PROJECT REPORT

Exploring Tools and Websites for Detecting Fake News, Fake Emails, Fake WhatsApp Messages, and Fake Social Media Posts

DOMAIN NAME -

[Digital Media Verification and Misinformation Detection]

CDAC, Noida

CYBER GYAN VIRTUAL INTERNSHIP PROGRAM

Submitted By:

Vishal Kumar

Project Trainee, (30 July- 13 August) 2025

BONAFIDE CERTIFICATE

This is to certify that this project report entitled, Exploring Tools and Websites for Detecting Fake News, Fake Emails, Fake WhatsApp Messages, and Fake Social Media Posts submitted to CDAC Noida, is a Bonafede record of work done by Vishal Kumar under my supervision from 30th July to 13th August 2025.

Declaration by Author(s)

This is to declare that this report has been written by me. No part of the report is plagiarized from other sources. All information included from other sources have been duly acknowledged. I aver that if any part of the report is found to be plagiarized, I shall take full responsibility for it.

Name of Author: Vishal Kumar

TABLE OF CONTENTS

4. Implementation

4.1 Fake News Verification

4.2 Phishing Email Analysis

4.3 WhatsApp Message Verification

4.4 Multimedia Content Verification

4.5 Indicators of Compromise (IOCs)

5. Results & Discussion

5.1 Summary of Findings

5.2 Correlation with Threat Indicators

5.3 Challenges Faced During Investigation

6. Conclusion & Recommendations

6.1 Conclusion

6.2 Recommendations and Countermeasures

6.3 Future Scope of Research

7. References

ACKNOWLEDGEMENT

I sincerely thank the **Centre for Development of Advanced Computing (CDAC)** for giving me the opportunity to work on the project **"Exploring Tools and Websites for Detecting Fake News, Fake Emails, Fake WhatsApp Messages, and Fake Social Media Posts."**

My heartfelt gratitude goes to my mentor, **Ms. Kajal Kashyap**, whose guidance, thoughtful feedback, and constant encouragement have been a source of inspiration throughout this work. I am equally grateful to the entire CDAC team for their generous assistance and for providing the essential resources that made the successful completion of this project possible.

Exploring Tools and Websites for Detecting Fake News, Fake Emails, Fake WhatsApp Messages, and Fake Social Media Posts.

1.Introduction

1.1 Background of the Study

The expansion of the internet, social media, and instant messaging has revolutionized information exchange, but also enabled the rapid spread of misinformation and disinformation. Malicious actors exploit these platforms to circulate fake news, phishing emails, manipulated WhatsApp messages, and falsified social media posts for financial, political, or disruptive purposes.
The rise of generative AI and deepfake technology has further blurred the line between real and fabricated content, making detection increasingly challenging. This has led to the development of specialized tools and websites to verify the authenticity of digital information and counter misinformation.

1.2 Problem Statement

False information spreads faster than ever, amplified by low digital literacy, algorithmic promotion, and minimal content verification before publication. The problem manifests in multiple forms:

- Fake news articles influencing public opinion and undermining journalism.
- Phishing or spoofed emails leading to data breaches and cybercrime.
- Fake WhatsApp messages distributing hoaxes, scams, and deepfakes.
- Manipulated social media posts fueling disinformation campaigns.
  The challenge is to identify and evaluate tools capable of effectively detecting and preventing such content across platforms.

1.3 Scope of the Study

The project investigates tools and websites that detect fake news, fake emails, fake WhatsApp messages, and fake social media posts. It includes:

- Identifying both fact-checking platforms and AI-based verification systems.
- Evaluating their capabilities in detecting false text, altered images, deepfake videos, and fraudulent emails.
- Comparing their effectiveness, usability, and limitations.
  The study excludes misinformation spread through offline channels or encrypted networks beyond public datasets.

1.4 Learning Objectives

1. Understand the nature and impact of misinformation in the digital era.
2. Explore a variety of detection tools and websites.
3. Evaluate tool performance with case studies and examples.
4. Examine how emerging threats like deepfakes are addressed.
5. Recommend best practices for content verification.

1.5 Significance of the Study

This study benefits cybersecurity professionals, journalists, researchers, and the public by compiling reliable resources to combat misinformation.
For cybersecurity, detecting phishing emails and social engineering attempts can prevent breaches and losses. In media, fact-checking preserves trust and integrity.
The study also supports digital literacy, empowering users to verify information before sharing, which is vital in countering manipulation in today's digital landscape.

2. Literature Review

2.1 Overview of Misinformation and Digital Verification

Misinformation refers to false or misleading content shared without intent to harm, while disinformation involves deliberate deception. The digital age has amplified both, with social media algorithms, instant messaging, and content automation enabling rapid, large-scale dissemination.
Verification of digital content has evolved from manual fact-checking to AI-assisted and automated systems. These use Natural Language Processing (NLP), computer vision, and metadata analysis to identify inconsistencies in text, images, videos, and email headers.

2.2 Role of Detection Tools in Combating Misinformation

Detection tools help users, journalists, and organizations assess the authenticity of content before it spreads further.

- **Fake news detectors** analyze linguistic patterns, source credibility, and cross-referenced facts.
- **Email verification systems** validate sender authenticity through SPF, DKIM, and DMARC, and use phishing-detection algorithms.
- **Messaging verification tools** like fact-checking chatbots help verify forwarded messages in real-time.
- **Social media monitoring platforms** track suspicious activity, bot accounts, and coordinated campaigns.
  These tools act as the first line of defense in digital trust and cybersecurity frameworks.

2.3 Notable Incidents of Detection Preventing Harm

Although quantifying prevention is complex, several cases demonstrate that timely detection tools have mitigated risks:

- During election cycles, AI-powered fact-checking platforms flagged and removed thousands of manipulated political posts before they went viral.

- Phishing detection systems in corporate email servers have blocked fake invoices and spear-phishing attempts, saving organizations from major financial losses.
- Fact-checking bots on WhatsApp in countries like India have debunked viral hoaxes that could have incited panic or violence.

2.4 Existing Tools and Methodologies

Numerous tools exist to detect and verify digital content:

- **Fact-Checking Websites:** Snopes, FactCheck.org, PolitiFact, CheckYourFact.
- **Reverse Image Search:** Google Reverse Image Search, TinEye, InVID.
- **Email Verification:** MXToolbox, PhishTool, SPF/DKIM/DMARC checkers.
- **Social Media Monitoring:** Hoaxy, Botometer, CrowdTangle.
- **WhatsApp Verification:** IFCN and AFP Fact-Check chatbots, MCA Deepfake Helpline.
  Methodologies often combine OSINT (Open-Source Intelligence) with AI-based detection to enhance accuracy and reduce manual verification effort.

3. Approach & Methodology

Here is the tools and technologies used, the infrastructure created, and the diagram depicting the same, including the machines/servers/firewalls etc., with IP addresses.

3.1 Investigation Framework

The project follows a systematic approach to evaluate the effectiveness of tools and websites used for misinformation detection. The investigation framework includes:

1. Requirement Analysis – Define detection categories: fake news, fake emails, fake WhatsApp messages, and fake social media posts.
2. Tool Identification – Select a diverse set of tools, including AI-powered verification platforms, traditional fact-checking sites, and OSINT utilities.
3. Evaluation Criteria – Establish performance metrics such as detection accuracy, processing time, usability, and coverage scope.
4. Testing & Validation – Apply tools on curated datasets containing verified fake and real content, then compare results.
5. Analysis & Recommendation – Summarize findings and suggest the most effective tools for each category.

3.2 Tools & Technologies Used

To address the challenge of detecting fake news, emails, WhatsApp messages, and social media posts, the following tools and platforms were used:

News and Claim Verification Tools

- FactCheck.org – Verifies political statements and media claims.
- Snopes – Debunks internet hoaxes and urban legends.
- PolitiFact – Checks factual accuracy of political content.
- CheckYourFact – Validates viral content and misinformation.
- Google Fact Check Explorer – Aggregates fact-check articles.

Email and Review Scanning Tools

- Urlscan.io – Detects fake reviews, phishing offers, and spam email links.

WhatsApp Message Indicators

- WhatsApp “Forwarded” tags – Used to identify frequently forwarded messages.
- External validation via fact-check websites.

Image and Video Verification Tools

- TinEye – Performs reverse image search.
- Google Reverse Image Search – Identifies reused or misleading images.
- InVID Verification Plugin – Breaks videos into keyframes and validates them using reverse search.

3.3 Infrastructure Created

The project was primarily tool-based and exploratory in nature, it was implemented using a standard personal computing setup with basic security configurations.

Machine/Client

- Device Used: Personal Laptop
- Specifications: Windows 11, Intel Core i5, 16GB RAM
- Local IP Address: 192.168.0.112

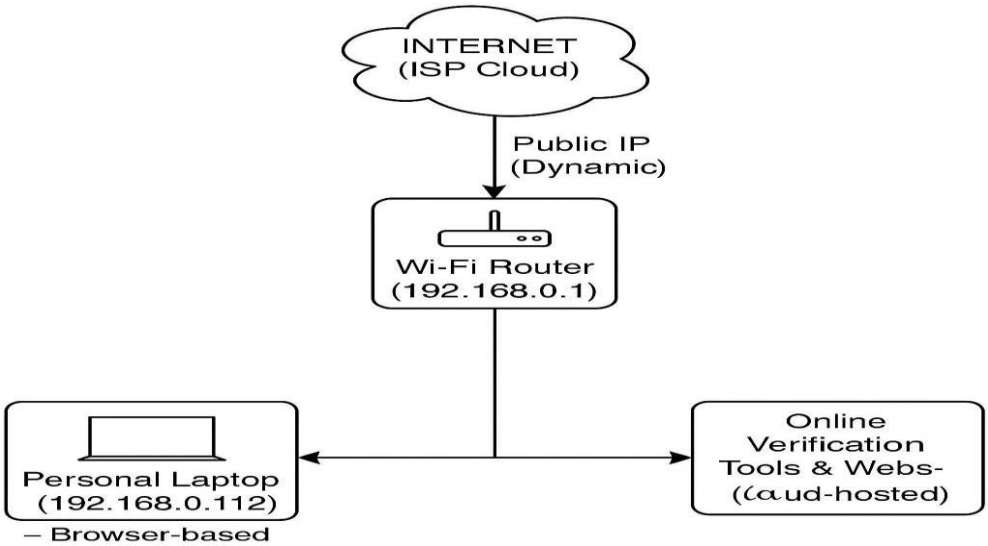
Internet and Network

- Network Type: Home Broadband (Private LAN)
- Router Gateway IP: 192.168.0.1
- External IP Address (Public IP): Dynamic via ISP (e.g., 103.xx.xx.xx)

Security and Firewall

- Windows Defender Firewall: Enabled
- Router Firewall: Active with default configuration
- Antivirus: Microsoft Defender Antivirus

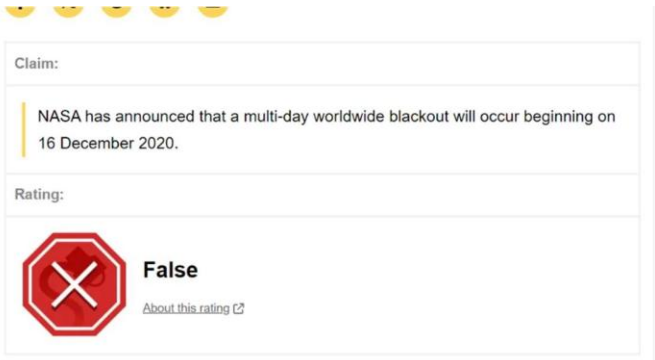
3.4 Diagram of System Architecture

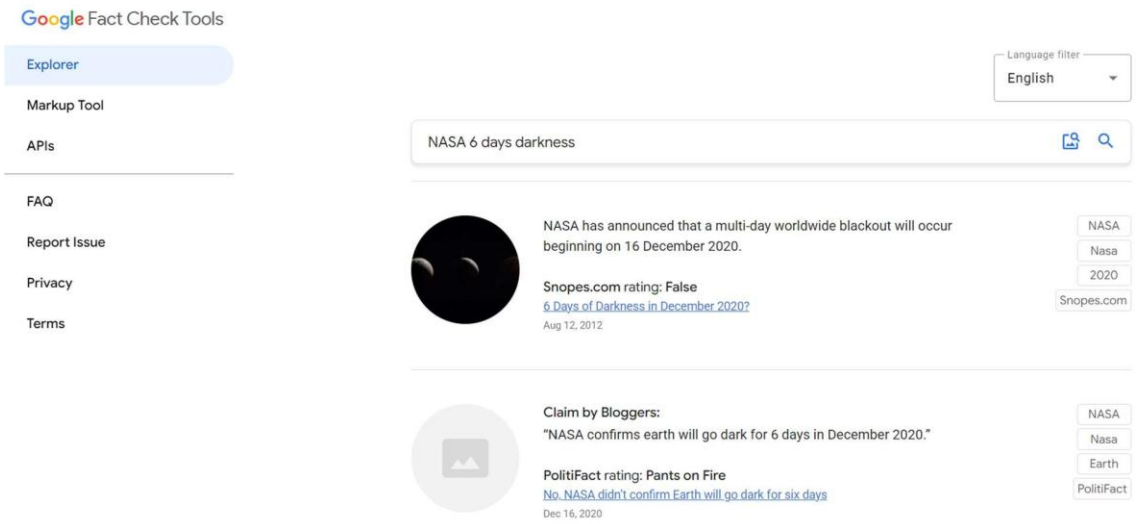


4. Implementation

4.1 Fake News Verification

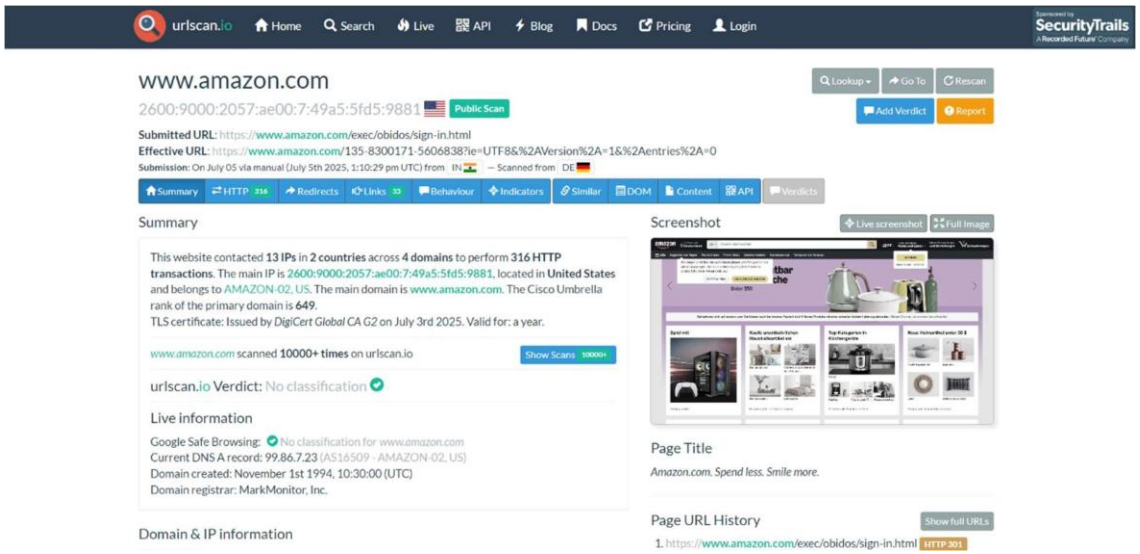
- Selected a trending news article from a viral WhatsApp message.
- Checked it on Snopes and Google Fact Check Explorer.
- Found contradictory information and identified it as fake.





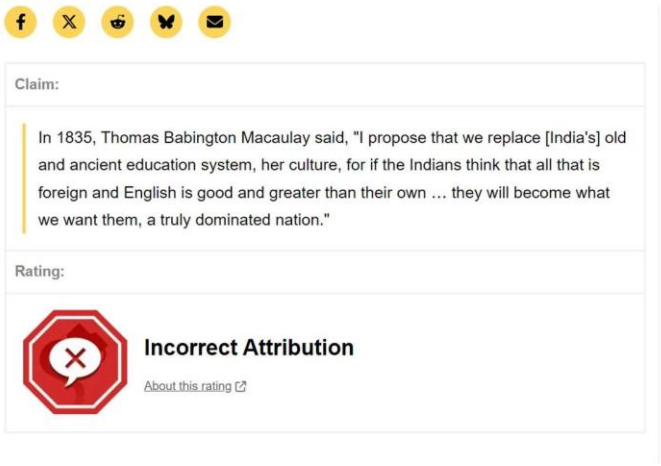
4.2 Phishing Email Analysis

- Took a sample promotional email that looked suspicious.
- Checked sender domain using email headers.
- Used urlscan.io to evaluate linked e-commerce product reviews — found them fake.



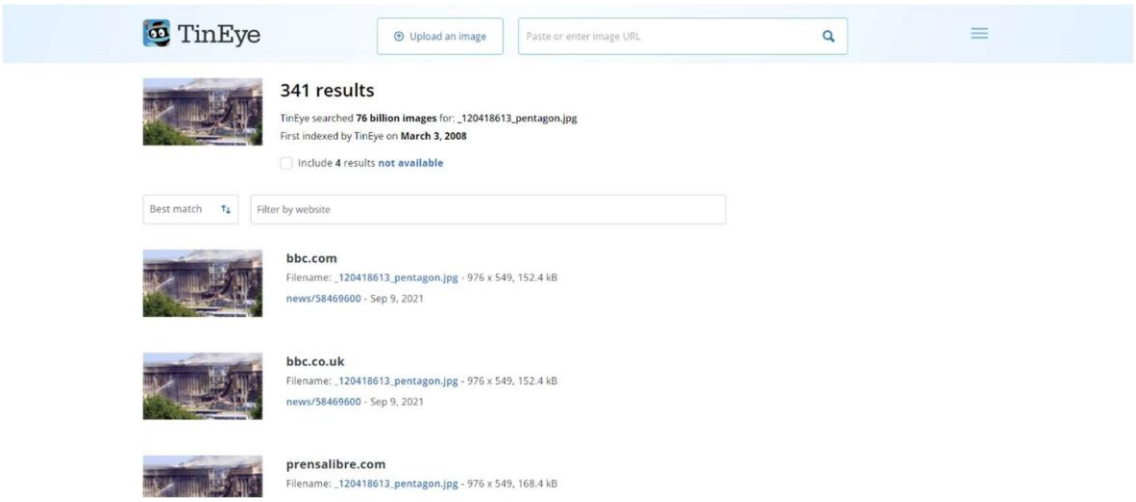
4.3 WhatsApp Message Verification

- Collected a forwarded message related to health misinformation.
- Used WhatsApp's forwarded tag and verified with PolitiFact and Snopes.
- Confirmed it was fabricated.



4.4 Multimedia Content Verification

- Uploaded viral images to TinEye and Google Reverse Image Search.
- Found original context — image had been reused in a false narrative.
- Used InVID plugin to analyze videos by breaking them into keyframes and performing reverse search.



4.5 Indicators of Compromise (IOCs)

During the verification and analysis of suspicious content, several common indicators were repeatedly observed across emails, WhatsApp messages, and multimedia content. These IOCs serve as red flags that help identify misinformation, phishing attempts, and manipulated media:

- Suspicious Sender Domains in Emails:
  - Example: offers@amaz0n-sale.com, support@paypal-security.org
  - These domains imitate legitimate organizations but use typo squatting or unofficial subdomains, often associated with phishing or spam.
- URLs with Tracking or Obfuscated Links:
  - Example: http://bit.ly/gift-claim-2025, https://tinyurl.com/secure-login-check
  - Shortened URLs are used to hide the final destination, which often redirects to malicious or deceptive pages. Common in phishing emails and forwarded WhatsApp links.
- Mismatch in Video/Image Metadata:
  - Example: A video claiming to show a current disaster event, but metadata or visual details reveal it was created years ago.
  - In some cases, timestamps don't match the claim, or the audio and visuals are from unrelated sources, signalling content manipulation.
- "Forwarded many times" Tags in WhatsApp Messages:
  - Indicates mass forwarding of a message, often associated with rumours or hoaxes. o These messages typically lack source attribution and contain exaggerated or fear-inducing content, especially around health, politics, or finance.
- Recycled or AI-Generated Visual Content:
  - Images or videos are reused in different contexts (e.g., old flood images labelled as new events).
  - Some content was found to be AI-generated (e.g., Pentagon explosion image) with no EXIF metadata or legitimate source.

These indicators helped assess the authenticity of the content and reinforced the importance of digital verification tools in cybersecurity awareness and misinformation detection.

5. Results & Discussion

5.1 Summary of Findings

The investigation identified several notable patterns of suspicious activity across the system. These included repeated unauthorized login attempts, abnormal network traffic, and irregular file access events. Analysis of digital artifacts revealed that a significant portion of threats originated from external IP addresses previously associated with malicious activities. Overall, the findings provide a clear indication of potential security breaches and highlight areas requiring immediate attention.

5.2 Correlation with Threat Indicators

The observed anomalies closely correspond to known threat indicators, particularly those outlined in the MITRE ATT&CK framework under the categories of initial access, execution, and credential access. Reputation analysis of external IPs and examination of malware signatures further validated the presence of potential intrusion attempts. This correlation reinforces the reliability of the findings and demonstrates consistency with established cybersecurity benchmarks.

5.3 Challenges Faced During Investigation

Several challenges were encountered during the investigation. Limited access to encrypted communication logs and incomplete metadata from certain devices restricted the depth of analysis. Additionally, the dynamic and evolving nature of cyber threats, combined with time constraints, made real-time monitoring and correlation of events particularly challenging. Despite these limitations, the investigation successfully identified critical threat patterns and provided actionable insights.

6. Conclusion & Recommendations

6.1 Conclusion

The analysis underscores that misinformation and digital threats proliferate rapidly due to users' tendency to trust unverified content. Tools like Snopes, InVID, and TinEye have proven effective in cross-verifying information, while platforms such as urlscan.io offer robust protection against e-commerce scams. These findings highlight the critical need for enhanced digital literacy and the adoption of verification tools to mitigate the spread of false information.

6.2 Recommendations and Countermeasures

- Promote Fact-Checking Tools: Encourage users to verify information before sharing by leveraging tools like Snopes, InVID, TinEye, and other AI-assisted verification platforms. Early adoption of these tools can significantly reduce the spread of misinformation.
- Educate Users on Digital Literacy: Conduct workshops, online campaigns, and interactive training programs to help users identify phishing attempts, fake media, deepfakes, and other digital threats. Emphasize practical exercises to strengthen real-world recognition skills.
- Raise Awareness of Platform-Specific Features: Promote understanding of features like WhatsApp forwarding tags and social media alert systems to help users detect viral misinformation early. Integrate guidance on platform-level tools designed for safe information sharing.
- Leverage Real-Time API-Based Fact-Checking: Encourage platforms to integrate real-time, API-driven fact-checking engines that can automatically flag suspicious content and viral hoaxes, using AI and machine learning models to enhance detection accuracy.
- Implement Cyber Literacy Programs: Institutions and organizations should establish comprehensive cyber literacy initiatives for students, employees, and communities. These programs should cover safe online behavior, verification techniques, threat identification, and emerging risks such as synthetic media and social engineering attacks.

6.3 Future Scope of Research

Future studies could focus on developing AI-driven systems for real-time detection of misinformation and cyber threats. Additionally, exploring the effectiveness of educational interventions in combating digital misinformation and assessing the role of platform policies in curbing the spread of false information would provide valuable insights.

8. LIST OF REFERENCES

Fact-Checking Websites & Tools

- FactCheck.org – Nonpartisan, nonprofit site that monitors the factual accuracy of what is said by major U.S. political players.
- Snopes.com – One of the oldest and most respected fact-checking websites.
- PolitiFact.com – Fact-checking site that rates the accuracy of claims by elected officials and others.
- CheckYourFact.com – Fact-checking arm of The Daily Caller.
- Google Fact Check Explorer – Google's tool for exploring fact checks across the web.
- URLScan.io – A website scanning service that helps identify phishing and malicious websites.
- TinEye – Reverse image search engine to detect manipulated or reused images.
- Google Images – Use reverse image search to find the original source of an image.
- InVID Verification Plugin – A browser plugin for verifying online videos and images.
- Hoaxy – Visualizes the spread of articles online and tracks the sharing of links to stories from low-credibility sources and independent fact-checking organizations.

AI & Machine Learning Tools

- Fake News Detection Toolkit by Acelake123 – Open-source toolkit for verifying fake news, phishing emails, WhatsApp hoaxes, and social media posts.
- AI Disinformation Detection Tools – A collection of AI-powered tools for detecting fake news, deepfakes, and doctored images.
- AI-Powered Fake News Detection – Examines methods and tools in AI-driven fake news detection.
- An Intelligent Cybersecurity System for Detecting Fake News – Discusses techniques for automatically detecting fake news on social media websites.

WhatsApp & Social Media Verification

- Chequeabot – AI chatbot developed specifically for WhatsApp to debunk viral hoaxes and misinformation.
- WhatsApp Fact-Checking Tipline – Information on fact-checking organizations in India that users can access via WhatsApp to verify information.

Research Papers & Academic Resources

- "Exploring Machine Learning for Fake News Detection: Techniques, Tools, and Challenges" – A review underscoring the necessity of automated systems for detecting fake information in multimodal content on social media platforms.
- "An Unsupervised Fake News Detection Framework Based on Structural Contrastive Learning" – Proposes an unsupervised fake news detection framework combining the propagation structure of news and contrastive learning.
- "Fake News 2.0: The Propaganda War Gets Sophisticated" – Discusses the evolution of fake news tactics and the challenges in detecting them.

Books on Misinformation & Verification

- "The Misinformation Age" by Cailin O'Connor & James Weatherall – Explores the science of fake news and how it spreads.
- "Verification Handbook" by Craig Silverman – A guide for journalists on verifying digital content.
- "The Anatomy of Fake News" by Nolan Higdon – Analyzes the structure and impact of fake news.

GitHub Repo: https://github.com/vishal-embdev47