计算机学院科研团队情况介绍表

团队名称	可信数据智能		团队负责人	蔡亮	
联系人	蔡亮	Email	leoncai@zju.edu.cn	电话	
个人主页	https://person.zju.edu.cn/0002380				

主要情况介绍：

蔡亮，研究员，博士生导师，国家级高层次人才计划领军层次入选者，中国计算机学会杰出会员，浙江大学国家制度研究院特聘研究员、可信数据研究中心主任，浙江大学区块链研究中心常务副主任，兼任浙江省区块链技术研究院院长、浙江省可信数据要素研究院院长、工信部人工智能工作委员会人机对齐工作组组长。

课题组聚焦可信数据要素与人工智能人机对齐关键技术研究与产业化，致力于我国可信数据体系、数据基础设施、可信人工智能体系建设，推动数据跨领域、跨行业的加工利用，促进要素与实体经济深度融合，确保人工智能的价值观对齐。团队已承担多个国家、省/部级研究任务，围绕“建设数据安全流通体系，赋能数据要素价值发挥，护航人工智能的价值观对齐”的目标，开展相关基础研究、技术攻关与产业应用。

1.可信数据要素关键技术主要围绕可信数据流转的底层区块链支持技术、数据产品研发的共性关键技术、数据价值评估与度量技术等方向开展研究，研究内容包括异步拜占庭性共识协议、区块链分片技术、智能合约并行与投机执行、数据应用成效评估、多源数据场景下的贡献度评估等等。团队牵头建设了全国性的数据产业联盟“全球数贸会数据智能专委会”，引

入了 40 多家数据密集型央企和特大型国企以及互联网头部企业，设计了多个跨行业、跨领域的数据产品，为多个世界 500 强企业设计了数据要素市场化、产业化的顶层规划，构建了覆盖数据集质量、数据贡献度的综合评估指标体系，开发了相应的评估工具，系统推进了场景化应用的成效评估与实践验证。

2.人工智能人机对齐研究方向主要面向意识形态安全和价值观对齐问题，针对越狱攻击技术与电子围栏技术开展研究，研究内容包括大模型越狱机制研究、跨模态一致性研究、多模态越狱与防御技术研究等。团队长期支撑中央网信办、国家广电总局、浙江网信办等监管机构关于大模型备案的合规测评；同时，受工信部电子司委托，牵头成立了全国人工智能人机对齐工作组，致力于筑牢 AI 伦理与安全防线、维护社会稳定与公共利益，同时推动技术融合、构建国际共识，以期构建开放、合作、负责任的 AI 发展环境，为 AI 技术的健康、负责任发展保驾护航。