

项目简述

作品名称：个人隐私信息泄露风险评估系统

英文名称：Personal privacy information leakage risk assessment system

作品介绍：

随着大数据技术和人工智能的迅速发展，数据的创建量呈指数级增长。数据的应用场景也愈加广泛，这使得用户不再享有对数据的绝对控制权，数据的安全性受到威胁，从而频发个人隐私泄露事件。数据安全是数字经济发展的压舱石，是国家安全的重要组成部分，数据安全已成为事关国家安全和经济社会发展的重大议题。为了能够对数据内容的潜在隐私风险做出评估，设计了个人隐私信息泄露风险评估系统，针对多源数据进行个人信息层面的泄露风险评估，用于保证数据安全，此前提下充分发掘数据要素价值及推动生产力发展。

系统创新性的依据去年国家发布实施的《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》彰显了我国高度重视数据安全保护工作的相关法规，以法为基准指定风险泄露标准，帮助发现数据存在的风险隐患，能够有效预防信息泄露情况的发生。此系统面向个人隐私保护领域，符合A-ST专项指南范围。针对现有的检测工具（例如Firefox monitor）依据多样但不明确，且检测的数据部分来自存储着已泄露数据的数据库，并不能反映个人隐私信息的泄漏情况的问题，系统收集指定目标的多源数据，可以根据不同的数据隐私需要，匹配不同事物数据项，并调整不同数据项的风险等级。系统相较于其他检测方案更加符合国家法律法规，例如在新冠肺炎流调数据个人隐私泄露检测的应用场景中，系统可以在数据发布前做出风险评估，数据发布方可以根据评估结果对数据做出适当修正，从而减少隐私泄露情况的发生。

系统针对重要待检测目标，从数据源中利用深度学习、文件解析、分布式爬虫等技术获取数据，提取可能伴随隐私泄露风险的数据并归类。随后通过检测模型对收集的数据进行泄露风险评估，模型对多条数据类型进行建模，结合基于人工经验的固定规则和基于机器学习的关联规则算法推断，实现对风险等级的科学评估。风险评估模型以现有的法律法规为依据，根据不同数据之间的关联程度最终给出评估结果，也可以根据用户需求进行调整，满足不同源数据存在的差异化风险需求。并根据实际网络部署情况，分布式部署爬虫和检测程序，在一定程度上提高了系统运行的效率和稳定性，帮助解决了数据孤岛带来了检测问题，加强了数据检测过程中的安全性。

