Content Generation and Summarization with Generative AI

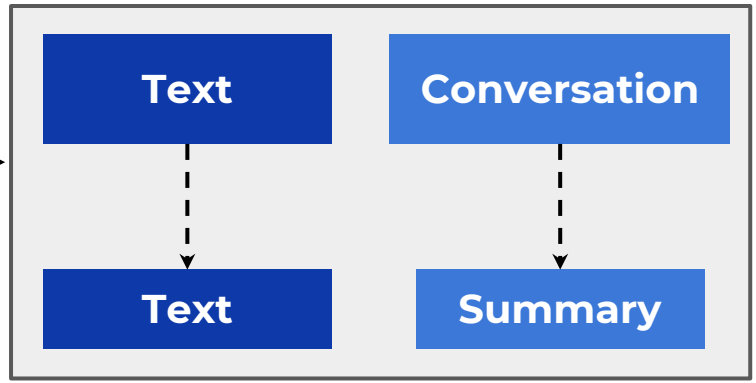
Agenda

In this session, we will :

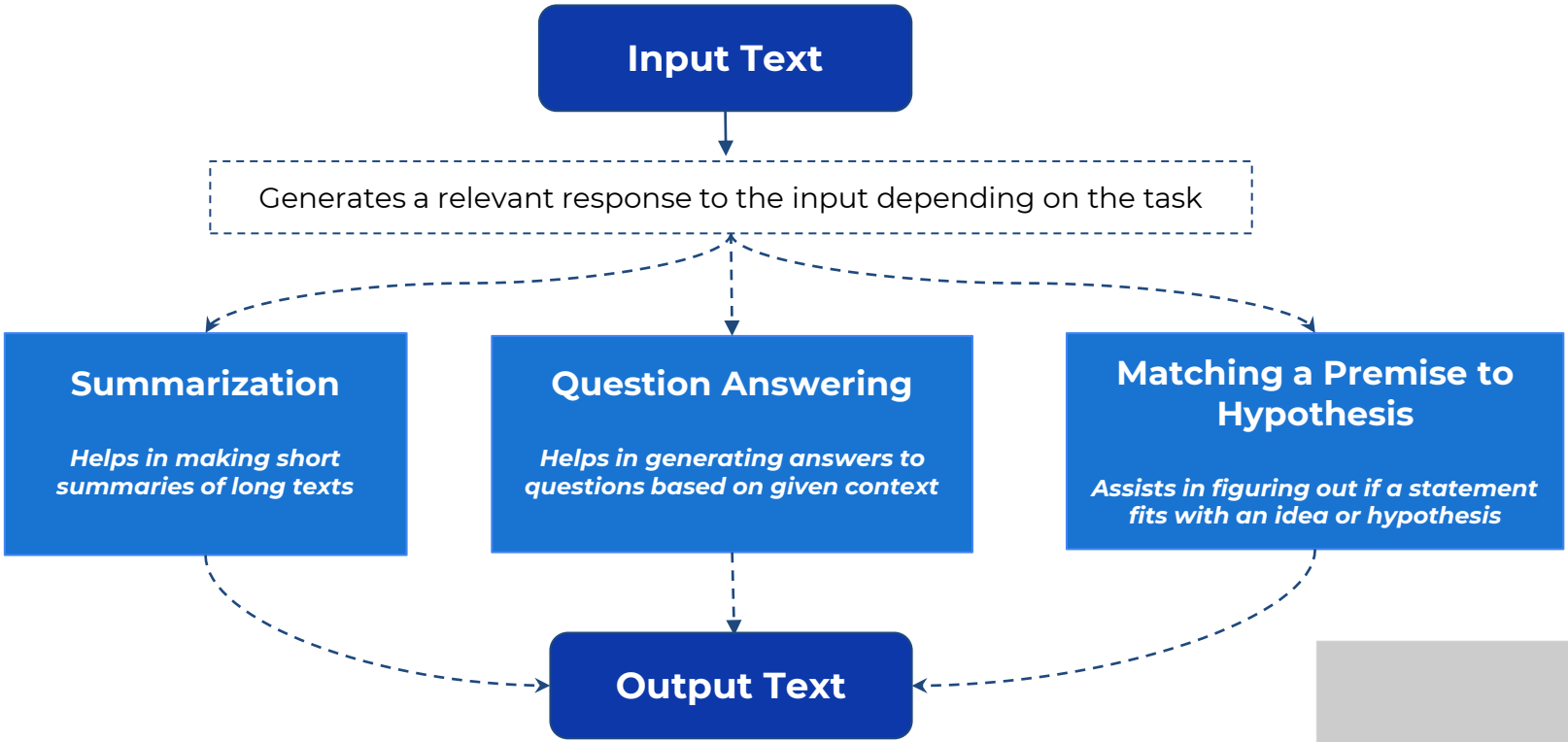
- Overview of Text-to-Text Generation
- Structure of Text Generation Tasks (Summarization) in Generative AI
- Data Preparation with respect to Text Generation Task
- Prompt Structure with respect to Text Generation Task
- Evaluation with respect to Bert and Rouge Score in Text Generation Task

Text-to-Text Generation

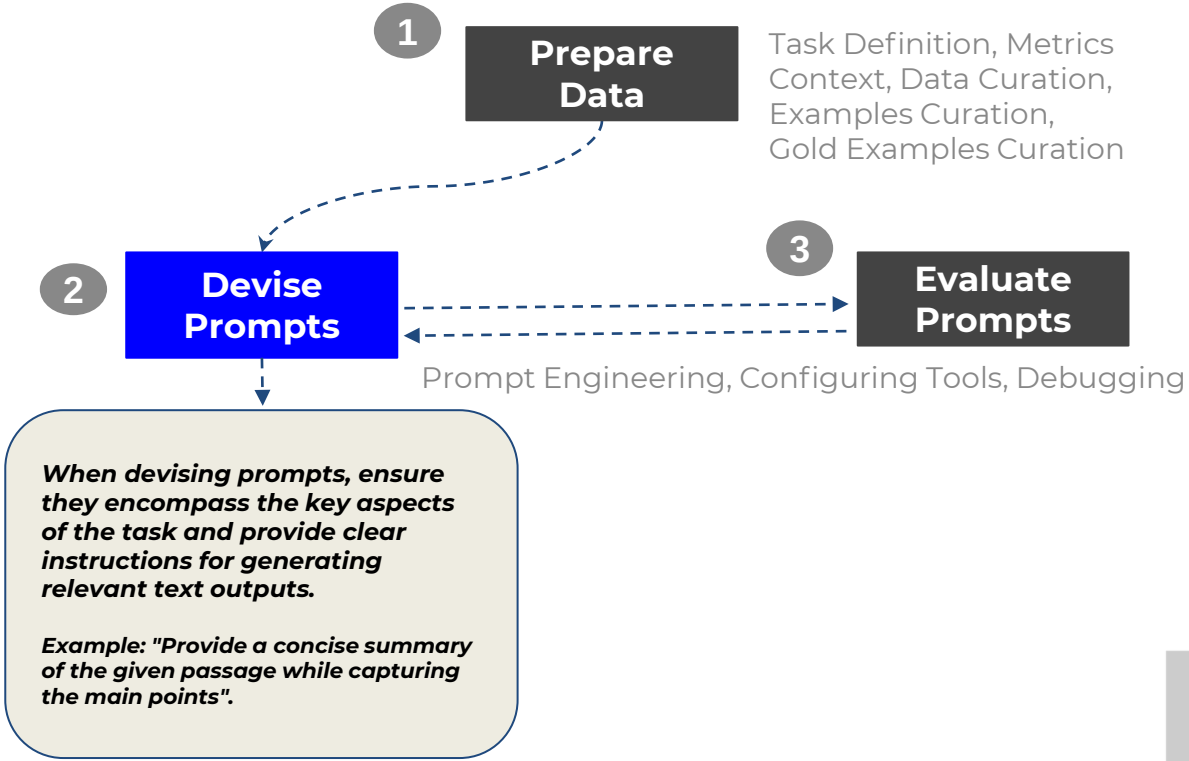
Text-to-Text involves the generation of a relevant response to the input depending on the task



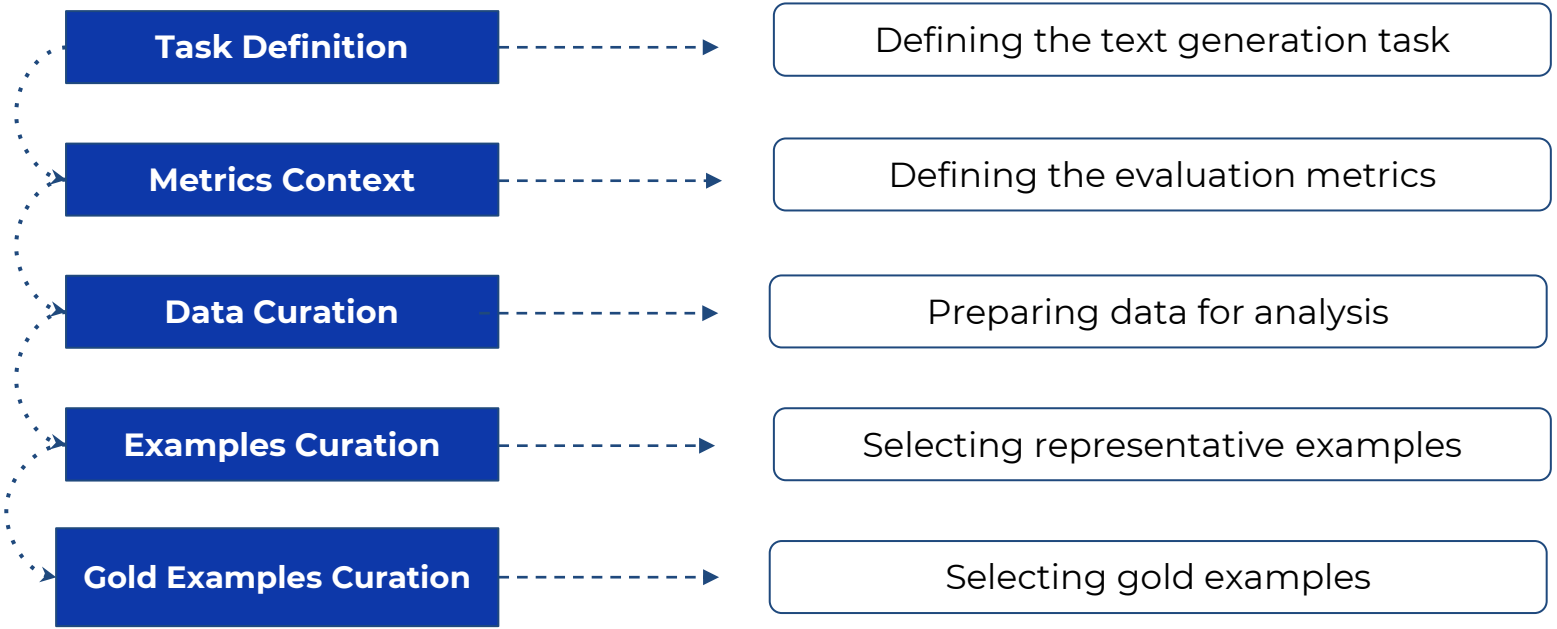
Text-to-Text Generation



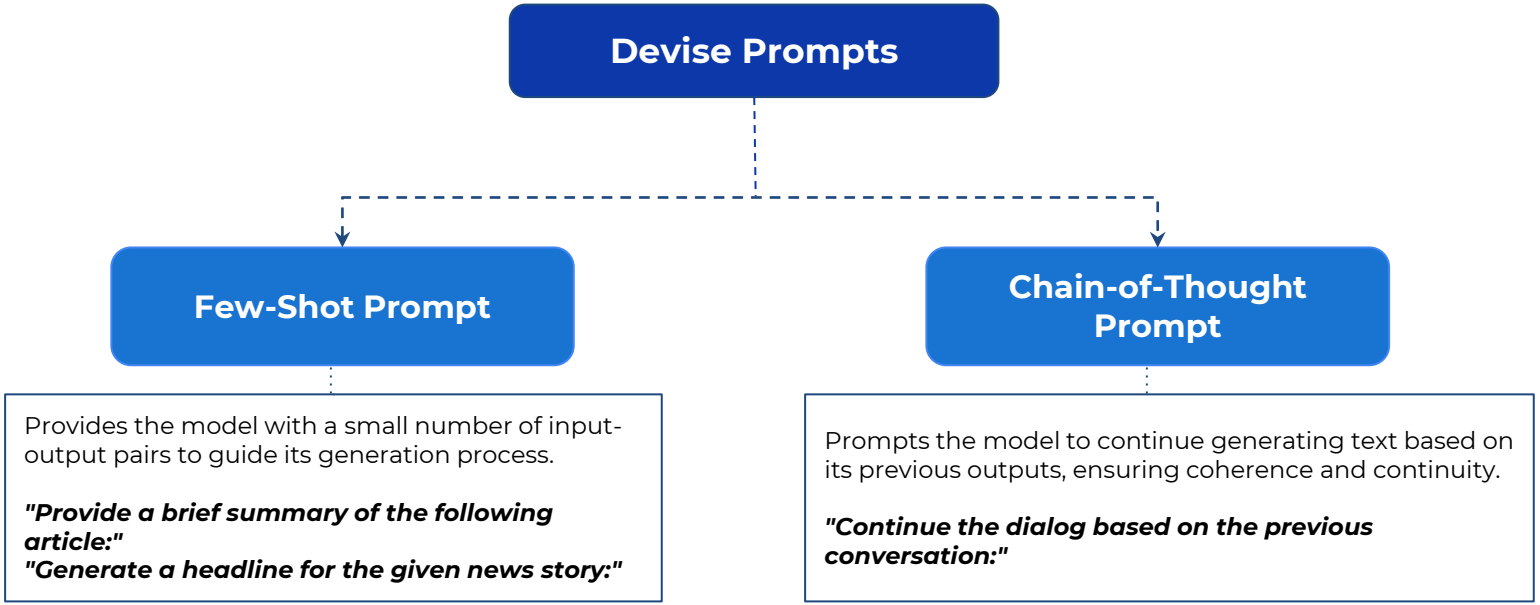
Structure of Text Generation Tasks



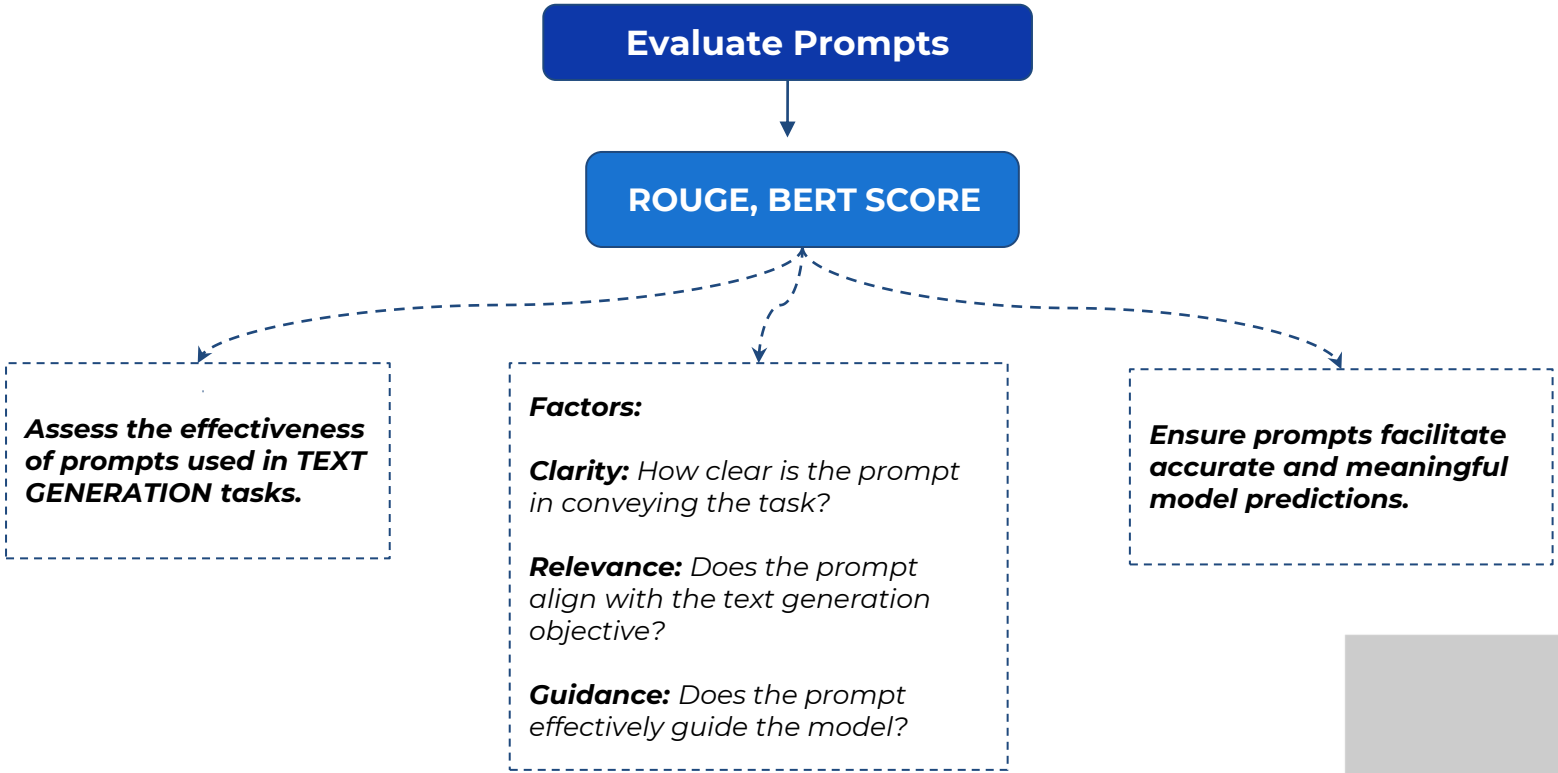
Preparing Data in Text Generation Task



Devise Prompts in Text Generation Task

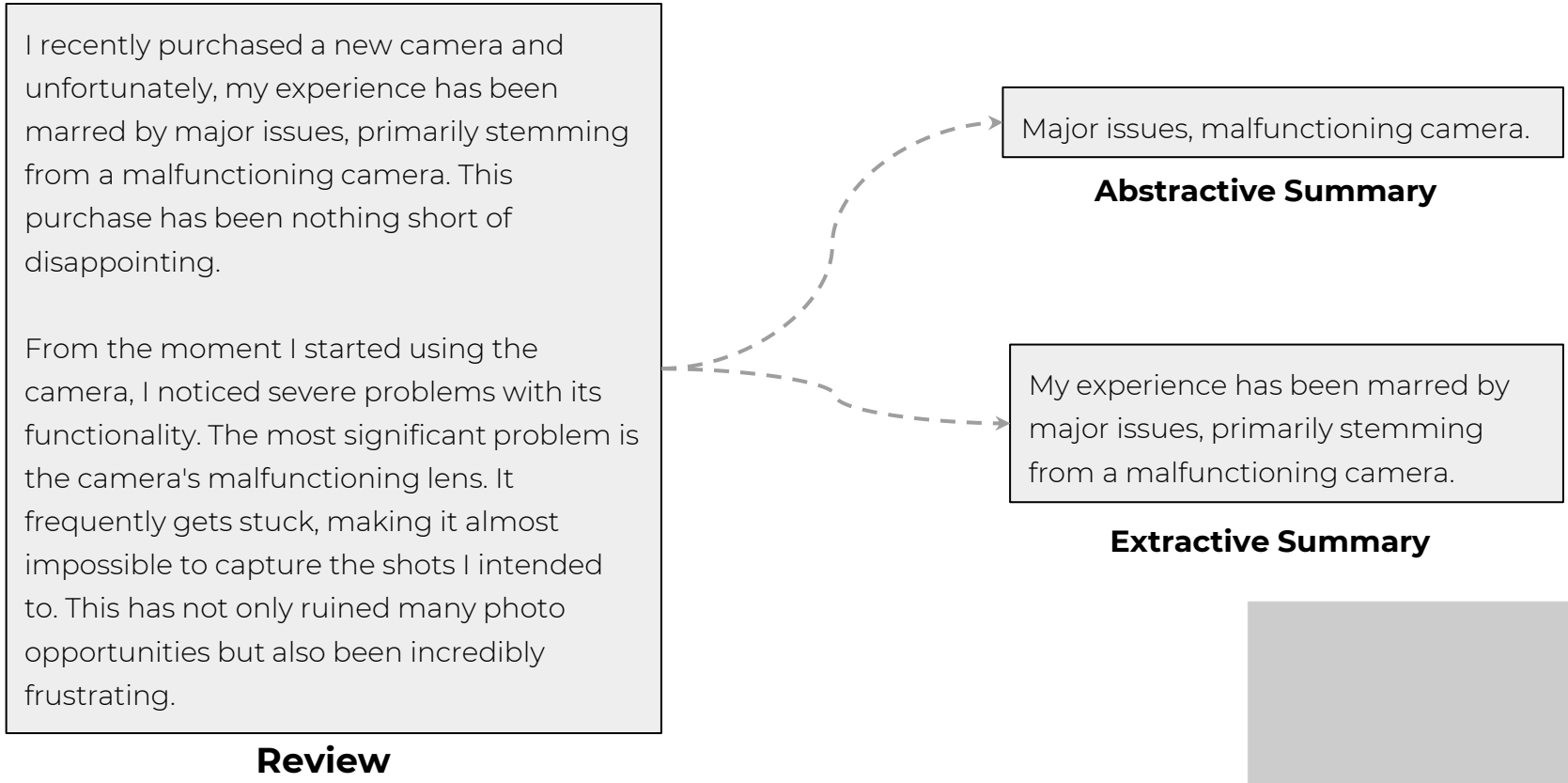


Evaluate Prompts in Text Generation Task



Step 1: Defining Objectives & Metrics in Summarization

Defining Objectives in Summarization



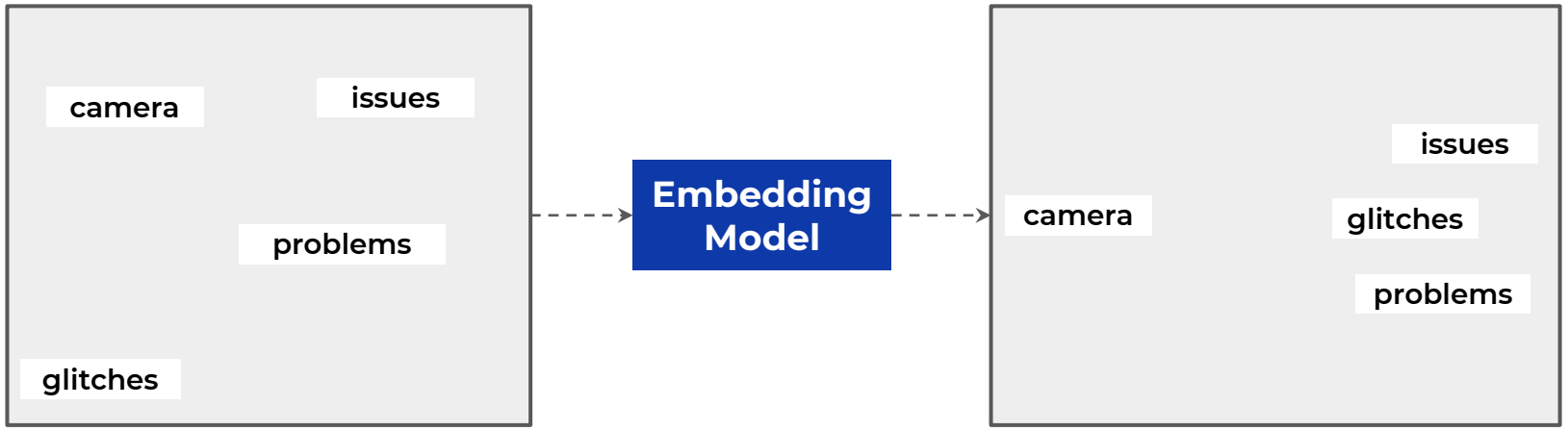
Defining Metrics: ROUGE: n-grams

AI-Generated Summary
Colossal disappointment with constant glitches, defects, and a malfunctioning camera.

Human-Generated Summary
Incredibly let down by the never-ending issues – glitches, defects, and a camera that just won't cooperate.

	Length of Matches
Unigram (1-gram) Matches glitches, defects, and, a, camera.	5
Bigram (2-gram) Matches (glitches, defects), (defects, and), (and, a).	3
Trigram (3-gram) Matches (glitches, defects, and), (defects, and, a).	2
4-gram Matches (glitches, defects, and, a).	1
Longest Common Subsequence (glitches, defects, and a), (camera).	5

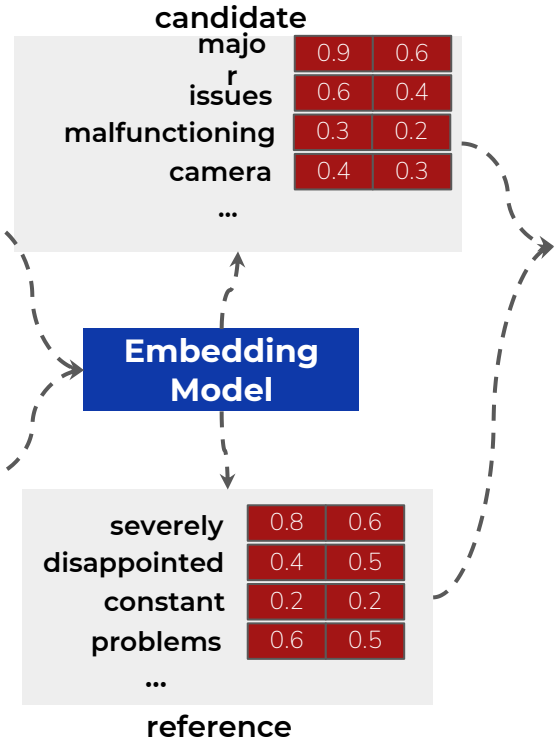
Embeddings (Used in Bert Score)

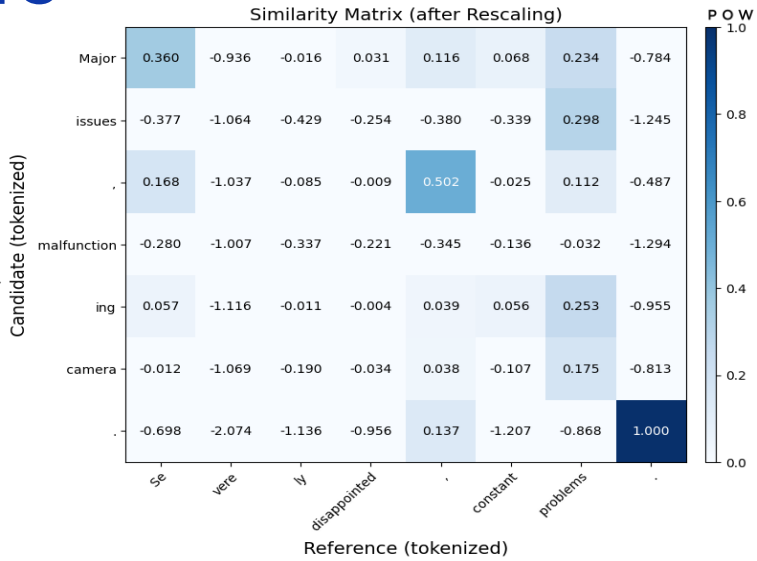


Defining Metrics: BERT Score

AI-Generated Summary
Major issues, malfunctioning camera

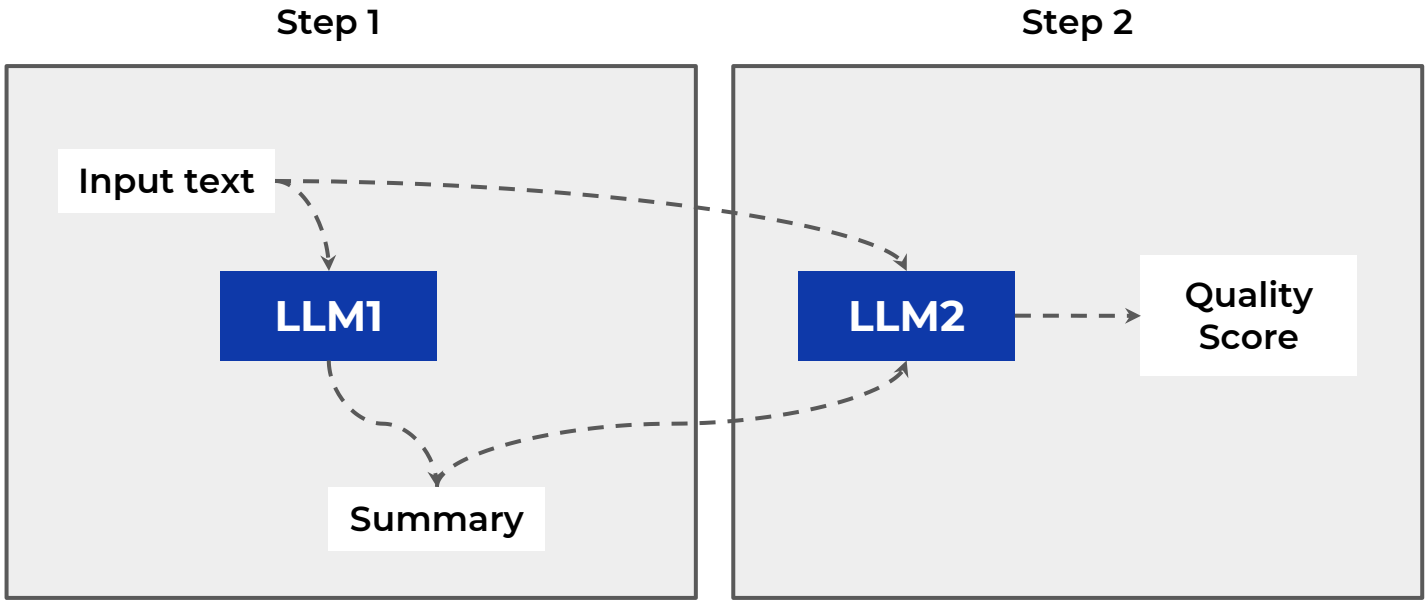
Human-Generated Summary
Severely disappointed, constant problems





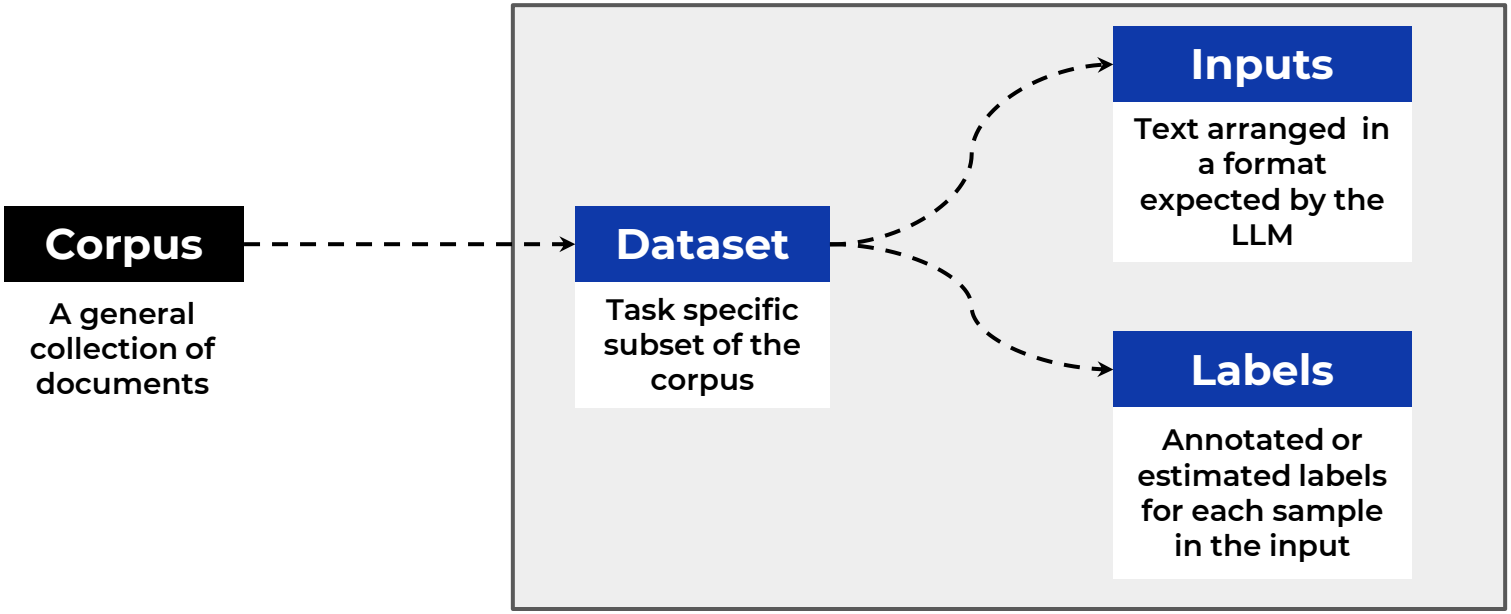
P = (0.360 + 0.298 + 0.502 + ...)/7 = 0.36
R = (0.360 - 0.936 -0.011 + ...)/8 = 0.17
F1 = (2 * P * R)/(P + R) = 0.23

Defining Metrics: LLM Rating

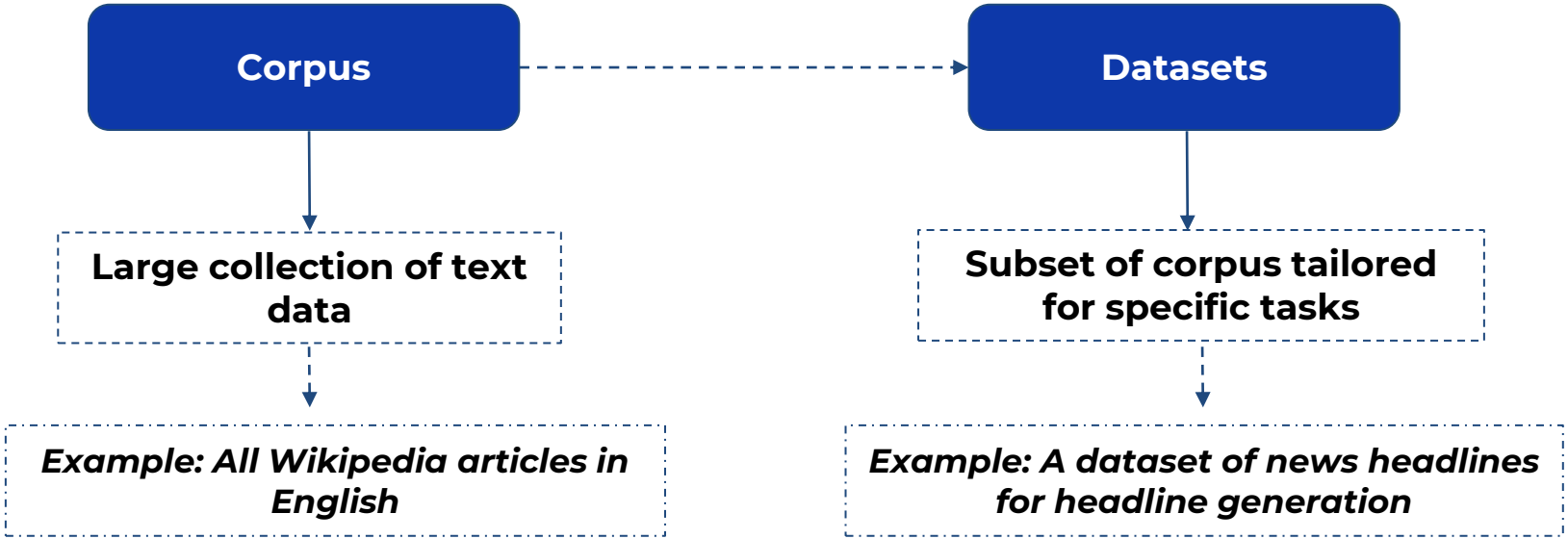


Step 2 : Assemble Data

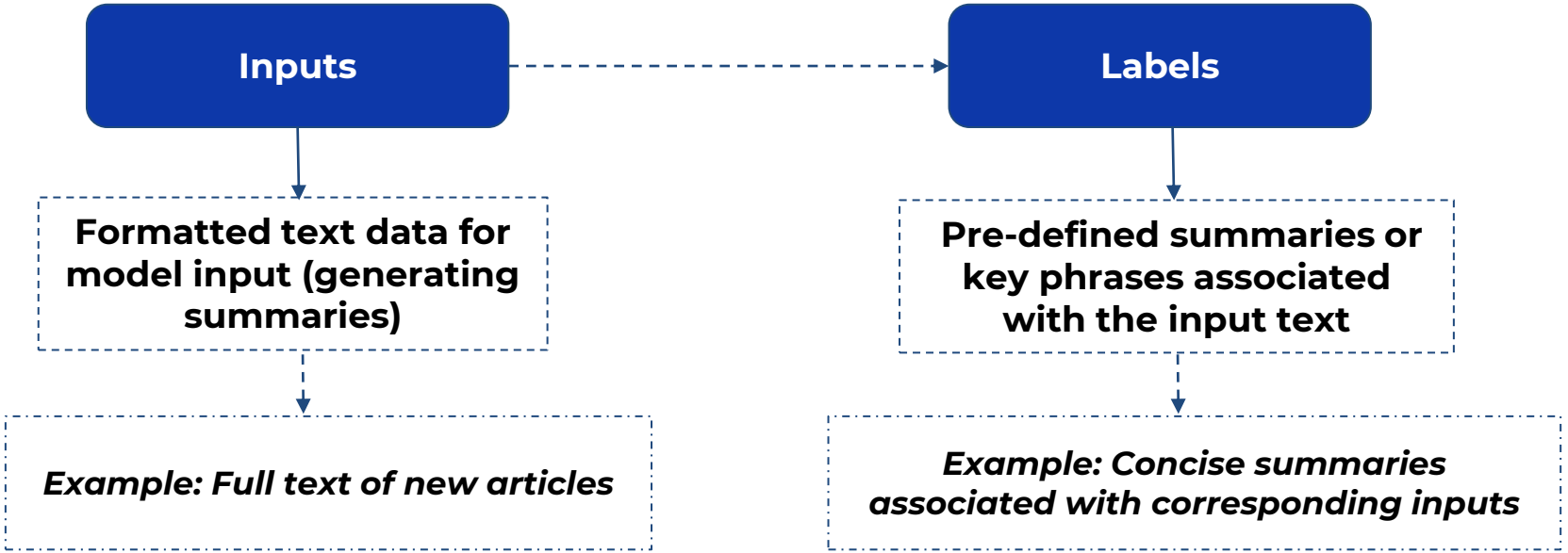
Data Preparation Process



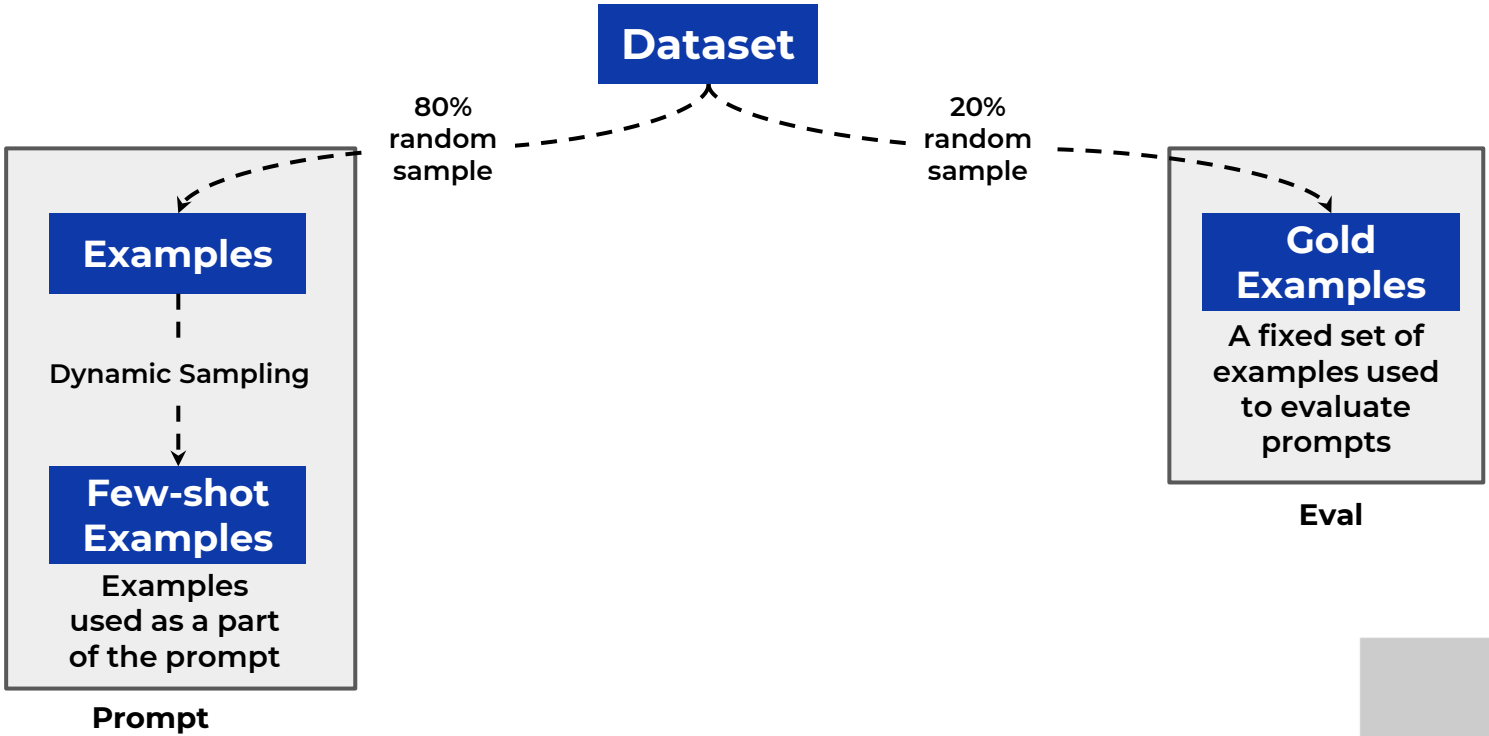
What are Corpus and Datasets?



What are Inputs and Labels?

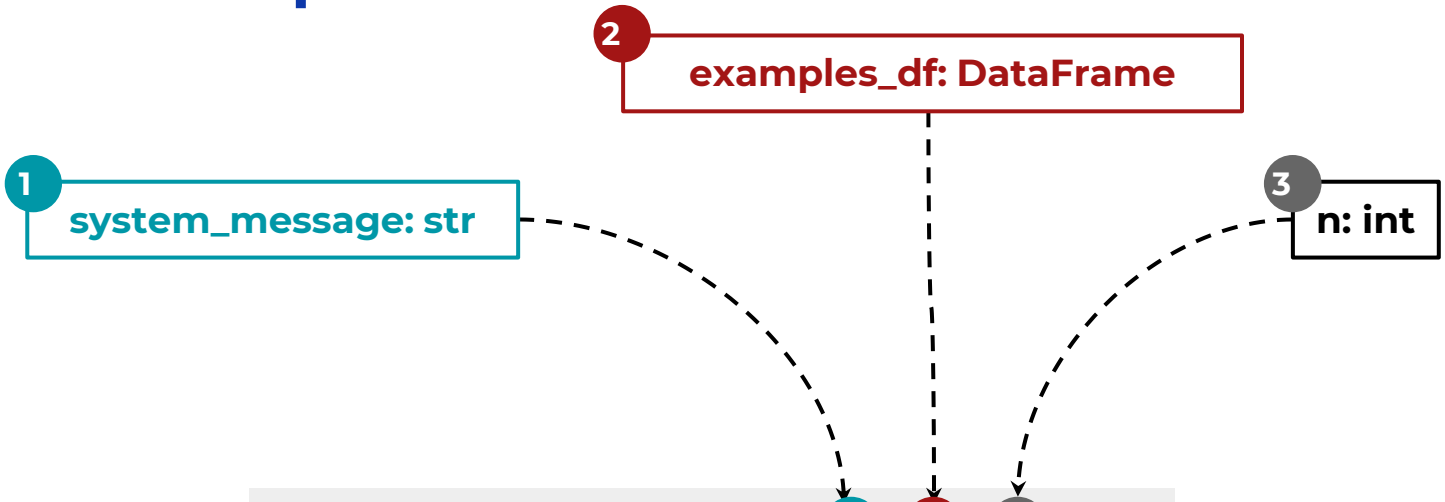


Assembling Data in Text Generation Task



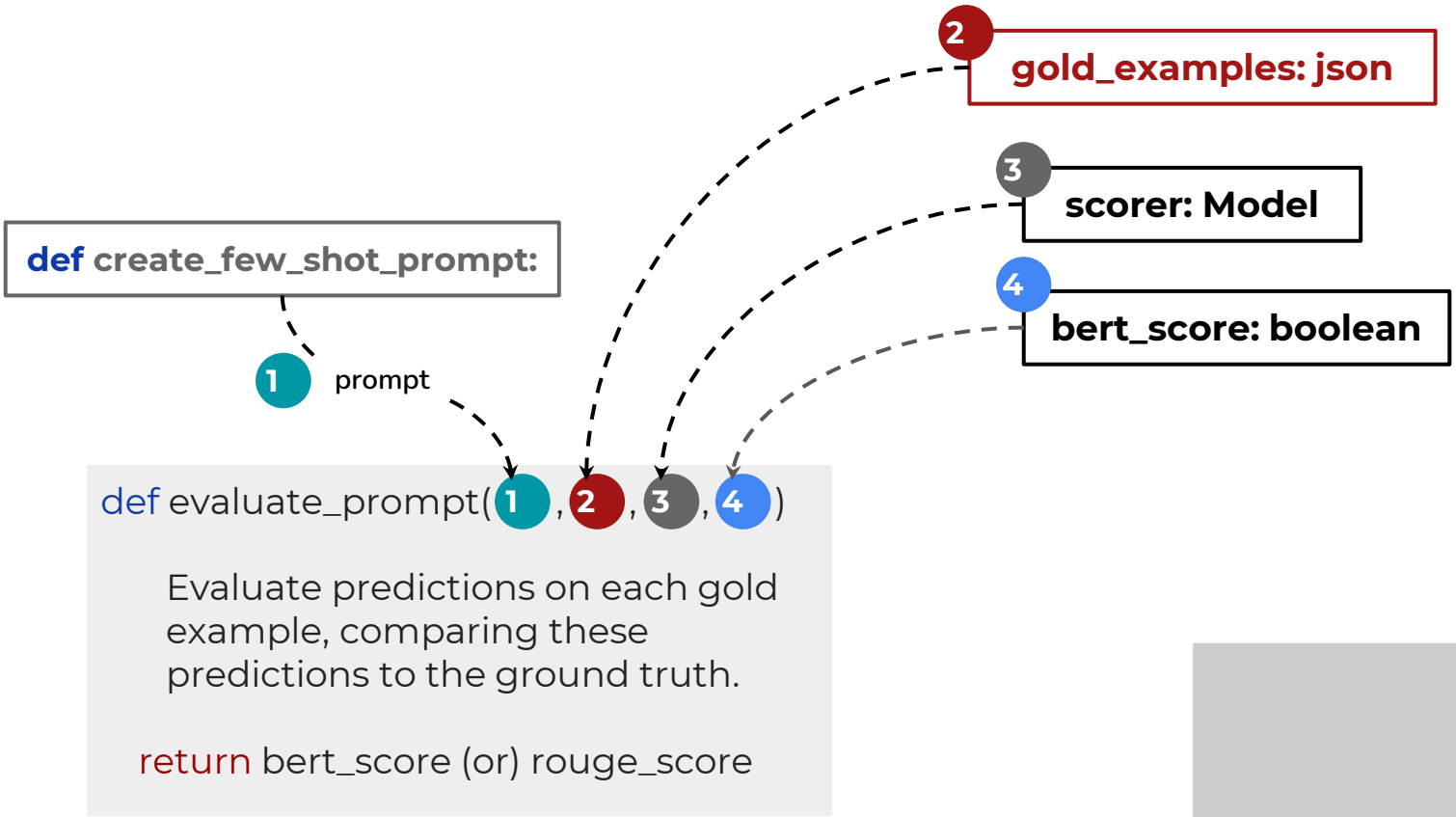
Step 3: Derive Prompts

Derive Prompts



Step 4: Evaluation of Prompts

Understanding the Evaluation Process



Summary of Text-to-Text Tasks