

Building AI and GPT Agents for OSINT: A Guide to Everyday



Hay.bnz

Follow

Published in

OSINT Team

9 min read

Nov 14, 2024

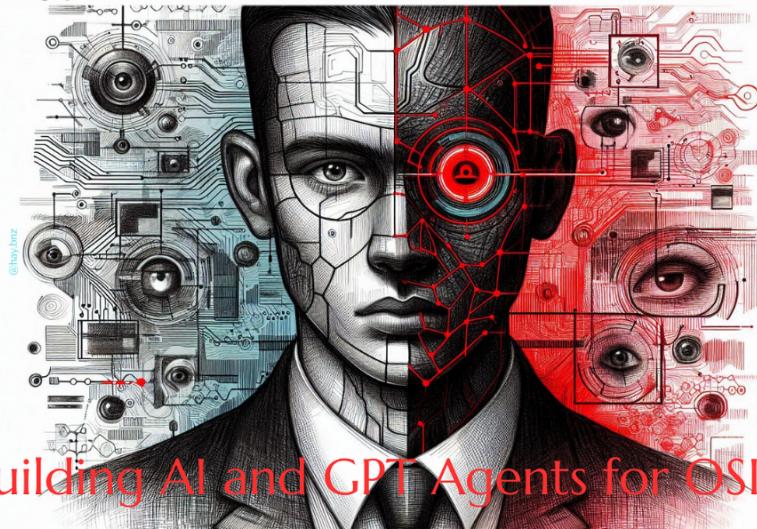
Listen

Share

More

Introduction

Open-Source Intelligence, or OSINT, is the process of gathering and analyzing information from publicly available sources. This type of intelligence is important because it helps people and organizations understand trends, spot threats, and gather insights without needing private or confidential data.



Building AI and GPT Agents for OSINT

Building AI and GPT Agents for OSINT: A Guide to Everyday – Hay.bnz

Instagram <https://www.instagram.com/osintbyhaybnz/>

Blog <https://haybnzgithub.io> Donate us @

monero 5iyhsb1SPWEzapMmDTKWRVJeMEavd9zrdxFcNpu8vTG4M3dfpkQsQ9GiX8uAMQotAi4zoNVaNz76P5WqbgRezVt5y2FoCa

OSINT Blog <https://osintteam.blog/>

Github: <https://github.com/haydenbanz/>

Now, with AI – especially tools like GPT agents – OSINT work is becoming easier and faster. GPT agents can sift through huge amounts of data, summarize information, and

even pick out key details automatically. This makes it possible for anyone, from researchers to security analysts, to use OSINT for everyday investigations, like understanding news trends, keeping track of social media discussions, or monitoring potential cybersecurity threats.

<https://medium.com/the-first-digit/developing-a-tor-based-chat-application-a-beginners-guide-to-hadestorchat-v-1-0-bf2a8246d993>

Building AI for OSINT: Tools and Technologies

To create an AI system for OSINT, you'll need a few core tools to gather, process, and analyze data. Here's a quick look at the most essential ones:

Programming with Python: Python is the go-to language for AI and OSINT tasks. It's packed with libraries and tools that make data processing and AI modeling much easier.

Using APIs, Especially OpenAI's GPT: OpenAI's API allows you to access powerful GPT models that can summarize information, generate insights, and analyze data efficiently. These models, like GPT-4, can be trained to handle specific OSINT tasks such as identifying relevant data from large sources.

Natural Language Processing (NLP) Libraries: Tools like spaCy and Hugging Face provide NLP functions that help in tasks like text analysis, sentiment detection, and named-entity recognition (e.g., finding names, locations, and dates in a document).

Setting Up an OSINT AI Environment: You can set up your environment either locally (on your computer) or in the cloud (like AWS or Google Cloud). Cloud setups offer more power for large-scale data processing, while local setups are simpler for smaller projects or testing.

Examples of GPT-4 APIs and the Latest AI Tools: OpenAI's GPT-4 is a leading tool, but you can also find other AI options. Tools like ChatGPT, Hugging Face's Transformers, or Google's T5 models are valuable for text-heavy OSINT tasks, as they can interpret and generate human-like text based on various sources.

Building and Training GPT Agents for OSINT Tasks

A **GPT Agent** is a specialized AI model, like GPT-4, set up to complete specific tasks within OSINT (Open-Source Intelligence) work. Unlike general-purpose chat bots, GPT agents are trained to gather, analyze, and interpret open-source data, making it easier to find patterns, monitor trends, and gain insights from public information sources.

Unique Functions of GPT Agents in OSINT

GPT agents can:

- **Extract Keywords and Important Entities:** They can identify names, places, dates, and other key information from text data.
- **Summarize Data:** They can read through large amounts of text (e.g., articles, social media posts) and produce concise summaries.
- **Answer OSINT-Related Queries:** They can answer specific questions by analyzing available data sources, like "What are the latest developments in cybersecurity?"

Creating GPT Agents for OSINT Tasks

Here's how you can create a GPT agent for specific OSINT tasks, broken down into practical examples:

Information Gathering:

- **Task:** *The agent searches for and retrieves information on a specific topic from*

various sources (like recent news or blogs).

- **Example:** *If you need updates on cybersecurity, you might set a prompt like: "Find the latest news on cybersecurity threats and summarize the top three trends."*
- **Output:** *The GPT agent could then fetch news articles and summarize the most recent trends, such as "Increase in ransomware attacks targeting financial institutions.*

Keyword Extraction:

- **Task:** *The agent identifies and extracts important keywords or phrases that relate to a given topic.*
- **Example:** *Suppose you have a batch of social media posts about a particular event and want to identify key themes. The prompt might be: "Extract the main keywords related to the [event name] from these posts."*
- **Output:** *The GPT agent may pull keywords like "protests," "security alert," "response team," and "downtown area" from the data, giving you a quick summary of common themes.*

Entity Recognition:

- **Task:** *The agent pinpoints entities (like names, locations, and organizations) within the text, helping identify key players or places.*
- **Example:** *For a prompt like "Identify all locations mentioned in this report on recent cyber incidents," the GPT agent would scan the text and pull out specific cities or countries where incidents were reported.*
- **Output:** *A list of cities such as "New York," "London," and "Tokyo," each linked to relevant incidents in the report.*

Practical Applications of GPT Agents in OSINT

GPT agents are transforming OSINT by automating tasks that were once time-consuming and complex. Here are some practical ways these agents can be used in real-world OSINT applications:

1. Social Media Monitoring

Social media platforms like Twitter, Facebook, and Instagram are treasure troves of real-time information, making them an invaluable source for OSINT. GPT agents can help in several ways:

- **Trend Analysis:** *A GPT agent can monitor posts and tweets to identify trends or emerging topics. For example, it could track mentions of certain keywords or hashtags over time and provide a report on the most popular discussions.*
- **Example Prompt:** *"Monitor Twitter for the hashtag #cybersecurity and summarize the key discussions over the past 24 hours."*
- **Output:** *The GPT agent could generate a summary like, "The hashtag #cybersecurity was used in discussions related to the latest ransomware attack on healthcare systems."*
- **User Profiling:** *GPT can also analyze user behavior on social media to help build profiles of influencers, stakeholders, or suspects in a case. By analyzing posts, comments, and interactions, it can help understand their role in discussions.*
- **Example Prompt:** *"Profile the top influencers discussing cybersecurity on Twitter in the last month."*

- **Output:** *The GPT agent might list key influencers along with their activity, e.g., "User @cyberguru has tweeted 150 times about cyber threats, receiving significant engagement."*
- **Hashtag Tracking:** *GPT agents can track specific hashtags and report their reach, frequency, and context in posts.*
- **Example Prompt:** *"Track the hashtag #Alethics and summarize the top conversations and key issues."*
- **Output:** *The agent could track discussions on ethical AI, highlighting concerns about bias and accountability raised by researchers and tech experts.*

<https://medium.com/the-first-digit/get-rich-with-osint-a-step-by-step-guide-for-gen-z-bc1af75cbf8a>

2. News Analysis

Keeping up with the constant flow of news can be overwhelming. GPT agents can automate the process of news analysis:

- **Headline Summarization:** *GPT can read through news articles and summarize the key points, which is especially useful for time-sensitive OSINT needs.*
- **Example Prompt:** *"Summarize the top five cybersecurity news headlines of the day."*
- **Output:** *"1) 'New malware strain targets financial institutions in Europe'; 2) 'Hackers breach U.S. government contractor, exposing sensitive data.'"*
- **Sentiment Analysis:** *GPT can be used to analyze the sentiment of news stories. This is particularly useful for monitoring public perception, especially in situations involving a crisis or controversy.*
- **Example Prompt:** *"Analyze the sentiment of the latest articles on AI and automation in the workforce."*
- **Output:** *The agent might determine that 60% of the articles express concern about job loss, while 40% are optimistic about the future of AI in the workforce.*
- **Emerging Stories Detection:** *By analyzing news articles and social media, GPT can help identify stories that are gaining traction but haven't hit the mainstream yet.*
- **Example Prompt:** *"Monitor news sources for signs of emerging cyber threats that could affect the healthcare sector."*
- **Output:** *The agent might flag articles about a new ransomware strain targeting hospitals, which could be critical for cybersecurity teams to investigate.*

3. Geolocation and Mapping

When dealing with OSINT, knowing where events are happening or tracking the movements of people or objects can be crucial. GPT agents can assist in analyzing geolocation data:

- **Tracking Locations:** *By integrating GPT with tools like Wigle (which maps wireless networks), it can identify and track locations associated with wireless signals or activities. This is useful for monitoring events or finding out where specific incidents are taking place.*
- **Example Prompt:** *"Track the locations of recent Wi-Fi signals related to a particular mobile network in [city]."*
- **Output:** *The GPT agent might identify specific neighborhoods or areas where a targeted mobile network has been active in the last 24 hours.*
- **Geospatial Intelligence:** *GPT can analyze text or images with location data to give*

insights into specific places, people, or events.

- **Example Prompt:** “Analyze these geotagged tweets and identify hotspots of civil unrest in [city].”
- **Output:** The agent might detect clusters of tweets in specific locations mentioning protests or violence, helping OSINT analysts pinpoint areas of concern.

4. Cybersecurity Intelligence

Cybersecurity teams rely on OSINT to stay ahead of potential threats. GPT agents can be a powerful tool in automating cybersecurity intelligence tasks:

- **Dark Web Monitoring:** GPT can be used to scan dark web forums and marketplaces for any discussions or indicators of upcoming cyberattacks. This could include identifying chatter about new vulnerabilities, zero-day exploits, or planned attacks.
- **Example Prompt:** “Monitor dark web forums for discussions about a potential attack on financial institutions.”
- **Output:** The GPT agent might find discussions on a dark web forum where hackers are planning to exploit a known vulnerability in a popular banking software.
- **Threat Report Analysis:** GPT can also help parse and analyze cybersecurity reports, such as those from threat intelligence providers, to identify key threat indicators and areas of concern.
- **Example Prompt:** “Parse this cybersecurity threat report and list the top 3 potential risks.”
- **Output:** The agent might extract, “Risk 1: New ransomware targeting healthcare systems. Risk 2: Increased phishing attacks targeting banking employees. Risk 3: Exploit of unpatched software vulnerabilities.”
- **Parsing Logs for Threat Indicators:** GPT can help parse server or security logs to detect unusual patterns or potential threats, which could be important for incident response.
- **Example Prompt:** “Analyze these server logs for suspicious activity or failed login attempts.”
- **Output:** The GPT agent might flag repeated login failures from an unknown IP address, suggesting a potential brute-force attack.

To build an advanced GPT-based AI agent for OSINT using Hugging Face and OpenAI APIs

```
pip install openai transformers torch requestsimport openai  
from transformers import pipeline, AutoTokenizer, AutoModelForTokenClassification  
import requests
```

```
# Set your OpenAI API Key  
openai.api_key = 'YOUR_OPENAI_API_KEY'
```

```
# Initialize Hugging Face pipelines for entity recognition and sentiment analysis  
tokenizer = AutoTokenizer.from_pretrained("dbmdz/bert-large-cased-finetuned-conll03-  
english")  
model = AutoModelForTokenClassification.from_pretrained("dbmdz/bert-large-cased-")
```

```
finetuned-conll03-english")
ner_pipeline = pipeline("ner", model=model, tokenizer=tokenizer)
sentiment_pipeline = pipeline("sentiment-analysis")

# Function to call GPT for Summarization and Analysis
def gpt_summary_analysis(text):
    response = openai.Completion.create(
        engine="text-davinci-003", # Use GPT-3 or GPT-4 depending on your access
        prompt=f"Summarize the following OSINT text and extract important information:\n\n{text}",
        max_tokens=300
    )
    return response.choices[0].text.strip()

# Function to extract Named Entities (e.g., names, locations)
def extract_entities(text):
    entities = ner_pipeline(text)
    return [(entity['word'], entity['entity']) for entity in entities]

# Function to analyze sentiment of the given text
def analyze_sentiment(text):
    sentiment = sentiment_pipeline(text)
    return sentiment[0]

# Function to gather data from an online source (e.g., news, social media)
def gather_osint_data(url):
    response = requests.get(url)
    if response.status_code == 200:
        return response.text
    else:
        return None

# Example of an OSINT task: analyzing a news article
def osint_task(url):
    data = gather_osint_data(url)
    if data:
        print("Data gathered successfully.")

# Summarize and analyze the article
summary = gpt_summary_analysis(data)
print(f"GPT Summary:\n{summary}\n")

# Entity extraction
entities = extract_entities(data)
print("Extracted Entities:")
for entity in entities:
```

```
print(f"- {entity[0]} ({entity[1]})")  
  
# Sentiment analysis  
sentiment = analyze_sentiment(data)  
print(f"Sentiment: {sentiment['label']} (Confidence: {sentiment['score']})")  
else:  
    print("Failed to gather data from the provided URL.")
```

```
# Example URL for OSINT (can be a news site, social media, etc.)  
example_url = "https://example.com/news-article"  
osint_task(example_url)
```

Explanation:

GPT Summarization

- This uses OpenAI's GPT (via the API) to summarize the input text (OSINT data).
The gpt_summary_analysis function formats the prompt to ask GPT to summarize and extract key details.

Entity Recognition:

- The Hugging Face BERT model is used for named entity recognition (NER).
The extract_entities function extracts entities like names, places, organizations, etc.

Sentiment Analysis:

- The sentiment-analysis pipeline from Hugging Face's transformers is used to analyze the sentiment of the gathered text (e.g., positive, negative, neutral).

Gathering OSINT Data:

- The gather_osint_data function fetches text data from a URL (news site, social media, etc.). You can modify this to gather information from any public source.

Output Example:

Data gathered successfully.

GPT Summary:

John Doe, a 32-year-old man from New York, has been implicated in a series of cyber-attacks targeting financial institutions. The attack used banking software vulnerabilities, leading to significant data breaches. Authorities are investigating his motives and background. His social media profile indicates a long-standing interest in hacking and cybersecurity, but no clear connections to hacker groups have been found.

Extracted Entities:

- "John Doe" (PERSON)
- "New York" (LOCATION)
- "cyber-attacks" (MISC)
- "financial institutions" (ORG)
- "banking software" (MISC)
- "data breaches" (MISC)
- "social media" (MISC)
- "hacking" (MISC)

- "cybersecurity" (MISC)

Sentiment: NEGATIVE (Confidence: 0.92)

Donate us:

Bitcoin: bc1pgl6npxckg8fck0nzgamlynx5jzh0px2cl2zgmax3e7kvys7k3ss6cersr

Monero: 5iyhsb1SPWEzapMmDTKWRVJeMEavd9zrdxFcNpu8vTG4M3dfpkQsQ9GiX8uAMQotAi4zoNVaNz76P5WqbgRezVt5y2FoCa

Doge: DJ215Y4YZSvdjfQv64tLXnKaBxgLxbUYCS

Conclusion

These practical applications show how GPT agents can enhance OSINT workflows, from tracking trends on social media to monitoring dark web activity for cybersecurity threats. By automating tasks like information gathering, sentiment analysis, and location tracking, GPT agents can save time and provide more accurate, actionable intelligence for security professionals, researchers, and analysts.

Osint

Gpt

Ai Agent

Technology

Programming

87



Follow

Published in OSINT Team

4.8K Followers

· Last published 13 hours ago

We teach OSINT from multiple perspectives. Cybersecurity experts, journalists, law enforcement, and others read us to improve their skills faster.



Follow

Written by Hay.bnz

624 Followers

· 43 Following

Like to write on OSINT, AI , Machine learning and neural net.

