NTNU | Kunnskap for ei betre verd

AI-Assisted Assurance Profile Creation for System Security Assurance

Muhammad Mudassar Yamin, Shao-Fang Wen, and Basel Katt
ESORICS(SecAssure) 20th August 2024, Bydgoszcz, Poland

Whoami

Professional Title

- Associate Professor NTNU (Norwegian University of Science and Technology)

Expertise

- Cybersecurity: System security, penetration testing, intrusion detection
- AI & Machine Learning: Applications in cybersecurity
- Cyber Range Training: Development and implementation

Certifications

- OSCE, OSCP, LPT-MASTER
- CEH, ,CPTE, CHFI, CPTE, CISSO, CBP, CCNA-CyebrOPS, KLSE

Awards

- INTERPOL Medal for Innovation
- 50+ HoF (Microsoft,TrendMIcro,US DOD)

Key Projects

- Open Cyber Range, CR14 NATO CCDCOE
- Towards a common ECSC roadmap, ENISA

Publications

- "Cyber ranges and security testbeds: Scenarios, functions, tools and architecture." Computers & Security 88 (2020): 101636.
- "Weaponized AI for cyber attacks" (Journal of Information Security and Applications 57 (2021): 102722)
- "Modeling and executing cyber security exercise scenarios in cyber ranges." Computers & Security 116 (2022): 102635.

Contact

- Email: Muhammad.m.yamin@ntnu.no
- LinkedIn:https://www.linkedin.com/in/mudassaryamin/

Outline

- Introduction and Background
- Research Objectives
- Methodology
- System Design
- Case Study and Results
- Comparative Analysis
- Limitations and Challenges
- Future Work
- Conclusions

Introduction and Background

System Security Assurance (SSA) is critical in the evolving digital threat landscape

SSA helps organizations[1]:

- Proactively identify and address potential security vulnerabilities
- Meet regulatory requirements
- Maintain positive reputation with stakeholders

Assurance Profiles (APs) are structured frameworks outlining:

- Specific security objectives
- Requirements for achieving defined SSA levels

Traditional AP creation challenges:

- Labor-intensive process
- Requires extensive domain knowledge
- Difficulty in maintaining consistency
- Keeping up with rapidly evolving security threats

Research Objectives

Develop an AI-assisted method for creating security assurance profiles

Leverage Large Language Models (LLMs) to automate and enhance AP generation

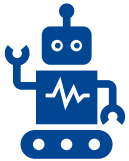
Implement Retrieval-Augmented Generation (RAG) techniques for comprehensive compliance

Compare AI-generated profiles with human-created profiles

Identify advantages and limitations of the AI-assisted approach

Methodology

Applied experimentation approach [2]



Retrieval-Augmented Generation (RAG) [3]:

Enhances LLM output with external authoritative knowledge base

Improves relevance and accuracy without model retraining

Mitigates LLM hallucination by providing authoritative context

Case study:

Comparative analysis of AI-generated vs. human-created profiles

Qualitative evaluation of comprehensiveness, accuracy, and compliance

System design

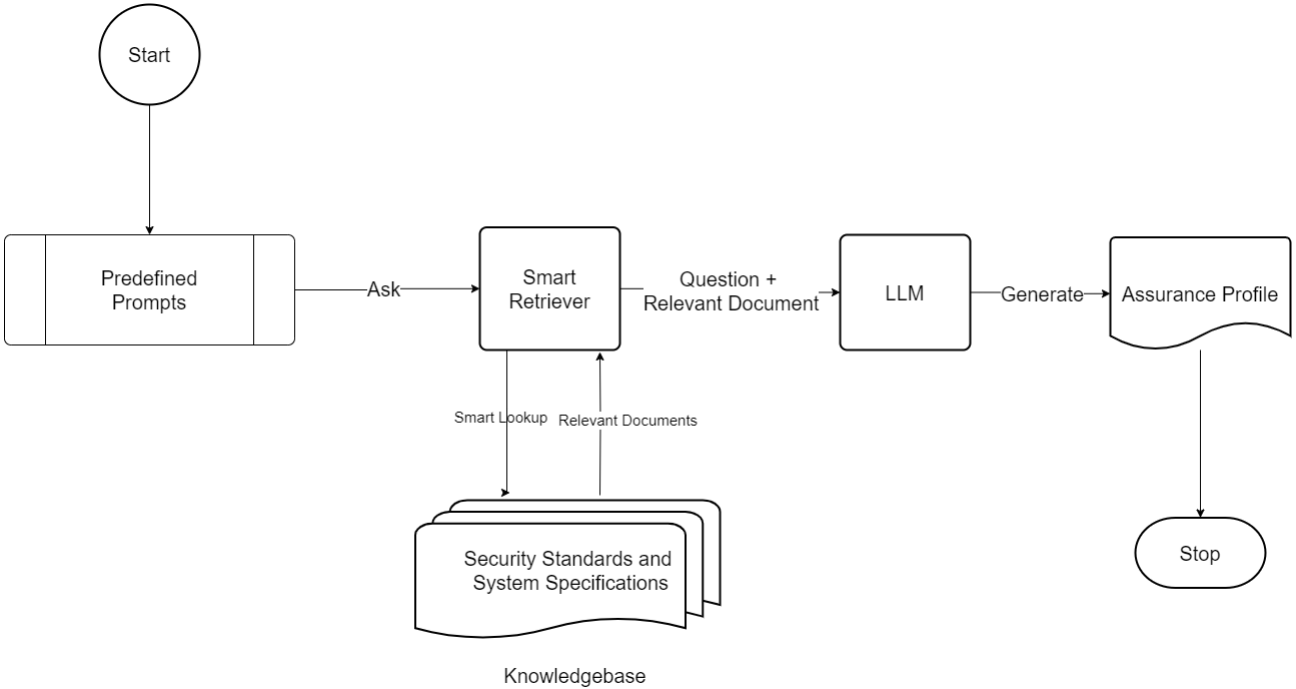


Fig. 1. Illustration of the system design

System design(2)

Components:

Input: System details, standards documents (OWASP ASVS [4], GDPR [5]), system requirements

Local LLM implementation:

- Meta Llama 2 model (70B parameters) [6]
- Ollama framework

Hardware: Intel Xeon processor (32 cores, 64GB RAM)

Output: Generated Assurance Profile

Security considerations:

Local operation ensures confidentiality of system being tested

Avoids sharing sensitive information over the internet

Case Study Setup

Subject: Medical facility system with IoT and edge communication components

Standards considered:

- OWASP Application Security Verification Standard (ASVS) [4]
- General Data Protection Regulation (GDPR) [5]

Evaluation criteria:

- Comprehensive coverage of security requirements
- Granularity of security-related questions
- Inclusion of specific vulnerabilities
- Adherence to compliance standards

AI-Generated Profile

Security Requirement	Security Questions
1. Access Control	
1.1 Identity Management	- Are individuals identified by a single, unique identity within the system? - What mechanisms are in place to manage user identities securely? - How is access granted and revoked?
1.2 Authentication	- What methods are used for user authentication? - Are multi-factor authentication (MFA) methods implemented? - How is the authentication data stored and protected?
1.3 Authorization	- How are user roles and permissions defined and enforced? - Are there mechanisms to prevent unauthorized access to sensitive data and functions? - How is access monitored and audited?
2. Data Protection	
2.1 Data Encryption	- Is data encrypted at rest and in transit? - What encryption standards are used? - How are encryption keys managed and stored?
2.2 Data Anonymization	- Are there mechanisms to anonymize sensitive data where necessary? - How is anonymized data verified to ensure it cannot be re-identified?
2.3 Data Integrity	- What measures are in place to ensure data integrity during transmission and storage? - How is data integrity verified and validated?
3. Compliance with Standards	
3.1 GDPR Compliance	- How does the system ensure compliance with GDPR regulations? - What processes are in place to manage and document data subject consent? - How is personal data handled in accordance with GDPR requirements?

Continued on next page

Security Requirement	Security Questions
3.2 Normen Compliance	- Does the system meet the "Normen" requirements for patient and privacy protection? - What specific measures have been implemented to comply with these standards?
4. System Integrity	
4.1 Technical Descriptions	- Is there a comprehensive technical description of all system elements, including software and hardware components? - How does the system architecture support overall security and integrity? - Are there any identified vulnerabilities in the technical components?
4.2 Information Flow	- Is there a detailed chart documenting the information flow and all involved components? - How is data flow monitored and controlled to ensure security? - Are there checkpoints or controls in place to detect and mitigate security breaches?
5. Vulnerability Management	
5.1 Vulnerability Identification	- How are vulnerabilities identified within the system? - What tools and techniques are used for vulnerability scanning?
5.2 Vulnerability Mitigation	- What processes are in place for vulnerability remediation? - How are patches and updates managed and deployed?
5.3 Incident Response	- What is the incident response plan for handling security breaches? - How are incidents reported, documented, and resolved?
6. Specific Vulnerabilities	
6.1 Broken Access Control	- VUC-ACC-01: How is Directory Traversal File Include prevented and monitored? - VUC-ACC-02: What measures are in place to prevent Bypassing Authorization Schema? - VUC-ACC-03: How is Privilege Escalation detected and mitigated? - VUC-ACC-04: What protections exist against Insecure Direct Object References?
6.2 Cryptographic Failures	- VUC-CRY-01: How is Weak Transport Layer Security addressed and improved? - What encryption protocols are used to prevent cryptographic failures?

Table 1: Generated assurance profile

AI-Generated Profile(2)

Comprehensive coverage of security domains:

- Access Control (Identity Management, Authentication, Authorization)
- Data Protection (Encryption, Anonymization, Integrity)
- Compliance with Standards (GDPR, Normen)
- System Integrity (Technical Descriptions, Information Flow)
- Vulnerability Management (Identification, Mitigation, Incident Response)

Specific vulnerabilities addressed (e.g., Broken Access Control, Cryptographic Failures)

Detailed security questions for each requirement

Alignment with ASVS levels (1, 2, and 3)

Comparative Analysis

AI-generated profile advantages:

- Broader view of general security requirements
- More comprehensive and detailed security questions
- Holistic view of system architecture and information flow
- Detailed compliance standard questions (e.g., GDPR)
- Inclusion of incident response and mitigation strategies

Human-created profile limitations:

- Heavy focus on specific vulnerabilities, lacking broader view
- Incomplete coverage of general security requirements
- Missing details on compliance standards
- Lack of information on incident response and mitigation

Limitations and Challenges

Context Length:	• Limited to analyzing ~1,600 pages of documentation • Potential issue for systems with extensive documentation
Iterative Process:	• Requires specific prompts for certain requirements • Potential for human bias in prompt creation
Hallucination:	• LLMs can generate false information • Mitigation through RAG and appropriate controls
Validation Approach:	• Current qualitative assessment may introduce subjectivity • Need for more systematic, quantitative validation methods

Future Work

Multi-system testing:

- Apply the approach to diverse systems for more robust results

Automated vulnerability test case generation:

- Extend the system to generate and validate test cases

Advanced RAG techniques:

- Explore multi-query RAG for automated prompt generation

Enhanced reasoning:

- Investigate "chain of thoughts" technique for improved argumentation

Quantitative evaluation methods:

- Develop metrics for systematic comparison of AI and human-generated profiles

Conclusions

AI-assisted AP generation shows significant promise:

- Matches or exceeds human-generated profiles in comprehensiveness
- Reduces potential for human error and inconsistency
- Efficiently handles complex compliance requirements

Potential impact:

- Streamlined SSA processes
- Improved efficiency in cybersecurity practices
- Enhanced ability to adapt to evolving security landscapes

Challenges remain in validation and fine-tuning of AI-generated content

Further research needed to fully realize the potential of AI in SSA

Acknowledgments and References

ASCERT project

Norwegian Research Council (Grant number: 329062)

Selected References:

1. Katt, B., & Prasher, N. (2019). Quantitative security assurance.
2. Edgar, T. W., & Manz, D. O. (2017). Research methods for cyber security.
3. Lewis, P., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks.
4. Shaees, S. (2022). Software Security Assessment and Analysis using OWASP ASVS.
5. Voigt, P., & Von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR).
6. Touvron, H., et al. (2023). Llama 2: Open foundation and fine-tuned chat models.

Questions are guaranteed in life, answers aren't👽

Muhammad.m.yamin@ntnu.no