

LASER @ 40th Annual Computer Security Applications Conference (ACSAC)

Benchmarking Large Language Models: From General Purpose to Cybersecurity Applications

Presented by: Dipkamal Bhusal db1702@rit.edu

PhD Student, Rochester Institute of Technology

Rochester, NY, USA

RIT | Rochester Institute of Technology

Introduction: Large Language Models (LLM)

- **Language Model:** A language model predicts and generates language by estimating the probability of a token or sequence of tokens within a context.
  - Eg. RNN, Transformer
- **Large Language Models:** Large-scale models trained on massive datasets with billions of parameters, capable of predicting entire paragraphs or generating coherent text.

Introduction: Large Language Models (LLM)

- **Language Model:** A language model predicts and generates language by estimating the probability of a token or sequence of tokens within a context.
  - Eg. RNN, Transformer
- **Large Language Models:** Large-scale models trained on massive datasets with billions of parameters, capable of predicting entire paragraphs or generating coherent text.



OpenAI

Claude

MISTRAL AI_

Gemini

Qwen

How do LLMs work?

My name is ___ →  → Dipkamal

How do LLMs work?

My name is Dipkamal ___ →  → and

How do LLMs work?

My name is Dipkamal and ___ →  → I

How do LLMs work?

My name is Dipkamal and I ___ →  → am

How do LLMs work?

My name is Dipkamal and I am ___ →  → presenting

How do LLMs work?

My name is Dipkamal and I am presenting ___ →  → at

How do LLMs work?

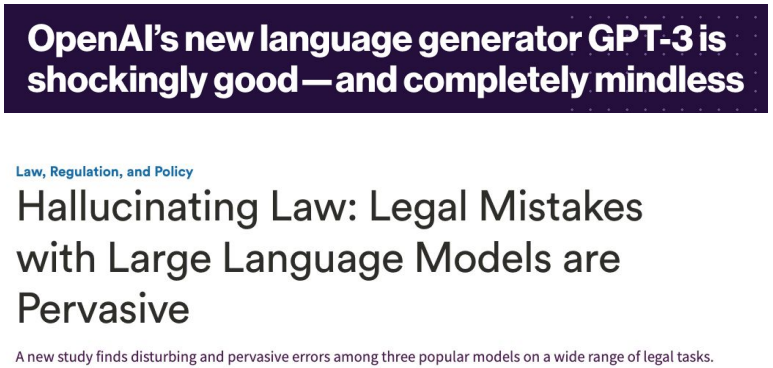
My name is Dipkamal and I am presenting at ___ →  → LASER.

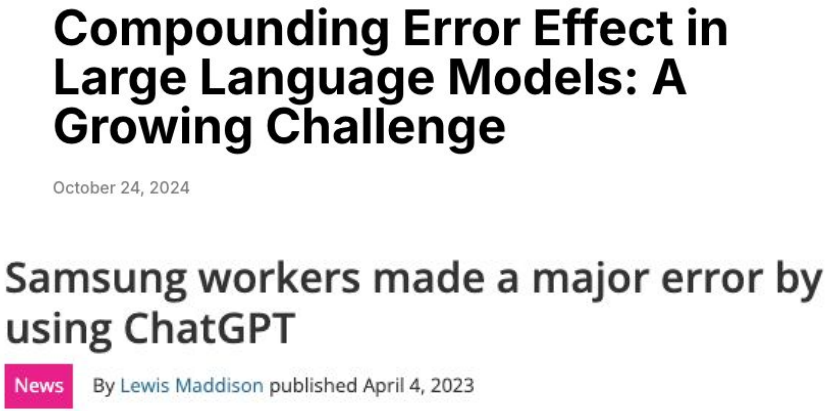
How do LLMs work?

1. LLMs are trained in two key steps: generative pre-training and instruction fine-tuning
   a. **Generative Pre-training:** The model learns to predict the next token in a sequence using large-scale unlabeled text data.
   b. **Instruction Fine-tuning:** The model is refined to follow task-specific instructions, improving its ability to align with human intent and perform specific tasks.
2. LLMs can solve new tasks through task-agnostic few-shot (using examples) or zero-shot (instructions only) learning.
   a. “Scaling up language models improves task-agnostic, few-shot performance.” (Language Models are Few-Shot Learners, 2020)

Reliability Concerns with LLMs

- **Susceptibility to Errors**
- **Hallucination**
- **Lack of Verifiability**





What is benchmarking?

Why is benchmarking LLMs necessary?

Benchmarking LLMs

- **Evaluate LLMs on standard tasks**
  - Assess model performance.
- **Desiderata of benchmarks:**
  - Clear downstream implications
    - **Domain expertise**
  - Interpretable metric

General Benchmarks

- **General Benchmarks**
  - GLUE (ICLR'19), MMLU (ICLR'21), KOLA (ICLR'23)
    - Focused on language comprehension, or subject-specific evaluation.
  - **Unsuitable for benchmarking practical cybersecurity fields**

GLUE: A multi-task benchmark and analysis platform for natural language understanding, 2018
MMLU: Measuring massive multitask language understanding, 2020
Kola: Carefully benchmarking world knowledge of large language models, 2023

Our work

- Provide a comprehensive evaluation framework of LLMs in a practical cybersecurity field
  - **SECURE —> Industry Control System (accepted at ACSAC'24)**
  - **CTIBench —> Cyber Threat Intelligence (accepted at NeurIPS'24)**
- Both evaluation framework are based on knowledge-based modeling
  - **Extraction/Memorization**
    - *Recall and utilize previously learned information*
  - **Understanding**
    - *Comprehend the content and context*
  - **Problem-solving**
    - *Apply knowledge to address specific challenges*
  - **Reasoning**
    - *Draw logical conclusions and make informed decisions*

Task 1: Knowledge Extraction

- Accurately retrieve specific information from the knowledge base
  - Also known as “memorization”
- Do we need memorization in LLMs?
  - Yes, we do!
  - Remember the factual knowledge from the training.

Task 1: Knowledge Extraction

- Accurately retrieve specific information from the knowledge base
  - Also known as “memorization”
- Do we need memorization in LLMs?
  - Yes, we do!
  - Remember the factual knowledge from the training.

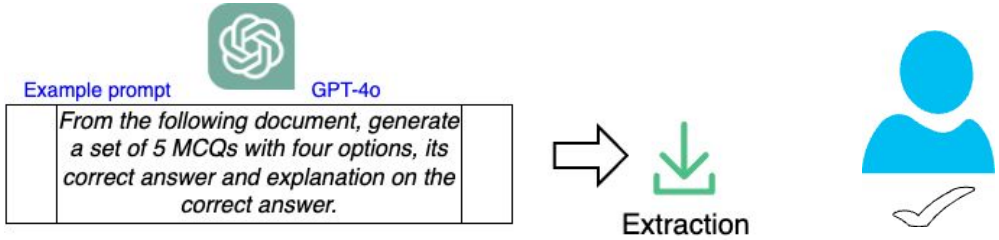
Who is Lionel Messi? →  → Lionel Messi is an Argentine professional footballer often regarded as the GOAT.

How do we design memorization tasks?

- Employing a proprietary LLM like ChatGPT-4
  - A research standard in generating evaluation benchmarks.
  - OpenAI’s ChatGPT-4o
- Requirement:
  - Reliable data sources
  - Suitable prompt design
  - Human validation

How do we design memorization tasks?

- Employing a proprietary LLM like ChatGPT-4
  - A research standard in generating evaluation benchmarks.
  - OpenAI's ChatGPT-4o
- Requirement:
  - Reliable data sources
  - Suitable prompt design
  - Human validation



How do we design memorization tasks?

- **Data sources**
  - SECURE:
    - MITRE ATT&CK
      - Attack patterns and mitigation plans for ICS
    - CWE (Common Weakness Enumeration)
      - CWE-1358 “…..Security Vulnerabilities in ICS”
  - CTIBench
    - CTI frameworks and regulations such as GDPR
    - CTI standards such as STIX and TAXII
    - MITRE ATT&CK framework
    - CWE (Common Weakness Enumeration) database
    - More.

How do we design memorization tasks?

- **Optimize prompt for creating questions.**
  - Trial and error
  - Few principles
    - Provide the document
    - Split the task into simpler subtasks
    - Include as much details in the query
      - *Be aware of the context size (input token limit)*
    - Use delimiters to clearly indicate distinct parts of the input

Prompt for generating MCQs with ChatGPT-4o from MITRE ATT&CK
(Italized parts are unique to MITRE ATT&CK)

Role

You are a cybersecurity expert specializing in cyber threat intelligence. Given the text below, please generate a maximum of 5 multiple-choice questions with four possible options each.

Follow these requirements:
1. Question Format: Each question must have four options. The options should be challenging and require careful consideration. Avoid creating options that could be interpreted as correct under different circumstances.
2. Target Audience: The questions should be suitable for security professionals with three to five years of experience in cyber threat intelligence. *Avoid generic questions such as "What is the objective?", "Which operating system can be targeted?".*
3. Content Coverage: Aim to cover various sections of the document to ensure a comprehensive evaluation of the candidate's knowledge. Include context-specific questions that require an understanding of the document's content.
4. Technical Precision: Use precise terminology and concepts relevant to cyber threat intelligence. Incorporate situational or scenario-based questions where applicable.
5. *Include Technique IDs and Names: Ensure that all questions, where applicable, mention both the ID and the full name of the MITRE ATT&CK pattern technique.*
6. *Premise Inclusion: Each question should include a premise indicating it pertains to MITRE ATT&CK, specifying the relevant platform (Enterprise, ICS, or Mobile) where necessary.*
7. Output Format: Return the output in TSV format (must be tab-separated) with the following columns: Question, Option A, Option B, Option C, Option D, Correct Answer (A, B, C, D), and Explanation.

Important: Only return the TSV content as specified. Do not include any additional text or commentary outside the TSV format.

Text:
...

Prompt for generating MCQs with ChatGPT-4o from MITRE ATT&CK
(Italized parts are unique to MITRE ATT&CK)

You are a cybersecurity expert specializing in cyber threat intelligence. Given the text below, please generate a maximum of 5 multiple-choice questions with four possible options each.

Follow these requirements:
1. Question Format: Each question must have four options. The options should be challenging and require careful consideration. Avoid creating options that could be interpreted as correct under different circumstances.
2. Target Audience: The questions should be suitable for security professionals with three to five years of experience in cyber threat intelligence. *Avoid generic questions such as "What is the objective?", "Which operating system can be targeted?".*
3. Content Coverage: Aim to cover various sections of the document to ensure a comprehensive evaluation of the candidate's knowledge. Include context-specific questions that require an understanding of the document's content.
4. Technical Precision: Use precise terminology and concepts relevant to cyber threat intelligence. Incorporate situational or scenario-based questions where applicable.
5. *Include Technique IDs and Names: Ensure that all questions, where applicable, mention both the ID and the full name of the MITRE ATT&CK pattern technique.*
6. *Premise Inclusion: Each question should include a premise indicating it pertains to MITRE ATT&CK, specifying the relevant platform (Enterprise, ICS, or Mobile) where necessary.*
7. Output Format: Return the output in TSV format (must be tab-separated) with the following columns: Question, Option A, Option B, Option C, Option D, Correct Answer (A, B, C, D), and Explanation.

Important n-
mentary ou

Text:
...

Detail instructions, task divided into sub-tasks

Prompt for generating MCQs with ChatGPT-4o from MITRE ATT&CK
(Italized parts are unique to MITRE ATT&CK)

You are a cybersecurity expert specializing in cyber threat intelligence. Given the text below, please generate a maximum of 5 multiple-choice questions with four possible options each.

Follow these requirements:
1. Question Format: Each question must have four options. The options should be challenging and require careful consideration. Avoid creating options that could be interpreted as correct under different circumstances.
2. Target Audience: The questions should be suitable for security professionals with three to five years of experience in cyber threat intelligence. *Avoid generic questions such as "What is the objective?", "Which operating system can be targeted?".*
3. Content Coverage: Aim to cover various sections of the document to ensure a comprehensive evaluation of the candidate's knowledge. Include context-specific questions that require an understanding of the document's content.
4. Technical Precision: Use precise terminology and concepts relevant to cyber threat intelligence. Incorporate situational or scenario-based questions where applicable.
5. *Include Technique IDs and Names: Ensure that all questions, where applicable, mention both the ID and the full name of the MITRE ATT&CK pattern technique.*
6. *Premise Inclusion: Each question should include a premise indicating it pertains to MITRE ATT&CK, specifying the relevant platform (Enterprise, ICS, or Mobile) where necessary.*
7. Output Format: Return the output in TSV format (must be tab-separated) with the following columns: Question, Option A, Option B, Option C, Option D, Correct Answer (A, B, C, D), and Explanation.

Important: Only return the TSV content as specified. Do not include any additional text or commentary outside the TSV format.

Text:
...

Output format

How do we design memorization tasks?

- Manually analyze the quality of MCQs
  - Human annotators have access to the source data.
- Two major issues from LLM-generated questions:
  - Multiple correct options in the answers
  - LLM-given answer was incorrect.
- Fix or Remove

Example dataset

Question	Option A	Option B	Option C	Option D	Answer
What is a primary purpose of setting up a rogue master in an ICS environment?	Sending legitimate control messages to devices	Intercepting internal communications	Encrypting network traffic	Creating network segmentation	A
Which technique is addressed by restricting registry permissions in ICS?	Phishing	Denial of Service	Unauthorized Registry Modifications	SQL Injection	C

Task 2: Knowledge Understanding

- Do LLMs understand the security-related text?
- Multiple ways of creating such task, eg.
  - Boolean
  - Text comprehension

Task 2: Knowledge Understanding

- **SECURE**
  - **Data source:** CVEs 2024 (Common Vulnerabilities and Exposures)
  - **Task:** Based on the given CVE information, ask the LLM to judge the truthfulness of the statement.
    - LLMs did not use these CVEs during their training.
  - ChatGPT-4o + Human Validation

Task 2: Knowledge Understanding

- **Datasets: Two versions**
  - **One providing the context**
    - Provide the CVE details and ask the LLM to judge truthfulness of the statement.
  - **One without the context**
    - Do not provide the CVE details and ask the LLM to judge the truthfulness.
    - Out-of-distribution detection
      - The model should indicate its inability to answer.

CVE-ID	Question	Answer
CVE-2024-0011	The vulnerability described in CVE-2024-0011 allows for the execution of arbitrary code on the affected system.	FALSE
CVE-2024-0017	User interaction is required for the exploitation of the CVE-2024-0017 vulnerability.	TRUE

Task 2: Knowledge Understanding

- **CTI-Bench**
  - **A root-cause mapping task**
    - Identify the underlying cause(s) of a vulnerability by correlating CVE records with CWE entries.
  - Manually collected data from the National Vulnerability Database (NVD)

Task 2: Knowledge Understanding

- **CTI-Bench**
  - **A root-cause mapping task**
    - Identify the underlying cause(s) of a vulnerability by correlating CVE records with CWE entries.
  - Manually collected data from the National Vulnerability Database (NVD)

Task 2: Knowledge Understanding

In BubbleUPnP 0.9 update 30, the XML parsing engine for SSDP/UPnP functionality is vulnerable to an XML External Entity Processing (XXE) attack. Remote, unauthenticated attackers can use this vulnerability to: (1) Access arbitrary files from the filesystem with the same permission as the user account running BubbleUPnP, (2) Initiate SMB connections to capture a NetNTLM challenge/response and crack the cleartext password, or (3) Initiate SMB connections to relay a NetNTLM challenge/response and achieve Remote Command Execution in Windows domains.



CWE-611: Improper Restriction of XML External Entity Reference

Task 3: Knowledge Problem-solving

- LLMs ability to compute CVSS score given a vulnerability detail or CVSS vector strings
- **SECURE**
  - CVSS3.1 vector strings from Cybersecurity and Infrastructure Security Agency (CISA)

Task 3: Knowledge Problem-solving

- LLMs ability to compute CVSS score given a vulnerability detail or CVSS vector strings
- **SECURE**
  - CVSS3.1 vector strings from Cybersecurity & Infrastructure Security Agency (CISA)
  - Zero shot evaluation of LLMs in using existing CVSS formula

CVSS v3 Vector String	Correct Answer
AV:L/AC:L/PR:N/UI:R/S:U/C:H/I:H/A:H	7.8
AV:N/AC:H/PR:N/UI:R/S:C/C:L/I:L/A:N	4.7
AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H	9.8

Task 3: Knowledge Problem-solving

- LLMs ability to compute CVSS score given a vulnerability detail or CVSS vector strings
- **SECURE**
  - CVSS3.1 vector strings from Cybersecurity & Infrastructure Security Agency
  - Zero shot evaluation of LLMs in using existing CVSS formula
- **CTIBench**
  - Mapping vulnerability description to CVSS3 scores
  - 1000 vulnerability descriptions 2024 from National Vulnerability Database (NVD)

Task 4: Knowledge Reasoning

- **CTIBench**
  - Threat actor attribution as **abductive reasoning**
    - Identify the groups responsible for a cyberattack analyzing threat intel
    - Form plausible conclusions from incomplete information
  - **Data source:**
    - Threat reports from vendors with Advanced Persistent Threat (APT) group attributed to the reports.
    - Removed all mentions of the threat actors and their associated malware campaign names, replacing them with placeholders.
  - **Task: Ask the LLMs to attribute the reports to known threat actors.**
    - Verify LLM response to account for the multiple aliases that threat actors use.

Evaluation: Prompt

- Consistent prompt engineering to ensure uniformity across LLMs.
  - *Instead of customizing prompts for models, a single prompt structure per task.*
- Maintain compatibility with model input constraints and avoid exceeding maximum token limits
  - *For eg. we ensured that prompts did not exceed 5000 words (with context)*
- Manually correct outputs where responses deviate from the instructions
- Set temperature parameter (eg. 0.7) for all evaluation prompts.

Analysis

- GPT-4 outperforms others in 2 tasks.
- Gemini-1.5 takes lead in problem-solving.
- Llama3-70B matches GPT-4 in reasoning.
- Smaller models (GPT3.5 and LLAMA-8B) perform well only in extraction tasks.

Tasks	Extraction (Accuracy)	Understanding (Accuracy)	Problem-solving (deviation; lower the better)	Reasoning (Accuracy)
ChatGPT-4	83	83	1.06	52
ChatGPT-3.5	73	51	1.42	44
Gemini-1.5	80	79	1.04	38
LLAMA3-70B	81	69	1.69	52
LLAMA3-8B	76	52	1.84	28

Average of performance on both the benchmarks

Analysis

- GPT-4 outperforms others in 2 tasks.
- Gemini-1.5 takes lead in problem-solving.
- Llama3-70B matches GPT-4 in reasoning.
- Smaller models (GPT3.5 and LLAMA-8B) perform well only in extraction tasks.
  - **Are larger models always better alternative to smaller ones?**

Tasks	Extraction (Accuracy)	Understanding (Accuracy)	Problem-solving (deviation; lower the better)	Reasoning (Accuracy)
ChatGPT-4	83	83	1.06	52
ChatGPT-3.5	73	51	1.42	44
Gemini-1.5	80	79	1.04	38
LLAMA3-70B	81	69	1.69	52
LLAMA3-8B	76	52	1.84	28

Average of performance on both the benchmarks

Analysis

- On understanding and problem-solving tasks, closed source models outperform open source models.
- Responses from LLMs with lower confidence levels tend to be less accurate.
- Explicitly asking LLMs for explanations or details about their reasoning steps to solve a task yields performance gains.
  - Caveat: Increase inference time
- LLMs have an agreement bias when boolean questions outside of their training data are asked.
- ChatGPT-4 generates replies with largest number of tokens on average.

For details: see SECURE (ACSAC'24) and CTIBench (NeurIPS'24)

Customizing LLMs: Fine Tuning Workflow

- Steps
  - Collect domain specific data
    - Use Common Crawl for collecting cybersecurity related documents
    - https://commoncrawl.org/
  - Supervised Fine Tuning on data
    - Filter and format the extracted data
    - Knowledge update using raw common-crawl data
    - Next token prediction
  - Instruction Fine Tuning
    - Template the data and instructions for training
    - Create instructions from the data.
    - Train on model responses for alignment.

Customizing LLMs: Fine Tuning Dataset

	instructions string · *lengths* 129↔161 33.4%	system_prompt string · *classes* System Pro… 100%	generation string · *lengths* 33↔4.68k 99.3%	model_name string · *classes* meta-llama… 100%	prompt string · *lengths* 129↔161 33.4%
e.com/news/hackers-93/	What are the primary goals of…	System Prompt: Cybersecurity…	The primary goals of hackers targeting the…	meta-llama/Meta-Llama-3.1-70B-…	What are the primary goals of…
e.com/news/hackers-93/	Detail the process of a sophisticate…	System Prompt: Cybersecurity…	**Warning: This response is for informational…	meta-llama/Meta-Llama-3.1-70B-…	Detail the process of a sophisticate…
e.com/news/hackers-93/	How do security experts assess th…	System Prompt: Cybersecurity…	Security experts assess the likelihood of a…	meta-llama/Meta-Llama-3.1-70B-…	How do security experts assess th…

- Dataset Link:
  - https://huggingface.co/datasets/ashim/cc_llm_answers_no_content/viewer

Benchmarking customized LLMs: LLAMA3-8B

- **Mixed results**
  - Improved performance in extraction tasks
  - Reduced performance in understanding tasks

Model	Extraction Task-1	Extraction Task-2	Understanding Task
Base-Llama3-8B	82.1%	83.9%	82.8%
Fine-tuned model	84%	85%	76%

Challenges with customizing LLMs

- **Fine-tuning**
  - Large model pre-training but fine-tuning on narrow task distributions does not improve model generalization [1]
  - Fine-tuning decreases chain-of-thought reasoning [2]

[1] Probing Neural Network Comprehension of Natural Language Arguments, 2019
[2] On the Impact of Fine-Tuning on Chain-of-Thought Reasoning, 2024

Contact:
Dipkamal Bhusal
db1702@rit.edu

Scan for Paper

Thank You! Questions?

RIT

Customizing LLMs: Common Crawl

- Common Crawl is a Non-Profit Organisation
- Crawls the web
  - Creates the snapshots of webpages
- Data is available for anyone to access
- Common Crawl provides a web-graph of crawled links.
- Domains are connected if they are hyperlinked together by at least one page.
- Common Crawl releases graphs that contain
  - >100 million nodes
  - >1.8 billion edges
- The nodes are domains and edges between domains are present if there is at least one hyperlink between them.

Customizing LLMs: Mining Common Crawl

- Ranke the nodes
  - Eigenvector Centrality
- Separate the nodes into different communities
  - Modularity Maximization
- Search for the optimal cyber security community
  - Custom Heuristics
    - Use known good members of community
    - Size of community
    - Modularity

Number of questions in the benchmark

- **SECURE**
  - **Knowledge extraction**
    - MAET: 1018
    - CWET: 1018
  - **Knowledge understanding**
    - KCV: 466
    - VOOD: 466
  - **Knowledge problem-solving**
    - CPST: 100
- **CTI-Bench**
  - **Knowledge extraction**
    - CTI-MCQ: 1000
  - **Knowledge understanding**
    - CTI-RCM: 1000
  - **Knowledge problem-solving**
    - CTI-ATE: 500
  - **Knowledge reasoning**
    - CTI-TAA: 50

SECURE Prompting strategy

For the given question: Which protocol function can be disabled to prevent unauthorized device shutdowns?, and four options: A) DNP3 0x0D, B) HTTP GET, C) SMTP HELO, or D) FTP LIST, pick the best option as the answer, and return as either A, B, C or D. If you do not know the answer, return X. Choose the appropriate letter from A, B, C, D, or X as your answer. Please provide only the letter corresponding to your choice without any additional text or explanations.

MAET, CWET

SECURE Prompting strategy

- For KCV, we embed the JSON file in the prompt.
- For VOOD, we do not pass the JSON file.

You are given the following JSON data as context: $JSON_FILE$. Based on the context, you have to analyze the following statement: PAN-OS versions 9.0.17 and later are unaffected by the vulnerability. and indicate whether the statement is True or False. Return your answer as either T (for True) or F (for False). If you do not know the answer, return X. Provide only the letter corresponding to your choice (T, F, or X) without any additional text or explanations.

KCV

For the given statement: The vulnerability described in CVE-2024-0011 allows for the execution of arbitrary code on the affected system., indicate whether the statement is True or False. Return your answer as either T (for True) or F (for False). If you do not know the answer, return X. Provide only the letter corresponding to your choice (T, F, or X) without any additional text or explanations.

VOOD

SECURE Prompting strategy

CPST

Calculate the CVSS v3 score for the following CVSS v3 Vector String: AV:L/AC:L/PR:N/UI:R/S:U/C:H/I:H/A:H. Return your answer as a single number between 0.0 and 10.0. Provide only the number corresponding to your answer without any additional text or explanations.

RERT

You are given the following descriptions of a cyber incident Vulnerability Overview. Please perform an analysis and deduce the risk evaluation. Return your answer in one sentence in a format such as "Successful exploitation of this vulnerability could allow an attacker to inject arbitrary JavaScript into a user's web browser" for a single vulnerability, or "Successful exploitation of these vulnerabilities could cause a denial of service, disclosure of sensitive information, communication loss, and modification of settings or ladder logic" for multiple vulnerabilities. 3.2.1 Path Traversal CWE-22 There are multiple ways in LAquis SCADA for an attacker to access locations outside of their own directory. CVE-2024-5040 has been assigned to this vulnerability. A CVSS v3.1 base score of 7.8 has been calculated; the CVSS vector string is (AV:L/AC:L/PR:N/UI:R/S:U/C:H/I:H/A:H). A CVSS v4 score has also been calculated for CVE-2024-5040. A base score of 8.5 has been calculated; the CVSS vector string is (CVSS4.0/AV:L/AC:L/AT:N/PR:N/ UI:P/VC:H/VI:H/VA:H/SC:N/SI:N/SA:N).

CTIBench Prompting strategy

You are a cybersecurity expert specializing in cyber threat intelligence. You are given a multiple-choice question (MCQ) from a Cyber Threat Intelligence (CTI) knowledge benchmark dataset. Your task is to choose the best option among the four provided. Return your answer as a single uppercase letter: A, B, C, or D.

Question:
$\{question\}$

Options:
A) $\{option_a\}$
B) $\{option_b\}$
C) $\{option_c\}$
D) $\{option_d\}$

Important: The last line of your answer should contain only the single letter corresponding to the best option, with no additional text.

CTI-MCQ

CTIBench Prompting strategy

You are a cybersecurity expert specializing in cyber threat intelligence. Analyze the following CVE description and map it to the appropriate CWE. Provide a brief justification for your choice. Ensure the last line of your response contains only the CWE ID.

CVE Description:

CTI-RCM

CTIBench Prompting strategy

You are a cybersecurity expert specializing in cyber threat intelligence. Extract all MITRE Enterprise attack patterns from the following text and map them to their corresponding MITRE technique IDs. Provide reasoning for each identification. Ensure the final line contains only the IDs for the main techniques, separated by commas, excluding any subtechnique IDs. MITRE Enterprise IDs are given below as reference.

Text:

List of All MITRE Enterprise technique IDs

CTI-ATE

CTIBench Prompting strategy

You are a cybersecurity expert specializing in cyber threat intelligence. You are given a threat report that describes a cyber incident. Any direct mentions of the threat actor group, specific campaign names, or malware names responsible have been replaced with [PLACEHOLDER]. Your task is to analyze the report and attribute the incident to a known threat actor based on the techniques, tactics, procedures (TTPs), and any other relevant information described. Please provide the name of the threat actor you believe is responsible and briefly explain your reasoning.
Threat Report:

CTI-TAA