# **AI Copilot Suite — Project Documentation**

**Aman Singh Ghusinga**

## **Project Overview**

The **AI Copilot Suite** is an end-to-end AI assistant developed to simplify and accelerate real-world business tasks. It provides users with a single point of interaction for two powerful capabilities: scam detection and data cleaning. These tools are designed to save time, reduce errors, and make advanced AI accessible to non-technical users.

At the heart of this system lies the **Agentic AI Copilot**, a command-line interface that serves as the intelligent coordinator of all tasks. Instead of requiring users to run separate scripts for each functionality, the Agentic Copilot interprets user instructions and seamlessly calls the right processes. This integration makes it possible for users to execute either data cleaning, scam detection, or both in a single, unified workflow.

## **Purpose & Business Value**

Businesses today deal with significant challenges around data and communication security. Teams spend countless hours cleaning messy datasets before they can be used for analysis or reporting. Meanwhile, the risk of phishing, scam texts, and fraudulent emails continues to grow, threatening both individuals and organizations.

The AI Copilot Suite addresses these challenges by leveraging advanced AI models to automate routine but critical tasks. By transforming natural language instructions into executable actions, the suite makes it possible for users of all skill levels to clean data and detect scams without writing a single line of code. This results in tangible business value, including significant time savings, improved data accuracy, and better protection against fraudulent activities.

## **Architecture Overview**

The suite is implemented as a set of modular Python programs, each with a specific role:

1. **scam\_detector.py**

This module focuses solely on analysing text messages to determine whether they are scams. It can process both individual messages and entire columns of data from a CSV file. Its results include a scam classification and an explanation describing why the message may be suspicious.

1. **data\_cleaning\_copilot.py**

This module is responsible for processing and transforming user-uploaded CSV files. Users can describe the cleaning steps they wish to perform in plain English, and the tool uses AI to generate cleaning suggestions. It then applies safe, predefined Python transformations to produce a clean dataset.

1. **agentic\_copilot.py**

This is the core orchestration layer and the main entry point for users. Instead of running multiple scripts separately, users launch this single program. It analyses the user’s input and decides whether to call the scam detection pipeline, the data cleaning pipeline, or both. This design creates a smooth user experience and keeps the entire system modular and maintainable.

## **How the System Works**

When a user runs the **Agentic AI Copilot**, they are greeted with a prompt asking how they would like to proceed. Users can type natural language instructions like:

“I want to clean my CSV file and check it for scams.”

or

“I just want to scan a single message for scams.”

The system interprets this text using a language model and determines the user’s intent. If data cleaning is requested, it asks the user to upload a CSV file. The system displays the columns and a preview of the data, prompting the user to describe any transformations they want to perform, such as converting date formats or dropping rows with missing values.

Once the data cleaning is complete, the system saves a new, cleaned CSV file. If scam detection is requested, either on its own or as a follow-up step after cleaning, the system processes messages either one at a time or in bulk from the CSV. Each message is classified as either SCAM or LEGITIMATE, with an explanation provided for transparency.

This pipeline ensures that even non-technical users can handle complex data workflows, from transforming raw datasets to identifying potentially fraudulent messages, all without writing code.

## **Scam Detection Copilot**

The **Scam Detection Copilot** is designed to protect users from fraudulent schemes that can arrive via email, SMS, or other communication channels. Users can analyse either single messages or batches of messages stored in CSV files. This functionality is crucial for businesses that regularly handle large volumes of communication and want to identify phishing attempts or other scams.

The tool uses a language model to analyse the content of messages and classify them as either LEGITIMATE or SCAM. It goes beyond simple keyword detection by evaluating the message’s context and structure, helping to detect sophisticated fraud attempts that might evade traditional rule-based filters. In addition to classification, the model provides an explanation describing why it considered a message suspicious.

This capability can be used independently or as part of a larger workflow, making it highly flexible for real-world applications.

**How It Works:**

* Accepts a text message as input
* Sends the message to GPT-3.5 Turbo with a specialized prompt describing various known scam types
* Receives back:
  + Whether the message is a **SCAM** or **LEGITIMATE**
  + An explanation if it’s a scam

**Key Features:**

* Detects over 15 different scam types, such as:
  + Prize scams
  + Government or IRS scams
  + Bank verification scams
  + Crypto scams
* Can operate on:
  + A single text input
  + An entire CSV file’s column of messages
* Outputs:
  + Scam label
  + Explanation of reasoning
* Helps protect businesses from fraud and builds trust with customers

This file is entirely focused on the task of **text classification for scam detection.**

**Example**

A screen shot of a computer

AI-generated content may be incorrect.

This shows how the copilot not only labels the message but also educates users on why a message is suspicious.

A screenshot of a computer

AI-generated content may be incorrect.

**Business Value**

Organizations can integrate this tool into their operations to proactively monitor incoming communications. This protects customers, reduces financial losses, and helps teams respond quickly to potential fraud, ultimately increasing trust in digital communication channels.

## **Data Cleaning Copilot**

The **Data Cleaning Copilot** enables users to transform messy, inconsistent datasets into clean, analysis-ready data. Rather than requiring users to manually code transformations, this tool allows them to describe what they want in natural language.

For example, a user might say:

“Convert date columns to YYYY-MM-DD format and drop rows where revenue is missing.”

The system uses a language model to generate cleaning suggestions, helping users think of additional improvements they might not have considered. It then executes Python-based cleaning routines, addressing issues like mixed date formats, null values, and inconsistent text.

The cleaned data is saved as a new CSV file, ensuring that the original data remains intact for traceability and auditing.

**How It Works:**

* Asks the user for:
  + CSV file path
  + Natural-language instructions describing the desired cleaning steps
* Displays:
  + Columns in the CSV
  + Sample rows for user review
* Sends the user’s instructions and data preview to GPT-3.5 Turbo
* Receives:
  + Suggestions for additional cleaning tasks
* Executes cleaning operations in Python, such as:
  + Converting date formats to YYYY-MM-DD
  + Dropping rows with missing values
  + Stripping whitespace and cleaning text
* Outputs a new cleaned CSV file (cleaned\_data\_v2.csv)

**Key Features:**

* Handles complex date parsing, including:
  + Mixed date formats (e.g. “12/04/1985”, “July 2, 1983”)
* Keeps the original data safe by writing cleaned data to a new file
* Allows business users to clean data without writing code

**Example**

* Original Data:

A screenshot of a computer

AI-generated content may be incorrect.

* User Instruction:

“Convert all dates to YYYY-MM-DD and drop rows with missing revenue.”

* Cleaned Output:

A screenshot of a computer

AI-generated content may be incorrect.

**Business Value**

By reducing the need for manual intervention, this tool saves significant time and minimizes errors in data preparation. It empowers business analysts and non-technical staff to clean data quickly, enabling faster decision-making and more accurate analysis.

## **Agentic AI Copilot**

The **Agentic AI Copilot** is the intelligent brain of the system, capable of interpreting user instructions and orchestrating both scam detection and data cleaning workflows.

When launched, it engages users in a conversation to determine what they want to achieve. Based on the user’s input, it makes decisions about which processes to run and in which order. For example, if a user wants to clean a CSV file and then check the messages for scams, the copilot ensures the cleaning pipeline runs first, followed seamlessly by scam detection.

While the system currently relies on user instructions for each step, the long-term vision is to make it fully agentic. In the future, the copilot could proactively analyse uploaded data, detect issues like mixed date formats or missing values, and suggest solutions without being prompted.

**How It Works:**

* Launches as a single script (python agentic\_copilot.py)
* Prompts the user:

**“How can I help you?”**

* **Uses GPT-3.5 Turbo to interpret the user’s instructions and decide:**
  + Should we run data cleaning?
  + Should we run scam detection?
  + Should we run both?
* **Dynamically calls:**
  + data\_cleaning\_copilot.py for data cleaning
  + scam\_detector.py for scam detection
* **Supports:**
  + Single-message scam checks
  + Bulk scam checks in CSVs
  + Chaining cleaning → scam detection for integrated workflows
* **Outputs results into one or more CSV files**

**Key Features:**

* Allows the user to:
  + Clean data only
  + Detect scams only
  + Run both processes in sequence
* Makes the entire suite accessible through one conversational script
* Future-ready for more proactive, agentic features

This file turns your two separate modules into one seamless pipeline where the user simply talks to the system.

**Example Workflow**

A screen shot of a computer

AI-generated content may be incorrect. A screen shot of a computer

AI-generated content may be incorrect.

*Figure: Agentic Copilot conversational interface*

Consider this scenario:

A company receives thousands of customer messages stored in a CSV file. The data includes:

* Dates in different formats
* Missing values in some rows
* Potentially fraudulent messages mixed in

With the AI Copilot Suite, the process looks like this:

1. The user runs:

python agentic\_copilot.py

1. The copilot asks:

“How can I help you?”

1. The user replies:

“I want to clean my CSV and check it for scams.”

1. The copilot:
   * Loads the CSV
   * Displays sample data
   * Accepts instructions like:
     + “Convert all dates to YYYY-MM-DD”
     + “Drop rows where revenue is missing”
   * Cleans the CSV and saves a new file
   * Offers to run scam detection on the cleaned file
   * Scans each message and adds:
     + Scam label (SCAM or LEGITIMATE)
     + Explanation for why a message was flagged
2. The system saves the results in a CSV like:

| **id** | **message** | **date\_sent** | **scam\_label** | **explanation** |
| --- | --- | --- | --- | --- |
| 1 | Your account locked. Click here. | 2023-10-01 | SCAM | This resembles a bank verification scam. |
| 2 | Thank you for your purchase. | 2023-10-01 | LEGITIMATE | Normal transaction message. |

This integrated workflow saves time, reduces errors, and helps keep businesses secure.

## **Current Capabilities**

Today, the AI Copilot Suite can:

* Accept plain-language instructions for cleaning data
* Detect scams across many message types
* Handle:
  + Mixed date formats
  + Missing data
  + Bulk or single-message scanning
* Save cleaned and labelled data into new CSV files
* Run either single tasks or fully chained workflows through one script

## **Future Vision**

While the current system successfully automates many tasks, the long-term goal is to make it truly agentic. Future improvements might include:

* Automatic scanning of CSVs to detect:
  + Percentage of missing values
  + Mixed data types
  + Duplicate rows
* Proactive suggestions for cleaning or analysis
* Automatic generation of dashboards and reports
* Fully automated end-to-end workflows without constant user prompts

These features would transform the Copilot from a responsive tool into an intelligent AI collaborator capable of managing complex tasks independently.

## **Technical Stack**

**This project uses:**

* Python
* Pandas
* OpenAI GPT-3.5 Turbo
* dotenv for environment and API key management

## **Skills Demonstrated**

**The AI Copilot Suite demonstrates:**

* Prompt engineering
* Modular Python architecture
* Safe integration of AI-generated code
* End-to-end workflow design
* Transforming natural-language instructions into business actions
* Solving real-world problems with AI

**Summary**The AI Copilot Suite successfully turns advanced AI models into practical, user-friendly tools. It enables businesses to clean their data and detect scams—all through one script—showcasing how AI can drive real efficiency and security in business operations.