AI HALLUCINATION AND CITATION VERIFICATION SYSTEM

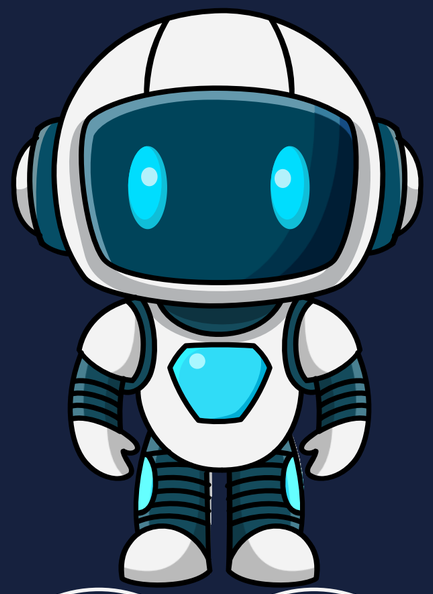
Presented by the Apex-1

PROBLEM STATEMENT :

Generative AI models are widely used for research, learning, and decision making. However, these models often generate factually incorrect information presented with high confidence. A critical extension of this problem is the creation of fake citations, non-existent references, and broken links, which appear legitimate but cannot be verified. This makes it difficult for users to trust AI-generated content and may lead to misinformation, legal risks, and ethical concerns.

Build a system that can detect, flag, and verify factual claims and citations generated by AI models, helping users distinguish between reliable and unreliable AI-produced information.

Proposed Solution



We propose an AI Claim Verification System that:

- Accepts a natural language claim as input
- Extracts key entities using NLP
- Detects negation and misleading phrasing
- Retrieves evidence from trusted knowledge sources
- Verifies whether the claim is True or False
- Outputs a confidence score and explanation

Architecture

Input Claim (Text)

⬇

Text Preprocessing & Normalization

⬇

Named Entity Recognition (NER)

⬇

Negation & Question Detection

⬇

Evidence Retrieval (Wikipedia API)

⬇

Rule-Based + Heuristic Verification Engine

⬇

Claim Verdict (True / False + Confidence + Explanation)

INNOVATION

- Claim-level verification, not just keyword search
- Combines NLP + external knowledge instead of pure LLM guessing
- Handles negation, false confidence, and misleading claims
- Uses explainable rules, not black-box decisions
- Lightweight, fast, and suitable for real-time use

SOCIAL IMPACT

- Reduces spread of misinformation
- Improves trust in AI-generated content
- Supports students, researchers, and journalists
- Helps fight fake news and AI hallucinations

MARKET IMPACT

Can be integrated into:

- AI chatbots, search engines , New platforms
- Educational tools, Enterprise AI systems

Valuable for:

- EdTech, Media & publishing
- Legal & compliance tools
- AI governance platforms

Technology Stack

Core Technologies

- Python
- SpaCy (NLP & NER)
- Wikipedia API
- Regex-based logic
- JSON structured output

Future Enhancements

- Google Fact Check API
- Knowledge Graph integration
- LLM-based reasoning layer

Platform & Tools

- Google Colab
- GitHub
- REST-ready architecture

THANK YOU