

Skeleton of Thoughts Prompting

Instructor

Bhaskarjit Sarmah

Vice President, Blackrock



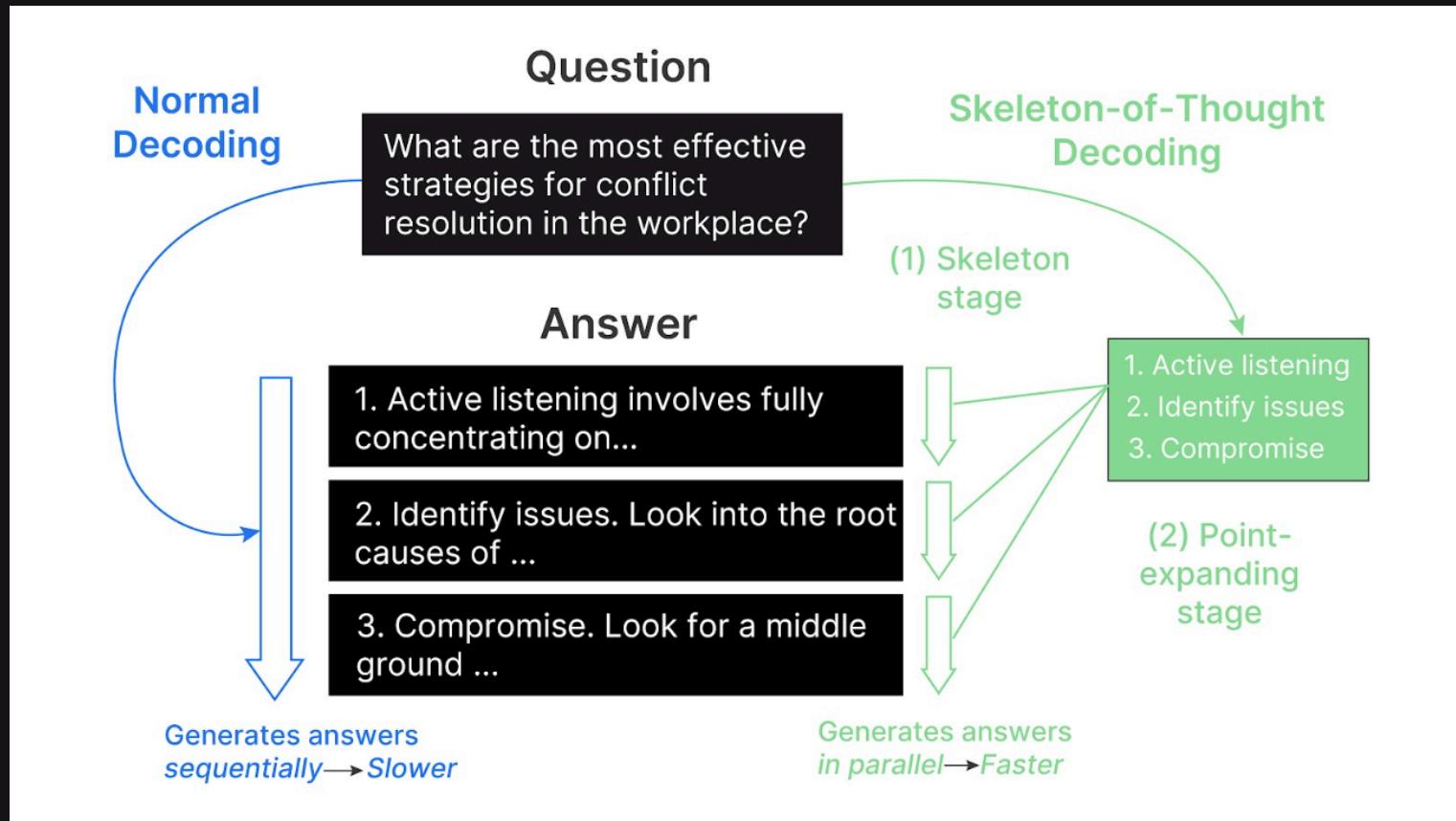
Definition

- The **Skeleton-of-Thought** (SoT) is a prompting method that outlines answers first and then fills in details in parallel, speeding up the response.

How Does it Work

- **Two-Phase Efficiency:** Begins with structuring an answer skeleton, then expands points in tandem for rapid completion.
- **Parallel Processing:** Enhances speed by detailing all parts of the outline at once, streamlining the thinking process.

How Does it Work



Skeleton-of-Thoughts: Example

Prompt 1. Skeleton prompt Template

[User:] You're an organizer responsible for only giving the skeleton (not the full content) for answering the question. Provide the skeleton in a list of points (numbered 1., 2., 3., , etc.) to answer the question. Instead of writing a full sentence, each skeleton point should be very short with only 3 to 5 words. Generally, the skeleton should have 3 to 10 points. Now, please provide for the following question.

{ Question }

Skeleton:

[Assistant:] 1

Skeleton-of-Thoughts: Example

Prompt 2. Point-Expanding prompt Template

[User:] You're responsible for continuing the writing of one and only one points in the overall answer to the following question.

{ Question }

The skeleton of the answer is

{ skeleton}

Continue and only continue the writing of point {Point index}. Write it **very shortly** in 1 to 2 sentences and do not continue with other points!

[Assistant:] {Point index}. {Point skeleton}

Cons: Skeleton of Thoughts Prompting

Might not understand how to link answer parts, leading to less coherent responses.

With pay-per-word services, the outlining process could increase costs.

Speed benefits vary with the prompt and model, which could affect practicality.

Thank You
