# CREDIT: Credit-Driven Evidence Selection for Multi-Source Retrieval-Augmented Generation

## Abstract

Retrieval-augmented generation (RAG) succeeds or fails on how well it chooses what to read. When evidence arrives from disparate repositories—APIs, web pages, manuals, tables—uniform aggregation or fixed routing across sources blurs complementary signals and magnifies duplication. At the same time, relevance-oriented re-rankers optimize proxy scores that may not track the generator's loss. We introduce CREDIT, a training-free procedure that treats retrieval as credit assignment to an evidence portfolio. From a single reverse-mode sweep through the language model, CREDIT computes first-order, loss-aligned attributions that estimate each candidate's marginal utility for the generator. These attributions drive a two-level policy: (i) allocate a read budget across sources via a diversity-regularized utility that balances query affinity with cross-source novelty; (ii) within each source, admit documents by greedy accumulation of estimated loss reduction until the budget is exhausted. Our analysis connects these attributions to a smooth, monotone surrogate that yields a near-submodular objective, for which greedy selection enjoys standard approximation and regret bounds; we also bound the excess risk relative to an oracle chooser. On multi-dataset QA and open-ended generation, CREDIT consistently improves answer quality and reduces evidence volume compared to uniform pooling and single-source baselines, indicating that objective-coupled credit assignment is crucial for multi-source RAG.