# AI-Crafted Custom Resources for Threat Intelligence & Hunting

# This project:

It's a AI automated Threat Intelligence and hunting that creates custom resources. This will do the collecting and automating for you and provide some AI content which the AI has gone through. This project can be expanded to be scraping on a bigger scale or adding extra things to improve it.

### To note:

- This was done by running n8n on my local machine
- This was also using ChatGPT to vibe code JavaScript for each node
- This does take some time to run because it's going to go through more than 10 news and depending that it contains key words
- If it does pass and that it contains key words the ai might output nothing of the "use case" or "query"

## Tools used:

- n8n
- thehackernews.com
- ChatGPT
- DeepSeek R1
- Google sheets

### API/connections used:

- Jina AI (jina.ai)
- OpenRouter (openrouter.ai)
- Google sheets

## Programming and Markup language:

- JavaScript
- HTML

## **Problems:**

- This has only be test for than 20 times but you can come across maybe either false positive (FP) and false negative (FN).
- It only scrapes one page it cannot go to the next page where there are more news article so if a bunch of news came in and there are 2 pages of todays news. It will only scrape the first page
- Sometimes the code node will give blanks in either title, description or MITRE (That's more of the coding end)
- It uses a lot of tokens for this to run. Running it around 3 or 4 times a day should be fine. But if it uses too much tokens the API service will not be able to be used for the next 24 hours (I was using a free version).

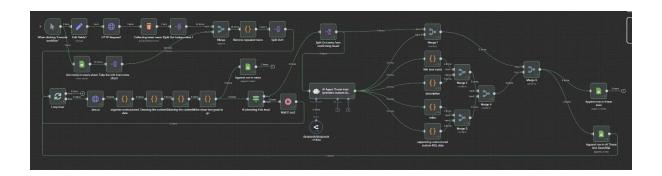
## How does it work:

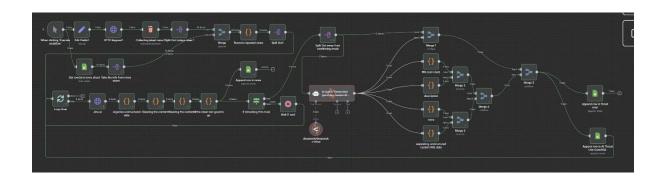
- 1. This will go through "https://thehackernews.com/" site
- 2. It will then scrape all the news on the first page of "https://thehackernews.com/"
- 3. The scraping would be getting the HTML
- 4. This would be treated as Temporary data
- 5. This Temporary data contains the URL
- 6. This Temporary URL will be compare to Permanent URL in the excel sheet
- 7. If "Temporary URL == Permanent URL" the Temporary URL will not go through, only the ones that are not equal to the Permanent URL will go through
- 8. It will then go into a Loop which will go through the number of URL there are
- 9. Each URL news will go in one at a time
- 10. Jina AI will extract the content from the URL
- 11. That content will go through a cleaning process
- 12. The cleaning process (the code) will get rid as much of the unnecessary stuff as possible
- 13. It will then go through the new content and look for key words
- 14. Once its finished find or not find the keyword it will append the excel sheet

- 15. It appends the sheet for the reason that if this is run again you will not get 2 of the same information
- 16. Back to find the key word, once found or not it will be put through a IF statement
- 17. If a key word was found in the content it will give a true value and move the content to the AI agent
- 18. If it was false, it will loop back again and start doing the next URL
- 19. Once the ai is finished with the content it output the "AI content"
- 20. The AI content will go through 4 codes that each one will do either getting the title, description, MITRE ATT&CK, and KQL Queries.
- 21. They will all be merge together one by one
- 22. Once finished it will be append to the excel sheets
- 23. To note that all JavaScript code was vibe code

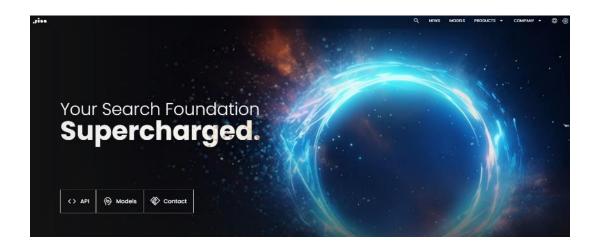
The Process:

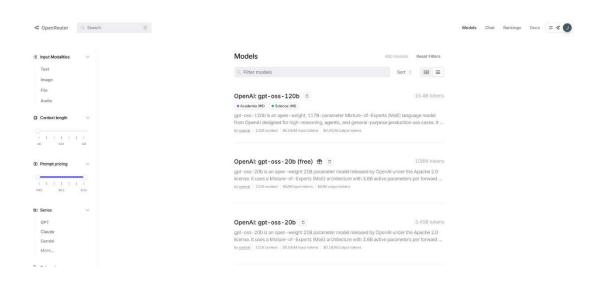
This is the workflow:



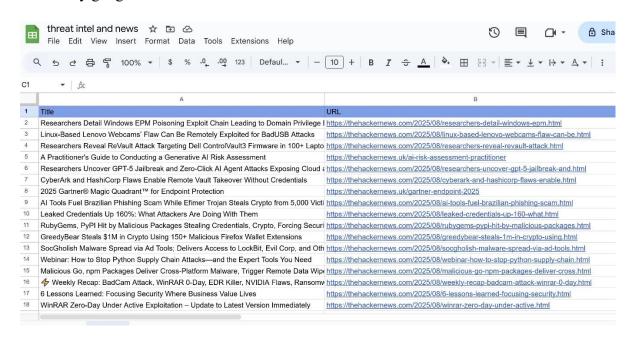


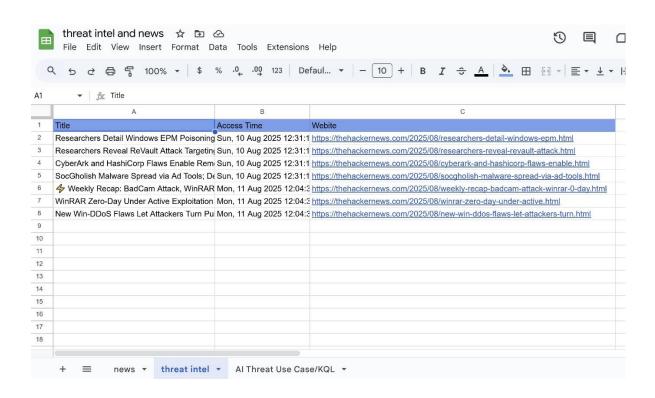
These are the 2 sites where I got my API from. (jina ai and openrouter):

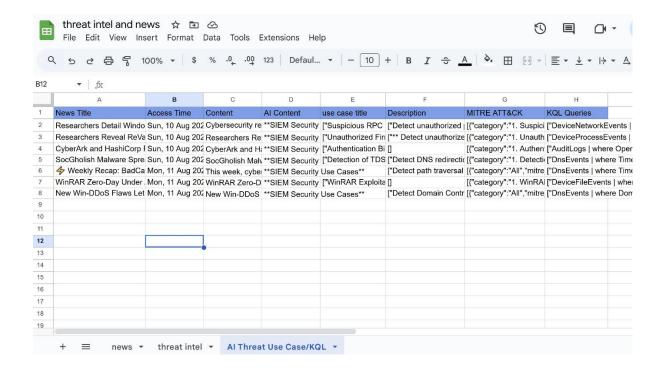




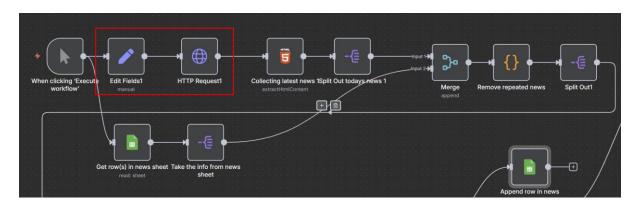
#### This is my google sheets result:



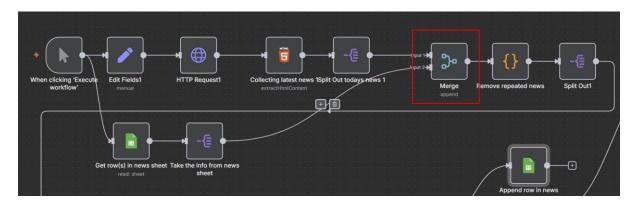




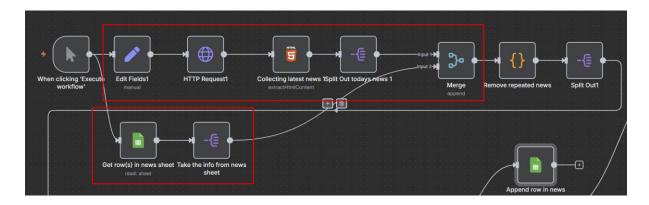
This contains the URL of "https://thehackernews.com/":



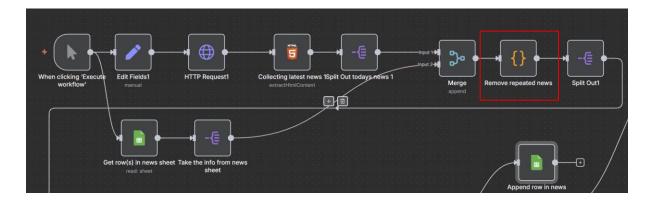
This extract the HTML content of the website only for "https://thehackernews.com/" which then gets to the merge node:



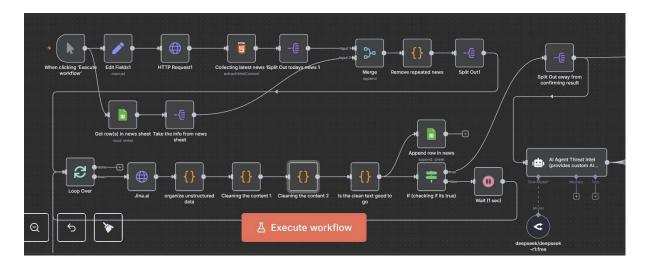
Same for the google sheet that its just getting the URL and having Temporary URL and Permeant URL:



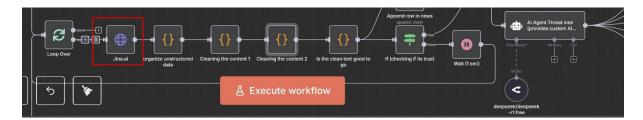
Then it gets pushed in a code node (JavaScript) where it outputs the URL that are not equal to each other:



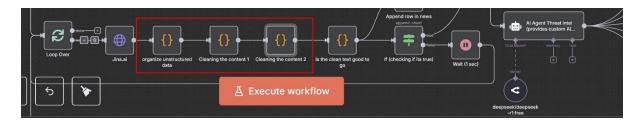
The list will of URL will go through a Loop:



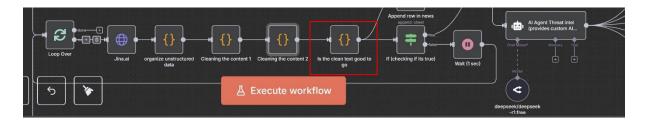
This will go into AI scrape called Jina AI which the node and AI is setup in a way that it will go through each news one at a time:



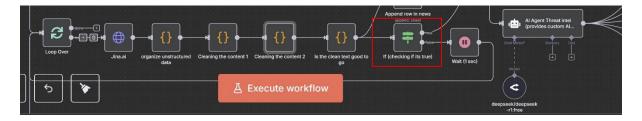
It then goes through 3 code nodes (JavaScript) which the 1st nodes separates the content from everything else. The 2nd and 3rd is trying to get as much of the tiny detail because this content will be pushed into the AI agent.



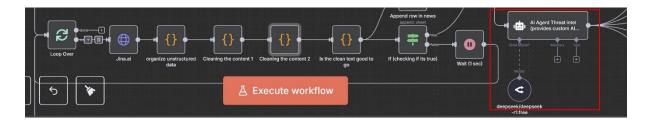
The Last node will check in the cleaned content any key words which could be domains, URL or CVE. If contains any of the threat hunting keywords then it will give a true, if not false:



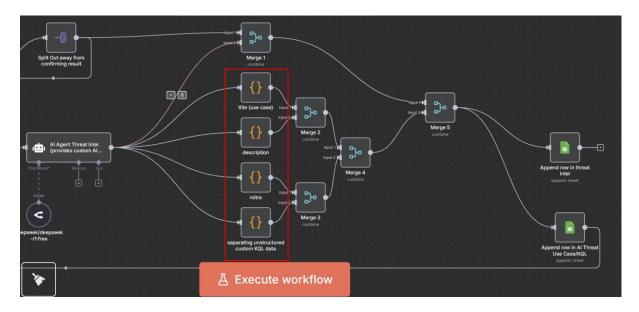
This then gets append into a google sheet which become the database where this will help the 1st google sheet if there are any duplicates in the future. It can filter out the URLs.



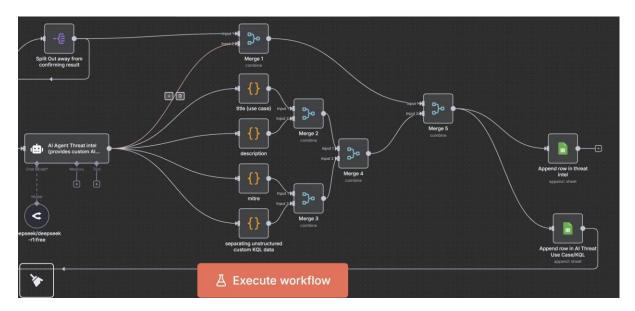
The IF statement checks if it's true or false. If true get the content pushed to the AI Agent. If false move onto the next URL.



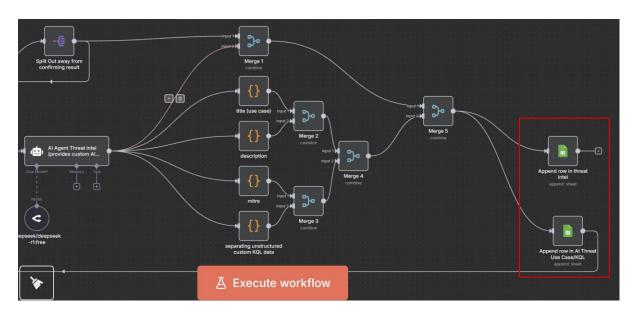
It gets pass to the AI agent to analyse the content and produce the AI content, title, description, MITRE ATT&CK, and KQL queries:



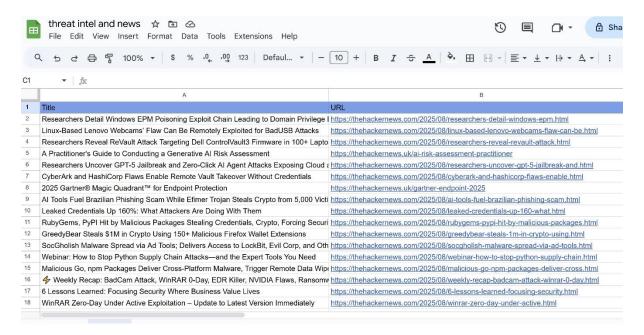
Once everything is done, everything gets put into 1 row. The reason there is not just 1 merge node is that it will not create 1 row but multiple which a lot with give "[undefined]":



#### Once its all in one row it gets append into the excel sheets



This was the first excel to get append information which just keeps all the news. It does not matter if it was pass by ai or not:



This sheet is for the ones that have been pass through the AI agent:

