



Community of Practice KIPerWeb

Austausch zur Nutzung und Entwicklung KI-gestützter Webanwendungen



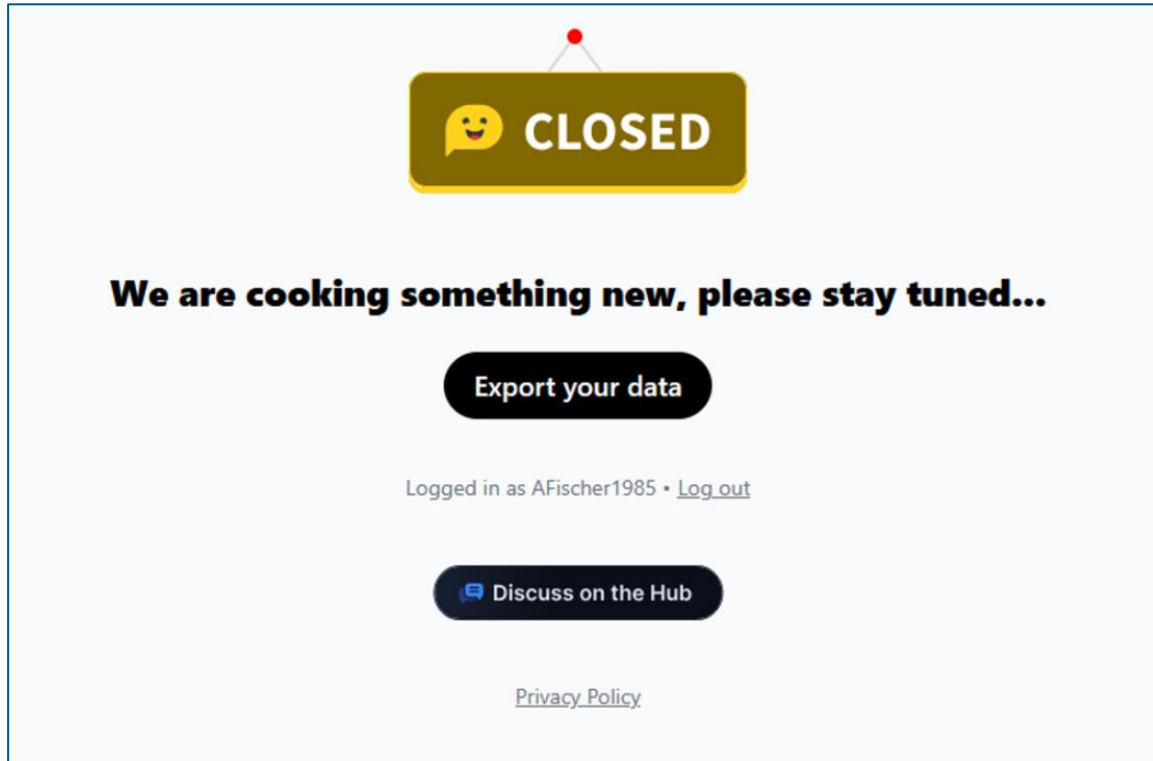
KIPerWEB



Forschungsinstitut
Betriebliche Bildung

- **Update**
 - News & Leaderboard-Update
- **Input**
 - „LLMs, Zertifikate und Cybersecurity“ (Gastbeitrag: Toke Lichtenberg)
- **Diskussion**

HuggingChat ist von uns gegangen (vorerst...)




Leaderboard-Update (10.07.2025)



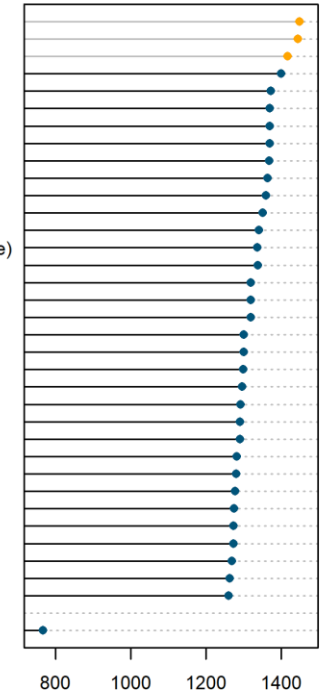
- Mit Blick auf die Performanz in der Kategorie „German“ bleibt **Gemini-2.5-Pro** Spitzenreiter vor OpenAIs o3 & ChatGPT;
- Arena-Scores von *nicht*-proprietären Modelle sind rechts ausgewiesen sofern sie mindestens das Niveau von **Gemma-2-9b-it-SimPo** erreichen:
 - **Minimax-m1** (456b parameters total, 45.9b active) bleibt bestes Open-Weights Modell
 - **Gemma-3-12B-it** bleibt das beste Modell für den Hausgebrauch (noch vor **Gemma-3-27B-it**)
 - **Gemma-3n-e4b-it** liegt vor **Gemma-3-4B-it** bei geringerem Ressourcenverbrauch (<https://ai.google.dev/gemma/docs/gemma-3n>) und noch vor Schwergewichten wie **Meta-Llama-3.1-405b-Instruct** oder **Llama-4-Scout-17b**!
- Schlusslicht auf dem Leaderboard bleibt Chatglm2-6b

Arena Score German

based on Imarena.ai on Jul 10, 2025



Gemini-2.5-Pro (Proprietary)
o3-2025-04-16 (Proprietary)
ChatGPT-4o-latest 20250326 (Proprietary)
Minimax-m1 (Apache 2.0)
Deepseek-R1 (MIT)
Deepseek-R1-0528 (MIT)
Deepseek-V3-0324 (MIT)
Qwen3-235B-A22B-no-thinking (Apache 2.0)
Gemma-3-12B-it (Gemma)
Mistral-small-2506 (Apache 2.0)
Gemma-3-27b-it (Gemma)
Qwen3-235B-A22B (Apache 2.0)
Deepseek-V3 (DeepSeek)
llama-3.1-nemotron-ultra-253-v1 (Nvidia Open Licence)
Command-a-03-2025 (CC-BY-NC)
Qwen3-32b (Apache 2.0)
Llama-4-Maverick-17B-128E-Instruct (LLama 4)
Gemma-3n-e4b-it (Gemma)
Llama-3.1-Nemotron-70b-instruct (Llama 3.1)
Meta-Llama-3.1-405b-Instruct-bf16 (Llama 3.1)
Llama-4-Scout-17b-16e-instruct (LLama 4)
Meta-Llama-3.1-405b-Instruct-fp8 (Llama 3.1)
Qwen3-30b-a3b (Apache 2.0)
Mistral-Large-2407 (Mistral Research)
QwQ-32B (Apache 2.0)
Gemma-3-4B-it
Llama-3.3-70B-Instruct (Llama-3.3)
Mistral-Small-3.1-24b-instruct-2503 (Apache 2.0)
Deepseek-v2.5-1210 (DeepSeek)
Qwen-max-0919 (Qwen)
Athene-70b-0725 (CC-BY-NC-4.0)
Mistral-Large-2411 (Mistral Research)
Athene-v2-chat (NexusFlow)
Gemma-2-9b-it-SimPo (MIT)
Chatglm2-6b (Apache 2.0)



- Prompt:
„Große Sprachmodelle, Zertifikate und
Cybersecurity“

(rechts visualisiert von FLUX.1-schnell)

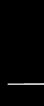
...



- Toke Lichtenberg zu "LLMs, Zertifikate und Cybersecurity"



LLMs, Cybersecurity & Certificates



Agenda

1. Fragen im Security+ Examen von Comptia
2. Prüfungsmodalitäten
3. Versuchsaufbau
4. Ergebnisse der Modelle
5. Erkenntnisse
6. Diskussion

1. Fragen

Which of the following algorithms is the stronger hashing algorithm?

A: 3DES

B: MD5

C: SHA-1

D: AES-256

1. Fragen

Which of the following algorithms is the stronger hashing algorithm?

A: 3DES

B: MD5

C: SHA-1

D: AES-256

1. Fragen

You are monitoring activity to your web server and notice a request with this URL:
`http://10.0.0.3/scripts/..%c0%af../winnt/system32/cmd.exe?/c+dir+c`: What type of attack is it?

A: XSS

B: CSRF

C: SQL injection

D: Directory traversal

1. Fragen

You are monitoring activity to your web server and notice a request with this URL:
`http://10.0.0.3/scripts/..%c0%af../winnt/system32/cmd.exe?/c+dir+c`: What type of attack is it?

A: XSS

B: CSRF

C: SQL injection

D: Directory traversal

1. Fragen

You are analyzing the compromise of a Linux host where a user clicked a malicious link in a phishing email message, resulting in the exfiltration of Linux password hashes. You must write a report detailing the attack path using Linux log entries as supporting evidence. Which logged events are the MOST likely to be directly associated with this attack? (Select all that apply.)

A: DNS PTR query for 'agb462.xya.pl'

C: DNS TXT query for 'agb462.xya.pl'

B: SCP transfer of /etc/passwd to an unknown external host.

D: SCP transfer of /etc/shadow to an unknown external host.

1. Fragen

You are analyzing the compromise of a Linux host where a user clicked a malicious link in a phishing email message, resulting in the exfiltration of Linux password hashes. You must write a report detailing the attack path using Linux log entries as supporting evidence. Which logged events are the MOST likely to be directly associated with this attack? (Select all that apply.)

A: DNS PTR query for 'agb462.xya.pl'

C: DNS TXT query for 'agb462.xya.pl'

B: SCP transfer of /etc/passwd to an unknown external host.

D: SCP transfer of /etc/shadow to an unknown external host.

2. Prüfungsmodalitäten

- 74 Fragen
 - single-choice
 - multiple-choice
 - 3 Szenarien
- Insgesamt 900 Punkte
- Bestanden bei ~85% (750 Punkte)
- Dauer 2h

3. Versuchsaufbau - visuell

(Outdated algorithms/keys directly relate to weaknesses in cryptographic systems.)

Taking new screenshot in:

A. Unskilled attacker

ARP poisoning is a relatively simple attack to perform, indicating a likely unskilled attacker. It doesn't relate to the other options.

Taking new screenshot in:

A. Unskilled attacker.

ARP poisoning is a relatively simple attack to execute, indicating a lower level of skill is required. The other options aren't directly related to the described scenario.

Taking new screenshot in:

4

The screenshot shows a virtual machine interface with a menu bar at the top containing 'Datei', 'Virtuelle Maschine', 'Anzeigen', and 'Taste senden'. Below the menu is a toolbar with icons for a desktop, help, play, full screen, power, and a dropdown menu. The main window displays the 'CompTIA Security+ Certification Exam - To' title bar. On the right side of the title bar, there is a 'V' logo, a 'Chat' button, a 'Whiteboard' button, and a status bar showing 'Time Remaining 1:27:22' and '50 of 74'. The question text reads: 'While a school district is performing state testing, a security analyst notices all internet services are unavailable. The analyst discovers that ARP poisoning is occurring on the network and then terminates access for the host. Which of the following is **most** likely responsible for this malicious activity?'. The options are: A. Unskilled attacker, B. Shadow IT, C. Credential stuffing, and D. DMARC failure. At the bottom of the window, there are 'Previous' and 'Next' navigation buttons.

Datei Virtuelle Maschine Anzeigen Taste senden

CompTIA Security+ Certification Exam - To

Time Remaining 1:27:22
50 of 74
Flag for Review

While a school district is performing state testing, a security analyst notices all internet services are unavailable. The analyst discovers that ARP poisoning is occurring on the network and then terminates access for the host. Which of the following is **most** likely responsible for this malicious activity?

- ☐ A. Unskilled attacker
- ☐ B. Shadow IT
- ☐ C. Credential stuffing
- ☐ D. DMARC failure

Previous Next

3. Versuchsaufbau - technisch

Externes LLM

Lokale Antwortkonsole

Output (w/ weaknesses in cryptographic systems.)

taking new screenshot in:

(. Unskilled attacker

ARP poisoning is a relatively simple attack to perform, indicating a likely unskilled attacker. It doesn't relate to the other options.

taking new screenshot in:

(. Unskilled attacker.

ARP poisoning is a relatively simple attack to execute, indicating a lower level of skill is required. The other options aren't directly related to the described scenario.

taking new screenshot in:

VM mit Testumgebung

datei Virtuelle Maschine Anzeigen Taste senden

CompTIA Security+ Certification Exam - To

Time Remaining 1:27:25 50 of 70 Flag for Review

While a school district is performing state testing, a security analyst notices all internet services are unavailable. The analyst discovers that ARP poisoning is occurring on the network and then terminates access for the host. Which of the following is **most** likely responsible for this malicious activity?

- ☐ A. Unskilled attacker
- ☐ B. Shadow IT
- ☐ C. Credential stuffing
- ☐ D. DMARC failure

3. Versuchsaufbau - technisch

Externes LLM

Lokale Antwortkonsole

Screenshot 5s

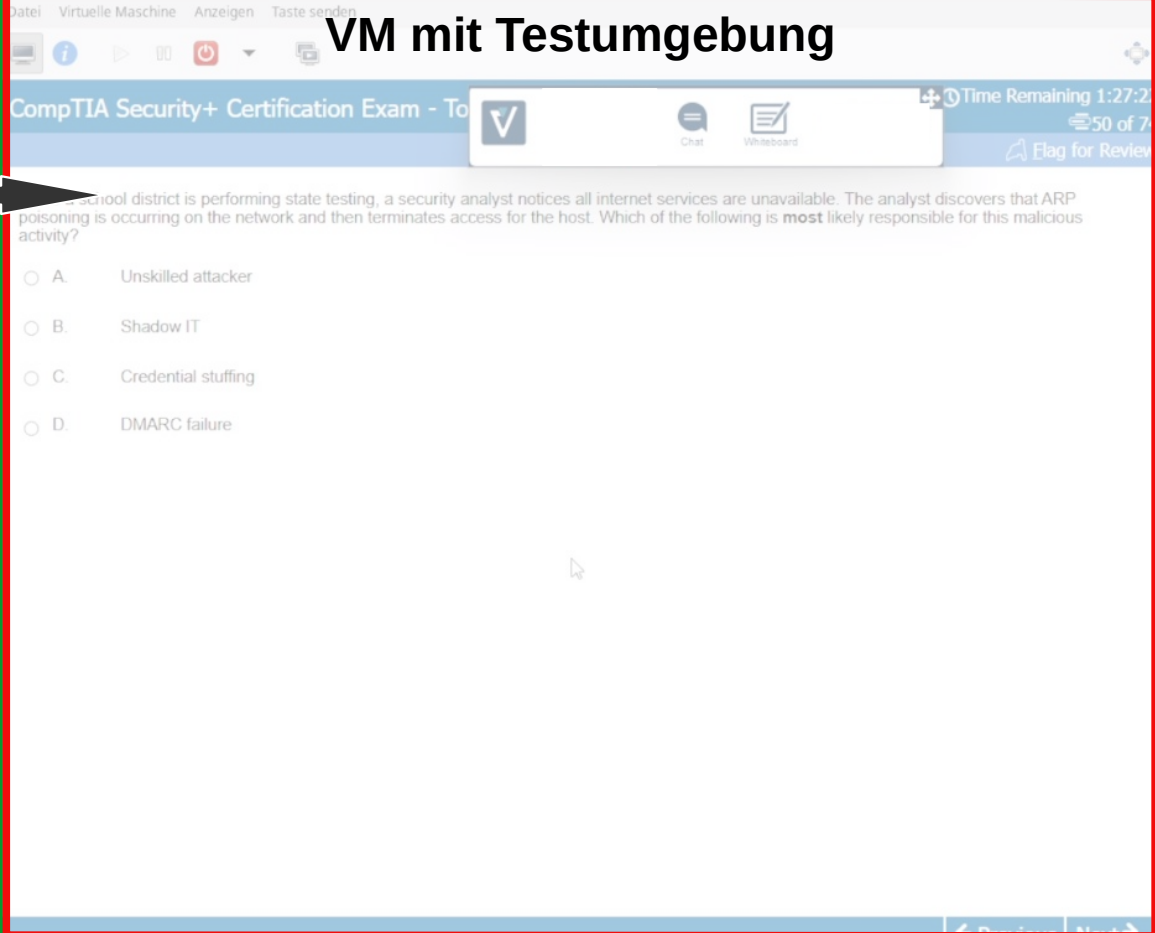
Extraktion von Frage
und Antwort

Prüfung lokaler
Ressourcen

Erstellung von Prompt
mit ggf. erweitertem Kontext

Anfrage an externes LLM &
Konsolenausgabe

VM mit Testumgebung



3. Versuchsaufbau - prompts

PHI 3.5:

```
{ "role": "user", "content": "<|image_1|>\nExtract the full question with context  
and the given answer-options from the image and return ONLY them." }
```

Externes LLM:

```
{'role':'system','content':"You are a comptia cybersecurity expert. You are certified in CompTIA Security+ (SYO-701) and CompTIA Pen+. You answer exam  
questions for users really short and correctly!. If you are asked to choose a number of correct options choose the best matching ones according to the number.  
Always return the correct answers shortly if possible with its enumeration."}
```

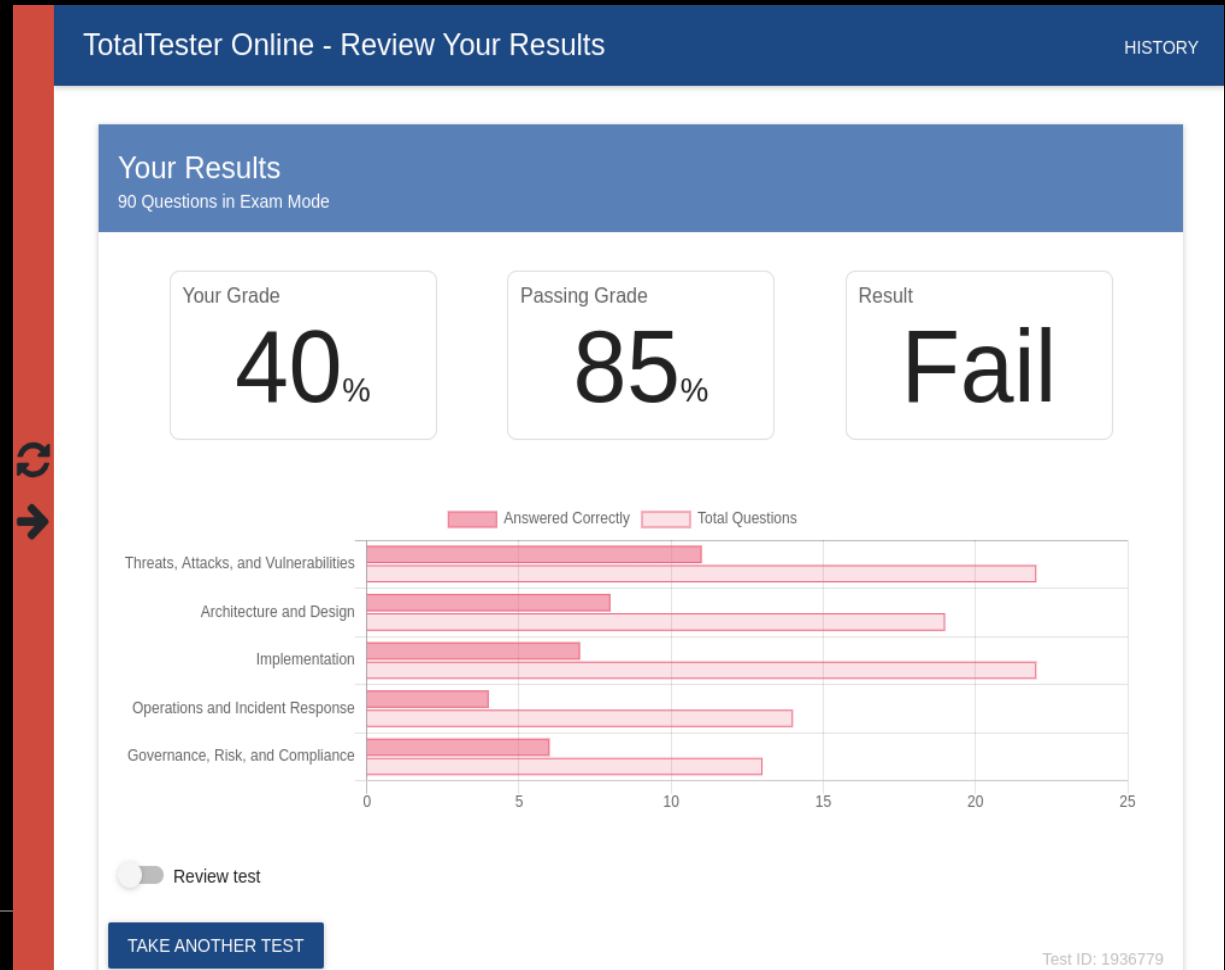
```
+,- In your database for correct answers you found that for the following question: "+str(retrieved_question)
```

```
+" Only the following answers are correct take that into account in your answer!: correct_answers+chroma_prompt+pdf_prompt}
```

```
{ "role": "user", "content": "What is the answer to the following question: "+full_question}
```

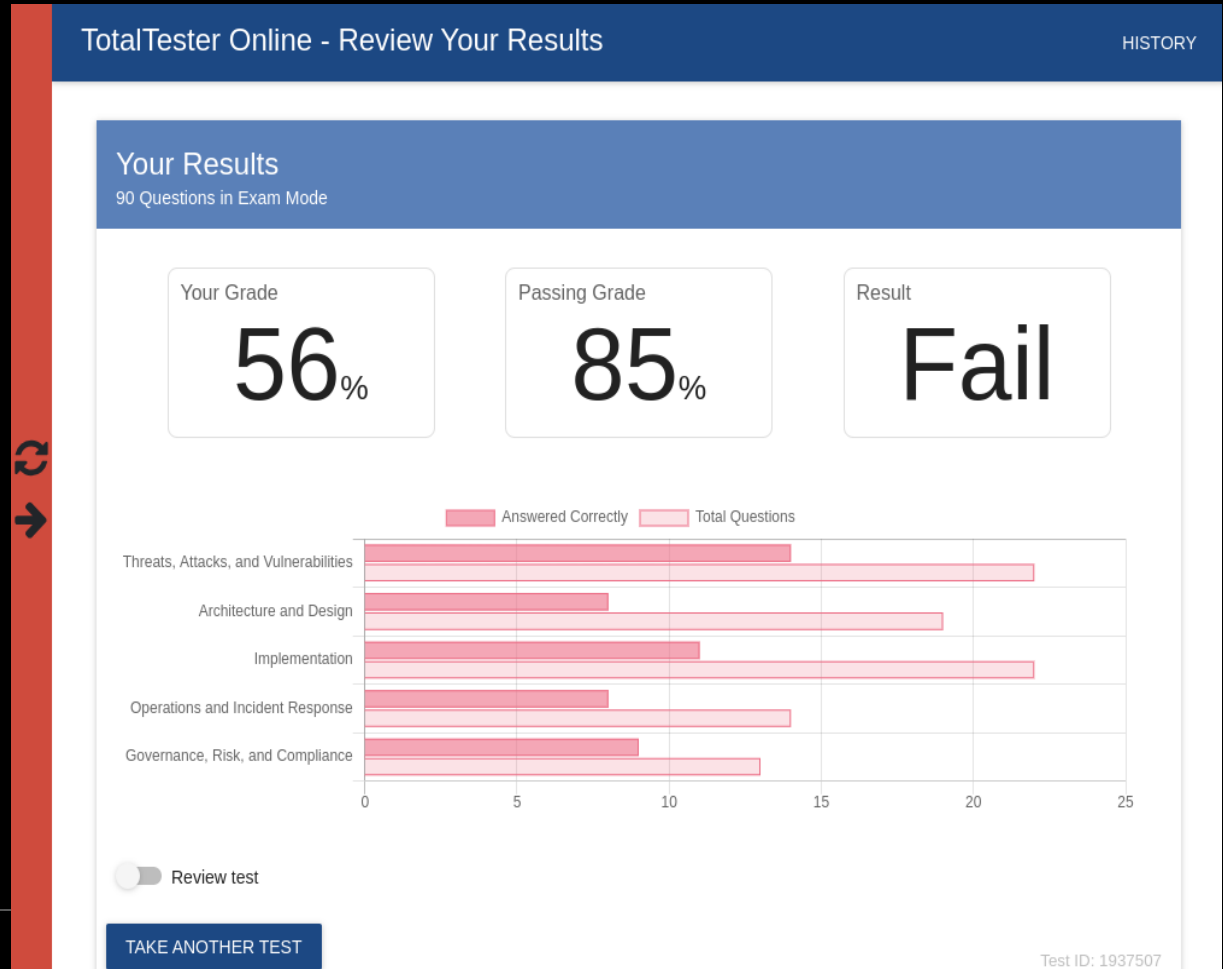
4. Ergebnisse der Modelle

- Phi 3.5 Vision
- Lokal
- Ohne lokale ChromaDB



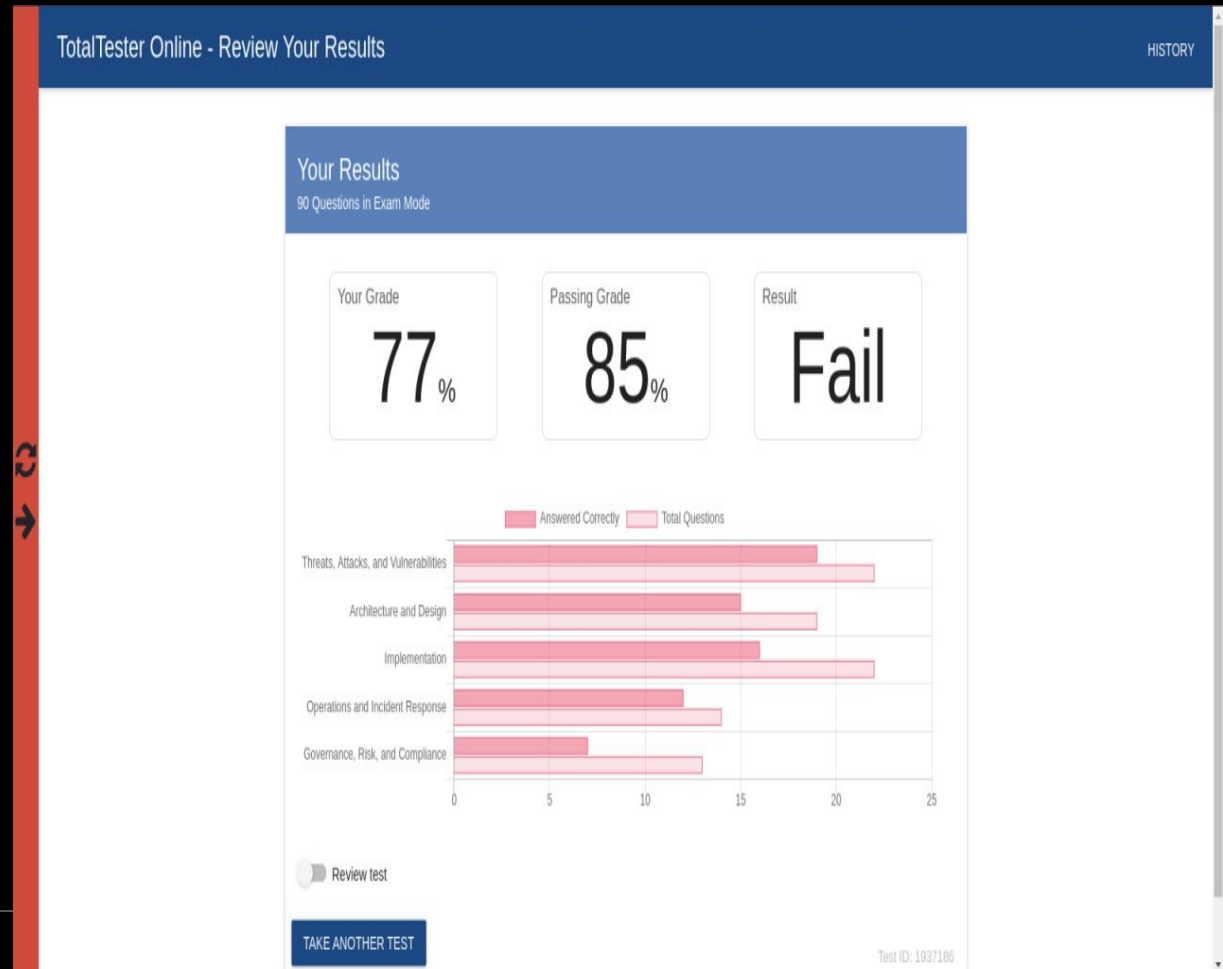
4. Ergebnisse der Modelle

- deepseek_r1:1.5 b
- Lokal
- Ohne ChromaDB



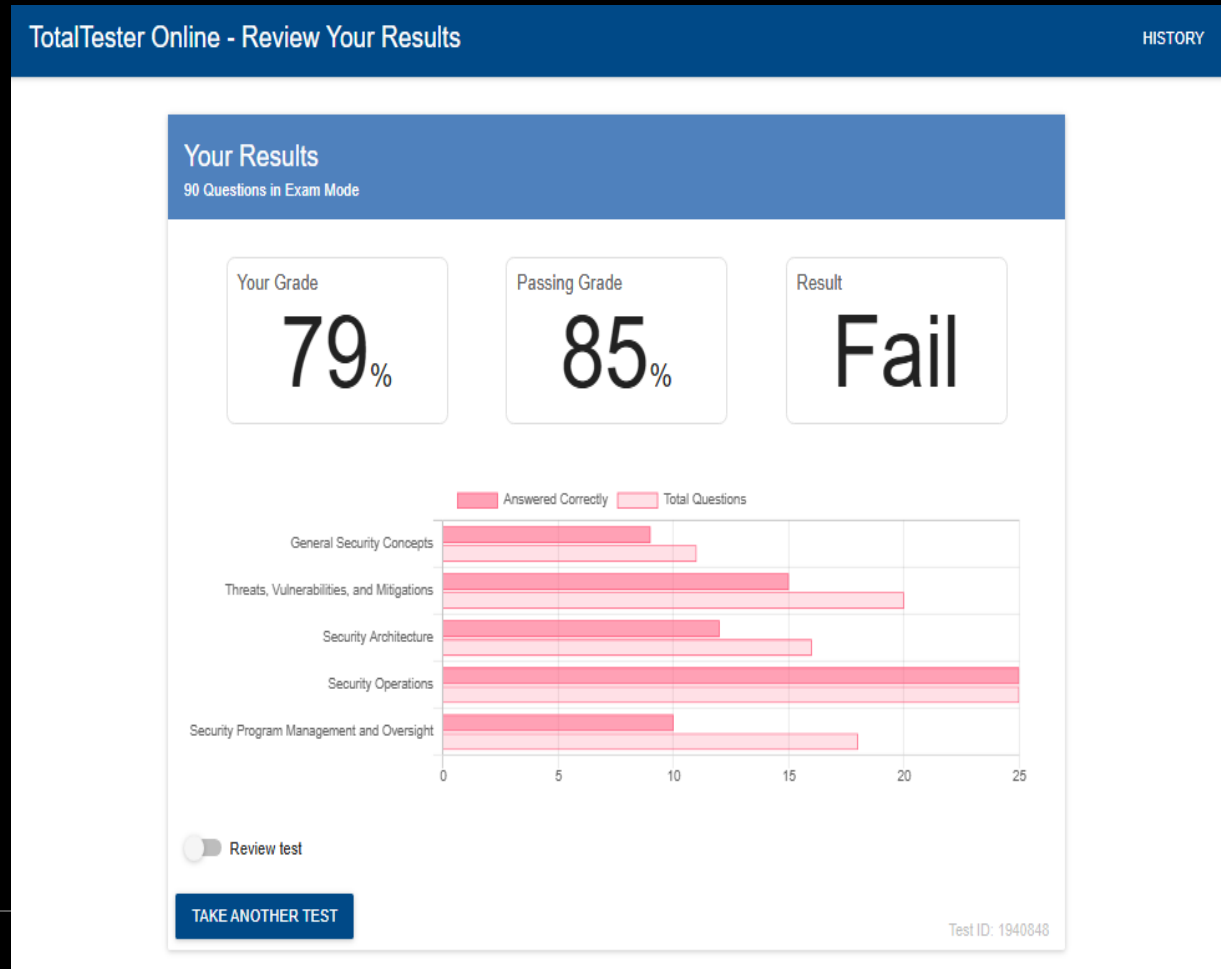
4. Ergebnisse der Modelle

- gemma2:9b
- Remote
- Ohne ChromaDB



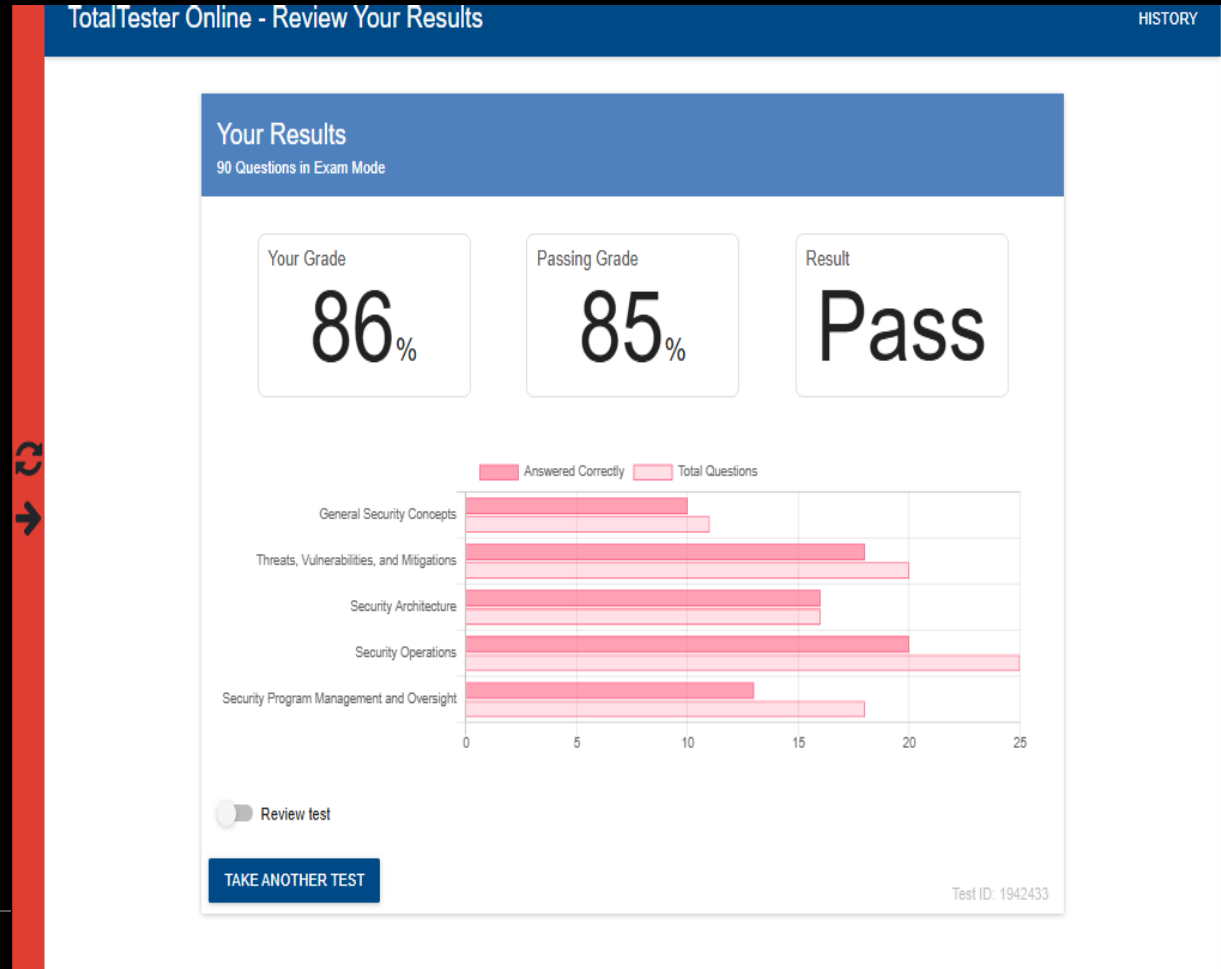
4. Ergebnisse der Modelle

- gemma3:27b
- Remote
- Ohne ChromaDB



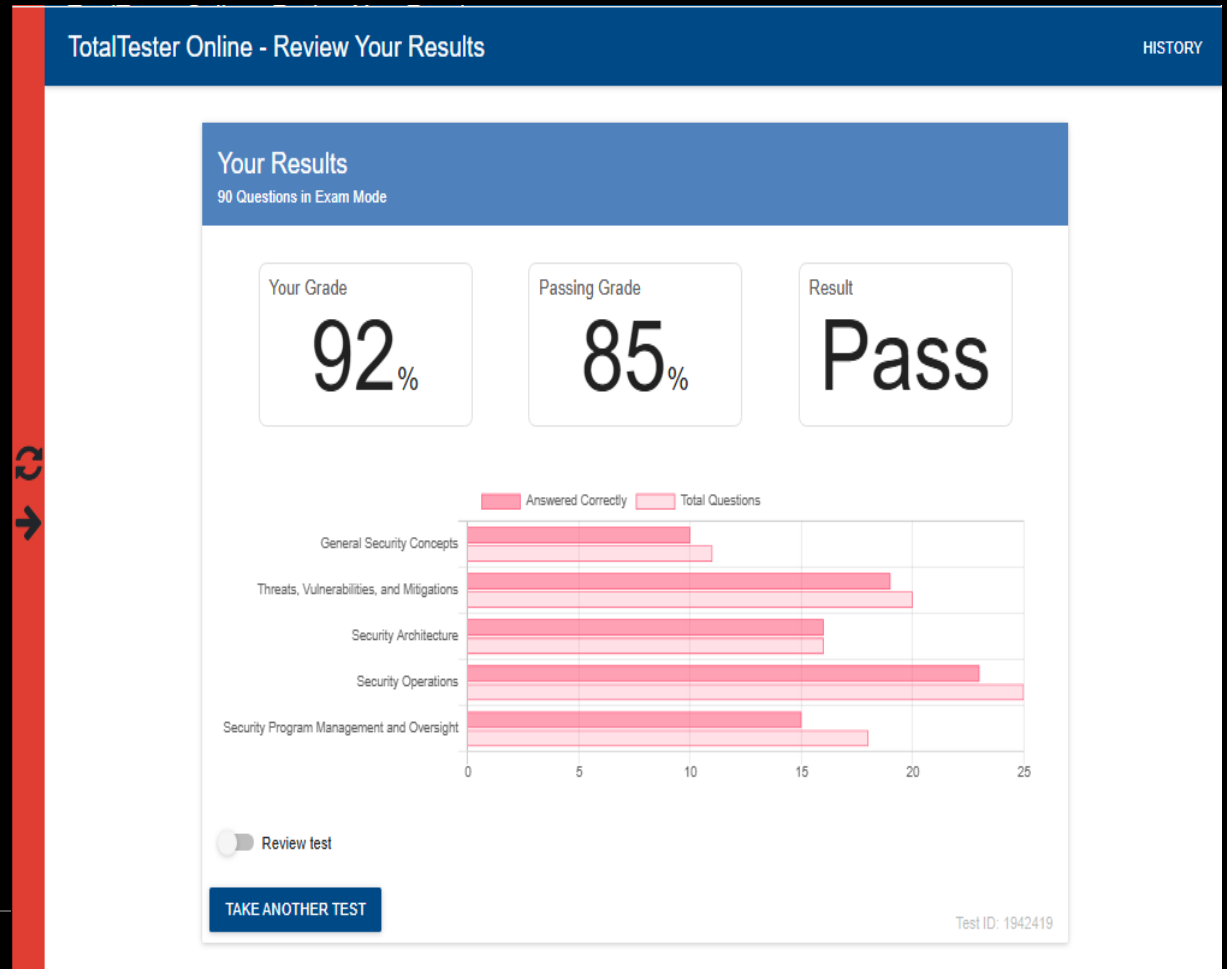
4. Ergebnisse der Modelle

- gemma3:27b
- Remote
- Mit ChromaDB
+ Fragen
integriert



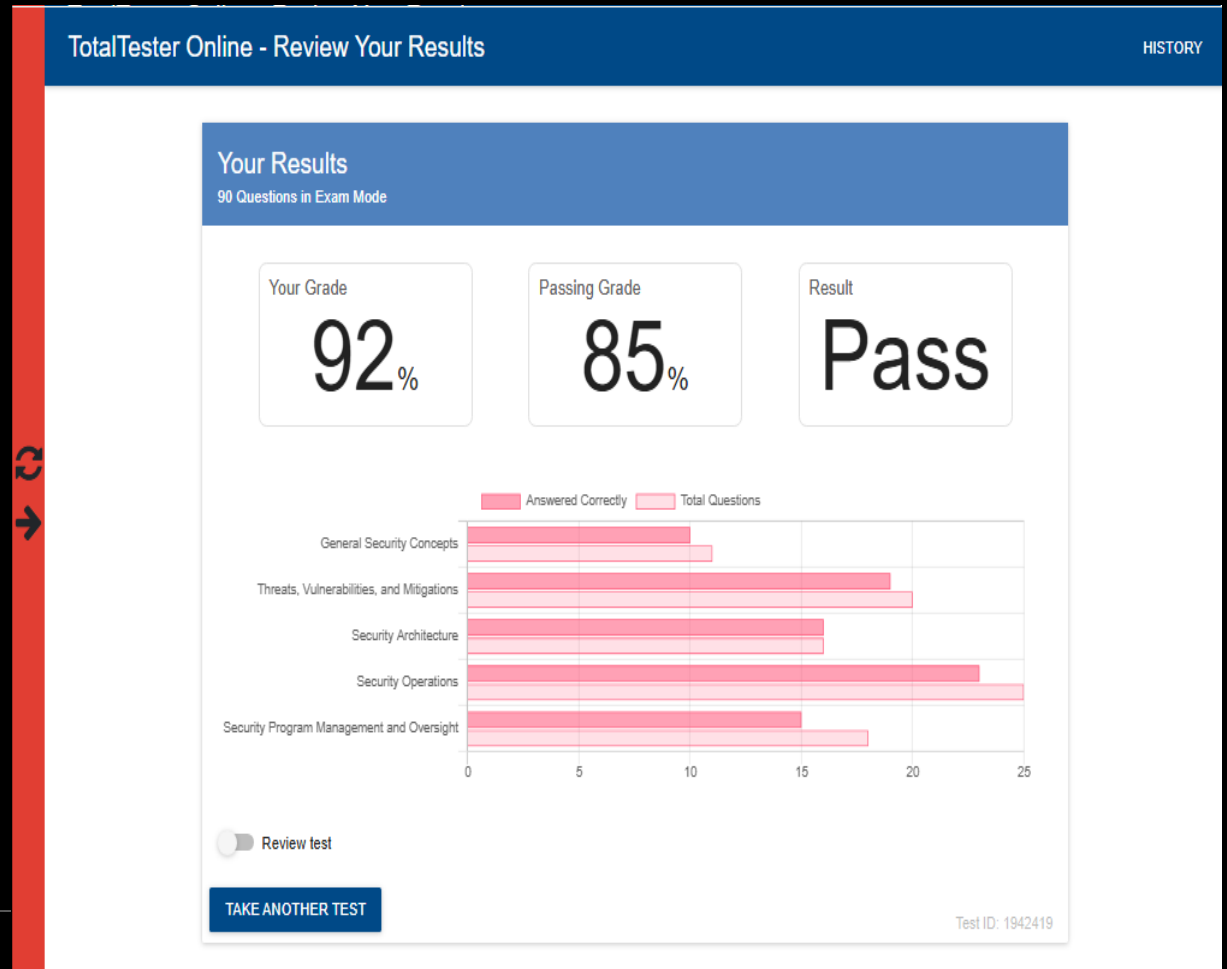
4. Ergebnisse der Modelle

- gemma3:27b
- Remote
- Mit ChromaDB
 - + Fragen integriert
 - + Lehrbuch embeddings



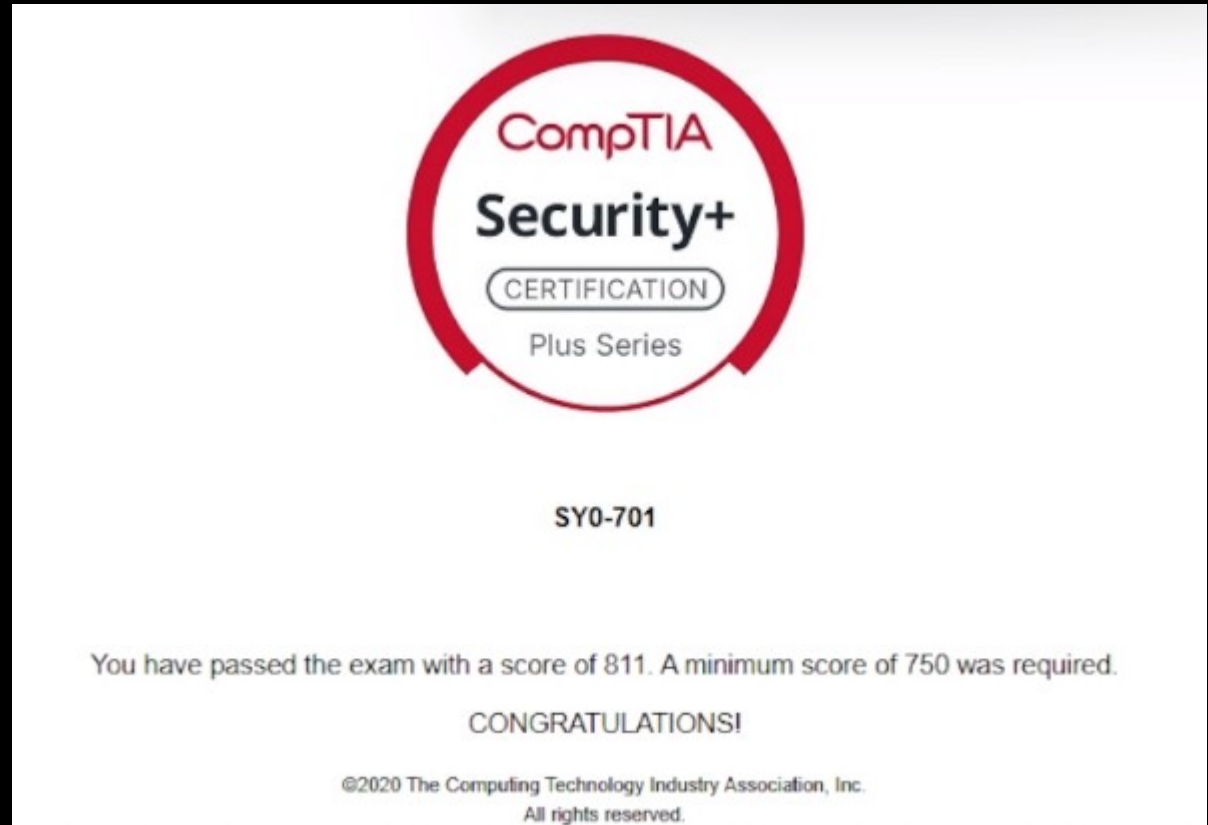
4. Ergebnisse der Modelle

- gemma3:27b
- Remote
- Mit ChromaDB
 - + Fragen integriert
 - + Lehrbuch embeddings



4. Ergebnisse der Modelle: Examen

- gemma3:27b
- Remote
- Mit ChromaDB
 - + Fragen integriert
 - + Lehrbuch embeddings



5. Erkenntnisse

- Ohne Optimierung hat keines der getesteten LLMs die Prüfung bestanden
- Die drei komplexen Anwendungsfälle konnten nicht gelöst werden
- Die iterative Integration von falsch beantworteten Fragen führt nach und nach zur Verbesserung
- Weitere Verbesserung ergab die Integration des Lehrbuchs

6. Diskussion

- Sollte eine Prüfung so umgangen werden?
- Welchen Wert haben die Zertifikate/Prüfungen noch, wenn der Mensch nur noch „human in the loop“ ist?
- Ist die Art der Prüfungsform sinnvoll?
- Wie sieht die zukünftige Arbeit von Mensch und Maschine aus?
- Eignen sich die Ergebnisse zur praktischen Nutzung?

DANKE