

# Oracle-RLAIF and RLHF Robustness in Noisy Retrieval on DiDeMo Benchmark

Assignee Research

May 30, 2026

## Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does the robustness of Oracle-RLAIF compare to reinforcement learning from human feedback (RLHF) when evaluated on the DiDeMo benchmark's retrieval accuracy under noisy or adversarial input. Spoken query retrieval is an important interaction mode in modern information retrieval. However, existing evaluation datasets are often limited to simple queries under constrained noise conditions, making them inadequate for assessing the robustness of spoken query retrieval. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: SQuTR: A Robustness Benchmark for Spoken Query to Text Retrieval under Acoustic Noise. Research question: How does the robustness of Oracle-RLAIF compare to reinforcement learning from human feedback (RLHF) when evaluated on the DiDeMo benchmark's retrieval accuracy under noisy or adversarial input conditions?.

## 2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.5/10.

### **3 Results**

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.5/10.

### **4 Limitations**

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

### **References**

- <http://arxiv.org/abs/2510.02561v1>
- <http://arxiv.org/abs/2602.12783v2>
- <http://arxiv.org/abs/2312.11456v4>