

We thank the reviewers for their constructive feedback. We have revised the manuscript to address all raised points. New changes in the revised manuscript are highlighted for your convenience.

Below is a summary of how we addressed each specific comment.

### **Response to Reviewer 1**

1. Comment regarding weak logic and comparison to specialized models (e.g., o1-mini). Response: Addressed in the Discussion section. We have added text acknowledging the simplicity of the symbolic triples while emphasizing that the generated Python code allows for complex logic. We also added a comparison highlighting that while models like o1-mini are effective, NSAR provides superior auditability and interpretability for high-stakes domains.
2. Comment regarding short Related Works section. Response: Addressed in Section 2. We have expanded the section to include the historical context of Cross-Lingual Information Retrieval (CLIR) and probabilistic structured queries.
3. Comment regarding the definition of symbol N in Cost Efficiency Analysis. Response: Addressed in Section 4.6. We have explicitly defined N as the total number of sentences in the haystack.
4. Comments regarding formatting (bolding, figure legends, BGE-M3 capitalization). Response: Addressed throughout the manuscript. We have removed excessive bolding, standardized the capitalization of "BGE-M3," and updated Figure 1 caption to explicitly describe the solid vs. dashed lines.

### **Response to Reviewer 2**

1. Comment regarding the novelty of CROSS vs. existing RAG methods. Response: Addressed in Section 3.1. We have clarified that CROSS is a RAG backbone specifically optimized for sentence-level granularity and massive scale (512k tokens), serving as the necessary prerequisite for the neurosymbolic module.
2. Question regarding the "LLM-only" baseline setup. Response: Addressed in Section 4 (Results). We added a "Baseline Definition" paragraph explicitly stating that the "LLM-only" baseline has access to the full document text (up to the context limit), ensuring a fair comparison.
3. Comment suggesting separation of CLIR vs. Long-Context challenges. Response: Addressed in Section 1 (Introduction). We now distinguish between "needle-in-a-haystack" (attention degradation) challenges and "cross-lingual barrier" (semantic misalignment) challenges.

4. Question regarding how NSAR+3 combines strategies and if baselines use RAG. Response: Addressed in Section 4.7. We clarified that all baselines (CoT, ReAct, Self-Reflection) utilize the CROSS retrieval backbone. Regarding NSAR+3, we detailed the integrated prompting sequence.

5. Minor comments on citation formatting. Response: Corrected throughout the manuscript.