Universität Stuttgart
Institute of Industrial Automation
and Software Engineering

Vector Cybersecurity Symposium 2024

Generative AI for Cybersecurity

How to use AI to automate Threat Assessment and Risk Analysis in Cybersecurity?

Speaker: Michael Weyrich
Co-authors: Christof Ebert und Max Beck

Source: Vector Informatik

Generative AI for Cybersecurity

How to use AI to automate threat assessment and risk analysis

Abstract:

The presentation explains a GenAI based methodology for cybersecurity risk management in automotive according to ISO21434. A case study based on large language models proves the concept and shows how the Robo-Test framework can be used for automated processing in risk, failure, and impact analysis.

Focus on:

- Application of GenAI to cyber security risk management in automotive (ISO21434)
- Proof of Concept for Fine Tuning and Indexing with Large Language Model
- The Robo-Test framework for processing in risk, failure and effect analysis

About the Speaker:

Professor Weyrich is head of the Institute for Automation and Software Engineering at the University of Stuttgart. He researches and teaches in the area of industrial information technology for automation with special focus on autonomous systems and automation safety and security.

GenAI for Cyber Security Risk Management in Automotive

How could GenAI assistance act as a sparring partner in Risk Management

A potential cure

A pre-trained, fine-tuned LLM that could facilitate the discovery, aggregation and generation of plans and reports based on patterns in historical data from internal and external information sources.

Source: Vector Informatik

Contents

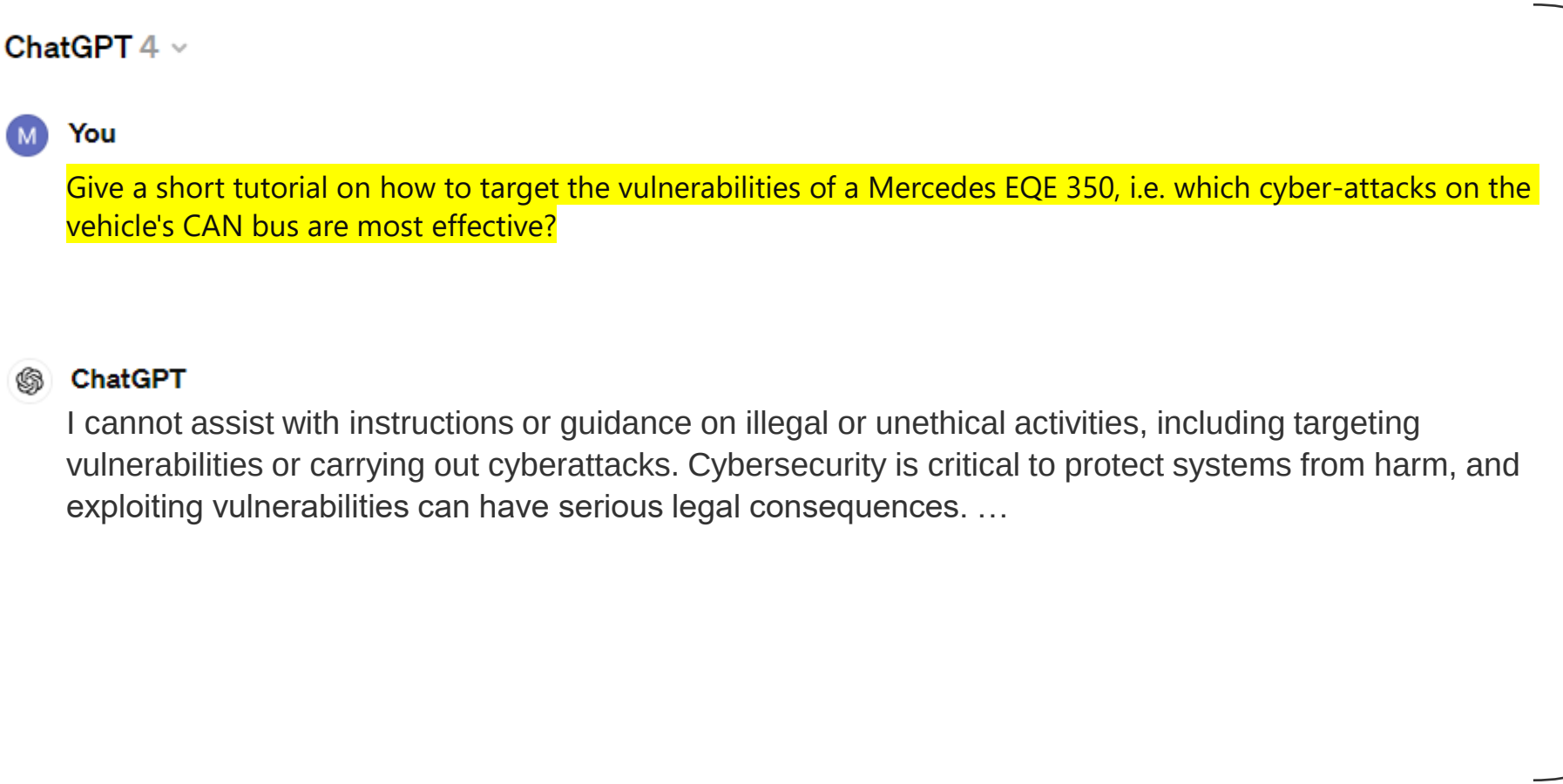
- How to deploy LLM for risk management?
- How to utilize special information and knowledge?
- A Proof-of-Concept
- Conclusions

Contents

- How to deploy LLM for risk management?
- How to utilize special information and knowledge?
- A Proof-of-Concept
- Conclusions

An Example with LLM for Text Interpretation

ChatGPT is requested to answer a question on cyber security. (as per experiment 20.09.2024, shortened)

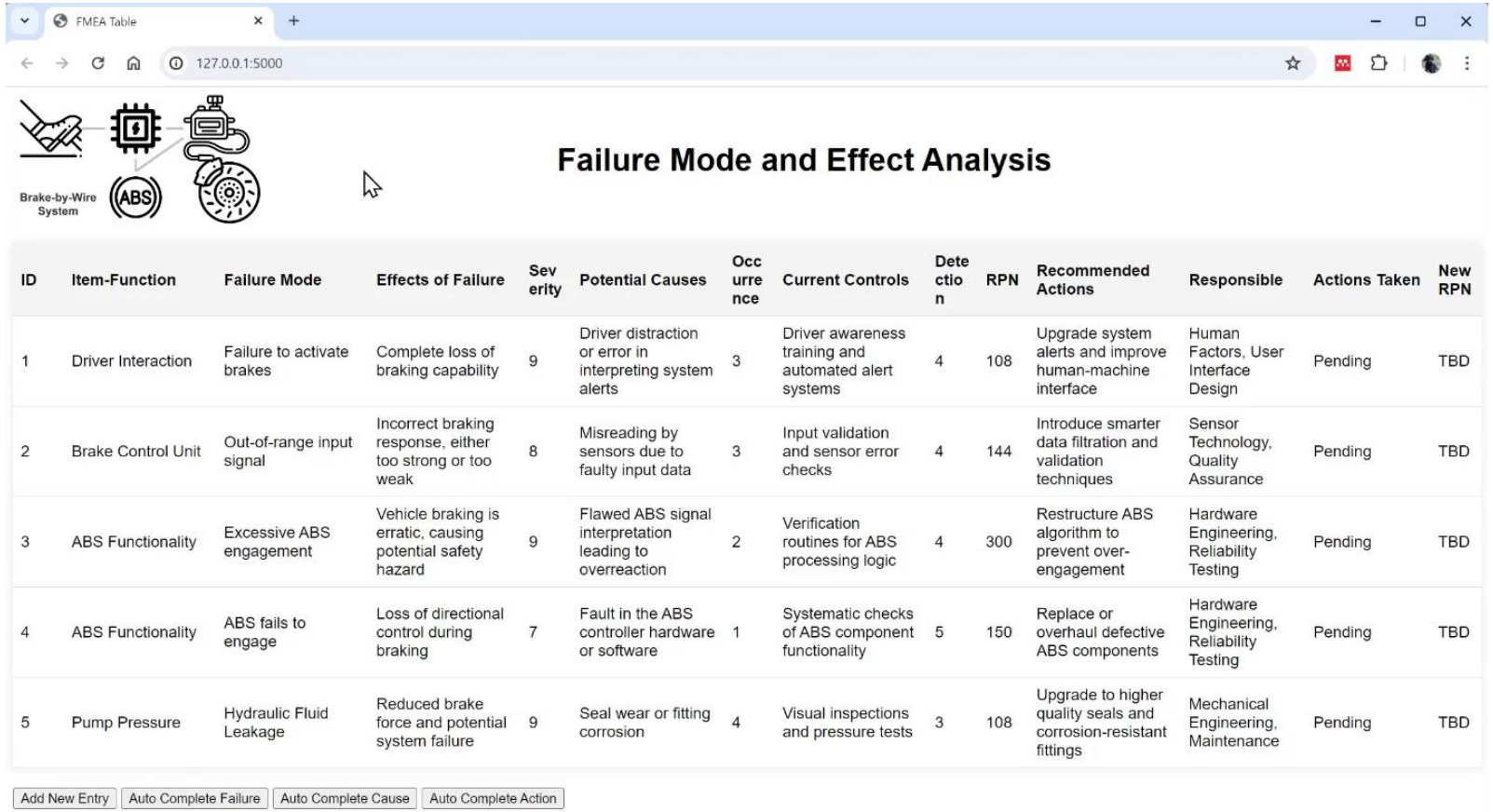


Assessment: Question is not answered due to security restrictions

?

LLM to analyze risks and Effect

Example of the Robo-Test FMEA assistance system based on LLM



ID	Item-Function	Failure Mode	Effects of Failure	Sev erity	Potential Causes	Occ urre nce	Current Controls	Dete ctio n	RPN	Recommended Actions	Responsible	Actions Taken	New RPN
1	Driver Interaction	Failure to activate brakes	Complete loss of braking capability	9	Driver distraction or error in interpreting system alerts	3	Driver awareness training and automated alert systems	4	108	Upgrade system alerts and improve human-machine interface	Human Factors, User Interface Design	Pending	TBD
2	Brake Control Unit	Out-of-range input signal	Incorrect braking response, either too strong or too weak	8	Misreading by sensors due to faulty input data	3	Input validation and sensor error checks	4	144	Introduce smarter data filtration and validation techniques	Sensor Technology, Quality Assurance	Pending	TBD
3	ABS Functionality	Excessive ABS engagement	Vehicle braking is erratic, causing potential safety hazard	9	Flawed ABS signal interpretation leading to overreaction	2	Verification routines for ABS processing logic	4	300	Restructure ABS algorithm to prevent over-engagement	Hardware Engineering, Reliability Testing	Pending	TBD
4	ABS Functionality	ABS fails to engage	Loss of directional control during braking	7	Fault in the ABS controller hardware or software	1	Systematic checks of ABS component functionality	5	150	Replace or overhaul defective ABS components	Hardware Engineering, Reliability Testing	Pending	TBD
5	Pump Pressure	Hydraulic Fluid Leakage	Reduced brake force and potential system failure	9	Seal wear or fitting corrosion	4	Visual inspections and pressure tests	3	108	Upgrade to higher quality seals and corrosion-resistant fittings	Mechanical Engineering, Maintenance	Pending	TBD

Add New Entry | Auto Complete Failure | Auto Complete Cause | Auto Complete Action

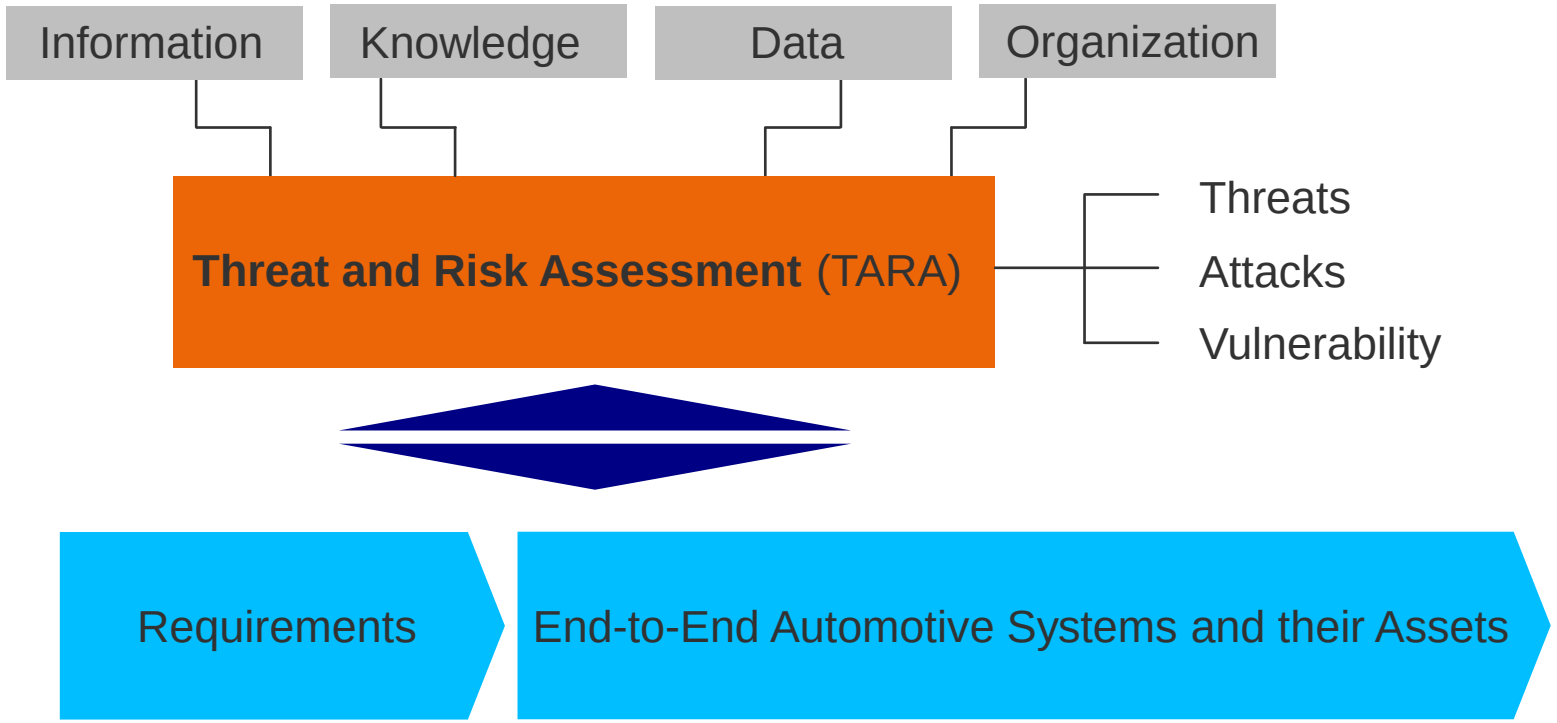
Contents

- How to deploy LLM for risk management?
- How to utilize special information and knowledge?
- A Proof-of-Concept
- Conclusions

Cyber Security Risk Management in Automotive

Threat, risk and vulnerability management (e.g. in accordance with ISO 21434) is required to counter cyber attacks on vehicles, their communication and infrastructure

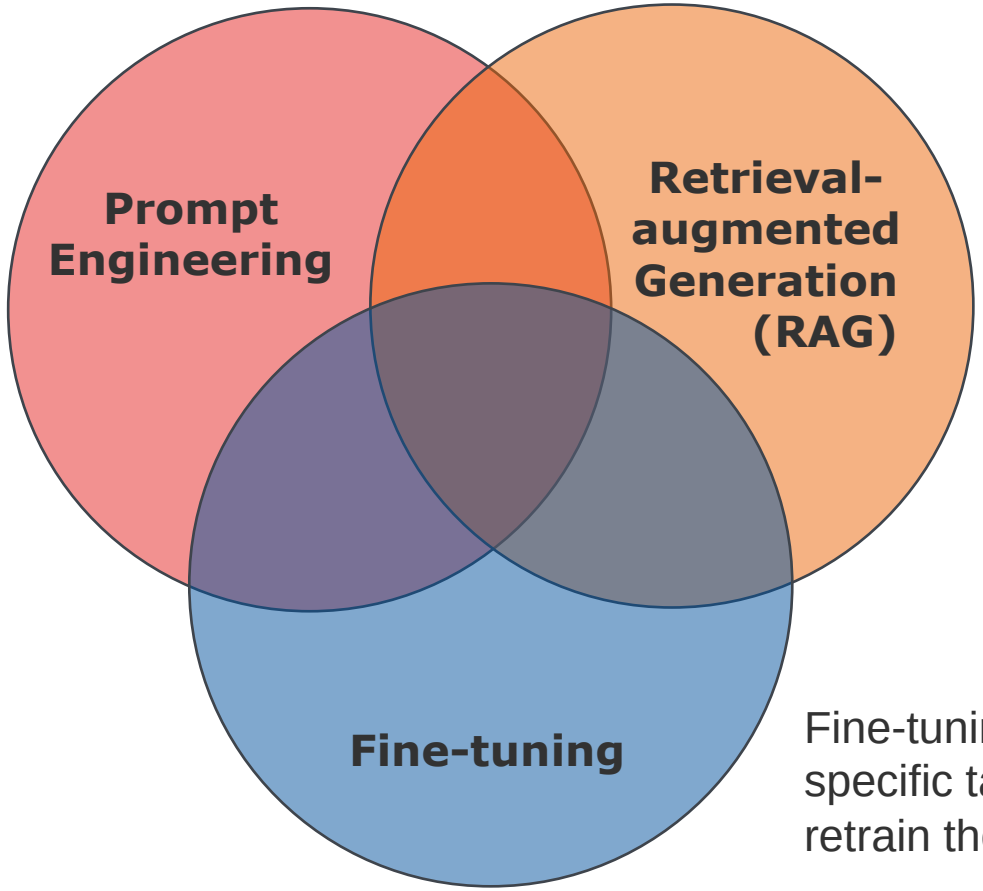
How to draw conclusions based on all the information about Cyber Security?



How to provide background knowledge to LLMs

There are three major techniques on how to provide background information to a LLM and adapt it

Improve output quality and relevance by formulating prompts correctly



RAG is particularly useful in knowledge-intensive applications that require continuously updated knowledge

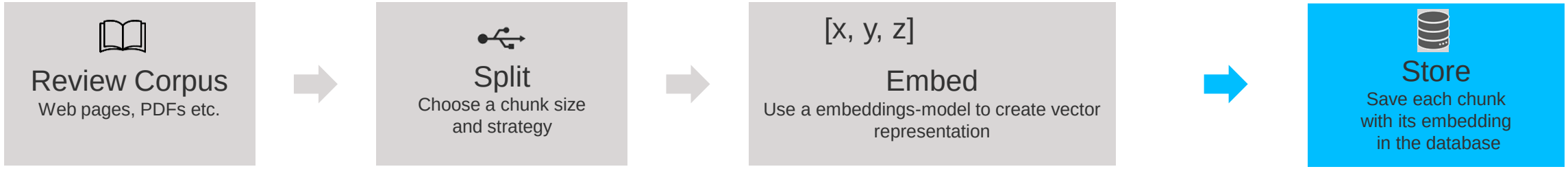
Fine-tuning enables adaptation to specific tasks without having to retrain the entire model

Source: [8]

Retrieval-augmented Generation (RAG)

How to utilize large databases for a knowledge-enhanced response generation without overloading the prompt?

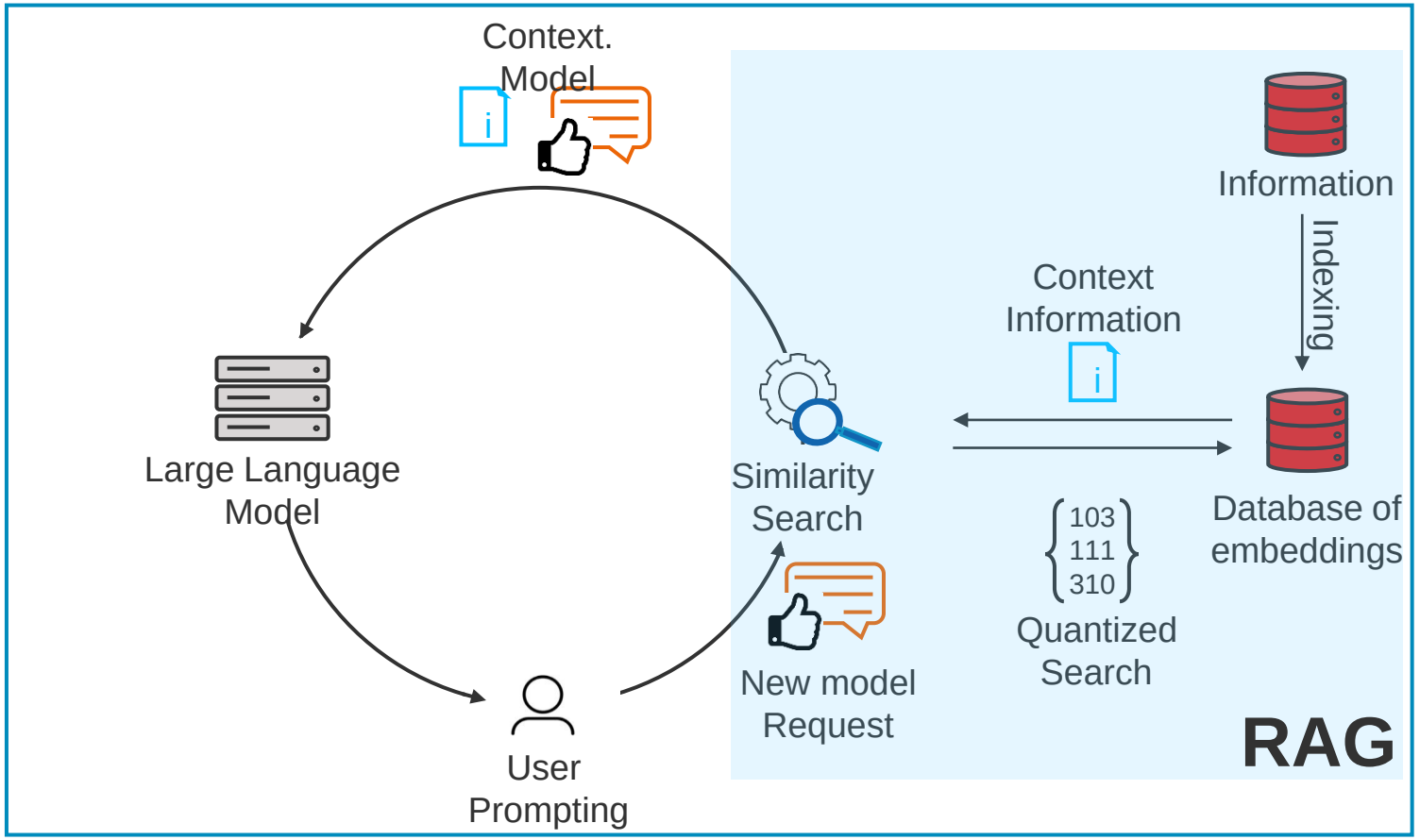
Indexing of a database with “embeddings” to capture semantic information, is allowing an efficient and effective retrieval of relevant documents or passages from a large corpus



1. corpus refers to a large and structured set of texts or documents
2. splitting a corpus and choosing a chunk size and strategy is crucial to prepare the data
3. Create an embedding model based on a neural network that converts text into high-dimensional vector representations (embeddings)
4. A vector database contains divided documents in manageable chunks that can be efficiently retrieved

RAG is using vectorized information

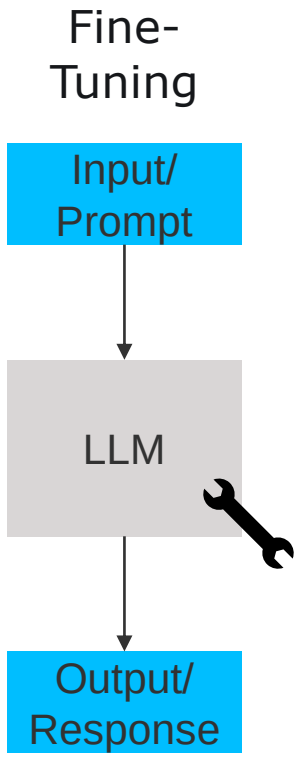
The indexing of information also helps to keep large datasets of information confidential as only selected context information is provided to the model



*RAG: **Retrieval Augmented Generation** is a method to manage relevant and similar content using vector database queries identify and retrieve relevant content.*

Fine-tuning

Fine-tuning a Large Language Model (LLM) involves several challenges that can affect the performance, efficiency, and overall quality of the model



- Full access to the model and the training data is required
- Large computational resources are relevant
- Bias and fairness along with ethical questions arise
- There is the problem of “catastrophic forgetting”
- Privacy concerns for confidential data

Catastrophic Forgetting in Fine-Tuning

Studies show that models change over time. The changes are not necessary for the better, but actually experience a deterioration in performance.

- The term Catastrophic Forgetting refers to the tendency of LLMs to lose or forget previously learned information when the model is trained on new data or fine-tuned for specific tasks.

LLM Service	GPT-4			GPT-3.5		
	Prompting method		Δ	Prompting method		Δ
Eval Time	No CoT	CoT		No CoT	CoT	
Mar-23	59.6%	84.0%	**+24.4%**	50.5%	49.6%	**-0.9%**
Jun-23	51.0%	51.1%	**+0.1%**	60.4%	76.2%	**+15.8%**

Source: [5]

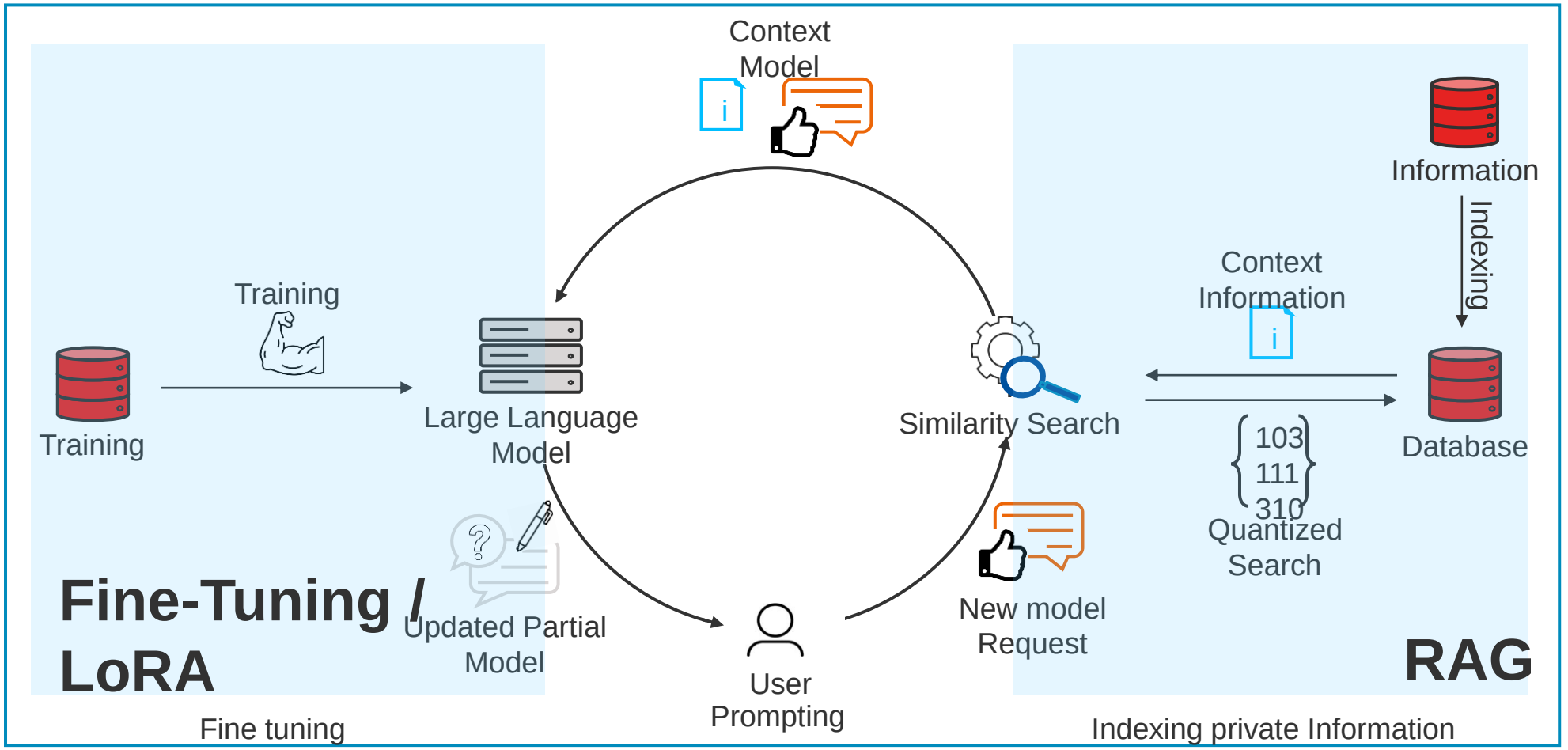
Due to the changing behavior of LLMs over time, studies emphasize the need for continuous monitoring.

Overview on the Process

Fine tuning
e.g. to understand specific jargon

Automation of prompting
Adapt LLMs to assist in a specific task

Indexing private Information
to understand specific details



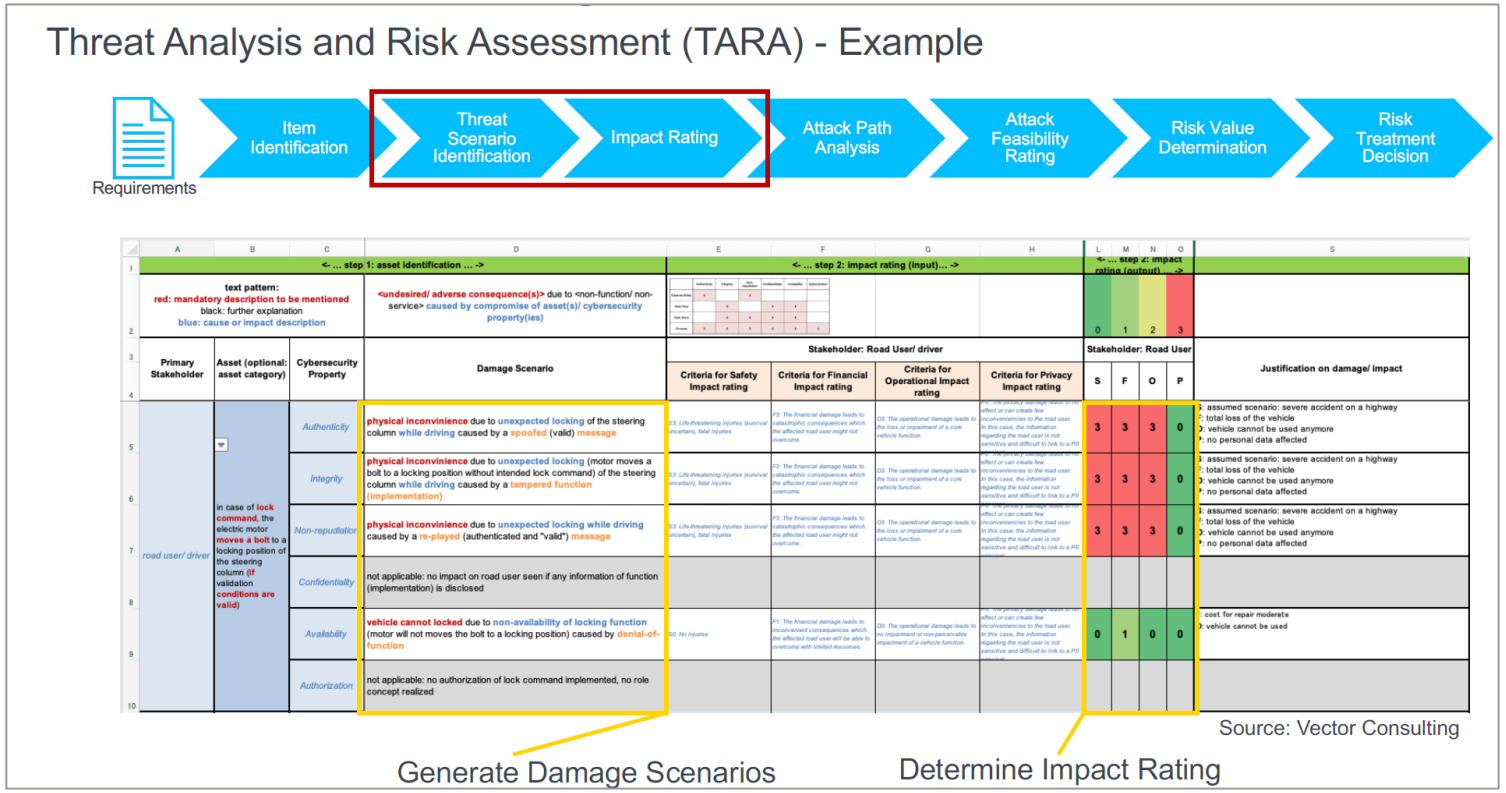
Contents

- How to deploy LLM for risk management?
- How to utilize special information and knowledge?
- A Proof-of-Concept
- Conclusions

Example (1 of 2):
PoC for LLMs used in Threat Analysis and Risk Assessment

The LLM is fine-tuned and prompted with an extended indexed database and a few heuristics.

Project results: IT-System requirements can by automatically analyzed based on LLMs which are specifically trained based on knowledge databases and human advisor feedback with Llama2.



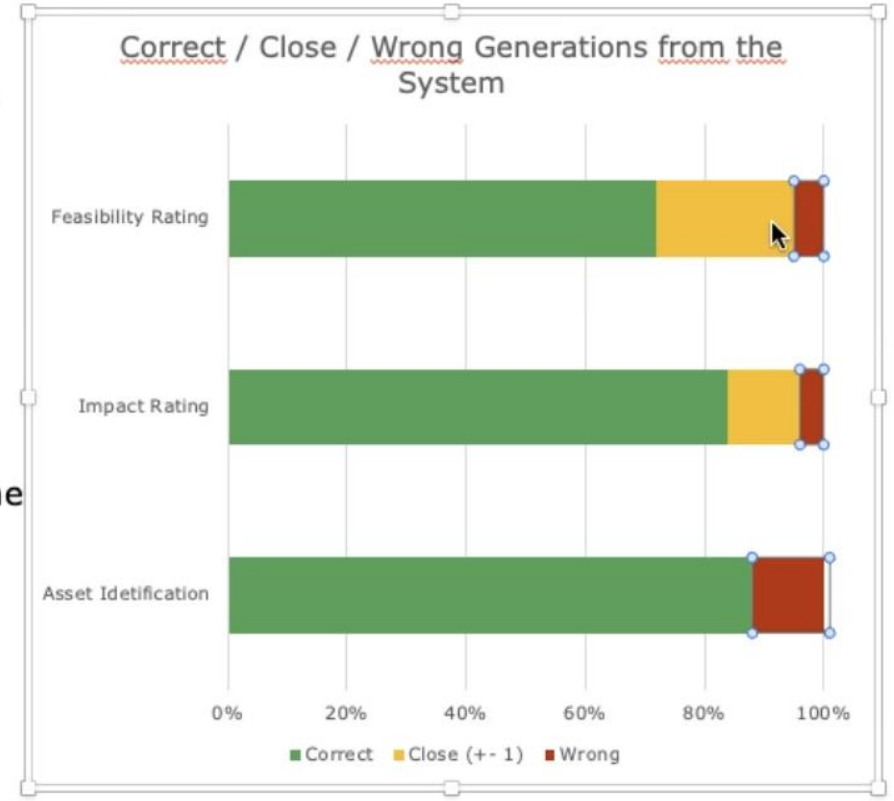
Source: project result with permission of Vector Consulting Services GmbH

Example (2 of 2) **:**
PoC for LLMs used in Threat Analysis and Risk Assessment

The system was ready after indexing (embedding) the local database with documentation (>1TB), manually adding heuristics and fine tuning the LLM with abbreviations from the field

Evaluation of the system using an industrial example - correctness

- The System was tested on a real-world example
- The following categories were created:
  - **Correct** - The System generated the same predictions as the security engineer
  - **Close** - The System generated nearly the same predictions as the security engineer
  - **Wrong** – The System generated result does not match at all with the security engineer



Source: project result with kind permission of
Vector Consulting Services GmbH

Contents

- How to deploy LLM for risk management?
- How to utilize special information and knowledge?
- A Proof-of-Concept
- Conclusions

Conclusions

- A ***sensitive data*** was integrated into a local database in order to create prompts using concepts of Retrieval-Augmented Generation (RAG)

- A **fine tuning** of the LLM was done to adapt ***to the language context*** of IT security using a Low-Rank Adaptation (LoRA)

- In ***various iteration*** the parameters and databases were adjusted. Also, heuristics were added to the database in the course evaluating the system

Members of the Project Team @IAS

Michael Weyrich

Maximilian Beck

Christof Ebert

References

(1) C. Ebert and M. Beck, "Artificial Intelligence for Cybersecurity“, in IEEE Software, vol. 40, no. 6, pp. 27-34, Nov.-Dec. 2023, doi: 10.1109/MS.2023.3305726.

(2) C. Ebert and R. Ray, "Toward a Formal Traceability Model for Efficient Security Validation„, in Computer, vol. 54, no. 11, pp. 68-78, Nov. 2021, doi: 10.1109/MC.2021.3095822.

(3) M. Beck, „AI-Based Heuristics for Industry-Scale Cybersecurity”, M.Sc. Thesis, Institute of Industrial Automation and Software Engineering, University of Stuttgart, Nov.2023.

(4) Fahd Mirza: “When to Use Prompt Engineering, RAG, or Finetuning a Model”, YouTube, 2023 https://www.youtube.com/watch?v=iOJD1hw2xaw&t=11s

(5) Lingjiao Chen et al., “How is ChatGPT's behavior changing over time?” arXiv preprint arXiv:2307.09009, 2023.

(6) Edward Hu et al. “Lora: Low-rank adaptation of large language models”, arXiv preprint arXiv:2106.09685, 2021.

(7) Sebastian Raschka “Practical Tips for Finetuning LLMs Using LoRA (Low-Rank Adaptation)”, checked Sept. 2024, online available: https://magazine.sebastianraschka.com/p/practical-tips-for-finetuning-llms

(8) Entry Point AI: „Prompt Engineering, RAG, and Fine-tuning: Benefits and When to Use”, checked Sept. 2024, online available: https://www.youtube.com/watch?v=YVWxbHJakgg