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

(Domain: Digital Media Verification and Misinformation Detection)

TABLE OF CONTENTS

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

4.1. PROBLEM STATEMENT:

With the increasing spread of misinformation across digital platforms, identifying and verifying fake news, emails, WhatsApp messages, and social media posts has become a critical need. This project addresses the challenge of detecting such misinformation using various online tools and verification platforms.

4.2. LEARNING OBJECTIVE

This project aims to investigate various tools and websites designed to detect and combat the spread of misinformation across different online platforms. The project will involve researching and evaluating tools specifically tailored to identifying fake news articles, fake emails, fake WhatsApp messages, and fake social media posts. The selected tools will be tested for their effectiveness in detecting and verifying the authenticity of digital content across multiple platforms.

5. APPROACH:

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.

5.1. Tools and 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.

5.2. 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, 8GB RAM
- **Local IP Address**: 192.168.0.112

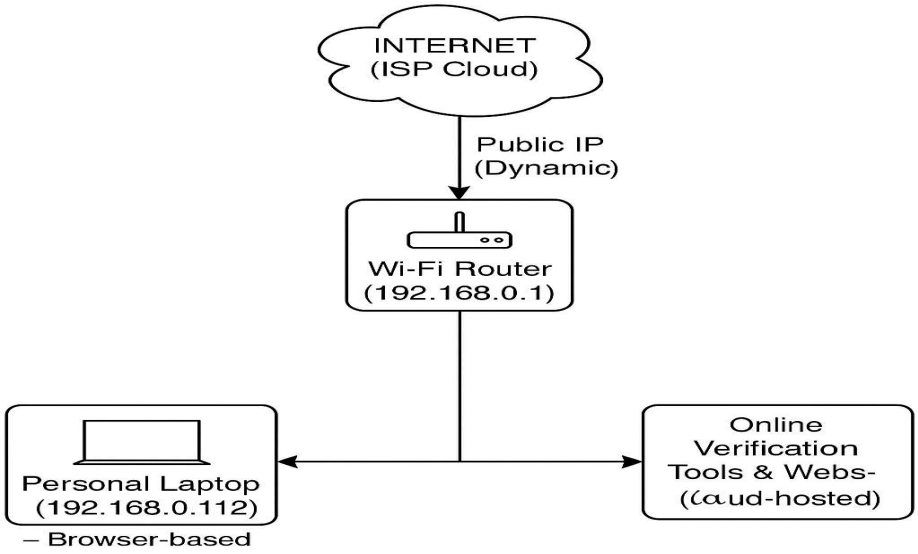
Internet and Network

- **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

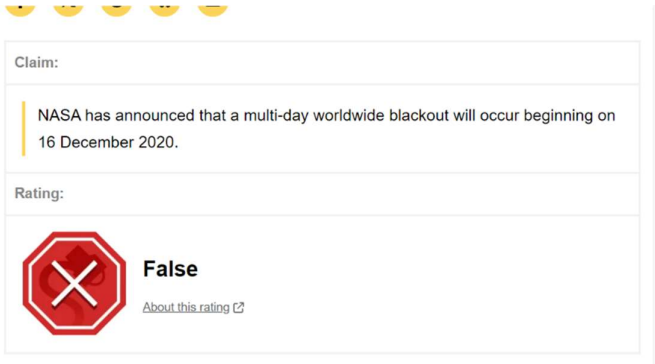
5.3. Diagram of System Architecture

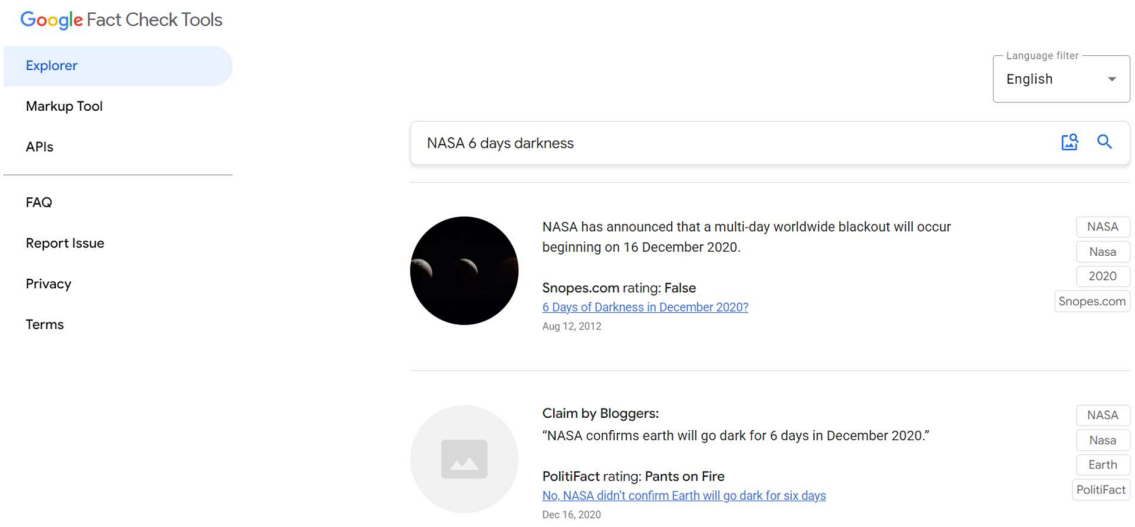


6. IMPLEMENTATION:

6.1. Verifying Fake News Headlines

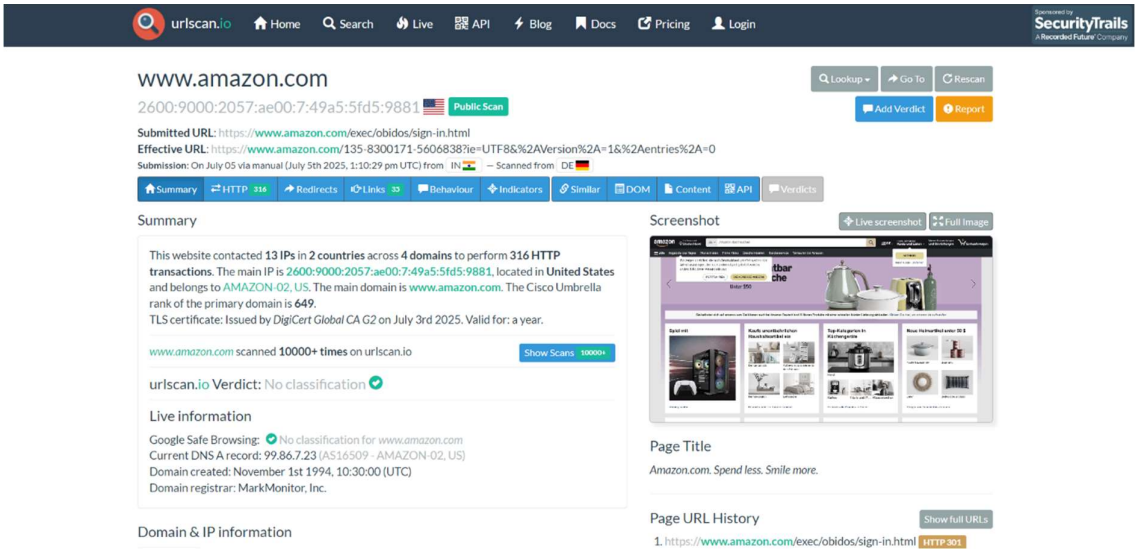
- Selected a trending news article from a viral WhatsApp message.
- Checked it on **Snopes** and **Google Fact Check Explorer**.
- Found contradictory information and identified it as fake.





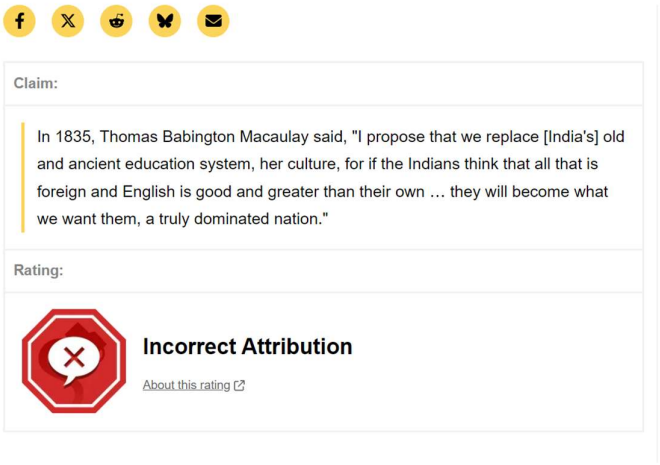
6.2. Analyzing Phishing/Scam Emails

- 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.



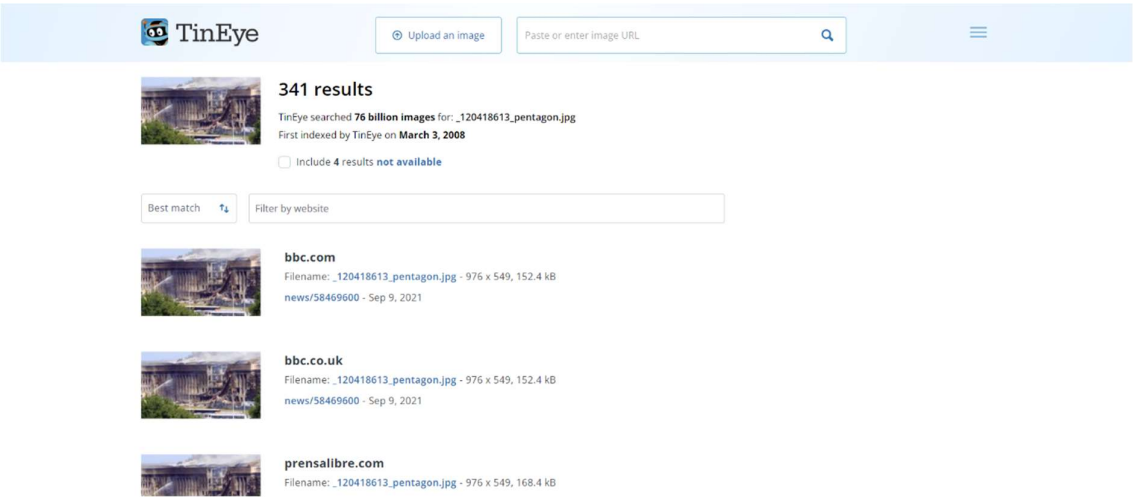
6.3. WhatsApp Message Verification

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



6.4. 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.



6.5. Indicators of Compromise (IOC) Observed:

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 typosquatting 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**, signaling content manipulation.
- **"Forwarded many times" Tags in WhatsApp Messages:**
    - Indicates mass forwarding of a message, often associated with rumors or hoaxes.
    - 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 labeled 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.

7. CONCLUSION & RECOMMENDATIONS:

Key Findings:

- Most misinformation spreads rapidly because users trust content without verification.
- Tools like Snopes, InVID, and TinEye were highly effective for cross-verification.
- Some tools like urlscan.io offer strong protection against e-commerce scams.

Recommendations & Countermeasures:

- Encourage the use of fact-checking tools before sharing any information.
- Educate users on recognizing phishing indicators and fake media.
- Promote awareness of WhatsApp forwarding tags and misinformation detection practices.
- Platforms should integrate real-time API-based fact-checking engines to flag viral hoaxes.
- Institutions should conduct cyber literacy programs for students and employees.

8. LIST OF REFERENCES:

- https://www.factcheck.org
- https://www.snopes.com
- https://www.politifact.com
- https://checkyourfact.com
- https://toolbox.google.com/factcheck/explorer
- https://www.urlscan.io
- https://www.tineye.com
- https://images.google.com
- https://www.invid-project.eu/tools-and-services/invid-verification-plugin
- **"The Misinformation Age"** by Cailin O'Connor & James Weatherall
- **"Verification Handbook"** by Craig Silverman
- **"The Anatomy of Fake News"** by Nolan Higdon