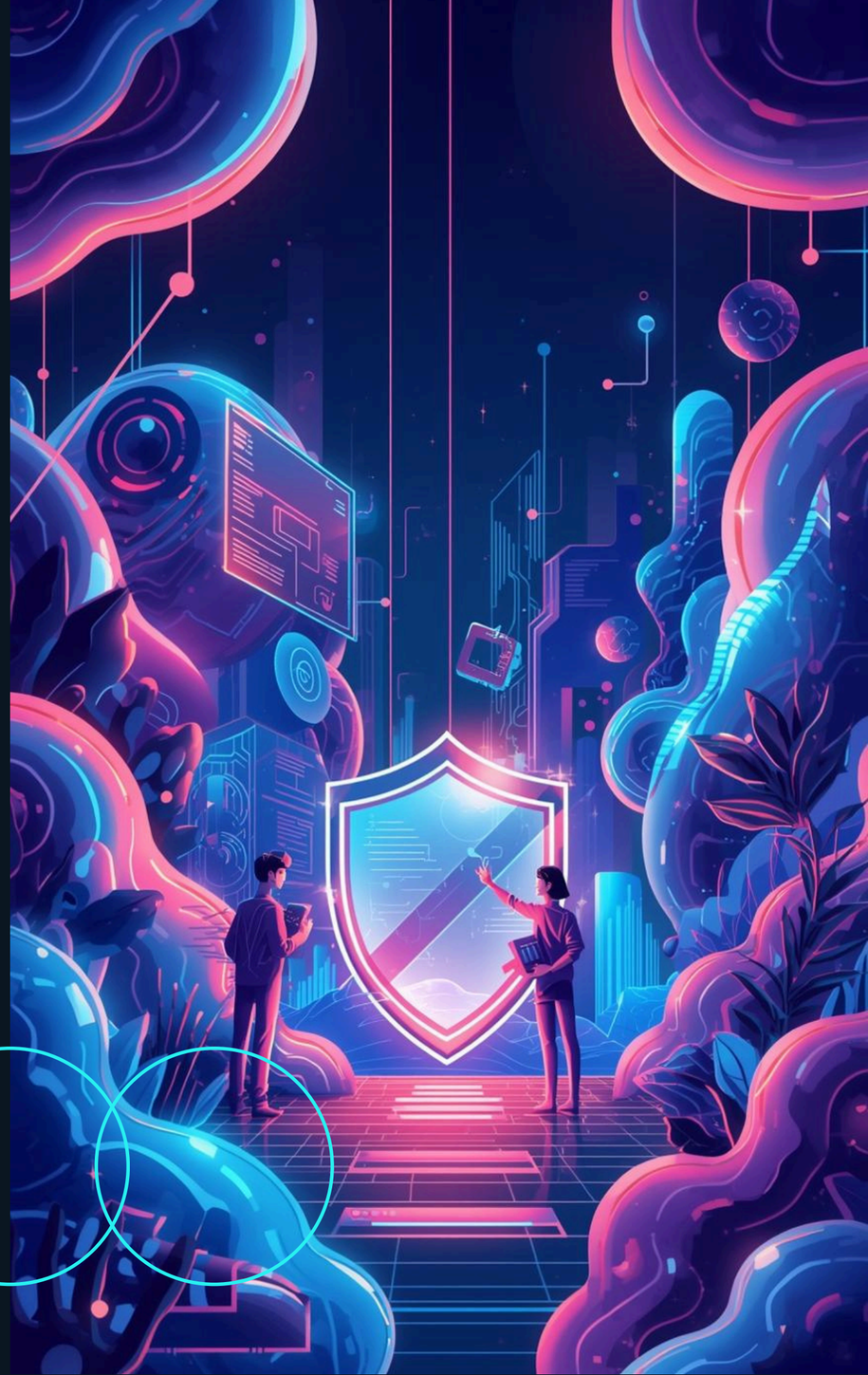
MISINFORMATION MITIGATION TECHNOLOGIES

Fake News & Misinformation Detection System

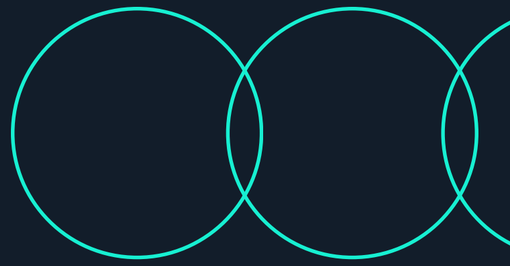
Presented by
Loshitha K,
Dhanasri R,
Devika M.



Emotional News Sharing

The Urgency of Verification

Users often share emotionally triggering news impulsively on social media, leading to widespread misinformation. This behavior underscores the need for effective detection and prevention systems to combat false narratives.



The Threat of Fake News

Understanding the Indian Context

Social Unrest

Fake news can incite **social unrest** in communities, leading to violence and conflict. When misinformation spreads, it can create panic among citizens and escalate tensions between different groups.

Financial Scams

Many scams are disguised as news, misleading users into sharing personal information or money. **Financial scams** thrive on emotional manipulation, making people vulnerable to malicious intentions and deceit.

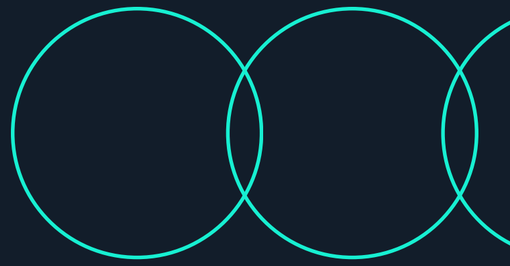
Medical Misinformation

The spread of **medical misinformation** can have dire consequences for public health. Unverified claims about treatments or vaccines can lead to harmful behaviors, putting individuals and communities at risk.

Key Insight: Real-Time Solution

No Current Fake News Blocking

While spam call blocking protects users from unwanted communication, **the absence of real-time fake news blocking** leaves individuals vulnerable to emotionally manipulative content, threatening informed decision-making and social stability.



Core AI Solution

Analyzing news for accuracy

Safe News

The AI system allows the **sharing of verified news** while blocking likely fake content. This dual approach encourages responsible sharing and ensures that users are informed of their decisions, maintaining a clean information ecosystem.

User Alerts

When fake news is detected, users receive a warning message such as "High Message Blocked – This content is likely misinformation." This feature serves as an immediate feedback mechanism, empowering users to think critically before sharing information further.

AI Detection Methods

Understanding how AI identifies fake news effectively

Language Patterns

The system analyzes word usage and syntax to detect potential misinformation, differentiating between credible sources and dubious content.

Emotional Triggers

By identifying emotionally charged language, the AI can flag content designed to provoke strong reactions, reducing impulsive sharing.

Credibility Signals

The AI evaluates the reliability of news sources and historical sharing patterns, allowing it to assign a confidence score for each article.

Emotional Manipulation

Detecting triggers that amplify misinformation risks

Fear Triggers

Fear triggers evoke **immediate reactions**, making individuals more likely to share without verifying the content's accuracy.

Anger Triggers

Anger triggers lead to impulsive sharing, as users tend to spread content that resonates with their emotional state.

Sympathy Triggers

Sympathy triggers engage a sense of compassion, often resulting in the uncritical sharing of emotionally charged stories without verification.

Source Analysis

Evaluating sharer behavior to improve misinformation detection

Origin Evaluation

This system assesses the source of information, helping to identify patterns in sharing behavior that indicate potential misinformation.

Offender Tracking

Repeated offenders are flagged for review, with actions taken to restrict or block their ability to spread misinformation further.

User Alerts

Users receive alerts about suspicious sources, empowering them to be more discerning in their sharing and consumption of news.

Fairness in Blocking

User Credibility

The system evaluates user credibility based on past verified sharing to ensure informed decision-making in blocking content.

Trust Labels

Trusted users receive caution labels to indicate potential misinformation without outright blocking their content, fostering responsible sharing.

Repeat Offenders

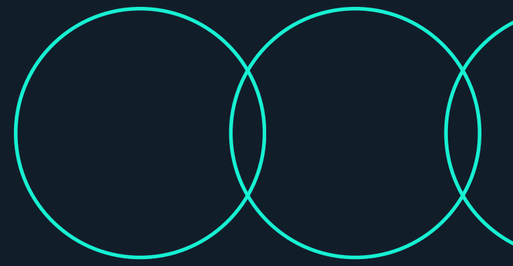
Users consistently flagged for sharing misinformation are restricted or blocked, minimizing the spread of harmful content within the community.

Accuracy Assurance

The system guarantees fairness and accuracy in enforcement, continually refining algorithms to improve decision-making and user experience.

Fake News Scenarios

Comparing blocked vs. shared news



Application Areas

Focus on platforms for combating misinformation effectively

Messaging Platforms

Implementing the AI-powered system in messaging platforms can **significantly reduce** the spread of fake news, allowing users to share content confidently without the fear of misinformation.

Social Media

By integrating the solution into social media networks, we can **proactively monitor** and block emotionally triggering misinformation, enhancing the overall credibility of shared information among users.

Social Impact

Benefits to Society

Misinformation Reduction

Implementing an AI-powered system significantly decreases the spread of misinformation, ensuring that users encounter factual content rather than emotionally charged falsehoods that can mislead and manipulate.

Emotional Protection

The system safeguards individuals from emotional manipulation by identifying emotionally charged news, preventing the sharing of content that could cause unnecessary distress or provoke irrational behavior among users.

Support for Fact-Checkers

By minimizing the circulation of false information, the system empowers fact-checkers, allowing them to focus on verifying genuine content and enhancing public trust in accurate information sources.

Future Enhancements

Multilingual Support

Offering support in Tamil, Hindi, and Kannada will enhance accessibility and usability for diverse Indian audiences, promoting wider adoption.

Voice & Video Detection

Implementing detection for voice and video misinformation will address the growing threats posed by multimedia content in social media networks.

Deepfake Analysis

Advanced deepfake detection capabilities will help identify manipulated media, ensuring users are aware of potential misinformation in visual content.

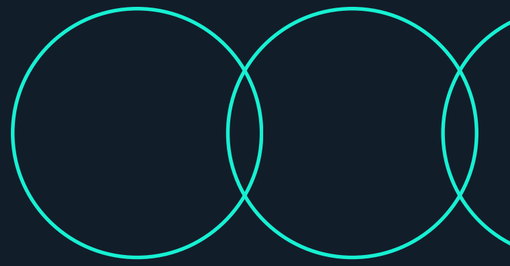
Real-time Monitoring

Integrating real-time rumor monitoring will enable swift identification and action against emerging misinformation trends, fostering a more informed public.

The Speed of Emotion

Truth vs. Emotional Appeal

Misinformation spreads rapidly because it triggers **emotional responses** more effectively than facts. Users often share emotionally charged content impulsively, prioritizing emotional engagement over logical reasoning, leading to widespread misinformation.



Pause and Verify

Encouraging Thoughtful Sharing

This AI solution empowers users to critically assess news items, cultivating a culture of verification. By prompting reflection before sharing, it helps combat misinformation and promotes informed engagement.

Innovation Through Observation

Meaningful Technology Development

By closely analyzing real-world challenges, we can develop **impactful technological solutions** that address pressing issues, ensuring our innovations genuinely resonate with societal needs and foster positive change.