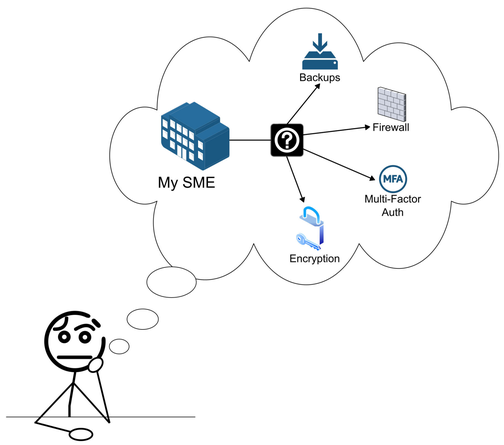
Towards an AI-based Security Consultant for SMEs

Zurich University of Applied Sciences (ZHAW) | Institute of Computer Science | Information Security Group (ISE)

Prof. Dr. Ariane Trammell, Prof. Dr. Marc Rennhard, Maurice Amon, Louie Wolf

Motivation

SMEs in Switzerland are increasingly affected by cyberattacks [1]. A survey of small business managers by gfs-zürich [2] found that 11% of companies in Switzerland experienced cyberattacks require considerable effort to repair the damage.
Established cybersecurity frameworks such as NIST, ISO 27001, and CIS provide guidelines to improve information security. However, these frameworks are intentionally abstract to enable cross-industry application. The lack of company-specific risk references and concrete recommendations makes implementation difficult for small and medium-sized enterprises [3].



This research project aims to automate basic tasks of security consultants for Swiss SMEs using a data-driven approach. The architecture collects security-relevant company information to create a Digital Twin of the Organization (DTO). This digital replica serves as a model where an LLM is tailored to perform a company-specific security assessment. Instructions from tools and remedial measures are analyzed via a Retrieval-Augmented Generation (RAG) system to generate concrete implementation steps. The goal is to equip Swiss SMEs with a tool to identify major risks and obtain specific actions to mitigate them.

Research Questions

- **RQ1:** What information must be included in a Digital Twin to comprehensively cover all relevant security areas and controls, and how can this information be collected with a high degree of automation?
- **RQ2:** How can we identify the correct threats and security controls given the information about a company, in a completely automated fashion?
- **RQ3:** How can we generate easy-to-understand guidelines that provide SMEs with all required information to implement the right security controls?

Call to Action

Get Involved – Partner with Us!

We are looking for:

Pilot companies interested in taking advantage of an automated threat detection system

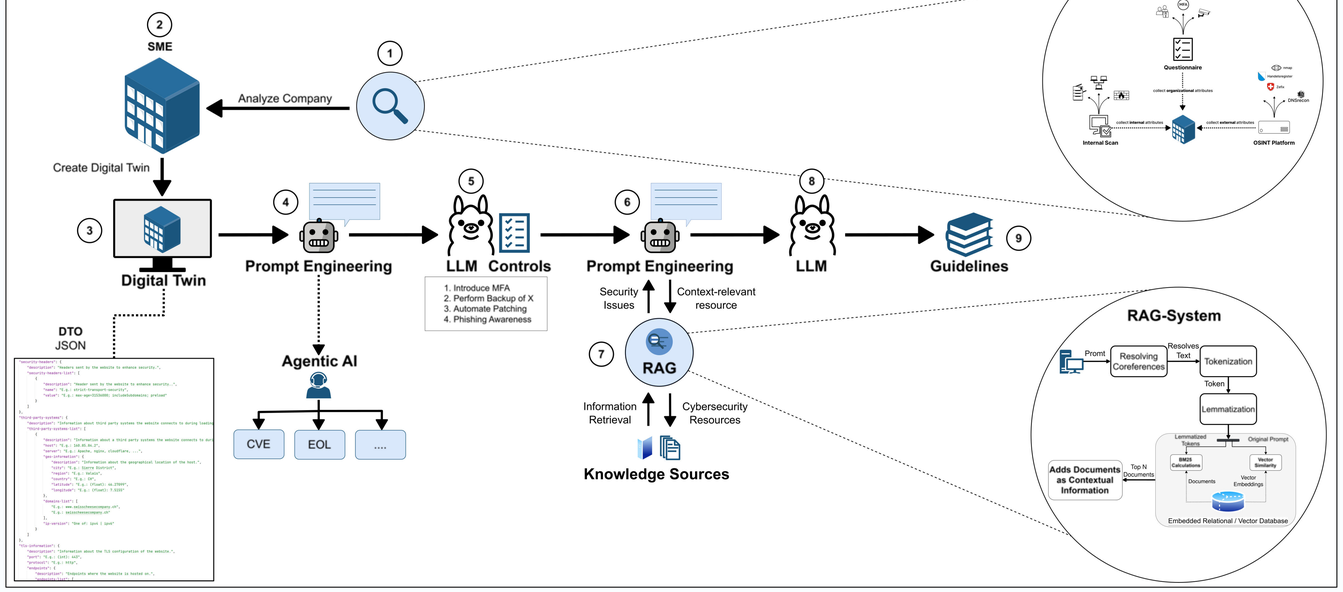
Industry partners interested in building a business case backed by applied research

Be part of an innovative collaboration – and help advance the cybersecurity of SMEs through practical solutions.

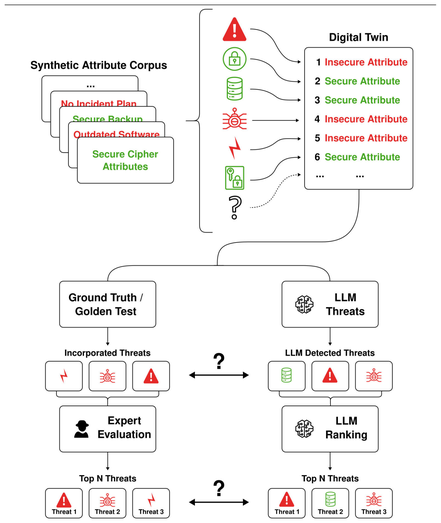
👉 **Interested?** Talk to us today, leave your business card, or reach out via email: **ariane.trammell@zhaw.ch**

Architecture

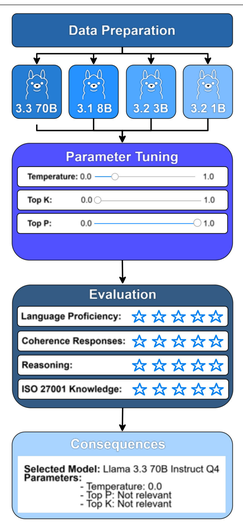
1. DT Data Collector: Uses various tools and APIs to gather as much security-relevant data regarding a certain company as possible.
2. Based on the collected data, the system then creates a DTO of the company.
3. The data contained in the DTO is structured in a JSON-File.
4. Leveraging a mapping between the security controls and their corresponding paths within the DTO's JSON structure, the system constructs a meta-prompt that formulates a query regarding the fulfillment of a specific control. This meta-prompt also incorporates the relevant DTO paths required to assess whether the company has implemented the control in question.
5. The meta-prompt will be forwarded to the Nginx server which redirects the request to the Ollama-backend.
6. Once the threats have been identified, the system formulates another meta-prompt designed to generate a set of mitigation instructions tailored to the detected risks. These instructions are grounded in the company's specific context. The RAG system may also supply supplementary materials to enhance accuracy and reduce the risk of hallucinations.
7. The RAG-System is in charge of matching the input-query with the Top N resources.
8. The meta-prompt, that has been enhanced with contextual information from the RAG-System, will be forwarded to the Nginx server which redirects the request to the Ollama-backend.
9. The last step in the systems pipeline: The guidelines and instruction manuals for the user.



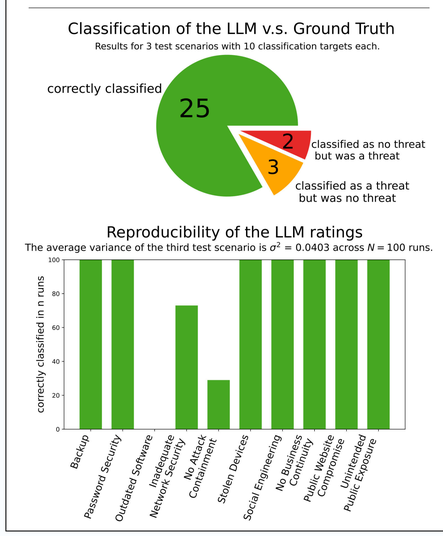
Evaluation: Testsuite



Model Selection

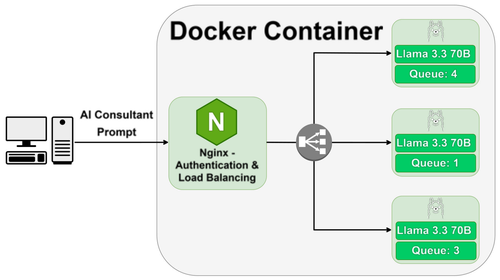


Interim Results

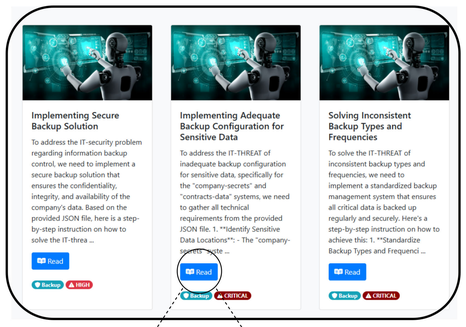


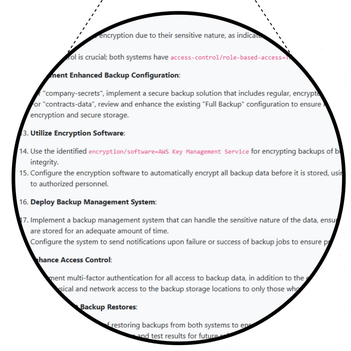
Implementation

LLM Deployment



Example Analysis AI-Consultant





References

[1] Bundesamt für Statistik. Polizeiliche Kriminalstatistik (PKS). 34847183.
[2] Karin Mändli Lerch und Mara Huber. Digitalisierung und Cybersicherheit in Schweizer KMU.
[3] Ariane Trammell u. a. „Towards Automated Information Security Governance"

Acknowledgements

This work was partly done within the project CYREN ZH, which is funded by the Digitalization Initiative of the Zurich Higher Education Institutions (DIZH).