

Agent AI, Basta Parser Extraordinaire

By Joshua Platt

Published: 2025-02-28 · Archived: 2026-04-05 21:33:21 UTC



4 min read

Feb 28, 2025

Black Basta is a ransomware group that has spent the past couple of years attacking global networks. Their activities are well known in the cybersecurity space. Some might even say prolific at this point.

On Feb 11, 2025 internal communication amongst the group was publicly leaked¹. The leaked data consisted of matrix server chat logs with information pertaining to the day to day operations of the group. But in order to better understand the communication of the group, you first need to inspect and parse the leaked file. While the dataset has been analyzed publicly, including a BlackBastaGPT² release for public use, this post delves into the utilization of AI to parse and further enable the investigation of the dataset.

After acquiring the leaked data from a public repository³ a cursory check of the file was conducted. The file is over 47MB in size, which makes it less than ideal for text editors.

```
MD5: 2f95cf2c7a2dc364b8530b7cc03d13ec
SHA1: e23008b0cc8bb8916b1c7bfaa4777f253fe2bcb7
SHA256: 5d8d88da1086475546d551a5735c1d46df0ef659b5cd549f84d944641a050fbb
```

The file output appears below with a few characteristics. It appears to be Unicode, UTF-8 text and has very long lines. Oddly enough, it does not detect the file as JSON text data.

```
file blackbasta_chats.json
blackbasta_chats.json: Unicode text, UTF-8 text, with very long lines (469)
```

Let's check the file using python. The output from python's json tool is unable to properly parse the file, which is a second indication that the file is going to need some work.

```
python3 -m json.tool blackbasta_chats.json
Expecting property name enclosed in double quotes: line 2 column 5 (char 6)
```

Using the head command, we can extract the first few lines of the file and verify the structure.

```
head -n20 blackbasta_chats.json
{
timestamp: 2023-09-18 13:35:07,
chat_id: !VdvDXHFZwWdpIAtpCj:matrix.bestflowers247.online,
sender_alias: @usnamenn:matrix.bestflowers247.online,
message: BAZA
}
{
timestamp: 2023-09-18 13:50:31,
chat_id: !uJZKZVgGmmSiNvobZH:matrix.bestflowers247.online,
sender_alias: @usnamess:matrix.bestflowers247.online,
message: !!!
}
```

We have a few options here. We can take the file and write a parser ourselves. We could attempt to properly enclose the key pairs in quotes. But let's see what GPT has to offer that might make this all a bit faster.

Press enter or click to view image in full size

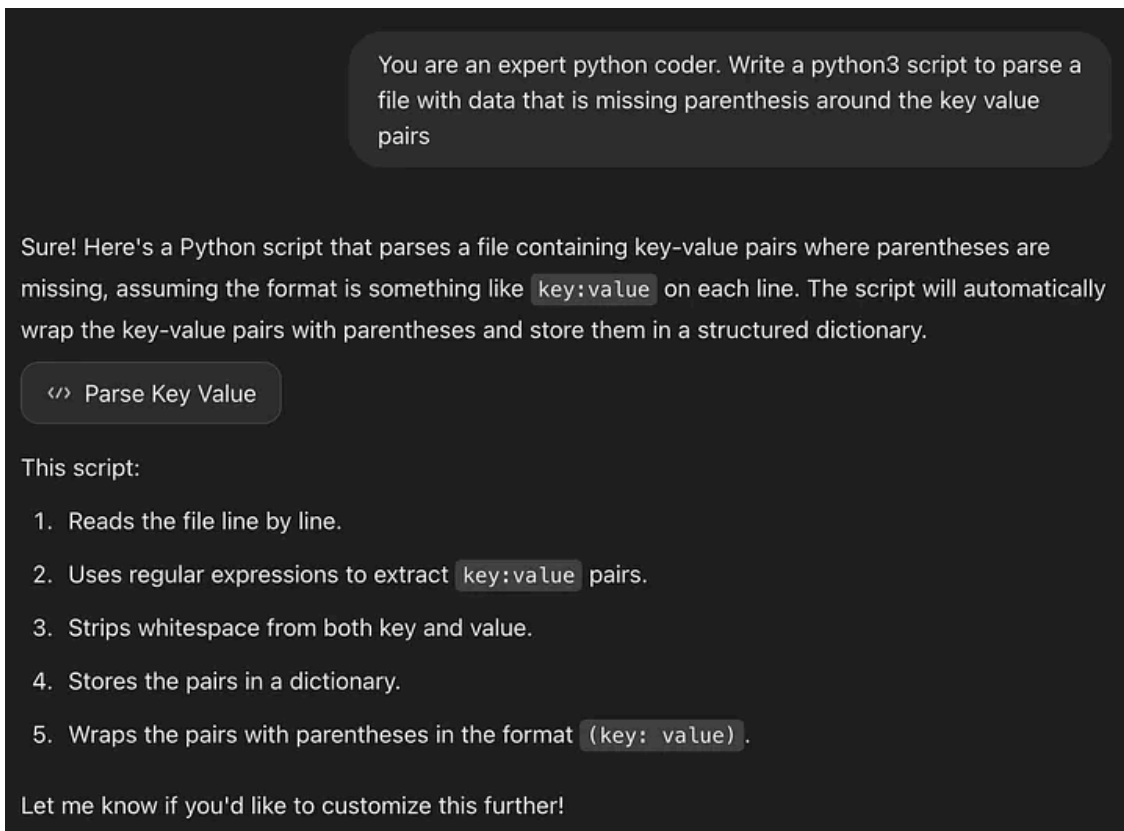


Image1. Initial Prompt

Next we suggest modifications to write the output to file along with fixing issues with the initial output.

Press enter or click to view image in full size

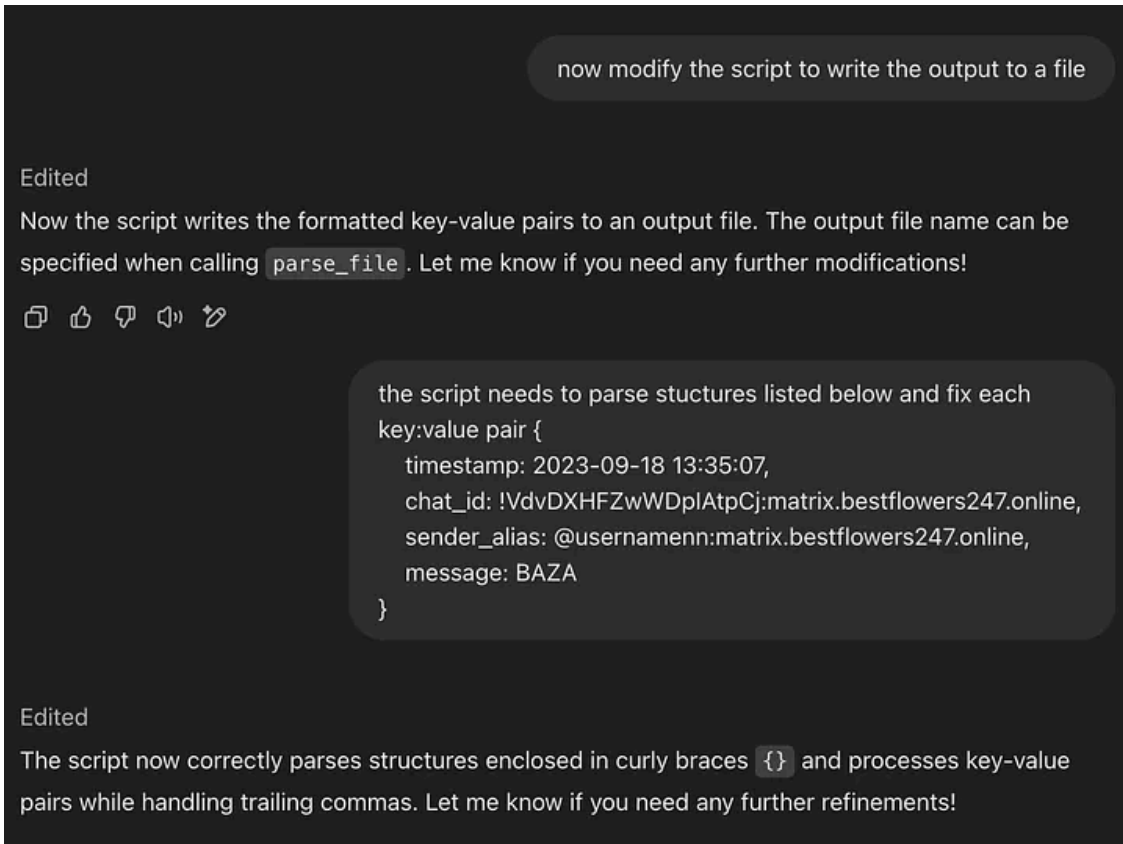


Image 2: Modifications

After inspecting the code, GPT utilized a regex for matching data inside the brackets, which was not efficient at all.

```
match = re.match(r'(\w+):\s*(.*)', line)
```

We can prompt GPT to remove the regex and utilize the comma delimiter instead.

Press enter or click to view image in full size

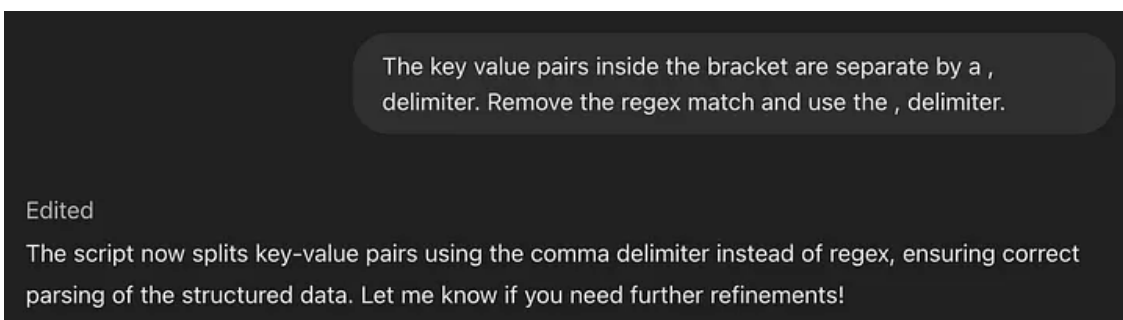


Image 3: Modify parsing.

After training the model for sorting through syntax related irregularities in the dataset, it was time to output the dataset into a Sqlite database.

Press enter or click to view image in full size

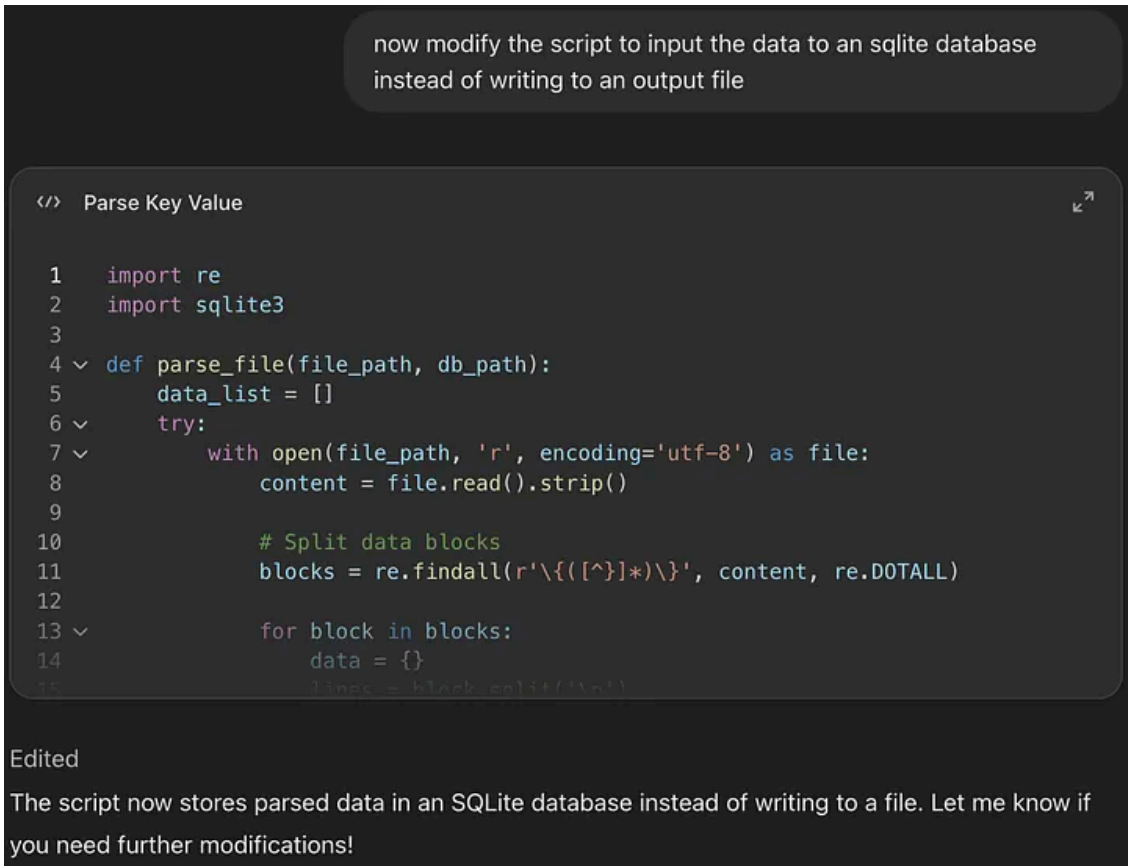


Image 4: Script to output Sqlite database

The sqlite database from the generated python script is shown below.

Press enter or click to view image in full size

	id	timestamp	chat_id	sender_alias	message
1	1	2023-09-18 13:35:07,	!VdvDXHFZw0pIAtPcj:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	BAZA
2	2	2023-09-18 13:50:31,	!uJZKZVgGmSINvobZH:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	!!!
3	3	2023-09-18 17:43:18,	!FtoGkSqUPiGjGNkK0L:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	cpu 2core 2.4 ghz, 4 gb ram, 100 gb ...
4	4	2023-09-18 17:44:42,	!FtoGkSqUPiGjGNkK0L:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	'''
5	5	2023-09-18 17:47:48,	!kJVcUcyUsQhw@CuIPD:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	1
6	6	2023-09-18 17:47:56,	!kJVcUcyUsQhw@CuIPD:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	'''
7	7	2023-09-18 17:48:02,	!kJVcUcyUsQhw@CuIPD:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	matrix.bestflowers247.online
8	8	2023-09-18 17:48:07,	!kJVcUcyUsQhw@CuIPD:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	cpu 2core 2.4 ghz, 4 gb ram, 100 gb ...
9	9	2023-09-18 17:49:52,	!RMdGGuCKLBreGJPvLH:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	!
10	10	2023-09-18 17:50:16,	!RMdGGuCKLBreGJPvLH:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	https://torguard.net/
11	11	2023-09-18 17:59:32,	!kJVcUcyUsQhw@CuIPD:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	ку
12	12	2023-09-18 17:59:37,	!kJVcUcyUsQhw@CuIPD:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	привет
13	13	2023-09-18 17:59:45,	!RMdGGuCKLBreGJPvLH:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	привет
14	14	2023-09-18 17:59:57,	!B0pqkyIMhBRfCPXwod:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	привет
15	15	2023-09-18 18:00:08,	!RMdGGuCKLBreGJPvLH:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	Привет!
16	16	2023-09-18 18:00:35,	!RMdGGuCKLBreGJPvLH:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	bclq9ee7wtrvjeu7vancjgdup3kcy66cyx74...
17	17	2023-09-18 18:08:11,	!kJVcUcyUsQhw@CuIPD:matrix.bestflowers247.online,	@username:matrix.bestflowers247.online,	Запросы на платный валом хешей

Image 5: Parsed messages in stored in sqlite database

The prompts below were used to further refine the structure of the database.

Get Joshua Platt's stories in your inbox

Join Medium for free to get updates from this writer.

Remember me for faster sign in

PROMPT:

separate the chat_id into two separate columns in the database using the : delimiter. Name the first column room and the second column room_server. Separate the sender_alias into two columns using the : delimiter. Name the first column sender_user and the second column sender_server.

PROMPT:

now create a second message column named translated_message. Using google translate, the script needs to translate any Russian language messages to us english and insert them into the translated message column.

Press enter or click to view image in full size

	id	timestamp	room	room_server	sender_user	sender_server	message
	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1	2023-09-18 13:35:07,	!VdyvX0HFzW0pIAtPcJ	matrix.bestflowers247.online,	@usernameenn	matrix.bestflowers247.online,	BAZA
2	2	2023-09-18 13:50:31,	!uJZKZVgGmSiNvobZH	matrix.bestflowers247.online,	@usernameess	matrix.bestflowers247.online,	!!!
3	3	2023-09-18 17:43:18,	!Fto0KsqUPiGjGNkx0L	matrix.bestflowers247.online,	@usernameyy	matrix.bestflowers247.online,	cpu 2core 2.4 ghz, 4 gb ram, 100 gb _
4	4	2023-09-18 17:44:42,	!Fto0KsqUPiGjGNkx0L	matrix.bestflowers247.online,	@usernameyy	matrix.bestflowers247.online,	'''
5	5	2023-09-18 17:47:48,	!k3VclcyUs0hw@CuIPD	matrix.bestflowers247.online,	@usernameyy	matrix.bestflowers247.online,	1
6	6	2023-09-18 17:47:56,	!k3VclcyUs0hw@CuIPD	matrix.bestflowers247.online,	@usernameyy	matrix.bestflowers247.online,	'''
7	7	2023-09-18 17:48:02,	!k3VclcyUs0hw@CuIPD	matrix.bestflowers247.online,	@usernameyy	matrix.bestflowers247.online,	matrix.bestflowers247.online
8	8	2023-09-18 17:48:07,	!k3VclcyUs0hw@CuIPD	matrix.bestflowers247.online,	@usernameyy	matrix.bestflowers247.online,	cpu 2core 2.4 ghz, 4 gb ram, 100 gb _
9	9	2023-09-18 17:49:52,	!RMjGGuCKLBreGjPwLH	matrix.bestflowers247.online,	@usernameess	matrix.bestflowers247.online,	!
10	10	2023-09-18 17:50:16,	!RMjGGuCKLBreGjPwLH	matrix.bestflowers247.online,	@usernameess	matrix.bestflowers247.online,	https://torguard.net/
11	11	2023-09-18 17:59:32,	!k3VclcyUs0hw@CuIPD	matrix.bestflowers247.online,	@usernameegg	matrix.bestflowers247.online,	ку
12	12	2023-09-18 17:59:37,	!k3VclcyUs0hw@CuIPD	matrix.bestflowers247.online,	@usernameegg	matrix.bestflowers247.online,	привет
13	13	2023-09-18 17:59:45,	!RMjGGuCKLBreGjPwLH	matrix.bestflowers247.online,	@usernameegg	matrix.bestflowers247.online,	привет
14	14	2023-09-18 17:59:57,	!B0pqkyIMnBRfCPxwod	matrix.bestflowers247.online,	@usernameegg	matrix.bestflowers247.online,	привет
15	15	2023-09-18 18:00:08,	!RMjGGuCKLBreGjPwLH	matrix.bestflowers247.online,	@usernameess	matrix.bestflowers247.online,	Привет!!
16	16	2023-09-18 18:00:35,	!RMjGGuCKLBreGjPwLH	matrix.bestflowers247.online,	@usernameess	matrix.bestflowers247.online,	bclq9ee7wtrvjeu7vancgdup3kcy66cyx74n3f_
17	17	2023-09-18 18:08:11,	!k3VclcyUs0hw@CuIPD	matrix.bestflowers247.online,	@usernameess	matrix.bestflowers247.online,	Запросы на платный взлом хешей

Image 6: Modified Database

For one final task, the script was modified to adjust the table and include a column for translated messages along with converting the messages to English prior to storing them.

PROMPT:

now create a second message column named translated_message. Using google translate, the script needs to translate any Russian language messages to English and insert them into the translated message column. The python script should ignore any messages with ip addresses or emails.

Results may vary and the prompts here can definitely be improved. Overall, AI was highly effective in cutting down the time necessary to properly format and store the data for analysis. Incorporating AI into your workflow can save a substantial amount of time and improve your overall ability to leverage larger datasets.

Source: https://medium.com/walmartglobaltech/agent-ai-basta-parser-extraordinaire-24edfc59992a