Media and Communication
2025 • Volume 13 • Article 9543
https://doi.org/10.17645/mac.9543

ARTICLE

Open Access Journal

SMART 2.0: Social Media Analytics and Reporting Tool Applied to Misinformation Tracking

Mahmoud Mousa Hamad, Gopichandh Danala, Wolfgang Jentner, and David Ebert

Data Institute for Societal Challenges, University of Oklahoma, USA

Correspondence: Mahmoud Mousa Hamad (mmh@ou.edu)

Submitted: 30 October 2024 **Accepted:** 5 February 2025 **Published:** 3 April 2025

Issue: This article is part of the issue "AI, Media, and People: The Changing Landscape of User Experiences and Behaviors" edited by Jeong-Nam Kim (University of Oklahoma) and Jaemin Jung (Korea Advanced Institute of Science and Technology), fully open access at https://doi.org/10.17645/mac.i475

Abstract

The rapid proliferation of social media has created new data stemming from users' thoughts, feelings, and interests. However, this unprecedented growth has led to the widespread dissemination of misinformation—deliberately or inadvertently false content that can trigger dangerous societal ramifications. Visual analytics combines advanced data analytics and interactive visualizations to explore data and mine insights. This article introduces the Social Media Analytics and Reporting Tool (SMART) 2.0, detailing its application in tracking misinformation on social media. An updated version of its predecessor, SMART 2.0 enables analysts to conduct real-time surveillance of social media content along with complementary data streams, including weather patterns, traffic conditions, and emergency service reports. SMART 2.0 offers enhanced capabilities like map-based, interactive, and AI-powered features that enable researchers to visualize and understand situational changes by assessing public social posts and comments. As a misinformation classification and tracking case study, we collected public, geo-tagged tweets from multiple cities in the UK during the 2024 riots. We showcased the effectiveness of SMART 2.0's misinformation detection and tracking capabilities. Our findings show that SMART 2.0 effectively tracks and classifies misinformation using a human-in-the-loop approach.

Keywords

machine learning; misinformation; SMART; SMART 2.0; social media; surveillance; visual analytics

1. Introduction

This article adheres to the following definitions: misinformation is misleading information shared without intent to deceive, whereas disinformation is shared with deliberate intent to mislead (Treen et al., 2020).

Consistent use of these terms ensures clarity when discussing social media data analysis. In the context of social media, users may unknowingly share disinformation, making it difficult to distinguish between those spreading false information unintentionally versus deliberately.

Individuals spread misinformation for a specific goal, such as for a political agenda, or unknowingly, which can have many dangerous ramifications. Misinformation can have serious consequences, as seen in the false claim that a Muslim killed three children in Southport, England, which led to a mosque being firebombed in Northern Ireland ("Ards mosque community," 2024). Misinformation, such as climate change denial spread through social media, can also hinder progress by creating public confusion, fostering misplaced criticism, and fueling protests against policies meant to combat societal issues (Treen et al., 2020).

The Social Media Analytics and Reporting Tool (SMART), a research tool that uses social media, was developed for situational awareness, event monitoring, and public sentiment analysis (Snyder, Karimzadeh, Stober, & Ebert, 2019). This article introduces SMART 2.0, the updated version of SMART, which helps stakeholders, researchers, and community partners visualize and analyze social media data. We focus on SMART 2.0's machine learning-powered capabilities to track misinformation. We explore its features and applications for media research, showcasing a case study demonstrating its effectiveness in monitoring misinformation. Both SMART and SMART 2.0 are individual systems that we have developed over the years. This article aims to describe SMART and introduce SMART 2.0, highlighting its enhancements and new features.

SMART, developed with a team that includes members at Purdue University and Penn State, has been widely used since 2013. Initially employed by the US Coast Guard and later by many public safety agencies, it played a key role in supporting public safety during college football games, Fleet Week, and presidential inaugurations. Its use expanded to other agencies, including local law enforcement, intelligence centers, and organizations like the American Red Cross, proving its versatility in planned and emergent situations (Snyder, Karimzadeh, Chen, & Ebert, 2019; Snyder, Karimzadeh, Stober, & Ebert, 2019). SMART's integration into the North Atlantic Treaty Organization's network for alerting and managing public safety and resilience (NATO REACT) project led to the creation of SMART 2.0, further demonstrating its global applicability, helping to monitor social media for misinformation and enhancing the crisis response (Ilia State University, n.d.). The objectives of the NATO REACT research project were to build upon SMART using interactive machine learning for real-time human–computer collaborative decision-making, multicultural and multilingual support, human-in-the-loop misinformation identification and information filtering, and information fusing of social and environmental sensing data.

SMART 2.0 extends and improves SMART by including new features and improved performance. These new features and improvements include the following: machine-learning language translation of the user interface, social media data and all other existing visualization tools, misinformation detection and filtering, environmental data sensing and tracking, and improved performance for social media data fetching, as well as an improved software architecture of the system.

2. Related Work and Literature Review

2.1. Related Work

Nowadays, social media has become part of everyday life. It has also become a critical resource for situational awareness and event monitoring. In computational social science, gathering and analyzing data is integral to research that uses social media data to answer specific questions. The ubiquity of social media usage has led to a rise in the number of data services, tools, and analytics platforms for academic research and enterprise usage (Batrinca & Treleaven, 2015).

However, several challenges accompany social media data collection and analytics. These challenges include data scraping, cleansing, holistic data sources, protection, analytics, and visualization (Batrinca & Treleaven, 2015).

Despite these challenges, we recognize the value of the insights that social media data might provide, especially in a real-time setting. The social media revolution has created unprecedented opportunities to assess public responses and critiques of a multitude of social issues and events; this data can be provided and evaluated in real-time. We can use data science practices to harness this information to identify trustworthy information, reduce false claims, and take actionable steps for research and practical purposes, such as real-time surveillance and monitoring.

Therefore, several tools and platforms have been developed to address the challenges of social media data analysis, such as: Netlytic, which analyzes social networks and summarizes large text volumes; Gephi, an open-source graph visualization and analysis software; and NodeXL, an open-source template for Microsoft Excel for network analysis (Bastian et al., 2009; Hansen et al., 2010; Quan-Haase & Sloan, 2022).

2.2. Misinformation Theory

Social media has significantly accelerated the spread of misinformation, making it easier for false information to reach a broad audience (Treen et al., 2020). Research has shown that health misinformation can spread quickly through social media (Vosoughi et al., 2018). In 2013, a tweet from a hacked *Associated Press* account falsely reported an injury to then-President Obama, causing a $130 billion drop in stock value within minutes (Rapoza, 2017). This highlights the influence of social media on the rapid dissemination of misinformation.

Recent research has focused on developing methods to track and analyze misinformation on social media. One example is bot detection, used to identify automated accounts that spread misinformation (Ferrara et al., 2016). Network analysis is also used for examining the structure of information diffusion networks to identify patterns of misinformation spread (Shao et al., 2018). Furthermore, machine learning and natural language processing (NLP) techniques are used to classify and detect misinformation in text content (Shu et al., 2017). Researchers have also analyzed fact-checking integration, where automated systems are combined with human fact-checkers to verify information (Hassan et al., 2017). To add to this recent research, this study presents a novel approach for using SMART 2.0 to conduct surveillance of social media activity to identify misinformation in tweets.

3. Methodology

The methodology of this article encompasses three key components: data collection, analysis techniques, and user interface design. This section details our approach to gathering social media data despite recent restrictions on application programming interfaces (APIs). We describe here the machine learning and NLP techniques employed for data analysis and misinformation detection, and outline the interactive visualization features developed to support real-time decision-making. Our methodology prioritizes both technical robustness and user accessibility, with particular emphasis on overcoming data access limitations through innovative solutions like geolocation prediction and interactive machine learning models.

3.1. Data Collection

In the past, accessing real-time data from platforms like X (formerly Twitter) and Instagram was easy and free through APIs, but recent changes have restricted access. X's free API access was eliminated, and only costly paid plans are now available for meaningful data (Calma, 2023; Stokel-Walker, 2024). Similarly, Meta has limited data access on Facebook and Instagram to public profiles with over 25,000 followers, and it has a lengthy application process (Ryan-Mosley, 2023). Although scraping data using bots can be effective for small datasets, it is becoming increasingly difficult due to bot detection mechanisms, making this method unsustainable and arguably unethical (Chiapponi et al., 2022).

SMART 2.0 relies on geo-tagging to display data on the map. While SMART 2.0 uses methods to scrape Instagram and X for geo-specific posts, the limitations of scraping led to a reliance on paid APIs like X's. To overcome the shortage of geo-tagged data, SMART 2.0 is powered by a machine learning-powered geolocation prediction tool, which predicts the location of non-geo-tagged posts using deep learning models trained on geo-tagged data (Snyder, Karimzadeh, Chen, & Ebert, 2019). This increases the number of geo-tagged data available and increases SMART 2.0's value.

3.2. Analysis Techniques

SMART 2.0 handles multiple data types—textual, geographical, and environmental—each requiring specific analysis methods. SMART 2.0 offers classification, searching, and filtering tools for text-based data. Users can define categories using keywords, and SMART 2.0 supports complex criteria for filtering. Including deep learning features, SMART 2.0 can refine searches by removing irrelevant results, with machine learning models adapting to user preferences (Snyder et al., 2020). SMART 2.0 applies NLP techniques, such as latent Dirichlet allocation (LDA), to extract topics and classify data automatically (Snyder, Karimzadeh, Stober, & Ebert, 2019).

To identify and classify misleading information, SMART 2.0 utilizes interactive machine learning. Using a human-in-the-loop approach, the model corrects misclassifications by updating the model in real-time and it includes a built-in misinformation detection feature that classifies data as misinformation or not, using machine learning models trained on thousands of labeled tweets. This model classifies data points and can be interactively updated in real-time, allowing efficient data filtering, especially during emergencies, where vast amounts of information must be sifted for accuracy.

The misinformation models were initially pre-trained on a diverse dataset of over 10,200 tweets, including categories such as weather, news, traffic, Covid-19, security, trending, and random topics. This initial pool of tweets was manually labeled to validate the presence of misinformation, with care being taken to ensure a relatively balanced representation of both classes. The data was preprocessed using stop word removal, lemmatization, stemming, and then vectorization, which converts words or tokens into numbers that a computer can understand. Then, a random split of 80–20 was used to divide the data into training and testing subsets. The training samples were further leveraged to undergo a 5-fold cross-fold validation to train several models and identify the optimal model parameters. Finally, the unseen test data was used to evaluate each model's performance. This process achieved testing classification accuracies between 70–78%, with the passive-aggressive classifier performing best. The best model configuration is saved for future interactive learning.

During the real-time usage of SMART 2.0, users can provide feedback on possible wrongly classified tweets within the system, thereby triggering the lightweight partial training to update the label and train the model weights. This iterative and interactive misinformation training mechanism allows it to learn new patterns and improve its detection capabilities.

By supporting multilingual training data, including English, Italian, and Georgian, SMART 2.0 broadens its application across different languages. Figure 1 shows flowcharts of the machine learning model lifecycle used to implement misinformation classification. The top flowchart in Figure 1 shows how the misinformation classification model was constructed, trained, deployed, run, and updated; and the bottom flowchart illustrates the update process of a tweet's misinformation label by the user.

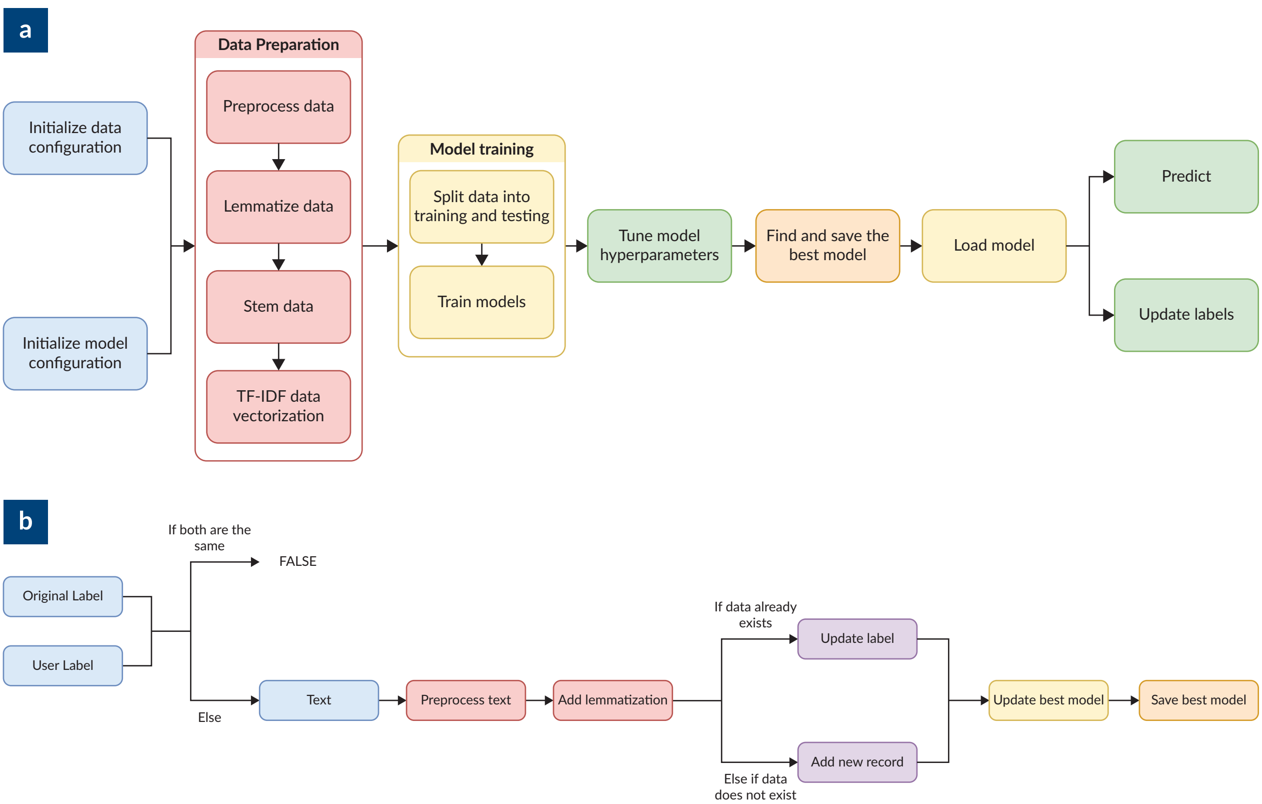

Figure 1. Misinformation flowchart: (a) steps to train, run, and update the misinformation classifier; (b) process of updating the misinformation label of a single tweet. Note: TF-IDF = term frequency–inverse document frequency.

3.3. User Interface

An interactive data visualization and analysis tool, SMART 2.0 enables the understanding of several data types for real-time decision-making as it communicates information using cutting-edge data visualization. From graph-based to map-based visualizations, the SMART 2.0 design allows us to optimally communicate the relevant data to the user. Figure 2 showcases a high-level overview of the user interface.

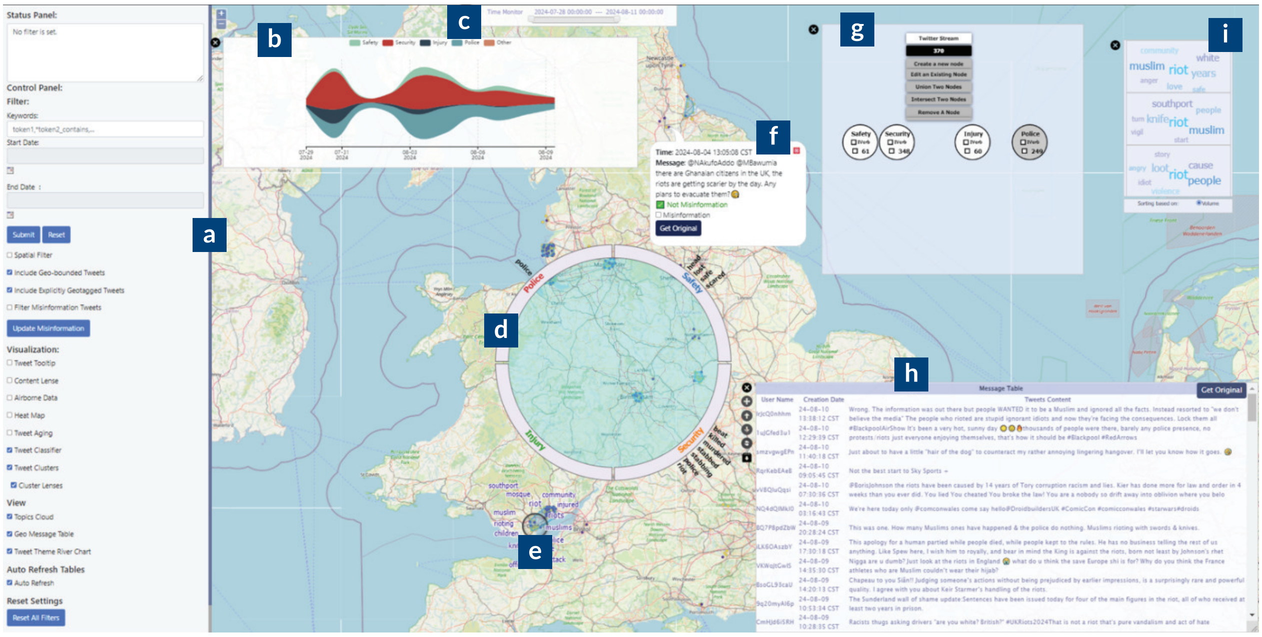

Figure 2. The SMART 2.0 user interface showing an array of panels and features of the toolkit: (a) side panel with controls and filters for various features; (b) theme river chart shows the number of tweets with time for each user-defined class; (c) the time monitor filters tweets based on the creation time; (d) the cluster lens clusters tweets by distance and displays any class-defining keywords present in any cluster under the lens; (e) the content lens displays common keywords in any tweet under the lens as the user hovers over the data; (f) the tweet tooltip displays details about the tweet such as time, message, and misinformation label, and it includes controls for fetching the original (non-translated) text and updating the misinformation label if necessary; (g) the classifier window is the classifier feature of the tool where the user can create, edit, delete, unionize, intersect, filter, and otherwise manage classes that are used elsewhere in the system; (h) the tweet data table shows the tweets displayed on the map and their details and allows the user to switch between the original (non-translated) text and the translated text of the tweets; (i) the topics cloud displays clusters of keywords generated using LDA, where each cluster of keywords represents a topic that exists in the current session, and clicking on a keyword filters the data on the map using the clicked keyword.

Figure 3 is a screenshot from SMART 2.0 showing the user interface for the misinformation classification of tweets.

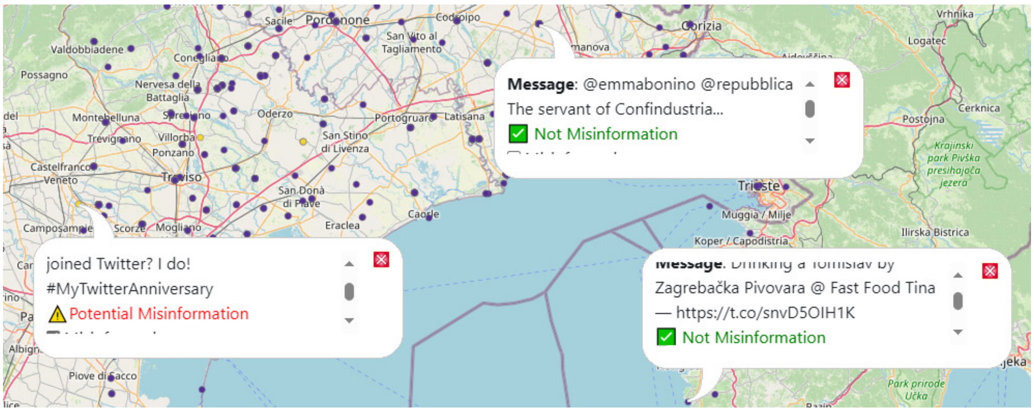

Figure 3. SMART 2.0 screenshot showing a sample of social data records in Venice with their tooltips showing their misinformation classification labels.

3.4. SMART Capabilities

SMART 2.0 enables the user to search for data and filter it using sophisticated full-text search functionality, as in Figure 4(a). A date and time filter is also available and works with keyword filtering. In addition to keyword and date filters, a spatial filter is available, which filters data on the map. The user draws a polygon on the map, and only data inside it is displayed, as seen in Figure 4(b). The misinformation filter allows users to toggle between hiding and showing the data classified as containing misinformation with the data labeled as containing misinformation being yellow, and the dark blue data points on the map being labeled as not containing misinformation.

The tooltip feature allows the user to click on a data point and display its data and metadata, as in Figure 4(c). The content lens displays common tokens among a data group as the user hovers over the map and it integrates with translation capability. The user can toggle back and forth between the original and translated tokens. Figure 4(d) shows the content lens in SMART 2.0. Moreover, the user can add multiple content lenses by freezing a content lens in the desired location. The heat map feature visualizes the spatial distribution of tweets. Figure 4(e) shows that the areas of higher intensity will appear yellow/greenish, and areas of lower intensity will appear blue.

SMART 2.0 features an interactive interface that allows users to search, classify, and filter tweets using "union," "intersection," and "not" filters, with users being able to create or edit classifiers, modify keywords, and apply these changes in real-time. As shown in Figure 4(f), a separate window manages filters, and changes are immediately reflected in the main interface. The "not" filter, displayed in grey, excludes tweets containing specific keywords.

Additionally, SMART 2.0 includes an intentional verb filter, which allows secondary filtering based on verbs that indicate human intent, such as "need," "want," or "attempt." Users can easily remove classifier nodes by selecting a node and clicking the "remove a node" button.

The theme river view is a time-series visualization that shows the number of tweets in different classifiers over the last two hours, providing an intuitive visualization of the temporal evolution of topics through a river metaphor. The theme river view is automatically linked with the classifiers. As shown in Figure 4(g), when the user creates a new classifier, it shows instantaneously in the theme river view. Users can click on classifiers in the legend to remove them from the theme river.

The tweets cluster view visualizes groups (clusters) of tweets located closely in geographic space. The clusters are visualized using a polygon-based representation, as shown in Figure 4(h). They are zoom-adaptive, which means when zooming in, the clusters split into small-scale ones. This feature aims to enable effective multi-scale exploration.

Figure 4. Visualization features: (a) keywords and date filters; (b) spatial filter; (c) tooltip; (d) content lens; (e) heat map; (f) classifier nodes window; (g) theme river chart; (h) cluster view; (i) cluster lens; (j) topic cloud.

The topic lens feature is a secondary feature in the cluster view and can be enabled by clicking the "cluster lens" option. According to Figure 4(i), the topic lens filters clusters within the lens and visualizes the keywords related to the current classifiers in a radial layout. The user can move the underlying map to investigate regions of interest while the position of the lens is fixed on the screen.

Figure 4(j) shows the LDA topic model window. LDA allows SMART 2.0 to automatically extract topics and define these topics using filter-enabled tokens, aiding in data comprehension.

4. Misinformation Tracking Case Study Using SMART 2.0

4.1. Introducing the Case Study and Its Significance

Our case study will focus on using SMART 2.0 to track and classify misinformation. The events that we studied took place in England between late July and early August 2024. The events primarily occurred in Liverpool, where riots took place, vehicles, shops, and buildings were set ablaze, and individuals were assaulted, harassed, and abused (Frayer, 2024; Otis, 2024). Similar riots took place in dozens of other towns across England, Wales, Scotland, and Northern Ireland (Ahmed, 2024). The riots were the aftermath of a fatal stabbing of three little girls in Southport (Lawless, 2024) that took place during a Taylor Swift-themed dance and yoga workshop in the northern English town on July 29, 2024, and caused the injury of several others (Lawless, 2024).

This case study was chosen because it is recent and relevant. The UK riots came about mainly due to misinformation and a perceived mistrust in legacy media (Frayer, 2024) and are documented to have been instigated by misinformation about the identity and religion of the perpetrator (Syed/London, 2024). While this case study effectively demonstrates SMART 2.0's capabilities, its scope is limited to a specific event. Future studies should explore broader applications, such as public health misinformation or disaster response, to validate the tool's versatility. The fact that misinformation was instigated in the social media realm makes this case study an appropriate choice for studying misinformation using SMART 2.0.

In this case study, we will attempt to use SMART 2.0's visualization, data science, and machine learning features to understand and track any misinformation associated with this event. To understand the riots and the events, we collected tweets from around Liverpool, Manchester, and other towns in the UK. We will use SMART 2.0's interactive misinformation classification and tracking capabilities to showcase its features.

The perpetrator's name and details were initially not disclosed to the public, given that he was a minor. Despite that, posts spread rapidly online, claiming he was a Muslim and an illegal migrant. The spread of this unfounded information led to riots against Muslims and migrants across the UK. Amid the riots and the violence, the court decided to release the name of the perpetrator to calm far-right public opinion and decrease violence against Muslims and other minorities.

The perpetrator's name was Axel Rudakubana, a 17-year-old Chirsitian UK-born teenager of Rwandan descent. He was not Muslim nor a migrant. Despite these facts, misinformation continued spreading across social media that he was Muslim and a migrant and was given the name "Ali al-Shakati" by social media users online without an official source.

4.2. Time Frame and Geographical Scope

The UK riots, which lasted from July 29 to August 5, 2024, were sparked by the stabbing of three young girls in Southport and spread across various towns in England, Wales, and Northern Ireland. Influential social media figures, such as Andrew Tate and Tommy Robinson, fueled the riots by falsely claiming the attacker was a Muslim immigrant, exacerbating tensions and inciting violence. Even after the perpetrator's identity was revealed, far-right narratives persisted, promoting distrust in the media and justifying continued violence. The riots led to the destruction of mosques and businesses and assaults on police, with over 400 people arrested (Syed/London, 2024).

In this case study, we will look at tweets from July 28, 2024, to August 10, 2024, from various towns in the UK, such as Liverpool, London, Cardiff, Leeds, and more. Our goal is to track the misinformation on X during the initial phase of the riots and then see how misinformation might have changed after the perpetrator's identity was revealed.

4.3. Case Study Methodology

4.3.1. Data Collection

SMART 2.0 was set up for this case study by creating a new historical event in the system and then loading the event into a new session. We use SMART 2.0's features to filter data, locate data points, and explore tweets. We also used the following features to study this case: event creation, event loading, keyword filtering, spatial filtering, date filtering, keyword filtering, tooltip, heatmap, and more. We also used SMART 2.0's interactive misinformation detection and classification capabilities to identify and track misinformation.

The system initially loads pre-trained misinformation classification models fine-tuned in previous iterations as the base model. The system is designed to be interactive and continuously improving, featuring a client-server architecture that allows users to provide feedback on the model's classifications in real-time. When users encounter a tweet, they can see the model's classification through a tooltip and correct any misclassifications, which then feed back into the model through partial fitting—an efficient technique that allows the model to learn from new data without complete retraining. This user feedback loop helps improve the model's accuracy over time, though one noted limitation is the need to preserve user-provided labels better when the model undergoes complete retraining. The system also includes practical features like the ability to filter out content classified as misinformation and batch update classifications across multiple posts in a single session.

Our data collection process used a custom script to scrape X for location-based tweets across the UK, focusing on tweets containing keywords related to the Southport incident and subsequent riots. The script collected tweets from July 28 to August 10, 2024, gathering a total of 370 tweets. The dataset for this case study is limited in size (370 tweets), which may not fully capture the breadth of misinformation surrounding an event of this scale. Furthermore, the reliance on geo-tagged tweets introduces biases toward users who enabled location sharing, potentially excluding significant portions of the population. Future work should explore methods to scale SMART 2.0 to larger datasets and reduce biases introduced during data collection. Another limitation is the potential oversaturation of certain hashtags or keywords, which may skew the dataset towards narratives. These limitations should be considered when interpreting the results of our analysis.

4.3.2. Initial Misinformation Classification Results

Figure 5 shows the initial misinformation classification of the tweets in Southport (a) and Liverpool (b). As shown in Figure 5, many of the data points in Liverpool and Southport have been classified as containing misinformation. This represents the initial classification result of the model without any input from the user. Initially, the data in this case study is foreign to the misinformation classification model.

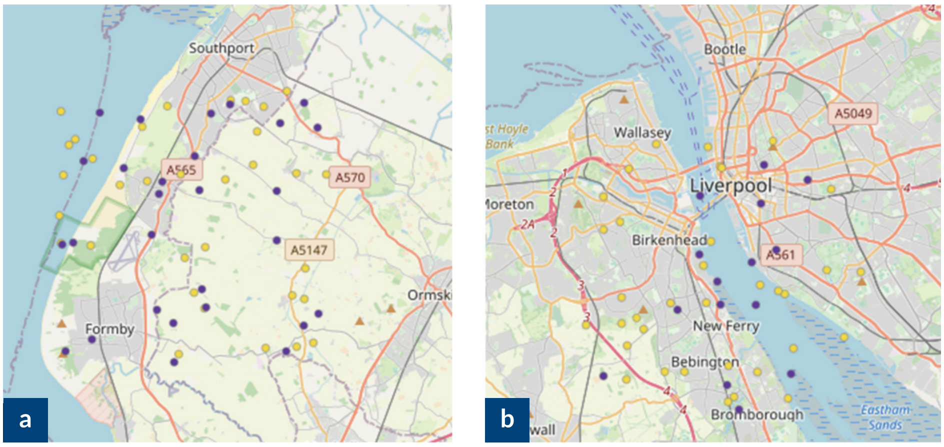

Figure 5. Misinformation-labeled data (yellow) in (a) Southport and (b) Liverpool.

4.3.3. Using Interactive Machine Learning for Refining Classifications

SMART 2.0's misinformation capabilities can be refined to specific events and use cases depending on the user's views and conception of what misinformation entails. Since we have established the base truth from multiple trustworthy and official sources that the perpetrator is not a Muslim and not a migrant, then we searched for and selected tweets that claim that he is Muslim. However, our model mislabeled them as not containing misinformation, so we then corrected them and updated the labels for all the data.

After identifying a few similar mislabeled tweets and correcting the model's predictions, we updated the data labels. After updating the misinformation label of 50% of the tweets, we got the model fine-tuned to the data in the current session. We used the fine-tuned model to update the labels for all the data. We found better accuracy in detecting misinformation in tweets that still asserted that the perpetrator was a Muslim. The initial accuracy of the model without any updates was 53.7%. After updating the misinformation labels of 50% of the tweets (185 tweets), the accuracy increased to 82.7%. We chose to update 50% of the data because it is a midpoint that provides enough data for the model to adapt to the case study. We conducted an experiment to evaluate the following hypothesis: the more the user updates the misinformation classification model, the better its accuracy will be. We will discuss this experiment in the next section of the article.

4.4. Key Findings

The findings regarding misinformation are according to the misinformation labels as classified by the misinformation classifier after updating the labels of 50% of the tweets.

4.4.1. Patterns in Misinformation Spread

We have noticed that much misinformation is concentrated in Liverpool and Newport, near Cardiff, where the riots against Muslims took place. Figure 6a and Figure 6c demonstrate this trend.

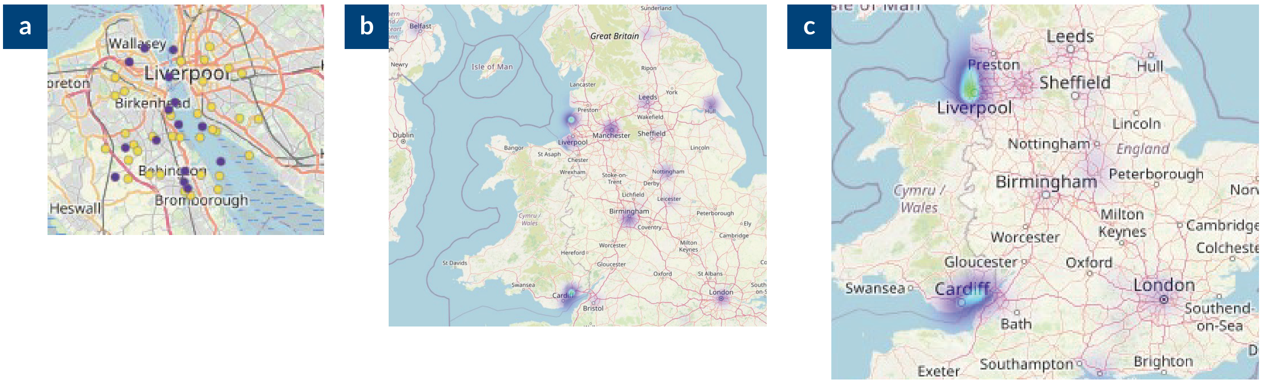

Figure 6. Patterns of misinformation distribution: (a) prevalence of misinformation in Liverpool according to the misinformation classifier; (b) concentration of tweets that do not contain misinformation in UK cities where riots took place; (c) heat map of the tweets that were classified as containing misinformation by the model. Note: Yellow dots are tweets classified as misinformation by the classifier.

On the other hand, in Southport, where the stabbing took place, we find there is a concentration of tweets that do not contain misinformation. Figure 6b shows a heatmap of the spatial distribution of tweets that were classified as not containing misinformation. This trend illustrates the premise that misinformation about a spatially located event, such as in this case study, is less common near the event. We can see that north of Liverpool, where Southport is, contains a high concentration of data that does not contain misinformation. Interestingly, we also notice that Newport, located northeast of Cardiff, contains a high concentration of tweets that were classified as containing misinformation and those that were classified as not containing misinformation by the model. Table 1 below shows the distribution of tweets in multiple major cities in the UK.

During the exploration of the tweets, we found that there were not many tweets about riots right after the fatal stabbing on July 29, 2024. Instead, the tweets that talked about riots started pouring in after a particular incident took place that sparked the riots against a mosque in Liverpool. Figure 7 shows a line chart that shows the number of tweets from July 29, 2024, to July 31, 2024.

Table 1. Tweets in each city, including the number of misinformation tweets.

City	Newport	Southport	Liverpool	London	Manchester	Birmingham	Cardiff	Leicester	Hull	Belfast	Leeds	Sunderland	Blackpool	Plymouth	Bristol
Total Tweets	102	61	50	19	19	17	14	11	10	9	6	5	4	4	3
Not Misinformation	69	47	14	16	17	17	14	8	8	9	5	5	4	4	3
Misinformation	33	14	36	3	2	0	0	3	2	0	1	0	0	0	0

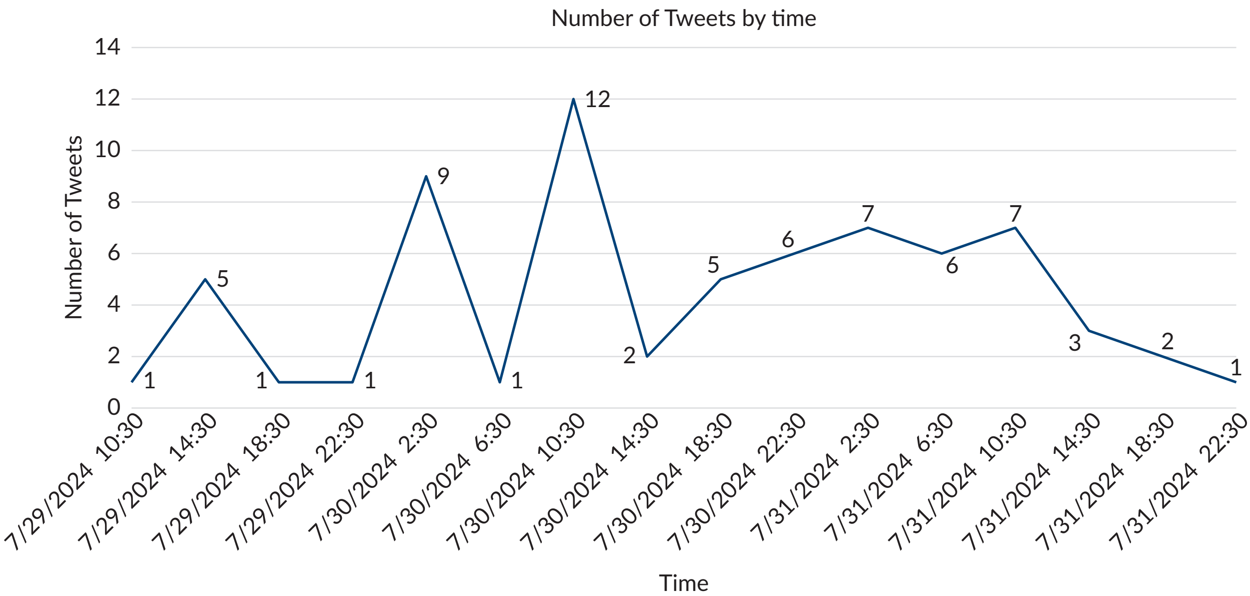

Figure 7. Line chart showing the incidence of tweet data from July 29, 2024, to July 31, 2024.

We noticed a pattern where the number of tweets about riots increased noticeably after an incident where a man with a knife was arrested in Liverpool near a vigil for the girls killed in the stabbing incident in Southport. In the tweets, many X users identify the person as a Muslim. However, according to the BBC, the man arrested with a knife near the vigil is called Jordan Davies and is not a Muslim (O'Neill & PA News, 2024). Instead, he was on his way to joining a mob that was about to stir up chaos in the area (O'Neill & PA News, 2024).

From Figure 7, we can see that there were only seven tweets in total right after the stabbing incident which took place on July 29. However, the number of tweets increased very quickly on July 30 to a total of 35 tweets. We attribute this increase to the incident where a man with a knife was arrested near the victims' vigil near Southport. Misinformation spread on X, claiming that the man with the knife was Muslim, but later reports from the BBC clarified that he is neither Muslim nor far-right (O'Neill & PA News, 2024).

After updating the misinformation model with the true labels, we identified a set of keywords that are common between the tweets that the model labeled as containing misinformation. Figure 8 displays a word cloud that includes these keywords.

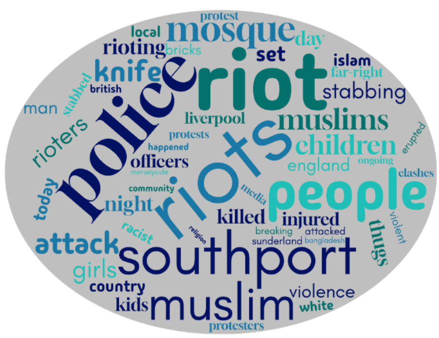

Figure 8. Word cloud with the most common keywords in the tweets that the model labeled as containing misinformation.

4.4.2. Effectiveness of Misinformation Identification

We have observed that SMART 2.0 initially identified misinformation without human input, with the initial accuracy being 53.7%. Some of these errors were remedied by correcting the labels of some of these tweets and then updating the labels for all the data using the human-in-the-loop approach.

Evaluating the interactive misinformation classification model in this case study is imperative. We want to assess the correctness and accuracy of the misinformation classification model as the user interactively updates the misinformation labels of the data. We hypothesize that the more the user corrects the misinformation labels of the data, the higher the model's accuracy would be at classifying the data's misinformation labels. To test this hypothesis, we conducted an experiment where we manually labeled the tweets in the case study for misinformation using a set of ground truths and guidelines.

We used the following ground truths: the Southport stabbing suspect is not a Muslim; the Southport stabbing suspect is not an immigrant; and the Southport stabbing suspect is Christian and was born and raised in the UK. In regards to guidelines, we used the following in the labeling process: tweets with racist, rude, and/or inappropriate undertones are not necessarily labeled as containing misinformation; and tweets that oppose ground truths are labeled as containing misinformation.

We manually labeled all the 370 tweets in this case study using the ground truths and guidelines described above. We wrote a computer program that starts with the un-updated model with the dataset being split into four equal parts. We used a stepwise iterative training approach to allow the models to learn new patterns related to the case study. In each iteration, a subset of the data is used to train and update model weights, followed by an evaluation of their performance. This process is repeated at 25%, 50%, 75%, and 100% of the training data. Figure 9 shows a graphical representation of the results.

The computer program that we created simulates user actions in SMART 2.0 when the user updates the misinformation label of a given data point. The program also simulates the retrieval of the label of any given data point in the user's session. Doing so enables us to conduct this experiment automatically, making it drastically faster to run, alter, and improve.

As seen in Figure 9(a), we find that as the user updates the misinformation classification model, it becomes more accurate, as seen in the increase in accuracy per iteration. The initial accuracy of the model (without any human input) was around 0.53. Although accuracy improved from 53% to 95.4% through iterative updates, this evaluation was conducted only on the case study dataset, which limits the generalizability of the results. Future work should include cross-validation on larger, unseen datasets to ensure robustness and to test the model's ability to generalize to diverse scenarios. Figure 9(b) also shows a similar trend of increasing precision, recall, and F1 score, meaning that the model's performance improved significantly with each iteration. Figure 9(c) also shows improvements in the model's classification results by the apparent decrease in the number of false positives and false negatives the more the model is updated.

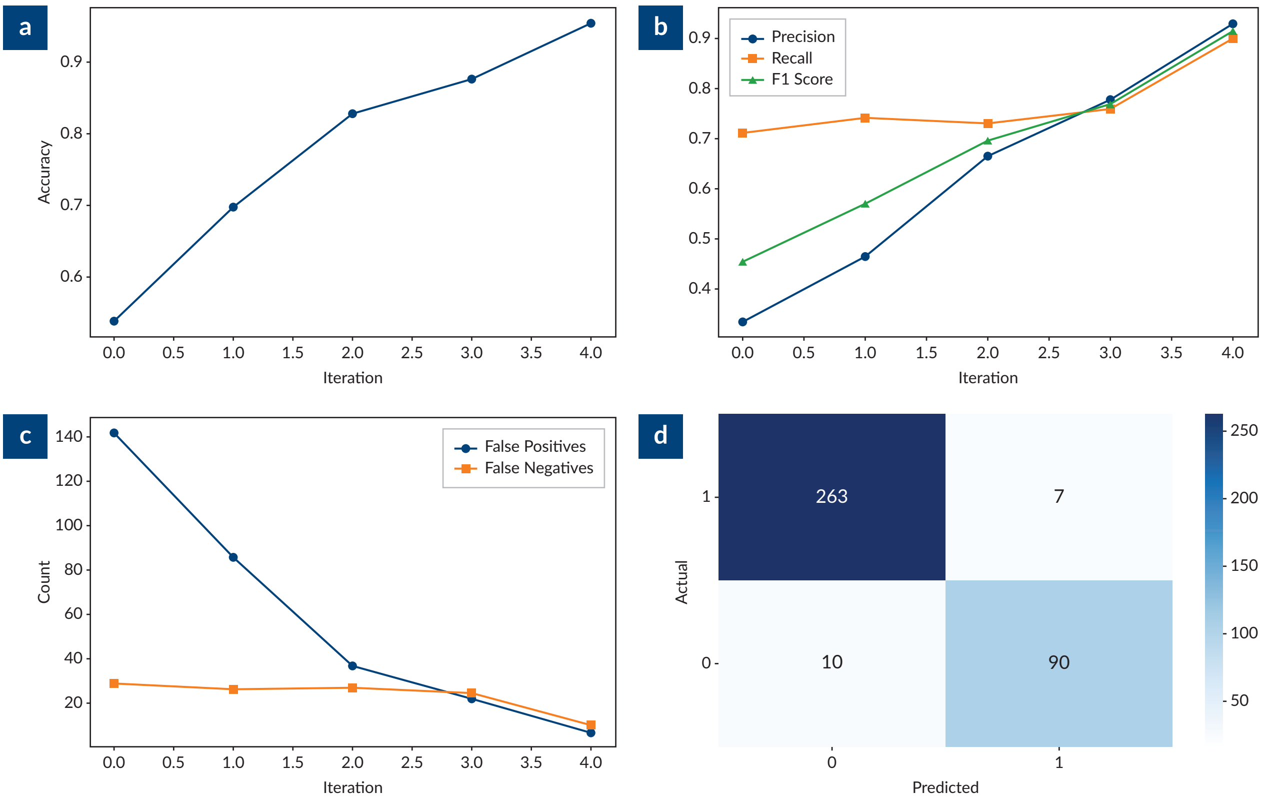

Figure 9. Evaluation results of the interactive misinformation model: (a) accuracy of the model for each quarter of the dataset; (b) model's precision, recall, and F1 score at each iteration; (c) error rates, including the number of false positives and the number of false negatives at each iteration while updating the model; (d) final confusion matrix of the model, created after the model was updated with the true labels.

4.4.3. Notable Results

We were surprised to discover that a lot of the data points and tweets were about the man with the knife who was arrested near the vigil in Southport. We assumed that most of the riots and the misinformation were due to the initial fatal stabbing of the three girls, but that turned out not to be the case.

5. Discussion and Future Improvements

The case study of the UK riots in 2024 demonstrates the powerful capabilities of SMART 2.0 in tracking and analyzing misinformation spread during a critical event. The study revealed how quickly false information about a perpetrator's identity and background spread on social media, particularly on X, highlighting the need for real-time monitoring and analysis tools like SMART 2.0. Contrasted with more accurate information in Southport and Newport, the concentration of misinformation in Liverpool shows how misinformation can have localized effects, underscoring the importance of SMART 2.0's geospatial analysis capabilities. We found a trend that illustrates the premise that misinformation about a spatially located event is less common near the event. However, this premise was not tested well and should be expanded on in future research.

The spike in misinformation following the arrest of a man with a knife near the victims' vigil illustrates how secondary events can amplify and reshape the dissemination of misinformation, and SMART 2.0's temporal analysis features were crucial in identifying this pattern. It proved valuable in improving misinformation classification accuracy, and the ability to refine the misinformation classification model through user input was essential for context-specific misinformation. Our experiment showed that the more the user updates and interacts with the misinformation classification model, the better its accuracy and performance. However, one of the main downsides of our analysis and experiment is the low number of data points. In future research, we need to have more data sources and data points to evaluate the performance of the misinformation classifier effectively.

The case study demonstrated the importance of cross-referencing social media data with official sources, like BBC reports, to establish ground truth and identify misinformation accurately. These findings underscore the complex nature of misinformation spread during crisis events and the value of tools like SMART 2.0 in providing real-time insights to researchers, journalists, and policymakers. Based on the experiences and insights gained from this case study, we identified several areas for future improvement of SMART 2.0, including:

1. Enhanced language models: Develop more sophisticated NLP models to better understand context and implicit references in text-based data.
2. Cross-platform integration: Expand SMART 2.0's capabilities to integrate data from multiple platforms, providing a more comprehensive view of misinformation dissemination.
3. Automated fact-checking: Implement automated fact-checking features to cross-reference claims in social media posts with reliable news sources and official statements in real time.
4. Trend prediction: Develop predictive models that can forecast potential misinformation trends based on early signals in social media data.
5. User network analysis: Incorporate features to analyze the networks of users spreading misinformation, identifying key influences and bot networks.
6. Extend multilingual support: Expand language support to better track misinformation in diverse linguistic contexts, particularly in multilingual regions.
7. Integration with traditional media monitoring: Develop features to correlate social media misinformation trends with coverage in traditional media outlets, such as TV stations and news sites.

6. Conclusion

The SMART 2.0 has demonstrated its effectiveness in tracking and analyzing misinformation during critical events, as seen in the case of the 2024 UK riots. SMART 2.0's insights into the spread of false information, its geographical patterns, and the role of trigger events highlight its data visualization and multi-dimensional analysis capabilities. The tool adapts quickly through user feedback, supports multiple languages, and helps researchers understand the complex nature of misinformation.

However, the case study also revealed potential areas for improvement, such as enhanced language understanding and misinformation classification, and increasing the number of data points. As social media continues to shape public discourse, tools like SMART 2.0 are crucial for combating misinformation and supporting evidence-based decision-making and real-time monitoring.

Acknowledgments

This work is supported in part by the North Atlantic Treaty Organization (MYP-REACT G5812) and the Data Institute for Societal Challenges, University of Oklahoma. The authors would also like to acknowledge the support of Wagner Rosa for the initial creation of the misinformation model and the English editing support by Dr Yessenia Torres.

Funding

SMART 2.0 was partially funded by NATO REACT, Grant # MYP-REACT G5812.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability

Data used in this study will be made available upon reasonable request, provided the request is legitimate and academically relevant. Contact the corresponding author for more details.

References

Ahmed, J. (2024, August 6). Mapped: Violent protests grip the country with fears of more to come. *The Independent*. https://www.the-independent.com/news/uk/crime/uk-riot-locations-protests-map-far-right-b2591874.html

Ards mosque community "nervous" after overnight attack. (2024, August 11). *BBC News*. https://www.bbc.com/news/articles/c984dzxg6eko

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*, *3*(1), Article 1. https://doi.org/10.1609/icwsm.v3i1.13937

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: A survey of techniques, tools and platforms. *AI & SOCIETY*, *30*(1), 89–116. https://doi.org/10.1007/s00146-014-0549-4

Calma, J. (2023, May 31). Twitter just closed the book on academic research. *The Verge*. https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research

Chiapponi, E., Dacier, M., Thonnard, O., Fangar, M., Mattsson, M., & Rigal, V. (2022). An industrial perspective on web scraping characteristics and open issues. In Y. Amir & C. Nita-Rotaru (Eds.), *2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks—Supplemental Volume (DSN-S)* (pp. 5–8). IEEE. https://doi.org/10.1109/DSN-S54099.2022.00012

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96–104. https://doi.org/10.1145/2818717

Frayer, L. (2024, August 5). Misinformation online fueled all-out race riots in the United Kingdom. *NPR*. https://www.npr.org/2024/08/05/nx-s1-5063345/misinformation-online-fueled-all-out-race-riots-in-the-united-kingdom

Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Morgan Kaufmann.

Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017). Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster. In S. Matwin, S. Yu, & F. Farooq (Eds.), *KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1803–1812). ACM. https://doi.org/10.1145/3097983.3098131

Ilia State University. (n.d.). *Network for alerting and managing public safety and resilience—REACT*. https://iliauni.edu.ge/en/siaxleebi-8/gonisdziebebi-346/network-for-alerting-and-managing-public-safety-and-resilience-react4.page

Lawless, J. (2024, July 31). British police charge 17-year-old with murder over a stabbing attack that killed 3 children. *AP News*. https://apnews.com/article/uk-stabbing-attack-southport-far-right-violence-a2e43d0d49776c138790d083713873f7

O'Neill, L., & PA News. (2024, August 9). Man caught with knife near vigil for Southport stabbing victims jailed. *BBC*. https://www.bbc.com/news/articles/c4gz79dln5xo

Otis, J. (2024, August 9). Covering the U.K. riots amid disorder and misinformation. *The New York Times*. https://www.nytimes.com/2024/08/09/insider/uk-riots.html

Quan-Haase, A. & Sloan, L. (Eds.). (2022). *The Sage handbook of social media research methods* (2nd ed). Sage.

Rapoza, K. (2017, February 26). Can "fake news" impact the stock market? *Forbes*. https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market

Ryan-Mosley, T. (2023, November 21). Meta is giving researchers more access to Facebook and Instagram data. *MIT Technology Review*. https://www.technologyreview.com/2023/11/21/1083760/meta-transparency-research-database-nick-clegg

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*(1), Article 4787. https://doi.org/10.1038/s41467-018-06930-7

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36. https://doi.org/10.1145/3137597.3137600

Snyder, L. S., Karimzadeh, M., Chen, R., & Ebert, D. S. (2019). City-level geolocation of tweets for real-time visual analytics. In S. Gao, S. Newsam, L. Zhao, D. Lunga, Y. Hu, B. Martins, X. Zhou, & F. Chen (Eds.), *GeoAI '19: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery* (pp. 85–88). ACM. https://doi.org/10.1145/3356471.3365243

Snyder, L. S., Karimzadeh, M., Stober, C., & Ebert, D. S. (2019). Situational awareness enhanced through social media analytics: A survey of first responders. In *2019 IEEE International Symposium on Technologies for Homeland Security (HST)* (pp. 1–8). IEEE. https://doi.org/10.1109/HST47167.2019.9033003

Snyder, L. S., Lin, Y.-S., Karimzadeh, M., Goldwasser, D., & Ebert, D. S. (2020). Interactive learning for identifying relevant tweets to support real-time situational awareness. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 558–568. https://doi.org/10.1109/TVCG.2019.2934614

Stokel-Walker, C. (2024, February 27). Under Elon Musk, X is denying API access to academics who study misinformation. *Fast Company*. https://www.fastcompany.com/91040397/under-elon-musk-x-is-denying-api-access-to-academics-who-study-misinformation

Syed/London, A. (2024, August 5). How online misinformation stoked anti-migrant riots in Britain. *TIME*. https://time.com/7007925/misinformation-violence-riots-britain

Treen, K. M. I., Williams, H. T. P., & O'Neill, S. J. (2020). Online misinformation about climate change. *WIREs Climate Change*, *11*(5), Article e665. https://doi.org/10.1002/wcc.665

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

About the Authors

Mahmoud Mousa Hamad is a software developer at the Data Institute for Societal Challenges at the University of Oklahoma. Originally from Palestine, Mousa Hamad earned his BS and MS in computer science from the University of Oklahoma. His expertise spans full-stack web applications development, database management, mobile applications development, and machine learning model creation. For the past three years, he has served as the lead engineer on the Social Media Analytics and Reporting Tool (SMART), and he is currently focused on developing a new version that will make the tool more modern, scalable, and user-friendly.

Gopichandh Danala is a research scientist at the Data Institute for Societal Challenges at the University of Oklahoma. He received an MS and PhD in electrical and computer engineering from the University of Oklahoma, focused on medical informatics in developing computer-aided design (CAD) tools to improve diagnosis and patient care using computer vision, machine learning, and decision-making environments. His research interests are healthcare analytics, medical informatics, data science, and visual analytics to develop decision-making systems that address real-world problems and enhance societal well-being.

Wolfgang Jentner is a research scientist at the Data Institute for Societal Challenges at the University of Oklahoma. He received his PhD in visual analytics, a research discipline at the intersection of machine learning, interactive visualization, and human–computer interaction, from the University of Konstanz, Germany. His research interests include social media text analysis, explainable AI, visual analytics for civil security and critical infrastructure, and One Health.

David Ebert is a Gallogly chair professor of electrical and computer engineering (ECE) and director of the Data Institute for Societal Challenges at the University of Oklahoma. He is an IEEE fellow, recipient of the 2017 IEEE Computer Society vgTC Technical Achievement Award, and an adjunct professor at Purdue University. He performs research in visual analytics, novel visualization techniques, interactive machine learning and explainable AI, human–computer teaming, advanced predictive analytics, and procedural abstraction of complex, massive data.