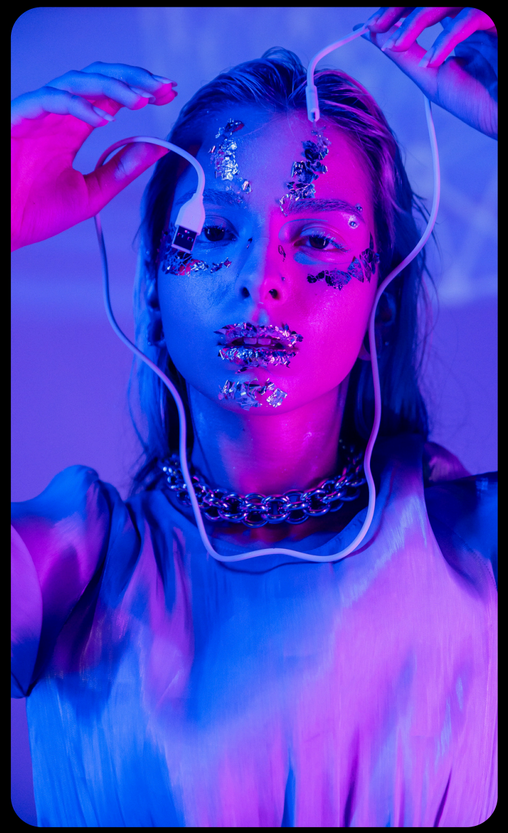
HALLUCINATION CORRECTION ASSISTANT

AI-GENERATED HALLUCINATIONS USING A FINE-TUNED GPT-3.5 TURBO MODEL

Presented By: Darious Brown

Bellevue University

www.reallygreatsite.com



Why Hallucinations Matter

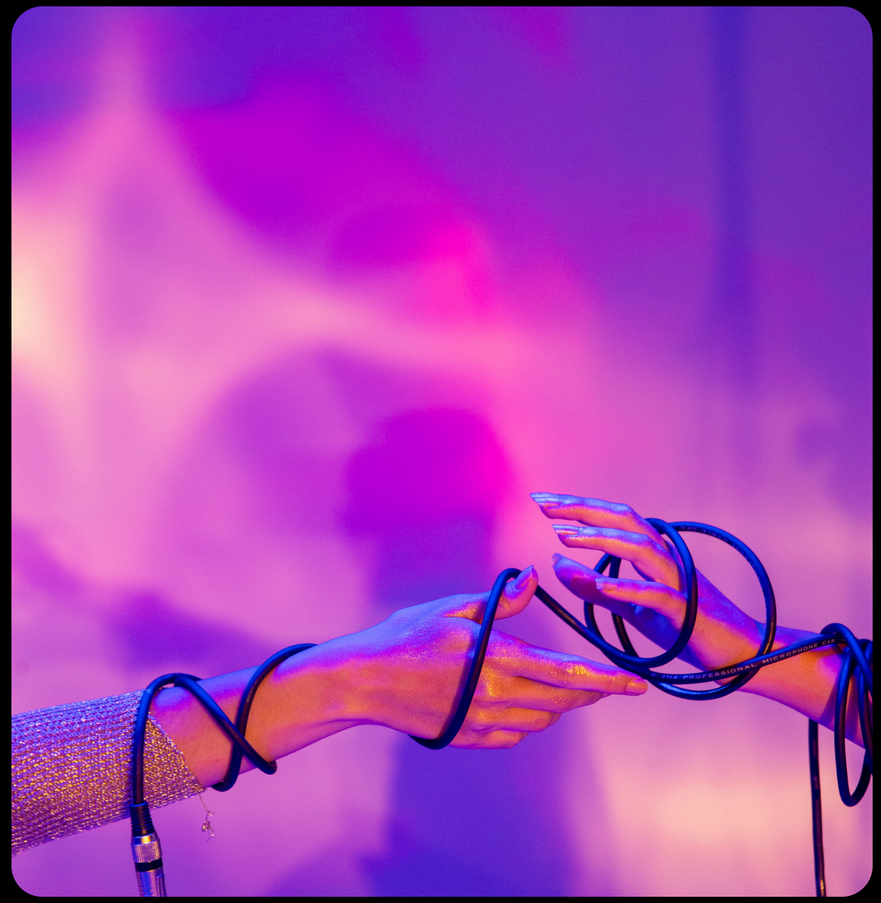
- Large Language Models (LLMs) often generate false but fluent-sounding outputs
- These “hallucinations” pose risks in education, healthcare, and enterprise AI
- Need for tools to verify and correct hallucinated content

Bellevue University

DARIOUS BROWN

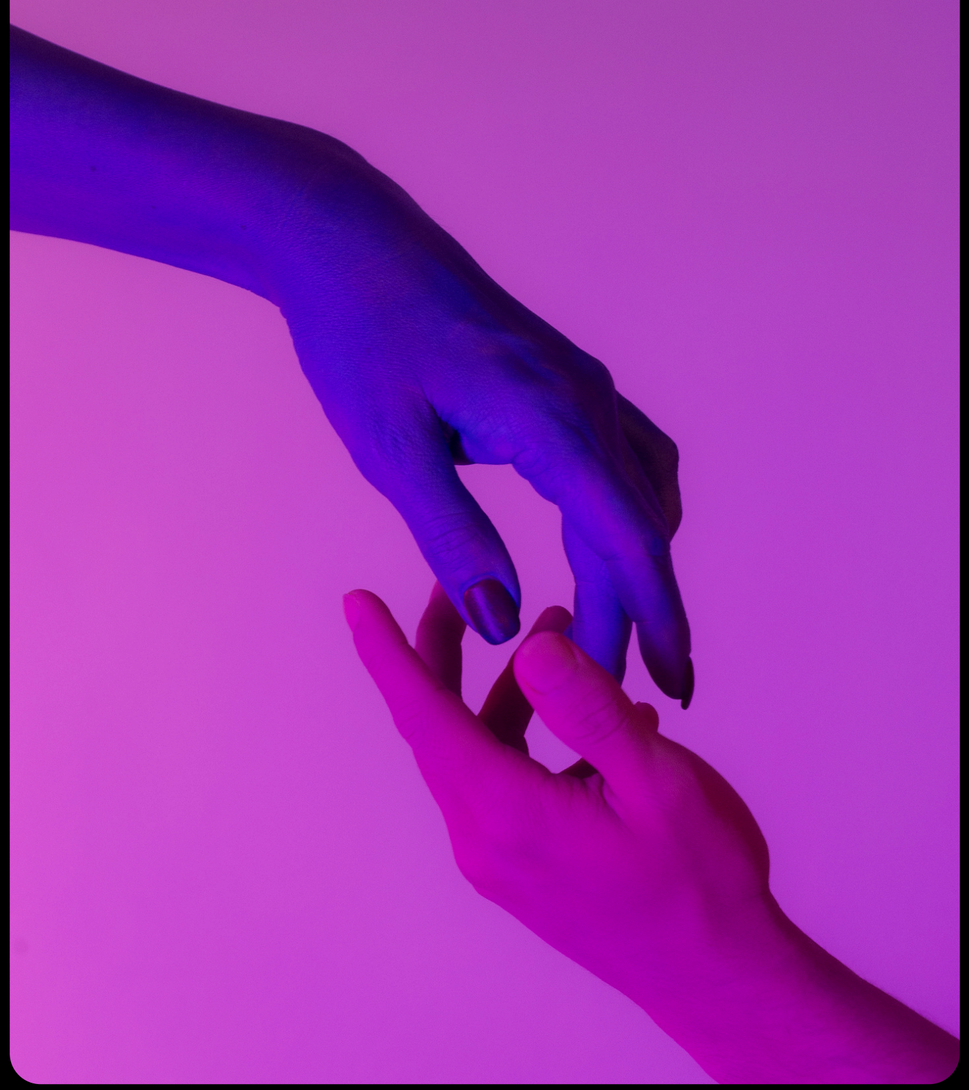
Project Goals

- Detect and correct hallucinations using a fine-tuned GPT model
- Offer clear and accurate outputs
- Measure and visualize model confidence
- Compare model responses with external LLMs for validation



Bellevue University

www.reallygreatsite.com



Model Architecture

- **Fine-Tuned Model Name:** ft:gpt-3.5-turbo-0125:dares-apis:hallucination-logical-v1:BvwmRqwc
- **Tools Used:**

1. OpenAI fine-tuning
2. Streamlit for UI
3. Fuzzy string matching for response similarity

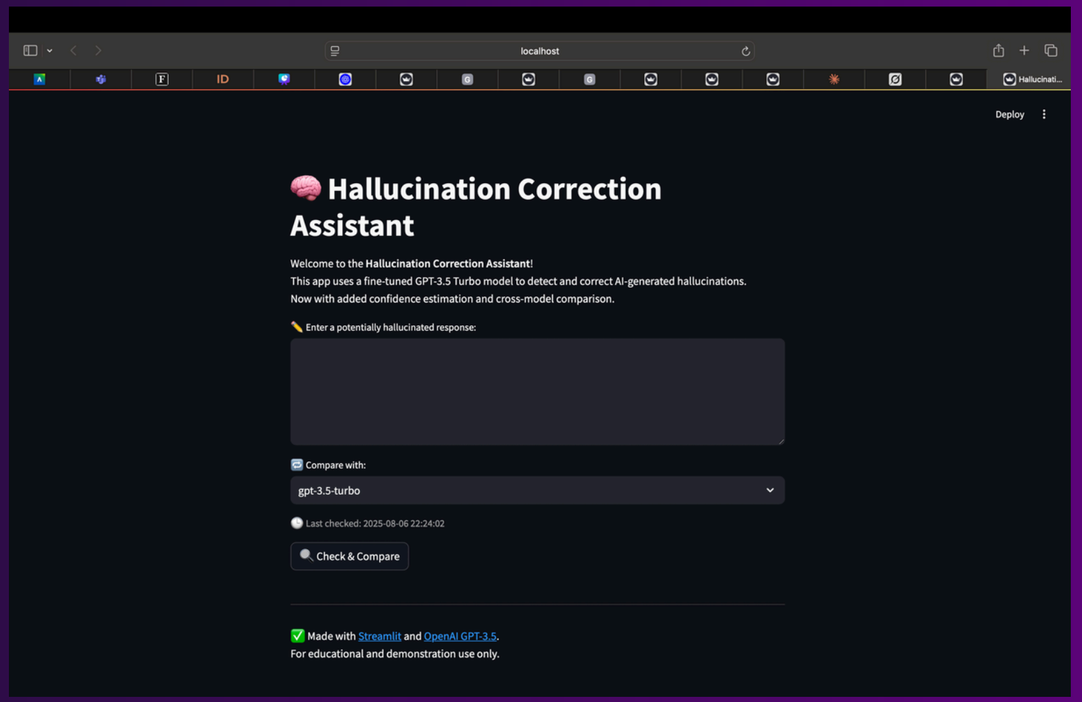
- **Diagram Flow:** Simple flow: User Input → Model → Correction + Confidence Score

DARIOUS BROWN

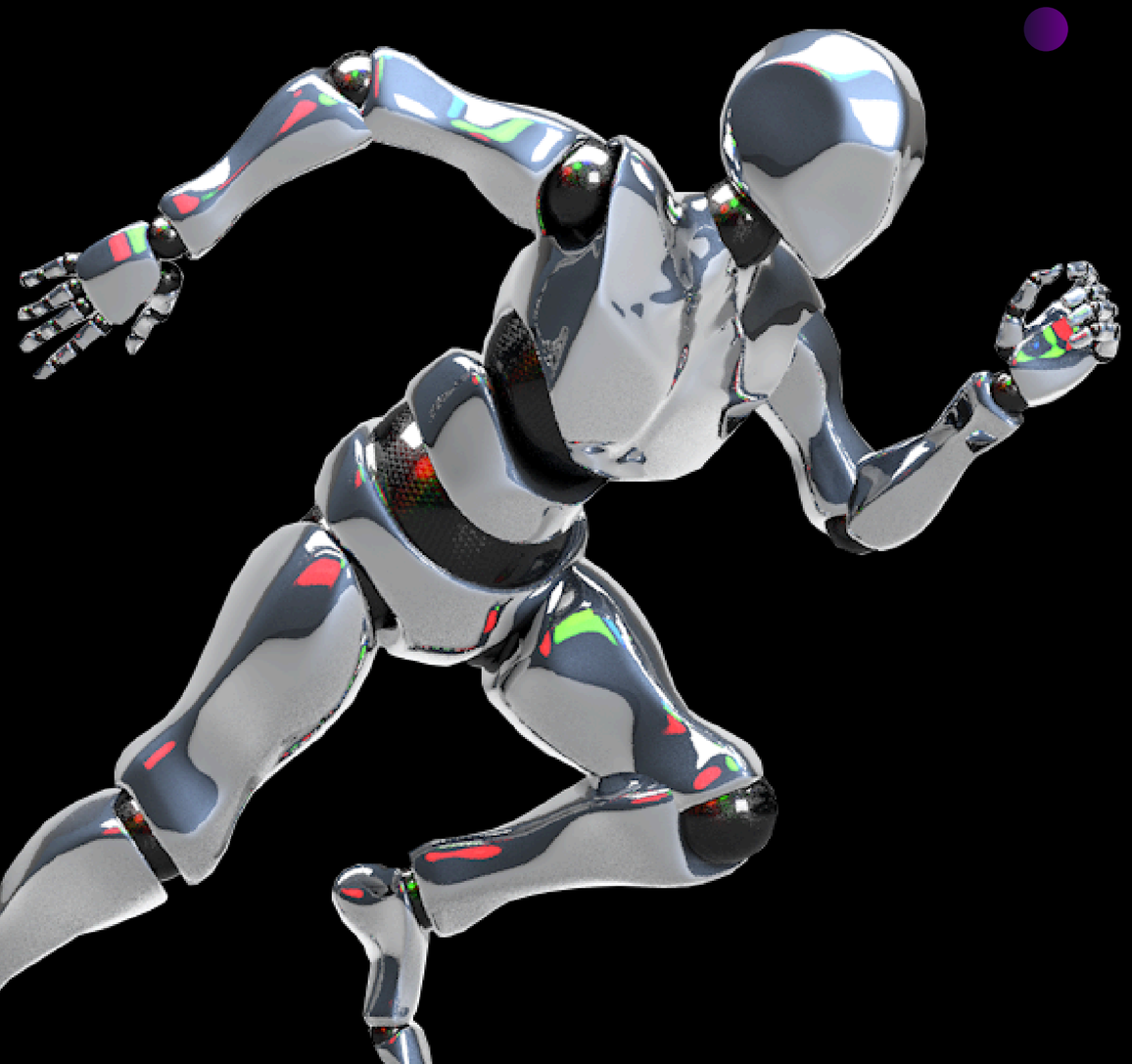
User Interface (Streamlit)

Labels for:

- Input field
- Corrected output
- Confidence percentages
- Compare button
- Mention: Real-time interaction, no coding required



DARIOUS BROWN

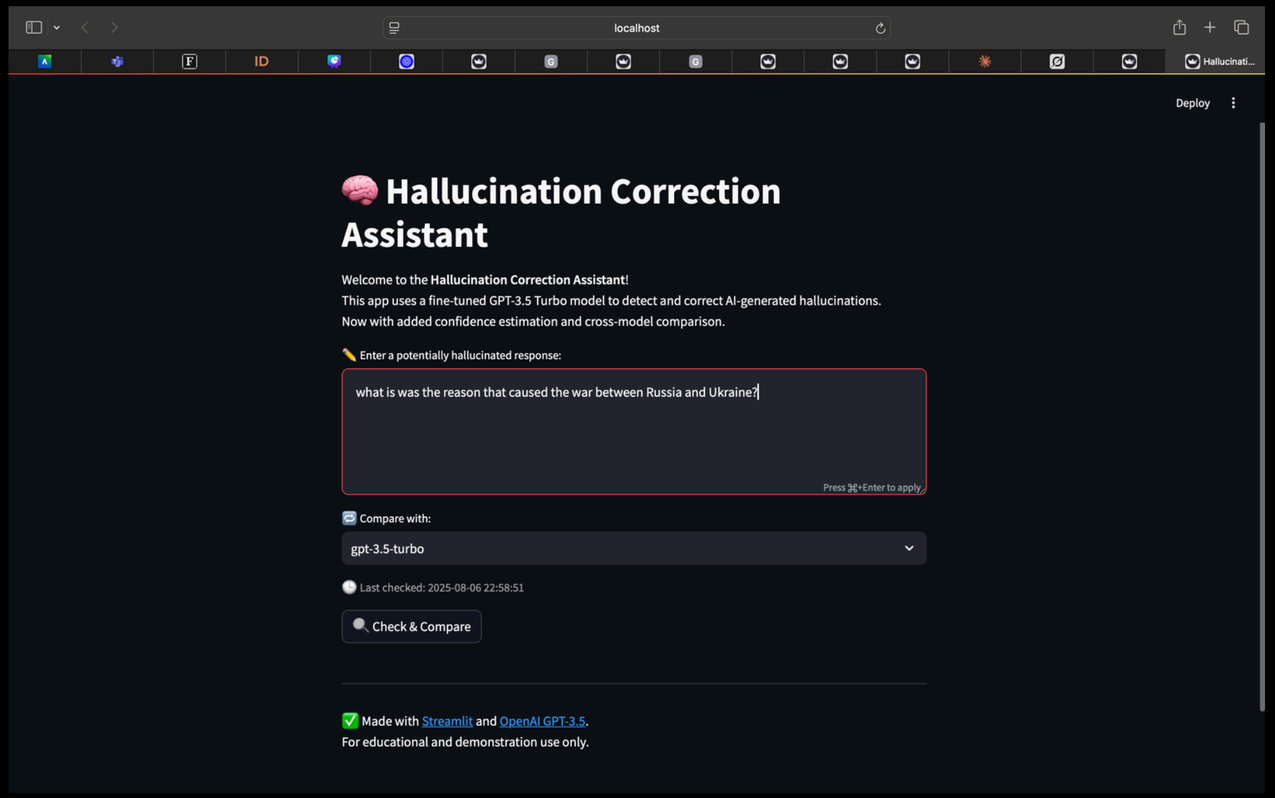


Enhancements Summary

- Confidence score based on fuzzy string matching
- External model response comparison
- Timestamped analysis for auditability
- Error handling for clean UX
- OpenAI best practices (separating roles, safety layers)

DARIOUS BROWN

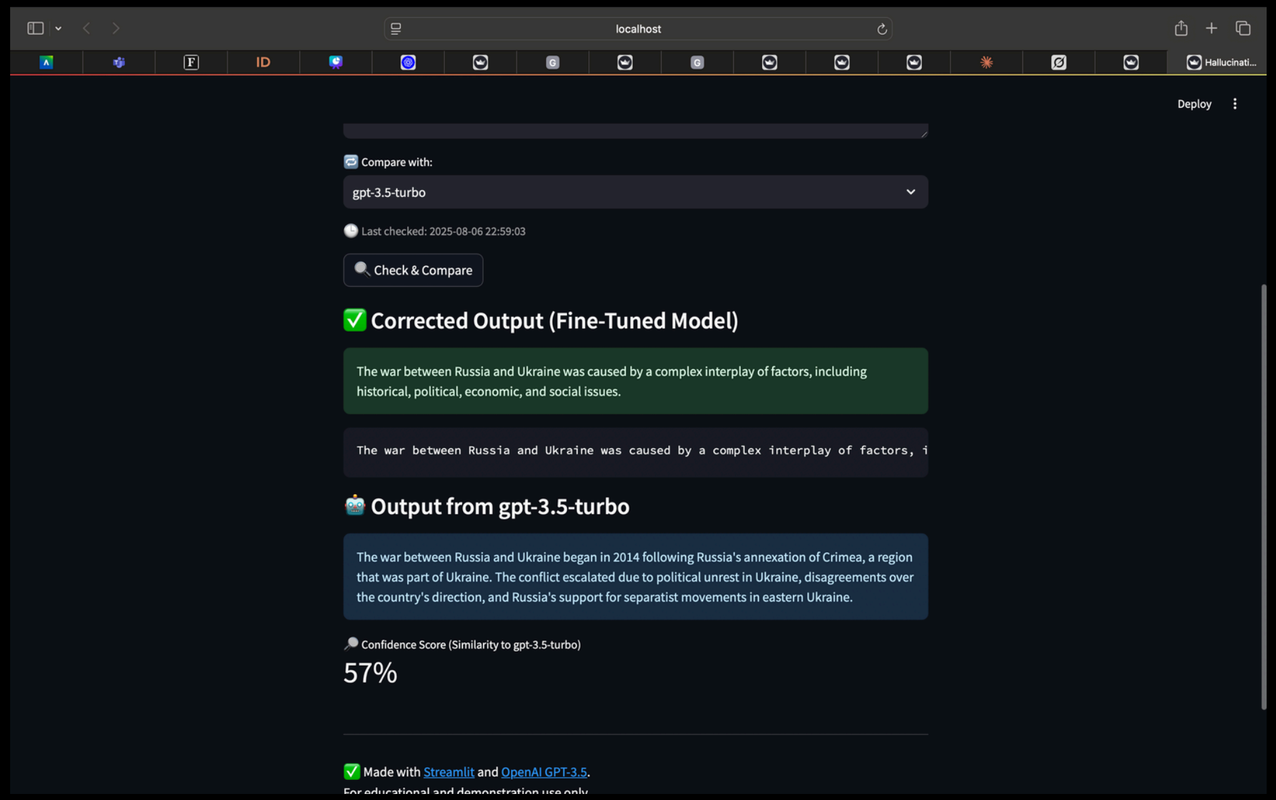
EXAMPLE INPUT



www.reallygreatsite.com

DARIOUS BROWN

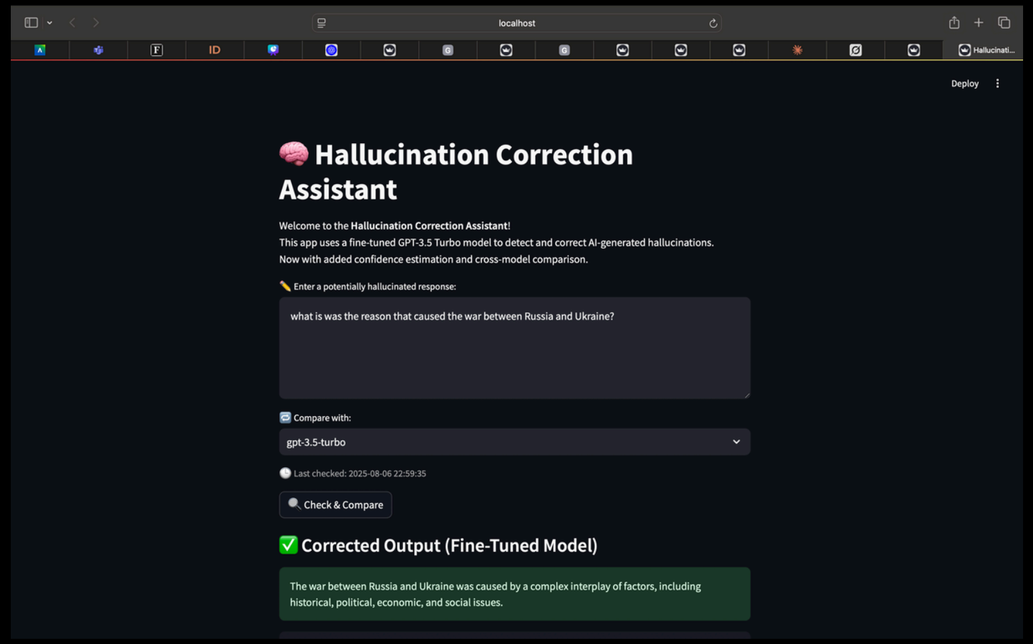
EXAMPLE OUTPUT

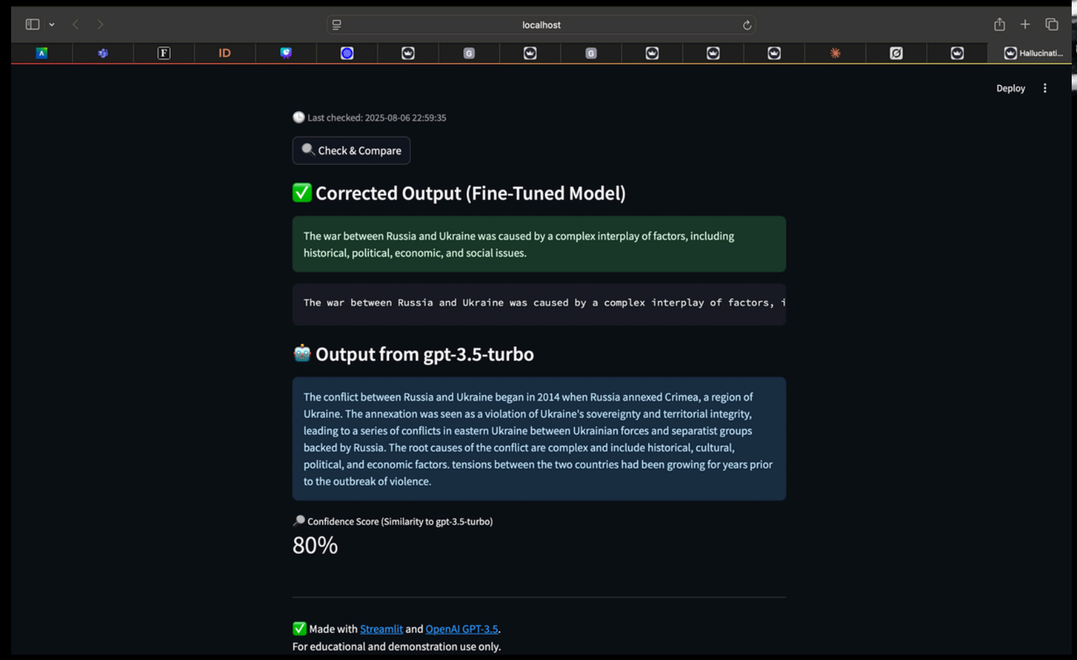


www.reallygreatsite.com

Model Growth Through RAG

DARIOUS BROWN





As questions were repeated the ,model showed a direct positive relationship through an increased confidence percentage

Bellevue Univeristy

www.reallygreatsite.com

Conclusion

- **AI hallucination correction is critical in LLM deployment**
- **Our app demonstrates a practical, user-friendly solution**
- **Future work:**
  - Support multi-sentence documents
  - Cross-model benchmarking
  - Extend to medical/financial domains

Get in Touch

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed at ipsum vitae lacus lobortis lacinia. Donec tristique arcu massa, at.

Phone Number
267-309-5099

Email
darious.brown3@icloud.com

Website
www.reallygreatsite.com

GitHub Repository & App Link
@reallygreatsite