

Step 1: View Construction



Input

Anchor a

Candidates c_1 c_2



Preprocessing

clean/noremalize/sentence split

Theme
raw text

Action
verb sequence

Outcome
ending sentence



Output: 3 view texts (for a, c_1, c_2)

Step 2: Similarity + Routing



E5-Mistral-7B-instruct
Encoder



Fusion

$$v = 0.5v^{\text{theme}} + 0.25v^{\text{action}} + 0.25v^{\text{out}}$$



Similarity & Confidence

$$S(a, c_1) \quad S(a, c_2)$$

$$\Delta = |S(a, c_1) - S(a, c_2)|$$

$$\Delta \geq \tau?$$



Step 3: Inference + Decision

Prompt Construction
anchor + two candidates
view snippets
constrained A/B output

Gemini-3-Flash
Sampling + Voting



sample 3 times
majority voting

Final Output
Answer = A or B



Argmax $S(a, c)$