Enhancing Public Safety in Smart Cities through LLM-based Misinformation Detection

Saeed H. Al-Shamrani
Dept. of Computer Science
University of Bisha, Saudi Arabia
University of Alabama at Birmingham
Birmingham, AL 35294-1241
shalsham@uab.edu

Ragib Hasan
Dept. of Computer Science
University of Alabama at Birmingham
Birmingham, AL 35294-1241
ragib@uab.edu

***Abstract*—The paradigm of smart cities has great potential to improve urban management. Using modern information and communication technologies (ICTs), smart cities can efficiently improve the quality of life of their inhabitants in multiple domains, including public safety. While smart cities are constantly employing technologies such as Internet of Things (IoT) sensors and cameras to make informed decisions, the lack of infrastructure and resources has led them to resort to crowd-sourcing data from citizens. This increasing dependence on a wide range of data sources also creates a wide range of issues. One particular area of concern is the difficulty of ensuring the trustworthiness of every piece of information. Data in smart cities can be susceptible to a variety of issues that can compromise the trustworthiness of data, hence, decisions can be flawed, which can indeed affect public safety. While information obtained from ICTs can be guarded using technical solutions, information provided by citizens can contain misinformation. Furthermore, most public safety platforms depend on users' reports through designated apps or social media, which allows for the possibility of misinformation, hence compromising the trustworthiness of data. In this paper, we propose a fine-tuned Large Language Model (LLM) using misinformation datasets to detect misinformation in texts within the context of public safety. We achieve an accuracy of 98% in detecting misinformation. We also highlight the open issues in this area and briefly discuss future directions.**

***Index Terms*—Smart Cities, Large Language Models, Misinformation detection**

I. Introduction

Smart cities leverage Information and Communication Technologies (ICTs) to optimize urban management. By using advanced intelligent technologies, improvements can be seen across a wide range of applications in efficiency, sustainability, and even safety. While these smart technologies enhance the citizens' quality of life, their safety is a crucial requirement. Public safety is one of the most critical components in today's metropolitan areas, which include protecting people's lives from threats like crime, accidents, and natural hazards.

In recent years, the dissemination of safety alerts has undergone a transformation. Intelligent systems now collect and aggregate data from diverse sources, process it using advanced analytics, and deliver timely alerts to the public. These systems rely on inputs from IoT sensors, CCTV cameras, crowdsourced data, and user-generated reports via dedicated mobile applications and social media platforms.

However, the trustworthiness of such data is a huge concern since these applications affect people's safety. Data trustworthiness is a concept concerning the reliability and accuracy of data, hence, it is a vital component in informed decision-making. While data from ICT infrastructure can be safeguarded through technical measures, citizen-contributed information is more vulnerable to misinformation. Ensuring the trustworthiness of this data is essential, as flawed information can lead to poor decision-making and jeopardize public safety.

There have been a lot of approaches to evaluate and assess the trustworthiness of crowdsourced data against misinformation, such as machine learning, deep learning, and blockchain. However, recent technological advancements in the sector of Large Language Models (LLM) have made them superior in a variety of applications. These models presented high potential in advanced text processing, which made them suitable for processing and analyzing textual data from all kinds of sources. LLMs can understand context, content, work with fewer to no features, and comprehend unstructured data. We argue that a fine-tuned LLM can be very effective and efficient to ensure the trustworthiness of data in smart cities. Their adaptability and contextual understanding make them well-suited for detecting misinformation in public safety applications.

Contributions: The contributions of this paper are as follows:

1) We present a model that shows the promise that LLMs hold for misinformation detection in smart cities' public safety applications.
2) We fine-tune compact LLMs that are suitable for deployment in low-resource infrastructure, such as the edge in smart cities, using publicly available misinformation detection datasets.
3) We also discuss and highlight the open challenges and limitations in this domain.

Organization: The rest of the paper is organized as follows: in Section II, we provide a brief background about smart cities, crowd-sourcing, public safety applications in smart cities, and an overview of misinformation detection techniques. In Section III, we explore the issues of misinformation in existing literature and how it affects the trustworthiness of data. In Section IV, we propose our model, discuss our experimental setup, demonstrate the result, and provide a brief discussion. In Section V, we highlight the open challenges in misinformation detection techniques in smart cities. We conclude and discuss future directions in Section VI.

II. Background

A. Smart Cities

Smart cities are cities that utilize Information and Communication Technologies to enhance urban management [9]. There is a wide range of technologies involved in smart cities, including but not limited to IoT, Cloud & Edge computing, Blockchain, AI, and Big Data. The applications are diverse, covering almost every aspect of any city, such as energy management, traffic control, waste recycling, and even governance. The ultimate goal is to improve efficiency, utilize resources, and, of course, enhance the citizens' quality of life [9].

B. Crowdsourcing

While smart cities heavily rely on IoT sensors to gather data, there has been a trend within recent years towards collecting data from people. Smart city citizens can act as "human sensors" providing data for various smart city applications [21]. The proliferation of smartphones and Internet access allowed citizens to actively contribute data and participate in the development and operation of smart cities [16]. There are other factors as well that led to the evolution of crowdsourcing. One significant factor is the rise of social media platforms, which established a channel for information sharing and community engagement. Additionally, recent advances in data analytics have helped process large volumes of information.

C. Public Safety and Misinformation Threats

Smart cities aim to enhance citizens' quality of life; hence, public safety is a crucial element. Public safety conditions include any event that directly affect the lives of the city inhabitants, including crimes, disasters, and diseases. While smart cities utilize many objects to monitor and detect abnormal activities, it is generally challenging to monitor everything. Citizen participation becomes a necessity in this case [16]. Public safety and security applications leverage crowdsourced information to identify, respond to, and prevent safety threats in urban environments. Since information exists in this application that highly impacts people's lives, it is necessary to ensure every piece of information is trustworthy in order to make the right and needed response. The one particular issue here is that citizens' feed of information might contain misinformation, which not only wastes resources, but also might delay the help for those who really need it during emergencies [7]. Hence, it is important to combat misinformation on these platforms.

While the information can be generated using data from various data sources, such as sensors and cameras, citizens can actively post feeds to the smart application or through social media. The possibility for citizens to share misinformation through these platforms is a major concern [15], [21].

D. Misinformation Detection

Public safety is a major aspect of smart cities. Safety platforms can be used for emergency reporting, incident tracking, and community policing, which are integral parts of smart cities initiatives. Public safety applications rely on data from various sources such as the city's ICT infrastructure and official reporting channels. These sources can be limited as resource limitations can affect the coverage limit of sensors and cameras, which encourages reliance on crowdsourced data. However, citizen participation comes with a wide range of trustworthiness issues, such as misinformation. These issues not only affect the reliability of these platforms but also waste resources, impact public trust, and pose safety risks. Furthermore, the scale of smart cities amplifies not only the opportunities but also the challenges of such applications. Below, we highlight misinformation detection techniques, and Figure 1 summarizes these techniques:

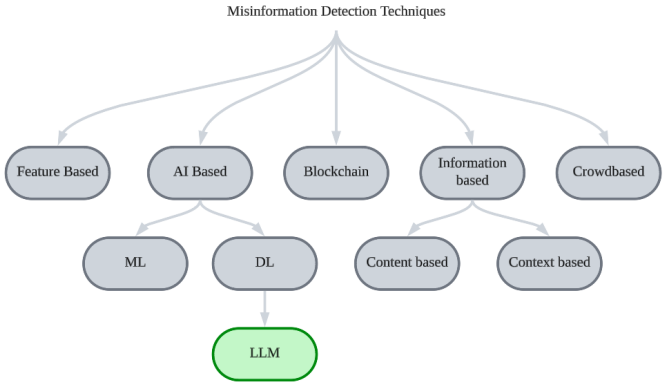

Figure 1. Misinformation Detection Techniques

1) **Feature Based** These methods depend on text analysis in order to identify misinformation through text features. These methods work by extracting linguistic patterns, statistical properties, and structural features. The reliance on a predefined set of features provides a straightforward approach to detect misinformation in texts. It is particularly effective when linguistic and statistical markers are consistent, however, it is not adaptable and can be challenging to work with dynamic and evolving texts [17].
2) **Artificial Intelligence-Based**
   *a) **Machine Learning-Based***
   It depends on algorithms trained on labeled datasets to classify new content as true or false. These algorithms include Support Vector Machines (SVMs), decision trees, random forests, and logistic regression. These methods rely on temporal, structural, and linguistic features to predict the authenticity of the text. While these approaches can offer simplicity and interpretability, they require extensive feature engineering, which lacks flexibility, and it can be challenging to handle complex texts [1], [17]. There is also an issue with scalability as performance is noticed to drop with noisy and high-dimensional data, such as texts that exist on social media [1].
   *b) **Deep Learning-Based Approaches***
   These methods utilize advanced neural network architectures to process complex and unstructured data to identify falsehood patterns. Deep learning Techniques include: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks

(LSTMs), and Graph Neural Networks (GNNs). These methods are more sophisticated, hence they outperform traditional methods and provide excellent analysis and work across various data formats such as texts, images, and videos. While these methods do not require extensive feature engineering, they demand substantial computing resources and lack interpretability [1], [17]. Large Language Models (LLMs) fall under this category, however, they can better understand linguistic patterns and contextual relationships [14], which makes them more effective in this specfici area.

3) **Blockchain and Decentralized Approaches** These methods utilize blockchain to verify information authenticity and enforce trustworthiness in digital content. Blockchain can also ensure transparency, verify provenance, and help establish trust in information systems. Blockchain-based solutions include verifying provenance to track and validate the origin of data [12], reputation-based assessment through utilizing smart contracts to ensure accountability [3], and also through rewards which act as an incentive to reward honest sources and penalize untrustworthy entities [10].
4) **Information Based**
   *a) **Content-Based***
   These methods focus on the intrinsic properties of the text itself to detect misinformation. They analyze linguistic features, semantic patterns, and stylistic elements within the content. Content-based techniques rely on the assumption that misinformation often exhibits detectable differences in language use, such as emotional polarity or deceptive phrasing. These properties make them effective for static text analysis but less adaptable to external factors like propagation dynamics [17].
   *b) **Context-Based***
   These methods emphasize the external environment surrounding the text, such as social interactions, user behavior, and propagation patterns. Context-based approaches work well in dynamic settings like social media, where misinformation's spread and user engagement provide critical clues. However, they require additional data beyond the text itself which makes them more complex [17]. content-based approaches are limited by their focus on text and lack of adaptability, while context-based approaches are constrained by data dependency, complexity, and potential misinterpretation of external signals.
5) **Crowd-Based**
   The crowd-based approach to misinformation detection integrates human input, such as crowdsourced evaluations or expert fact-checking. This approach shines in handling subjective or culturally specific content. They depend on human understanding beyond content or context-based techniques. However, the process is time-consuming and costly. It is also limited by the availability of evaluators. Human input can also be biased or inconsistent, which might complicate the validation process. Ultimately, this approach is less efficient and less scalable than automated techniques, despite having qualitative strength [17].

III. Related Works

Currin *et al.* in [6] addresses the issue of managing public safety in developing nations and highlights the potential of using mobile phones to crowdsource public safety reports from citizens. They argue that the dependence on automated sensors and quantitative data in smart cities might not be always feasible especially when resources are limited. They contributes a Smart City Qualitative Data Analysis (SCQDA) model to analyze unstructured citizen reports while emphasizing the challenges of Natural Language Processing (NLP) with semantics. Especially with colloquialisms and slang which could compromise the reliability of data analysis.

Huang *et al.* explored the use of open crime data and crowdsourcing to build a smart app for public safety [11]. Their works is based on the availability of open crime data through law enforcement agencies. On the other hand, they also address the limitations of such data, especially the lack of real-time accuracy. They drew a concept called "Nudging" to encourage public participation. It is basically a choice of architecture that predictably influences behavior. They have one design dilemma which is how to balance between protecting users privacy and ensuring the accuracy and quality of information.

Pereira *at al.* [19] proposed a mobile app design to improve public safety through the integration of open crime data, crowdsourcing, and users' personal information. The design requires users to form a "neighbors group" and approve new users using a vote-based approach. While the app supports community policing and include private and government initiatives, they state that citizens might be encouraged to intervene which could jeopardize their safety. There are also some concerns about the reliability of shared information,

Sarzaeim *et al.* in [23] addresses the demand for crime prediction and prevention by introducing Large Language Models (LLMs) integration into smart policing systems. They utilize various LLMs and methods to prove their effectiveness in crime classification using experimental scenarios. While their framework offers a foundation for developing LLM-assisted tools for smart policing, they also share the concerns of complexity in integrating LLMs into existing smart policing systems.

IV. Methodology

Large Language Models (LLMs) can effectively overcome the limitations within current misinformation detection techniques. They can capture semantics without the need for predefined features as in the case of feature based misinformation detection [8]. LLMs can also learn representations from raw texts unlike ML techniques which requires extensive feature engineering [27]. Additionally, LLMs can understand distant connections in text, learn from less labeled data and run efficiently using attention mechanisms. They can also combine different types of information easily, exceeding deep learning techniques [2]. While LLMs suffer from challenges like hallucinations and computational costs, their adaptability and

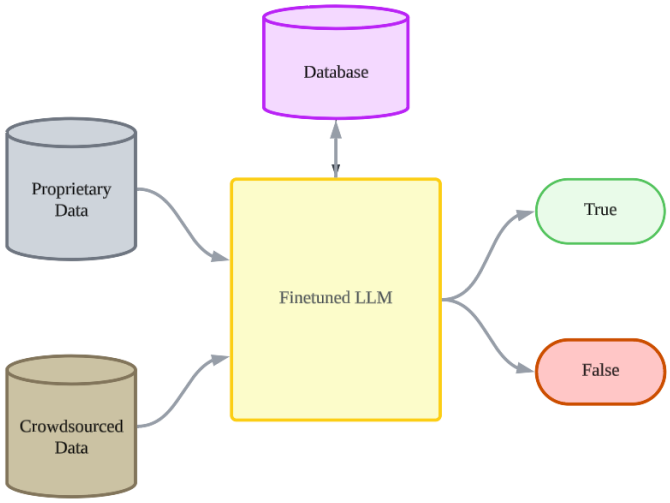

Figure 2. System Architecture

robustness make them a powerful solution for misinformation detection. In this section, we will explain the system model we used in our work, describe the datasets we utilized, show, and discuss the results.

A. Architecture and System Design

The system architecture as presented in Figure 2 explains the components of our model. We assume there are two modules responsible for data collection. The crowdsourcer collects data actively from social media. The crowdsourcers sweep social media websites looking for posts with special keywords. Those keywords are words that indicate something has happened in a given city. The second module is collecting data from public safety apps. The targeted data is the textual information given by the user when submitting an alert. Our classifier is simply a pretrained LLM which we finetuned using labeled misinformation datasets. More details about our choice of models and datasets are discussed later in this section. The outputs of the model are single labels which can either be "True" or "False". If the data is found to be trustworthy, an alert will be generated, otherwise, it is considered false.

B. Data Description

To fine-tune our model, we used various misinformation datasets that are publicly available online. While there are no datasets at least publicly available for the specific context of public safety, we tried to utilize datasets that are similar in presentation such as fake news datasets. It is generally difficult to find datasets for a very specific aspect. The other issue with broader context datasets is that they are often insufficient for advanced applications like those potentially tied to public safety due to issues like limited scale, lack of diversity, or absence of comprehensive features [26]. Despite not having datasets that are specific to the purpose of misinformation detection in public safety applications, we made sure to use only datasets that at least have some aspects concerning this purpose. Below are the datasets we used in this work. Table I has some characteristics of these datasets.

- **COAID Dataset**: contains a collection of texts related to COVID-19 obtained from news articles, social media websites, and fact-checking websites [5]. It has more than 5,000 claims labeled true, false, or other(uncertain). While it is considered a COVID-19 healthcare misinformation dataset, it has additional annotations covering various categories, including public safety.
- **LIAR Dataset**: contains around 12,800 short claims from a fact-checking organization called PolitiFact [28]. These claims are exclusively in a political context and labeled across six categories: "pants on fire" "false" "barely true" "half true" "mostly true" and "true".
- **PHEME Dataset**: contains more than 20,000 entries from Twitter previously (Now X) that are labeled either true or false [18]. It contains tweets that were posted during breaking news, some of which are rumors and some are not. This dataset covers a wide range of topics, including politics, public health, crime, and entertainment. This dataset is designed specifically to detect rumors on social media posts.

Table I
DATASET STATISTICS

Dataset	Entries	Labels	Reference
COAID	5,216	3	[5]
LIAR	12,836	5	[28]
PHEME	20,000	2	[18]

C. LLM Models

In our work, we fine-tuned three pre-trained Large Language Models. We exclusively chose models that are suitable for edge deployment in order to make our experiment realistic for smart cities. Below is a list of our pre-trained models. A summary of these models is presented in Table II.

Table II
COMPARISON OF DISTILBERT, MOBILEBERT, AND TINYBERT

Attribute	DistilBERT	MobileBERT	TinyBERT
Parameters (M)	66	25.3	14.5
Layers	6	24	4
Hidden Size	768	512	312
Inference Speed (rel.)	60% faster than BERT	5.5× faster than BERT	9.4× faster than BERT
Compression Method	Distillation	Distillation	Distillation
Use Case	General NLP	Mobile/Edge	Resource-Constrained

- **DistilBert:** DistilBert is a lightweight version of BERT, created by distilling the larger model to retain 97% of its performance while reducing size and computational requirements by 40% [22].
- **MobileBERT:** MobileBERT is another compact and distilled BERT variation. It is four times smaller and five times faster than BERT. It is designed for resource-limited devices with competitive NLP performance. MobileBert is optimized for mobile and edge deployment [24].
- **TinyBERT:** TinyBERT is an ultra-small BERT variant with 4-layers. It is seven times smaller and nine times

faster than BERT. It achieves 98% performance while being the fastest and most suitable for resource-constrained devices [13].

D. Experimental Setup

We fine-tuned our models for a single-label classification task using a standard training/test split. All inputs were tokenized with a maximum length of 512. Training was performed over 3 epochs with commonly used settings (batch size 8, warmup steps, weight decay). The main hardware was a *Tesla P100 GPU*. We selected the best-performing model based on validation results, without including a separate baseline for comparison.

E. Evaluation Metrics

Model performance was evaluated using accuracy, precision, recall, and F1-score via a classification report, with accuracy also computed separately as the mean of correct predictions. The formula for each metric used is shown in Table

Table III
EVALUATION METRICS FORMULAS

Metric	Formula
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1 Score	$\frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$

F. Results

Our three models (DistilBert, MobileBert, and TinyBert) showed almost similar performance across datasets. The COAID dataset (5,000 entries, 3 labels) yielded 0.95 accuracy. On the second dataset (12,000 entries, 5 labels), DistilBert and MobileBert achieved 0.74 accuracy while TinyBert performed lower at 0.67. All three models excelled on the third dataset (20,000 entries, 3 labels) with 0.97 and 0.98 accuracy. These findings suggest that these models handle varying classification tasks effectively, while model TinyBert slightly struggles with more complex multi-label scenarios. The number of classification labels impacts performance across all models. Table IV below shows the results of our models.

Table IV
LLM PERFORMANCE ACROSS DATASETS

Dataset	Model	Metrics			
		Accuracy	Precision	Recall	F1
COAID	*DistilBert*	0.95	0.92	0.92	0.92
	MobileBert	0.93	0.93	0.93	0.93
	TinyBert	0.93	0.93	0.93	0.93
Liar	*DistilBert*	0.74	0.74	0.74	0.74
	MobileBert	0.74	0.74	0.74	0.74
	TinyBert	0.67	0.67	0.67	0.67
Pheme	*DistilBert*	0.98	0.98	0.98	0.98
	MobileBert	0.97	0.97	0.97	0.97
	TinyBert	0.97	0.97	0.97	0.97

V. CHALLENGES

Large Language Models (LLMs) hold a promise in combating misinformation in public safety applications. However, they face significant challenges and limitations. Below, we explore some of these challenges.

Data Quality: The accuracy of LLM-based misinformation detectors relies on the quality of their training and fine-tuning data. There is a clear lack in datasets for specific applications such as public safety [26]. We used fake news, rumors, and public health datasets in this research, however, data in these subjects can be highly biased, incomplete, and influenced by demographic factors. Crowdsourced data is often noisy, biased, and incomplete, which can cause flaws in detection [25]. In a smart city scale with large volumes of data crowdsourced, these issues can be amplified, which complicates the accuracy of detecting misinformation.

Real Time Processing: Real-time processing is a cornerstone of misinformation detection in smart city public safety systems, but LLMs face significant hurdles in achieving the speed and efficiency required for timely interventions [10]. Network latency and bandwidth limitations further hinder the ability to fetch and process data from distributed sources [4]. Optimizing LLMs for real-time use requires techniques like model pruning, quantization, and edge-cloud hybrid architectures to balance speed and accuracy [10].

Scalability: Scaling systems to detect misinformation in smart cities faces major challenges with large data streams, as these cities produce a massive amount of daily data from various sources [20], which most existing models cannot handle efficiently. This requires expensive and energy-intensive computing infrastructure [10]. Additionally, as user numbers and data types grow, it is challenging to keep performance consistent, especially with different languages and formats. Using distributed models can help, but can lead to problems with synchronization and consistency [20], [29]. To effectively scale these systems, it's important to use modular designs and methods that distribute work while maintaining reliability.

Interpretability: Similar to deep learning techniques, LLMs are often considered "Black Boxes" [17]. The decisions made by these modes can be difficult to explain. The lack of explainability can affect trustworthiness and accountability, which are crucial concepts in public safety apps.

Evolving Threats: Misinformation in public safety applications is growing as a challenge since adversaries are continuously adapting and evolving their strategies. One particular area of concern is the use of advanced AI and LLMs to outpace the current detection techniques [4], [10].

AI Generated Content: LLMs can offer a solid solution in detecting misinformation, however, they can also be part of the problem. LLM can also be trained and employed to spread misinformation in social media or maliciously target public safety apps [17].

VI. CONCLUSION AND FUTURE DIRECTIONS

In this work, we demonstrated the potential of Large Language Models (LLMs) in detecting misinformation within

public safety applications in smart cities. We fine-tuned three compact LLMs; DistilBERT, MobileBERT, and TinyBERT on publicly available misinformation datasets, we achieved promising results, including up to 98% accuracy in certain scenarios. These models are particularly well-suited for deployment in resource-constrained environments, such as edge devices commonly used in smart city infrastructure. The results are promising despite the lack of publicly available datasets in this domain. Furthermore, we highlighted some of the limitations and open issues.

Although LLMs can achieve high accuracy in misinformation detection, the lack of context-specific datasets is a critical challenge. Without having a dataset for public safety misinformation detection, it is hard to determine whether the model will work effectively in a real-life setting. However, the absence of data sets does not necessarily mean that there are no alternative solutions. One possible approach is to crowdsource relevant information from social media websites using specific keywords and use an LLM to label these texts by accessing trustworthy online news platforms. While the assumption that an incident is False because there is no news about it online might not be true, it will help build a more relevant dataset for our purpose. Human experts can also be hired to fulfill that purpose, but the process can be costly and time-consuming.

Another solution is to utilize the multimodality of LLMs. Having various data sources is a challenge but also an opportunity that can be used to enhance data trustworthiness. One of the future directions and a possible robust solution is using cross-referencing across various data sources in smart cities to ensure data trustworthiness. The multimodality of LLMs can help these models process various types of input, e.g., texts, images, audio, and numbers to achieve data trustworthiness between different data types and sources.

In conclusion, while LLMs offer a powerful tool for enhancing public safety in smart cities, their success depends on addressing data limitations, ensuring real-time performance, and building trust through transparency and human oversight. With continued research and development, these models can play a central role in creating safer, more resilient urban environments.

REFERENCES

[1] A. Bondielli and F. Marcelloni. A survey on fake news and rumour detection techniques. *Information sciences*, 497:38–55, 2019.

[2] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[3] C. N. Buțincu and A. Alexandrescu. Blockchain-based platform to fight disinformation using crowd wisdom and artificial intelligence. *Applied Sciences*, 13(10), 2023.

[4] C. Chen and K. Shu. Combating misinformation in the age of llms: Opportunities and challenges. *AI Magazine*, 45(3):354–368, 2024.

[5] L. Cui and D. Lee. Coaid: COVID-19 healthcare misinformation dataset. *CoRR*, abs/2006.00885, 2020.

[6] A. Currin, S. Flowerday, E. de la Rey, K. van der Schyff, and G. Foster. A smart city qualitative data analysis model: Participatory crowdsourcing of public safety reports in south africa. *The Electronic Journal of Information Systems in Developing Countries*, 88(6):e12232, 2022.

[7] A. Dabbous, A. Tarhini, and A. Harfouche. Circulation of fake news: Threat analysis model to assess the impact on society and public safety. In *2023 IEEE International Symposium on Technology and Society (ISTAS)*, pages 1–9. IEEE, 2023.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019.

[9] J. S. Gracias, G. S. Parnell, E. Specking, E. A. Pohl, and R. Buchanan. Smart cities—a structured literature review. *Smart Cities*, 6(4):1719–1743, 2023.

[10] K. Harrison and A. Leopold. How blockchain can help combat disinformation, Jul 2021.

[11] Y. Huang, Y. Wang, and C. White. Designing a mobile system for public safety using open crime data and crowdsourcing. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 67–70, 2014.

[12] S. Huckle and M. White. Fake news: A technological approach to proving the origins of content, using blockchains. *Big data*, 5(4):356–371, 2017.

[13] X. Jiao, Y. Yin, L. Shang, X. Jiang, X. Chen, L. Li, F. Wang, and Q. Liu. Tinybert: Distilling bert for natural language understanding. *arXiv preprint arXiv:1909.10351*, 2019.

[14] W. Khan, A. Daud, K. Khan, S. Muhammad, and R. Haq. Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends. *Natural Language Processing Journal*, 4:100026, 2023.

[15] E. Kochkina, M. Liakata, and A. Zubiaga. All-in-one: Multi-task learning for rumour verification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3402–3413, 2018.

[16] X. Kong, X. Liu, B. Jedari, M. Li, L. Wan, and F. Xia. Mobile crowdsourcing in smart cities: Technologies, applications, and future challenges. *IEEE Internet of Things Journal*, 6(5):8095–8113, 2019.

[17] S. Kwon and B. Jang. A comprehensive survey of fake text detection on misinformation and lm-generated texts. *IEEE Access*, 2025.

[18] M. Liakata, A. Zubiaga, R. Procter, G. W. S. Hoi, and P. Tolmie. Pheme dataset for rumour detection and veracity classification. https://figshare.com/articles/dataset/PHEME_dataset_for_Rumour_Detection_and_Veracity_Classification/6392078, June 2018.

[19] A. G. Pereira, E. Estevez, and P. R. Fillottrani. An innovative mobile app integrating relevant and crowdsourced information for improving citizen's safety. In *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance*, pages 217–225, 2018.

[20] S. Raghavan, B. Y. L. Simon, Y. L. Lee, W. L. Tan, and K. K. Kee. Data integration for smart cities: opportunities and challenges. *Computational Science and Technology: 6th ICCST 2019, Kota Kinabalu, Malaysia, 29-30 August 2019*, pages 393–403, 2020.

[21] M. Romano, T. Onorati, I. Aedo, and P. Diaz. Designing mobile applications for emergency response: citizens acting as human sensors. *Sensors*, 16(3):406, 2016.

[22] V. Sanh, L. Debut, J. Chaumond, and T. Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.

[23] P. Sarzaeim, Q. H. Mahmoud, and A. Azim. A framework for llm-assisted smart policing system. *IEEE Access*, 2024.

[24] Z. Sun, H. Yu, X. Song, R. Liu, Y. Yang, and D. Zhou. Mobilebert: a compact task-agnostic bert for resource-limited devices. *arXiv preprint arXiv:2004.02984*, 2020.

[25] C. Thibault, G. Peloquin-Skulski, J.-J. Tian, F. Laflamme, Y. Guan, R. Rabbany, J.-F. Godbout, and K. Pelrine. A guide to misinformation detection datasets. *arXiv preprint arXiv:2411.05060*, 2024.

[26] F. Torabi Asr and M. Taboada. Big data and quality data for fake news and misinformation detection. *Big data & society*, 6(1):2053951719843310, 2019.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[28] W. Y. Wang. "liar, liar pants on fire": A new benchmark dataset for fake news detection. *CoRR*, abs/1705.00648, 2017.

[29] J. Zhang, H. Bu, H. Wen, Y. Liu, H. Fei, R. Xi, L. Li, Y. Yang, H. Zhu, and D. Meng. When llms meet cybersecurity: A systematic literature review. *Cybersecurity*, 8(1):1–41, 2025.