Airtable_Agent Code Injection Remote Code Execution Vulnerability

Critical **igor-magun-wd** published **GHSA-v38x-c887-992f** last week

Package

flowise (npm)

Affected versions	Patched versions
<= 3.0.13	3.1.0

flowise-components (npm)

<= 3.0.13	3.1.0

Description

Please find POC file here : https://trendmicro-my.sharepoint.com/:u:/p/kholoud_altookhy/IQDrSbbuHq51TJXkfDZY3L26AYM9Jw8yPQNK5zXvpQ3p9rQ?e=IS7yPv

ZDI-CAN-29412: FlowiseAI Flowise Airtable_Agent Code Injection Remote Code Execution Vulnerability

-- CVSS ----------

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

-- ABSTRACT ----------

Trend Micro's Zero Day Initiative has identified a vulnerability affecting the following products:
Flowise - Flowise

-- VULNERABILITY DETAILS ----------

- Version tested: 3.0.13
- Installer file: hxxps://github.com/FlowiseAI/Flowise
- Platform tested: Ubuntu 25.10

Analysis

FlowiseAI Flowise Airtable Agent pythonCode Prompt Injection Remote Code Execution Vulnerability

This vulnerability allows remote attackers to execute arbitrary code on affected installations of FlowiseAI Flowise. Authentication is not required to exploit this vulnerability.

The specific flaw exists within the run method of the Airtable_Agents class. The issue results from the lack of proper sandboxing when evaluating an LLM generated python script. An attacker can leverage this vulnerability to execute code in the context of the user running the server.

Product information

FlowiseAI Flowise version 3.0.13 (https://github.com/FlowiseAI/Flowise)

Setup Instructions

npm install -g flowise@3.0.13
npx flowise start

Root Cause Analysis

FlowiseAI Flowise is an open source low-code tool for developers to build customized large language model (LLM) applications and AI agents. It supports integration with various LLMs, data sources, and tools in order to facilitate rapid development and deployment of AI solutions. Flowise offers a web interface with a drag-and-drop editor, as well as an API, through an Express web server accessible over HTTP on port 3000/TCP.

One such feature of Flowise is the ability to create chatflows. Chatflows use a drag and drop editor that allow a developer to place nodes which control how an interaction with a LLM will occur. One such node is the Airtable Agent node that represents an Agent used to answer queries on a provided Airtable table.

When a user makes a query against a chatflow using the Airtable Agent node, the run method of the Airtable_Agents class will be called. This method will first request the contents of the Airtable table passed to the node and convert it to a base64 string. It will then set up a pyodide environment and create a python script to be executed in this environment. This python script will use pandas to extract the column names and their types from the Airtable table. The method will then create a system prompt for an LLM using this data as follows:

```
You are working with a pandas dataframe in Python. The name of the dataframe is df

The columns and data types of a dataframe are given below as a Python dictionary with
keys showing column names and values showing the data types.
```

```
{dict}

I will ask question, and you will output the Python code using pandas dataframe to
answer my question. Do not provide any explanations. Do not respond with anything
except the output of the code.

Security: Output ONLY pandas/numpy operations on the dataframe (df). Do not use
import, exec, eval, open, os, subprocess, or any other system or file operations. The
code will be validated and rejected if it contains such constructs.

Question: {question}
Output Code:
```

Where {dict} is the extracted column names and {question} is the initial prompt provided by the user.

This system prompt will be sent to an LLM in order for it to generate a python script based on the user's prompt, and the LLM generated response will be stored in a variable name pythonCode. The method will then evaluate the pythonCode variable in a pyodide environment.

While the LLM-generated Python script is evaluated in a non-sandboxed environment, there is a list of forbidden patterns that are checked for before the script is executed on the server. The function validatePythonCodeForDataFrame() enumerates through a list, named FORBIDDEN_PATTERNS, which contains pairs of regex pattern and reasons. Each regex pattern is run against the Python script, and if the pattern is found in the script, the script is invalidated and is not run, responding to the request with a reason for rejection.

The input validation can be bypassed, which can still lead to running arbitrary OS commands on the server. An example of this is the pattern `/\bimport\s+(?!pandas|numpy\b)/g`, which intends to search for lines of code which import a module other than pandas or numpy. This can be bypassed by importing along with pandas or numpy. For example, consider the following lines of code:

```
import pandas as np, os as pandas
pandas.system("xcalc")
```

pandas is imported, but so is the os module, with pandas as its alias. OS commands can then be invoked with `pandas.system()`.

Using prompt injection techniques, an unauthenticated attacker with the ability to send prompts to a chatflow using the Airtable Agent node may convince an LLM to respond with a malicious python script that executes attacker controlled commands on the flowise server.

An attacker can use this vulnerability to execute arbitrary python code, which can lead to arbitrary system commands being executed on the target server.

It is also possible for an authenticated attacker to exploit this vulnerability by specifying an attacker controlled server in a chatflow. This server would respond to prompts with an attacker controlled python script instead of an LLM generated response, which would then be evaluated on the server.

It is also possible for an authenticated attacker to exploit this vulnerability by specifying an attacker controlled Airtable table in a chatflow. This airtable table would contain columns whose name contain prompt injections, that are later passed to an LLM to use when generating a python script.

comments documenting the issue have been added to the following code snippet. Added comments are prepended with "!!!".

From packages/components/nodes/agents/AirtableAgent/core.ts

```
import type { PyodideInterface } from 'pyodide'
import * as path from 'path'
import { getUserHome } from '../../../src/utils'

let pyodideInstance: PyodideInterface | undefined

export async function LoadPyodide(): Promise<PyodideInterface> {
    if (pyodideInstance === undefined) {
        const { loadPyodide } = await import('pyodide')
        const obj: any = { packageCacheDir: path.join(getUserHome(), '.flowise', 'pyodid
        pyodideInstance = await loadPyodide(obj)
        await pyodideInstance.loadPackage(['pandas', 'numpy'])
    }

    return pyodideInstance
}

export const systemPrompt = `You are working with a pandas dataframe in Python. The name

The columns and data types of a dataframe are given below as a Python dictionary with ke
{dict}

I will ask question, and you will output the Python code using pandas dataframe to answe

Security: Output ONLY pandas/numpy operations on the dataframe (df). Do not use import,

Question: {question}
Output Code:`

export const finalSystemPrompt = `You are given the question: {question}. You have an an
Standalone Answer:`
```

From packages/components/nodes/agents/AirtableAgent/AirtableAgent.ts

```
import axios from 'axios'
import { BaseLanguageModel } from '@langchain/core/language_models/base'
import { AgentExecutor } from 'langchain/agents'
import { LLMChain } from 'langchain/chains'
import { ICommonObject, INode, INodeData, INodeParams, IServerSideEventStreamer, PromptT
import { getBaseClasses, getCredentialData, getCredentialParam } from '../../../src/util
import { ConsoleCallbackHandler, CustomChainHandler, additionalCallbacks } from '../../.
import { LoadPyodide, finalSystemPrompt, systemPrompt } from './core'
import { validatePythonCodeForDataFrame } from '../../../src/pythonCodeValidator'
```

```
import { checkInputs, Moderation } from '../../moderation/Moderation'
import { formatResponse } from '../../outputparsers/OutputParserHelpers'

class Airtable_Agents implements INode {
    label: string
    name: string
    version: number
    description: string
    type: string
    icon: string
    category: string
    baseClasses: string[]
    credential: INodeParams
    inputs: INodeParams[]

    // !!! [... Truncated for Readability ...]

    // !!! input variable holds prompt from user
    async run(nodeData: INodeData, input: string, options: ICommonObject): Promise<strin
        const model = nodeData.inputs?.model as BaseLanguageModel
        const baseId = nodeData.inputs?.baseId as string
        const tableId = nodeData.inputs?.tableId as string
        const returnAll = nodeData.inputs?.returnAll as boolean
        const limit = nodeData.inputs?.limit as string
        const moderations = nodeData.inputs?.inputModeration as Moderation[]

        // !!! the chatflow may contain moderation nodes that search for prompt injectio
        if (moderations && moderations.length > 0) {
            try {
                // Use the output of the moderation chain as input for the Vectara chain
                input = await checkInputs(moderations, input)
            } catch (e) {
                await new Promise((resolve) => setTimeout(resolve, 500))
                // if (options.shouldStreamResponse) {
                //     streamResponse(options.sseStreamer, options.chatId, e.message)
                // }
                return formatResponse(e.message)
            }
        }

        const shouldStreamResponse = options.shouldStreamResponse
        const sseStreamer: IServerSideEventStreamer = options.sseStreamer as IServerSide
        const chatId = options.chatId

        const credentialData = await getCredentialData(nodeData.credential ?? '', option
        const accessToken = getCredentialParam('accessToken', credentialData, nodeData)

        let airtableData: ICommonObject[] = []

        // !!! Get Airtable data
        if (returnAll) {
            airtableData = await loadAll(baseId, tableId, accessToken)
        } else {
            airtableData = await loadLimit(limit ? parseInt(limit, 10) : 100, baseId, ta
        }
```

```
        let base64String = Buffer.from(JSON.stringify(airtableData)).toString('base64')

        const loggerHandler = new ConsoleCallbackHandler(options.logger, options?.orgId)
        const callbacks = await additionalCallbacks(nodeData, options)

        const pyodide = await LoadPyodide()

        // First load the csv file and get the dataframe dictionary of column types
        // For example using titanic.csv: {'PassengerId': 'int64', 'Survived': 'int64',
        let dataframeColDict = ''
        try {
            const code = `import pandas as pd
import base64
import json

base64_string = "${base64String}"

decoded_data = base64.b64decode(base64_string)

json_data = json.loads(decoded_data)

df = pd.DataFrame(json_data)
my_dict = df.dtypes.astype(str).to_dict()
print(my_dict)
json.dumps(my_dict)`
            dataframeColDict = await pyodide.runPythonAsync(code)
        } catch (error) {
            throw new Error(error)
        }

        // !!! ask LLM to come up with python script...
        // Then tell GPT to come out with ONLY python code
        // For example: len(df), df[df['SibSp'] > 3]['PassengerId'].count()
        let pythonCode = ''
        if (dataframeColDict) {
            const chain = new LLMChain({
                llm: model,
                // !!! prompt passed to LLM
                prompt: PromptTemplate.fromTemplate(systemPrompt),
                verbose: process.env.DEBUG === 'true' ? true : false
            })
            const inputs = {
                // !!! Airtable column names are also subbed into the system prompt whic
                dict: dataframeColDict,
                // !!! question, which is later subbed into the system prompt, is given
                question: input
            }
            const res = await chain.call(inputs, [loggerHandler, ...callbacks])
            // !!! the LLM's responce is assigned to the pythonCode variable
            pythonCode = res?.text
            // Regex to get rid of markdown code blocks syntax
            pythonCode = pythonCode.replace(/^```[a-z]+\n|\n```$/gm, '')
        }

        // Then run the code using Pyodide (only after validating to prevent RCE)
        let finalResult = ''
```

```
        if (pythonCode) {
            const validation = validatePythonCodeForDataFrame(pythonCode)
            if (!validation.valid) {
                throw new Error(
                    `Generated code was rejected for security reasons (${
                        validation.reason ?? 'unsafe construct'
                    }). Please rephrase your question to use only pandas DataFrame opera
                )
            }
            try {
                // !!! The python code is evaluated in a non-sandboxed environment
                const code = `import pandas as pd\n${pythonCode}`
                // TODO: get print console output
                finalResult = await pyodide.runPythonAsync(code)
            } catch (error) {
                throw new Error(`Sorry, I'm unable to find answer for question: "${input
            }
        }

        // Finally, return a complete answer
        if (finalResult) {
            const chain = new LLMChain({
                llm: model,
                prompt: PromptTemplate.fromTemplate(finalSystemPrompt),
                verbose: process.env.DEBUG === 'true' ? true : false
            })
            const inputs = {
                question: input,
                answer: finalResult
            }

            if (options.shouldStreamResponse) {
                const handler = new CustomChainHandler(shouldStreamResponse ? sseStreame
                const result = await chain.call(inputs, [loggerHandler, handler, ...call
                return result?.text
            } else {
                const result = await chain.call(inputs, [loggerHandler, ...callbacks])
                return result?.text
            }
        }

        return pythonCode
    }
}
```

Proof of Concept

A proof of concept for this vulnerability is provided in ./poc.py. It expects the following syntax:

```
    python3 poc.py --method [server OR chatflow OR prompt_injection] [--user <USER
-passwd <password> --host <HOST> --r_host <R_HOST> --r_port <R_PORT> --l_port
<L_PORT> --port <PORT> --cmd <CMD> --chatflow_id <CHAT_ID> --airtable_token
<AIRTABLE_TOKEN> --base_id <BASE_ID> --table_id <TABLE_ID>]
```

Where USER is a username of a user on the server, PASSWORD is the user's password, HOST is the ip address of the vulnerable flowise server, R_HOST is the ip address of a malicious server started by this poc, R_PORT is the port a malicious server started by this poc is listening on (default: 5000), L_PORT is the port a malicious server started by this poc should listening on (default: 5000), PORT is the port the vulnerable flowise server is listening on (default: 3000), CMD is the command to execute on the flowise server (default: xcalc), CHAT_ID is the chatflow id of a chatflow using the Airflow Agent node, AIRTABLE_TOKEN is an api key for an Airtable account, BASE_ID is the base id to find an airflow table in, and TABLE_ID is the airflow table to use.

This poc has three modes of operation controlled by the method argument. The method argument may have any of the values "server" OR "chatflow" OR "prompt_injection".

method = "server"

By default the poc will start a malicious server listening on the port specified by the <L_PORT> value. This server will respond to requests made to the "/api/chat" endpoint with a JSON object containing an LLM response that contains a malicious python script. This python script will execute a command specified by the value.

method = "chatflow"

By default, the poc will first establish an authenticated session on the server using the and arguments. It will then send a POST request to the "/api/v1/chatflow" endpoint with a JSON body containing a crafted chatflow using an Airflow Agent node and a ChatOllama node configured with a server specified by the <R_HOST> and <R_PORT> arguments. The Airflow Agent node will be configured using the <AIRTABLE_TOKEN>, <BASE_ID>, and <TABLE_ID> arguments. The default values of these arguments will be valid for 14 days from 2026-02-22. The poc will then send a POST request to the "/api/v1/internal-prediction/" endpoint in order to trigger a prediction using the chatflow.

It is intended that the server specified in the chatflow is a server started using the server method of this poc. When making a prediction against this chatflow, flowise will send a request to the specified server in order to generate an LLM response. The response recieved by flowise will be evaluated as a python script. Upon successful exploitation, The argument passed to the server method of this poc will be executed on the vulnerable flowise server.

method = "prompt_injection"

By default, the poc will send a POST request to the "/api/v1/prediction/*chat_id*" endpoint, where *chat_id* is a vulnerable chatflow id specified by the <CHAT_ID> parameter. The JSON body of this request will contain a question member whose value will be a prompt containing a prompt injection. Upon successful exploitation, The argument will be executed on the vulnerable flowise server.

Due to the nature of LLM responses, it may take multiple attempts to be successful or require a different prompt injection technique depending on the model used.

Testing Environment

The provided proof of concept was tested using FlowiseAI Flowise version 3.0.13 runing on a Ubuntu 25.10 VM. The prompt injection method was tested using the Llama3.2 model running in Ollama.

Credits

Dre Cura (@dre_cura) and Nicholas Zubrisky (@NZubrisky) of TrendAI Research

-- CREDIT ----------
This vulnerability was discovered by:
Dre Cura (@dre_cura) and Nicholas Zubrisky (@NZubrisky) of TrendAI Research

-- FURTHER DETAILS ----------

Supporting files:

If supporting files were contained with this report they are provided within a password protected ZIP file. The password is the ZDI candidate number in the form: ZDI-CAN-XXXX where XXXX is the ID number.

Please confirm receipt of this report. We expect all vendors to remediate ZDI vulnerabilities within 120 days of the reported date. If you are ready to release a patch at any point leading up to the deadline, please coordinate with us so that we may release our advisory detailing the issue. If the 120-day deadline is reached and no patch has been made available we will release a limited public advisory with our own mitigations, so that the public can protect themselves in the absence of a patch. Please keep us updated regarding the status of this issue and feel free to contact us at any time:

Zero Day Initiative
zdi-disclosures@trendmicro.com

The PGP key used for all ZDI vendor communications is available from:

http://www.zerodayinitiative.com/documents/disclosures-pgp-key.asc

-- INFORMATION ABOUT THE ZDI ----------
Established by TippingPoint and acquired by Trend Micro, the Zero Day Initiative (ZDI) neither re-sells vulnerability details nor exploit code. Instead, upon notifying the affected product vendor, the ZDI provides its Trend Micro TippingPoint customers with zero day protection through its intrusion prevention technology. Explicit details regarding the specifics of the vulnerability are not exposed to any parties until an official vendor patch is publicly available.

Please contact us for further details or refer to:

http://www.zerodayinitiative.com

-- DISCLOSURE POLICY ----------

Our vulnerability disclosure policy is available online at:

http://www.zerodayinitiative.com/advisories/disclosure_policy/

How This Differs from CVE-2026-41138

CVE-2026-41138 introduced sanitization for the Pyodide code ran by the Airtable Agent. This advisory demonstrates a bypass of that sanitization, which we addressed separately by disallowing imports outright.

Severity

Critical 9.2 / 10

CVSS v4 base metrics

Exploitability Metrics	
Attack Vector	Network
Attack Complexity	High
Attack Requirements	Present
Privileges Required	None
User interaction	None
Vulnerable System Impact Metrics	
Confidentiality	High
Integrity	High
Availability	High
Subsequent System Impact Metrics	
Confidentiality	None
Integrity	None
Availability	None

Learn more about base metrics

CVSS:4.0/AV:N/AC:H/AT:P/PR:N/UI:N/VC:H/VI:H/VA:H/SC:N/SI:N/SA:N

CVE ID

CVE-2026-41265

Weaknesses

No CWEs

Credits

zdi-disclosures Reporter