A Review of Fine-Tuning for QA and a Proposal for a New Annotated Corpus

Alice Breton*†[1] and Adam Faci‡[1]

[1]Huma-Num : l'IR* des humanités numériques (Huma-Num IR*) – Aix Marseille Université, Centre National de la Recherche Scientifique, Établissement public Campus Condorcet – France

Résumé

Developing reliable Retrieval-Augmented Generation (RAG) systems for specific domains like social sciences and humanities (SSH) requires question-answering (QA) datasets tailored to the linguistic and epistemological specificities of these fields. This necessity motivates a comprehensive review of existing QA datasets and corpora, ranging from general benchmarks (e.g., SQuAD, Rajpurkar et al., 2016; Natural Questions, Kwiatkowski et al., 2019; HotpotQA, Yang et al., 2018) to domain-specific corpora. However, these resources remain inadequate for SSH, due to a lack of domain-specific linguistic nuances, interdisciplinary complexity, and epistemological diversity, which limits the development of accurate and contextually grounded RAG systems. This review therefore highlights their limited coverage of SSH and the unique challenges posed by this linguistic context. To address this gap, we introduce a new synthetically generated QA corpus derived from real user queries in a documentary search engine, supplemented by human annotation to ensure the accuracy and relevance of questions and answers. This approach aligns with recent work on automatic QA data generation using large language models (e.g., HistoriQA-ThirdRepublic, Pellet et al., 2026). Beyond proposing this new corpus, we review state-of-the-art fine-tuning methods for adapting language models to QA tasks, such as supervised fine-tuning, instruction tuning, and lightweight adaptation techniques like LoRA, discussing their strengths and limitations for interpretable and responsible use in SSH. This work contributes to the broader discussion on models, methods, and applications of interpretable AI in SSH, while examining trust, explainability, and biases in generation systems when relying on domain-specific corpora and annotations.

Mots-Clés: RAG, QA, SSH, Fine, tuning, State of the art, Benchmark

*Intervenant

†Auteur correspondant: alice.breton@huma-num.fr

‡Auteur correspondant: adam.faci@huma-num.fr