Uporaba alata NotebookLM-a za pomoć pri rješavanju CTF-ova

Datum: 6.10.2024

Procedura istraživanja:

- predstavljanje alata (što je, kakav je to alat, čiji je, kome je namijenjen)
- predstavljanje mogućnosti (što alat može/koje su mu mogućnosti, prednosti/mane)
- *edge* testing (testiranje rubnih slučajeva, mogućnosti)
- prezentacija (prolazak jednog CTF natjecanja i testiranje alata u stvarnoj situaciji - emuliranje CTF-a)

Valjalo bi objasniti što je **NotebookLM**.

Izvor: https://support.google.com/notebooklm/answer/14273541?hl=en
*NotebookLM gives you a personalized AI collaborator that helps you do your best thinking. After uploading your documents, NotebookLM becomes an instant expert in those sources so you can read, take notes, and collaborate with it to refine and organize your ideas.

Izvor: https://www.fastcompany.com/91143291/googles-notebooklm-is-a-great-tool-for-adding-ai-to-your-notes

After querying NotebookLM, you can save its responses and build on them to make new notes. These reside on your noteboard, along with other AI queries and new notes you make. If you're writing something new or preparing a presentation, you can use NotebookLM to assist you in exploring your materials. You can have a dialogue with your own notes.

Osnovno o alatu:

- Project Tailwind (u zadnji trenutak je ime izmjenjeno)
- prototip
- izgrađen na osnovi Gemini-a (Googlov LLM, v.1.5 Pro)
- prva verzija u javnost "puštena" sredinom 2023 no samo za Američko tržište (https://www.theverge.com/23845856/google-notebooklm-tailwind-ai-notes-research)
- u Lipnju ove godine proširen je support za ostale zemlje (https://medium.com/@HacktheCost/googles-updated-ai-powered-notebooklm-now-available-across-india-uk-and-200-nations-8382a02a5395)
- namijenjen za studente i generalno ljude kojima treba pomoć izvući korisne informacije iz velike količine podataka

Značajke (https://support.google.com/notebooklm/answer/14276468?hl=en&ref_topic=14272891,14272180,&visit_id=638638406349495188-1158215411&rd=1):

- podržava 50 izvora
- parsira izvore do 500 000 riječi ili 200MB za uploadane datoteke

- izvori mogu biti:
    - Google Docs
    - Google Slides
    - PDF, Text i Markodown datoteke
    - linkovi (samo će tekst sa stranica biti parsiran, slike i videi neće, stranice koje nemaju web scrapping ili "*paywall*" neće biti parsirane)
    - Zalijepljen tekst
    - YouTube linkovi (transkript javnih videa koji ne krše ToS YT-a će biti parsiran i to tek za videe starije od 72 sata)
    - Audio datoteke
- odgovori NotebookLM-a mogu biti spremljeni zajedno u bilješke u samoj aplikaciji
- citati spremljeni u bilješke, citati vode do dijelova iz kojih su izvedeni u izvornom tekstu
- nakon uploada izvora i dalje možemo limitirati na što će se LM fokusirati
- bilježnice se mogu podijeliti sa drugim osobama (kao viewer/editor)
- LM može sugestirati pitanja na temelju izvora i prethodnih pitanja

U daljnjem testiranju:

- napraviti nekoliko svojih bilježnica te testirati koliko toga LM:
    - zna bez ikakvih izvora
    - što sve može uz pružene izvore
    - za kraj napraviti "needle-in-a-haystack" test (test u kojem se unutar velikog teksta vezanog za jednu temu ubaci informacija nevezana za tu temu - https://arize.com/blog-course/the-needle-in-a-haystack-test-evaluating-the-performance-of-llm-rag-systems/)

Pri istraživanju teme se kao dobra alternativa pokazao alat AnythingLLM koji lokalno pokreće LM izbora (Llama 3 model bi bio dobar kandidat) https://anythingllm.com/ te u svojoj suštini radi sve isto što i NotebookLM bez potencijalnih ograničenja (naravno sada osobno računalo preuzima obrađivački dio posla)

Datum: 7.10.2024

Prije svega valjalo bi testirati osnovno znanje LM-a kako na engleskom tako i na hrvatskom jeziku.

Nije moguce razgovarati sa LM-om bez postavljanja izvora pa sam prije svega postavio datoteku sa brojem 22 u nadi da cu dobiti nesto vise od samo toga.
Testiranje predznanja na engleskom jeziku:
JA > "Disregard my source file. What can you tell me about yourself?"
LM > "The sources and our conversation history do not contain information about you, so I cannot answer your question."

JA > "Sto imas u svojim izvorima?"
LM > "The information available in the sources you have provided is limited to a single excerpt:

- The excerpt from "Pasted Text" contains only the number **22**. "

JA > "Odgovaraj mi na hrvatskom"
LM > "The sources you provided do not contain any information that would allow me to respond to your query in Croatian.
The available information in our conversation history only confirms that your sources contain a single numerical excerpt, "22", from a source labeled "Pasted Text"[1]"

Daljnji pokuaji da odgovori na hrvatskom su zavrsili neuspjehom iako razumije upite na hrvatskom.

U nastavku pokusavamo dobiti znanje koje nije eksplicitno upisano.

JA > Write a short script that results with said number.
LM > [kratki igrokaz o kapetanu i inzinjeru na svemirskom brodu]

JA > script in python and bash
LM >
Based on your previous input, the number you are interested in is **22** . Below are scripts in Python and Bash that will output this number:

Python

```
print(22)
```

This script uses the `print()` function to output the number 22 directly.

Bash

```
echo 22
```

This script utilizes the `echo` command to print the number 22 to the console.
""

Znaci bez dodatnih izvora LM ima nekakvo predznanje o jezicima.

JA - Can you offer some sources (links, books, videos...) that mention 22
LM -
The sources and our conversation history do not contain any links, books, or videos that mention 22. The only available information is the number **22** itself, originally provided by you.

Osim toga ne može ništa više.

Datum 11.10.2024

Rješavanje primjernog međuispita 2022 iz SRS-a koristeći prošlogodišnje prezentacije.

Nakon predanih prezentacija i PDF-a ispita vraćen je sljedeći odgovor:

```
Let's work through the questions in the document you provided, drawing on the
provided source materials, to enhance your understanding of the subject.

## Solution to "srs_mi_vjezba[1].pdf" Questions

### Question 1

**a) When one of the security requirements is no longer met, we say that an
________________ has occurred.**

**Answer:** Incident [2]

**Explanation:** This question is taken directly from the document.

**b) What are the three basic security requirements?**

**Answer:** Confidentiality, Integrity, Availability [2]

**Explanation:** These three security requirements are often referred to as the
**CIA Triad**. They form the foundation of security principles.

- **Confidentiality** ensures that information is accessible only to authorized
individuals.
- **Integrity** guarantees the accuracy and trustworthiness of information,
ensuring it hasn't been tampered with.
- **Availability** ensures that information and resources are accessible to
authorized users when needed.

**c) The security requirement that data must not be seen by unauthorized persons
is called ___________________.**

**Answer:** Confidentiality [2]

**Explanation:** This question builds on the previous question about the CIA
Triad.

### Question 2

**A backup tape stores business data and is then transported by courier to a
backup location to preserve the data in the event of a disaster at the primary
site. By doing so, which security requirement is met?**

**Answer:** Availability [2]

**Explanation:** By storing backup data at a separate location, the company
ensures data accessibility even if a disaster strikes the primary site. This
```

```
directly addresses the **availability** aspect of the CIA Triad.

### Question 3

**Match the external sources of security threats to their corresponding
descriptions:**

**a) Organized crime groups:**

**Answer:** Motivated by financial gain [3]

**Explanation:** This is a common understanding of organized crime. While the
sources don't explicitly link organized crime to financial gain in this context,
it's widely accepted that financial profit is a primary motivator for such groups.

**b) Cybercrime:**

**Answer:** Criminal activities that target computer networks or devices. [3]

**Explanation:** Cybercrime encompasses a wide range of criminal activities
carried out using computers and networks. These activities often aim to steal
data, disrupt services, or cause financial harm.

**c) Hacktivists:**

**Answer:** Motivated by political or social goals [3]

**Explanation:** Hacktivists use hacking techniques to promote a political or
social agenda. Their actions can range from website defacement to data leaks and
service disruptions.

### Question 4

**Static analysis is:**

**Answer:** Analysis of source code to detect security vulnerabilities [3]

**Explanation:** Static analysis involves examining the source code of a program
without executing it. This allows security professionals to identify potential
vulnerabilities and weaknesses in the code structure.

### Question 5

**You wrote the following code in a programming language:**

x = 3;
```

y = 'SRS';

z = x + y;

If this is a strongly typed programming language that performs runtime type checking, what will happen?

Answer: Runtime error [3]

Explanation: Strongly typed languages enforce strict data type rules. In this case, adding an integer (x) to a string (y) would violate these rules, resulting in a runtime error.

Question 6

You wrote the code:

1: ispitProlaz = true;

2: try {

3: points = provjeriBodove();

4: if (bodovi < 50){

5: ispitProlaz = false;

6: }

7: }

8: catch (Exception ex)

9: {

10: //write error

11: }

If the method provjeriBodove() can throw an error, what should be fixed in the code to make it work correctly and to satisfy the Fail Securely principle?

Answer: We should replace the expression on line 1 with ispitProlaz = false; and add the condition else { ispitProlaz = true; } after line 6. [4]

Explanation: The "Fail Securely" principle dictates that in the event of an error or failure, the system should default to a secure state. By initializing ispitProlaz to false and only changing it to true after successful execution, we ensure that any error during the process will maintain a secure state (failing the exam). The else statement added after the check (bodovi < 50) ensures that if the check passes (meaning the score is 50 or above), the ispitProlaz flag is set to true.

Question 7

Listing files you see that the file srs_ispit.txt has permissions r w - r- - r - x set. Will external users (others) be able to read the file?

Answer: Yes [5]

Explanation: In Unix-like operating systems, file permissions are represented by a series of characters. The permissions are divided into three groups: Owner, Group, and Other. The "r" signifies read access. In this case, the file permissions "r w - r- - r - x" indicate that external users ("others") have read permission ("r").

Question 8

Young Stefan wrote the following message on his blog: "I have developed a new super-secret cryptographic algorithm that is completely secure because no one knows how it works." Based on the excerpt, we can conclude that young Stefan is not aware of:

Answer: Kerckhoffs's principle [6]

Explanation: Kerckhoffs's principle states that the security of a cryptosystem should rely solely on the secrecy of the key, not the secrecy of the algorithm. Stefan's belief that his algorithm is secure because no one knows how it works directly contradicts this principle.

Question 9

Which of the following are valid key pairs for the "Plain RSA" cryptosystem?

Answer:

- **(5,65),(1,65)**
- **(3,65),(29,65)** [6]

Explanation: Valid key pairs for the "Plain RSA" cryptosystem are those where the public exponent (e) and private exponent (d) satisfy the following equation:

(e * d) MOD φ(N) = 1

where N is the modulus and φ(N) is Euler's totient function of N. For the given options, only the following pairs satisfy the equation:

- **(5,65),(1,65)**: Here, N = 65, φ(N) = 48. You can check that both (5 * 1) MOD 48 = 1 and (1 * 5) MOD 48 = 1 hold true.
- **(3,65),(29,65)**: Here, N = 65, φ(N) = 48. You can check that (3 * 29) MOD 48 = 1 holds true.

Question 10

Given is the following pseudo-code which aims to encrypt a message of arbitrary length with a symmetric cipher:

Input:

- m is the message to be encrypted
- k is the symmetric key of size 16 bytes

Encryption procedure:

1. Pad the message m so that its size is a multiple of 16 bytes;
2. Calculate SHA256(m) and append those 32 bytes to the end of the message m;
3. iv = randomly selected 16 bytes;
4. m1, m2, …, mk = decomposition of m into blocks of size 16 bytes;
5. for each i=1, …, k calculate ci = AES128(iv+i, k) XOR mi (+ denotes ordinary addition);
6. join the blocks iv, c1, c2, …, ck into ciphertext c;

Questions:

a) Describe the corresponding decryption procedure.

Answer:

1. Extract iv (first 16 bytes) and ciphertext blocks c1, c2, …, ck from the ciphertext c.
2. For each block i = 1, ..., k, calculate mi = AES128(iv + i, k) XOR ci.
3. Concatenate blocks m1, m2, …, mk to obtain m'.
4. Extract the last 32 bytes of m' as h', which is the purported hash.
5. Remove the last 32 bytes from m' to obtain the original message m.
6. Calculate h = SHA256(m).
7. Compare h and h'. If they are equal, the decryption is successful and the message integrity is confirmed. Otherwise, the message has been tampered with.

b) Does this encryption procedure provide confidentiality? If yes, explain why, if not, describe in detail a scenario where confidentiality is compromised.

Answer: Yes, this encryption procedure provides confidentiality because the message is encrypted using AES128, a strong symmetric encryption algorithm. The key k is used in the encryption process, and without knowledge of this key, an attacker cannot decrypt the ciphertext. The use of a randomly generated initialization vector (iv) ensures that identical plaintext blocks encrypt to different ciphertext blocks, further enhancing confidentiality.

c) Does this encryption procedure provide integrity? If yes, explain why, if not, describe one scenario where integrity is compromised.

Answer: Yes, this encryption procedure provides integrity through the use of the SHA256 hash function. Appending the hash of the message to the message itself allows the receiver to verify if the message has been altered during transmission. Upon decryption, the receiver calculates the hash of the decrypted message and compares it with the received hash. If the hashes match, the integrity of the message is confirmed.

d) If the procedure does not provide both properties, describe one way of encrypting a message of arbitrary length so that both properties are satisfied.

Answer: The provided procedure already provides both confidentiality and integrity. However, if we wanted to explore alternative ways to achieve both properties, one common approach is to use **authenticated encryption** modes like GCM (Galois/Counter Mode) or EAX (Extended Authentication Code). These modes combine encryption and authentication into a single step, simplifying the process and offering robust security guarantees.

Question 11

Given is the following pseudo-code which aims to encrypt a message of arbitrary length with an asymmetric cipher:

Input:

- m is the message to be encrypted
- (e, N) is the public RSA key, N has 2048 bits

Encryption Procedure:

1. Pad the message m so that its size is a multiple of 16 bytes;
2. m1, m2, …, mk = decomposition of m into blocks of size 16 bytes;
3. for each i=1, …, k calculate ci = mi^e MOD N;

```
4. join the blocks c1, c2, …, ck into c;

**Questions:**

**a) Describe the corresponding decryption procedure.**

**Answer:**

1. Extract ciphertext blocks c1, c2, …, ck from the ciphertext c.
2. For each block i = 1, ..., k, calculate mi = ci^d MOD N where d is the private RSA key.
3. Concatenate blocks m1, m2, …, mk to obtain the original message m.

**Important Notes:**

- The sources provided don't cover every concept discussed in the exercise document, especially regarding operational questions about specific network protocols and tools.
- For further understanding of these topics, it's recommended to consult additional resources.

This detailed breakdown should enhance your understanding of the concepts and solutions presented in "srs_mi_vjezba[1].pdf".
```

Sljedeći tjedan:

- testiranje rubnih slučajeva (needle-in-a-haystack, common sense, kontradiktorni izvori)
- izrada prezentacije

Datum 18.10.2024

Testiranje rubnih slučajeva:

Needle-In-A-Haystack

https://github.com/gkamradt/LLMTest_NeedleInAHaystack/tree/main/needlehaystack/PaulGrahamEssays (pristup. 18.10.)
Korišteni tekstovi sa ove stranice:

- addiction.txt
- love.txt

U oba teksta umetnuta izjava: "*Green apples are commonly known as the healthiest apples of the bunch.*"

Tekstovi zasebno testirani
Ispitno pitanje: Which apples are the healthiest?

U love.txt izraz je ubačen nakon 1. rečenice sedmog (7.) poglavlja.
Odgovor:
`The sources provided focus on career advice and do not contain information about the health benefits of different apple varieties.`

U addiction.txt izraz je ubečen nakon 1. rečenice petog (5.) poglavlja.
Odgovor:
`The sources provided discuss the increasing addictiveness of the world due to technological progress, but they do not mention anything about the health benefits of different types of apples.1234... Source 5 does mention that green apples are commonly known as the healthiest apples, but it does not provide any evidence or further explanation for this claim.10

Kao što je vidljivo u ovom (kraćem tekstu) LM može pronaći "needle" u tekstu.

Razlike u tekstovima:
addiction.txt (prije dodavanja fraze):

- znakova: 7321
- riječi: 1254
- linija: 118

love.txt (prije dodavanja fraze):

- znakova: 25142
- riječi: 4600
- linija: 378

Razlike:

- znakova: 17821
- riječi: 3346
- linija: 260

Očito u većim tekstovima LM teže pronalazi "iglu".
Nama to ne bi trebao predstavljati problem jer ionako ne namjeravamo predstavljati prevelike tekstove.

Common sense - zdrav razum

https://deepgram.com/learn/hellaswag-llm-benchmark-guide
by Brad Nikkel (pristupljeno 18.10.)

Koristiti ćemo *HellaSwag* benchmark test.
Pitanja kakva se nalaze na takvom testu:

```
John opened the refrigerator and saw a carton of milk. He picked it up and:
a) poured it into his car's gas tank.
b) used it to water his houseplants.
c) sniffed it to check if it was still fresh.
d) spread it on his toast like butter.

Sarah was running late for her flight. As she arrived at the airport, she:
a) decided to take a nap in the parking lot.
b) rushed to the check-in counter with her luggage.
c) started cooking a three-course meal.
d) began painting a landscape of the runway.

The chef was preparing a salad and needed to cut some tomatoes. He reached for:
a) a hammer and some nails.
b) a sharp kitchen knife.
c) a pair of scissors and glue.
d) a blowtorch and safety goggles.

As the sun began to set, the campers realized they needed light. They decided to:
a) plant some trees and wait for them to grow.
b) paint the tent with glow-in-the-dark paint.
c) start a small, controlled campfire.
d) call a taxi to drive them to the nearest city.

The little boy's tooth was loose and about to fall out. His parents told him to:
a) tie it to a rocket and launch it into space.
b) swallow it with his breakfast cereal.
c) wiggle it gently and wait for it to come out naturally.
d) use superglue to keep it in place permanently.

During the winter storm, the power went out. The family immediately:
a) started a bonfire in the living room.
b) got out flashlights and candles for light.
c) called the zoo to borrow some penguins.
d) began digging a tunnel to the neighbor's house.

At the fancy restaurant, the waiter approached with the bill. The customer:
a) ate the bill with a knife and fork.
b) used it as a paper airplane and flew it back to the kitchen.
c) reviewed it and prepared to pay with a credit card.
d) requested that it be translated into hieroglyphics.

As the car's engine started making strange noises, the driver:
a) turned up the radio to drown out the sound.
b) pulled over to the side of the road safely.
c) opened the hood and poured in some orange juice.
```

```
d) called a psychic for car advice.

When the teacher asked for homework, one student realized he had forgotten his.
He:
a) claimed his dog had eaten it, then barked to demonstrate.
b) offered to perform an interpretive dance instead.
c) apologized and asked for an extension to complete it.
d) suggested the class take a field trip to find it.

As the basketball player prepared for a free throw, he:
a) took off his shoes and threw them at the hoop.
b) dribbled the ball and focused on his shot.
c) asked a spectator to stand on his shoulders.
d) began reciting the periodic table of elements
```

Očito ljudi na ova pitanja mogu jednostavno odgovoriti no LLM-ovi sa ovakvim tipom pitanja znaju imati problema (ironično pitanja je generirao claude.ai tj. LLM).
Gore navedena pitanja pružena kao izvor.

Odgovor:

```
Here are the answers to the questions in the sources:

- **John would most likely sniff the milk to check if it was still fresh.** This
is the most logical action out of the given options.

- **Sarah would rush to the check-in counter with her luggage because she was late
for her flight.** The other options are not actions someone would take if they
were late for a flight.

- **The chef would reach for a sharp kitchen knife to cut the tomatoes for the
salad.** A knife is the appropriate tool for cutting tomatoes.

- **The campers would likely start a small, controlled campfire for light as the
sun began to set.** This is a common and safe way for campers to provide light and
warmth.

- **The little boy's parents would advise him to wiggle it gently and wait for his
loose tooth to come out naturally.** This is the typical advice given to children
with loose teeth.

- **The family would immediately get out flashlights and candles for light when
the power went out during a winter storm.** This is a practical response to a
power outage.

- **At a fancy restaurant, the customer would review the bill and prepare to pay
```

```
with a credit card when the waiter brings it.** This is the standard procedure for
paying at a restaurant.

- **As the car's engine starts making strange noises, the driver would pull over
to the side of the road safely.** This is a safe and responsible action to take
when a car malfunctions.

- **The student who forgot his homework would apologize and ask for an extension
to complete it.** This is a responsible way to handle the situation.

- **As the basketball player prepared for a free throw, he would dribble the ball
and focus on his shot.** This is the typical routine for a basketball player
taking a free throw.
```

Iznenađujuće točni i razumni odgovori sa argumentima.

Kontradiktorni izvori

U izvorima ćemo predstaviti dvije kontradiktorne tvrdnje koje nisu dio "zdravog razuma" što znači da LM neće imati način provjeriti istinitost tvrdnji.
Tvrdnje:

- Actions speak louder than words" vs. "The pen is mightier than the sword"
- "Look before you leap" vs. "He who hesitates is lost"
- "Absence makes the heart grow fonder" vs. "Out of sight, out of mind"

Svaka tvrdnja biti će stavljena kao zasebni izvor (sve u istu bilježnicu).

Pitanja:
Which one matters more: Actions or words?
`It depends on the context.`

Should I take a leap of faith on a decision or explore all possibilities?
`It depends on the situation.`

Does absence of a thing or a person makes me love it/them more?
`It likely depends on various factors such as the nature of the relationship, the individuals involved, and the reasons for the absence.`

Uglavnom LM nema bias te ne teži ni jednom izvoru više nego nekom drugom.
Tretira sve izvore jednako.

Napomena: zapravo se dosta još može testirati (kako naslovi izvora utječu na izbor i parsiranje izvora, politička pristranost, ilegalni izvori poput povjerljivih dokumenata?....)
No kako to može ići u nedogled to nećemo raditi.

Za napraviti:

- prezentacija koja objedinjuje cjelokupno istraživanje
- pokazati primjere na prezentaciji (osmisliti protok testiranja uz prezentiranje)