Automating the Assessment of Japanese Cyber-Security Technical Assessment Requirements Using Large Language Models

Kento Hasegawa[1,*][a], Yuka Ikegami[2,*], Seira Hidano[1], Kazuhide Fukushima[1][b], Kazuo Hashimoto[2] and Nozomu Togawa[2]

[1] *KDDI Research, Inc., 2-1-15, Ohara, Fujimino-shi, Saitama, Japan*

[2] *Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo, Japan*

Keywords: Internet of Things, Security, Large Language Models, Retrieval-Augmented Generation, JC-STAR.

Abstract: Several countries, including the U.S. and European nations, are implementing security assessment programs for IoT devices. Reducing human effort in security assessment has great importance in terms of increasing the efficiency of the assessment process. In this paper, we propose a method of automating the conformance assessment of security requirements based on Japanese program called JC-STAR. The proposed method performs document analysis and device testing. In document analysis, the use of rewrite-retrieve-read and chain of thought within retrieval-augmented generation (RAG) increases the assessment accuracy for documents that have limited detailed descriptions related to security requirements. In device testing, conformance with security requirements is assessed by applying tools and interpreting the results with a large language model. The experimental results show that the proposed method assesses conformance with security requirements with an accuracy of 95% in the best case.

1 INTRODUCTION

Cybersecurity measures for IoT devices have become increasingly important. The cybersecurity requirements for IoT devices are standardized in ETSI EN 303 645 (ETSI, 2024) and NISTIR 8425 (Fagan et al., 2022). Several countries, including the United States and European nations, are implementing different systems based on standard testing specifications. Singapore launched a system called the Cybersecurity Labelling Scheme (CLS) in 2020, allowing vendors of consumer IoT devices to self-declare that their products meet security standards by affixing labels to them. Similar systems are being considered in the UK, the United States, and the EU, and some have already been implemented or are planned for implementation.

In line with trends in several countries, Japan is planning a system called Japan Cyber STAR (JC-STAR) to be implemented in 2025 (Information-technology Promotion Agency, 2024). This system will establish four levels of security requirements, with the first level (Star 1) set to be launched in 2025. Similar to the CLS, under Star 1, IoT device vendors will assess conformance with the security requirements. If the requirements are met, vendors can affix a conformance label to their products, thereby demonstrating to consumers that they meet the security requirements.

Various tools can be used for the security evaluation of IoT devices. For example, well-known tools, such as Nmap and Nikto, are available for anyone to use. However, according to a study by Kaksonen et al., relying solely on these tools is not sufficient to conduct a thorough security evaluation (Kaksonen et al., 2024). This is attributed to the diversity of IoT device functions, which makes it challenging to conduct a standardized security evaluation.

To reduce human effort in security assessment, the utilization of large language models (LLMs) can be considered. However, there are challenges in applying these models due to hallucinations, in which LLMs generate false or misleading responses. Determining how to effectively utilize LLMs in security assessment has significant importance in terms of increasing the efficiency of the assessment process.

This paper proposes a method of automating security assessment, particularly security assessment in JC-STAR Star 1, using LLMs. The target security assessment process consists of document analysis and physical testing. In the document analysis step,

[a] https://orcid.org/0000-0002-6517-1703

[b] https://orcid.org/0000-0003-2571-0116

*Kento Hasegawa and Yuka Ikegami contributed equally to this work.

Hasegawa, K., Ikegami, Y., Hidano, S., Fukushima, K., Hashimoto, K. and Togawa, N.
Automating the Assessment of Japanese Cyber-Security Technical Assessment Requirements Using Large Language Models.
DOI: 10.5220/0013345300003944

In *Proceedings of the 10th International Conference on Internet of Things, Big Data and Security (IoTBDS 2025)*, pages 305-312
ISBN: 978-989-758-750-4; ISSN: 2184-4976

retrieval-augmented generation (RAG) is utilized to analyze the user manuals of IoT devices. In the physical testing step, the results of operating the devices with specific tools are analyzed with the LLM. By combining these approaches, security verification can be automated.

The contributions of this paper can be summarized as follows:

- By utilizing RAG, a technique is established to extract explanations regarding security requirements from the IoT device user manuals and to evaluate compliance with those requirements.
- A method is developed for automatically analyzing the results of physical testing by combining security verification tools with LLMs.
- In the evaluation experiments, an average accuracy rate of 83% was achieved in document evaluation.

2 PRELIMINARIES

2.1 JC-STAR

The Japanese government has decided to start a new program called "JC-STAR: Labeling Scheme based on Japan Cyber-Security Technical Assessment Requirements" in 2025. This program verifies that IoT devices conform to technical security requirements. This program begins at the Star 1 level. At this level, the common minimum security requirements for IoT devices as products are defined. IoT device vendors can either conduct their own security assessment of the devices according to the compliance criteria and the assessment procedures established at Star 1 or request assessment from third-party assessment organizations. If the assessment confirms that the security requirements are met, the vendor can obtain a label for the IoT device. By attaching this label to their products, IoT vendors can assert that their products meet the security requirements.

A total of 16 items are established as security requirements for Star 1. Table 1 shows an overview of the security requirements. The security standards include elements similar to international security standards, such as ETSI EN 303 645. For each security requirement, two assessment approaches are used: document evaluation and device testing.

In document evaluation, compliance with security requirements is verified by referring to the descriptions in IoT device specifications or user manuals. The challenge is to determine security compliance from documents that may not be adequately detailed regarding security.

In device testing, compliance with security requirements is verified by observing the behavior of the IoT device. Existing tools such as Nmap and Nikto can potentially be used in physical testing. However, in security compliance evaluation, it is difficult to determine whether the outputs from these tools confirm compliance with security requirements.

Note that, in Table 1, items in the "Device Testing" column marked with an asterisk "*" are considered challenging to automate using software because physical inspection of the IoT device, such as checking the label attached to the body, is needed. Thus, they are excluded from the scope of this paper. Additionally, item #4 involves performing a brute-force verification attack on passwords, and since existing tools are available for this purpose, it is also excluded from consideration.

2.2 Integration with Large Language Models

Large language models (LLMs) have made remarkable advancements. However, although LLMs are trained on a vast range of data sources, they lack detailed knowledge concerning specific domains. In particular, the security field requires deep computer-related knowledge, making it difficult to interpret security requirements using only LLMs. In this paper, we address problems that cannot be addressed by using LLMs alone by introducing several techniques outlined below.

2.2.1 Retrieval-Augmented Generation

RAG (Lewis et al., 2020) is a technology that allows LLMs to interpret and respond to information in specific domains by providing knowledge from external sources. In RAG, a collection of passages related to external knowledge is prepared in advance as an external data source. When a prompt is given by a user, the RAG system first retrieves passages related to the prompt from the data source and compiles them as context. The LLM is then instructed to generate a response to the prompt based on this context. This allows the LLM to produce responses grounded in external knowledge, using context as a clue.

2.2.2 Rewrite–Retrieve–Read

In RAG, the accuracy of the retrieval process affects the accuracy of response generation. Since the passages in the data source and the sentences input as

Table 1: Requirements of JC-STAR level 1.

#	Category	ETSI EN 303 645	Assessment Approach Document Evaluation	Device Testing
1	No vulnerable authentication/authorization mechanism	5.1-3	Yes	-
2	No vulnerable authentication/authorization mechanism	5.1-2	Yes (①, ②)	-
3	No vulnerable authentication/authorization mechanism	5.1-4	Yes	-
4	No vulnerable authentication/authorization mechanism	5.1-5	-	Yes*
5	Managing vulnerability reports	5.2-1	Yes (①, ②, ③)	-
6	Keep software updated	5.3-1	-	Yes (①, ②, ③*)
7	Keep software updated	5.3-3	Yes	-
8	Keep software updated	5.3-7	Yes	-
9	Keep software updated	5.3-8	Yes	-
10	Keep software updated	5.3-16	-	Yes (①*, ②)
11	Securely store sensitive parameters	5.4-1	Yes	-
12	Communicate securely	5.5-1	Yes (① or ②)	-
13	Minimize exposed attack surfaces	5.6-1	Yes (①)	Yes (①, ②)
14	Resilience to outages	5.9-1	-	Yes*
15	Delete user data	5.11-1	Yes (①)	Yes (①*, ②*)
16	Provide information on products	5.12-2	Yes (①–⑤)	-

prompts can differ in style and vocabulary, the challenge lies in effectively searching for passages related to the prompt. Rewrite–retrieve–read (Ma et al., 2023) is a technique that improves retrieval accuracy by rewriting queries with the LLM based on the prompt to search for relevant passages via RAG.

2.2.3 Chain of Thought

Chain of Thought (CoT) is a technique that increases the accuracy of LLM responses by breaking down complex instructions into step-by-step explanations (Wei et al., 2022). Since the passages obtained through RAG typically explain the operating procedures or specifications of IoT devices, they often contain limited detailed descriptions of security requirements. By applying CoT, it becomes possible for the LLM to accurately assess compliance with security requirements.

3 PROPOSED METHOD

3.1 Overview

In this paper, we propose a method for automating the security assessment procedure of JC-STAR. As explained in Section 2, the key points for automation are as follows:

- Document analysis on the basis of documents with insufficient security-related descriptions.
- Determination of security requirements on the basis of the logs output by security tools during device testing.

The following section presents proposals for automating the security assessment of JC-STAR via LLMs; these proposals are divided into two parts: document analysis and device testing.

3.2 Part 1: Document Analysis

In document analysis, the use of RAG, rewrite-retrieve-read, and CoT allows for the assessment of conformance with security requirements based on the documentation of IoT devices. Algorithm 1 outlines the document analysis procedure.

3.2.1 Step 1-1: Vector Database Construction

In this step, a vector database is constructed based on IoT device documents, such as specifications and user manuals. First, the documents are divided into chunks, specifically either by paragraph or by using fixed-length strings. Next, an embedding model is used to obtain embedding vectors for each chunk. By associating the original chunk strings with their corresponding embedding vectors, a vector database is constructed.

3.2.2 Step 1-2: Query Generation

The style of security requirements differs from that of IoT device documents. Therefore, merely searching the vector database based on the phrasing of security requirements might not yield the expected results. To address this, the rewrite-retrieve-read approach is used to rephrase the security requirements and generate queries for searching the vector database. In the rephrasing step, an LLM is used to produce the rephrased text. Prompt 1 shows a template for the

Algorithm 1: Part 1: Document Analysis for Security Requirements.

Input: Document **d**, Security requirements $\mathcal{S}$, Embedding model E, LLM M.
Output: Evaluation results $\mathcal{R}_{\text{DA}}$.

Step 1-1: Vector Database Construction ▷
  $\mathcal{D}_{\text{chunk}} \leftarrow$ Split the given document **d** into several chunks.
  $\mathcal{D}_{\text{emb}} \leftarrow \{E(c) : \forall c \in \mathcal{D}_{\text{chunk}}\}$. // Obtain an embedding for each chunk.
$\mathcal{R}_{\text{DA}} \leftarrow \emptyset$
for $\mathbf{s} \in \mathcal{S}$ **do**
    // Process for each security requirement.
  Step 1-2: Query Generation ▷
    $\mathbf{q} \leftarrow$ Generate a query based on the security requirement **s** using the rewrite-retrieve-read method.
  Step 1-3: Context Retrieval ▷
    $\mathcal{C} \leftarrow$ Retrieve relevant chunks from the database $\mathcal{D}_{\text{emb}}$ based on the query **q**.
  Step 1-4: Conformance Assessment ▷
    $\mathcal{C}' \leftarrow$ Refine a set of the retrieved context $\mathcal{C}$ using LLM M.
    $r \leftarrow$ Provide a conformance assessment prompt to the LLM M and obtain its response.
    $\mathcal{R}_{\text{DA}} \leftarrow \mathcal{R}_{\text{DA}} \cup \{r\}$.
end for
return $\mathcal{R}_{\text{DA}}$

prompt provided to the model based on the security requirements phrasing.

Prompt 1: Step 1-2

> What information in the user manual of an IoT device would help you to determine if the device meets the following conformance criteria? Please answer with examples of specific functions and specifications.
>
> Conformance criteria:
> { Security requirement **s** }
>
> Answer:

3.2.3 Step 1-3: Context Retrieval

Based on the queries obtained in Step 1-2, the context is retrieved from the vector database constructed in Step 1-1. To consider both keyword occurrence in the chunks and the semantic meaning of the sentences, reranking (Xiao et al., 2024) is utilized. In the reranking process, first, a list of top similar contexts is obtained for the query from both the keyword search and the embedding search. The lists from both searches are then merged, removing duplicates, to create a combined list. This list is provided to a reranking model, which determines the ranking of the search results for the query. By providing this list to the reranking model, a sorted list of search results for the query is obtained, taking into account the ranking determined by the model.

3.2.4 Step 1-4: Conformance Assessment

Based on the list of contexts obtained in Step 1-3, conformity to security requirements is assessed. In Step 1-4 ①, the contexts are rephrased. The contexts extracted from IoT device documents may lack sufficient information to determine the conformity of security requirements. Thus, it is necessary to augment the information via an LLM. Prompt 2 shows the prompt used in this step.

Prompt 2: Step 1-4 ①

> You are an expert in IoT devices. List what you can find out from the following documents extracted from the user manual of an IoT device.
>
> Extracted documents:
> { List of relevant chunks }

Continuing from Step 1-4 ①, in Step 1-4 ②, it is confirmed whether the content described in the IoT device documentation meets the security standards. Prompt 3 shows the prompt used for this step.

Prompt 3: Step 1-4 ②

> Answer Yes or No, followed by a one-sentence reason.
> Question:
> Does the IoT device meet the following conformance criteria?
> { Security requirement **s** }

3.3 Part 2: Device Testing

In device testing, security tools are applied to the device. Based on the execution results, an LLM is used to evaluate conformance with security requirements. Algorithm 2 shows an overview of the procedure in Part 2.

Algorithm 2: Part 2: Device Testing for Security Requirements.

Input: Security requirements $\mathcal{S}$.
Output: Evaluation results $\mathcal{R}_{\mathrm{DT}}$.

$\mathcal{R}_{\mathrm{DT}} \leftarrow \emptyset$
for $\mathbf{s} \in \mathcal{S}$ **do**
    // Process for each security requirement.
  Step 2-1: Context Retrieval ▷
    $\mathbf{o} \leftarrow$ Retrieve the relevant context. The procedures are the same as in Steps 1-1, 1-2, and 1-3.
  Step 2-2: Tool Execution ▷
    $t \leftarrow$ Select the appropriate security tools to assess security requirement s.
    $\mathbf{o} \leftarrow$ Obtain the output (or report) of performing the securty tool on the target IoT device.
  Step 2-3: Conformance Assessment ▷
    $r \leftarrow$ Provide a conformance assessment prompt based on security requirement $\mathbf{s}$ and tool output $\mathbf{o}$ to the LLM M and obtain its response.
    $\mathcal{R}_{\mathrm{DT}} \leftarrow \mathcal{R}_{\mathrm{DT}} \cup \{r\}$.
end for
return $\mathcal{R}_{\mathrm{DT}}$

3.3.1 Step 2-1: Context Retrieval

Even in the case of actual device testing, there might be instances where it is necessary to refer to the documentation of the IoT device. In such situations, the same procedures as in Steps 1-1, 1-2, and 1-3 should be followed to obtain context related to the security requirements.

3.3.2 Step 2-2: Tool Execution

In this step, security tools are applied to the actual IoT device, and the results are obtained. Each security requirement is prematched with appropriate security tools. For example, security requirement #13-① specifies disabling unnecessary interfaces. Here, we focus on TCP/IP ports for simplicity. In this case, a tool such as Nmap can be used to perform the assessment. In another case, for security requirement #10, which requires that users be able to confirm the model number of an IoT device, a script can be used to crawl the Web-based management interface of the IoT device and obtain the displayed information.

3.3.3 Step 2-3: Conformance Assessment

Based on the tool outputs, an LLM is employed to evaluate the conformance with security requirements. Since the output content varies by tool, the wording of a prompt is slightly adjusted for each security requirement. For example, for security requirement #13-①, whether unnecessary ports are disabled is determined via the context obtained from the documentation and the output of a port scan tool. Therefore, the following prompt can be used:

Prompt 4: Step 2-3 for #13-①

You are an expert in IoT devices. Answer Yes or No, followed by a one-sentence reason.

Question:
The following is the result of a port scan performed by Nmap on an IoT device. Are there any irrelevant ports with reference to the extracted IoT documents?

Extracted documents:
{ List of relevant chunks }

Port scan result:
{ Output of the security tool }

4 EXPERIMENTS

4.1 Setup

We evaluate the automation of the conformance assessment for the Star 1 security requirements of JC-STAR. In the experiment, we compare the accuracy of automated conformance assessment with that of human evaluations to determine how closely they align. The IoT device used for the evaluation is an IP camera running in the emulation environment FirmAE (Kim et al., 2020).

The experimental program is implemented in Python. For the vector database, FAISS (Douze et al., 2024) is employed; the reranker used is BAAI/bge-reranker-large (Xiao et al., 2024). The embedding model is stsb-xlm-r-multilingual (Reimers and Gurevych, 2019), and the LLM is Mistral-7B-Instruct-v0.2 (Jiang et al., 2023).

4.2 Experiment 1: Document Analysis

In this experiment, we aim to confirm how accurately the method proposed in Part 1 can generate responses

Table 2: Experimental results of document analysis.

#	Subitem	# of Appropriate Responses	Could be assessed
1		5	✓
2	①	4	✓
	②	5	✓
3		1	✓
5	①	0	-
	②	5	✓
	③	5	✓
7		2	✓
8		4	✓
9		5	✓
11		5	✓
12	①	5	✓
	②	5	✓
13	①	3	✓
15	①	4	✓
16	①	5	✓
	②	5	✓
	③	5	✓
	④	5	✓
	⑤	5	✓
Average		4.15	19 / 20

compared with human evaluations. To account for the randomness inherent in LLM generation, the results were generated five times.

Table 2 shows an overview of the results of Experiment 1. In the experiment, evaluations are conducted for each specific subitem if the subitem is available. The third column indicates the number of appropriate responses obtained among the five attempts. A checkmark ✓ in the fourth column indicates that at least one appropriate response was obtained in those five attempts. As shown in the table, appropriate responses were achieved for 19 out of the 20 evaluation items, resulting in an accuracy of 95%. This demonstrates that using the proposed method yields a very high level of accuracy.

As an example of the experimental results, we present the conformance assessment conducted based on security requirement #2-①. This security requirement involves verifying whether the default password is unique for each device, consists of at least six characters, and is not easily guessable. First, in Step 1-4 of the proposed method, the context obtained from the documentation is refined. Prompt 5 and Response 1 show the prompt and response used in this experiment. When entering Prompt 6, we obtained Response 2. In this sequence, as indicated by the underlined text in Prompt 5 and Response 1, both the username and password for the device are “admin,” which means they do not meet the security requirement. Consequently, the result is determined as “No” (not compliant), as shown in Response 2.

Prompt 5: Step 1-4 ① for #2-①

You are an expert in IoT devices. List what you can find out from the following documents extracted from the user manual of an IoT device.

Extracted documents:
...
Document 4:On the login window, enter the User name/Password and click OK to login. The default User name/Password is admin /admin . If you wish to save the login information, please select Auto Login .
...

Response 1: Step 1-4 ① for #2-①

From the given documents, I can find out the following information related to an IoT device:

...
4. The device has a login window for HTTP server with default username and password as admin/admin.
...

Prompt 6: Step 1-4 ② for #2-①

Answer Yes or No, followed by a one-sentence reason.

Question:
Does the IoT device meet the following conformance criteria?

For products that use passwords or passcodes in the user authentication mechanism via the network for the product or in the client authentication mechanism at the initial setup of the product, the following criteria (1) shall be satisfied when a default password is used at the time of product installation. (1) The default password shall be a unique value that differs for each product and shall be at least 6 characters in length and not easily guessable.

Table 3: Experimental results of device testing.

#	Subitem	Tool	Could be assessed
6	①	†	✓
	②	†	✓
10	②	†	✓
13	①	Nmap	✓
	②	Nikto	✓
Total			5 / 5

Response 2: Step 1-4 ② for #2-①

No, the IoT device does not meet the conformance criteria stated as the default passwords, as mentioned in the user manual, are not unique or not mentioned to be at least 6 characters in length and not easily guessable.

However, the proposed method was unable to generate an appropriate response for security requirement #5, which states "Is there a published point of contact for reporting security issues with the product (e.g., URL of the manufacturer's website, phone number, email address)?" In the documentation of the targeted IoT device, this information was provided in an image format. Since only text information from the documents in this experiment, the necessary information was not gathered when the vector database was constructed in Step 1-1. To address this, implementing optical character recognition processing or introducing multimodal language models to interpret images could be potential solutions, as well as searching for product information directly from the web.

4.3 Experiment 2: Device Testing

In Experiment 2, actual verification was conducted using an IoT device running in an emulation environment. The "Tool" column in Table 3 lists the tools used in the experiment. The "†" symbol denotes a custom tool. This tool automatically retrieves the contents of the IoT device's management interface. Specifically, it uses Playwright[a] for Python to automate web browser control to capture the content displayed on the management interface. Nmap[b] is a well-known port scanning tool, and Nikto[c] is a well-known vulnerability scanner.

The "Could be assessed" column in Table 3 shows the results of Experiment 2. As shown here, the proposed method could successfully assesses the security requirements listed in the table.

[a] https://playwright.dev/python/

[b] https://nmap.org/

[c] https://github.com/sullo/nikto

As an example of the experimental results, Prompt 7 and Response 3 show the prompt and response for security requirement #13-①. In #13-①, it is confirmed whether unnecessary ports are disabled. As indicated by the underlined text in Prompt 7, the documentation for the IoT device states that port 80 is used. However, the port scan result with a wavy line indicates that the port 8080 is being used. This discrepancy is due to the fact that the port number of the management interface was changed for execution in the emulation environment.

The evaluation via the proposed method, as shown in Response 3, successfully identifies the inconsistency between the port numbers described in the documentation and those that were actually used. Therefore, the compliance evaluation of security requirements effectively highlights this concern.

Prompt 7: Step 2-3 for #13-①

You are an expert in IoT devices. Answer Yes or No, followed by a one-sentence reason.

Question:
The following is the result of a port scan performed by Nmap on an IoT device. Are there any irrelevant ports with reference to the extracted IoT documents?

Extracted documents:
...
Document 2:Frame rate: 1 30 fps Compression: 5 levels Recording Recording type: continuous, schedule or motion detection with software Port Settings HTTP port: 80 (default) ...

Port scan result:
PORT STATE SERVICE
xxx/tcp open xxx
8080/tcp open http-proxy
...

Response 3: Step 2-3 #13-①

Based on the provided documents, the relevant ports for the IoT device are 80 (Document 2)...
The port scan result shows open ports 8080, ... None of these ports match the ports mentioned in the documents, therefore they are considered irrelevant to the IoT device in question.

4.4 Discussion

4.4.1 Limitations

In the conformance assessment, several tasks require manual intervention. For example, verifying that a label with the product model number is on the body of an IoT device is challenging to fully automate. However, the evaluation of most security requirements can be automated. Therefore, in streamlining manual work, the proposed automation tool is effective.

4.4.2 Collection of Technical Information

In both document analysis and device testing, it is necessary to verify the technical documentation related to the IoT device. However, user manuals and other documents often lack sufficiently detailed information, making it difficult to evaluate conformance with security requirements. Therefore, it is crucial to collect a wide variety of technical information related to the IoT devices under inspection. Determining how to automatically gather this relevant information remains an issue for future work.

5 CONCLUSION

In this paper, we propose a method of automating the conformance assessment of security requirements based on JC-STAR. The proposed method consists of document analysis and device testing. In document analysis, the use of rewrite-retrieve-read and CoT within RAG increases the assessment accuracy. In device testing, conformance with security requirements is assessed by applying tools and interpreting the results with an LLM. The experimental results show that the proposed method assesses conformance with security requirements with an accuracy of 95% in the best case.

Future work includes quantifying the certainty of evaluation results and providing rationales. This is expected to enhance the reliability of inspection results. Additionally, examining differences arising from language variations in security requirements and user manuals also included.

ACKNOWLEDGEMENTS

The results of this research were obtained in part through a contract research project (JPJ012368C08101) sponsored by the National Institute of Information and Communications Technology (NICT).

REFERENCES

Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P.-E., Lomeli, M., Hosseini, L., and Jégou, H. (2024). The faiss library.

ETSI (2024). Cyber security for consumer internet of things: Baseline requirements, ETSI EN 303 645 V3.1.3.

Fagan, M., Megas, K., Watrobski, P., Marron, J., and Cuthill, B. (2022). Profile of the iot core baseline for consumer iot products, NIST IR 8425.

Information-technology Promotion Agency (2024). Japan cyber star (jc-star). https://www.ipa.go.jp/en/security/jc-star/index.html.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L. R., Lachaux, M.-A., Stock, P., Scao, T. L., Lavril, T., Wang, T., Lacroix, T., and Sayed, W. E. (2023). Mistral 7b.

Kaksonen, R., Halunen, K., Laakso, M., and Röning, J. (2024). Automating iot security standard testing by common security tools. In *Proceedings of the 10th International Conference on Information Systems Security and Privacy*, pages 42–53. SCITEPRESS.

Kim, M., Kim, D., Kim, E., Kim, S., Jang, Y., and Kim, Y. (2020). Firmae: Towards large-scale emulation of iot firmware for dynamic analysis. In *Proceedings of the 36th Annual Computer Security Applications Conference*, pages 733–745. Association for Computing Machinery.

Lewis, P. S. H., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., tau Yih, W., Rocktäschel, T., Riedel, S., and Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. In *NeurIPS*.

Ma, X., Gong, Y., He, P., Zhao, H., and Duan, N. (2023). Query rewriting in retrieval-augmented large language models. In *Empirical Methods in Natural Language Processing, EMNLP*, pages 5303–5315.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Xiao, S., Liu, Z., Zhang, P., Muennighoff, N., Lian, D., and Nie, J.-Y. (2024). C-pack: Packed resources for general chinese embeddings. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 641–649. Association for Computing Machinery.