

Detecting Misinformation & AI-Generated Content on Social Media Platforms

Amy Louise Harrison | Student ID: 2133368

Supervisor: Mobolanle Bello

6B+

social media users worldwide exposed to misinformation daily

70%

of users cannot reliably identify AI-generated content

<3s

target response time for credibility analysis in the prototype

01 - BACKGROUND

Introduction

Social media has transformed global communication, yet simultaneously accelerated the spread of misinformation. The rise of large language models (LLMs) has compounded this challenge — AI tools can now generate highly convincing fake news, images, and videos that are increasingly indistinguishable from authentic content. Research confirms that false information spreads faster than corrections, and many users continue to judge credibility by appearance rather than accuracy. Despite extensive academic research on misinformation detection, few practical, accessible tools exist for everyday users.

RESEARCH QUESTION

"How can a lightweight, easy-to-use website help users detect potential misinformation and AI-generated content on social media platforms?"

02 - LITERATURE

Key Findings from Literature

- Tasnim et al. (2021) — COVID-19 misinformation spread rapidly on Facebook & Twitter; emotionally charged content prioritised by algorithms (Pulido et al., 2023).
- O'Hara & Shackman (2024) — LLMs generate convincing fake articles; existing AI detectors show inconsistent accuracy (Howard et al., 2024).
- Wang (2025) — Scientometric mapping reveals need for explainable, applied detection tools, not just theoretical models.
- Baker & Mehra (2024) — Complex detection tools are inaccessible to non-expert users; simple rule-based approaches remain effective.
- Lam & Ng (2025) — User awareness and digital literacy are as important as automated detection in combating misinformation.

03 - OBJECTIVES

Research Aim & Objectives

To design and evaluate a simple web-based tool that helps users identify potential misinformation and AI-generated content on social media platforms.

- Analyse existing approaches to detecting misinformation and AI-generated content.
- Design and develop a prototype website (HTML, CSS, JavaScript) providing a credibility score and fact-checking links.
- Evaluate the prototype through user testing, focusing on usability and perceived usefulness.
- Provide recommendations for improving online information literacy based on findings.

08 - CONCLUSIONS

Conclusions & Future Work

This project demonstrates that a lightweight, rule-based web tool can meaningfully support everyday users in assessing social media credibility. While not claiming complete detection accuracy, the prototype successfully raises awareness and encourages critical engagement with online information. Future developments could incorporate machine learning models, real-time database verification, and multilingual content analysis to significantly expand impact and accuracy.

04 - METHODOLOGY

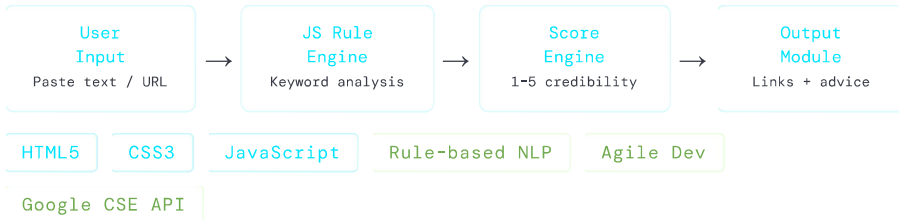
Development Approach

An Agile iterative development approach was adopted, enabling continuous refinement through cycles of build → test → evaluate.



05 - SYSTEM

System Architecture



06 - ETHICS & RISKS

Ethical & Legal Considerations

- Only publicly available social media content analysed — no private data accessed.
- No personal user data stored; GDPR-compliant design throughout.
- Results clearly framed as advisory — tool does not claim definitive factual accuracy.
- Potential keyword bias acknowledged; diverse test datasets used to mitigate.
- Validated against established fact-checking platforms: Snopes and PolitiFact.

07 - OUTCOMES

Key Outcomes & Impact

- A fully functional prototype website with real-time credibility scoring for social media content.
- User evaluation demonstrating improved critical thinking before sharing online content.
- A set of recommendations for improving digital literacy among everyday social media users.
- Foundation for future work: advanced ML detection, multilingual support, and educational integration.