

AI Agent for Misinformation Detection

Ashley Lescarbeau, University of New Hampshire



Background & Introduction

As of early 2026, an estimated 71% of images on social media are AI-generated, accelerating the spread of misinformation across many platforms.

Traditional fact-checking methods, while valuable, are too slow to keep up with the rapid increase in false contents. Automated, intelligent systems that can draw from reliable and trusted sources are essential to helping users assess the credibility of the media they consume daily.

Research Objectives

1. Build an agentic AI system to classify news articles as Real, Fake, or Unable to Determine.
2. Integrate retrieval-augmented generation (RAG) with a live news vector store.
3. Enable the agent to learn source trustworthiness over time via reinforcement-style feedback.

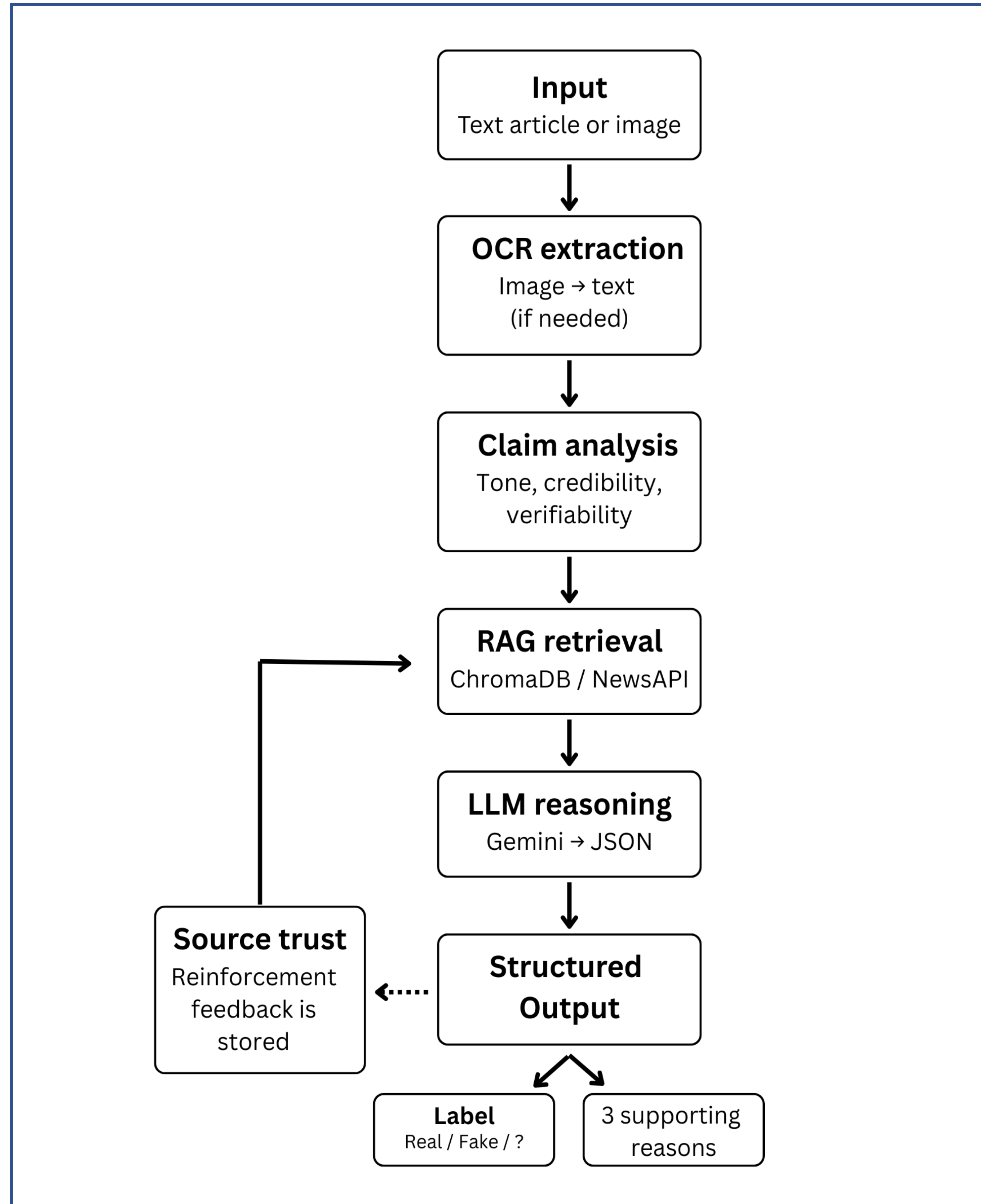
Key Design Choices

Google Gemini: Multimodal LLM for text generation / embeddings

ChromaDB Vector Store: Stores News API article embeddings. Retrieves $k = 3$ semantically similar articles to ground LLM

Structured Output: Agent always returns reasons with label to ensure explainability

System Architecture



Reinforcement Feedback

"Trust" of a source starts at 0.5 for all new sources.

Correct prediction: **+0.1 (max 1.0)**

Wrong prediction: **-0.1 (min 0.0)**

Unknownblog.com	Fake (T)	0.5 -> 0.6
Bbcnews.com	Fake (F)	0.5 -> 0.4
Who.org	Real (T)	0.5 -> 0.6

Conclusions

The agent successfully combines LLM reasoning with RAG retrieval to classify news credibility with structured explanations. The reinforcement-style trust update mechanism allows the agent to learn from experience and adapt over time. This provides a foundation for a scalable, explainable misinformation detection system.

Future Work

Image Detection: Extend OCR pipeline to classify AI-generated or manipulated social media images

Expanded Source Coverage: NewsAPI free tier limits volume for testing. Integrating additional APIs would broaden vector store and increase accuracy and reliability.

User Interface: Build a web UI so students, educators, and social media users can paste or upload content for real-time analysis.

Resources & Acknowledgements

Marr, B. (2025). 15 Mind-Blowing AI Statistics Everyone Must Know. *Forbes*.

Lewis, P. et al. (2020). Retrieval-Augmented Generation for NLP. *NeurIPS*.

CS 675: Machine Learning in Practice, (2026). Matthew Magnusson, UNH

**This poster template provided courtesy of
UNH ESRC Poster Printing Services**

Trust us to make your poster look **GREAT!**

Website: <http://posters.unh.edu>

Poster Guide: <http://goo.gl/1E7TJY>



DELETE THIS SLIDE BEFORE PRINTING