

# 从英国到中国的人工智能 伦理与安全概览



之江实验室  
ZHEJIANG LAB

## 构建面向可问责、负责任、透明的人工智能

2021-4-9 (周五) 14:00 | 人工智能小镇15号楼2层206室



### 钟蕙心

英国巴斯大学可问责、负责任、透明的人工智能博士研究生在读

Huixin Zhong is a Ph.D. candidate in accountable, responsible, and transparent AI at the University of Bath. She received an M.Res in accountable, responsible, and transparent AI at the University of Bath (UK) and an M.Sc. in social psychology at Lancaster University (UK); and a B.Sc. in applied psychology and medical law at Chongqing Medical University. Before joining the Ph.D. program in 2019, Huixin worked in the AI industry for two years. The past projects she worked on includes affective computing and rehabilitation robots for children with autism. Huixin's Ph.D. research focuses on influencing human decision making and judgements with AI. This research combines AI and behavior economics. The results are expected to be useful in understanding the relationship between humans and AI in many settings and in developing guidance and standards for the development and use of human-in-the-loop AI systems.

### ► 讲座简介

人工智能带来的诸如算法偏见和歧视，公众对于新技术缺乏信任、算法操纵、数据隐私等伦理问题日益凸显。人工智能的伦理与安全问题，已经成为人工智能发展无法回避，以及各国政府着力解决的重要问题。英国政府于2018、2019年分别针对数据以及人工智能安全与伦理出台了相关指导。在国内，国务院及国家各部委从2017年起也相继出台了一系列人工智能安全与伦理原则。中英两国在人工智能发展的指导性原则上存在诸多相似之处，然而文化、制度及社会环境差异将深刻地影响两国对相关原则的解释以及具体行业标准和法律法规的制定，并进而影响未来国际法有关人工智能伦理与安全领域法律法规的制定。在本次讲座中我将分析人工智能伦理和安全领域存在的挑战，介绍英国及中国的数据及人工智能伦理安全政策、通过对比其相似与差异探讨未来构建具有中国特色的人工智能伦理与安全治理框架、以及全球人工智能治理的发展方向。