

AI Agent — Documentation

Project Overview

This project demonstrates an **AI Content Repurposing Agent** that automatically extracts information from existing content sources — such as **blogs, YouTube videos, PDFs, or text files** — and **repurposes it** into new formats like:

- ✨ Short summary
- 📁 LinkedIn post
- 🐦 Tweet thread
- 🎬 60-second video script

It helps creators and professionals save time by automating content transformation using **Python, LangChain, and Hugging Face** — all **without OpenAI API keys**.

Tools & Libraries Used

- **Python 3.9+**
- **LangChain** – for chaining NLP tasks together
- **Transformers (Hugging Face)** – for summarization & text generation
- **Requests + BeautifulSoup4** – for extracting blog content
- **YouTube Transcript API** – for fetching video transcripts
- **PyPDF2** – for reading PDFs

Step-by-Step Workflow

1 Input

You can provide any one of the following:

- Blog URL (e.g., <https://realpython.com/python-basics/>)
- YouTube video ID (e.g., [dQw4w9WgXcQ](#))
- Local [.txt](#) or [.pdf](#) file inside the [inputs/](#) folder

2 Content Extraction

- If it's a **blog URL**, the script fetches the page and extracts readable text using **BeautifulSoup**.
- For **YouTube**, it fetches the transcript via **youtube-transcript-api**.
- For **text/PDF**, it reads the content directly from your local file.

3 Chunking

The text is divided into smaller **chunks** to handle token limits efficiently and ensure smoother summarization.

4 Summarization

Each chunk is summarized using a **Hugging Face model** (e.g., **facebook/bart-large-cnn** or **flan-t5-small**), and all partial summaries are merged into one clean, structured summary.

5 Repurposing

The merged summary is then repurposed into:

- A **LinkedIn post** (professional tone, paragraph style)
- A **Tweet thread** (short, concise points)
- A **Short video script** (dialogue-friendly narrative style)

6 Output

The generated outputs are automatically saved in the **outputs/** folder as a single text file named

Each file contains:

- ♦ **Summary**
- ♦ **LinkedIn Post**
- ♦ **Tweet Thread**
- ♦ **Short Video Script**

💡 Challenges Faced

- Handling large content required **text chunking** to stay within model token limits.
- Summarization quality varied based on input format – cleaned text improved results.
- Large models (like bart-large-cnn) can be slow on CPU, so lighter models (flan-t5-small) were used for better speed.

🔧 Possible Improvements

- Add a simple **web UI** for users to paste URLs or upload files.
- Integrate caching to skip re-processing the same source twice.
- Add options for **multi-language** summarization.
- Export each repurposed version into **separate files** (e.g., linkedin.txt, tweets.txt, script.txt).

🏁 Conclusion

This AI Agent automates the repurposing of digital content – from extraction to summarization and multi-format generation – showcasing how **LangChain + Hugging Face** can work together to simulate **real-world AI automation**.

It's lightweight, modular, and easy to extend for creators, marketers, or developers who want to **turn long-form content into ready-to-publish snippets** automatically.