# Allergen Extractor Exercise

## Goal

Build a modular, clean-code system that takes a URL of an online cookbook, scrapes its recipes, and outputs:

* The original text content
* A deduplicated list of allergenic ingredients (JSON or plain text)

## Core Requirements

Input: URL of a webpage containing cookbook/recipes:

Example: https://www.allrecipes.com/recipe/46822/indian-chicken-curry-ii/

Process:  
• Fetch and scrape recipe text from the provided URL  
• Parse ingredients from the scraped text  
• Call Azure OpenAI to analyze the text and return a list of allergenic ingredients

## Architecture Guidelines

- AIClient is mandatory for allergen extraction  
- Use interfaces/abstract classes  
- Load config (keys, endpoints, user-agent, etc.) from env or config file  
- Handle HTTP errors, invalid URLs, and log at appropriate levels

## Azure OpenAI Placeholder

endpoint = "https://muvan-o1-mini.openai.azure.com/"  
api\_version = "2024-12-01-preview"  
deployment\_name = "gpt-4.1-2"  
model\_name = "gpt-4.1"  
subscription\_key = “50bU6AQnSEuu3mA4kEYLb9neDods1UmAzK3zwYPCp8SR7RNTQlbyJQQJ99ALACHYHv6XJ3w3AAABACOGnNKS”

## Deliverables

* Well-structured source repository