



Advanced Fake News Detection System

Website & Chrome Extension to Detect and Combat
Fake News on Multiple Social Media Platforms

Domains: Web Development & Machine Learning

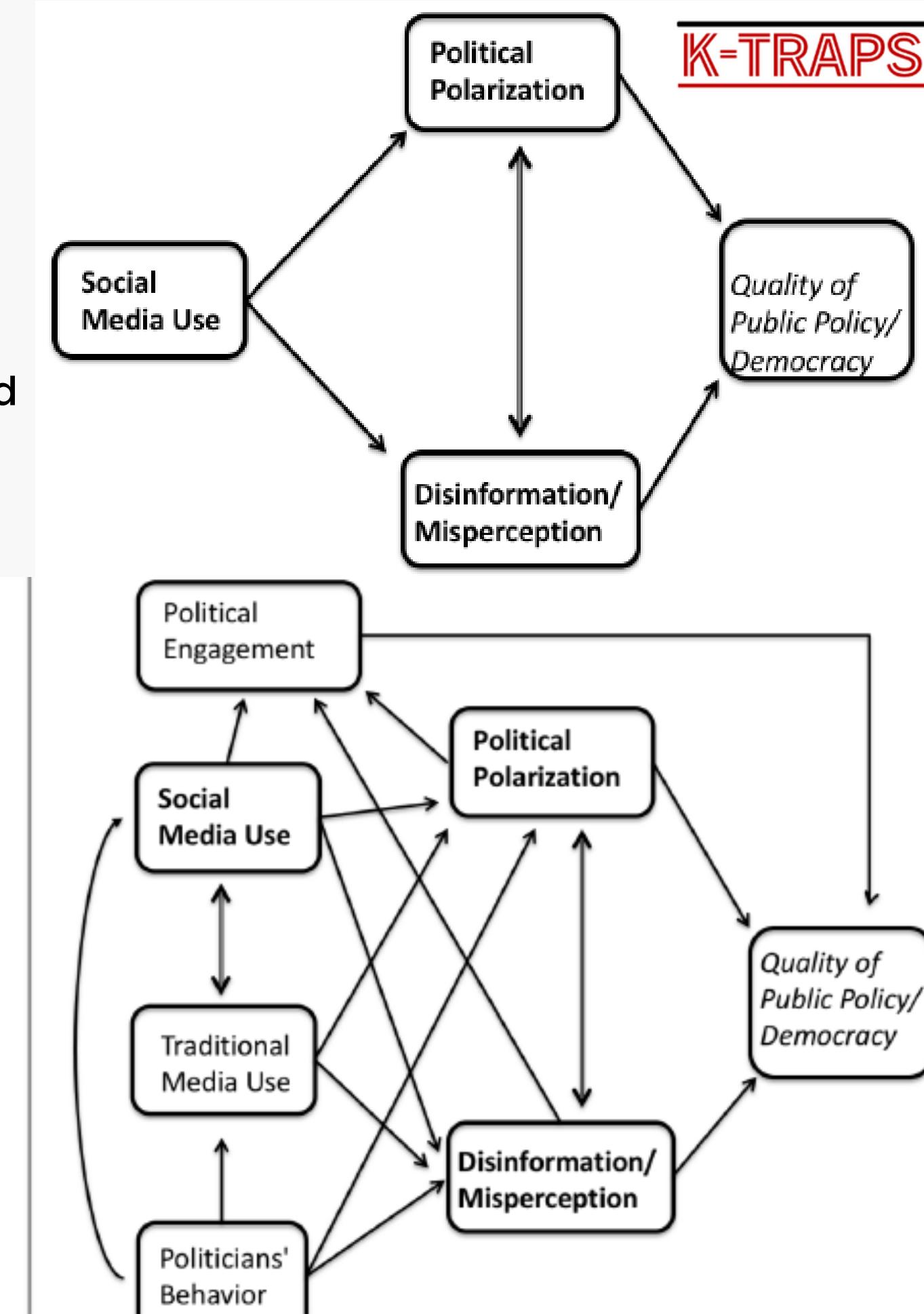
Tech Stack: Tailwind CSS, Material UI , Next JS, Node JS, Flask, Express, Pytorch, Pytorch Geometric, NetworkX

k-traps

FAKE NEWS AS A CYBERSECURITY THREAT

- Fake news perception and impact can be measured on a psychological level with real implications in the real world
Examples: #pizzagate, the Jan 6th riots, Cambridge Analytica Scandal
- Misinformation impacts individuals, markets, economies, enterprises and governments. This is increasingly creating violence, law and order situations and incitement
- Modern Warfare must include misinformation and data collection as a field as they can be used maliciously against a person's psychology. Fake news being a primary source of information reinforces people's personal belief and contributes to further division of ideals and values. This has real life implications and is threat to democracy
- A study in Science Advances found that false news stories are 70% more likely to be retweeted on Twitter than true stories. A survey by the Pew Research Center found that 64% of U.S. adults say that fake news has caused a great deal of confusion about the basic facts of current events
- According to a report by the Global Disinformation Index, fake news websites make an estimated \$235 million per year from advertisements

Impact of Fake News on Political Polarization and Democracy



Problem Statement:

Design and develop a technological solution/software tool for Tracking & Tracing Fake News and its origin using official sources as the input filter. The solution should have a mechanism to mitigate the impact of the spread of Fake News by auto-populating the fake news spreaders' inboxes with the official/authenticated news content.

WE HAVE CONSIDERED THE FAKE NEWS DETECTION SYSTEM AS A TWO PART PROBLEM BASED ON THE PROBLEM STATEMENT-OFFENCE AND DEFENCE

OFFENCE:

- Design and develop a technological solution/software tool for Tracking & Tracing Fake News and its origin using official sources as the input filter.

DEFENCE:

- The solution should have a mechanism to mitigate the impact of the spread of Fake News by auto-populating the fake news spreaders' inboxes with the official/authenticated news content.

PROBLEMS WITH EXISTING SYSTEMS

- The speed at which fake news spreads: Fake news can spread rapidly through social media, misinformation is spread six times faster than any factual news on social media
- The use of advanced technology: Some fake news creators use sophisticated tools and techniques to create convincing fake news, making it difficult to detect using traditional methods.
- The use of opinion-based content: Some fake news is designed to appeal to emotions and opinions, rather than being fact-based. This makes it harder to detect using traditional fact-checking methods.
- Exploitation of user: Large multinational corporations or the ‘Big-Techs’ mould the truth or real news for monetary gains based on mass-data collection and psychological factors
- Lack of measures to combat the impact of fake news in the cyberspace: Currently no such measure exists to combat prevailing fake news in cycle and its impact on psychology of people

Features of k-traps System

K-TRAPS

- State-of-the-Art fake news detection system having timely updated information, with great accuracy and information retrieval system
- Advanced data analytics and insights:

With the help of natural language processing models we performed zero-shot-classification on the input URL/text of the user to get analysis related to category, clickbait probability, keywords, sentiment analysis, political views, etc. of a particular post/text/URL

- Community Detection:

With the help of user activity and analysis of posts we are able to generate social network graphs that links activity of users based on type of opinions on particular topics and their views. This helps to identify communities that may be harmful to a particular user and prevent political polarization

- Pre-emptive strike of real news:

Based on community detection we have identified certain communities that have a probability for fake news generation and spread and also susceptibility to it. This will help us strike them by spreading the real news to the users with most activity. Ours being the primary source of information will help people second guess the fake news and be less susceptible to it

- Recommender system:

Based on community detection and echo chamber analysis we will prevent extreme views being recommended by social media user feeds. This will prevent political polarization of the people and help them have their own opinions based on all the view-points recommended to them by us rather than just the extreme views.

OFFENCE

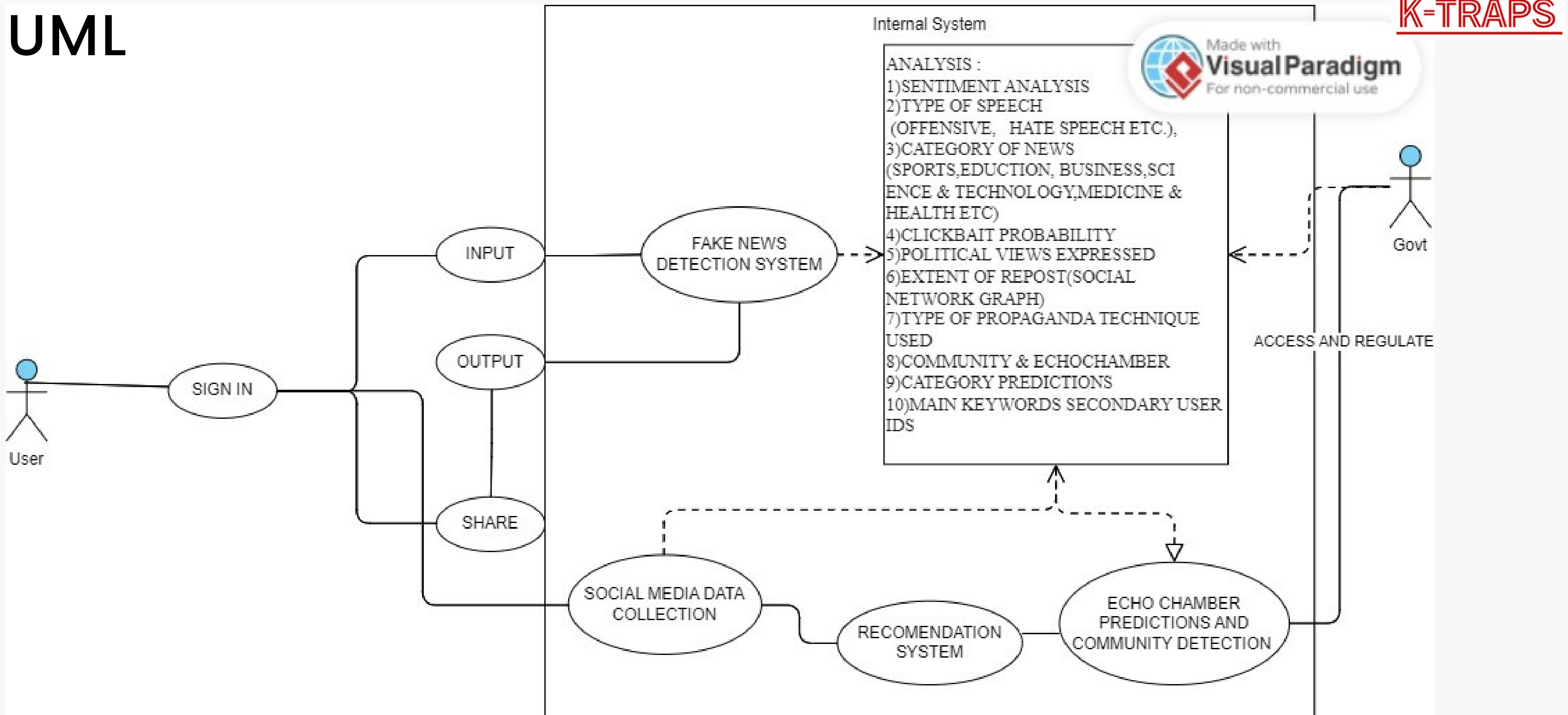
K-TRAPS

- Checking the authenticity of a news article and backing it up with official government statistics, reports and articles
- Filtering out and tracking Fake News and Misinformation on the multiple Social Media Platforms
- Maintaining the access to fake news to extremists and limiting the amount of hate-speech content on said social media platforms
- Using Machine Learning Algorithms to perform zero-shot classification on information provided by the user.
- Data analysis and inference is done on category of text, hate-speech detection, clickbait analysis, activity of repost etc.
- This is done to know the person's inference of the news and perform an impact analysis of his/her retweeting of said information

DEFENCE

- Creating community detection networks to predict which communities are likely to generate fake news and which are susceptible to it, and preventing the spread of misinformation by spreading the correct news to them first.
- Ours being a primary source of information prevents the user to be confused about information and making them less susceptible to misinformation and highly polarized news
- Based on user information and analysis of the posts recommended by the particular social media platform to the user, we tweak the type of posts consumed and exposed to the user
- If the post and content in it is determined as fake we filter it out. For the rest of the posts tweaking is done in terms of fake news, clickbait probability, extreme-political beliefs, hate-speech etc.
- This is done to prevent political polarization and spread of misinformation and will thus curb amount of hate-speech

UML



can search information and share results

can access analysis based on user activity and social media activity and regulate type and flow of information by recommender system



How our Website Works

User-Side:

- A web-based user interface can be used for verifying or validating any piece of text or content of an article published on the internet.
- The user can access our website by signing in through Google or any Social Media Platform account. On the website, the user will be provided an input field where any Text Query or URL of a sample article can be placed. Upon populating the input field, a response is sent.

The response contains :

- Authenticity / Validity of the content of sample news article or text query i.e fake or not along with information about the publisher of the article if available.
- Authentic information extracted from various genuine sources about the topic searched by the user.
- Links or references to genuine resources that have published authentic information about the topic that is being queried by the user.

User Interface

K-TRAPS

Response

The screenshot shows a web application titled "News Analyser". At the top, there is a navigation bar with links for "Home", "Users", and "Sign In →". Below the title, there is a large circular placeholder image. The main content area contains the following text:

News Analyser

A software tool for Tracking & Tracing Fake News and its origin using official sources as the input filter.

[Start Now](#) [Learn more](#)

Enter URL or Text Query

You will be provided with accurate information as well as the percentage accuracy of the input article will be highlighted

[Submit](#)

Start Now!

The screenshot shows a response from the K-TRAPS system. At the top, there is a navigation bar with links for "Home", "Users", and "Sign Out →". The main content area displays the following text:

This is not fake news. The source of the article is a legitimate news source, Hindustan Times. The article cites Omarosa Manigault Newman, former aide to President Donald Trump, as its source. There are also several other reputable sources that corroborate her claims, such as The New York Times and The Washington Post.

The statement: "The United States has the highest rate of gun violence in the developed world." According to a 2019 report from the Small Arms Survey, the United States has the highest rate of civilian-held firearms per capita in the developed world with 120.5 guns per 100 people. The same report found that the United States also had the highest absolute number of gun homicides among the countries studied, with 11,208 deaths in 2017. Furthermore, an analysis conducted by BBC News in 2018 found that the United States accounted for 29 percent of all mass shootings worldwide between 1966 and 2012, which is more than five times as many as the next country on the list, the Philippines. These statistics indicate that the United States does indeed have the highest rate of gun violence in the developed world. Sources: Small Arms Survey: <https://www.smallarmssurvey.org/fileadmin/docs/F-Working-papers/SAS-WP25-US-Firearms-Industry.pdf> BBC News: <https://www.bbc.com/news/world-us-canada-42675809>

Admin/Government Interface

K-TRAPS

- **Trend Analysis of Current Affairs:**

An analysis of the current affairs, basically the news articles and topics that are frequently queried on our website is visually represented in the form of a graph.

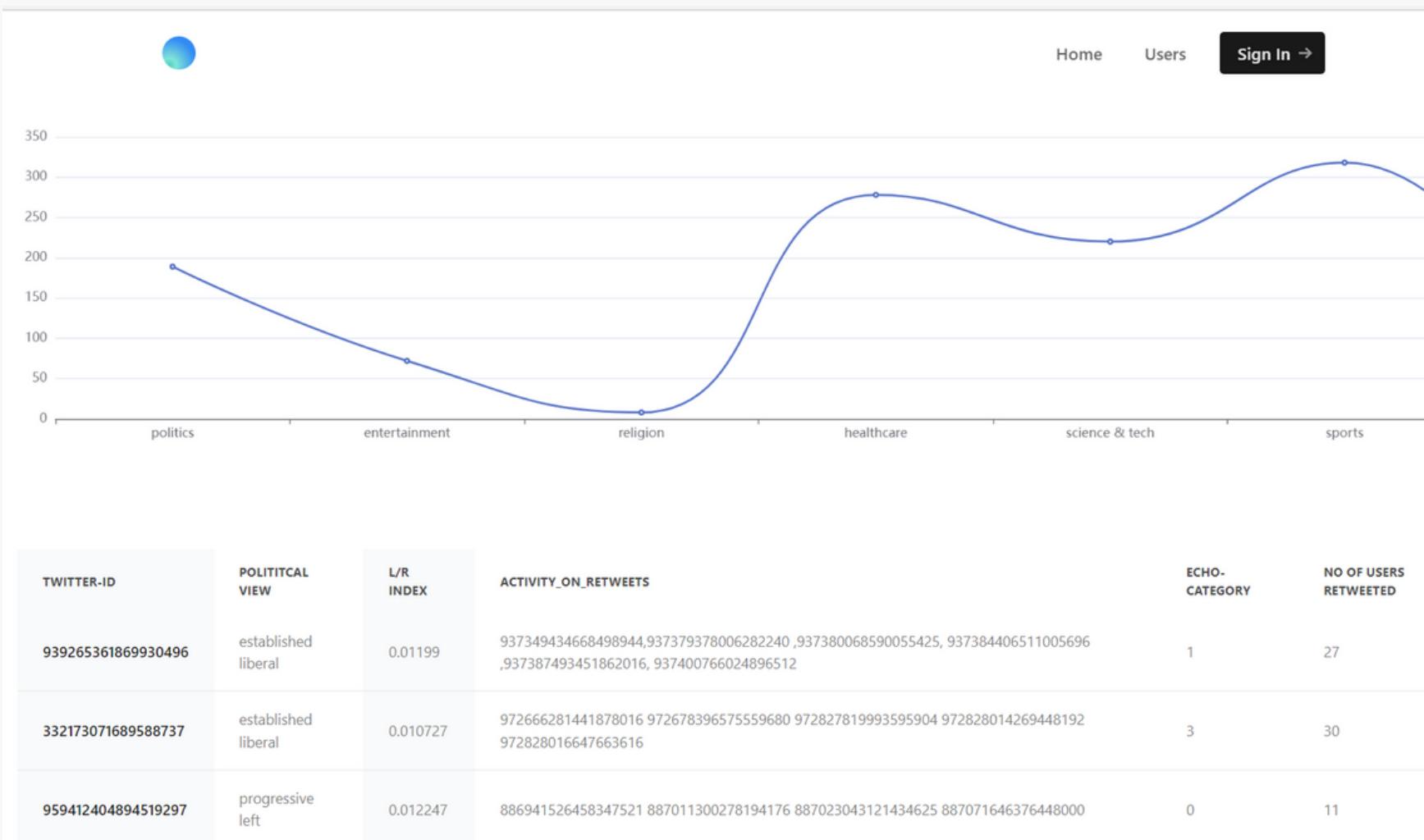
- **User Catalogue**

An analysis of the latest search query of all users, the political affiliation associated with the query along with its authenticity and category is visually represented in the form of a graph.

- **User Sketch:**

An analysis of all the queries of a user on our website is useful in generating their sketch/profile. The search queries are represented along with their authenticity, political affiliation, sentiments expressed, type of speech expressed in the query/news article, clickbait probability, list of users that have forwarded that message in the form of tweets, which in turn helps in determining the community they belong to, are cited tabularly as well as graphically

Analytics for Admin /Government



Real-time-Graph of community detection, political views and category of news



Real-time-Graph of latest search by a particular user, the output (binary form) and political views



CHROME EXTENSION

- The chrome extension can be used to validate text that is selected on a website and implicitly generate a post on any social media platform containing authentic information about the selected text along with various resources referring to the text
- The chrome extension will also filter the data on the user's social media feed in terms of Content Filtering and Information Filtering.
- We are using a feed filter system via Chrome extension that will extract contents recommended by the social media platform recommender system. The information or content extracted will be first scrutinized in terms of fake news. If revealed positively that post will automatically be discarded from the user's feed.
- This way we are limiting the exposure and subsequent impact of fake news propagation



CHROME EXTENSION

K-TRAPS

We will also preemptively decrease the amount of hate speech and misinformation spread by randomizing the content shown by the social media recommender to the user with our own recommendations based on analysis of user's profile and community relation

Select text and right-click search on the extension

Inbox (105) - adwai... Imported R Node Modules at... Classes Learn the blockchai... What's the Differen... ex E

HT Premium Web Stories Trending Quiz Videos Photos Tech Business Spec

kept in an "infamous vault" by the owners of The National Enquirer.

Read more: [Top Russian officials can't leave country until PM's nod: Report](#)

Omarosa Newman who served as the director of communications for the Office of Public Liaison during the Trump administration shared her perspective on Donald Trump's legal troubles as the former President pleaded not guilty to 34 charges of falsifying business records in a Manhattan court in case related to hush money payments to adult film star Stormy Daniels.

Reading the indictment conf
Donald Trump, Omarosa Ma

Read more: [Man sneaks into arrested: Report](#)

Copy Ctrl+C
Search Google for "Omarosa Newman who served as the..."
Print... Ctrl+P
Brave
Search 'Omarosa Newman who served as the...' on my website
Inspect

Get Response without visiting website

This is not fake news. The article is from the Hindustan Times, a reputable news article is based on claims from former White House aide Omarosa Manigault Ne cites sources such as CBS News and the Associated Press. Sources: <https://www.hindustantimes.com/world-news/donald-trump-news-omarosa-newman-donald-trump-s-secrets-and-tapes-kept-in-infamous-vault-ex-aide-c101681023691922.html> CBS News: <https://www.cbsnews.com/news/omarosa-newman-claims-white-house-had-secret-vault-of-president-trumps-tapes/> Associated Press: <https://apnews.com/014f02ea5466401ba8344e7d8a6a41b7>

This statement: "The number of homeless people in the United States has increased since 2017" According to the U.S. Department of Housing and Urban Development's Homeless Assessment Report, the number of homeless people in the United States increased from 553,742 in January 2017 and 647,258 in January 2018, an increase of 17%. Other sources include the National Alliance to End Homelessness: <https://endhomelessness.org/homelessness-in-the-united-states/>

Share Responses on Social Medias

twitter.com/intent/tweet?text=This%20is%20not%20fa... [Download](#) [Link](#) [Report](#) [Unmute](#) [N](#)

Imported R Node Modules at... Classes Learn the blockchai... What's the Differen... ex Express r

Everyone

This is not fake news. The story was reported by multiple news outlets, including NBC News, CNN, and ABC News. Here are some of the sources: NBC News: <https://www.nbcnews.com/politics/donald-trump/omarosa-says-trump-s-legal-troubles-are-symptom-his-insecurity-n1162176CNN>: <https://www.cnn.com/2020/03/09/politics/omarosa-manigault-newman-trump-legal-troubles/index.html> ABC News: <https://abcnews.go.com/Politics/omarosa-manigault-newman-discusses-trumps-legal-troubles/story?id=69847316>

Everyone can reply

[Image](#) [GIF](#) [Link](#) [Smile](#) [Reply](#) [Retweet](#) [Tweet](#)

DATA COLLECTION

TYPES OF USERS:

USER OF OUR SYSTEM->MAIN USER

PEOPLE THAT HAVE PERFORMED ACTIVITY ON THE MAIN USERS POST->SECONDARY USERS

THE USER DATA COLLECTED CAN BE CATEGORISED

- ACTIVITY ON OUR WEBSITE
- SOCIAL MEDIA ACTIVITY

ACTIVITY ON OUR WEBSITE:

- THE SEARCHES DONE USING OUR SYSTEM DETERMINES THAT THE USER IS QUESTIONING THE CONTENT AND AUTHENTICITY OF THE NEWS ARTICLE PRESENTED TO THEM.
- FROM THAT WE CAN EXTRACT/INFER THAT THEY ARE QUESTIONING THIER STANCE ON THE PARTICULAR TOPIC AND TYPE OF PROPAGANDA TECHNIQUE THAT DOESN'T WORK ON THEM

DATA COLLECTION

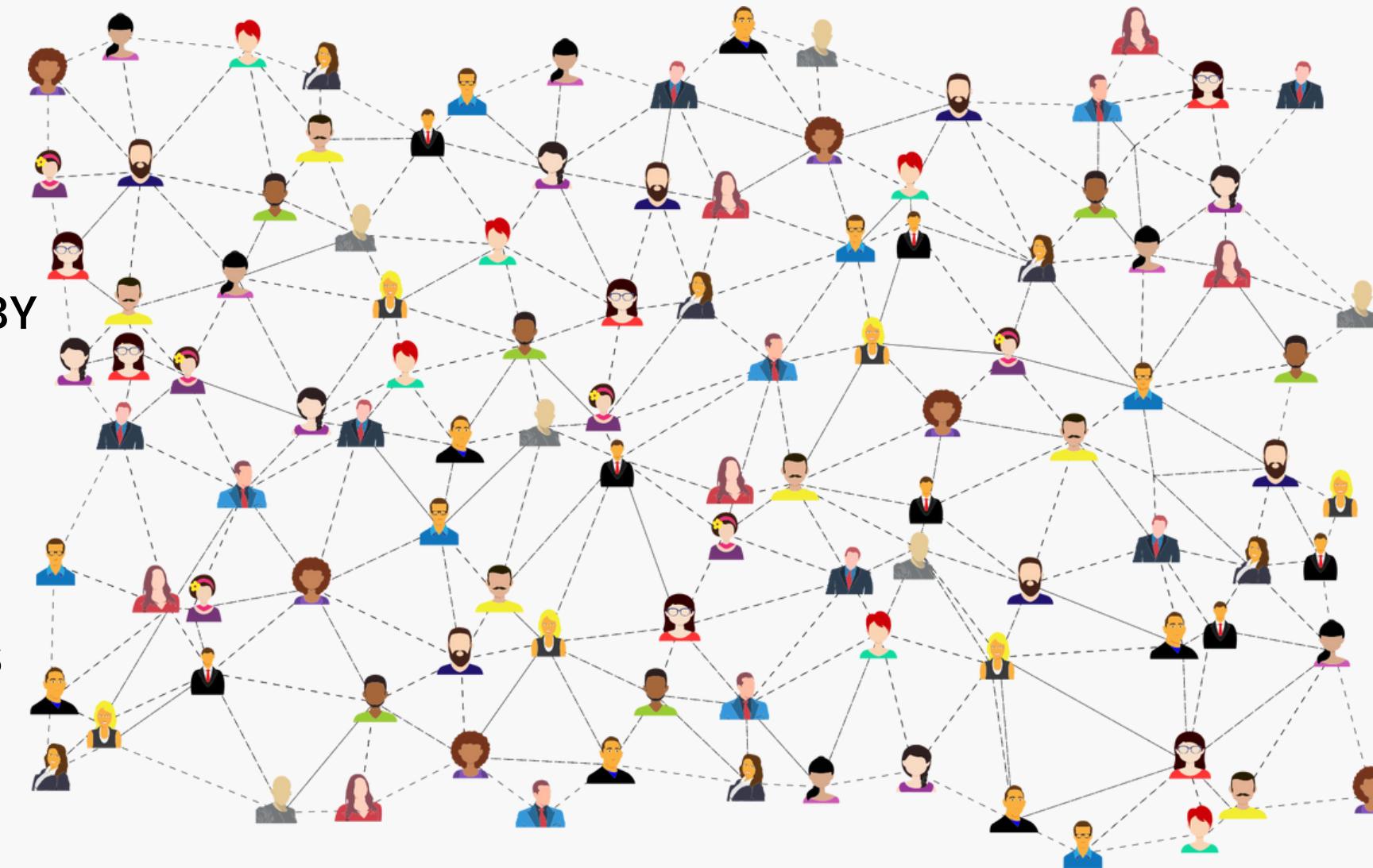
K-TRAPS

SOCIAL MEDIA ACTIVITY:

- ONCE OUR EXTENSION HAS BEEN ACTIVATED THE DATA WILL BE COLLECTED USING THE USERNAME AND USER ID OF THE PERSON
- TIME SPENT ON A PARTICULAR TWEET/POST BY THE MAIN USER ALONG WITH ITS CONTENT
- REPOSTED CONTENT AND ACTIVITY (LIKE, COMMENTS, REPOSTS) BY THE MAIN USER OR SECONDARY USER'S POSTS IN THE THREADS

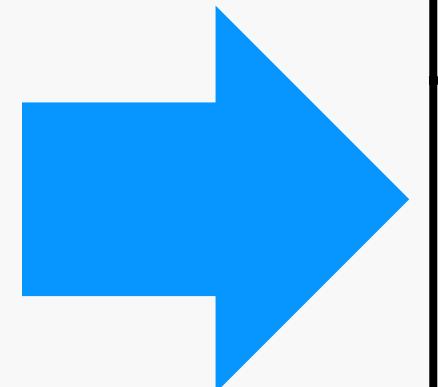
USER-FEED FILTER:

- ACTUAL RECOMMENDATIONS MADE BY SOCIAL MEDIA PLATFORMS WILL BE SEQUENTIALLY SENT FOR FACT-CHECKING AND A RECOMMENDER SYSTEM FOR TAILORED RECOMMENDATIONS
- SECONDARY USERS THAT HAVE PERFORMED SOME ACTIVITY(COMMENTS/LIKES) ALONG WITH THEIR USER IDS AND SOCIAL MEDIA HISTORY OF THEIR ACTIVITY



ANALYSIS

ACTIVITY ON OUR WEBSITE	SOCIAL MEDIA ACTIVITY
USERNAME & USER_ID	USERNAME & USER_ID
TYPE OF INFORMATION SEARCH	CURRENT USER FEED AS PER SOCIAL MEDIA
ALL QUERIES AND FOLLOW UP QUESTIONS	ACTIVITY BY THE USER AND SECONDARY USERS ON USER'S ACTIVITY



Performing Analysis

URL ON POST	SENTIMENT ANALYSIS
TEXT ON THE POST	TYPE OF SPEECH (OFFENSIVE,HATESPEECH ETC.)
CATEGORY OF NEWS (SPORTS,EDUCTION,BUISNESS,SCIENC E & TECHNOLOY,MEDICINE & HEALTH)	CLICKBAIT PROBABILITY
POLITICAL VIEW EXPRESSED	EXTENT OF REPOST (NO OF ACTITVITY ON POST)
TYPE OF PORPAGANDA TECHNIQUE USED	COMMUNITY & ECHOCHAMBER CATEGORY PREDICTIONS
KEYWORDS	SECONDARY USER IDS

FAKE NEWS DETECTION SYSTEM

K-TRAPS

Current System: GPT-3

Alternative LLMs: Llama Alpaca

Details:

- temperature=0.7,
 - max_tokens=2000,
 - top_p=1.0,
 - frequency_penalty=0.0,
 - presence_penalty=1
-
- most up-to-date information(till 2023)
 - faster information retrieval of sources
 - huge database of information
(>175 billion parameters and >410 billion tokens)
 - better accuracy than alternate systems
(76% to 86.4% accuracy)

Input:

Piece of text or URL of Article or Headline

Output:

- **Verification**

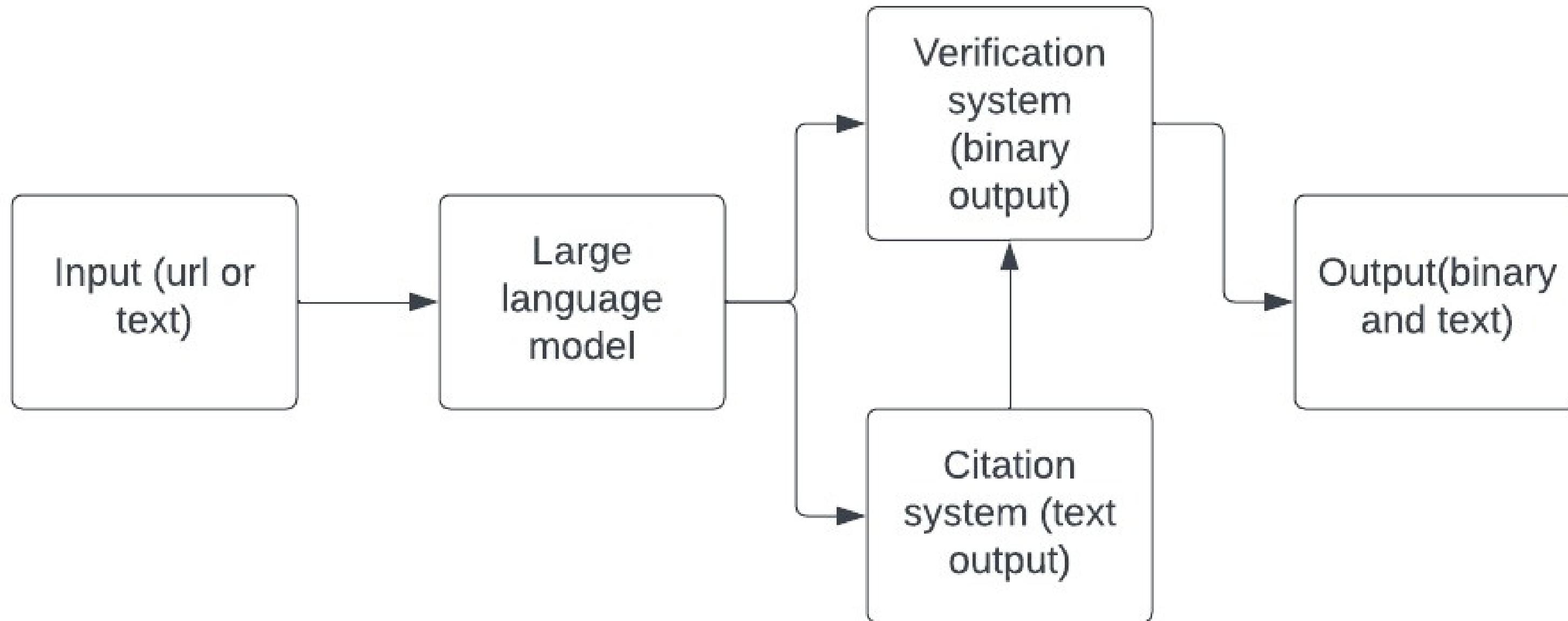
Determines whether the piece of text is having correct information and gives result as a binary output(true or false)

- **Citations or Sources**

Based on the input we find similar correct information on the GPT-3 knowledge graph reaffirm the information through the verification stage and return to the user only if the content is correct or true

FAKE NEWS DETECTION SYSTEM

UNDERLYING WORKING



ALTERNATE FAKE NEWS DETECTION SYSTEM

K-TRAPS

DATABASE:

We are creating a centralized database of information related to the government and its activities

This will form as a base knowledge graph for governments own fact checking Large Language Model (LLM)

This database will contain the following data:

- Wikipedia, Wikimedia and Google Knowledge Graphs
- live twitter data scraping of government & ministries accounts
- data scraping of reports across all government websites like Press Information Bureau (PIB)
- non-classified reports from ministry of defence and armed forces
- and many more databases that will be considered as ground truth for the fact checking large language model

The government can regulate the ground truth with the inclusion of knowledge graphs and later can be made decentralized with verification systems in place for making the fact checking system 'of the people, for the people and by the people'

ALTERNATE FAKE NEWS DETECTION SYSTEM K-TRAPS

1. Multiple Factchecking LLMS

Using GPT-4, Alpaca, GPT-3.5, LLaMA all fine tuned on our customized database and knowledge graph for fact checking

2. Hybrid Graph Neural Network and Transformer based Large Language Model

Graph Neural Networks:

- Advanced graph neural networks like Bi-Directional Graph Convolutional Networks(BiDGCN), Graph Convolutional Networks for Fake News(GCNFN), User Preference-aware Fake News Detection (UPFD)
- all have proven very good results on rumor detection on social media and fake news probability detection based on user history and social graphs
- Fine-tuning using our centralized database the customized Large Language Model will be able to detect fake news and trace it very quickly

Transformer Based Models

By using transformer based models like bert and roberta with siamese networks to measure proximity of content of credible sources from our database and knowledge graphs with the content presented, we will also be able to present logical justification for the incorrect news along with cited sources using graph networks

ALTERNATE FAKE NEWS DETECTION SYSTEM

UPFD MODEL ARCHITECTURE

GCNFN Model Architecture

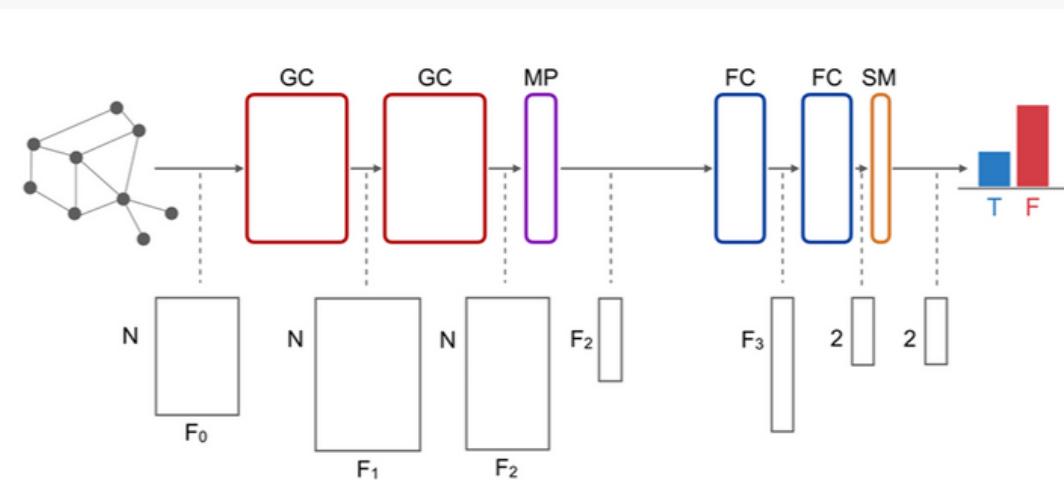


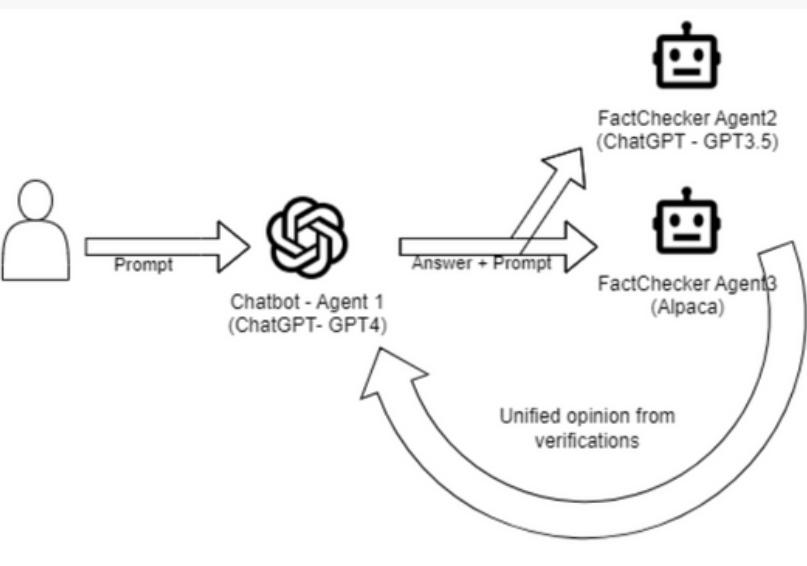
Figure 5: The architecture of our neural network model. Top row: GC = Graph Convolution, MP = Mean Pooling, FC = Fully Connected, SM = SoftMax layer. Bottom row: input/output tensors received/produced by each layer.

Given a URL u (or a cascade c arising from u) with corresponding tweets $T_u = \{t_u^1, t_u^2, \dots, t_u^N\}$ mentioning it, we describe u in terms of graph G_u . G_u has tweets in T_u as nodes and estimated news diffusion paths plus social relations as edges. In other words, given two nodes i and j , edge $(i, j) \in G_u$ iff at least one of the following holds: i follows j (i.e. the author of tweet i follows the author of tweet j), j follows i , news spreading occurs from i to j , or from j to i .

News diffusion paths defining *spreading trees* were estimated as in [42] by jointly considering the timestamps of involved (re)tweets and the social connections between their authors. Given t_u^n – the retweet of a cascade related to URL u , and $\{t_u^0 \dots t_u^{n-1}\}$ – the immediately preceding (re)tweets belonging to the same cascade and authored by users $\{a_u^0, \dots, a_u^n\}$, then:

1. if a_u^n follows at least one user in $\{a_u^0, \dots, a_u^{n-1}\}$, we estimate news spreading to t_u^n from the very last tweet in $\{t_u^0 \dots t_u^{n-1}\}$ whose author is followed by a_u^n ;
2. if a_u^n does not follow any of the users in $\{a_u^0, \dots, a_u^{n-1}\}$, we conservatively estimate news spreading to t_u^n from the user in $\{a_u^0, \dots, a_u^{n-1}\}$ having the largest number of followers (i.e. the most popular one).

Finally, nodes and edges of graph G_u have features describing them. Nodes, representing tweets and their authors, were characterized with all the features presented in Section 2⁴. As for edges, we used features representing the membership to each of the aforementioned four relations (*following* and *news spreading*, both directions). Our approach to defining graph connectivity and edge features allows, in graph convolution, to spread information independently of the relation direction while potentially giving different importance to the types of connections. Features of edge (i, j) are concatenated to those of nodes i and j in the attention projection layer to achieve such behavior.



Using Multiple LLMs like GPT-4, GPT-3.5, Alpaca etc

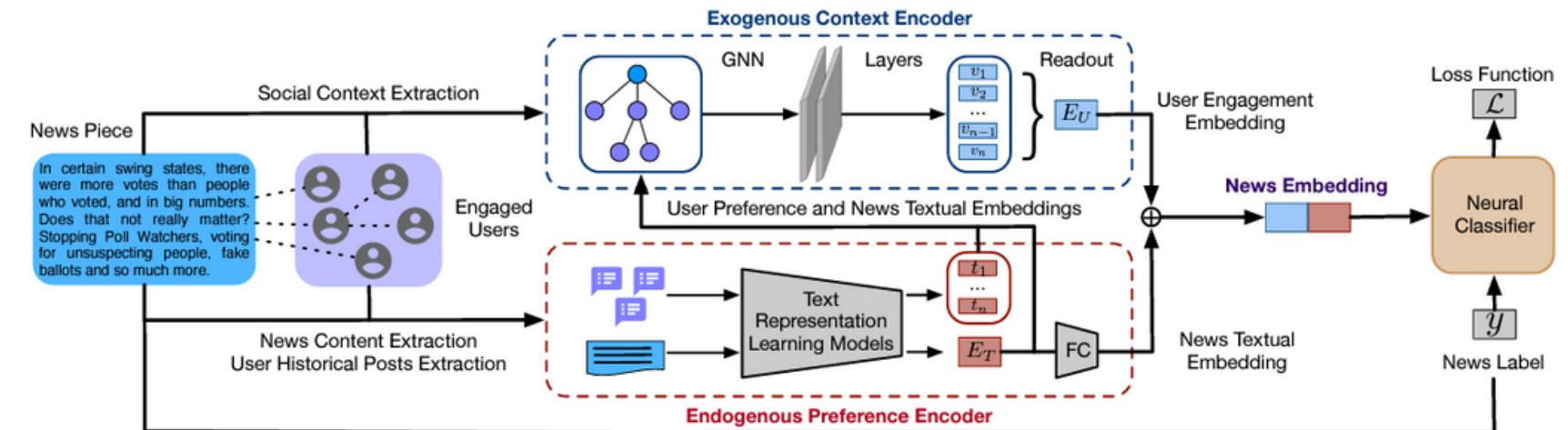


Figure 1: The proposed UPFD framework for user preference-aware fake news detection. Given the news piece and its engaged users on social media, we extract the exogenous context as a news propagation graph and encode the endogenous information based on user historical posts and news texts. The endogenous and exogenous information are fused using a GNN encoder. The final news embedding, composed of user engagement embedding and news textual embedding, is fed into the neural classifier to predict the news' credibility.

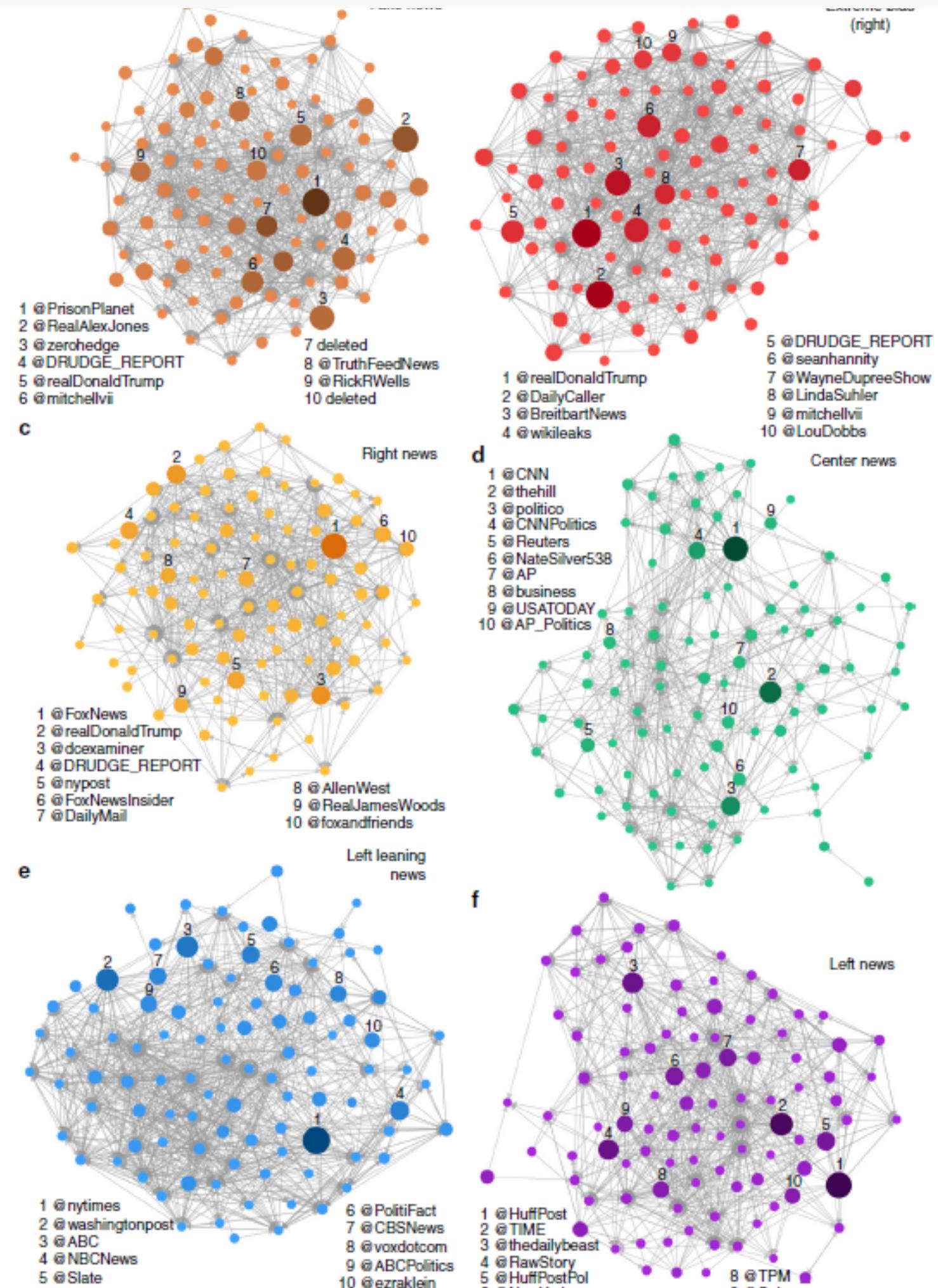
ID	Statement	Justification	label	S	S+M	SJ	S+MJ
1	We have the highest tax rate anywhere in the world.	Trump, while lamenting the condition of the middle class, said the U.S. has "the highest tax rate anywhere in the world." All sets of data we examined for individual and family taxes prove him wrong. Statutory income tax rates in the U.S. fall around the end of the upper quarter of nations. More exhaustive measures - which compute overall tax burden per person and as a percentage of GDP - show the U.S. either is in the middle of the pack or on the lighter end of taxation compared with other advanced industrialized nations.	false	X	✓	✓	✓
2	"Says Rick Scott cut education to pay for even more tax breaks for big, powerful, well-connected corporations."	A TV ad by the Florida Democratic Party says Scott "cut education to pay for even more tax breaks for big, powerful, well-connected corporations." However, the ad exaggerates when it focuses attention on tax breaks for "big, powerful, well-connected corporations." Some such companies benefited, but so did many other types of businesses. And the question of whether the tax cuts and the education cuts had any causal relationship is murkier than the ad lets on.	half-true	X	X	✓	✓
3	Says Donald Trump has given more money to Democratic candidates than Republican candidates.	but public records show that the real estate tycoon has actually contributed around \$350,000 more to Republicans at the state and federal level than Democrats. That, however, is a recent development. Fergusons statement contains an element of truth but ignores critical facts.	mostly-false	X	X	✓	✓
4	Says out-of-state abortion clinics have marketed their services to minors in states with parental consent laws.	As Cousins clinic in New York told Yellow Page users in Pennsylvania, "No state consents." This is information the clinics wanted patients or potential patients to have, and paid money to help them have it. Whether it was to help persuade them to come in or not, it provided pertinent facts that could help them in their decision-making. It fit the definition of marketing.	true	X	X	X	✓
5	Obamacare provision will allow forced home inspections by government agents.	But the program they pointed to provides grants for voluntary help to at-risk families from trained staff like nurses and social workers. What bloggers describe would be an egregious abuse of the law not whats allowed by it.	pants-fire	X	X	X	✓
6	In the month of January, Canada created more new jobs than we did.	In November 2010, the U.S. economy created 93,000 jobs, compared to 15,200 for Canada. And in December 2010, the U.S. created 121,000 jobs, compared to 22,000 for Canada. "But on a per capita basis, in recent months U.S. job creation exceeded Canada's only in October." January happened to be a month when U.S. job creation was especially low and Canadian job creation was especially high, but it is the most recent month and it reflects the general pattern when you account for population.	true	X	X	X	X
7	There has been \$5 trillion in debt added over the last four years.	number is either slightly high or a little low, depending on the type of measurement used, and thats actually for a period short of a full four years. His implication that Obama and the Democrats are to blame has some merit, but it ignores the role Republicans have had.	mostly-true	X	X	X	X

SOCIAL NETWORK COMMUNITY DETECTION & ECHO CHAMBER DETECTION

- **Community detection** focuses on identifying clusters or groups of users with similar interests or behaviour patterns. Echo chambers refer to the tendency of individuals to seek out and consume information that confirms their existing beliefs and opinions.
- By analyzing the patterns of interaction and content sharing within these communities, we have generated various communities based on views and opinions on a particular topic. This is used as bed-rock in the recommender system and in the mitigation of correct news. This information will be used to develop preventive measures, such as targeted interventions and messaging campaigns, to disrupt the spread of fake news within these communities.
- **Echo chamber predictions** focus on identifying users who are most susceptible to being trapped in a "filter bubble" or "echo chamber", where they are only exposed to information and viewpoints that reinforce their existing beliefs and biases.

SOCIAL NETWORK COMMUNITY DETECTION & ECHO CHAMBER DETECTION

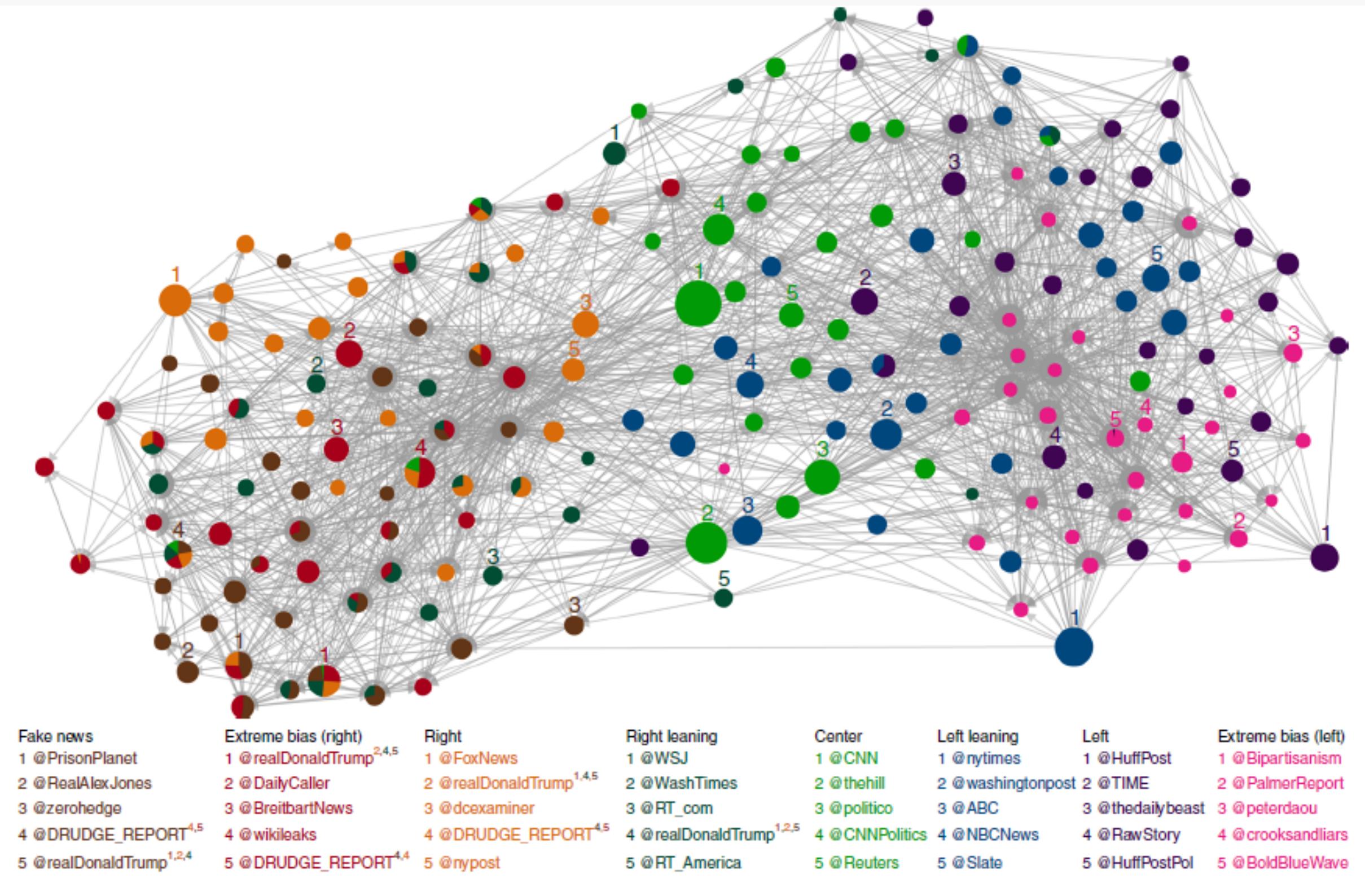
- With the help of our recommender system we prevent the user from being exposed to contents of only one type of political views. We offer logical and factual posts with increasing entropy of views of the posts. **This will help the user form their own opinion rather than getting sucked into a bubble.**
- Both of these used collaboratively along with the unique social and psychological factors that contribute to the formation of these information bubbles to preemptively work on spread of fake news.



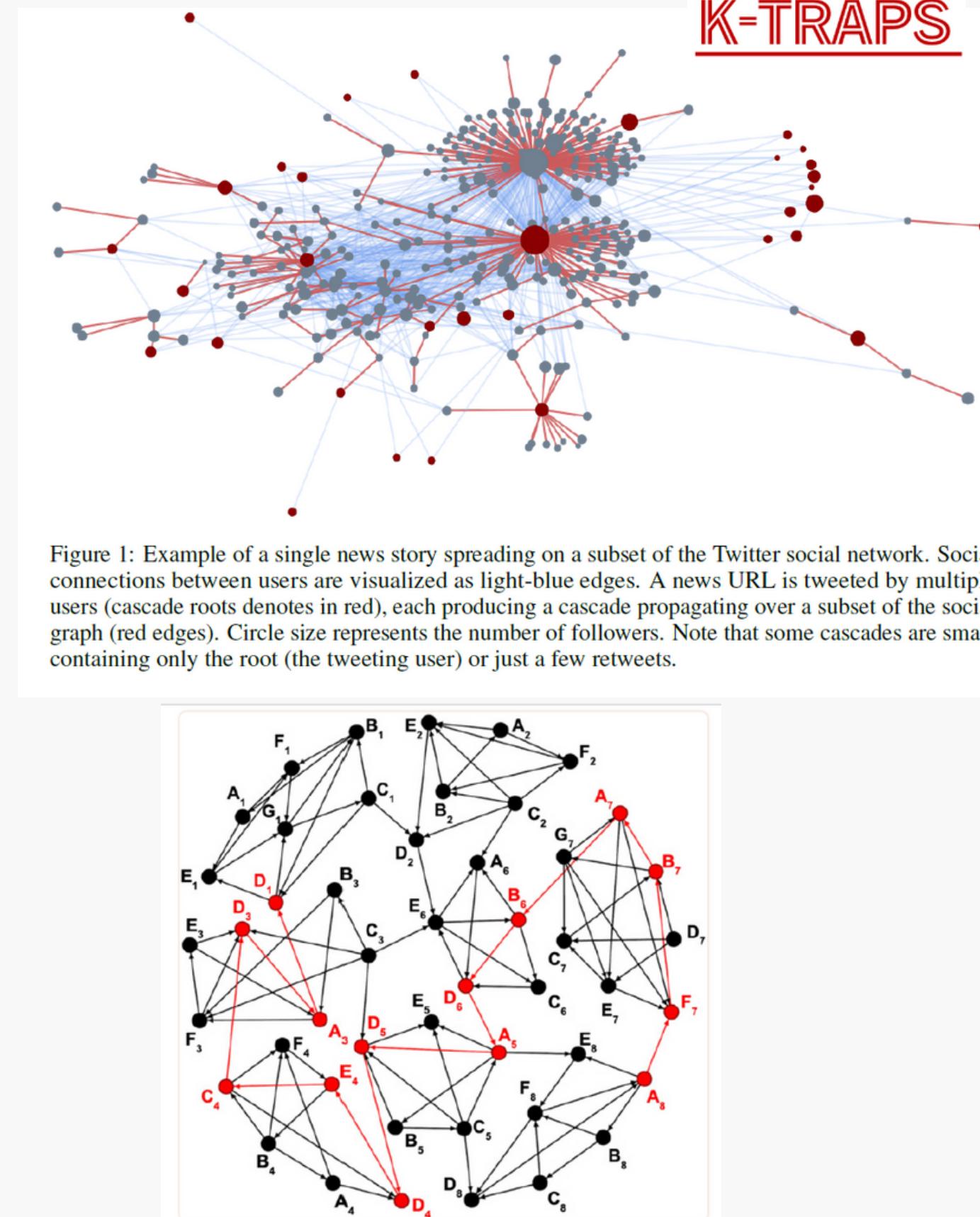
- Here we have a network graph of the top 100 twitter accounts that are the major influencers of news propagation related to a particular topic. This also includes a graph with the accounts propagating fake news and other political views
- Using community detection we have identified the probability of a community being the fake news originator for a particular topic. Based on trend analysis of the news, regardless of the topic it is, we know the likely community it originated from
- Using echo chamber analysis we also know which community will be affected or susceptible to it based on political views, psychological factors, propaganda techniques used etc.

Impact of misinformation spread and interactivity of various communities with fake news spreaders

K-TRAPS



Interactivity between multiple types of communities or 'echo-chambers'

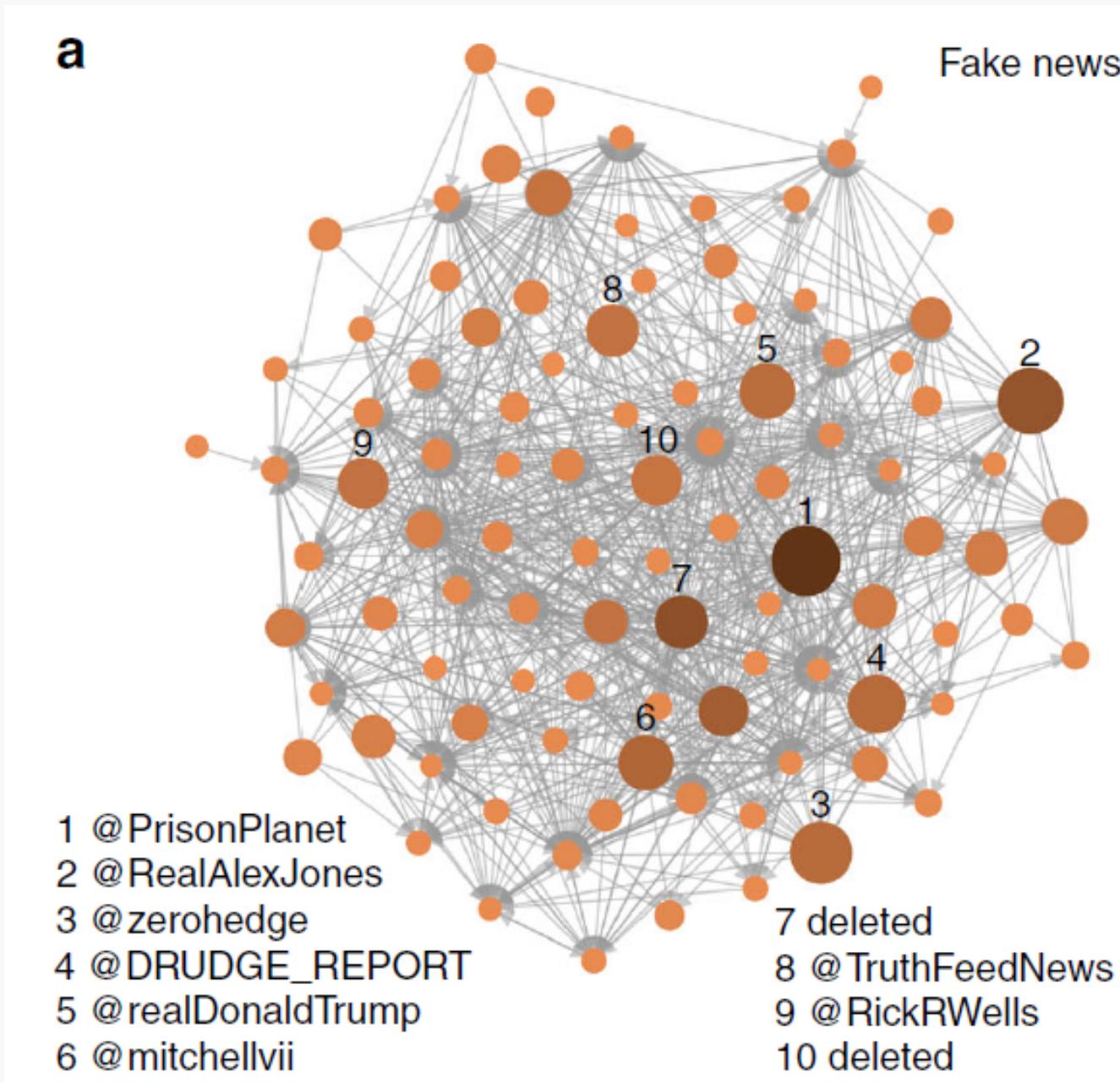


Red nodes denote fake news spreaders

Using this information how do we combat the fake news spreaders ?

K-TRAPS

Top 100 accounts spreading fake news related to a particular topic



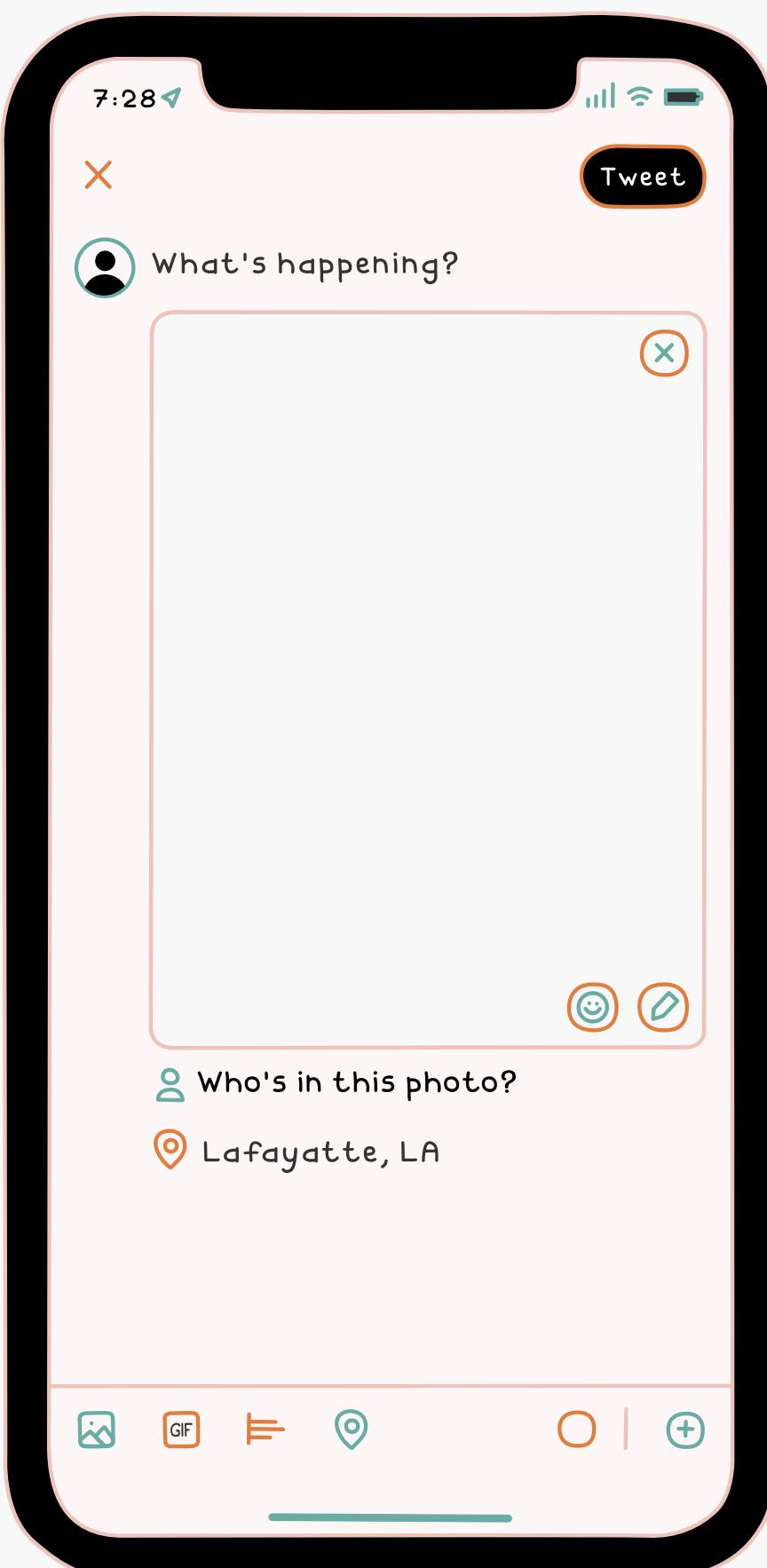
Bombard these accounts with posts from reccomender system with correct news related to the topic

- We use targeted interventions and mass messaging
- We have identified the community of fake news propagation and the community most susceptible to it
- We bombard those communities with factual news and logical posts related to those topics as soon as the trend analysis predicts that the topic is likely to be used for fake news propagation
- We do this using posts with factual news and propaganda techniques and psychological factors that work on those communities. This preemptive strike will deliver the factual and real news before the fake one as the primary source of information
- Bombarding these communities, especially the top users, will help us spread the correct news exponentially faster than the fake news and limit its propagation

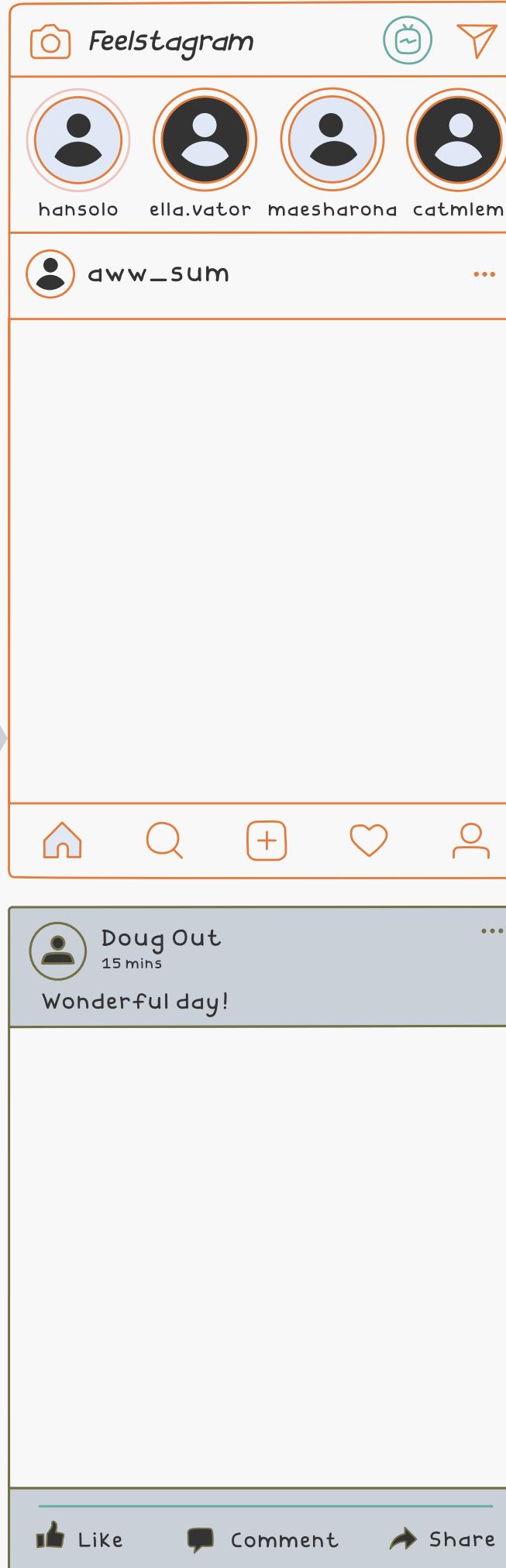
HYBRID RECOMMENDATION SYSTEM

- Using the chrome extension we are taking the social media defined user-feed and performing analysis on it. We are considering posts in batches of 10 for reducing computation time and response time
- From the posts we are extracting its contents in terms of images, text, and videos. The images and videos are converted to text using speech-to-text and video description generator
- We are first checking if any of those contents contains incorrect or fake news. If so we immediately discard that post
- Now on the rest of the posts we perform analysis in terms of category, political views, propaganda technique, clickbait probability, hate-speech probability and extreme-community content category etc.

SOCIAL MEDIA INFORMATION EXTRACTION



social media generated user feed posts extraction



Information extraction

TEXTUAL INFORMATION OF POSTS

VIDEOS
SPEECH-TO-TEXT OF AUDIOS AND VIDEO DESCRIPTION GENERATOR

Images
Extract any text in the image
Image Caption and Description Generator

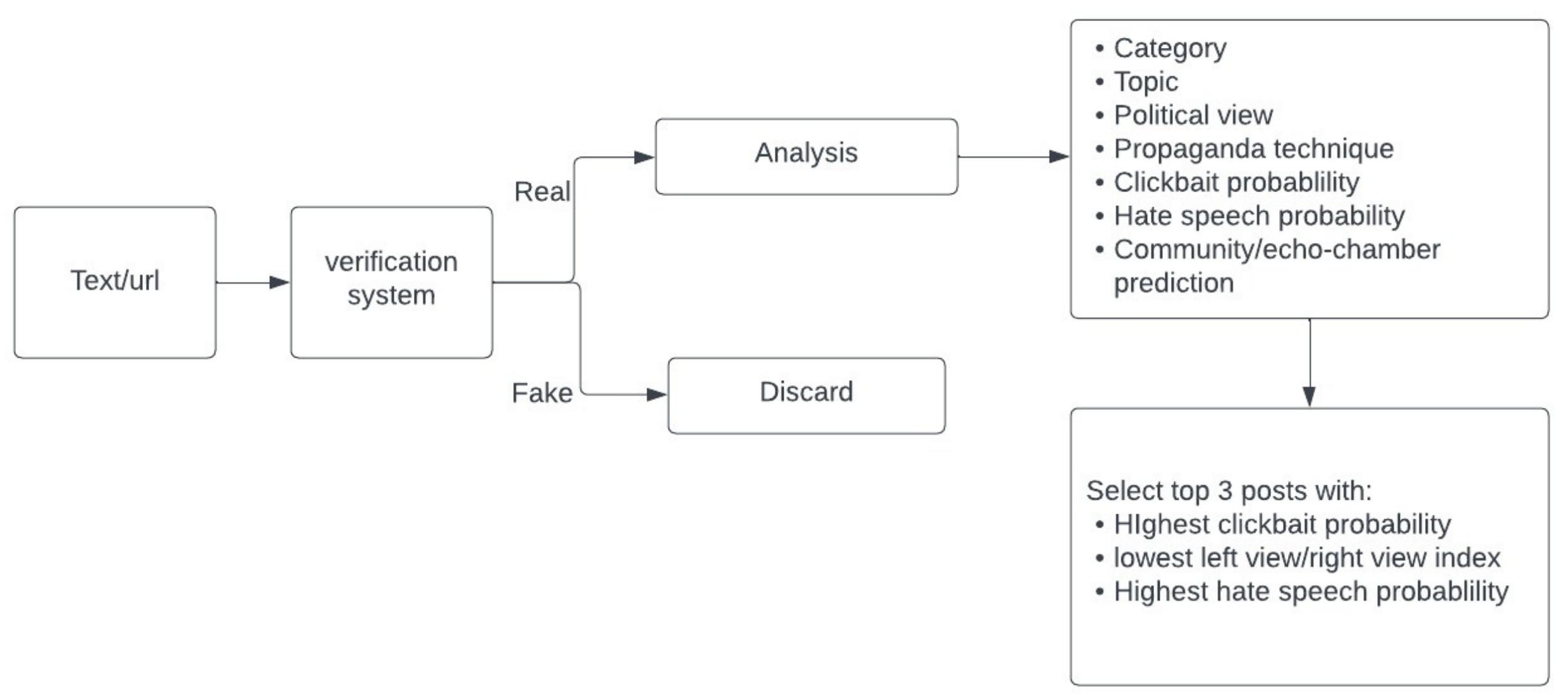
HOW WILL RECOMMENDATIONS BE MADE?

- Based on this analysis we select posts with the highest values in clickbait probability, hate-speech probability, and extreme-community content probability and create an index called 'l/r index' that measures the probability of posts having left views by right views. we select k top posts. k can be adjusted accordingly
- In parallel we select posts with the same topic and keywords of the k-top posts having the least clickbait probability, 5 percent increment of the user's l/r index, and least hate-speech content. the 5 percent increment is considered such that the user doesn't detect too much shift in political views in their content. This can be adjusted accordingly
- We mix these k 'recommended' posts with the rest of the (10-k) posts randomly and send this as content to the user
- On these posts we look user's activity and analyze the l/r index and other parameters and based on these we recommend the next 10 recommendations. This is done recursively based on newly computed l/r index

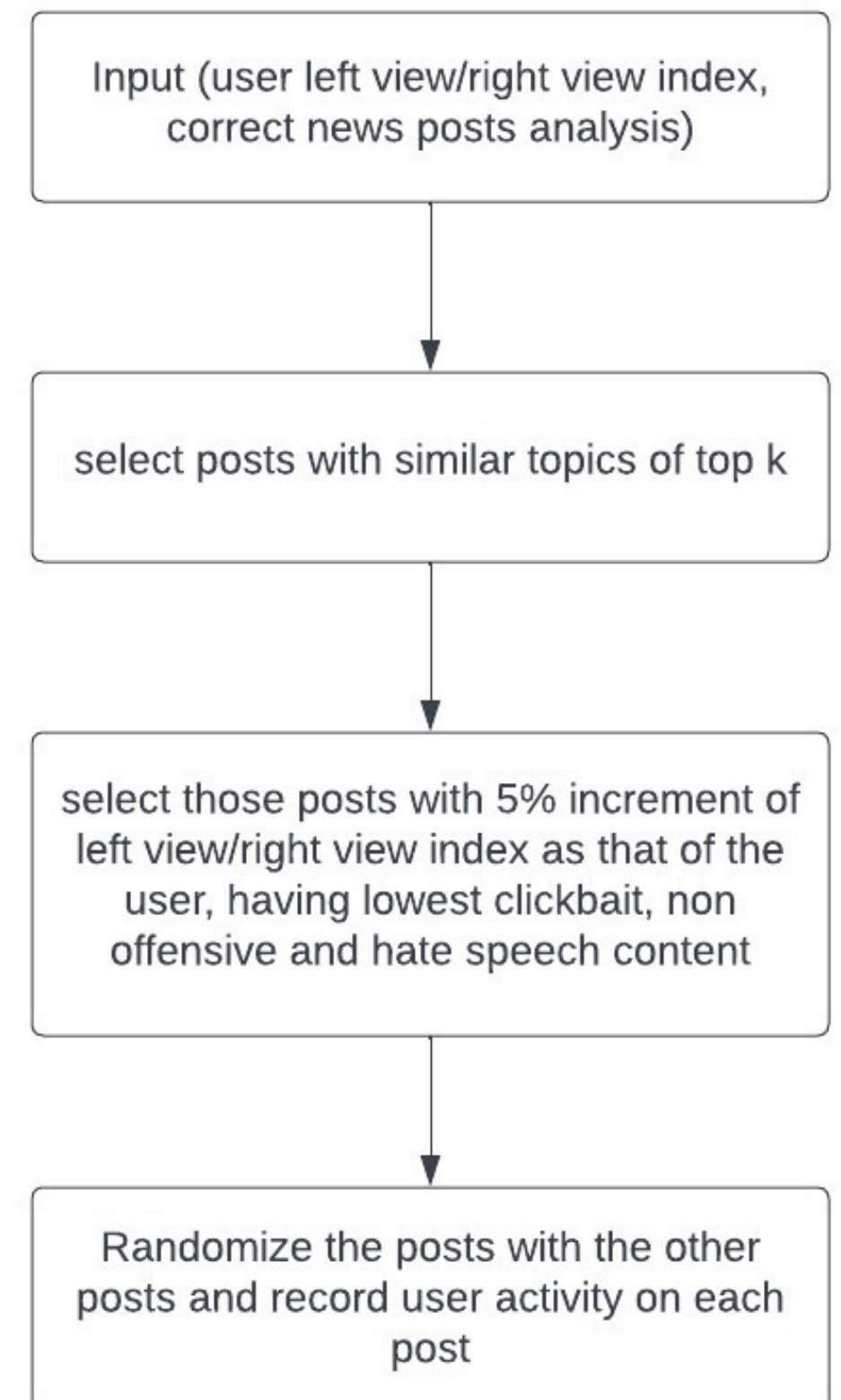
Personalized Reccomender System

Social Media User Feed

Here we are checking the authenticity of the posts and discarding the fake ones, and selecting those with extreme views to be replaced with posts having a neutral point of view



here we are selecting those real and factual news with same topic as that of the extreme views but showing a slightly different opnion, this to prevent a user form being sucked into an echo-chamber but having their opinon based on facts



IMPACT

- Increased trust in media:

If effective measures are taken to mitigate fake news, people may develop more trust in the media. This can lead to a more informed and engaged public.

- Reduced polarisation:

Fake news can exacerbate political polarisation by spreading false information that reinforces pre-existing beliefs. If people are exposed to more accurate and balanced reporting, it may help to reduce polarisation and promote more constructive dialogue.

- Using Analytics to make better policies and govern better

With the analysis and recommender system devised to reverse the impact of fake news and decrease political polarization. we can use these analysis to create better policies and laws that will benefit the people. This will also help the government in terms of better policing and enforce laws and policies to decrease cybercrime and cyber-bullying by tracing scams, hatespeech and much more.

- Perform threat analysis:

Performing threat analysis on data collected we will be helping the government identify communities that pose threat and take measure to prevent real life implications

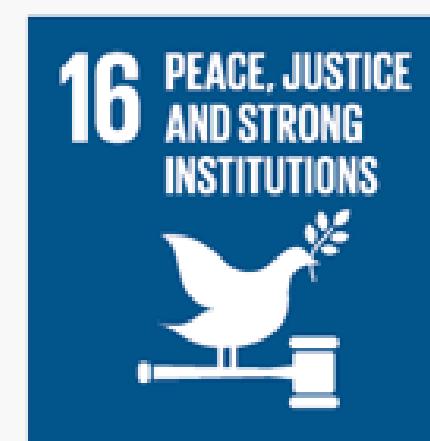
- Increased social trust:

Fake news can erode social trust by promoting distrust and suspicion of others. Mitigating fake news can help to rebuild social trust and promote a more cohesive and united society.

OUT OF THE 17 U.N. SUSTAINABLE GOALS OUR SOLUTION FOCUSES ON



SUSTAINABLE DEVELOPMENT GOALS



Preventing spread of fake news related to medicine and health services
De-stigmatizing crucial medical problems and preventing misinformation and hatespeech related to them

Our solution will decrease hate-speech and offensive posts on the platforms towards underrepresented and oppressed communities

We are preserving the right to privacy and preventing social media influence major opions and change views of people
keeping track and performing threat analysis of communities having oppressive views towards minority communities and the Government preventing real life implications
ex #pizzagate and January 6th Capitol riots

FUTURE SCOPE

- Make response time of information retrieval and fact checking faster
- Implement audio video deepfake news analysis for deepfake detection on news channels and video based social media platforms
- Cross information exchange between multiple social media platforms for better recommendation system and preventing political polarization
- Increasing the amount of information using credible sources as base truth. This will also help to make the justifications that are provided along with the detected fake news and the followup questions better

CONCLUSION

K-TRAPS

Problem	Our Solution
<p>Design and develop a technological solution/software tool for Tracking & Tracing Fake News and its origin using official sources as the input filter.</p>	<ul style="list-style-type: none">• usage of multiple approaches of fake news detection system with a proposed centralized database of information for generation of government's own large language model• tracing the origination of fake news with help of community detection and echo chamber predictions
<p>The solution should have a mechanism to mitigate the impact of the spread of Fake News by auto-populating the fake news spreaders' inboxes with the official/authenticated news content.</p>	<ul style="list-style-type: none">• the personalized recommendation system also prevents political polarization by preventing only post containing extreme views to be presented to the user and helps in managing the already accepted fake news and its impact• Using Graph Neural Networks to do impact analysis by echo chambers creation and tracking of information spread in them,in case of a new controversy correct information in terms of posts is spread throughout the chamber to prevent the fake news from being spread.Altering recommendations by removing fake-news and mixing posts of different viewpoints before recommending it to the user

We have created a sustainable solution with a large socio-economic impact on multiple social media platforms with great results

REFERENCES

1. Influence of fake news in Twitter during the 2016 US presidential election by Alexandre Bovet & Hernán A. Makse1 (doi:10.1038/s41467-018-07761-2)
2. User Preference-aware Fake News Detection by Yingtong Dou1, Kai Shu2, Congying Xia1, Philip S. Yu1, Lichao Sun3 (arXiv:2104.12259v1)
3. Fake news spreader detection using trust-based strategies in social networks without filtration by Bhavtosh Rath,Aadesh Salecha, and Jaideep Srivastava (doi:10.1007/s13278-022-00890-z)
4. Measuring exposure to information from political elites on Twitter by Mohsen Mosleh & David G. Rand (doi: 10.1038/s41467-022-34769-6)
5. Political polarization of news media and influencers on Twitter in the 2016 and 2020
6. US presidential elections by James Flaminio , Alessandro Galeazzi , Stuart Feldman4, Michael W. Macy , Brendan Cross , Zhenkun Zhou6, Matteo Serafino, Alexandre Bovet , Hernán A. Makse & Boleslaw K. Szymanski
7. <https://aclanthology.org/W18-5513.pdf>
8. <https://arxiv.org/pdf/2104.12259.pdf>
9. <https://arxiv.org/pdf/1902.06673.pdf>



Thank You

Name of Team Member	Position in team
Tanaya Shelke	Team Leader
Pranav Janjani	ML/DL
Adwait Mandge	Web Dev
Rohan Fatehchandka	Web Dev
Kunal Gaudani	Web Dev
Siddeshsingh Tanwar	Blockchain