

Detection of Fake News on Online Social Networking



Outline

- I. Problem Overview
- II. Current Solutions
- III. Limitations of Current Solutions
- IV. Conclusion
- V. Our Solution



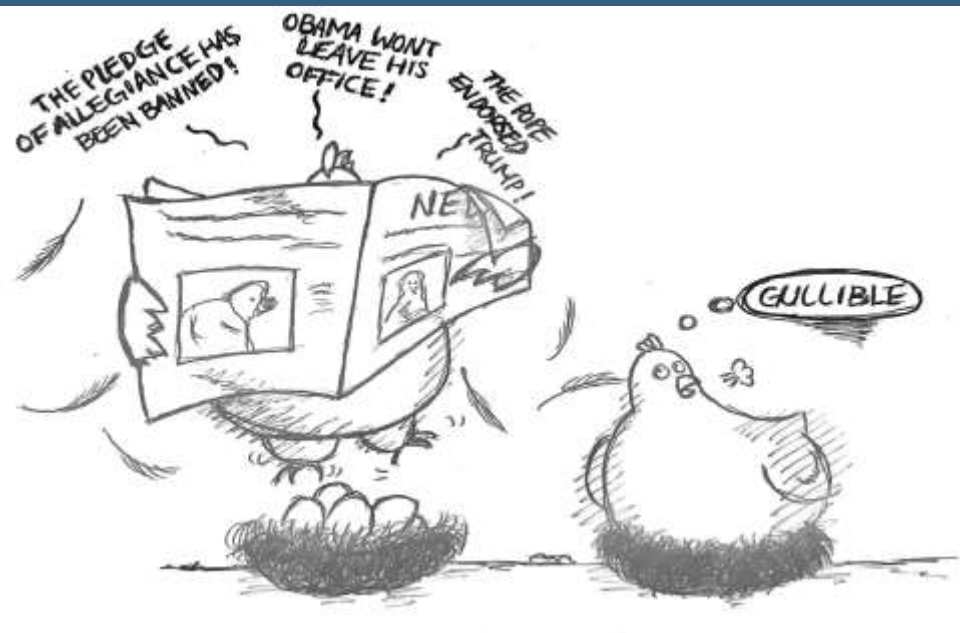
I. Problem Overview



What is Fake News?

Misinformation: False or incorrect information

Purpose: Affect the perception of people



Problem Overview:

- The large use of Online Social Networking has provided fertile soil for the emergence and fast spread of rumors.
- It is difficult to determine all of the messages or posts on social media are truthful.
- Fake news harms to real life.



Fake news – a definition

- Fake news [...] is where individuals or organisations intentionally publish hoaxes, propaganda and other misinformation and present it as factual.
- This can include blog and social media posts and fake online media releases.
- It does not include news satire sites as they are not presenting their content as legitimate factual news. Their intention is satire rather than misinformation.

It also does not include articles that are written from the perspective of a particular opinion or editorial standpoint, provided the information included is factually correct.



Intention?

- The intentionality of deception is also a requirement in Rubin et al.'s (2015) definition
- Whose intentionality?
 - *Creator of the news?*
- E.g. Governments (→ WMD example)
- The press
 - *Purveyor of the news?*
- The press
- Social media → you!
- How to capture the intention as a DM/ML feature?



Sweden signed
the deal to become
a member of
NATO??



Defend Misinformation!



PEPSI STOCK TANKS!
AFTER CEO ATTACKS TRUMP SUPPORTERS

PepsiCo CEO Indra
Nooyi told Trump
fans to "take their
business elsewhere"

How to detect misinformation?

- a. Human Intervention
by verifying information veracity
- b. Using Algorithm
by identifying fake content and validating the
information sources



II. Current Solutions



Current Solutions:

1)Linguistic approaches

- Data Representation
- Deep Syntax
- Semantic Analysis
- Rhetorical Structure and Discourse Analysis
- Classifiers



2) Network Approaches:

❑ Linked Data

- Fact-checking methods
- Leverages an existing body of collective human knowledge
- Query existing knowledge network, or publicly available structured data



a Barack Obama



U.S. President Barack Obama in front of the Resolute desk in the Oval Office of the White House, December 6, 2012

44th President of the United States

Incumbent

Assumed office
January 20, 2009

Vice President Joe Biden

Preceded by George W. Bush

United States Senator from Illinois

In office
January 3, 2005 – November 16, 2008

Preceded by Peter Fitzgerald

Succeeded by Roland Burris

Member of the Illinois Senate from the 13th District

In office
January 8, 1997 – November 4, 2004

Preceded by Alice Palmer

Succeeded by Kwame Raoul

Personal details

Born Barack Hussein Obama II
August 4, 1961 (age 52)
Honolulu, Hawaii, U.S.

Nationality American

Political party Democratic



How to recognise fake news

- **Read past the headline** – Many fake news creators don't put much effort into content beyond the headline as that is all many people will read before they share it.
- **Check who originally published it** – See who the original source of the story is. If you don't recognise it use a search engine to find out more about it.
- **Follow links and check sources** – Just because the article claims that something is a quote or a particular site verifies information doesn't mean they do.
- **Beware confirmation bias** – We are more likely to believe something that reinforces our pre-conceived ideas. Don't assume something is true because you want it to be.
- **Check other news outlets** – Is anyone else reporting this? If you can't find a major story anywhere else it probably isn't true.
- **Think before you share** – Fake news relies on many people sharing unverified information to get traction so if you aren't sure if something is real news or not, don't share it.



Fact checking

The only way to be sure that news is accurate is to check whether the facts they give are correct. Although this would be a difficult task for a single person, there are plenty of groups dedicated to checking this.

Snopes is a fact-checking website that looks at the legitimacy of online and offline news.

- www.snopes.com/

How to recognise fake news

→ (Human) strategies

- **Read past the headline** – Many fake news creators don't put much effort into content beyond the headline as that is all many people will read before they share it.
- **Check who originally published it** – See who the original source of the story is. If you don't recognise it use a search engine to find out more about it.
- **Follow links and check sources** – Just because the article claims that something is a quote or a particular site verifies information doesn't mean they do.
- **Beware confirmation bias** – We are more likely to believe something that reinforces our pre-conceived ideas. Don't assume something is true because you want it to be.
- **Check other news outlets** – Is anyone else reporting this? If you can't find a major story anywhere else it probably isn't true.
- **Think before you share** – Fake news relies on many people sharing unverified information to get traction so if you aren't sure if something is real news or not, don't share it.



Human strategies translate to various machine tasks



Strategy “Read past the headline”

- The goal of the **Fake News Detection** is to explore how artificial intelligence technologies, particularly machine learning and natural language processing, might be leveraged to combat the fake news problem. We believe that these AI technologies hold promise for significantly automating parts of the procedure human fact checkers use today to determine if a story is real or a hoax.
- Assessing the veracity of a news story is a complex and cumbersome task, even for trained experts ³. Fortunately, the process can be broken down into steps or stages. A helpful first step towards identifying fake news is to understand what other news organizations are saying about the topic. We believe automating this process, called **Stance Detection**, could serve as a useful building block in an AI-assisted fact-checking pipeline. So stage #1 of the **Fake News Challenge (FNC-1)** focuses on the task of Stance Detection.
- Stance Detection involves estimating the relative perspective (or stance) of two pieces of text relative to a topic, claim or issue. The version of Stance Detection we have selected for FNC-1 extends the work of Ferreira & Vlachos ⁴. For FNC-1 we have chosen the task of estimating the stance of a body text from a news article relative to a headline. Specifically, the body text may agree, disagree, discuss or be unrelated to the headline.



Task “stance detection”

FORMAL DEFINITION

Input

A headline and a body text - either from the same news article or from two different articles.

Output

Classify the stance of the body text relative to the claim made in the headline into one of four categories:

1. **Agrees:** The body text agrees with the headline.
2. **Disagrees:** The body text disagrees with the headline.
3. **Discusses:** The body text discuss the same topic as the headline, but does not take a position
4. **Unrelated:** The body text discusses a different topic than the headline

Stance detection n - Example



EXAMPLE HEADLINE

"Robert Plant Ripped up \$800M Led Zeppelin Reunion Contract"

EXAMPLE SNIPPETS FROM BODY TEXTS AND CORRECT CLASSIFICATIONS

"... Led Zeppelin's Robert Plant turned down £500 MILLION to reform supergroup. ..."

CORRECT CLASSIFICATION: AGREE

"... No, Robert Plant did not rip up an \$800 million deal to get Led Zeppelin back together. ..."

CORRECT CLASSIFICATION: DISAGREE

"... Robert Plant reportedly tore up an \$800 million Led Zeppelin reunion deal. ..."

CORRECT CLASSIFICATION: DISCUSSES

"... Richard Branson's Virgin Galactic is set to launch SpaceShipTwo today. ..."

CORRECT CLASSIFICATION: UNRELATED

<http://www.fakenewschallenge.org/>

Strategies “Follow links and check sources” and “Check other news outlets”

Task: Claim validation

Automated Fact Checking (Claim Validation)

Claim: Israel caused flooding in Gaza by opening river dams

Evidence documents:

The Daily Mail:

...Hundreds of Palestinians left homeless after Israel opens river dams and floods houses ...

Al Jazeera:

"Israel opened water dams, without warning, last night, causing serious damage to Gazan villages near the border, " General Al-Saudi told Al Jazeera. ...

The Jerusalem post:

...The Daily Mail published a story on Monday that originally accused Israel of intentionally opening dams in order to flood Gaza.

The only problem is, there are no dams in southern Israel. ... after amending the article's headline the pretense for the charge remained in the Daily Mail article...

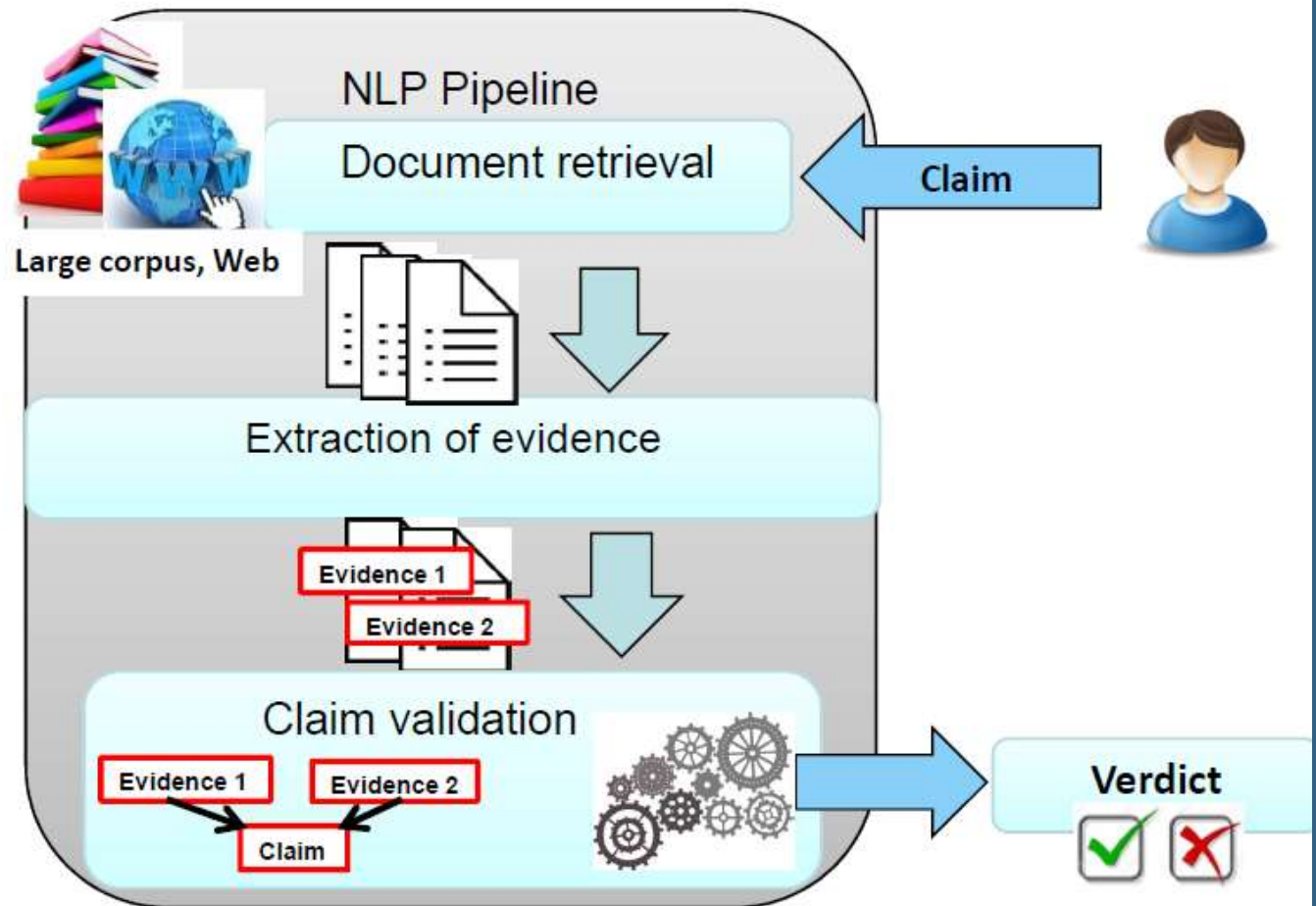
→ **Claim is false**



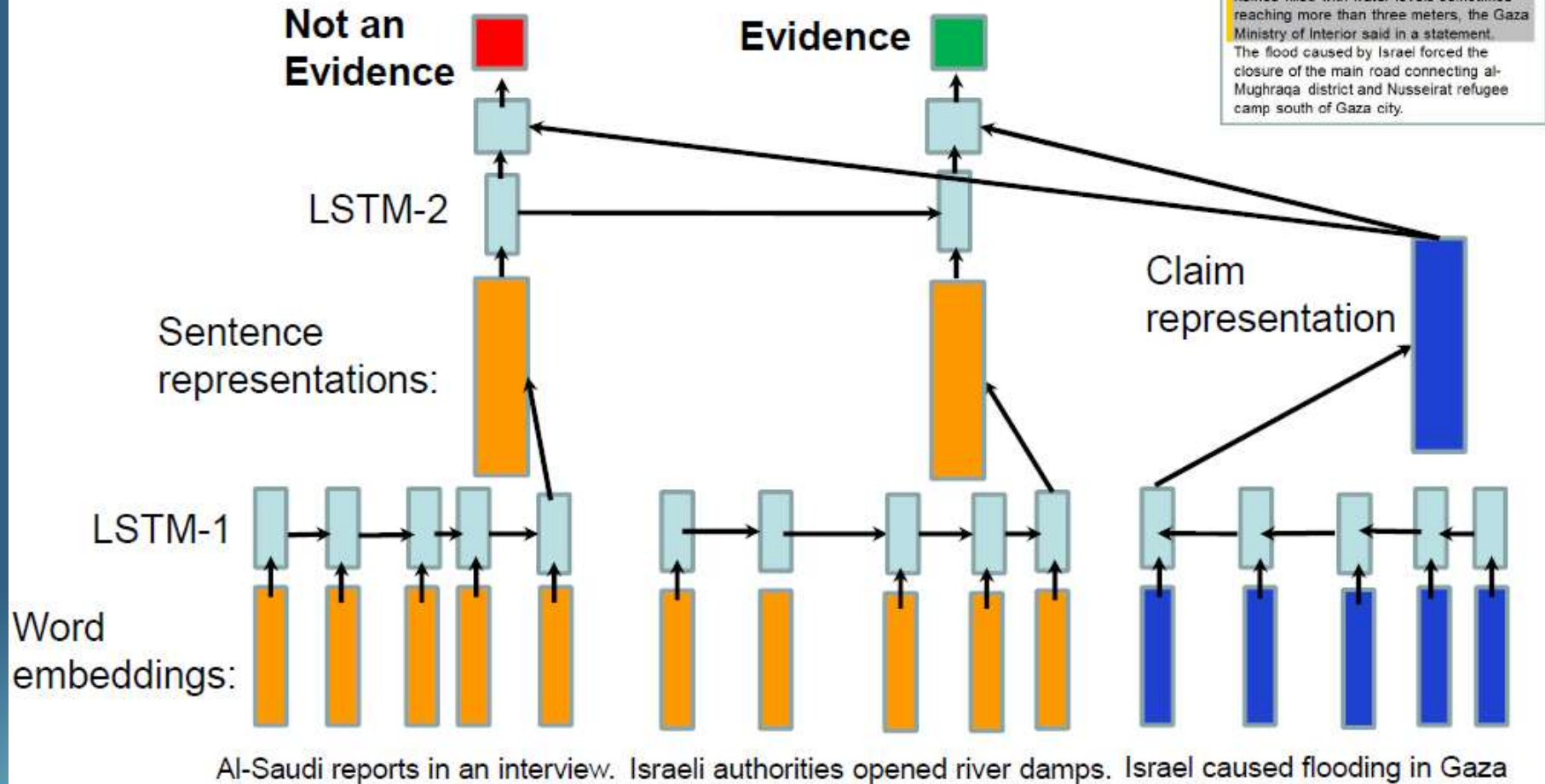
Automated Fact Checking (Claim Validation): NLP-Pipeline



TECHNISCHE
UNIVERSITÄT
DARMSTADT



Automated Fact Checking (Claim Validation): Extraction of Evidence using LSTMs



III. Limitations of Current Solutions



Time Sensitivity

- Quality vs Quickness
- Operate in a retrospective manner
- Results in the delay between the publication and detection of a rumor
- Latency aware rumor detection



- Clustering data by keywords using an ensemble method that combine user, propagation and content-based features could be effective.
- Computation of those features is efficient, but needs repeated responses by other users.
- Results in increased latency between publication and detection.



Accuracy

- Current studies focus on improving accuracy, but the accuracy of current techniques is still below 70%.
- Ambiguity in the language
- Evolving usage of Language:
e.g. Emoticons, Symbols
- Difficulty in classification



Other Drawbacks

- Most models are specific to some networks
- Identification of only small percentage of fake data
- Need more features



Technical Limitations

- Mainly concentrate on two specific technical problems
 1. How can we detect the signal of misinformation early?
 2. How can we improve the accuracy?



- Linguistic and network-based approaches have shown relatively high accuracy results in classification tasks within limited domains.
- Previous studies provide a basic topology of methods available
- New tool - Refine, Evolve and Design
- Hybrid System - Techniques arising from disparate approaches may be utilized together.

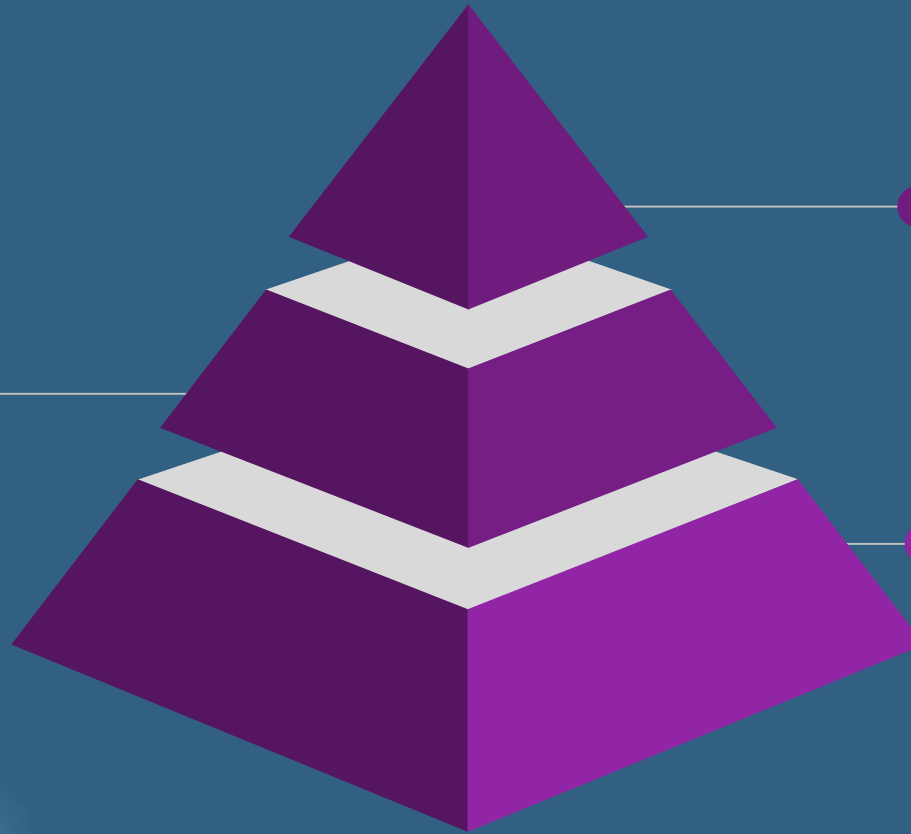


v. Our Solution



**Police Database:
User
Reports,Hashtag
Monitor,Tracing
Old tweets**

2



1

**User Level:
Plugin
Whatsapp Bot
Website**

3

**ML Model :
Text Analysis,
Fake News
Analysis,
Image and
Video Analysis**





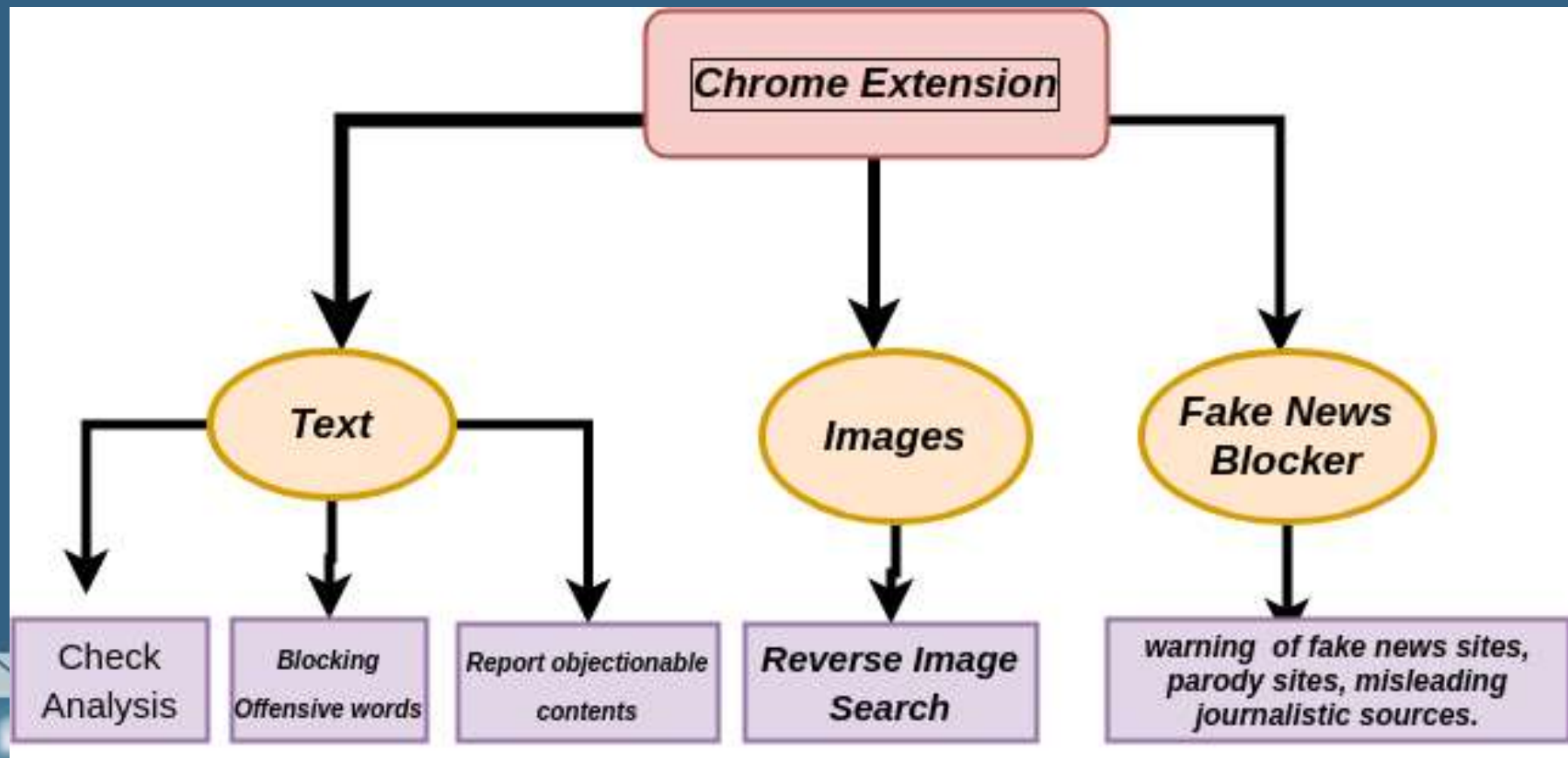
Chrome Extension

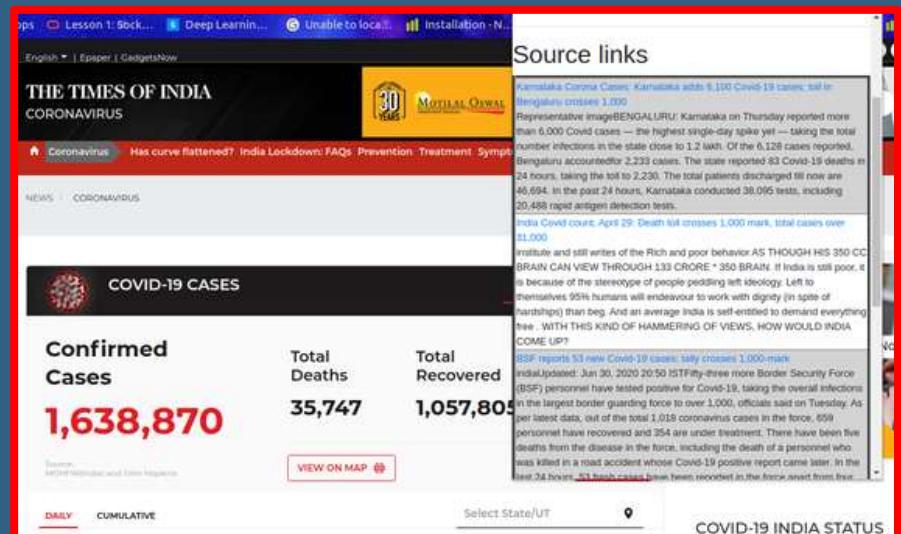
- Easy-to-use system
- Real-time analysis of the content in user's feed all website or social media
- fact checking of input links and reporting mechanism.

Website

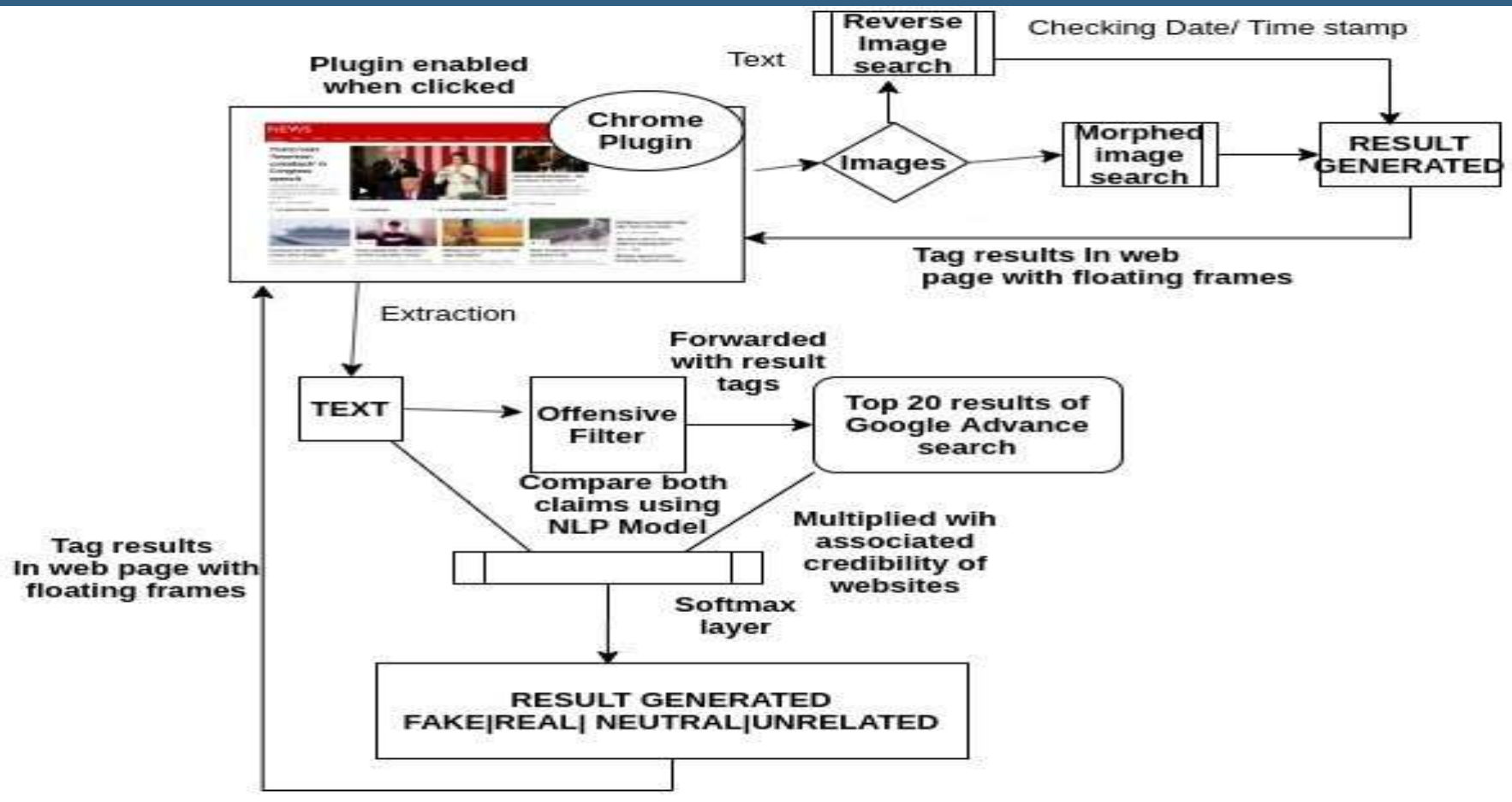
Whatsapp Bot







Proposed System



- To manage massive increase in demand for information during a time of crisis
- Inefficient and delayed information dissemination
- Despite efforts, misinformation about the virus has spread rapidly on various platforms.
- Spread of Rumors / Fake-News in Social Media Sites which could create panic among users





**Fake News
Checker
and Analysis
Results**

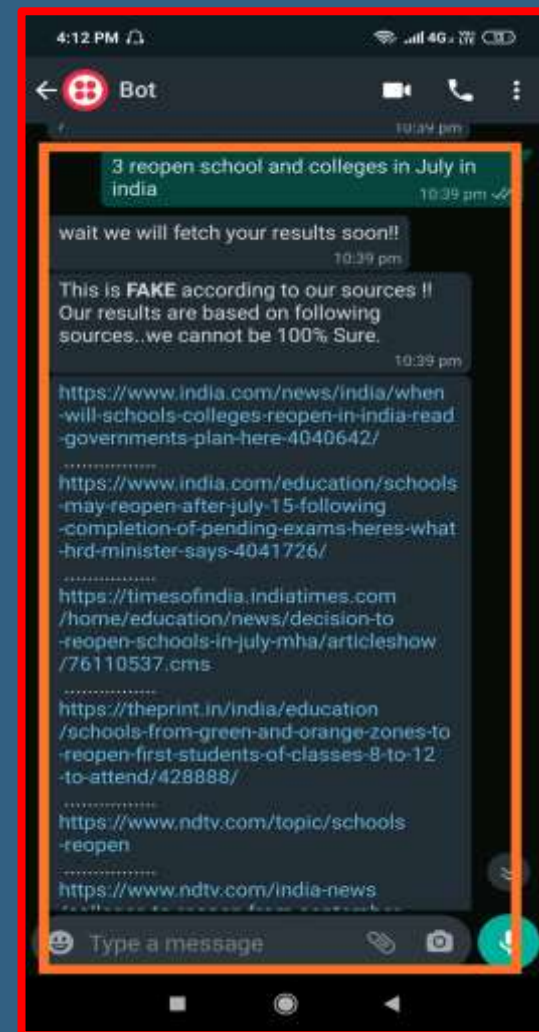


**List of viral
fake
news/rumours**



**Reporting
Mechanism
and news list
available**





Technology Stack

- **Website :** React.js , Java Script, JSON, HTML,CSS,BOOTSTRAP,IBM Cloud, APIs.
- **Plugin :** Javascript,HTML,CSS
- **REST API :** All Machine Learning Models are deployed as an API, FLASK
- **Back-end :** Hosted on web:gunicorn
- **Server :** Amazon EC2 instance
- **WharsApp Bot:** Python,APIs,DialogFlow,Flask,Heroku



Keras & Pytorch framework for Deep neural net implementation.

- Current state-of-the art model used:

- Optimizer Used: AdamW
- Loss Function used: Categorical Cross Entropy
- Batch Size : 8
- Transformers Model which we used for BERT embeddings:
- Bert ForSequenceClassification
- Epochs: 8
- Learning rate: $2e-5$
- Dropout: 0.3
- Results: F1 Score: 81.02

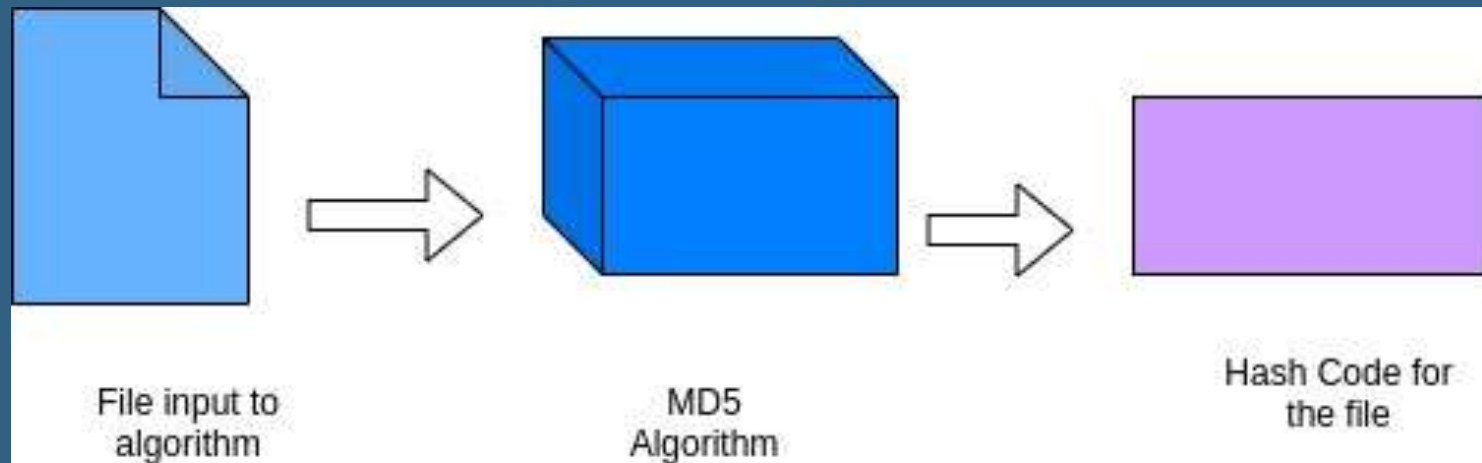


File Identification

- MD5 (Message Digest 5) sums can be used as a checksum to verify files or strings .
- MD5 Sums are 128-bit character strings (numerals and letters) resulting from running the MD5 algorithm against a specific file.
- Each file has a unique “Hash Value” .
- When we changes the file , the “Hash Value “ also changes
- So , copy of a file will have a “Hash Value” equal to the “Hash Value” of original file.
- In this way , we can know if the file is unique or not.



Flow chart for MD5 Algorithm :



Thank you!



Acknowledgement: For Design, Layout of slide
Sunghun Park, Venkat Kotha, Li Wang Wenzhi Cai