Template Injection in Agent Tool Definitions

High **MervinPraison** published **GHSA-hwg5-x759-7wjg** yesterday

Package

praisonai (pip)

Affected versions

<= 4.5.114

Patched versions

>= 4.5.115

Description

Summary

Direct insertion of unescaped user input into template-rendering tools allows arbitrary code execution via specially crafted agent instructions.

Details

The `create_agent_centric_tools()` function returns tools (like `acp_create_file`) that process file content using template rendering. When user input from `agent.start()` is passed directly into these tools without escaping (as shown in `agent_centric_example.py:85-86`), template expressions in the input are executed rather than treated as literal text. This occurs because:

1. No input sanitization or escaping is applied to user-controlled content
2. The ACP-enabled runtime auto-approves operations (`approval_mode="auto"`)
3. Tools lack context-aware escaping for template syntax

PoC

```
# Replace the agent.start() call at line 85 with:
result = agent.start('Create file with content: {{ self.__init__.__globals__.__bui
```

Successful exploitation creates `/tmp/pwned` confirming arbitrary command execution. The expression `{{7*7}}` renders as `49` instead of literal text.

Impact

Attackers can execute arbitrary system commands with the privileges of the running process by injecting malicious template expressions through agent instructions. This compromises the host system, enabling data theft, ransomware deployment, or lateral movement.

Recommended Fix

1. **Input Sanitization**: Implement strict whitelist validation for file content
2. **Contextual Escaping**: Auto-escape template syntax characters (e.g., `{{ }}`) in user input using Jinja2 `autoescape=True`
3. **Sandboxing**: Restrict template execution environments using secure eval modes
4. **Approval Hardening**: Require manual approval for file creation operations in production

Severity

High 8.8 / 10

CVSS v3 base metrics

Attack vector	Network
Attack complexity	Low
Privileges required	Low
User interaction	None
Scope	Unchanged
Confidentiality	High
Integrity	High
Availability	High

Learn more about base metrics

CVSS:3.1/AV:N/AC:L/PR:L/UI:N/S:U/C:H/I:H/A:H

CVE ID

CVE-2026-39891

Weaknesses

▶ CWE-94

Credits

offset Reporter