

How to Evaluate AI Prompts

Overview:

This document outlines a structured approach to evaluating AI-generated text prompts, especially when working with large language models (LLMs) such as ChatGPT.

Step 1: Define the Purpose

- Identify the goal of the prompt (e.g., informative, persuasive, narrative).
- Determine the desired tone, complexity, and ethical constraints.

Step 2: Establish Evaluation Criteria

Use the following core rubric dimensions:

- Accuracy: Is the output factually correct?
- Clarity: Is the response easy to read and understand?
- Completeness: Are all parts of the prompt addressed?
- Ethical Safety: Does the output avoid bias, hate speech, or misinformation?

Step 3: Assign Scores

- Use a 1-5 scale (or percentage) to rate each criterion. Provide short notes to explain ratings.

Step 4: Log Your Results

- Document scores using a CSV or Markdown table for analysis.
- Include the original prompt, model response, and reviewer comments.

Step 5: Iterate & Improve

- Refine prompts based on low-scoring criteria.

- Repeat evaluation for consistent scoring and quality control.

Tools & Tips:

- Use Python and the OpenAI API for automating evaluations.
- Store all evaluations in a centralized folder with standardized file names.
- Share and version-control documents using GitHub.

Created by Gerald Brown - July 2025