# RAC: Retrieval-Augmented Clarification for Faithful Conversational Search

Ahmed Rayane Kebir[1,2][0009–0009–2512–832X], Vincent Guigue[3][0000–0002–1450–5566], Lynda Said Lhadj[2][0009–0005–3850–9229], and Laure Soulier[1][0000–0001–9827–7400]

[1] Sorbonne Université, CNRS, ISIR, F-75005 Paris, France
[2] Ecole nationale Supérieure d'Informatique (ESI), Algeria
[3] AgroParisTech, UMR MIA-PS, Palaiseau, France

**Abstract.** Clarification questions help conversational search systems resolve ambiguous or underspecified user queries. While prior work has focused on fluency and alignment with user intent, especially through facet extraction, much less attention has been paid to grounding clarifications in the underlying corpus. Without such grounding, systems risk asking questions that cannot be answered from the available documents. We introduce RAC (**R**etrieval-**A**ugmented **C**larification), a framework for generating corpus-faithful clarification questions. After comparing several indexing strategies for retrieval, we fine-tune a large language model to make optimal use of research context and to encourage the generation of evidence-based question. We then apply contrastive preference optimization to favor questions supported by retrieved passages over ungrounded alternatives. Evaluated on four benchmarks, RAC demonstrate significant improvements over baselines. In addition to LLM-as-Judge assessments, we introduce novel metrics derived from NLI and data-to-text to assess how well questions are anchored in the context, and we demonstrate that our approach consistently enhances faithfulness.

**Keywords:** Conversational Search · Clarifying Questions · RAG.

## 1 Introduction

In open-domain information-seeking tasks, user queries are often short, ambiguous, or under-specified. Such characteristics make it difficult for traditional search systems to accurately capture user intent, as they typically provide only a ranked list of documents or passages without engaging in clarifying interactions [22]. Recent work has explored generating clarifying questions that are relevant, diverse, and human-plausible [8,28,30]. However, little attention has been given to whether these questions are grounded in the document corpus, even though unsupported clarifications may mislead users and harm retrieval effectiveness [10,18].

This is the author's version of the work. It is posted here for your personal use. The definitive version is published in:
*Proc of the 48th European Conference on Information Retrieval (ECIR '26), 29 March–2 April, 2026, Delft, Netherlands*

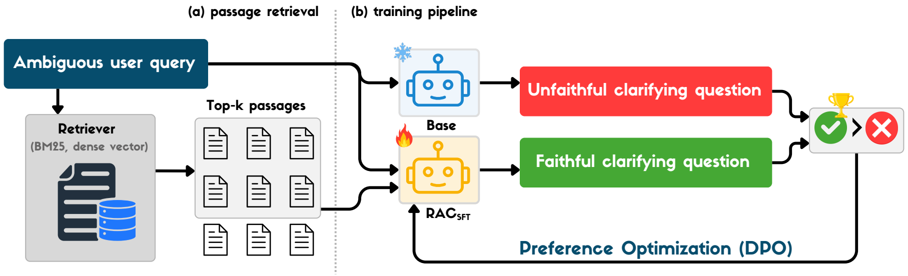

Fig. 1: Overview of RAC. Given an ambiguous user query, the system first retrieves the top-$k$ passages ((a) passage retrieval). A mixture of the fine-tuned model and the base model is then used to generate unfaithful clarifying questions. Both faithful and unfaithful clarifying questions are subsequently leveraged for preference optimization via the DPO algorithm ((b) training pipeline). During inference, the trained model directly generates faithful clarifying questions.

Early approaches to clarifying question generation in conversational search largely relied on facet-based methods. These methods extracted candidate facets from the document collection to produce clarifying questions via templates or sequence-to-sequence models [1,2]. While this offered a basic form of corpus grounding, the reliance on coarse-grained facets proved reductive.

The advent of large language models (LLMs) enabled more fluent generation, with systems either conditioning on extracted facets to produce natural clarifications or directly deriving facets from queries before turning them into questions. Yet the task remains split into two stages—facet identification and question generation—creating bottlenecks in facet extraction and risks of hallucination when clarifications introduce content unsupported by the corpus [27,28].

In this work, we build on the retrieval-augmented generation (RAG) paradigm [12] to ground clarifications directly in the corpus, focusing on answers supported by the documents. Facet extraction is performed implicitly by supplying the top-$k$ retrieved passages to the LLM, which then generates the clarifying question. The first contribution of this article is to propose a fine-tuning of conditional clarification generation, which greatly improves the quality of the questions. To further mitigate entity-level hallucinations, we also introduce a faithfulness reinforcement mechanism that steers the model to rely on the retrieved inputs rather than its internal knowledge, following the approach of [6].
Thus, we aim to address the following research questions:

**RQ1.** How can relevant passages be selected from the corpus, and how many should be used to optimally guide clarification?
**RQ2.** How does conditioning on these relevant passages affect the generation of clarifying questions?
**RQ3.** How can the faithfulness of clarifying question generation be improved when conditioned on relevant passages?

We introduce RAC, a framework for generating clarifying questions grounded in relevant retrieved passages, and train a large language model to prioritize faithful questions using preference tuning and contrastive learning, as illustrated in Fig. 1. We validate our approach on conversational search and open-domain Question Answering datasets through automatic metrics and LLM-as-Judge evaluations. Results show that RAC consistently enhances both the quality and faithfulness of clarifying questions, outperforming existing baselines.

## 2 Related Work

***Query Clarification in Conversational Search.*** Asking clarifying questions enables users to actively participate in query disambiguation, with the goal of better capturing their information intent [1,2,7,11]. Prior work in this area has primarily focused on two tasks: predicting the need for clarification and generating clarifying questions. In this paper, we focus on the latter. Recent studies have increasingly explored large language model based approaches. For instance, Sekulić et al. [27] conditioned an LLM on specific facets; however, such facets are not always readily available and often require external extraction tools. Siro et al. [28] leveraged temperature control and facet information to generate diverse clarifications, while Wang et al. [30] introduced a zero-shot clarifying-question generator using fixed templates and query facets. More recently, Tang et al. [29] proposed a prompting strategy grounded in an ambiguity taxonomy to improve handling of ambiguous queries. Although these methods produce plausible and diverse clarifications, they remain prone to hallucination, frequently generating questions about aspects unsupported by the underlying corpus. Additionally, the reliance on explicit facets limits applicability when facets are difficult to extract or unavailable.

***Retrieval Augmented Generation.*** Since the original article [12], several variants have been proposed, first for question answering [9] and later for clarification. Early studies primarily examined the role of the retriever in selecting corpus-grounded clarifications among candidate suggestions [18], whereas more recent work has shifted the focus toward generation [10], with particular attention to maximizing faithfulness during inference. In addition, [26] demonstrates that the RAG paradigm can be combined with knowledge bases to enhance disambiguation in domain-specific applications. However, these approaches rely on a zero-shot paradigm, whereas we demonstrate the benefit of fine-tuning the generator to better exploit the retrieved passages.

***Preference Tuning.*** Reinforcement learning from human feedback was introduced to align LLMs with human preferences [19], but reward-model methods were costly and often unstable. More recent techniques such as direct preference optimization (DPO) [23] and extensions [32] have improved efficiency by learning directly from pairwise comparisons. Beyond general alignment, generating both faithful and unfaithful baseline sentences allows contrastive learning algorithms