to be effectively applied for improving text generation. Such approaches have demonstrated strong performance in tasks such as automatic summarization [4] and data-to-text generation [6]. To be useful, text variants must be generated carefully, and previous work has relied on mixture-of-logits decoding. Such techniques are directly relevant to conversational search, where clarifying questions must remain faithful to the corpus.

Faithfulness Evaluation. Faithfulness measures whether generated text remains consistent with its input. In summarization, state-of-the-art approaches employ entailment-based metrics that leverage NLI models to score the consistency of summaries with source documents (e.g., RoBERTa-based entailment [16]). These methods provide fine-grained judgments of factual alignment on a continuous scale. In data-to-text generation, metrics such as PARENT [5] evaluate whether candidate outputs faithfully express entities and relations from structured inputs. By contrast, clarifying question generation has not been systematically assessed for faithfulness. Existing evaluations rely mainly on reference-based metrics (e.g., BLEU, METEOR) or indirect retrieval-based proxies [2,24], which do not directly measure factual consistency with the input context. In this work, we adapt entailment-based and data-grounding approaches from summarization and data-to-text to develop faithfulness evaluations tailored to clarifying question generation.

3 Methodology

Our RAC framework follows a two-stage training pipeline as illustrated in Fig. 2. In the first stage, a large language model p_{LM} is fine-tuned on existing clarification datasets along two axes to generate: (1) factual questions conditioned by user queries and retrieved passages p_{θ_0} and (2) less factual questions unconditioned by passages p_{uncond}. The two levels of question quality are then passed to a preference learning algorithm (contrastive) that encourages the model to rank faithful, evidence-grounded clarifications higher than unsupported or hallucinated alternatives.

We formulate the task of generating clarifying questions Cq as a retrieval-augmented generation task. The initial user query U_q enables the retrieval of a set of relevant passages $\mathcal{D} = \{d_1, \ldots, d_N\}$, which will be used as context for the generation. We assume all queries to be ambiguous, focusing on clarifying

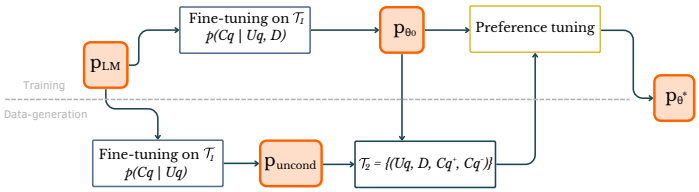

Fig. 2: Overview of our proposed training pipeline.

question generation rather than clarification need prediction [17]. Each passage may capture different semantic facets of the query, but we restrict to a single-turn setup, generating one clarifying question targeting the most useful facet.

3.1 Supervised Clarifying Question Generation

Retrieval-augmented generation (RAG) has shown that conditioning large language models on retrieved passages improves factual grounding and reduces reliance on parametric memory [9,12]. However, previous work has focused on generating direct zero-shot answers. Our contribution is to propose a fine-tuned model (twice) to better exploit the retrieved passages for the clarification task. To this end, we employ supervised fine-tuning (SFT) as the first stage of training: a large language model is trained to generate clarifying questions C_q conditioned on both the user query U_q and the corresponding retrieved passages $\mathcal{D}$ (leading to p_{θ_0}). Given a dataset $\mathcal{T}_1$ of query–passage–ground-truth-question tuples $(U_q, \mathcal{D}, C_q^+)$, the model is optimized with the negative log-likelihood objective:

$$\mathcal{L}_{\mathrm{SFT}}(\theta) = -\mathbb{E}_{\sim\mathcal{T}_1}\left[\sum_{t=1}^{|C_q^+|} \log p_\theta(C_{q,t} \,|\, U_q, \mathcal{D}, C_{q,<t})\right] \tag{1}$$

Here, each token of the clarifying question $C_{q,t}$ is predicted sequentially, conditioned on the user query, the retrieved passages, and the previously generated tokens (denoted $C_{q,<t}$).

SFT establishes a strong baseline for clarification. By learning to ask questions supported by retrieved passages, the model reduces ambiguity in user intent and provides an evidence-aligned starting point for the subsequent preference based alignment stage. This further improves faithfulness and mitigates hallucinations.

3.2 Faithfulness Alignment

Although the p_{θ_0} model is already fine-tuned to generate clarifying questions that are much more relevant than the initial p_{LM} model, one of its main limitations is its tendency to hallucinate: it may generate details that are absent from the retrieved passages $\mathcal{D}$.

Preference tuning. To mitigate this, we introduce a second training stage focused on faithfulness. We augment the training data with pairs of faithful (C_q^+) and unfaithful (C_q^-) clarifying questions and apply a contrastive learning approach. In particular, we employ DPO [23] over a dataset $\mathcal{T}_2 = \{(U_q, \mathcal{D}, C_q^+, C_q^-)\}$, where the model is explicitly trained to prefer faithful clarifying questions over unfaithful ones. In DPO, the learning objective (Eq. 2) aligns a policy model p_θ with a preference signal, favoring C_q^+ over C_q^-, given the same input $(U_q, \mathcal{D})$, as defined below:

$$\mathcal{L}_{\mathrm{DPO}}(\theta) = -\mathbb{E}_{\sim\mathcal{T}_2}\Big[\log\sigma\Big(\beta\log\frac{p_\theta(C_q^+ \mid U_q, \mathcal{D})}{p_{\theta_0}(C_q^+ \mid U_q, \mathcal{D})} - \beta\log\frac{p_\theta(C_q^- \mid U_q, \mathcal{D})}{p_{\theta_0}(C_q^- \mid U_q, \mathcal{D})}\Big)\Big] \tag{2}$$

Unfaithful clarifying questions generation. Preference-based alignment requires faithful–unfaithful question pairs, but manual creation is costly and automatic detection remains difficult. We propose an unsupervised method that simulates unfaithful questions by injecting controlled noise during decoding. Our method adapts the noisy decoding strategy of Duong et al. [6] to the clarification setting. The approach relies on two complementary models:
`Grounded model` p_{θ_0}`:` obtained by fine-tuning a pretrained base model p_{LM} on half of the dataset $\mathcal{T}_1$. Given a query and retrieved passages $(U_q, \mathcal{D})$, it outputs generally faithful clarifying questions $C_q \sim p_{\theta_0}(\cdot \mid U_q, \mathcal{D})$, though minor inaccuracies remain.
`Ungrounded model` $p_{\texttt{uncond}}$`:` obtained by fine-tuning the same base model but conditioned only on the user query U_q, i.e., $C_q \sim p_{uncond}(\cdot \mid U_q)$. It produces fluent and relevant clarifying questions, yet these are not guaranteed to be grounded in the retrieved passages $\mathcal{D}$.

While p_{uncond} produces overly unconstrained questions and p_{θ_0} tends to remain faithful, their combination yields plausible but unfaithful clarifying questions (the balance is critical, as highlighted by Duong et al. [6]). Specifically, we decode token-by-token from a mixture distribution (Eq. 3), using stochastic decoding (temperature and top-k sampling) to promote diversity and encourage hallucinated tokens.

$$C_{q,t} \sim (1-\alpha_t)\, p_{\theta_0}(\cdot \mid C_{q,<t}, U_q, \mathcal{D}) \;+\; \alpha_t\, p_{\text{uncond}}(\cdot \mid C_{q,<t}, U_q), \tag{3}$$

where $\alpha_t \sim \text{Bernoulli}(\alpha)$ controls the injection of ungrounded content. The noise parameter $\alpha \in [0,1]$ determines the faithfulness–fluency trade-off: $\alpha = 0$ recovers clarifying questions from p_{θ_0}, whereas $\alpha = 1$ generates ungrounded ones from p_{uncond}. The resulting questions remain fluent but contain ungrounded spans, yielding both intrinsic errors (contradictions with retrieved passages) and extrinsic hallucinations (additions not inferable from $\mathcal{D}$). These are used as unfaithful clarifying questions C_q^- in the augmented dataset $\mathcal{T}_2 = (U_q, \mathcal{D}, C_q^+, C_q^-)$, enabling preference optimization for faithfulness alignment.

3.3 Joint Training Objective

Supervised fine-tuning and preference optimization address complementary objectives: supervised fine-tuning operates at the token level, teaching the model to produce clarifying questions, while preference optimization encourages it to prefer faithful outputs over unfaithful ones. To leverage both, we propose a combined training objective: $\mathcal{L}_{\text{RAC}}(\theta) = \gamma \cdot \mathcal{L}_{\text{DPO}}(\theta) \;+\; (1-\gamma) \cdot \mathcal{L}_{\text{SFT}}(\theta)$.

4 Experimental Setup

4.1 Datasets and Evaluation

Datasets. We evaluate RAC on four datasets across conversational search and open-retrieval QA. For search, we use Qulac (derived from TREC Web Track