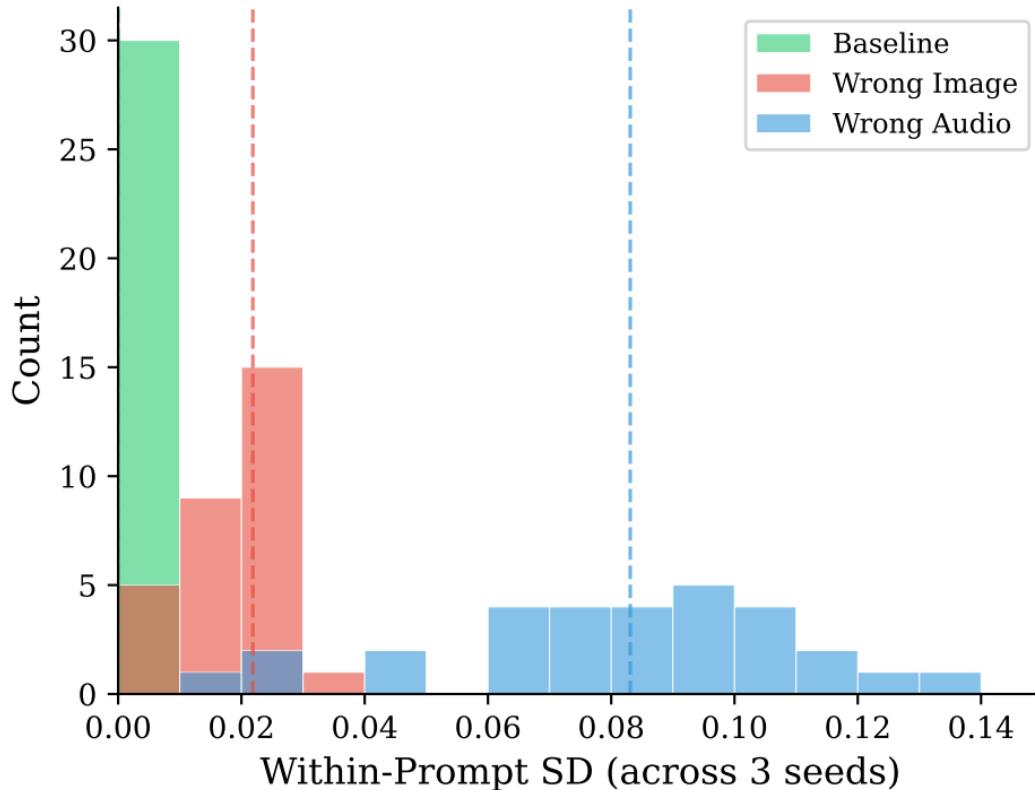


Reproducibility: Within-Prompt Variance Across Random Seeds

RQ1 Full Pipeline



RQ2 Planning Modes

