

Troubleshooting tree

No trace: stop and add one.

Wrong fact: was the span in the pack? If not, refuse or fix the index.

Tool error: the model is quoting your handler.

Should have refused: the gate failed, not the tone.

Slow phone turn: time ASR, model, and TTS separately.

Change one thing. Rescore the same tasks.

Sources and go deeper

Lewis and co-authors, Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, 2020 (paper)

last checked 2026-10

<https://arxiv.org/abs/2005.11401>

NIST AI Risk Management Framework (spec)

last checked 2026-10

<https://www.nist.gov/itl/ai-risk-management-framework>

MDN Web Docs, Web Speech API (community)

last checked 2026-10

https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API

Retrieval-augmented generation - Wikipedia (wikipedia)

last checked 2026-10

https://en.wikipedia.org/wiki/Retrieval-augmented_generation

This is an independent, free, learning-only student project. It respects every company and vendor, and every trademark belongs to its owner. It is not affiliated with or endorsed by any of them. Nothing here is sold. The author is happy to volunteer.

cyphermonkeylabs.com

No email address is printed on this sheet.