

AI-Assisted Opportunity Review

Prompt Engineering, Evaluation, and Workflow Design

Sanitized work sample demonstrating AI workflow design, prompt engineering, and evaluation.

Example AI-Assisted Review Workflow



User Defines Parameters in Internal Knowledge Base

Keywords, prompts, topics, are configured in Excel/txt files and stored in Internal Knowledge Base.

Inputs are loaded automatically into the Python script.



Automated Data Collection Script Pulls Opportunities

Custom Python code queries External Database API using configured keywords.

Results are filtered for status, dates, and other relevant variables.



Custom Generative AI Prompts Score and Summarize Opportunities

Azure Open AI or Claude is used with persona-based prompt templates to score and summarize each opportunity from 0-100 based on relevance to system modernization work for Internal Objectives



Results are Saved, Reviewed, and Evaluated

Outputs are saved to Excel files with results and metadata

Results are reviewed by humans for relevance and stored for historical tracking.

Objective

Help improve how our AI-assisted system finds and scores relevant opportunities by redefining the keywords inputs and prompt templates.

- What it searches for (priority-specific keywords)
- How it scores relevance (criteria keywords)
- How it's prompted to think and assess findings (prompt templates)

Prompt Engineering Workflow

01

Define Search Criteria

Create a list of search terms for catching all possible opportunities related to systems standardization

02

Define Evaluation Criteria

Develop keywords to assess the relevance of retrieved opportunities

List words or phrases that make a scraped opportunity relevant to the Objectives

03

Edit Prompt Templates for AI Instruction

Tweak or write prompts that guide the AI to summarize and score opportunity listings for relevance

04

Compare to Baseline and Evaluate

How does the change in input affect relevance of scraped results?

Evaluation

We will run a fixed dataset(the same scraped opportunities) through each engineer's config and compare:

- Effectiveness: Which sets of inputs provide the most relevant results compared to our current system?
- Prompt Quality: Do prompts appropriately guide the model while accounting for AI capabilities and limitations?
- Documentation & Reproducibility: Are prompt changes clearly documented so results can be interpreted and reproduced?

Glossary

- Relevance Scoring Keywords: Terms that signal whether a grant is aligned with our goals (i.e. Standardize and improve internal systems)
- Prompt Template: A written instruction that guides how the AI processes and scores an opportunity
- Relevance Score: A number between 1 –100 to rate how relevant the opportunity is to Organizational Priorities
- Prompt Engineering: The practice of writing and refining instructions for generative AI (such as ChatGPT, Copilot, Claude)

Outcomes

This work resulted in:

- Standardized prompt engineering guidance
- Structured evaluation criteria for AI outputs
- Training materials for new users
- A repeatable workflow supporting consistent human review