

Proposal of Tsuzumi's future: Detecting phishing sites

IOWN Pick Up NTT version Large Language Models



Background

The number of phishing attacks using generative AI has been increasing, with the number of victims per month reaching an all-time high. In response to this situation, we are working to reduce damage by detecting phishing sites using generative AI.

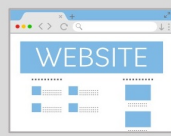
Summary

NTT has created technologies to collect and detect social engineering attacks on the Web that target human psychological weaknesses. This time, we realized high-precision phishing site detection using generative AI by linking with automatic web crawling technology.

Development Challenges



① Website information cannot be directly input to LLM

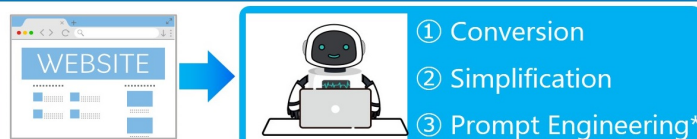


Is this a phishing site?

I don't know

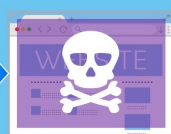
② Simple commands lead to many judgment errors

Technical Overview



- ① Convert website screenshot images to text
- ② Simplify HTML and text
- ③ Control the inference process by prompt engineering*

*the method of designing and optimizing commands to obtain a desired output.



This is a phishing site.

Detect phishing sites with over 98% accuracy*

*Using GPT-4

Features

- World's first detection technology utilizing advanced contextual understanding with generative AI to identify psychological tactics of phishing sites
- Ability to detect brands disguised as phishing sites and determine discrepancies with legitimate websites for accurate malicious determination
- Reduce the cost of manual monitoring and detection by collecting and analyzing a large number of phishing sites in a short time through automated website crawling

Future_benefits

Looking ahead to the future of many phishing attacks using generative AI, we will prevent damage with highly accurate phishing site detection technology that uses generative AI.

Exhibiting Company

NIPPON TELEGRAPH AND TELEPHONE CORPORATION, NTT Security Holdings Corporation

Contact

rdforum-exhibition@ml.ntt.com