ICS 35.030
CCS L 80

中华人民共和国国家标准

GB/T 42888—2023

信息安全技术
机器学习算法安全评估规范

Information security technology—
Assessment specification for security of machine learning algorithms

2023-08-06 发布　　2024-03-01 实施

国家市场监督管理总局
国家标准化管理委员会　发布

目　　次

前言 …… Ⅲ
1　范围 …… 1
2　规范性引用文件 …… 1
3　术语和定义 …… 1
4　概述 …… 2
4.1　安全原则 …… 2
4.2　安全要求分级 …… 2
5　机器学习算法技术安全要求和评估方法 …… 2
5.1　安全要求 …… 2
5.2　评估方法 …… 5
6　机器学习算法服务安全要求和评估方法 …… 9
6.1　安全要求 …… 9
6.2　评估方法 …… 9
7　机器学习算法安全评估流程 …… 11
7.1　流程要求 …… 11
7.2　评估准备 …… 11
7.3　评估方案 …… 11
7.4　评估执行 …… 12
7.5　评估结论 …… 12
7.6　评估报告 …… 12
附录 A（规范性）　算法推荐服务安全要求 …… 14
附录 B（规范性）　算法推荐服务评估方法 …… 21
参考文献 …… 29

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国信息安全标准化技术委员会(SAC/TC 260)提出并归口。

本文件起草单位：北京赛西科技发展有限责任公司、中国科学院计算技术研究所、清华大学、国家计算机网络应急技术处理协调中心、上海商汤智能科技有限公司、北京瑞莱智慧科技有限公司、阿里巴巴(中国)有限公司、中国科学院信息工程研究所、中国信息通信研究院、中国电子科技集团公司第十五研究所、国家信息技术安全研究中心、广州大学、北京大学、华东师范大学、北京航空航天大学、华为技术有限公司、北京旷视科技有限公司、北京百度网讯科技有限公司、深圳市腾讯计算机系统有限公司、浙江大学、北京奇虎科技有限公司、北京小桔科技有限公司、安徽工程大学、北京智者天下科技有限公司、北京交通大学、浙江工业大学、上海工业控制安全创新科技有限公司、中国人民公安大学、深圳市大数据研究院、北京计算机技术及应用研究所、中国科学院自动化研究所、上海燧原科技有限公司、烽台科技(北京)有限公司、中国电子技术标准化研究院。

本文件主要起草人：上官晓丽、郝春亮、许晓耕、胡影、陈钟、沈华伟、蒋慧、梅敬青、张宇光、彭骏涛、郭岩、李鹏霄、艾政阳、赵芸伟、韩晗、刘明、尹芷仪、庞亮、王晓诗、刘总真、周熙、孟国柱、景慧昀、张琳琳、朱纯超、霍珊珊、刘健、刘赫、苏航、金涛、刘吉强、任奎、张旭东、成瑾、朱红儒、杨韬、李钦、刘祥龙、王义飞、吴庚、赫然、顾钊铨、李实、曹晓琦、严敏瑞、付英波、郭颖、孙空军、唐家渝、刘曦泽、王哲麟、任璐、徐永太、张屹、秦湛、安泽亮、徐雨晴、李雪、李大海、徐光侠、包沉浮、郭建领、宣琦、张世天、赵涌鑫、王姣、王秉政、芦天亮、吴保元、韩磊、张雨桐、彭泉。

信息安全技术
机器学习算法安全评估规范

1 范围

本文件规定了机器学习算法技术和服务的安全要求和评估方法，以及机器学习算法安全评估流程。

本文件适用于指导机器学习算法提供者保障机器学习算法生存周期安全以及开展机器学习算法安全评估，也可为监管评估提供参考。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

机器学习算法 machine learning algorithm

功能单元通过学习新知识技能或整理已有知识技能以改进其性能的算法。

3.2

机器学习算法提供者 machine learning algorithm provider

利用机器学习算法实现特定功能的组织。

注：本文件中简称算法提供者，包括算法技术提供者和算法服务提供者。算法技术提供者是指算法技术的开发和提供方，算法服务提供者是指使用应用算法技术的服务提供方。

3.3

算法推荐服务 algorithmic recommendation service

互联网信息服务算法推荐 internet information service of algorithmic recommendation

应用算法推荐技术提供信息的服务。

注 1：应用算法推荐技术是指利用机器学习算法实现生成合成类、个性化推送类、排序精选类、检索过滤类、调度决策类等算法技术，向用户提供信息的活动。

注 2：本文件将生成合成类、个性化推送类、排序精选类、检索过滤类、调度决策类等算法统称为五类算法。

3.4

算法生存周期 algorithm lifecycle

机器学习算法从设计到退役的演进过程。

注 1：算法生存周期包括设计开发、验证确认、部署运行、维护升级、退役下线。

注 2：一般算法服务处于部署运行阶段。