 DOI: https://doi.org/10.12349/iser.v7i7.11397

Research on Enhancing the Efficiency of Cybersecurity Level Protection Assessment Driven by Artificial Intelligence

Weizhe Jing Rongze Zhang* Mengxiao Yuan Zhaosheng Zhan Yan Liu

Jinan Sanze Information Security Testing Co., Ltd., Jinan, Shandong, 250101, China

Abstract

This paper investigates AI-driven cybersecurity level protection assessment. The introduction section outlines the research background, significance, and objectives. The conceptual overview defines key related concepts and examines the current application status of artificial intelligence in this field. The theoretical analysis explores how critical technologies such as AI algorithms and data processing enhance assessment intelligence based on technical principles. The empirical design proposes an AI-powered assessment model with specified methods for data collection, feature extraction, and validation. Experimental results demonstrate AI-based approaches' superior performance through case comparisons. The conclusions and discussions summarize findings, identify existing limitations, and propose future improvement directions. Results indicate that artificial intelligence significantly enhances the intelligence level of cybersecurity level protection assessments and holds promising application prospects.

Keywords

Artificial Intelligence; tiered protection; evaluation model

人工智能驱动的网络安全等级保护测评效能提升研究

井维哲 张荣泽* 袁梦晓 展昭生 刘岩

济南三泽信息安全测评有限公司，中国 · 山东 济南 250101

摘 要

本文对人工智能驱动的网络安全等级保护测评进行研究，引言部分说明了研究背景和意义，并且确定了研究目标。概念简述部分界定了相关主要概念，并分析了人工智能在该领域的应用现状。理论分析部分从技术原理出发，分析了人工智能算法和数据处理等重要技术对测评智能化的支持作用。实证研究设计提出了一个利用人工智能的测评模型，并且明确了数据采集、特征提取和验证方法。实验结果与分析通过案例对比，证明了人工智能方法在测评性能上具有优势。结论和讨论部分总结了研究成果，指出了当前研究存在的不足，并且提出了未来进行优化的方向。研究结果表明，人工智能可以提高网络安全等级保护测评的智能化水平，并且拥有良好的应用前景。

关键词

人工智能；等级保护；测评模型

1 引言

在当今世界，网络快速发展，信息安全问题变得越来越关键。有相关研究将 AI 技术进行结合，提高风险评估的动态性和可解释性，给网络安全等级保护测评带来新的思路[1]。传统的等级保护测评尽管规范，然而常常耗时费力，并且严重依赖人工经验，容易出现疏漏情况。目前我国等级保护 2.0 标准已经全面实施，面对云计算、大数据等复杂的新型网络环境，传统的人工测评手段在效率与准确性方面显得有些不足，难以满足海量数据的处理需求。本文要研究人工智能驱动下网络安全等级保护测评效能的提高办法，尝试建立一种更为智能的测评模式。引入智能化手段后，既能减轻测评人员的工作负担，又能挖掘出传统方法难以察觉的隐蔽威胁。虽然当前 AI 在安全领域的应用已初步形成规模，然而针对等级保护测评的深度融合仍有待探索。Gong 等学者为解决等级保护测评中主观性较强和指标量化不够明确的问题，提出了将等保 2.0 与系统安全评估相结合的方法[2]。本文期望通过理论分析与模型建立，为相关领域的实践给予有价值的参考。

【作者简介】井维哲（1998–），男，中国山东东平人，本科，初级职称，从事网络安全研究。

【通讯作者】张荣泽（1982–），男，中国黑龙江黑河人，本科，高级，从事网络安全研究。

2 概念简述

人工智能技术能够模拟人类智能行为，在数据处理、模式识别领域展现出强大优势，为网络安全等级保护测评的

革新创造了机会。网络安全等级保护制度要求针对信息系统实施分等级的安全保护与测评工作，传统测评方法主要依靠人工访谈、核查和测试，存在工作量大、主观性强且难以应对海量数据等问题。研究说明，人工智能技术可以提高海量数据的处理能力，并且给网络安全等级保护测评带来了新的解决方式 [3]。当下，借助机器学习、自然语言处理等人工智能技术来提高测评效能已成为研究热点。智能自动化技术在提高网络安全测评效率上具有很大潜力 [4]。

3 理论分析

人工智能技术引入了机器学习、深度学习、自然语言处理等先进算法，为网络安全等级保护测评给予了有力的理论支持，促使测评模式从传统人工朝着智能化转变。有相关研究运用 AI 驱动的自动化方式，提高网络安全测评的效率和透明度，给等级保护测评带来新的技术方法 [5]。在数据采集与处理阶段，智能算法能够高效地清洗海量网络资产日志与流量数据并提取特征，。模式识别技术的应用让测评系统能够精准捕捉网络流量中的异常特征，自动识别潜在的安全漏洞与威胁行为，提高测评的准确性与效率。特别是在自动化测评报告生成与审查方面，基于大语言模型与知识图谱的技术已取得一定成效，像天帷智审系统的实践显示，AI 能有效解决人工审查标准不统一的问题。在应用生成式 AI 技术时，需要关注其可解释性风险。要严格依照《生成式人工智能服务管理暂行办法》的规定，落实算法备案、安全评估工作。如此一来，既能够提高效能，又能保证测评结果的客观性和可靠性。该系统通过建立知识库和自动化流程，提高测评效率、合规性判断的准确性 [6]。

表 1 理论分析表格

年份	类别	核心内容	关键参数 / 效能表现	数据来源
2021	合规标准	《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》正式施行	构建人工智能在等保测评数据处理环节的基础法律框架，明确各主体数据安全责任边界	国家市场监督管理总局
2023	合规标准	《生成式人工智能服务管理暂行办法》发布实施	规范等保测评中大语言模型等生成式 AI 技术的应用流程，明确算法备案与安全评估要求	求是网
2024	AI 技术应用	中检天帷推出国内首个 AI 核心等保测评报告审查工具（天帷智审 1.0）并投入商用	实现等保测评报告基础合规项自动化审查，解决传统人工审查标准不一、效率偏低问题	经济日报
2025	AI 技术应用	天帷智审完成 2.0 版本升级，全面适配公安部 2025 版等保测评报告模板	基于 5 万条标准条款知识图谱、4000 万条历史报告数据训练，将单份报告审核时长从数小时压缩至 10~30 分钟，综合效率提升超 300%，已落地应用于数十家测评机构及公安网安部门	澎湃新闻、经济日报

4 实证研究设计

实证研究部分要建立并验证人工智能驱动的网络安全等级保护测评模型。该方法利用有监督学习来提高测评效率，给有相关研究模型建立提供了主要参考 [7]。研究确定了包括数据采集、特征提取、模型建立和验证的全流程设计方案。有相关研究模拟攻击者角色进行安全评估，为提高等级保护测评效率提供了新思路 [8]。数据采集阶段着重获取网络设备日志、主机配置文件还有安全审计记录，借助大数据技术进行清洗与标准化处理。特征提取环节运用深度学习算法，从大量非结构化数据里自动识别关键安全控制点，形成高维特征向量。模型建立阶段整合机器学习与推理引擎，针对特定安全等级要求进行有监督训练，提高对未知威胁的检测能力。根据现有的数据分析，引入 AI 技术之后，单份测评报告的审核时长能够缩短到 10 至 30 分钟，综合测评效率的提高幅度达到了 400%，漏洞误报过滤率更是超过了 90%。通过和传统人工测评在耗时、人力投入、准确率方面的对比，本研究会对人工智能技术在提高网络安全等级保护测评效能方面的实际应用价值进行定量评估。

表 2 实证研究设计表格

测评效能指标	传统人工测评	AI 驱动测评	单位	数据来源
单份测评报告审核时长	120~180	10~30	分钟	经济日报 2026
政务云平台测评总投入	15	2	人天	CSDN 博客 2026
漏洞误报过滤率	40	≥90	%	CSDN 博客 2026
综合测评效率提升幅度	100（基准）	400	%	中国质量报 2026
金融机构复评时间节省	0（基准）	65	%	CSDN 博客 2026
资产梳理完成时长（百台服务器规模）	3~5	1	小时	CSDN 博客 2026

5 实验结果与分析

实际案例、实验数据能够表明，人工智能技术于网络安全等级保护测评里展现出了应用成效。乔路的研究说明，渗透测试方法可以提高网络安全等级保护测评的准确性和可靠性 [9]。传统人工审核一份报告一般要花费 120 至 240 分

钟，而且问题检出准确率仅仅为 62%，效率比较低并且容易遗漏安全隐患。与之相比，引入 AI 技术之后测评效能有了明显提高。就像中检天帷天帷智审 2.0 系统把单份审核时长缩短到 10 至 30 分钟，效率提高 4 倍，问题检出准确率高达 96%。安恒等保测评报告分析智能体更是让效率提高倍数达到 6.0，准确率维持在 94%。相关研究利用 LoRA 对大模型进行微调，以提高等保报告审核的准确性，这给有相关研究在智能体效率和准确率上的优化提供了方法参考 [10]。这些数据说明，人工智能不但大幅缩短了测评时间，还借助智能算法提高漏洞识别的精准度。

表 3 实验结果与分析表格

测评方法类型	单份等保测评报告审核时长	效率提升倍数（相对传统方法）	问题检出准确率	应用落地单位	数据来源
传统人工审核	120-240 分钟	1.0（基准）	62%	全行业各级测评机构	安恒信息 2025 行业分析
中检天帷天帷智审 2.0（AI）	10-30 分钟	4.0	96%	交通运输信息安全中心、华测检测等	经济参考网 2026
安恒等保测评报告分析智能体（AI）	20-48 分钟	6.0	94%	全行业测评机构开放试用	安恒信息 2025 官方发布

6 结论与讨论

本研究对人工智能技术在网络安全等级保护测评里的应用做了分析，证实其提高测评效能是可行的。当前的研究展现出 AI 技术具备相当可观的潜力，然而在实际应用进程当中，依旧面临着诸如模型泛化能力不足、数据隐私出现泄露等一系列挑战。未来进行优化时应当着重关注提高模型的鲁棒性，把联邦学习与区块链技术结合起来建立可信的数据共享、存证模式。借助持续不断地对算法架构加以迭代，能够建立起更为精准且高效的智能化测评模式，促使网络安全等级保护工作朝着更深层次发展。

参考文献

[1] ISLAM S, SARDAR B, KALOGERAKI E M, et al. Hybrid AI-Based dynamic risk assessment framework with explainable AI practices for composite product cybersecurity certification[J]. Int J Inf Secur, 2026, 25(2).

[2] GONG G J, YUAN Q L, YANG J X, et al. Safety Index, Evaluation Model, and Comprehensive Evaluation Method of Power Information System under Classified Protection 2.0[J]. Energies, 2023, 16(18).

[3] ZHAO W B, WANG P. AI-Driven Cybersecurity in IoT-Based Systems[J]. Sensors, 2025, 25(23).

[4] NEVES A, ARAUJO V M. Smart Automation for Enhancing Cybersecurity[J]. Foresight STI Gov, 2023, 17(1).

[5] ABBA S S, OLANIYI O M, OLADOYINBO O B, et al. AI-Driven Automation of Cybersecurity Certification Processes: Evaluating Efficiency, Transparency and Risk Mitigation in Digital Governance Systems[J]. J Eng Res Rep, 2025, 27(12).

[6] 张雨,赵慧,杨茂深,刘昊.基于推理模型的等级保护2.0自动化测评系统[J].计算机工程与设计,2023,44(11):3298-3304.

[7] 杜晓杰,梁承东,彭世强,成嘉轩.一种基于机器学习的等级保护自动测评方法[J].电子质量,2023(1):83-86.

[8] 陈宇飞.面向网络安全第三级等级保护的自动化测评系统设计与实现[D].成都:西南交通大学,2023.

[9] 乔路.渗透测试在网络安全等级保护测评中的应用研究[J].科技创新与应用,2022,12(14):159-162.

[10] 吴建华,刘振宇,曾睿,王文瑄,易勇,王世轶.基于LoRA微调大模型在网络安全等级保护测评报告质量检测中的应用研究[J].数据与计算发展前沿,2025,7(6):111-123.