《共享学习系统技术要求》（报批稿）编制说明

一、工作简况

1、起草单位：

浙江蚂蚁小微金融服务集团有限公司、中国联通股份有限公司、中国信息通信研究院、中国电信股份有限公司、阿里巴巴（中国）有限公司、北京大学、中和农信。

2、工作过程：

1）2019 年 5 月，建立标准编制组，进行相关标准研究

2019 年 5 月标准编制组成立后，随即展开广泛研究，摸清国内外的研究动向，为本标准的制定夯实牢固的基础。

2）2019 年 7 月 11 日，标准在 AIIA 中国人工智能产业发展联盟正式立项。

7 月 11 日在上海召开的 AIIA 中国人工智能产业发展联盟 2019 第二次全体大会标准化与推广工作组的会议上，蚂蚁牵头的《共享学习系统技术要求》联盟标准成功立项。参会公司包括微软，IBM，华为，中国联通，中国电信，信通院，高校等。IBM 专家提出希望将联邦学习包含在共享学习的标准中，以便达成业界共识。本次会后会联合联通网络研究院、电信研究院等继续标准撰写及推进工作。

3）2019年10月11日，标准通过意见征集稿。

中国人工智能产业发展联盟 2019 第三次全体大会标准化与推广工作组会议，推动共享学习系统技术要求标准内容意见征集，与会专家对于标准格式上进行了一定意见反馈，计划下次会议提交标准报批稿至总体组报批。

4）2020年1月14日，审议标准报批稿，提交至总体组报批。

中国人工智能产业发展联盟 2019 第四次全体大会标准化与推广工作组会议，会议审议本标准报批稿，同意报批。本次会议后，联盟标准化与推广工作组上报总体组审议，计划 2020 年 2 月对外正式发布。

3、主要起草人：

李克鹏、朴昕阳、周俊、王磊、王力、陈超超、加雄伟、孙明俊、洪澄、吴秉哲、李楠

二、标准编制原则和确定主要内容的论据及解决的主要问题

1、标准意义、目的及必要性

随着大数据技术逐步发展，数据孤岛问题日益突出，数据共享日益重要，但其中存在买卖、泄露和滥用等问题。公众和政府日益重视数据安全和隐私保护，随着欧盟 GDPR(General Data Protection Regulation,通用数据保护条例)法案正式实施及多国效仿，注重数据安全及隐私保护成为发展趋势。如何在满足数据安全、隐私保护和监管合规要求的前提下，设计一个机器学习框架，实现数据的多方协同和授权共享，得到更准确高效的模型和决策，进一步释放数据价值，是当前人工智能技术发展的一个难题。共享学习正是解决这一难题的技术方案。

共享学习是基于数据安全和隐私保护，在多个参与方之间通过共享模型或共享数据来进行机器学习的方案。共享学习方案包括：多方数据在可信执行环境（Trusted Execution Environment）中进行共享和融合学习的 TEE 方案，和多个参与方基于既定协议下通过交换不泄露隐私的非原始数据来进行共享学习的 MPC（Multi Party Computation，多方计算）方案。在 TEE 方案中，各个参与方的通过将本地数据加密后上传到可信执行环境中进行计算来实现数据安全和隐私保护；在 MPC 方案中，各方数据都保留在本地，通过交换不泄露隐私的非原始数据，来实现数据安全和隐私保护。在这两种方案中，数据都无法被平台或者其他参与方窥探，只能按照约定好的行为进行使用。

共享学习正逐渐被各个行业接受，并开始广泛应用于智能信贷、征信、智能医疗等场景中。

但是，目前行业内缺乏共享学习相关的标准，来规范化共享学习的定义、技术架构、技术流程、技术特性、安全要求等，来指导规避共享学习中存在的风险，不利于多参与方之间的数据共享学习及行业发展。因此行业内急需制定共享学习系统技术要求的标准，来指导共享学习系统的设计、开发、测试、使用、运维等，促进多参与方在满足数据安全、隐私保护和监管合规等要求下，实现基于数据协同和授权共享的机器学习，使共享学习更准确高效，进一步释放数据价值。

2、标准内容与适用范围

本标准项目用于制定共享学习系统的技术要求，来规范化共享学习的定义、技术框架及流程、技术特性、安全要求。

本标准项目适用于指导共享学习系统的设计、开发、测试、使用、运维管理等。

本标准项目的技术内容包括：

—共享学习的概述；

—基于可信执行环境的共享学习系统的技术框架、功能组件及技术流程；

—基于安全多方计算的共享学习系统的技术框架、功能组件及技术流程；

—共享学习系统的技术特性要求；

—共享学习系统的安全要求；

—附录：共享学习的使用场景与示例：智能风控及智能营销。

三、主要试验[或验证]情况分析

无。

四、知识产权情况说明

无。

五、产业化情况、推广应用论证和预期达到的经济效果

蚂蚁金服共享学习平台目前已应用于中和农信智能信贷等场景中，基于共享学习推出的“极速贷”产品已经累计放贷 14 亿，为 300 多贫困县提供普惠金融服务。

具体达到如下的业务效果：

（1）多方数据加密融合，实现风控效果最大化：对中和农信、蚂蚁金服等多方数据进行加密，共享学习挖掘出数据的最大价值；

（2）风险预估和线下基本一致：线上自动化审批运行，无重大逾期。

（3）模式/业务流程升级：从线下拜访和人工审批，变成“线上+线下”信贷工厂的模式；

蚂蚁金服共享学习平台在中和农信的智能信贷业务的实施效果如下：

（1）5 分钟极速放贷；

（2）为中和农信带来更快的余额规模增长，8 个月放款 14 亿；

（3）超过 300 个县享受普惠金融服务，为 10 万家以上农村客户提供金融服务。

大数据正在成为经济社会发展新的驱动力，随着公众对于数据安全和隐私保护意识的进一步加强，以及政府及行业的相关标准出台，共享学习必然会在更多行业领域发挥巨大价值。

六、采用国际标准和国外先进标准情况

经广泛调研，目前市场上缺乏关于共享学习的国际国内标准，但存在与共享学习概念类似的联邦学习相关产品级系统和国际标准。

联邦学习概念及产品级系统：2016 年由谷歌最先提出，用于解决安卓手机终端用户在本地更新模型的问题，其设计目标是在保障大数据交换时的信息安全、数据隐私、保证合法合规的前提下，在多参与方或多计算结点之间开展高效率的机器学习。2019 年，谷歌已实现首个产品级的联邦学习系统。

联邦学习相关的标准：IEEE P.3652.1 “Guide for Architectural Framework and Application of Federated Machine Learning”（联邦学习基础架构与应用标准）,主要定义了联邦学习体系结构框架及应用指南。

本立项提案提出的共享学习系统技术要求标准中提出的系统涵盖目前主流通用 MPC 方案的同时，也囊括了 TEE 方案：

——共享学习 MPC 方案：数据不出本地，系统通过交换不泄露隐私的非原始数据的方式，建立共享模型平台，触发协调本地学习模块。

——共享学习 TEE 方案：数据加密出本地，在可信计算环境下，多方数据融合学习输出最优模型。

目前这些技术方案，都缺乏相应的国内标准。

目前共享学习已在 ITU-T、IEEE 国际标准组织正式立项相关技术标准，并与业界同仁一道共建共享学习技术生态。

七、与现行相关法律、法规、规章及相关标准的协调性

本标准符合现有法律法规的要求。

八、重大分歧意见的处理经过和依据

无。

九、标准性质的建议

本标准非强制性要求标准。

十、贯彻标准的要求和措施建议

无。

十一、替代或废止现行相关标准的建议

无。

十二、其它应予说明的事项

无。

AIIA 标准《共享学习系统技术要求》编制工作组

2020-01-17