上海交通大学自动化系
Department of Automation
Shanghai Jiao Tong University

系统控制与信息处理教育部重点实验室
Key Laboratory of System Control and Information Processing, Ministry of Education

# 学术报告会

**时间：2025年3月31日 14:00-15:00**
**地点：电信群楼4号楼一楼E谷训练营**

## 端侧AI计算的方法与实践

**储洁宇**
**华为半导体业务部Kirin解决方案**
**AI高性能计算 & ISP亮色专家**

**摘要:**

端侧人工智能作为人工智能技术的重要分支，正推动智能终端设备向高效，安全，自主的方向快速演进。端侧AI具有低延迟响应，个性化定制，数据隐私保护以及离线场景支持等优势，成为边缘计算与AI融合的关键技术范式。本报告从技术和应用场景层面，聚焦轻量化模型设计（如MobileNet、EfficientNet）、模型压缩技术（量化、剪枝、知识蒸馏）、端侧大模型落地（LLM、AIGC）以及硬件加速架构（专用NPU、边缘计算芯片）的协同优化，阐明其如何突破终端设备的算力与能效瓶颈。

**简介:**

储洁宇2017毕业于上海交通大学软件学院，历任高性能计算团队PL，XM，多次获得公司级核心奖项。负责Kirin AI高性能NN计算，保证手机端到端NN计算竞争力，支撑端侧AI业务。从NPU首代商用开始，芯片AI性能榜单业界第一，Mate、P系列等多个AI亮点业务成功商用，包括隔空手势、AI拍照、AI视频等业界首创业务。负责ISP亮色效果，完成Mate 70系列红枫首商用，华为手机颜色效果获得突破性进展，领先业界。