快手安全
KUAISHOU SECURITY

人工智能应用中的隐私保护与伦理挑战

宋文宽

小米集团信息安全与隐私委员会秘书长

小米集团人工智能伦理委员会秘书长

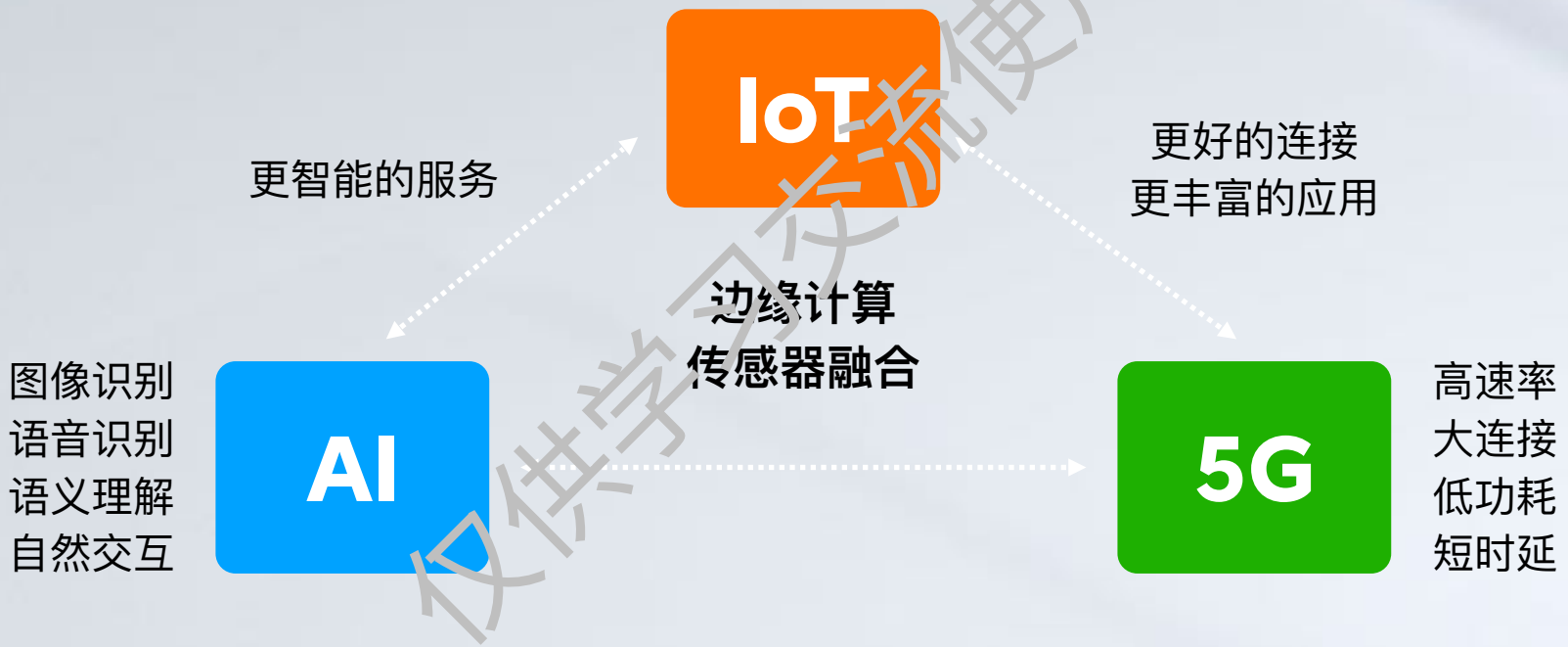
目录

Contents

01

5G+AIoT=人工智能时代

AI + IoT + 5G=万物互联智能时代



智能时代依赖AI技术处理海量数据

小米AI技术应用情况概览

02

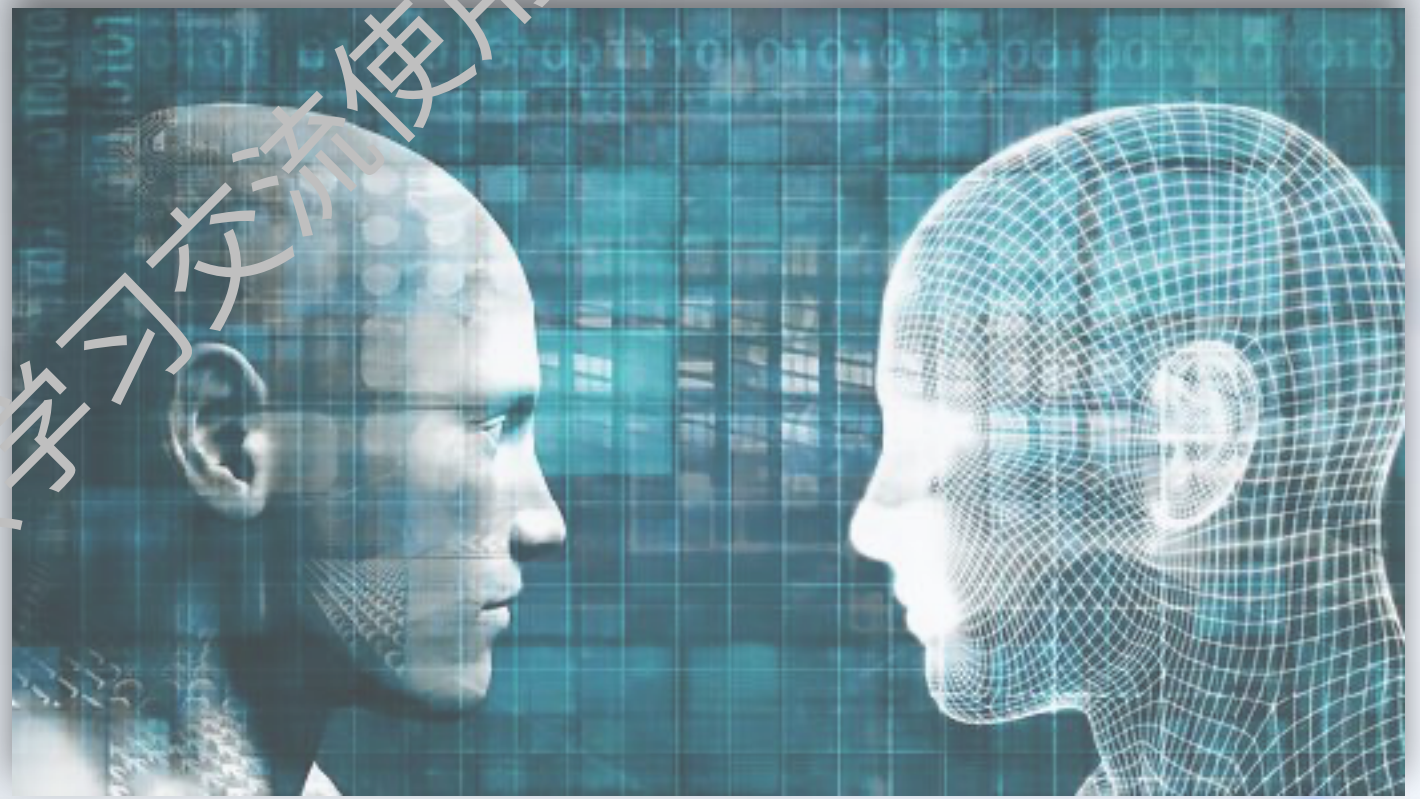
AI向善，拒绝信息茧房

人工智能（AI）面临的挑战

隐私保护

数据安全

AI伦理



越懂你，越需要更多个人数据?

产品

收集

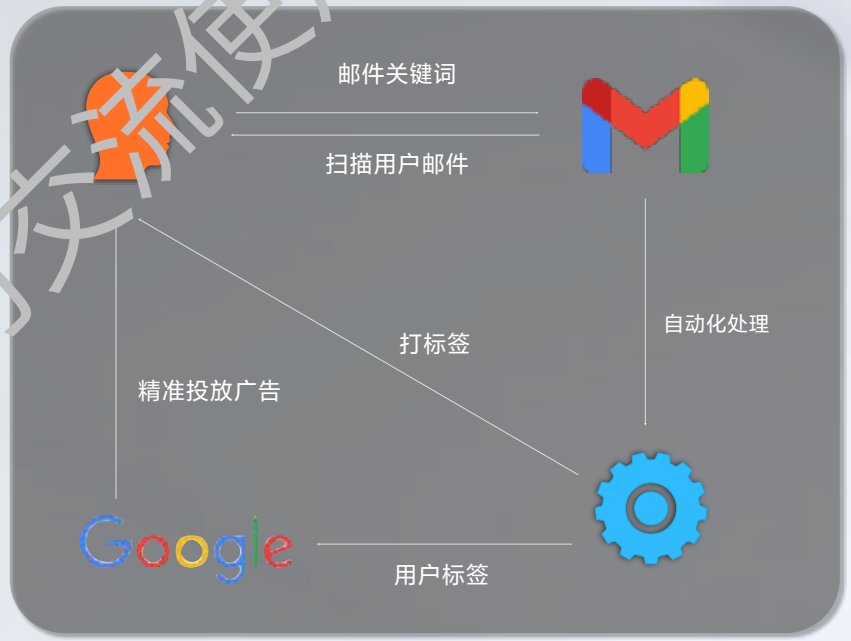
数据

训练

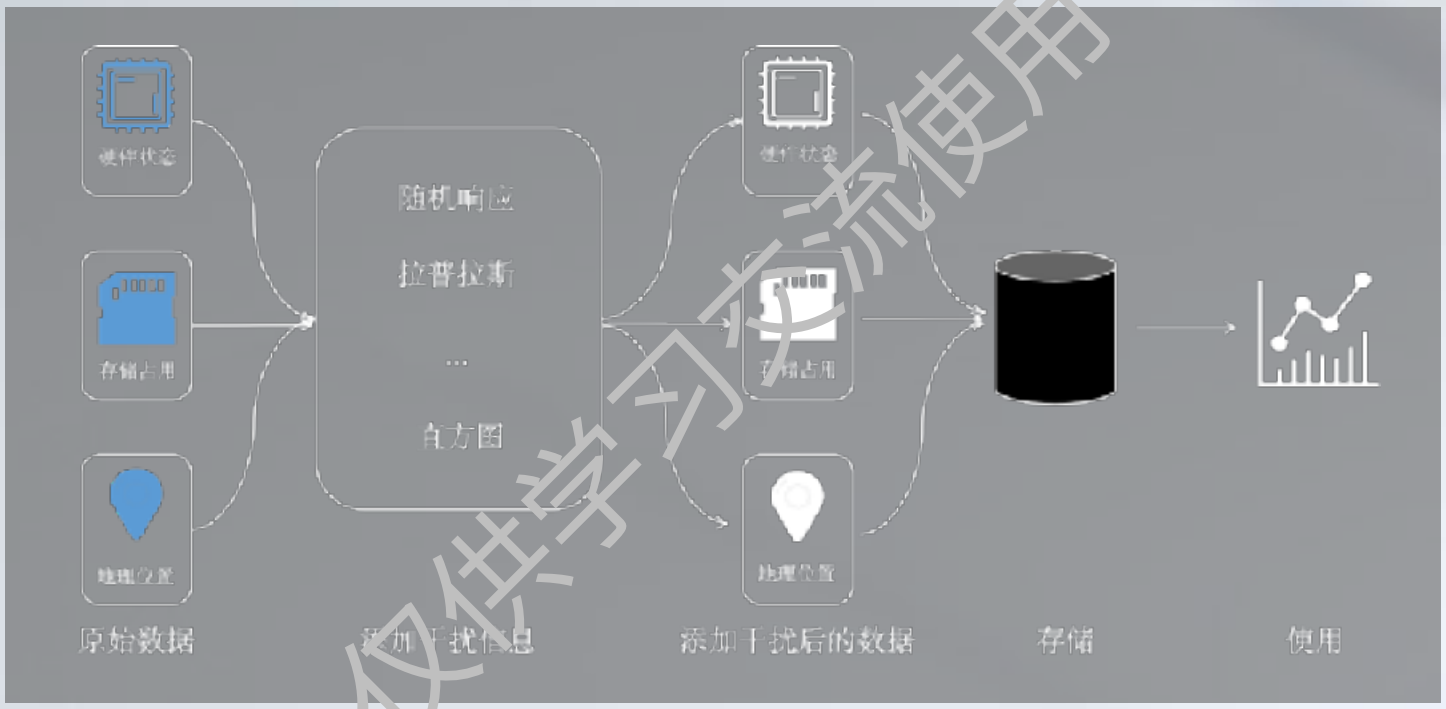
算法

算法陷阱，过度采集

商业利益驱使造成的隐私侵犯



小米差分隐私技术应用-最大化保护用户隐私



端侧数据收集源头，就增加干扰信息，确保任何人

都无法从中获取用户的准确信息，从而最大程度的保护用户隐私

小米AI视觉-手机AI图片功能隐私保护

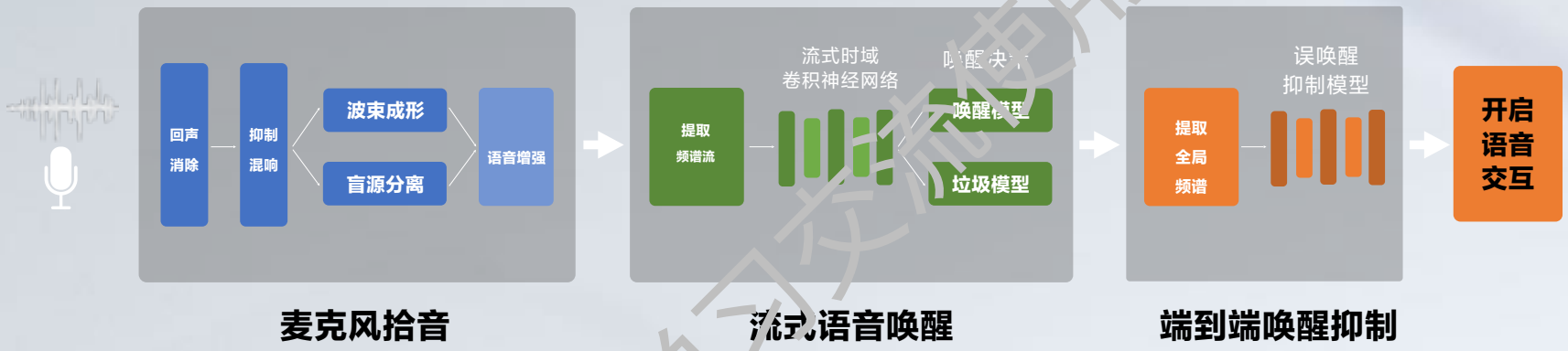
端侧AI算法

算法仅在设备本地执行，避免收集用户信息

训练数据合规来源

训练数据来源与开源图片库或正式采购训练结果安全防护，完善的标注流程

AI语音-不唤醒不触发在线服务，分级唤醒技术充分保护隐私



不唤醒不会触发在线语音服务

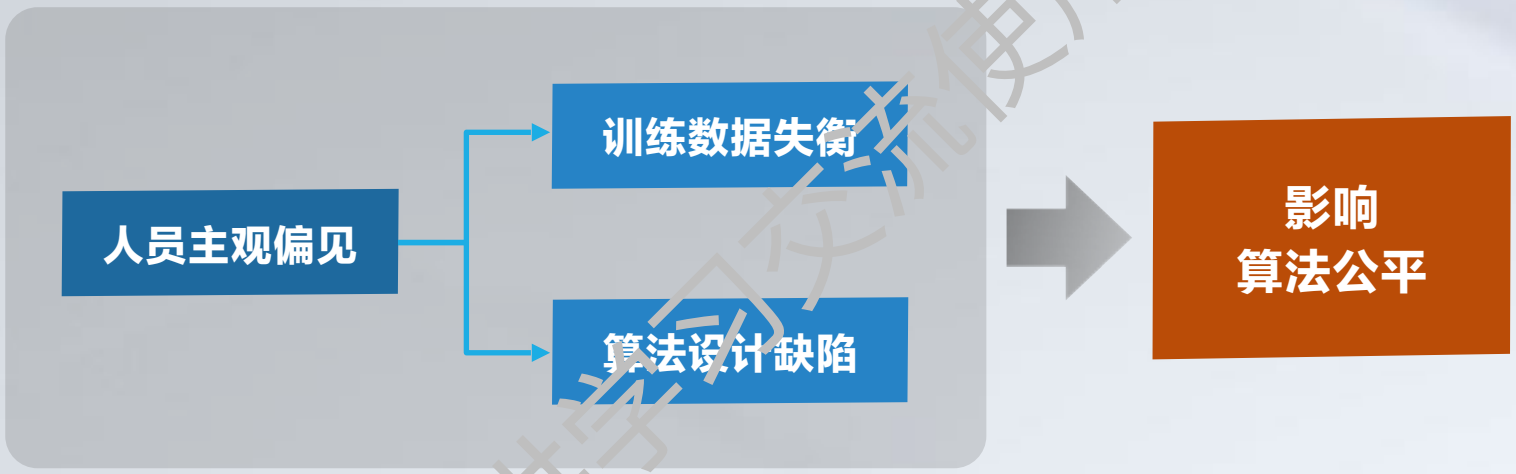
两级唤醒，利用不同特征模型视角、不同时间粒度、交叉验证抑制误唤醒，

在保证正常使用体验前提下，最大程度对用户隐私尊重和保护

03

算法公平，从源头做起

AI算法偏见是人的偏见还是技术的偏见？



AI算法的偏见，归根到底是人的偏见

AI算法偏见常见场景

性别偏见

种族偏见

年龄偏见

特殊人群偏见

人脸解锁算法公平性实践

小米可信AI原则及规范

贯穿AI模型设计、研发、测试、验收、上线、维护、销毁全周期

公平、全面、包容训练数据集

正规途径采购包含全球不同国家、年龄、性别、肤色等训练数据

多轮、全场景算法训练与测试

如：活体批处理、比对批处理、光线、角度、表情、面部遮挡等

AI模型集成在设备端运行

仅本地存储加密并不可逆的人脸特征值信息，人脸信息不离开设备

透明告知且可选择，充分尊重用户权利

提示人脸解锁风险，提供功能开关及删除人脸特征信息的选项

04

算法透明，需要可解释性

科技向善，发展可信任的人工智能

计算机或机器类人思考、类人行动

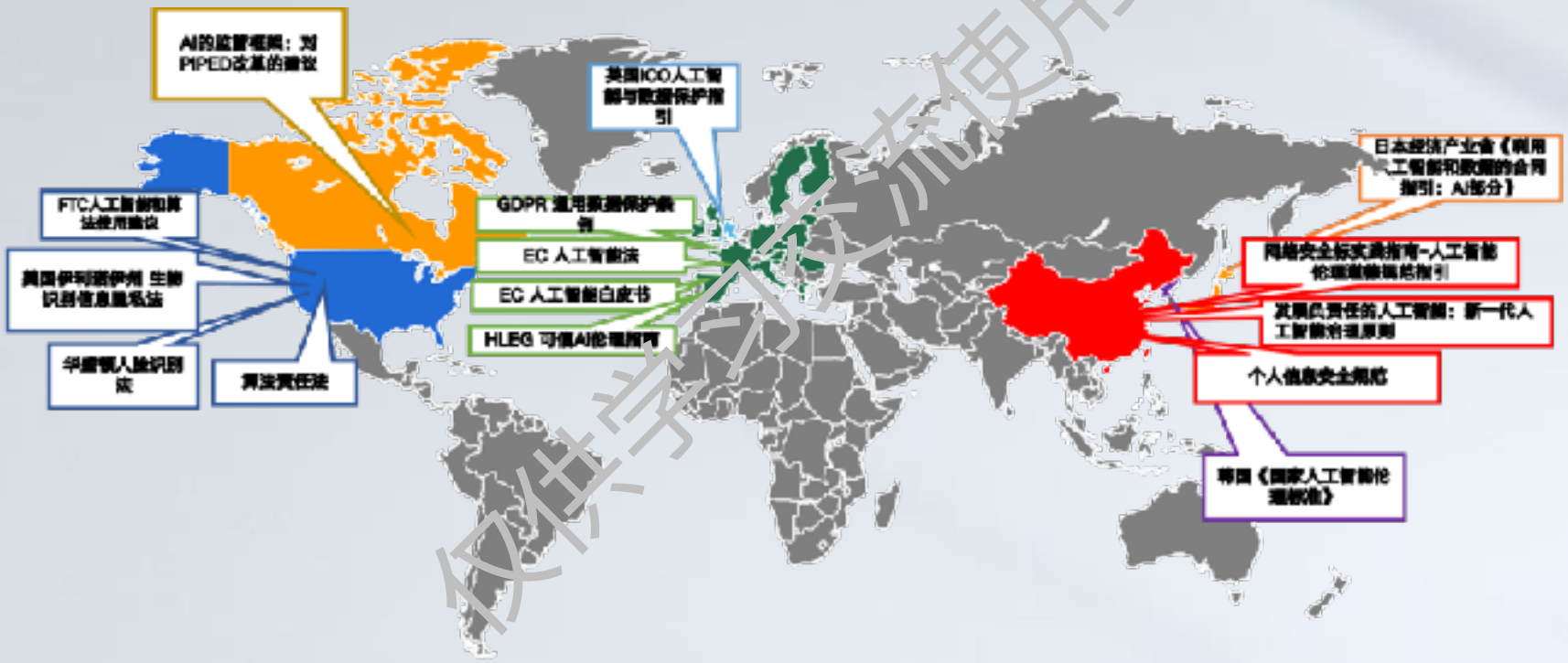
- 模式识别
- 机器学习
- 数据挖掘
- 智能算法

算法透明
可解释
可理解

算法“黑盒”，技术滥用等问题造成[illegible]凸显

- 数据过度索取造成的隐私侵犯
- 敏感数据采集引发安全问题
- 大数据“杀熟”、算法歧视等

AI治理，监管从理论探讨进入立法执法阶段



2021年4月欧盟发布关于《欧总一会和理事会关于制定人工智能统一规则（人工智能法）和修正某些欧盟立法的条例》提案是国际对AI监管动向的标志性事件

企业可信任人工智能框架

小米AI伦理四原则

欧盟人工智能伦理指南

欧盟7项AI伦理要求

- 尊重人类自主权
- 技术鲁棒性&安全性
- 隐私和数据治理
- 公平原则
- 公开透明
- 可追溯
- 社会福祉

小米可信AI四原则

- 安全性 Security and Safety
- 隐私保护 Privacy
- 公平性 Faireness
- 可解释性 Explainability

小米人工智能（AI）治理实践

小米人工智能伦理委员会

统一职责，制定和落地企业内标准规范

支撑业务，行业内探索最佳AI技术实践

风险掌控，及时发现和识别AI伦理风险

引导标准，AI伦理标准规范，法律法规

小米可信AI白皮书

永远相信美好的事情即将发生！