然集群等现象的启发，近年来，群体智能获得了广泛的关注和深入的研究，已经发展成一类十分有效的智能计算方法。本报告主要介绍群体智能的最新研究及其应用。首先介绍我们所提出并获得广泛关注和研究的一类新型群智能优化算法-烟花算法（FWA）及其最新研究进展；然后介绍群体机器人多目标搜索与群体协同学习的相关研究进展；接着介绍几个典型的群体智能应用进展；最后，指出群体智能研究的未来发展方向。

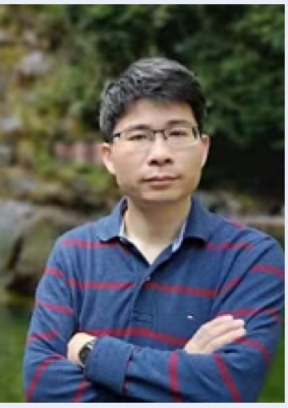


TAO Feng 陶锋
College of Philosophy, Nankai University, China 南开大学哲学院

Bio 简介：陶锋，男，南开大学哲学院副教授，清华大学哲学系博士，波恩大学访问学者。研究方向：人工智能哲学与美学，语言美学，在国内外核心期刊发表人工智能美学、哲学论文十多篇。被《中国社会科学文摘》《高等学校文科学术文摘》《人大复印资料》转载多篇。科技部“人工智能治理”“科技伦理”原则咨询专家、中国新一代人工智能发展战略研究院专家、天津社科联人工智能哲学专家
Website 主页： N/A

Title 标题: 人工智能模拟时代的艺术品
The Artworks in the Era of Artificial Intelligence Imitability

Abstract 摘要：与机械复制时代的表面形象复制不同，人工智能通过模拟人类的艺术机制来生成艺术品。虽然目前的人工智能并不是有意识地生成艺术，但是人工智能已经成为了艺术生成的“代理”式主体，这迫使我们需要打破艺术的“人类中心主义”。我们可以从以下几点来解析人工智能艺术，1.人工智能艺术的合法性何在？2.人工智能艺术生成的特点是1）行为上的，2）大数据-小任务的，3）内禀的，缺乏与外界的互动，没有社会-历史维度。3.人工智能艺术产生了“灵晕”再造、情感计算和真假判断三种变革，人工智能中的真假判断是主客符合意义上的真假，不具有“真理内涵”。4.我们可以从本雅明与阿多诺的论争来看人工智能与艺术民主性。5.还可以从艺术终结论来讨论，人工智能终结了某些艺术门类的先锋性以及艺术的欣赏价值，但是不会终结艺术。最后，从艺术-技术的关系看，人工智能艺术是两者的浅层融合，从海德格尔的艺术真理观、阿多诺的审美合理性概念出发，真理和审美理性才是艺术与技术融合的本质要素。



TU Liangchuan 涂良川
School of Marxism, South China Normal University, China 华南师范大学马克思主义学院

Bio 简介：涂良川，男，重庆云阳人。2000年毕业于东北师范大学数学教育专业，获得理学学士学位；2003年毕业于东北师范大学政法学院马克思主义哲学专业，获哲学硕士学位；2009年毕业于东北师范大学马克思主义理论与思想政治教育专业，获法学博士学位；2016-2017年加拿大女王大学哲学系访问学者；2016年12月被东北师范大学增列为博士生导师，2018年6月晋升为教授。
现任华南师范大学马克思主义学院学术委员会主任、教授、博士生导师，兼任中国辩证唯物主义学会理事、马克思主义哲学史学会理事。主要从事马克思主义哲学基本理论、马克思主义基本原理、人工智能哲学、马克思主义哲学与现实问题研究。主持国家社会科学基金青年基金、国家社会科学基金一般项目、国家社会科学基金重大项目子项目等项目；在《哲学研究》、《现代哲学》、《吉林大学学报》、《东北师大学报》、《南京社会科学》等学术期刊发表系列学术论文，出版学术专著4部；获省级社会科学优秀成果三等奖1项。

Website 主页 http://mks.scnu.edu.cn/a/20200317/591.html

Title 标题：人工智能“高阶自动化”的主体可能性——兼论人工智能奇点论的存在论追问
The Subject Possibility of "Higher Order Automation" in AI——An Inquiry on the Singularity of AI

Abstract 摘要： 人工智能是否可能获得主体性一直是一个极富争议的问题。乐观者基于技术的可能性从逻辑上肯认人工智能的主体可能性和“奇点”的逻辑必然性，悲观者则担忧人工智能获得主体性必然引发物种奇点的难题。事实上，人工智能作为高阶自动化的自治系统本质上是人的创造性产物，体现出了以反馈替代认知、以因果组织描述存在、以数安建构自我参照等类主体的能力。我们认为，虽然判定人工智能是否具有主体性是一个为时尚早、富有疑义的问题，但是认真面对人工智能的高阶自动化表现出来的类主体性，在人工智能时代认真面对人为理解我而创造出来的对象，并从这种对象的功能与效应中重新思考主体的本质，却是我们在人工智能时代理解人必须要解决的问题。

WANG Leye 王乐业
Center of Frontier Computing Science, Peking University 北京大学前沿计算研究中心

Bio 简介： Leye WANG is an assistant professor at Peking University. He obtained Ph.D. at Telecom SudParis and Université Pierre et Marie CURIE (UPMC), Paris, in 2016. Before joining Peking University, he was a postdoc research associate at the Hong Kong University of Science and Technology. His research interests include crowdsensing, urban computing, data privacy, and federated learning. He has published more than 70 papers with 3000+ citations according to Google Scholar. His work received the Honorable Mention award at UbiComp (2016) and the news highlight at Artificial Intelligence journal (2019).
Website 主页： https://wangleye.github.io/

Title 标题: Principle of Least Sensing and Computing: Building an Intelligent System with Minimum Data
最小感知计算原则：探索基于最小必要数据的智能系统构建

Abstract 摘要: With the worldwide emergence of data protection regulations, how to conduct law-regulated big data analytics becomes a challenging and fundamental problem. This talk introduces the principle of least sensing & computing, a promising paradigm toward law-regulated big data analytics. Under the guidance of this principle, various techniques including sparse sensing, differential privacy, and federated learning can be integrated to build an intelligent system with the minimum data.

WANG Yanjing 王彦晶
Department of Philosophy and Religious Studies, Peking University 北京大学哲学与宗教学系

Bio 简介： Yanjing Wang obtained his Ph.D. in logic from the University of Amsterdam, and is currently an associate professor at the Department of Philosophy and Religious Studies of Peking University, where he also serves as a vice-chair. His main research area is modal logic and its applications in philosophy, AI, and theoretical computer science. In recent years, Yanjing Wang has been initiating and promoting the systematic study of logics of know-wh (whether/how/why/who and so on) by introducing the so-called "bundled modalities" packing quantifiers and modalities together. It also leads to a family of new decidable fragments of first-order modal logic and a general approach to interpreting intuitionistic logic and various other non-classical logics as epistemic logics. He is also the director of the PKU Center for Logic, Language and Cognition, and an associate editor of the Journal of Philosophical Logic.
Website 主页： wangyanjing.com

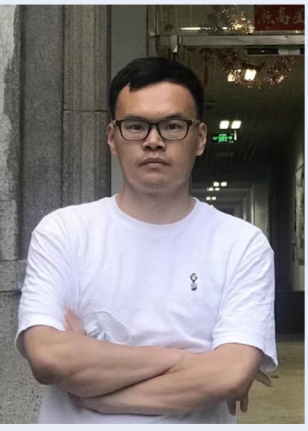
Title 标题: Knowing How to Plan
知道如何去计划

Abstract 摘要: In this talk, I aim to demonstrate how philosophy and AI can help each other with logic as the bridge, by an overview of our recent line of work on the logics of knowing how. The interaction between philosophy and AI in our work has two directions, which can be summarized as planning-based know-how and know-how-based planning. For the former, by using various notions of automated planning in AI, we give proper formal semantics of the know-how operator in our logic, fleshing out the key ideas of the intellectualist account of knowledge-how in epistemology. We axiomatize the resulting logics of knowing how, which raised interesting new philosophical questions. For the latter, we use our logics of knowing how to introduce higher-order epistemic planning in AI with highly non-trivial goals, and provide the algorithmic method to handle the apparently more complicated planning problem without paying the computational price, thanks to the logical framework.

WU Tung-Ying 吴东颖 （First Author 第一作者)
Institute of Philosophy, Chinese Academy of Sciences 中国科学院哲学研究所

Bio 简介：吴东颖，中国科学院哲学研究所讲师。复旦大学哲学学院博士后；密苏里大学哲学博士；密苏里大学圣路易分校硕士；台湾大学法律学士，其主要研究领域为科学哲学、认知与心灵哲学、因果模型与决策理论。论文发表于 Philosophy of Science, Philosophy of the Social Sciences, Philosophia 等国际一流哲学刊物。
Website 主页:
https://philpeople.org/profiles/tung-ying-wu

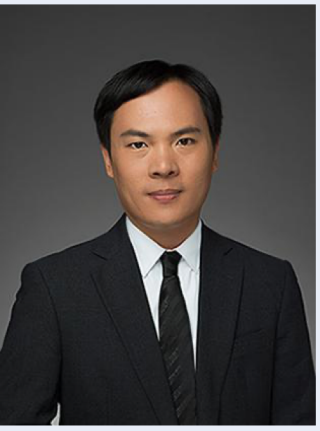
Wu Xiao'an 吴小安 (Coauthor 合作作者）
School of Marxism, Northwestern Polytechnical University, China 西北工业大学马克思主义学院



Bio 简介: 吴小安，西北工业大学马克思主义学院副教授，清华大学博士后，首都师范大学外国哲学博士，主要研究领域为因果性和条件句逻辑，在相关科学哲学杂志发表论文若干。
Website 主页: N/A

Title 标题：因果贡献度的困境与证成
The Dilemma and Justification of the Degree of Casual Contribution

Abstract 摘要：Recent papers have raised a number of key questions regarding causal contribution that can come in degrees. Specifically, Carolina Sartorio argues that we normally think about a cause as either a necessary condition or sufficient condition for its effect, but these two notions conflict in her examples. In one of her examples, the cause is necessary to its effect to a larger degree, but sufficient for its effect to a less degree. In her other example, the cause is necessary to its effect to a less extent, but sufficient for its effect to a larger extent. However, if we have only one notion of degrees of causal contribution, it is incoherent in these examples. Thus, she is skeptical about whether causal contribution can come in degrees. This paper suggests that an approach to degree of causal contribution that meets Sartorio's challenge. The approach measures a cause's causal contribution in terms of both the degree to which the cause is necessary for its effect, and the degree to which the cause is sufficient for its outcome.



YANG Yuchao 杨玉超
School of Integrated Circuits, Peking University 北京大学集成电路学院

Bio 简介: Yuchao Yang is a Professor with Tenure in School of Integrated Circuits and serves as Director of Center for Brain Inspired Chips at Peking University. His research interests include memristors, neuromorphic computing, and in-memory computing. He has published over 120 papers in refereed journals and conferences such as Nature Reviews Materials, Nature Electronics, Nature Communications, Nature Nanotechnology, Science Advances, IEDM, etc. as well as 5 book chapters. His papers have been cited >6000 times, with an H-index of 33. He was invited to give >30 keynote/invited talks on international conferences and serves as TPC chair or member for 9 international conferences. Yuchao Yang serves as the Associate Editor for 2 journals including Microelectronic Engineering and Nano Select, and editorial board member of Chip, National Science Review and Science China Information Sciences. He was invited to guest edit 3 special issues and write 12 News & Views, review articles, etc. He is a member of IEEE and MRS. He is a recipient of the National Outstanding Youth Science Fund, Qiu Shi Outstanding Young Scholar Award, Wiley Young Researcher Award, MIT Technology Review Innovators Under 35 in China, and the EXPLORER PRIZE.
Website 主页： http://yuchaolab.cn/

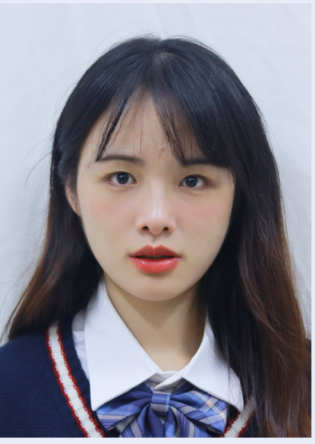
Title 标题：基于忆阻器的高效AI芯片与学习系统
Memristor Enabled High-efficiency AI Chips and Learning Systems

Abstract 摘要：Since the connection of the theoretical memristor concept with physical resistive switching devices in 2008, tremendous progress has been made in terms of material and device technology developments and their applications in memory and computing systems. The physical embodiments of memristors correspond to various resistive switching devices based on different mechanisms. These mechanisms endow the memristors with rich nonlinear dynamics, which is key to constructing biologically plausible dynamic computing systems. Memristor can be described as a set of differential equations that indicate how the internal state variables determine device characteristics and how external electrical stimulations influence these state variables. The increases in the number of state variables and internal dynamics have dramatically enriched the dynamics and functionality of memristors. Further exploration and control of such dynamics are essential for highly efficient information processing applications.

ZENG Yi 曾毅
Institute of Automation, Chinese Academy of Sciences 中国科学院自动化研究所

Bio 简介: Yi Zeng is a Professor and Deputy Director at Research Center for Brain-inspired Intelligence, and Director of the International Research Center for AI Ethics and Governance, both at the Institute of Automation, Chinese Academy of Sciences. Yi is also a Professor at Institute of Philosophy, Chinese Academy of Sciences. He is a board member for the National Governance Committee of Next Generation Artificial Intelligence, China. He is in the Expert Group on the Ethics/Governance of Artificial Intelligence in Health, World Health Organization (WHO), and is an expert in the UNESCO Adhoc Expert Group on AI Ethics and the UNESCO Expert Group on the Implementation of AI Ethics. He is the lead drafter of Beijing Artificial Intelligence Principles, and one of the major drafters for the National Governance Principles of New Generation Artificial Intelligence, China. He co-founded AI4SDGs Cooperation Network (http://www.ai-for-sdgs.academy/), and lead the efforts on Linking AI Principles (http://www.linking-ai-principles.org/), AI Governance Online (http://www.ai-governance.online/), and Cognitive Imitation Games (http://www.cognitive-imitation-games.ai/). His major research interests focus on Brain-inspired Artificial Intelligence models, AI Ethics and Governance, and AI for SDGs.
Website 主页： http://bii.ia.ac.cn/~yizeng/

Title 标题：Building and Living with Brain-inspired Ethical AI
构建符合伦理的类脑人工智能并与其共生
Abstract 摘要：N/A



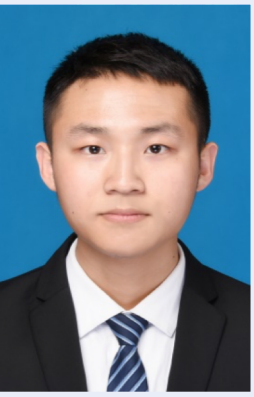
ZHANG Mengting 张梦婷
Department of Philosophy at Sun Yat-sen University 中山大学哲学系

Bio 简介：Mengting Zhang has recently received her master's degree in philosophy from the Department of Philosophy at Sun Yat-sen University and is currently applying for a doctoral position in Philosophy of Cognitive Science and Philosophy of Neuroscience. Her work focuses on the mechanisms of the predictive brain and on the spontaneous thinking activities, including dreams. She intends to apply researches in machine learning, such as the self attention mechanism and the methods of solving the over-fitting and under-fitting problem of the model, to the understanding of brain activity, especially to understanding spontaneous thoughts.

Website 主页：
https://www.researchgate.net/profile/Zhang_Mengting7

Title 标题： A New Understanding of the Role of Dreaming from the Perspective of Model
对梦的新解释：从模型的角度看

Abstract 摘要： This thesis proposes to provide a new interpretation of the evolutionary role of the phenomenon of dreaming, by combining the discussion of models in the field of machine learning. Recent overfitted brain hypothesis and prediction error minimization framework have offered a new explanation for the role of dreaming. They have understood the brain as a model, and considered dreaming as an activity of the brain model to preclude the over-fitting problem. Contrary to these views, this thesis states the hypothesis that the brain model should be understood as containing multiple sub-models and for the correspondent data of a specific domain, some sub-models may be over-fitting, some models may be under-fitting and other models may be valid. And on the basis of this hypothesis of the brain, the brain's activity during dreaming should be regarded as consisting of three parts: using the effective model to process the data, solving the under-fitting problem, and solving the over-fitting problem. And many dream experiences including vagueness, incongruity, discontinuity, and the precognitive dream can be explained in this kind of understanding of the brain, by referring to the methods of solving the under-fitting and over-fitting problem in machine learning. Also, other features of dreaming, such as the narrative feature, can be explained under the general framework of understanding the brain as a model.



ZHANG Kun 张琨
School of Health Humanities, Peking University 北京大学医学人文学院

Bio 简介: 张琨，现为北京大学医学人文学院医学伦理与法律系、北鹏前沿科技法律研究院助理研究员，在腾讯研究院法律研究中心工作期间，主要从事人工智能伦理、医疗大数据等领域的研究，目前主要研究方向为卫生法学、医学/生命伦理学、医疗人工智能、科技伦理。曾作为主要参与人参加并申请教育部哲学社会科学研究重大课题攻关项目《人类辅助生殖技术的法律规制研究》，在国内外重要学术期刊发表中外文学术论文多篇，曾参与编写并出版《中华人民共和国民法典释论》（法律出版社）关于“侵权责任编”的撰写。受邀在中华医学会医学伦理学分会2021年年会、中国法学会经济法学研究会2021年年年会等作主题发言。
Website 主页： N/A

Title 标题: 论医疗决策中人工智能技术运用的法律责任——以给药剂量为例
On Legal Liability for the Use of Artificial Intelligence Technology in Medical Decision Making: The Example of Drug Dosing

Abstract 摘要: 医疗决策是医患关系的主要方面，提出给药剂量建议是医疗决策的重要场景，具有必要性。医疗决策模式的智能化转向引起了对新兴技术的警惕。医疗人工智能参与给药剂量确定的医疗决策中，为避免讨论法律责任“方法论上的盲目飞行”，需先行“搜寻法律制度的哲学基础”。给药剂量的人工智能技术参与具有不同情境，对应不同的法律责任承担方式。《民法典》关于医疗损害责任的制度设计缺乏多元医疗决策模式的场景化嵌入，应当设计出医疗决策中运用人工智能法律责任的评价模型，在不同情境中，从知情决策、共享决策、家长制、工具性决策四种决策模式中进行场景化讨论，检视其情景化和场景化的归责基础。

ZHANG Ping 张萍
School of Law, Peking University 北京大学法学院

Bio 简介:
Website 主页： N/A

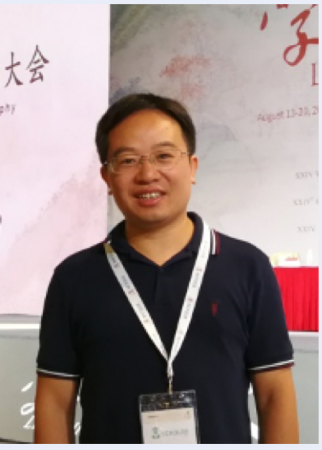
Title 标题:
Abstract 摘要:

ZHANG Weite 张伟特
Department of Philosophy, Tsinghua University
清华大学哲学系

Bio 简介：张伟特：清华大学哲学系&新雅书院助理教授、《清华西方哲学研究》副主编、仲英青年学者。2002-2006年清华大学比较文学专业本科，2006-2009清华大学中国古代文学专业硕士，2009-2016德国海德堡大学哲学专业博士。2017年加入清华大学。学术研究领域：系统性哲学（认识论、形而上学、元哲学），历史性哲学（笛卡尔、斯宾诺莎）等领域，聚焦哲学基础理论问题研究。学术代表作有专著Descartes' Doctrine of Clear and Distinct Perception（《笛卡尔的清楚分明感知学说》，2016），论文《笛卡尔的真理理论》（《哲学研究》2018）等。近年也深入参与清华大学本科教育教学改革和通识教育探索，担任清华大学通识改革专家工作组专家、新雅书院院长助理、哲学学堂班授课教师等工作。
Website 主页： N/A

Title 标题: 笛卡尔与人工智能："我思故我在"作为智能测试标准的可能性
Descartes and Artificial Intelligence: th Possibility of "I Think, Therefore I am" as a Standard for Intelligence Testing

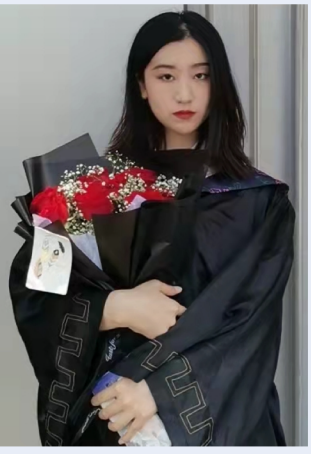
Abstract 摘要:笛卡尔作为西方近代哲学奠基人，开启了一种主体性哲学和认识论转向。笛卡尔对人和机器的本质有着系统的思考，提出了语言测试和理性行为测试作为判断智能的标准，但是他的机器理论和身心二元论框架使得他认为人工机器不可能通过这两个测试。因此，笛卡尔预言人工智能并无可能。本文认为，笛卡尔哲学具有极大的复杂性，这个不可能性预言并未耗尽笛卡尔哲学思考人工智能问题的理论潜力，他的身心三元论框架在理论上并未排除人工智能机器的可能性，而且最具启发性的是，《沉思集》还潜在地提供了第三个更有潜力的智能测试标准："我思故我在"测试。



ZHANG Xueyi 张学义 (First Author 第一作者)
School of Humanities, Southeast University, China 东南大学人文学院

Bio 简介: 张学义，哲学博士，东南大学哲学与科学系副教授，副系主任，硕士生导师，中国自然辩证法研究会问题哲学专业委员会副秘书长，中国知识论学会、江苏自然辩证法研究会理事。
2002年进入东南大学社会学专业学习，2006年获法学学士学位，同年转读科技哲学专业，师从马雷教授，先后于2009、2013年获得哲学硕士、博士学位，遂毕业留校任教；2011年10月 - 2012年1月期间，赴美国Rutgers大学短期访学，合作导师为美国文理科学学院院士Stephen Stich教授。

目前从事科技哲学的教学与科研工作，主要研究兴趣：科学哲学、实验哲学、科技伦理，同时涉猎问题哲学、心智哲学、认知科学哲学、人工智能哲学等领域。在《哲学动态》、《哲学分析》、《中国社会科学报》、《自然辩证法研究》、《自然辩证法通讯》、《科学技术哲学研究》、《东南大学学报（哲学社会科学版）》、《东北大学学报（社会科学版）》、《科技进步与对策》等期刊杂志上发表学术论文近20篇。其中，《行动者网络理论视阈下的物联网技术》、《实验哲学：一场新的哲学变革》、《实验的知识论——一种实验哲学的研究进路》、《专家直觉与大众直觉之辨——实验哲学的方法论基础新探》、《脑机融合技术的哲学审思》等论文先后被人大复印资料全文转载。目前，主持完成1项国家社科基金项目，参与多项国家及省部级课题工作。
Website 主页：N/A

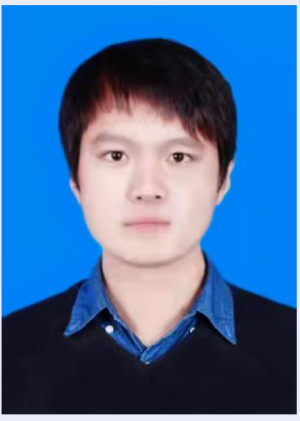


WANG Xiaoxue 王晓雪 (Coauthor 合作作者）
School of Humanities, Southeast University, China 东南大学人文学院

Bio 简介：王晓雪，东南大学本科生，2017年进入东南大学哲学专业学习，研究兴趣:实验哲学、科技伦理。
Website 主页：N/A

Title 标题： “伦理旋钮”：破解无人驾驶算法困境的密钥？
A Survey for Ethical Knob of Moral Algorithm of Driverless Car

Abstract 摘要： 针对无人驾驶的面临算法困境，意大利学者朱塞佩·康蒂斯萨等人提出了“伦理旋钮”算法理论：将无人驾驶汽车的算法设置权交付给车主，从而能够化解无人驾驶汽车在造成事故时所产生的道德-法律归责困境。但该算法可能会陷入集体性的“囚徒困境”：大多数车主为自保而在算法设置做出极端利己的选择，致使社会总伤亡程度增加。为此，本课题组运用实验哲学研究方法，将“伦理旋钮”与现有的伦理算法：罗尔斯算法、制动力学算法以及功利主义算法进行经验性比较，验证人们对各种伦理算法的接受度和对搭载了该伦理算法的无人驾驶汽车的购买欲求。数据表明：伦理旋钮虽然并没有完全消解道德-法律归责困境，但其在无人驾驶汽车事故归责方面要比其他算法更加明晰，具有明显的稳定性；该算法可能存在的“囚徒困境”在实际操作层面并未出现，且可以通过法律等途径对极端利己主义模式进行相应的控制；安装了伦理旋钮的无人驾驶汽车具有更大的市场可行性。“伦理旋钮”算法理论为破解无人驾驶算法困境提供了可能的解决方案，也是实验哲学研究方法在伦理学领域应用的有益尝试。



ZHAO Song 赵松
School of Politics and Public Administration, Qufu Normal University, China 曲阜师范大学政治与公共管理学院

Bio 简介：赵松，哲学博士，曲阜师范大学政治与公共管理学院讲师，研究方向为西方哲学、马克思主义哲学和人工智能哲学。
Website 主页： N/A

Title 标题: 哲学与人工智能的根本联系：生存论的前提反思 Exploration of the Fundamental Connection between Philosophy and Artificial Intelligence

Abstract 摘要:人工智能自身就包含着一种哲学理解，即把智能当作人的本质，因此旨在研究人类思维、心智和认知规律的认知科学处在哲学和人工智能跨学科研究的内在联系和核心地位上。但认知科学属于认识论范式，它始终带有近代哲学“我思”的意识哲学和理性主义的痕迹，无法解决主观的思维如何通达和切中思维之外的客观存在的问题。因此，把认知科学作为哲学和人工智能的内在联系实质上是把人工智能建立在了一个并不牢靠的根基之上。哲学在根底上作为关于对人之存在方式进行自觉反思的学说，其和人工智能的最根本的联系是在生存论层面上，即对人的存在的认识。人工智能诸范式的形成都依赖于如何理解智能，而人们对于智能理解的偏差归根结底在于对人的存在的认识的不同。

ZHAO Zhoukuan 赵周宽
School of Chinese Language and Literature, Xi'an International Studies University, China 西安外国语大学中国语言文学学院

Bio 简介： 赵周宽，陕西扶风人。1994年于陕西师范大学获文学学士学位，1997年于陕西师范大学获文学硕士学位。1997年6月至2019年6月，在西安外国语大学汉学院工作，主要从事留学生汉语教学工作和汉语国际教育专业本科、研究生教学工作。同时担任全校通识课《中西思想文化比较》课程的教学。在汉学院工作期间，于1999年10月至2000年6月，在北京语言文化大学进修基础阿拉伯语，2000年10月至2001年8月赴埃及开罗大学做访问学者，2003年3月至2005年5月担任阿拉伯联合酋长国迪拜警察总署汉语教学工作，以阿拉伯语教授汉语并编写汉语教材，2010年5月获文学博士学位（陕西师范大学），2012年12月至2015年10月在西安交通大学人文学院哲学博士后流动站从事比较哲学研究，2013年在中文学院担任本科生古代文论课、研究生西方美学史等课程的教学。现就任于西安外国语大学中国语言文学学院，副教授，硕士生导师，中华美学学会、陕西省外国文学学会会员，中国比较文学下属文学人类学学会理事，副秘书长。研究课题涉及比较文学、现象学、身体哲学和数字人文等领域。
Website 主页：
https://zwxy.xisu.edu.cn/info/1074/3781.htm

Title 标题: 后人类世伦理问题的基本思考
Basic Thoughts on Ethical Issues in the Post-Human Era

Abstract 摘要:本文将人工智能取得与人类平等伦理地位的未来可能世界作为"后人类世"的开端，并在此设定中预先对人机交互中的伦理问题做出思辨性探讨。这一研究与人工智能技术的当代发展实际状况相关，但却不以之为必要前提，而是用类似思想实验的方式对人工智能伦理主体的形成、身份、人格结构、程序伦理等问题展开辨析。由于人工智能伦理主体的形成，一种基于算法和程序的程序伦理将在人机交互中形成，程序伦理将在德性伦理与责任伦理之后构建起未来世界新的伦理交往规则。以程序伦理为基本规则的人机伦理交互，将在现实增强之后进一步实现世界的伦理增强。基于最新人工智能技术的哲学思辨，作为预流，开辟着未来的思想世界；反过来，人工智能技术的迅猛发展也促成了哲学范式的转变。

Zhou Lujun 周旅军
School of Social Work, China Women's University 中华女子学院社会工作学院

Bio 简介: 周旅军，北京大学社会学博士，研究领域为计算社会科学，主要方向包括可信赖的自主系统、新兴技术治理和数字人文研究。主持国家社会科学基金项目《人工智能促进社会性别平等研究》等，在数据库建设方面获得国家级数据库应用系统设计高级技术证书。常年为北京大学硕博士生教授大数据理论及其在社会科学中的应用。中国社会学会性别社会学专业委员会理事，中国社会学会计算社会学专业委员会理事，中国社会科学院国情调研与大数据研究中心特约研究员，中山大学国家治理研究院研究员，清华大学数据治理研究中心兼职研究员。

Website 主页：
http://www.cwu.edu.cn/xbyshfzxy/szdw/szms/33863.htm

Title 标题： 人工智能中的性别不平等治理何以可能？——基于文化堕距视角的算法规制路径分析

How is the Governance of Gender Inequality in Artificial Intelligence Possible: Analysis on Algorithm Regulation Path from the Perspective of Cultural Lag

Abstract 摘要：搜索引擎、短视频服务、共享经济众包平台等人工智能应用因其存在强化性别偏见的社会风险，不断引发社会关注和讨论。如何防止人工智能重演甚而加深现实社会中的性别问题，并充分利用新兴技术来促进性别平等措施的落地，是推进新兴技术治理的重要内容。本研究基于数据/知识、算法和算力的维度考察人工智能生态中存在的性别不平等问题，进而透过文化堕距的视角解析新兴技术治理的现实困境，指出科技伦理治理、法律政策规制、行业协同治理、企业自我治理等构成的科技治理体系，需要社会文化的转型才能更好地解决根植于历史中的不平等问题。同时，以人工智能平台企业的内部治理为例，结合对相关从业者的问卷调查，提出性别不平等治理的算法规制路径。

备注：本研究是“人工智能促进社会性别平等研究”（国家社会科学基金项目，编号20BSH040）的阶段性成果

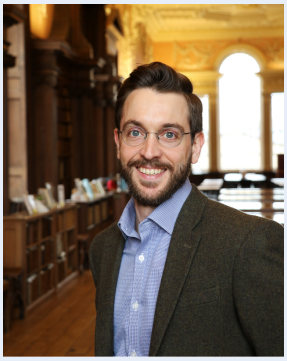
ZHU Jing 朱菁
College of Humanities, Xiamen University, China
厦门大学人文学院

Bio 简介 :朱菁，厦门大学哲学系教授，兼任厦门大学美育与通识教育中心副主任，教育部“长江学者”特聘教授（2008年入选）。研究领域为科学哲学、认知科学哲学、人工智能哲学和认识论等。
Website 主页：
https://phi.xmu.edu.cn/info/1116/3006.htm

Title 标题： 人工智能可解释性问题的哲学探索
A Philosophical Exploration of the Interpretability Problem of Artificial Intelligence

Abstract 摘要： 人工智能的可解释性近年来成为了一个颇受关注的问题，学术界、产业界和公共管理部门均对此给予重视。一方面由于以机器学习为代表的当代人工智能技术具有显著的“黑箱“特征，另一方面则因为人工智能技术和系统的日益扩展的应用范围，导致对智能系统的运作给出合适的解释说明，成为了一种看似合理的需求和规范。发展可解释的人工智能可以为推进负责任的、可问责的和可信任的人工智能提供必要的理论和伦理基础。本次报告力求阐明：（一）由于解释历来是哲学研究的重要课题，哲学对于理解和阐明人工智能的可解释性问题可以起到有益的作用，而关注人工智能的可解释性问题对于丰富和发展哲学关于解释之本质的探究也同样会有所裨益，哲学与人工智能研究在可解释性问题上加强互动交流，具有良好的合作共赢前景；（二）解释的多元主义或可为理解人工智能的可解释性提供合适的理论框架，有助于容纳面向智能系统的多目标、多层次和多种模式的解释需求与活动；（三）”以学习为中心的解释“可以为人工智能解释的多元主义提供一种具有整合性的指向和概念基础。

Organizers



Sebastian SUNDAY GRÈVE 王小塞
Assistant Professor, Department of Philosophy and Religious Studies, Peking University 北京大学哲学与宗教学系

Bio 简介： Sebastian's philosophical education began in Germany, at the University of Göttingen, and continued in England, first at the University of London, then Oxford. His research and teaching cover a large spectrum of issues ranging from aesthetics and epistemology to the history of analytic philosophy, logic, and the philosophy of mind. He is editor of *Wittgenstein and the Creativity of Language* (with Jakub Mácha; Palgrave, 2016), *Wittgenstein on Philosophy, Objectivity, and Meaning* (with James Conant; Cambridge University Press, 2019), and *Culture and Value after Wittgenstein* (Oxford University Press, 2022). During the academic year 2020–21, Sebastian was a Fellow of the Berggruen Research Center at Peking University. A recent essay on AI, 'Nietzsche and the Machines', is available open-access from the website of *The Philosophers' Magazine*. Another, 'Turing's Philosophy of AI', is forthcoming at *Aeon*.

Email 邮箱: ssg@pku.edu.cn

Website 主页: https://www.yhposolihp.com/

SUI Tingting 隋婷婷
Boya Post-doctoral Fellow/lecturer, Department of Philosophy and Religious Studies, Peking University 北京大学哲学与宗教学系

Bio 简介： Sui Tingting, a Ph.D. of Southeast University, is currently a Boya post-doctoral fellow/lecturer in the Department of Philosophy of Peking University. From 2017 to 2018, she was funded by the CSC Construction High-level University Postgraduate Program and went to the Rosenfeld Lab of department of psychology, Northwestern University, USA, to study EEG event-related potentials of cognitive psychology. Her contemporary research interests include artificial intelligence philosophy, cognitive neuroscience philosophy, experimental ethics and moral psychology. She has published a number of papers on "International Journal of Psychophysiology"(SCI/SSCI) and Chinese journals including "dialectics of Nature Communications", "Studies of Dialectics of Nature", "Philosophical Analysis" and "Social Sciences of Jiangsu Province". She also participated in the compilation of entries for the section of Philosophy of Science and Technology (Philosophical Issues of Life Sciences) in the third edition of Encyclopedia of China: Philosophy Volume, and wrote entries for "genetic psychology" and "philosophical problems in psychology".

Email 邮箱: suitingting@pku.edu.cn

Website 主页： N/A

Online Conference

Zoom	*腾讯会议* (Chinese parallel sessions only)
Day 1 ------- 主题：FAAI Day One 时间：2022年4月8日 13:30 Beijing https://us06web.zoom.us/j/86923498702?pwd=dnVOSTYwRXlvQWNxT1FyQ2c2ejVCUT09 会议号：869 2349 8702 密码：123321 Day 2 ------- 主题：FAAI Day Two 时间：2022年4月9日 08:00 Beijing https://us06web.zoom.us/j/81428946292?pwd=eFd2ZFNmdFJQL29DN3lKV0hURCttdz09 会议号：814 2894 6292 密码：123321 Day 3 ------- 主题：FAAI Day Three 时间：2022年4月10日 09:00 Beijing https://us06web.zoom.us/j/85888120260?pwd=b3EwLzJwQStVWTZtc0Z2Qm1oZGxidz09 会议号：858 8812 0260 密码：123321	Day 1 ------- 主题：FAAI Day One 时间：2022年4月8日 16:10 Beijing https://meeting.tencent.com/dm/W7sL39J6LxuE 会议号：319-391-790 密码：123321 Day 2 ------- 主题：FAAI Day Two 时间：2022年4月9日 10:10 Beijing https://meeting.tencent.com/dm/V1fmOT3VDhWq 会议号：508-491-798 密码：123321 Day 3 ------- 主题：FAAI Day Three 时间：2022年4月10日 11:10 Beijing https://meeting.tencent.com/dm/yx3b7916Fki5 会议号：395-446-674 密码：123321

Livestreams 直播

微信视频号：
WeChat Channel:

Bilibili
https://space.bilibili.com/342002870/
YouTube
https://youtu.be/05hG7lGRM1M
DAY1 YouTube
https://youtu.be/hcWqPkje7Ec
DAY 2 YouTube
https://youtu.be/2bd_Mtm16YY
DAY 3 YouTube
https://youtu.be/53mN3P3bg9M

Contact

- **Contact Us**

CHENG Yihui 程乙惠 Peking University
chengyihui@stu.pku.edu.cn
GONG Yao 公尧 Peking University
yao.gong@stu.pku.edu.cn
LUO Shiman 罗诗曼 Peking University
luolouna@gmail.com

- 如果您有任何问题，敬请与我们联系！

感谢参与！

Thanks for Participation!