The following publication Ngai EW, Lui AK, Kei BC (2025), "Natural language processing in government applications: a literature review and a case analysis". Industrial Management & Data Systems, Vol. 125 No. 6 pp. 2067–2104 is published by Emerald and is available at https://doi.org/10.1108/IMDS-07-2024-0711.

Natural language processing in government applications: A literature review and a case analysis

Abstract

This study explores the applications of Natural Language Processing (NLP) in the public sector, aiming to enhance governance, policy formulation, and service delivery. Through a systematic literature review, we synthesize recent advancements in NLP techniques and their implementation across various government functions. In addition, we present a case study of an NLP-based social media text analytics system designed for Cantonese in the Hong Kong context, highlighting the unique challenges and opportunities of applying NLP in a non-Western, multilingual environment. Our research introduces a novel conceptual framework that connects NLP techniques to specific government domains, offering a comprehensive view of how these technologies can improve governmental operations. We also address the ethical and practical challenges surrounding NLP adoption, including privacy concerns, algorithmic bias, and transparency issues. The study concludes with recommendations for policymakers on how to effectively integrate NLP into government operations and provides a roadmap for future research in this emerging field. This work contributes to both theoretical understanding and practical implementation, offering insights into the strategic, efficiency, and effectiveness benefits of NLP in the public sector.

Keywords: Natural language processing (NLP), Government applications, Literature review, Case analysis

1. Introduction

There is a growing trend in the development of natural language processing (NLP) applications for the public sector, making government applications of NLP a burgeoning topic across various fields, including computer science, public health, and public administration. Scholars and practitioners are increasingly recognizing the potential societal benefits that NLP technology can bring to the public through its application in government contexts. NLP is a branch of artificial intelligence (AI) that enables computers to comprehend, generate and manipulate human language. Despite having a cumulative history dating back to the 1950s (Kumar & Sharma, 2017), the true advancement and commercialization of NLP have only materialized in recent years.

NLP equips machines with the capability to understand and process both text and audio data (Misuraca, van Noordt, & Boukli, 2020). Through NLP, qualitative data in the form of text can be converted into quantitative data using metrics such as term frequencies (e.g., Hu, Deng, & Zhou, 2019). Once natural language is transformed into quantitative data, advanced analytical techniques and algorithms, such as sentiment analysis and text clustering, can be applied (Li, Thomas, & Liu, 2020). Common NLP applications encompass chatbots, virtual assistants (e.g., Siri), information extraction, sentiment analysis, and opinion mining (Cambria & White, 2014). Notably, NLP research has harnessed the capabilities of machine learning (ML) and deep learning (DL) techniques.

By empowering governments to analyze vast amounts of semi-structured or unstructured data, NLP has opened up significant opportunities for creating benefits, including cost savings, improved decision-making, and enhanced services (Pencheva, Esteve, & Mikhaylov, 2020). NLP plays a pivotal role in addressing critical issues in the public domain, spanning healthcare (Hu, et al., 2019), policy-making (Zavattaro, French, & Mohanty, 2015), and the legal sector (Medvedeva, Vols, & Wieling, 2020). For instance, Pencheva et al. (2020) highlighted the advantages of big data and advanced analytics, including NLP, throughout the policy cycle in the government sector. Additionally, Liu et al. (2019) employed NLP techniques, such as keyword extraction and sentiment analysis, to explore opinions on websites related to public transportation services in Shanghai.

As governments around the world increasingly embrace digitalization, the application of NLP technologies within the public sector has emerged as a promising avenue for improving service delivery, enhancing policy-making, and fostering citizen engagement. While NLP has seen broad application in commercial sectors, its adaptation to government contexts introduces unique challenges and opportunities. Unlike private organizations, government agencies are bound by specific ethical, operational, and societal considerations that necessitate a tailored approach to NLP applications. These differences not only highlight distinct challenges but also underscore the importance of tailoring NLP research to address specific governmental needs.

Despite its potential, our understanding of NLP's role within government remains limited. There is a pressing need for NLP to support more strategic, data-driven decision-making within government. By extracting insights from vast volumes of unstructured data, NLP can enable government agencies to respond more effectively to citizen needs and to anticipate societal issues. However, without a government-focused perspective,

these tools may fail to align with the unique decision-making processes of public institutions, limiting their value and impact.

While NLP technology has seen significant advances in fields like commerce, systematic research and application in the government sector remain underdeveloped. For instance, Kumar and Sharma (2017) conducted a systematic literature review focusing on the application of opinion mining in government intelligence. Kang et al. (2020) reviewed 24 business journals that utilized NLP in management research. Furthermore, while existing studies have made significant contributions to NLP techniques and algorithms (e.g., Liu et al. 2019), there is a lack of comprehensive reviews that synthesize findings across various domains and techniques, as well as a scarcity of research examining the practical implementation and societal impact of these technologies in government settings. Therefore, this study aims to fill the knowledge gaps in NLP applications for the government sector by conducting a systematic literature review to consolidate recent research efforts and a case study. The study has four primary objectives:

- Conduct a comprehensive literature review of NLP application in the government sector to provide an updated synthesis of existing research and identify gaps.
- Develop a classification framework that organizes the major government domains where NLP technologies are applied and outline key techniques supporting these applications.
- Use a real-life case analysis that demonstrate the design and implementation of an NLP-based social media text analytics system within a government department in Hong Kong, focusing on its practical value in decision-making.

- Offer insights for future research and practice by providing a societal value grid that highlights the strategic, efficiency, and effectiveness benefits of NLP applications for the public sector.

This study employs a two-stage research methodology, combining a systematic literature review with a case study, to provide a comprehensive understanding of NLP applications in government. The systematic review offers a comprehensive overview of the current state of research trends and theoretical frameworks, while the case study provides practical insights into the real-world implementation of NLP in a government setting. We chose this dual-method approach to provide a well-rounded perspective on NLP applications in government settings, addressing both theoretical gaps and practical implications.

This study provides unique insights into NLP applications within government settings by taking a comprehensive approach that addresses key gaps in the current literature. Unlike prior reviews that typically focus on specific applications like opinion mining or are limited to commercial and Western contexts, our study examines NLP's role across a broader spectrum of government functions. We especially focus on non-Western perspectives through an in-depth case study in Hong Kong, highlighting the linguistic and cultural nuances of Cantonese in the design and implementation of a social media text analytics system. This contextually rich examination offers practical, on-the-ground insights rarely covered in previous research.

Furthermore, we introduce a novel classification framework and a societal value grid that categorizes NLP applications according to their strategic, efficiency, and effectiveness benefits tailored to public sector needs. These tools provide both a synthesis of recent advancements and a roadmap for future NLP research within diverse government contexts. By addressing the distinctive needs of multilingual, multicultural

populations, this study expands the applicability of NLP research, offering both theoretical advancements and actionable implications for government-specific challenges.

Our study extends the existing body of knowledge by offering a more comprehensive, contextually diverse, and practically grounded exploration of NLP applications in government. These unique contributions not only fill existing gaps in the literature but also pave the way for future research and practical implementations of NLP technologies in diverse governmental contexts. We believe that these elements collectively demonstrate the distinct and valuable insights our study provides.

The structure of this paper is organized as follows. Section 2 outlines the research methodology employed in this study. Section 3 provides an extensive review of NLP research, covering its applications, techniques, and societal benefits, and introduces a conceptual framework for NLP applications. Section 4 presents a real-life case study to demonstrate the practical aspects of an NLP-based social media text analytics system for government applications. Section 5 discusses lessons learned from the implementation, challenges encountered, and the potential competitive advantages of using NLP in government contexts. Section 6 offers conclusions and directions for future research. Finally, Section 7 highlights the theoretical and managerial contributions of the study, discusses its influence on policy formation and implementation, and provides practical recommendations for policymakers.

2. Research methodology

In this research, a two-stage research methodology is employed (Ngai *et al*., 2015). The first stage involves a thorough literature review of articles focusing on "NLP applications in government" and the societal benefits derived from such applications. The

findings of the literature review contribute to the development of a conceptual framework and a societal value grid. The framework outlines major domains where NLP is applied and the key techniques supporting these applications, while the NLP societal value grid provides an overview of NLP's societal benefits in the government sector. The second stage comprises a case study of a government department currently utilizing NLP technologies to enhance government services. This case study not only validates the conceptual framework and societal value grid but also illustrates the design, development, and implementation of an NLP application system within a government department in Hong Kong, emphasizing the actual value of the system.

2.1. A comprehensive literature review of articles on NLP

We conducted a literature review of journal articles on NLP using a systematic approach inspired by Ngai and Wat (2002). Employing keywords such as "natural language processing," "NLP," and "government," we performed a literature search covering the period from 2017 to 2023. We selected the timeframe from 2017 to 2023 for the literature review for several key reasons. The year 2017 marked the introduction of the Transformer architecture (Vaswani et al., 2017), which catalyzed a significant leap in NLP capabilities and paved the way for subsequent models like BERT (Devlin et al., 2018) and GPT (Radford et al., 2019). This breakthrough fundamentally transformed NLP applications across various sectors, including government, making research from 2017 onward highly relevant for our study. The endpoint of 2023 was chosen because the literature search was completed in that year, and 2024 data was not yet available at the time of review. This ensures that the review captures the most recent and finalized studies while reflecting the latest advancements in NLP techniques and their applications in government. Thus, the 2017-2023 period offers a comprehensive and up-to-date view

of the significant transformations in NLP, particularly in the context of governmental applications.

Each selected paper underwent an individual examination to assess its relevance to the research topic. Subsequently, we analyzed and classified these papers based on the domains and techniques employed, as well as the underlying theories and models applied.

Our review primarily focused on studies from journals indexed in the Social Science Citation Index (SSCI), Science Citation Index (SCI), or Science Citation Index Expanded (SCIE) and refereed journal papers represent advanced research outputs (Ngai and Wat, 2002). Excluded from our review were doctoral dissertations, master's theses, conference papers, documentaries, and textbooks. Journals were used due to their presentation of advanced research outcomes and their wide acceptance by researchers and practitioners as reliable sources of information (Nord & Nord, 1995). We utilized seven prominent online databases, namely Scopus, Web of Science, Science Direct, Emerald Insight, Academic Search Premier, Business Source Premier, ProQuest, and EBSCO, covering a vast range of relevant academic journals.

2.2. Conceptual framework for NLP applications in government sector

The conceptual framework (Fig. 1) presented in this study was developed through a systematic analysis of the literature review findings. We identified recurring themes and patterns in NLP applications across various government domains, aligning them with the United Nations' Classification of Functions of Government (COFOG). Simultaneously, we cataloged the most frequently used NLP techniques and algorithms reported in the reviewed studies. This dual-layered approach allowed us to create a comprehensive framework that maps NLP techniques to specific government domains, providing a structured overview of the current state of NLP applications in government.

Figure 1 comprises two layers: the upper layer delineates the 10 domains of NLP implementation in government, namely, 1) general public service, 2) public order and safety, 3) defense, 4) economic affairs, 5) environmental protection, 6) housing and community amenities, 7) health, 8) recreation, culture, and religion, 9) education, and 10) social protection. These government sectors align with the COFOG issued by the United Nations, widely recognized and adopted in research, including by organizations such as the European Union (EU) (de Sousa, de Melo, Bermejo, Farias, & Gomes, 2019). The lower layer of the framework highlights six frequently used NLP classes, namely, feature extraction, text clustering, topic modeling, text classification, semantic analysis, and sentiment analysis, each supported by a set of algorithms. This framework serves as a guide for reviewing selected papers, assessing their relevance to the research topic, and classifying them based on the NLP technologies supporting applications in the government sector. The subsequent sections provide concise descriptions of these six NLP application classes, each complemented by a set of techniques facilitating relationship extraction from the data.

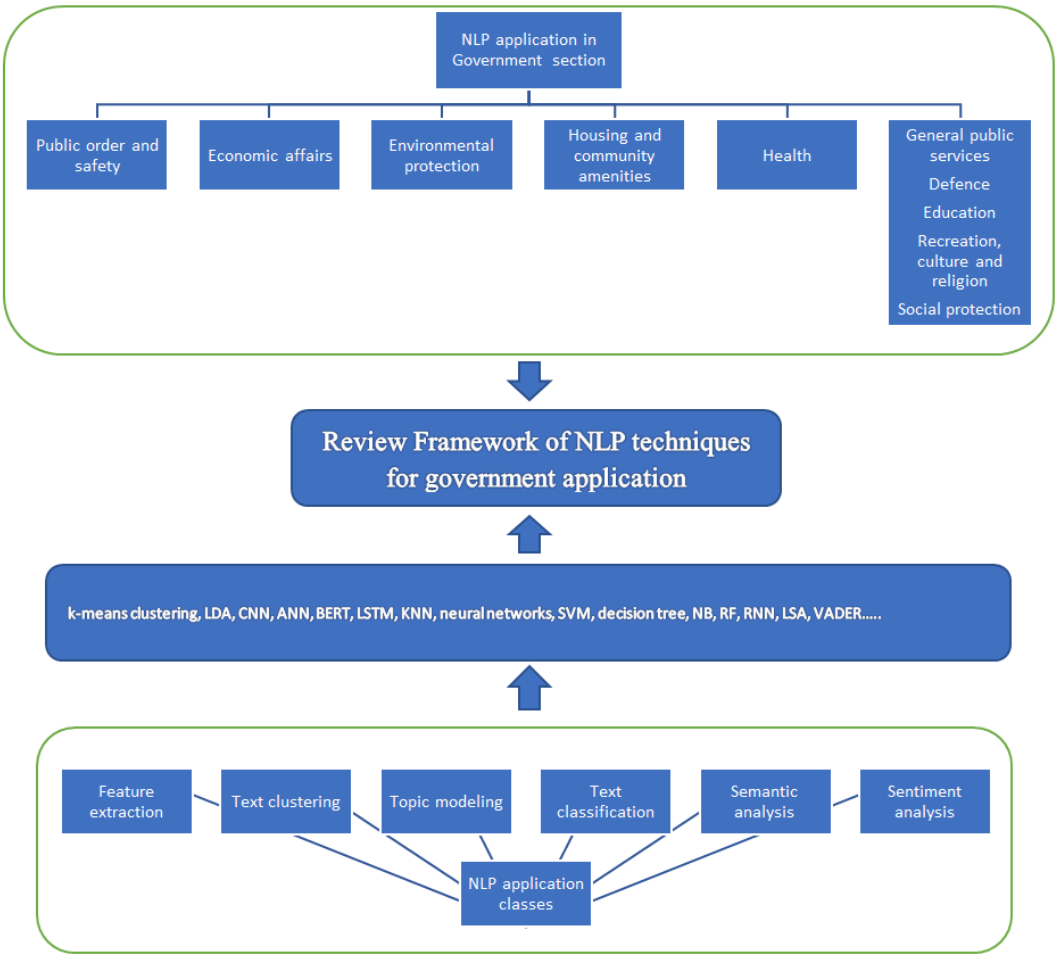

Fig. 1. NLP conceptual framework for NLP application in government sector.

(Source: Figure created by authors)

2.2.1. Feature extraction

Feature extraction is an essential NLP task focused on distilling specific information, such as relationships, entities, or events, from natural language text (Li, et al., 2020). This extracted data can then undergo further processing to adopt a structured format, facilitating efficient management and storage. Feature extraction plays a pivotal role in enabling systems or web users to derive meaningful insights from seemingly chaotic textual data. Some common feature extraction techniques encompass term frequency-inverse document frequency (TF-IDF), Bidirectional Encoder Representations from Transformers, and N-gram.

2.2.2. Text clustering

Text clustering is employed to group objects, such as paragraphs, documents, or sentences, based on the similarity between these objects (Tao, Yang, & Feng, 2020). Common techniques for text clustering include methods like k-means clustering.

2.2.3. Topic modeling

Topic modeling is a statistical modeling technique that falls under text clustering, designed to identify common topics within groups of documents, such as tweets (Tao, et al., 2020). The goal is to discover the main topics discussed in text, handling information without predefined labels (Zhai & Massung, 2016). A topic comprises a cluster of words that frequently appear together. For instance, a community might be sharing views on a topic related to public transportation, and topic modeling can help identify the primary topic without reading each review. Latent Dirichlet Allocation (LDA) is a widely used method for topic modeling.

2.2.4. Text classification

Text classification has garnered increasing attention due to the rapid growth of data in social media (Li, et al., 2020). Specifically, text classification is employed to categorize text files into different classes based on a list of training data with predefined labels (Tao, et al., 2020). Common text classification approaches encompass artificial neural network (ANN), convolutional neural network (CNN), BERT, long-short term memory (LSTM), K-nearest neighbors (KNN), neural networks, support vector machine (SVM), decision tree, naïve Bayes (NB), random forest (RF), and recurrent neural networks (RNN).

2.2.5. Semantic analysis

Semantic analysis involves interpreting the literal meaning of natural language that is not explicitly ascribed in syntactic form (Li, et al., 2020). Semantic-based NLP explores multi-word expressions and concepts that encompass various semantic knowledge representations. Semantic knowledge can be acquired intrinsically by constructing semantic structures within a corpus or extrinsically through semantic knowledge bases or ontologies using ML techniques (Cambria & White, 2014). Latent semantic analysis (LSA) is a commonly used technique for conducting semantic analysis of data.

2.2.6. Sentiment analysis

Sentiment analysis or opinion mining is a sub-discipline of text classification that focuses on identifying the emotions and opinions expressed by individuals (Hu, et al., 2019). Sentiment analysis involves tasks such as identifying sentiment sentences, discovering emotion-relevant keywords, determining the polarity and score (negative/positive/neutral) of words, and analyzing the emotional tendencies of the subjects of interest. Common sentiment analysis techniques include Valence Aware Dictionary for Sentiment Reasoning (VADER) and BERT.

3. Results

3.1. Results from the literature review

The search resulted in 103 articles from 36 journals. Each of these 103 articles were thoroughly reviewed and categorized based on the framework illustrated in Figure 1. In the subsequent sections, we analyze and discuss the findings from the selected articles.

3.1.1. Distribution of articles by the year of publication

Figure 2 illustrates the distribution of articles published from 2017 to 2023. There is a notable increase in research outputs, particularly in the year 2021.

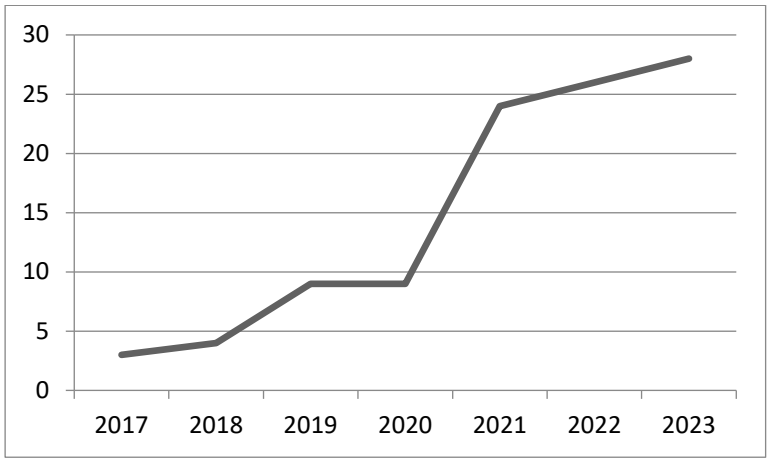

Fig. 2. Distribution of articles published by year

(Source: Figure created by authors)

Table 1 displays the distribution of studies categorized by application domain and publication year. It appears that NLP technology is predominantly employed in the health domain.

Table 1 The distribution of the articles by application domain and publication year.

(Source: Table created by authors)

	2017	2018	2019	2020	2021	2022	2023	Total No.	%
Economic affairs			1	1		3	3	8	7.8
Environmental protection		1		1	4	3	4	13	12.6
General public service						1	2	3	2.9
Health		1	3	3	17	17	11	52	50.5
Housing and community amenities				1			3	4	3.9
Public order and safety	3	2	5	3	3	2	5	23	22.3
Total	3	4	9	9	24	26	28	103	100

3.1.2. Distribution of articles per journal

In our findings, we identified a total of 48 different journals across various disciplines (e.g., computing, engineering, medical, etc.) that published articles related to NLP. As depicted in Table 2, the *Journal of Medical Internet Research* stands out with the highest

number of articles (13, 12.6%) addressing NLP topics. *Expert Systems with Applications* and *International Journal of Information Management* follow closely as the second and third highest contributors, each with 11 articles (10.7%). The majority of these journals are located in the UK (38, 36.9%), followed by the US (22, 21.4%), and Canada (20, 19.4%).

Table 2 Distribution of articles by journal title. *(Source: Table created by authors)*

. Rank	Journal title	Research Field	Country	No.	%
1	Journal of Medical Internet Research	Medical Informatics/Health Care Sciences & Services	Canada	13	12.6
2	Expert Systems with Applications	Engineering, Electrical & Electronic / Operations Research & Management Science / Computer Science, Artificial Intelligence	United Kingdom	11	10.7
3	International Journal of Information Management	Information Science & Library Science	United Kingdom	11	10.7
4	International Journal of Environmental Research and Public Health	Environmental Sciences / Public, Environmental & Occupational Health	Switzerland	8	7.8
5	JMIR Public Health and Surveillance	Public, Environmental & Occupational Health	Canada	7	6.8
6	IEEE Access	Telecommunications / Engineering, Electrical & Electronic / Computer Science, Information Systems	United States	4	3.9
7	Information Processing & Management	Computer Science / Decision Sciences / Engineering/Social Sciences	United Kingdom	4	3.9
8	PLoS ONE	Multidisciplinary Sciences	United States	3	2.9
9	Sustainability (Switzerland)	Environmental Science / Energy / Social Sciences	Switzerland	3	2.9
10	Computers & Security	Computer Science, Information Systems	United Kingdom	2	1.9
11	Data Technologies and Applications	Computer Science/Social Sciences	United Kingdom	2	1.9
12	Decision Support Systems	Business, Management and Accounting/Computer Science/Decision Sciences	Netherlands	2	1.9
13	Geo-Spatial Information Science	Remote Sensing	China	2	1.9
14	IEEE Journal of Biomedical and Health Informatics	Computer Science, Interdisciplinary Applications / Medical Informatics / Mathematical & Computational Biology / Computer Science, Information Systems	United States	2	1.9

15	IEEE Transactions on Industrial Informatics	Engineering, Industrial / Computer Science, Interdisciplinary Applications / Automation & Control Systems	United States	2	1.9
16	Journal of Ambient Intelligence and Humanized Computing	Telecommunications / Computer Science, Information Systems / Computer Science, Artificial Intelligence	Germany	2	1.9
17	Journal of Cleaner Production	Green & Sustainable Science & Technology | Engineering, Environmental / Environmental Sciences	United States	2	1.9
18	Journal of Computing in Civil Engineering	Engineering, Civil / Computer Science, Interdisciplinary Applications	United States	2	1.9
19	Resources, Conservation and Recycling	Economics, Econometrics and Finance / Environmental Science	Netherlands	2	1.9
20	Transportation Research Part D: Transport and Environment	Engineering / Environmental Science / Social Sciences	United Kingdom	2	1.9
21	Water Policy	Water Resources	United States	2	1.9
22	Accident Analysis and Prevention	Transportation / Social Sciences, Interdisciplinary/Ergonomics / Public, Environmental & Occupational Health	United Kingdom	1	1.0
23	American Journal of Clinical Nutrition	Nutrition & Dietetics	United States	1	1.0
24	Applied Sciences (Switzerland)	Chemical Engineering / Earth and Planetary Sciences / Engineering / environmental Science / Materials Science / Physics and Astronomy	Switzerland	1	1.0
25	Cities	Urban Studies	United Kingdom	1	1.0
26	Computers, Materials and Continua	Computer Science / Engineering / Mathematics / Materials Science	United States	1	1.0
27	Frontiers in Political Science	Operation of government	Switzerland	1	1.0
28	Frontiers in Psychology	Psychology, Multidisciplinary	Switzerland	1	1.0
29	Healthcare (Switzerland)	Health Professions / Medicine / Nursing	Switzerland	1	1.0
30	Humanities and Social Sciences Communications	Arts and Humanities / Business, Management and Accounting / Economics, Econometrics and Finance / Psychology / Social Sciences	United States	1	1.0
31	Infection Control and Hospital Epidemiology	Infectious Diseases / Public, Environmental & Occupational Health	United States	1	1.0
32	Journal of Information Science	Computer Science, Information Systems	United Kingdom	1	1.0
33	Journal of Intelligent Transportation Systems: Technology, Planning, and Operations	Computer Science / Engineering / Mathematics	United Kingdom	1	1.0

34	Journal of the American Medical Informatics Association	Computer Science, Interdisciplinary Applications / Medical Informatics / Health Care Sciences & Services / Computer Science, Information Systems	United Kingdom	1	1.0
35	Ocean and Coastal Management	environmental Science /Agricultural and Biological Sciences / Earth and Planetary Sciences	United Kingdom	1	1.0
36	World Wide Web	Computer Science	United States	1	1.0

3.1.3. Distribution of articles by NLP application classes and government domains

Table 3 presents the classification of all the articles according to their respective NLP application categories. It is noteworthy that government domains such as defense, recreation, culture, and religion, education, and social protection are not included in Table 3 due to the absence of articles found for these domains.

Table 3. Distribution of articles by NLP application classes and government domains**.**

(Source: Table created by authors)

NLP application class	Economic affairs	Environmental protection	General public services	Health	Housing and community amenities	Public order and safety	Total	%
Text classification	6	4		38	4	20	72	38.1
Text clustering				1			1	0.5
Feature extraction	1	1		5		3	10	5.3
Semantic analysis		2	1	1			4	2.1
Sentiment analysis	3	4	2	36	2	3	50	26.5
Topic modeling	1	13	1	24	1	12	52	27.5
Total	11	24	4	105	7	38	189*	100

*The total number is 189 because some of the 103 articles utilize more than one NLP application category.

3.1.4. Distribution of articles by NLP techniques

An increasing number of studies have indicated that organizations apply NLP technology, encompassing a wide range of techniques. Table 4 provides a summary of some reviewed articles corresponding to the classification scheme.

Table 4 Classification of some articles by NLP techniques. *(Source: Table created by authors)*

Application domains	NLP application class	Techniques	References
Economic affairs	Sentiment analysis	Kernel Density Estimation method	(Chang, Li, Huang, Zhang, & Chin, 2022)
		VADER	(Kabbani, Klumpenhouwer, El-Diraby, & Shalaby, 2022)
	Text classification	CNN	(Kim, Ha, Choi, & Moon, 2018; N. Kim & Hong, 2021)
	Feature extraction, text classification	Feature extraction (Word2Vec), text classification (k-nearest neighbor (KNN), logistic regression, and SVM)	(Amina & Azim, 2019)
	Sentiment analysis, topic modelling	Sentiment (VADER), topic modelling (LDA)	(Lock & Pettit, 2020)
Environmental protection	Semantic analysis	LSA, TF-IDF	(Stutzer, et al., 2021)
	Sentiment analysis	BERT	(Kim, Ganesan, Dickens, & Panda, 2021)
	Topic modeling	LDA, Hierarchical Dirichlet Process (HDP), Latent Semantic Indexing (LSI), Non-Negative Matrix Factorisation (NMF), Structural Topic Model (STM)	(Teh, Piao, Almansour, Ong, & Ahad, 2022)
		LDA	(Ghahramani, Galle, Ratti, & Pilla, 2021), (Bennett, Rachunok, Flage, & Nateghi, 2021), (Sun, et al., 2020),
		LDA, NMF, BERT	(Umamaheswaran, et al., 2023)
	Feature extraction, text classification	Feature extraction (TF-IDF), text classification (NB, SVM, decision tree, maximum entropy)	(D. Zhang, et al., 2018)
	Sentiment analysis, topic modelling	Sentiment analysis (NB), topic modeling (LDA)	(Pu, Jiang, & Fan, 2022)
Health	Sentiment analysis	Kernel Density Estimation method	(Shahrezaye, Suter, & Meckel)
		VADER	(Liew & Lee, 2021), (He, et al., 2021)
		BERT	(Chiang, et al., 2021), (Liang, et al., 2023)
		NB, CNN	(Roy & Ghosh, 2021)
		VADER, BERT, Textblog	(Hussain, et al., 2021)

	Tencent NLP	(G. Hu, Han, Zhou, & Liu, 2019), (Tan, et al., 2021).
Text classification	n-grams, LDA, NB, Random Forest (RF), SVM, BERT	(Cotfas, Delcea, & Gherai, 2021)
	BERT	(Y. Zhang, et al., 2022), (Tang, et al., 2021)
	CNN	(Luo, Wang, & Liu, 2022)
	SVM, RNN	(Ayers, et al., 2018)
Text clustering	k-means clustering	(Sameh Nagui Saleh, Lehmann, & Medford, 2021)
Topic modeling	LDA	(Alahmari, et al., 2022), (Kumar, Mukherjee, Choi, & Dhamotharan, 2022), (Zheng, Xue, Sun, & Zhu, 2021), (Lindquist, Thomas, Turner, Blankenship, & Kyle, 2021), (Boon-Itt & Skunkan, 2020)
	SVM	(Hasan, et al., 2022)
Emotion analysis, sentiment analysis	Sentiment analysis (VADER), emotion analysis (RNN)	(Zhang, Xu, Li, & Hu, 2021)
Emotion analysis, topic modeling	LDA	(Adikari, et al., 2021)
	Topic modeling (LDA), emotion (RNN)	(Sameh N Saleh, Lehmann, McDonald, Basit, & Medford, 2021)
Feature extractor, text classification	Feature extraction (BERT), text classification (RNN, LSTM)	(Essam, et al., 2021)
Feature extraction, sentiment analysis	Feature extraction (N-gram, TF-IDF), sentiment (SVM, NV, and KNN)	(Aljameel, et al., 2021)
Feature extraction, text classification, topic modeling	Feature extraction (Bert), topic modeling (LDA), text classification (CNN, LSTM, ANN)	(Shah, Yan, Tariq, & Ali, 2021)
	Feature extraction (TF-IDF), text classification (SVM, RF, and logistic regression), topic modeling (LDA),	(Chen, et al., 2019)
Semantic analysis, sentiment analysis	Sentiment (VADER), semantic (Named Entity Recognizer (NER))	(Pérez-Pérez, Pérez-Rodríguez, Fdez-Riverola, & Lourenço, 2019)
Sentiment analysis, topic modeling	topic modeling (LDA), sentiment (LSTM, RNN)	(Jelodar, Wang, Orji, & Huang, 2020)
	Sentiment analysis (VADER), topic modeling (LDA)	(Yin, Song, Yang, & Li, 2022)
	Topic modeling (LDA), sentiment (VADER)	(Xiong, Hswen, & Naslund, 2020)
Text classification, text representation	Text representation (GloVe, n-gram), text classification (SVM, LSTM, BERT)	(Miao, Last, & Litvak, 2022)

	Text Classification, topic modeling	Topic modeling (LDA), text classification (SVM, RF, NB, shallow neural network (NN), KNN, bidirectional long short-term memory, BERT)	(Yang, et al., 2021)
Housing and community amenities	Text classification	GloVe, BRNN, CNN	(Adikari & Alahakoon, 2020)
	Text classification, sentiment analysis	Naïve Bayes, Random Forest, VADER	(Duan, et al., 2023)
Public order and safety	Feature extraction	Named Entity Recognizer (NER)	(Zaeem, Manoharan, Yang, & Barber, 2017)
	Sentiment analysis	TextBlob	(Reynard & Shirgaokar, 2019)
		CNN	(Wu, Li, Shen, & He, 2020)
		Word2Vec	(Başarslan & Kayaalp 2023)
	Text classification	TF-IDF, PN, NB, KNN, SVM	(Alkhatib, El Barachi, & Shaalan, 2019)
		RF, NB, SVM, Logistics regression	(L. Li, Bensi, Cui, Baecher, & Huang, 2021)
		SVM	(Ragini, Anand, & Bhaskar, 2018), (Ragini, et al., 2018)
		RF, SVM, NB	(Amrit, Paauw, Aly, & Lavric, 2017)
		SVM	(Medvedeva, et al., 2020)
	Topic modeling	LDA	(Zhou, Kan, Huang, & Silbernagel, 2021), (Wang & Taylor, 2019), (Fersini, Messina, & Pozzi, 2017), (Basilio, Pereira, & Brum, 2019), (Son, Lee, Jin, & Lee, 2019)
	Feature extraction, topic modeling, text classification	Feature extraction (BERT), topic modeling (LDA), classification (CNN, LSTM, ANN)	(Wahid, et al., 2022)
	Feature extraction, text classification, topic modeling	Feature extraction (TF-IDF), text classification (BERT), topic modeling (LDA)	(Khatoon, et al., 2021)
	Topic modelling, Sentiment analysis	Topic modelling (LDA), sentiment (TextBlob)	(Aslan, Li, Çelebi, & Tian, 2020)

3.1.5. Distribution of articles by NLP techniques and government domains

Table 5 presents the main techniques used for NLP in the articles. LDA is the most popular, being used in 19.6% (37 of 189) of the studies reviewed, followed by SVM,

used in 10.1% (19 of 189), and then the Bidirectional Encoder Representations from Transformers (BERT), used in 9.5% (18 of 189) of studies.

Table 5 Distribution of articles by NLP techniques and government domains. *(Source: Table created by authors)*

NLP techniques	Economic affairs	Environmental protection	General public services	Health	Housing and community amenities	Public order and safety	Total
LDA	1	8		20		8	37
Support vector machines (SVM)	2	1		9		7	19
BERT		2		14		2	18
Naive Bayes (NB)		3		5	1	4	13
Valence Aware Dictionary for Sentiment Reasoningc (VADER)	2			9	1		12
CNN	1			4	1	3	9
LSTM				6		3	9
Neural network (NN)			1	4		2	7
Random forest				3	1	3	7
Term frequency-inverse document frequency		2		3		3	8
Logistic regression	1			3		1	5
Recurrent neural network (RNN)				5			5
K-nearest neighbors (KNN)	1			2		1	4
n-gram				4			4
Bidirectional long short-term memory				2		1	3
GloVe				2	1		3
TextBlob				1		2	3
ANN				1		1	2
Decision tree		1				1	2
Kernel Density Estimation method	1			1			2
LSA		1	1				2
Named Entity Recognizer (NER)				1		1	2
Non-Negative Matrix Factorisation (NMF)		2					2
Tencent NLP				2			2
Word2Vec	1					1	2
BRNN					1		1
Hierarchical Dirichlet Process (HDP)		1					1
k-means clustering				1			1
Latent Semantic Indexing (LSI)		1					1
Maximum entropy		1					1
Polynomial Networks (PN)						1	1

Structural Topic Model (STM)		1					1
Total	10	24	2	102	6	45	189*

*The total number is 189 because some of the 103 articles utilize more than one NLP application category.

3.1.6. Theories for NLP

Researchers apply theories to underpin their research design and describe a topic or phenomenon under study. The use of theoretical perspectives can help build an improved understanding of NLP applications in the public sector, including healthcare and crime detection. However, in NLP-related research, NLP often involves computational calculations, such as text classification and sentiment analysis, to obtain insights into weak theoretical contributions (Kang, Cai, Tan, Huang, & Liu, 2020; Kar & Dwivedi, 2020). Since 1990, the majority of NLP studies have been empirically-based (Deng & Liu, 2018). Such data-driven research shows a missing connection with theoretical perspectives from organization, management, systems, social science, and behavioral theories (Kar & Dwivedi, 2020).

Only a few studies in this field have applied theories to interpret results or phenomena. O'Leary (2016) used game theory to explain results generated by sentiment measures. This study demonstrated that game theory offers a potential rationale for understanding why popular ideas receive numerous comments. Based on media synchronicity theory, Son et al. (2019) used NLP techniques to examine the content of Twitter. Nistor et al. (2018) employed NLP and Bakhtin's theory of dialogism to explain how online communities reacted to newcomers' inquiries. Bakhtin's theory of dialogism, belonging to social theory, suggests that life is dialogic, and meaning evolves through interaction. Evans and Aceves (2016) explored the opportunities of using NLP and ML methods to mine text for social insights that could test or extend existing sociological theories. They also suggested that the availability of new data, including behaviors and social

relationships, provokes the possibilities of constructing new predictive and descriptive theories (Tavory & Timmermans, 2014).

3.1.7. NLP applications for pain points of government management

Our literature review, synthesizing findings from 89 studies, highlights effective NLP techniques such as sentiment analysis, topic modeling, and text classification. These techniques address critical pain points in government management, including crisis management, policy impact assessment, and misinformation detection. Table 6 provides an overview of these pain points/challenges, the corresponding NLP solutions, the number of studies supporting each theme, and example references.

The literature review identified six key themes where NLP technologies offer significant benefits:

1. Crisis management and rapid response: During crises like the COVID-19 pandemic, governments face challenges in swiftly processing information and responding effectively. NLP applications, such as real-time sentiment analysis and event detection from social media, enable rapid assessment of public sentiment and emerging issues, thereby supporting efficient crisis management. This is evidenced by 37 studies, including works by Hasan, et al. (2022) and Liang, et al. (2023).
2. Policy impact assessment: Governments often struggle to understand the impact and public reception of policies. NLP tools analyze public discourse, comments, and feedback, helping assess policy effectiveness and gather actionable insights. This theme is supported by 21 studies, such as those by Liew & Lee (2021) and Tang, et al. (2021).
3. Monitoring public sentiment and opinion: Gauging public emotions and opinions in real-time is challenging. NLP-powered sentiment analysis systems allow

governments to identify public sentiment trends, as demonstrated in 15 studies, including Başarslan & Kayaalp (2023) and Pu, et al. (2022).

4. Improved public services: Understanding and addressing citizen needs is crucial for effective governance. Sentiment analysis and feedback systems provide insights into citizen preferences and complaints, enabling targeted service improvements. This is highlighted in 8 studies, such as those by Alahmari, et al. (2022) and Yang, et al. (2021).
5. Misinformation detection and management: The proliferation of fake news affects public trust and governance. NLP tools detect misinformation by analyzing content and verifying sources, aiding in timely interventions. This is explored in 4 studies, including Shahrezaye, et al. (2022).
6. Resource allocation efficiency: Inefficiencies in resource distribution can be mitigated by analyzing data trends and patterns. NLP aids in optimizing resource allocation for public services, as shown in 4 studies, such as Chang, et al. (2022).

Table 6 Government management pain points/challenges and NLP solutions.

(Source: Table created by authors)

Pain points/Challenges	NLP Applications	No. of studies	Example of References
Crisis management and rapid Response	Real-time sentiment analysis and event detection from social media enable rapid assessment of public sentiment and emerging issues, supporting efficient crisis management.	37	(Adikari, et al., 2021), (Essam, et al., 2021), (Hasan, et al., 2022), (Kabbani, Klumpenhouwer, El-Diraby, & Shalaby, 2022), (Liang, et al., 2023)

Policy impact assessment	Analyzing public discourse, comments, and feedback helps assess policy effectiveness and gather actionable insights.	21	(Liew & Lee, 2021), (Lindquist, Thomas, Turner, Blankenship, & Kyle, 2021), (Tang, et al., 2021), (Sun, et al., 2020)
Monitoring public sentiment and opinion	NLP-powered tools such as sentiment analysis systems allow governments to identify public sentiment trends in real-time.	15	(Başarslan & Kayaalp 2023), (Pu, Jiang, & Fan, 2022), (Umamaheswaran, et al., 2023), (Yin, Song, Yang, & Li, 2022)
Improved public services	Sentiment analysis and feedback systems provide insights into citizen preferences and complaints, enabling targeted service improvements.	8	(Alahmari, et al., 2022), (Lock & Pettit, 2020), (Shah, Yan, Tariq, & Ali, 2021), (Yang, et al., 2021)
Misinformation detection and management	NLP-powered tools can detect misinformation by analyzing content and verifying sources, aiding in timely interventions.	4	(Shahrezaye, Suter, & Meckel, 2022), Xia et. a. 2023)
Resource allocation efficiency	Analyzing data trends and patterns aids in optimizing resource allocation for public services.	4	(Chang, Li, Huang, Zhang, & Chin, 2022), (Duan, et al., 2023)

3.1.8. Societal benefits of NLP applications

For the societal benefits of NLP technology, Table 7 presents examples of how government applications of NLP technology contribute to societal benefits, as identified in the literature review. Based on the results of our literature review on NLP applications

in the government sector, we propose the development of an NLP value grid to assess the technology's potential in creating value and enhancing the competitive advantages of entities within the government sector. Drawing on insights from Hammer and Mangurian (1987), Riggins (1999), and our literature review on NLP applications in the government sector, we have constructed an NLP value grid. This grid serves as an effect/value framework designed for evaluating the value of NLP technology within the government sector.

Informed by the findings of the literature review, we have integrated five dimensions of competitive factors into the value grid, namely, time (reducing the time required to manage data), interaction (enhancing communication with people), relationship (strengthening relationships with people), people (expanding reach to more individuals), and service (improving or introducing new services). Additionally, we have incorporated three benefits of value creation: effectiveness (enhanced management), efficiency (improved productivity), and strategic benefits (increased revenue by offering new products and services and entering new markets). This NLP value grid is both original and significant, developed specifically for the application of NLP in the government sector. The identified variables play a crucial role in the NLP societal value grid and are instrumental in enhancing entities' competitiveness within the government sector. Table 8 provides an overview of the NLP societal value grid for the government sector.

Table 7 Examples of societal benefits of NLP technology. *(Source: Table created by authors)*

• Different types of child abuse are identified and forecast through text mining and ML, thereby supporting children's health-related decisions (Amrit, et al., 2017).
• The use of NLP techniques to pre-processes online information related to identity theft can help identity threat behaviors, provide early warning, and avoid future crimes (Zaeem, et al., 2017).
• NLP technologies including sentiment analysis and topic modeling help to increase public participation and offer a reference in policy-making and project management (D. Zhang, et al., 2018)

- The use of NLP and DL–based techniques to forecast average sentiment trends, help to build more effective communication strategies and policies to maximize uptake of COVID-19 vaccines (Hussain, et al., 2021)
- LDA topics visualisation and validation provide information to the general public so that they can take wisely allocate resource and timely responsive strategies during natural disaster (Chang, et al., 2022)

Table 8 NLP societal value grid in the government sector. *(Source: Table created by authors)*

	Four Dimensions of the Applications of NLP Technologies	Value		
		Efficiency (Effi.)	Effectiveness (Effect.)	Strategy Benefits (SB)
Effect	Time (T)	Enables users to easily and accurately manage text from big data, thereby reducing administrative burdens and saving labor time	Enables the assessment of up-to-date and real-time information	Creates service excellence by improving information processing with decreasing error rates and costs
	Interaction (I)	Summarizes stakeholders' feedback for intelligent decision-making	Provides insights to improve services/policies that meet the target stakeholders' needs and concerns	Facilitates stakeholders engagement for information sharing and social interaction, while enabling organizations/agencies to substantially understand stakeholders' needs and make immediate responses
	Relationship (R)	Determines different stakeholders' actual opinions to facilitate understading of itself and stakeholders	Establishes improved stakeholder relationship using comments in big data	Monitors opinions, develops stakeholder relationship, and builds positive relationships with stakeholders by enhancing the efficiency and effectiveness of operations and information processes
	People (P)	Provides a wider population with access to big data from	Establishes stakeholder relationship with a wider population, and develops services to suit different stakeholders	Benefits citizens with global reach capabilities

	Service (S)	Automates tasks using NLP techniques	Improves services and provides news services to stakeholders (e.g., translation and e-mail filtering)	Bundles information and services

3.2. Discussion of the findings from the literature review

The findings of this study present a comprehensive overview of the current state of NLP applications in the government sector. The following subsections provide detailed analyses of the literature review, critical reflections, and implications.

3.2.1. Variability in NLP techniques across government domains

There is a noticeable variability in the techniques employed across different government domains. For instance, the health sector has adopted a wide range of NLP application classes and techniques, whereas the housing and community amenities sector has seen only limited adoption of NLP applications and techniques. This variability indicates the existence of resource and capability disparities among government sectors. Factors such as technical expertise, funding, and regulatory environments may influence the uneven distribution of NLP adoption. To address these issues, policymakers and practitioners may consider targeted strategies, such as providing funding for technology upgrades or creating frameworks to encourage the adoption of NLP across all sectors.

3.2.2. Development of theoretical frameworks

The literature reveals that NLP research in government applications has predominantly been empirical since the 1990s, there is a notable lack of theoretical foundations to explain the broader implications of these technologies. This gap restricts our understanding of the long-term impacts of NLP on government operations and

societal outcomes. To address this, our study emphasizes the need for developing theoretical framework/models, such as those based on organizational theories, to deepen our understanding of the interaction between NLP and government-citizen dynamics. These models can facilitate better decision-making frameworks and enhance citizen engagement.

3.2.3. Key trends and lack of Asian government context

Our review highlights that NLP techniques like text classification, sentiment analysis, and topic modelling are commonly used for addressing challenges such as policy analysis and citizen feedback management. However, these applications are predominantly found in the healthcare sector, likely due to resource disparities and difference in digital infrastructure across government departments. Importantly, the literature review identifies a significant research gap in the design, development, and real-world impact of NLP applications within Asian government contexts. This gap is critical as many non-Western governments are increasingly adopting NLP for governance, yet there is limited research on adapting NLP to their languages and contexts.

4. Phase two - Case study

Building on the insights from the literature review, the case study presented in this section illustrates the design and development of a social media text analytics system within a government department in Hong Kong. This case study directly addresses the gaps identified in the literature by showcasing a tailored implementation for Cantonese, aligning with the unique needs of non-Western government contexts.

The insights gained from the literature review directly informed the design and implementation of the case study in several ways. These include addressing the lack of

theoretical grounding identified in the literature, focusing non-western contexts, mitigating infrastructure challenges. This demonstrates how the literature review informed the case study and ensures its relevance and applicability to the government sector.

The case study demonstrates how challenges such as limited digital infrastructure can be mitigated through targeted design strategies. By focusing on a real-world application, the case study contributes to theoretical development by utilizing the Societal Value Grid to map the societal outcomes of NLP applications. This approach not only validates the findings from the literature review but also provides practical insights into the implementation of NLP technologies in government settings.

Case studies serve multiple purposes, including exploration of new areas, description of complex events or interventions, and explanation of complex phenomena (Lichtenthaler, 2004). This study employs a case investigation methodology to provide a detailed account of the development of a social media text analytics system — an exploration of a new area in government sector applications. The case study not only describes the design and development of the prototype system but also elucidates the value derived from its implementation (Ngai et. al., 2015).

The subject of this case study is a government department in Hong Kong that leverages advancements in information and communication technology to enhance and foster communication with various sectors of society. In this section, we present the outcomes of a case study focused on the implementation of the aforementioned framework within this government department. The discussion sheds light on issues related to the development of a social media text analytics system, illustrating its benefits and advantages. Government departments worldwide are increasingly adopting text mining, NLP, and AI to bridge the interaction gap between citizens and the government,

ultimately aiming to enhance the quality of government services. The examined department in Hong Kong is no exception to this trend. Notably, NLP works synergistically with text mining, facilitating the grouping and categorization of words to extract structured and unstructured data from social media platforms.

4.1. Research questions and objectives

The Internet has emerged as a central pillar of our information society, providing numerous channels for individuals in Hong Kong to express their opinions. Written Cantonese, prevalent in social media, serves as a key medium for the younger generation to voice their perspectives on diverse subjects, including social policy and government-related issues. Policymakers recognize the importance of collecting and analyzing online public opinion to inform more effective policy-making (Ngai, Lee, Choi, & Chai, 2018).

Cantonese, an essential dialect in Southern China (encompassing Hong Kong, Macau, and Guangdong) and among overseas Chinese communities (e.g., Canada and the United States), exhibits unique characteristics, including pronunciation, grammatical nuances, and a thousand additional characters created specifically for Cantonese (Cheung & Bauer, 2002). These distinctions render written Cantonese unintelligible to Mandarin speakers. Changes in media attitudes and public perceptions in Hong Kong over recent decades suggest an expanding social role for written Cantonese in the future (Snow, 2008). The rise of Web 2.0 has led local online users to frequently express ideas, opinions, and views in written Cantonese through social media platforms (Ngai, et al., 2018).

While sentiment classification research has been conducted in English, Chinese, and Japanese, there is a notable gap in exploring Cantonese sentiment classification for social media reviews, despite the global influence of Cantonese and its usage by 70 million people. Although various social media text analysis solutions (Abbasi, et. al., 2018) are

available, they predominantly focus on textual analysis for English, Chinese, and Japanese languages. Limited research exists on conducting textual analysis specifically for Cantonese.

Recognizing the increasing significance of written Cantonese and the analysis of written Cantonese in social media, we consider it a substantial endeavor to address the existing performance gap and unleash the potential of social media text analytics capabilities. Our objective is to design and develop an intelligent decision support system for social media sentiment analysis, specifically tailored to facilitate Cantonese textual analysis within the government sector. This system is intended to harness the power of text mining, NLP, and AI through the creation of a prototype that supports Cantonese textual analysis for government applications. The study seeks to address the following key questions:

1. How can NLP techniques be tailored to enhance public sector decision-making in non-Western contexts, specifically in regions with linguistic and cultural diversity (e.g., Cantonese in Hong Kong)?
    i) Existing studies on NLP in government have largely focused on English-language texts and Western-centric contexts, particularly in areas like policy analysis and public opinion monitoring. By addressing the need for NLP techniques in non-Western, linguistically and culturally diverse regions, this question directly fills a significant gap identified in our review. It will offer new insights into how NLP can be adapted to local languages (e.g., Cantonese) and cultural nuances, which is particularly urgent in global settings where multilingual and multicultural societies are the norm.

ii) This is an urgent area for investigation, as many governments in non-Western countries are increasingly relying on NLP for governance, yet the research on how to adapt NLP for these languages and contexts is limited.

2. What are the challenges and opportunities in integrating NLP-based text analytics systems in government departments with limited digital infrastructure, and how can these be addressed?

   i) While many studies have looked at NLP in well-resourced government agencies, fewer have examined how these technologies can be effectively implemented in low-resource environments. The literature review revealed a lack of research on overcoming the infrastructural and digital readiness barriers in less developed or emerging economies. This question brings attention to practical solutions and identifies specific strategies to address these challenges, thus advancing the understanding of NLP scalability in diverse governmental settings.

   ii) This question is critical because many governments face infrastructural constraints, and without addressing these, the potential of NLP in public sector applications could be stifled, particularly in developing economies.

3. What are the ethical implications of deploying NLP technologies in the public sector, particularly concerning privacy, bias, and accountability in data usage?

   i) Although there is some attention to ethics in the broader field of AI, there is insufficient exploration of ethical challenges unique to NLP in government. Our study highlighted that concerns related to privacy, bias, and transparency are not well addressed in the literature on NLP in the public sector. This question focuses on identifying the specific ethical challenges governments

face in implementing NLP, such as biases in algorithms and the handling of sensitive citizen data, which are critical for ensuring trust and accountability.

ii) With the rapid adoption of NLP in government decision-making and citizen engagement, addressing the ethical implications is urgent to avoid potential harms and public backlash. Moreover, these concerns are growing, particularly in light of global discussions on AI ethics and accountability.

4. What measurable societal impacts (e.g., faster service delivery, more inclusive policy-making) can be expected from the implementation of NLP systems in government, and how can these impacts be assessed?

   i) The literature largely focuses on the technical capabilities of NLP and its use in specific applications but lacks detailed studies on the societal outcomes of NLP implementation in the public sector. This question aims to address this gap by focusing on the societal value and public benefits of NLP, such as improving citizen engagement, transparency, and efficiency in government services. It also aims to propose frameworks for measuring these outcomes, which are critical for evaluating the true value of NLP in government.

   ii) As governments look to justify their investments in NLP, understanding the tangible benefits—both in terms of efficiency and public value—is critical. Measuring these impacts is essential for ensuring that NLP adoption leads to improved societal outcomes, thus making this research particularly timely.

4.2. Architectural framework of the social media text analytics system

In the subsequent section, we delineate the architectural framework of the social media analytics system, aiming to harness the potential of NLP for government sectors. A proficient system architecture serves as a blueprint for constructing the system, offering clarity on components, defining their functionalities, and specifying their

relationships and interactions (Nunamaker, Chen, & Purdin, 1990). The system architecture encompasses three primary components: the user interface, core system, and media data collection system (refer to Fig. 3).

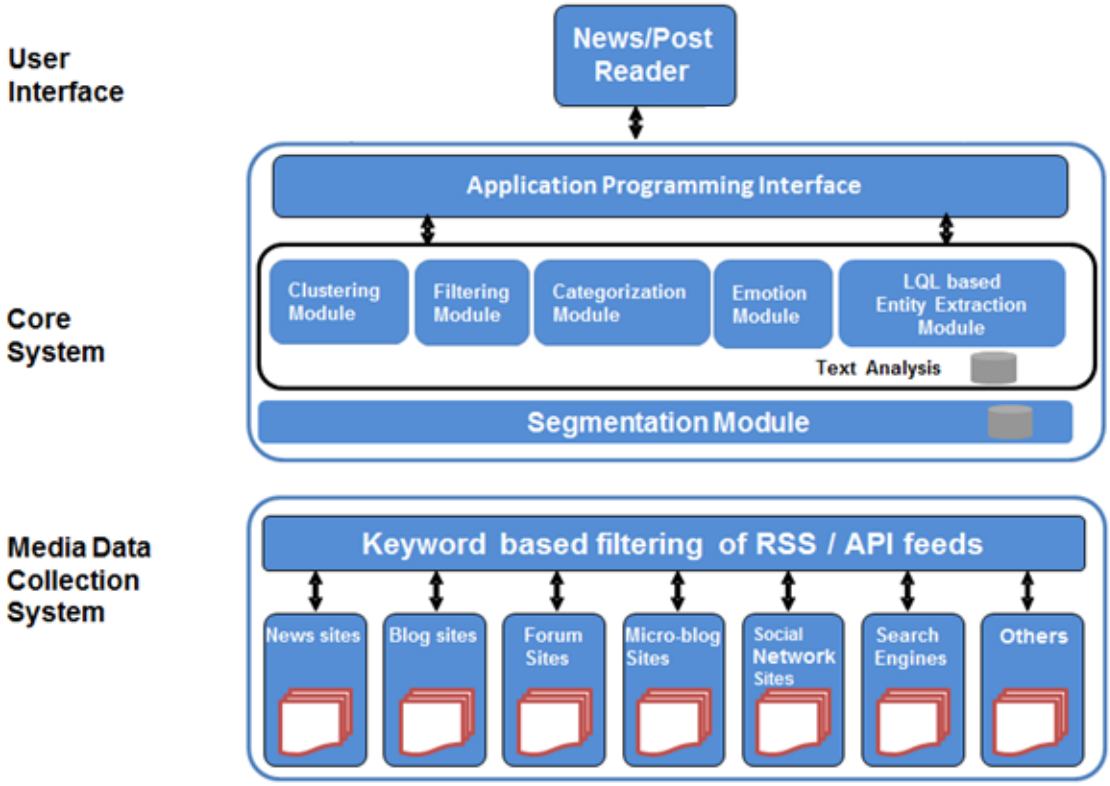

Fig. 3. Architectural Overview of the Social Media Text Analytics System
(Source: Figure created by authors)

1. Media data collection system: This component is tasked with capturing data from diverse sources. Textual content is retrieved from news websites, blogs, discussion forums, and social network platforms based on user-provided keywords' search results on the Internet. Additionally, specific websites incorporate a media data collection interface for extracting the necessary content. Subsequently, the collected content undergoes processing by a filter designed to eliminate unwanted noise before analysis.
2. Core system: The core system incorporates essential modules for data processing and analysis, outlined as follows:

- Segmentation module: This module undertakes cleansing and parsing tasks, including stop words removal, segmentation, stemming or lemmatization. It stores the text document in a structure, such as the Vector Space Model, which is more suitable for analysis. Key terms are determined as indexes based on their weights in the documents, with the weight calculated using TF–IDF.
- Clustering module: This module dynamically groups similar documents without pre-defined categories. The content's similarity is calculated, and documents are grouped together using an agglomerative hierarchical clustering method. This enables the identification of "hot-topics."
- Filtering module: As a content-based filter, this module removes junk and unwanted posts, such as advertisements, if specific words appear in the content and predefined rules are satisfied.
- Categorization module: After content analysis, this module categorizes text documents into different domains, such as social media (including web blogs, Facebook, and Twitter), news, and e-commerce, based on predefined rules.
- Emotion module: This module conducts sentiment analysis, also known as opinion mining, to gauge people's views, sentiments, evaluations, appraisals, attitudes, and emotions toward entities. It assesses entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. The sentiment analysis results enable the system to discern the satisfaction level of products and services or gauge the impact of a marketing campaign launch in the digital world. In this system, a three-stage classifier proposed by Ngai et al. (2018) is adopted. This sentiment classification combines three classifiers: SVM with NB features (NBSVM)-based classifier, a CNN-based classifier, and a lexicon-based classifier, as illustrated in Fig. 4. The results

from the three classifiers are combined using the weighted voting method. The proposed approach supports social media analysis in the Cantonese context and has been verified to outperform baseline approaches commonly used in the literature (Ngai et al., 2018).

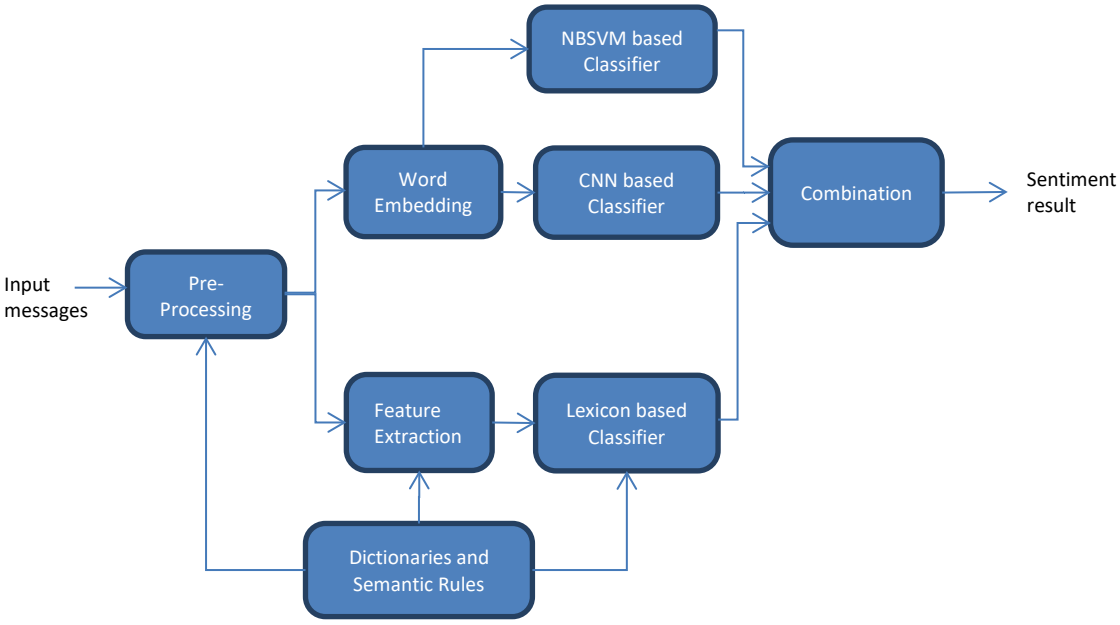

Fig. 4. The hybrid sentiment analysis framework (Ngai et. al., 2018)
(Source: Figure courtesy of Ngai et. al., 2018)

- Linguistic query language (LQL)-based entity extraction module: This module offers pre-defined linguistic rules in LQL for extracting features, including entities and topics. These rules are subsequently translated into computing rules and state machines to extract knowledge based on tagged and segmented text results. The LQL-based text analysis extraction method and system secured a Chinese Patent for Invention (Ngai, Choi, & Chai, 2017).

3. User interface: This system offers a user interface for accessing reports, refining rules and keywords, and ensuring proper usage. The interface encompasses various modules for citizen voice collection and analysis, as well as knowledge and hot topic discovery. Through a graphical user interface, posts from diverse

Internet data sources can be conveniently viewed on a single platform. Additionally, a tag cloud displays keywords identified in the posts. Users have the ability to further filter or conceal undesired posts by adding exclusive keywords, reporting junk domains, and linking to the original posts. The posts are categorized and can be displayed in different views, such as being grouped by districts in Hong Kong. Fig. 5 illustrates the menu screen for the "information map." The web-based interface furnishes a dashboard and infographic for summarizing data and information.

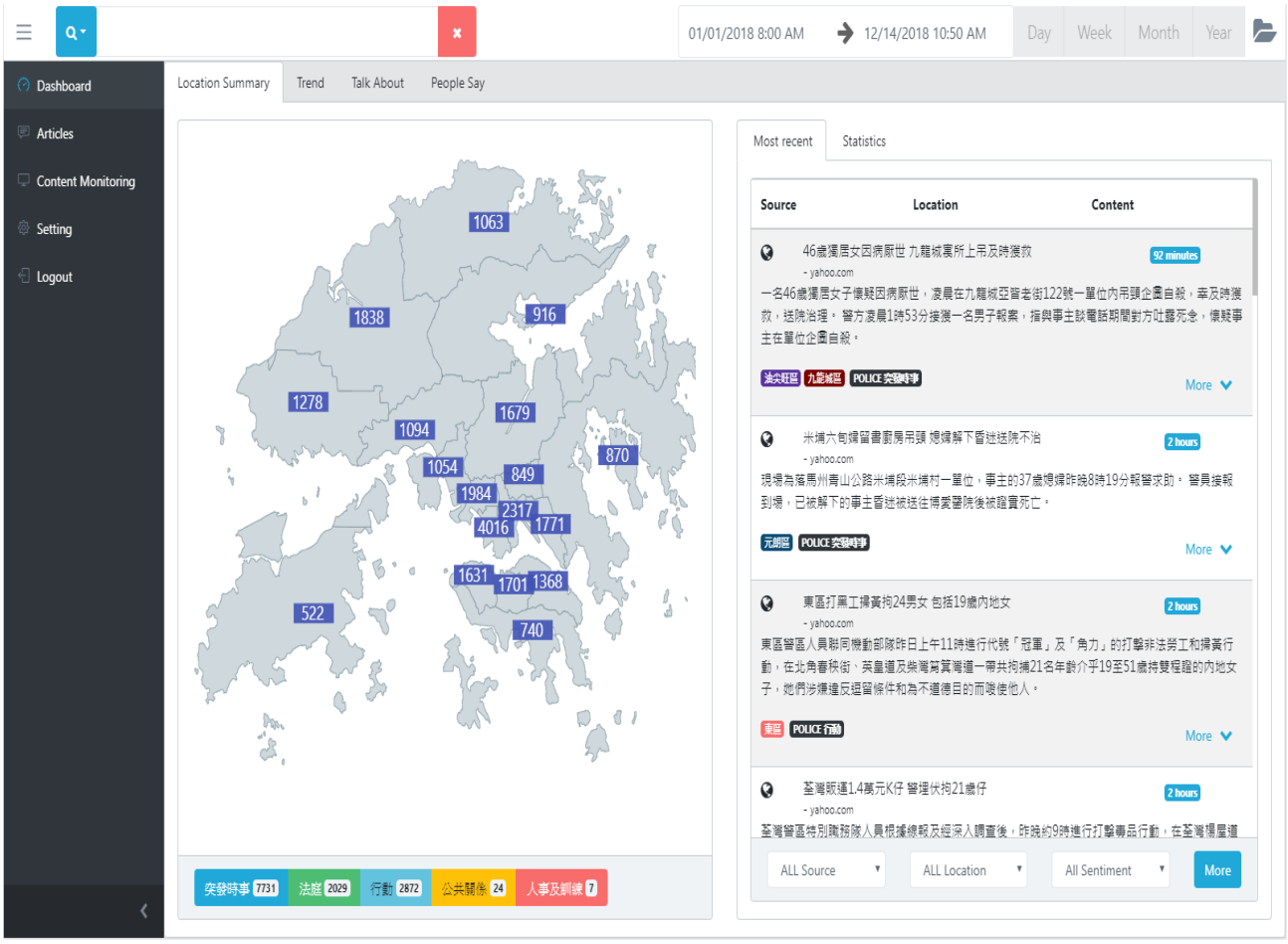

Fig. 5. "Information map" menu screen *(Source: Figure created by authors)*

4.3. Pilot test of the proposed system

4.3.1. System Prototype Building

Based on the forementioned social media text analytics system, we developed a prototype system for a pilot test. The development of a prototype system aligns with engineering concepts (Scott Morton, 1984) and is a prevalent practice in software engineering research (Nunamaker Jr, et al., 1990). The prototype is a web-based system constructed using Hyper Text Markup Language (HTML) and the Java programming language, ensuring portability and compatibility with most web browsers. As a web-based system with standard interfaces, it requires no special client software beyond a web browser like Google Chrome or Firefox. Addressing the crucial concern of data security on the Internet, the system employs an authentication mechanism to control access to user records. Security measures include password authentication and verification.

4.3.2. System observation and evaluation

To evaluate the effectiveness and usability of the NLP-based social media text analytics system, a pilot test was conducted with 30 participants, including both potential users and senior government officers. These participants were selected based on their roles in data management and processing tasks relevant to the system's implementation. The majority of participants had experience with basic NLP tools, while all had practical experience with the application of NLP systems in the public sector. This diverse mix of potential users and senior officers allowed for a comprehensive evaluation, incorporating both technical insights and user-centered perspectives on the system's usability and potential impact.

Quantitative data from the feedback received through the questionnaires has been recorded, and the summarized users' feedback is presented below:

i) Mean responses to system evaluation by potential users. This can be achieved through a questionnaire survey. The potential users were instructed to employ five-point scales (one = strongly disagree, three = undecided, and five = strongly

agree) to assess the two main aspects of the prototype system: its effectiveness and its usability. The results of analysis of questionnaire are shown in Table 9 (Effectiveness: mean = 4.11, S.D.= 0.24; Usability: mean = 4.24, S.D.= 0.25). In addition to reporting mean values and standard deviations, we have conducted a more thorough analysis of the reliability and internal consistency of the questionnaire items. Specifically, we performed a Cronbach's Alpha test to assess the internal consistency of the scale, ensuring that the items measuring the same constructs were correlated and reliable. The results of the reliability test showed acceptable values (Cronbach's Alpha > 0.7) for all key constructs, indicating that the questionnaire items were internally consistent.

Table 9 The questionnaire results of system evaluation. *(Source: Table created by authors)*

Effectiveness of the Prototype System	**Mean***	**Standard Deviation**
The social media text analytics prototype system can		
1. assist in gathering the opinions of public from different Chinese social media	4.38	0.64
2. assist in monitoring and listening to people's voices and also provide insights into their behavioral trends	3.89	0.54
3. identify key communities, specific topics and influencers	4.31	0.60
4. improve the decision-making process by analysis the people opinions	3.86	0.72
Mean =	4.11	S.D. = 0.24
Usability of the Prototype System		
The social media text analytics system		
5. is easy to use	4.30	0.76
5. is easy to use	4.30	0.76
6. is easy to learn to use it	4.54	0.53
7. is user friendly	4.29	0.78
8. is stable and reliable	3.88	0.61
9. can be used without any difficulty	4.10	0.95
10. I likely to recommend	4.35	0.88
Mean =	4.24	S.D.= 0.25

*Scale: 1 = strong disagree, 3= undecided, 5= strongly agree

ii) Qualitative feedback of system evaluation. Table 10 shows the data from the interview transcripts in the system evaluation. This table illustrates the initial

categorization of societal benefits linked to the government applications of NLP technology based on our proposed societal value grid (Table 8).

Table 10 Interviewees' feedback on the system (the qualitative feedback of system evaluation). *(Source: Table created by authors)*

Interviewee	Illustrative quotes in Societal Value Grid Dimension: (Time (T); Interaction (I); Relationship (R); People (P) and Service (S) Value Creation: (Efficiency (Effi.); Effectiveness (Effect.); Strategy Benefits (SB))
Interviewee 1	**(T) + (I) + (R) + (P) + (S) -> (Effi.) + (Effect.):** "The prototype system can provide what we needed to **identify patterns and some insights** from social media or online source, allowing us to improve **effectively and efficiently** in operations and public **services**. We may give a try".
Interviewee 2	**(T) + (I) + (R) + (P) -> (Effi.) + (Effect.) + (SB)**: "Analyzing public **feedback** in particular the unstructured context and **understand** citizen and business **concerns** and can be better serve the public. This is good to have a tool such that policy maker can use it to **connect with people** and measure public opinion on polices more **quickly and more effectively"**.
Interviewee 3	**(T) + (I) + (R) + (P) -> (Effi.)**: "The sentiment analytics tool can read behind **human mind** in the social media platforms, websites for **public comment** based on their AI models. It can be used like a "response tracker" to track **people's response** to dynamic local contents, hot topic among local people, etc".
Interviewee 4	**(T) + (I) + (R) + (P) -> (Effi.) + (Effect.) + (SB)**: "This system can help mine public sentiment and understand what citizen **preferences** and citizen **complaints** that can be classified into specific categories and provide to **appropriate** government department for making **immediate responses"**.
Interviewee 5	**(T) + (I) + (R) + (P) -> (Effi.)**: "I believe that public opinion may be computed by the system because it can gather the **opinions** of public through different social media channels. The system can identify their key communities, specific topics and influencers".
Interviewee 6	**(T) + (I) + (R) + (P) -> (Effi.) + (Effect.) + (SB)**: "We can **easily identify** patterns and classify content into meaningful topics so we can easily discover the **trends and opinion** in social media and online forums. We may even use it to active listening and identify noises. Hopefully, potential crisis and potential social unrest can be **read and managed effectively** through the system".
Interviewee 7	**(T) + (I) + (R) + (P) -> (Effi.) + (Effect.)**: "This system with Cantonese function has become indispensable in the monitoring and listening to people's **voices** and also provide **insights** into their **behavioral trends** as well as **public sentiments** in all popular sources of Chinese social media. Therefore, we can benefit by understanding people sentiments toward government **new policy"**.
Interviewee 8	**(T) + (I) + (R) + (P) -> (Effi.): "**It does not make sense to spend a lot of **time** to read the feedback and opinion in the social media platform. Nowadays, with using the text mining technologies like this social media text analytics system, we are pleased to see the **summarized results** from the system. The issue here is not only monitoring, but squeezing increasing **efficiency** out of the existing operations".

Interviewee 9	**(T) + (I) + (R) + (P) -> (Effi.) + (Effect.)**: "We have tried a sentiment analysis system but this one is a tailor-made for Cantonese speaker. With the **real-time data** and emotions capturing capabilities, it can identify **relationships**. We can use it to collect data and opinions from people. This can assist us in **decision-making"**.
Interviewee 10	**(T) + (I) + (P) + (S) -> (Effi.) + (Effect.) + (SB):** "We can improve the **decision-making** process by analyzing the **voices** of citizen and people in social media platform. We can reach and listen to **more people**. This is definitely useful to us to use the system to improve daily operations, **reduce costs** and increase the quality of **services**".

Several comments on the questionnaires are superficial, such as "the system is good" or "the system is useful". In general, users express satisfaction with the system. The following are user comments for the functional improvement of the prototype system.

- "The system should allow users to search news by the same date (i.e., same start and end date of the search range)".
- "There should be a filter of time range for searching the news / Facebook posts".
- "There should be a function to allow users to add new links / Facebook pages or groups".
- "The system should respond if users click on the hot tag once, instead of double click".
- "There should be an additional folder, "Team", which is owned by a user but not shared with others. Users can modify it via the settings page".
- "The "Sentiment Filter" and the sub-page of "Sentiment Statistics" should be removed in the Dashboard – "Local Summary" to improve user experience".

4.3.3. Model Evaluation

To substantiate the effectiveness of the proposed NLP-based social media text analytics system, we have incorporated an evaluation of the model's performance using widely accepted metrics such as precision, recall, and F1-score. These metrics provide a

comprehensive understanding of the system's ability to correctly classify and identify relevant information from social media data.

In this study, we evaluated the performance of our sentiment classification model for Cantonese content, focusing on its ability to accurately classify social media posts into one of three sentiment categories: positive, neutral, and negative. This task is framed as a single-label multi-class (SLMC) classification problem, where each message is assigned to exactly one class.

Sentiment Classification Performance: We assessed the model's effectiveness using a dataset comprising 3,156 social network posts. The model achieved the following performance metrics:

- Precision**:** 73.2%
- Recall**:** 77.5%
- F1 Score**:** 75.3%

These metrics indicate a balanced performance, with the F1 score reflecting the harmonic mean of precision and recall, thus providing a robust measure of the model's accuracy in sentiment classification.

Segmentation Performance: In addition to sentiment classification, our system includes a segmentation component that identifies and categorizes entities within the text, such as organizations, people, and locations. We evaluated the precision of this segmentation task by comparing the system's output against manually verified results. Precision was calculated using the formula:

Precision = Number of Correct Tags / (Total Number of Tags) ×100%

The precision results for each category are detailed in Table 11.

Table 11 Precision of the proposed system in segmentation. *(Source: Table created by authors)*

Category	No. of Tags	No. of Correct Tags	Precision (%)
Organizations	9,282	8,633	93.0
People	13,620	9,670	71.0
Locations	16,531	13,429	81.2
Overall	39,433	32,769	83.1

The segmentation component demonstrated high precision, particularly in recognizing organizations and locations, with overall precision reaching 83.1%. This indicates the system's strong capability in accurately identifying and categorizing entities within the text.

Overall, the evaluation results substantiate the effectiveness of our proposed model in both sentiment classification and text segmentation tasks, providing a solution for analyzing Cantonese social media content.

5. Discussion

5.1. Some lessons learned

As illustrated in this case study, social media text analytics system that a government department can adopt to offer better governance and enhance the citizen experience has the potential of fundamentally transforming the government department operations. Some lessons learned from this study can serve as a guideline that future implementers of text mining and NLP technologies can adapt to their situations. The following lessons learned are described as follows:

1. Government department should experiment on how an AI/text mining and NLP applications works and when and what part of operations require AI to facilitate the adoption of change. It constantly starts with a small scope because the lessons learned from the success of a small project can be easily transferred to the entire

department. For example, complaints from citizens can be automatically classified into specific categories and sent to appropriate person in charge in the department.

2. Co-operation of users and officers from departments (e.g., IT, legal, etc.) is important as expertise in different areas would be required in the field test. The IT department will need to set up a modern and updated IT infrastructure, and legal officers may help set up a legal framework related to AI considering ethics and privacy concerns.
3. The government sector has historically been slow to adopt new technologies. Usually, much of the IT budget went to maintaining or updating legacy systems instead of investing in innovative new solutions (Anand, 2019). We find that more innovative top management would be "willing to experiment AI." Obtaining top management support is one of the key success factors influencing the success of implementing the AI/text mining NLP-based systems. Top management support can ensure sufficient resource support for the project and set up the strategic direction. Moreover, management must understand that AI/text mining and NLP technologies will not magically make everything happen automatically. Management's expectations and requirements should be clearly identified at the proposal stage. Otherwise, excessive expectations can lead to disappointment of the management.
4. Pilot run and onsite testing are a few of the essential steps in assessing the readiness of implementation of the AI systems. For this system, we have piloted run in the government department for approximately 12 months with ongoing validation of the AI models in practice. Algorithms developed and validated in research and lab environments are not automatically applicable during deployment.

Moreover, the lack of specialization in AI among the government staff is a typical concern.

5.2. Challenges in implementation

While NLP technologies offer numerous benefits, such as enhancing efficiency, improving public services, and supporting decision-making processes, there are also potential challenges and risks associated with their implementation in the public sector. These include issues related to data privacy, algorithmic bias, and transparency, as well as the potential for misuse or misinterpretation of text analytics.

1. Data Privacy and Security: NLP applications in government frequently require the processing of sensitive or personal data from citizens, such as information extracted from social media, public records, or feedback channels. This raises significant concerns about data privacy and the potential for unauthorized access or breaches. Without stringent safeguards, there is a risk of violating individuals' privacy rights, which could lead to public distrust or legal ramifications. Ensuring robust data protection measures, such as encryption, anonymization, and access control, is essential to minimize these risks while leveraging NLP for improved public services.
2. Algorithmic Bias: NLP systems trained on large datasets may inadvertently reinforce or even amplify existing biases within the data. If the training datasets contain biased language, cultural stereotypes, or unequal representation, the resulting NLP applications might produce biased or discriminatory outcomes. In a government context, this could affect fairness in public services and policymaking, leading to unintended discriminatory practices or unequal resource distribution. Addressing bias mitigation is therefore crucial for ensuring that NLP systems support equitable and just governance.

3. Transparency and Accountability: Many advanced NLP models, such as those using deep learning, operate as "black boxes," where the decision-making process is opaque and challenging to interpret. This lack of transparency could undermine accountability, particularly in situations where NLP applications influence public policy or individual rights. Therefore, developing explainable NLP models and providing clear documentation about how NLP decisions are reached are essential for maintaining public trust and ensuring that government agencies remain accountable for their actions.
4. Misuse or Misinterpretation: The application of NLP in analyzing social media and public sentiment presents risks of misinterpretation or over-reliance on sentiment scores without context. For instance, NLP systems may misinterpret nuanced language such as sarcasm, regional idioms, or cultural expressions, which could lead to an inaccurate understanding of public sentiment. Additionally, the amplification of harmful or misleading content is a potential risk when governments rely on NLP-based social media analytics without sufficient checks and balances. Ethical guidelines and a cautious approach to interpreting NLP results are necessary to prevent unintended social tensions or misinformation.

5.3. Competitive advantage

In the public sector, the concept of "competitive advantage" traditionally used in the business context may appear out of place. However, we frame this idea within a comparative advantage perspective to better align with the unique goals and values of the public sector. Unlike the business market, where competitive advantage refers to outperforming rivals, in the public sector, comparative advantage refers to the ability of government agencies and public institutions to leverage technologies like NLP to more

effectively achieve public sector objectives. This advantage is rooted not in market competition, but in enhancing public service delivery, improving transparency, and addressing societal needs with greater efficiency.

Specifically, NLP technologies can provide comparative advantages in the following ways:

1. NLP systems, such as social media text analytics, enable government agencies to quickly process public sentiment and feedback. By efficiently identifying emerging issues and responding to citizens' concerns in real-time, governments can deliver services that are more timely, relevant, and aligned with public needs.

2. In government operations, NLP can streamline the processing of large volumes of text data, such as citizen communications, regulatory documents, and public policy reports. This leads to faster decision-making, reduced operational costs, and more transparent governmental functions—all of which serve to enhance public trust and accountability.

3. With NLP tools, governments can gain insights into the opinions and concerns of diverse citizen groups, allowing for more inclusive and evidence-based policy decisions. The comparative advantage here is the government's ability to make more informed decisions that reflect the true needs and preferences of the public.

4. The use of NLP in the public sector goes beyond efficiency—it enables the promotion of social justice and equity. By identifying disparities in service delivery or recognizing underserved populations, NLP can help ensure that government actions are inclusive and aligned with the principles of fairness and social welfare.

Thus, while the business sector may focus on market-driven competitive advantages, the public sector's "comparative advantage" lies in utilizing technologies like NLP to improve public administration, promote fairness, and meet the evolving needs of citizens

more effectively and efficiently. This approach ensures that NLP applications in the public sector are not only technologically advanced but also align with the broader objectives of public service and governance.

5.4. Challenges in Implementation

5.4.1 Cantonese Dictionary

One unique feature of the system is its support for text analysis of content written in Cantonese, including slang and idioms commonly used in Hong Kong. To build this capability, we developed a Cantonese dictionary that draws from standard references such as the *Guangzhou Cantonese Slang Dictionary* (“廣州話俗語詞典”) and the *Practical Cantonese Classification Dictionary* (“實用廣州話分類詞典”). Additionally, we incorporated entries from online resources specific to Hong Kong Cantonese, such as the *Common and Authentic Hong Kong Cantonese* dictionary (“香港常用及地道式廣東話”), *Yuet Dictionary (“粵典”)*, *Open Dictionary (“開放詞典”)*, and the *Hong Kong Net Culture Encyclopedia (“香港網絡大典”)*. Despite these resources, the initial dictionary proved insufficient for effective sentiment classification.

To enhance the dictionary, we recruited student assistants from a local university to label sentiment words from a vast corpus of social media posts. These entries were reviewed by experienced developers, and previously unlisted terms were added to the dictionary. Although this process was resource- and time-intensive, it significantly improved the system’s ability to perform accurate sentiment analysis. However, as Cantonese slang and idioms are constantly evolving, particularly in online discourse. To stay ahead of these changes, continuous monitoring and updating of the dictionary is essential. This challenge underscores the necessity of building dynamic and adaptable

language models for Cantonese that can evolve alongside language trends in the digital age.

5.4.2. Sentiment Classification Approach

Using a purely lexicon-based approach for sentiment classification initially led to performance limitations. While expanding the dictionary improved accuracy to some extent, results plateaued. To address this, we implemented a machine learning-based approach, incorporating Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) within the system (Ngai et al., 2018). Sentiment classification was conducted through a combined approach, where each classifier contributed to a majority-vote system. This ensemble method allowed us to leverage the strengths of individual classifiers, resulting in improved accuracy over single-method approaches.

However, even with machine learning integration, the challenge of effectively capturing the rich semantics of Cantonese text persisted. Recent advancements in NLP, particularly transformer-based models like BERT and its Cantonese-specific variants, offer promising solutions. BERT's contextualized word embeddings and its ability to understand sentence-level semantics could significantly improve sentiment classification by more accurately capturing subtleties such as tone, context, and colloquial expressions. Moving forward, we plan to integrate such advanced NLP models to further enhance the system's performance, as they provide a more linguistically appropriate framework for handling Cantonese text.

Given the rapid advancements in AI, particularly with large pre-trained models like LLMs, including BERT and GPT, the scope for improving sentiment analysis in Cantonese text is expanding. These models offer significant advantages over traditional lexicon-based and shallow machine learning methods, particularly in capturing the semantic and syntactic intricacies of a language as nuanced as Cantonese. By leveraging

pre-trained models fine-tuned for Cantonese, we could achieve more accurate sentiment classification, handle code-switching between Cantonese and English, and even process multimodal data such as images and videos that often accompany social media posts.

5.4.3. Individual Preference

As the machine learning classifiers and dictionaries were initially developed by the project team, they may reflect the sentiment preferences of the general public. This can sometimes lead to classifications that do not align with the needs of specific user groups, such as law enforcement. To address this, we implemented a customization mechanism for individual users. When users find a misclassified post, they can re-label it; once enough re-labeled posts are available to form a new training dataset, users can re-train the classifier to better align with their specific preferences. This feature allows the system to adapt continuously to individual needs over time.

5.4.4. Irrelevant Content Filtering

Keyword-based post collection often results in irrelevant content, such as advertisements, cluttering the data feed. To address this, we introduced an "aspect" classifier, which categorizes posts into user-defined themes, including an "irrelevant" category. By training the classifier with labeled examples for each aspect, the system can filter out irrelevant content and allow users to view posts grouped by topic. This not only refines post searchability but also enables categorization based on language patterns and writing styles rather than simple keyword matches, enhancing user efficiency and focus.

6. Conclusion and future research

NLP technologies have received considerable attention from both academics and industry professionals. NLP enables machines to process and comprehend text data,

automating tasks such as sentiment analysis. The shift from understanding individual words to comprehending entire sentences empowers machines to process unstructured text data and extract meaningful insights. Governments can utilize NLP and text analysis technologies to mine public sentiment and expert content, gaining insights into citizen preferences and information relevant to policy propositions or implementations.

After conducting a literature review and case analysis, establishing a connection between NLP technologies and government, we underscore the potential implications of AI, text mining, and NLP technologies for government bureaus and departments. This realization leads us to propose a research agenda, outlining areas of interest and suggesting subjects for future research. The primary questions arising from the intersection of AI, text mining, and NLP technologies with government department research are as follows:

- How can government departments leverage big data, text mining, NLP, and social media text analytics to enhance the relationship between government and the community?
- How impact will AI, text analytics, NLP, and social media text analytics have on the adoption of e-government practices?
- What challenges might arise in applying big data, NLP, and text analytics for government purposes?
- In what ways can big data and AI technologies be employed for government and continuous improvement?
- How can the social text analytics system integrate with existing internal legacy systems? Addressing this aspect is likely to pose a significant challenge.

- How can government departments utilize privacy-preserving NLP techniques, such as federated learning and differential privacy, to secure sensitive data while benefiting from text analytics?
- What impact would fair NLP algorithms, designed to detect and adjust for biases in text data, have on government transparency and accountability?
- In what ways can explainable AI (XAI) enhance the interpretability of NLP models in government applications, and how could these improvements increase public trust in government use of AI technologies?
- How can sentiment analysis models be adapted to recognize complex language features such as sarcasm, irony, and culturally specific expressions in government contexts?
- How might government NLP applications be enhanced to better handle multilingual and cross-cultural differences, and what is needed to ensure accurate and sensitive communication with diverse communities?
- What role should standardized ethical frameworks play in the application of NLP for government purposes, and how can they be designed to incorporate legal, moral, and social considerations?
- How can government departments develop robust AI systems for text analytics that address cross-cultural variations while maintaining high interpretability and public accountability?

These questions aim to guide future research toward creating practical, secure, and ethically sound AI and NLP solutions for government applications

This study has certain limitations. First, in the literature review, we only included journals in this research because journals provide the highest quality of study, and

researchers and practitioners frequently use journals to distribute new results and obtain information (Nord & Nord, 1995). Grey literature, including conference papers and white papers, was not included in our analysis. Future research may benefit from incorporating grey literature to gain insights into the latest developments in NLP.

Second, we designed and developed the social media text analytics system to support a government department and evaluated it as a proof of concept. While the case study illustrates the application of social media analytics using NLP technologies to monitor social opinions in the Cantonese community, particularly in Hong Kong, the prototype system does have some limitations. It's essential to recognize the inherent limitation of a single case study, implying that external generalizability is constrained. Future research could mitigate this limitation by examining the system in various government departments. Despite the satisfactory outcomes in terms of effectiveness and usability, it is crucial to view this prototype system as an initial exploration in this promising area. Further research and enhancements, particularly in the entity sentiment and sentiment classifier of the system, are necessary. The following directions are worth pursuing:

i) Sentiment classifier: To facilitate continuous performance improvement, the social media analytics system allows users to improve the Cantonese dictionary and re-train the ML based classifiers. If users find incorrect sentiment classification results, they can report the error in the system and mark the correct label. The re-labelled data can be fed into the system as the dataset for re-training the ML classifiers. On the other hand, a group of experts can review the re-labelled data for updating the dictionaries. This suggestion can enhance the accuracy for sentiment analysis that is continuously performed by the system.

ii) Entity Sentiment: Sentiment analysis operates by analyzing text at the document level, treating an entire news story or social media post as one cohesive piece of

text with a single sentiment prediction for the entire document. However, document-level sentiment analysis faces a challenge: many documents discuss multiple items, and diverse sentiments are often expressed about each of these items. Consequently, relying solely on document-level sentiment analysis may not yield results with sufficient granularity to achieve a nuanced understanding of sentiments. As a solution, we recommend incorporating entity-level sentiment analysis into the system for enhanced precision.

7. Theoretical and Managerial Implications

7.1. Theoretical Contributions

This study provides a novel conceptual framework for understanding the integration of NLP in the public sector, linking NLP techniques directly to specific government functions using the UN's COFOG. This framework expands on existing models of AI in government by providing a systematic categorization that connects specific NLP applications with broader governance goals. This theoretical contribution helps bridge the gap between NLP technologies and public administration theories, offering insights into how NLP can drive decision-making and policy outcomes.

Our study makes a significant theoretical contribution through the development of the Societal Value Grid. This framework synthesizes the potential benefits of NLP applications across three dimensions—strategic, efficiency, and effectiveness—that are directly relevant to public sector objectives. By categorizing NLP applications within these dimensions, we provide a structured theoretical model that positions NLP as an impactful tool in governmental operations, not only for technical advancements but also for aligning with broader public sector goals and societal needs. This grid adds depth to existing theoretical frameworks by linking technological potential with societal

outcomes, guiding researchers and policymakers in understanding how NLP implementations can meet specific public sector needs.

The study also provides a theoretical understanding of how NLP can support policy formation by improving data-driven insights and fostering public engagement. This contribution advances our knowledge of how NLP can enhance the role of data in policy-making processes, particularly in the context of real-time public feedback through social media and citizen engagement platforms.

7.2. Practical Contributions

One of the key practical contributions of this research is the demonstration of how NLP-based systems can be used for real-time monitoring and evaluation of public policy outcomes. For example, by analyzing social media data in real time, governments can gather public sentiment and identify emerging issues, enabling more dynamic and responsive policy-making. This offers a practical tool for improving the effectiveness and accountability of government policies.

The study includes a case study of a Cantonese-language NLP system implemented in a Hong Kong government department, which provides insights into the design, challenges, and social value of such systems. The case study offers a practical guide for policymakers and practitioners on how to implement NLP tools that cater to local languages and cultural contexts, addressing the practical challenges of scaling NLP solutions in non-English, diverse linguistic settings.

Furthermore, the research identifies strategies for overcoming the digital infrastructure challenges faced by many governments, particularly in low-resource environments. By providing recommendations for government agencies looking to implement NLP-based systems, this study helps public sector leaders better understand

the necessary organizational changes and investments required for successful NLP adoption.

Practically, the Societal Value Grid is designed as a tool for government managers and decision-makers to assess and prioritize NLP technologies in ways that are directly tied to public service goals. By offering a structured means to evaluate the potential impacts of NLP, it assists managers and officers in aligning technology investments with key objectives like policy formation, social value, and community engagement. For instance, it enables policymakers to make informed choices about adopting NLP for citizen sentiment analysis, resource allocation, or service delivery improvements. Moreover, this framework underscores the importance of ethical and socially responsible NLP applications, helping to guide the practical implementation of these technologies in ways that resonate with public expectations and democratic values.

7.3. Influence on Policy Formation and Implementation

This study specifically explores how NLP can enhance policy analysis and decision-making by automating the extraction of insights from large volumes of public data. By facilitating the processing of citizen feedback, government reports, and public discourse on social media, NLP tools can provide real-time insights that inform both policy development and policy evaluation. This addresses a significant gap in existing research, which has often overlooked the policy-making implications of NLP applications.

NLP can play a critical role in inclusive policy-making. The study highlights how NLP systems can help identify underrepresented voices in policy discussions, ensuring that diverse public opinions are considered. This aligns with public sector objectives of transparency, accountability, and inclusivity in decision-making processes.

7.4. Practical Recommendations for Policy Makers

This study also includes recommendations for integrating NLP into government workflows. For example, it suggests that governments looking to implement NLP-based systems prioritize areas where large-scale text data is already being generated, such as public services, social media, or open government data. By starting with these high-impact areas, governments can demonstrate the utility of NLP in policy formulation, leading to further adoption across other sectors.

Moreover, the paper offers guidelines for ensuring the ethical use of NLP technologies in the public sector, including principles for data privacy, algorithmic fairness, and transparency in NLP deployments. This helps policymakers balance the potential benefits of NLP with the need to uphold democratic values and protect citizens' rights.

Acknowledgements

We sincerely appreciate the insightful and constructive comments and guidance provided by the anonymous reviewers, which have greatly enhanced our manuscript. Ngai was supported in part by RIAIoT, PolyU, under Grant CD4T, and by RCTFF, PolyU, under Grants BBEY and BBFM. Kei was supported in part by PolyU under Grants R-ZDDM and ZDEX.

References

Abbasi, A. Zhou, Y., Deng, S. and Zhang, P. (2018), Text Analytics to Support Sense-Making in Social Media, *MIS Quarter*ly, 427 – 464.

Adikari, A., & Alahakoon, D. (2020). Understanding Citizens' Emotional Pulse in a Smart City Using Artificial Intelligence. *IEEE Transactions on Industrial Informatics, 17*, 2743-2751.

Adikari, A., Nawaratne, R., De Silva, D., Ranasinghe, S., Alahakoon, O., & Alahakoon, D. (2021). Emotions of COVID-19: content analysis of self-reported information using artificial intelligence. *Journal of Medical Internet Research, 23*, e27341.

Alahmari, N., Alswedani, S., Alzahrani, A., Katib, I., Albeshri, A., & Mehmood, R. (2022). Musawah: A Data-Driven AI Approach and Tool to Co-Create Healthcare Services with a Case Study on Cancer Disease in Saudi Arabia. *Sustainability, 14*, 3313.

Aljameel, S. S., Alabbad, D. A., Alzahrani, N. A., Alqarni, S. M., Alamoudi, F. A., Babili, L. M., Aljaafary, S. K., & Alshamrani, F. M. (2021). A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent COVID-19 outbreaks in Saudi Arabia. *International Journal of Environmental Research and Public Health, 18*, 218.

Alkhatib, M., El Barachi, M., & Shaalan, K. (2019). An Arabic social media based framework for incidents and events monitoring in smart cities. *Journal of Cleaner Production, 220*, 771-785.

Amina, B., & Azim, T. (2019). Scancpeclens: a framework for automatic lexicon generation and sentiment analysis of micro blogging data on China pakistan economic corridor. *IEEE Access, 7*, 133876-133887.

Amrit, C., Paauw, T., Aly, R., & Lavric, M. (2017). Identifying child abuse through text mining and machine learning. *Expert Systems with Applications, 88*, 402-418.

Anand, N. (2019), New decade, new tech: What government agencies will prioritize in 2020, https://federalnewsnetwork.com/commentary/2019/12/new-decade-new-tech-what-government-agencies-will-prioritize-in-2020/

Aslan, Ç. B., Li, S., Çelebi, F. V., & Tian, H. (2020). The world of defacers: Looking through the lens of their activities on Twitter. *IEEE Access, 8*, 204132-204143.

Ayers, J. W., Dredze, M., Leas, E. C., Caputi, T. L., Allem, J.-P., & Cohen, J. E. (2018). Next generation media monitoring: global coverage of electronic nicotine delivery systems (electronic cigarettes) on Bing, Google and Twitter, 2013-2018. *PLoS One, 13*, e0205822.

Başarslan, M. S. & Kayaalp F. (2023). MBi-GRUMCONV: A novel Multi Bi-GRU and Multi CNN-Based deep learning model for social media sentiment analysis. *Journal of Cloud Computing*, *12* (1), 5.

Basilio, M. P., Pereira, V., & Brum, G. (2019). Identification of operational demand in law enforcement agencies: an application based on a probabilistic model of topics. *Data Technologies and Applications, 53*, 333-372.

Bennett, J., Rachunok, B., Flage, R., & Nateghi, R. (2021). Mapping climate discourse to climate opinion: An approach for augmenting surveys with social media to enhance understandings of climate opinion in the United States. *PloS one, 16*, e0245319.

Bose, R. (2009), Advanced Anbalytics: Opportunities and Challenges, *Industrial Management & Data Systems*, 109, 2, 155 - 172.

Boon-Itt, S., & Skunkan, Y. (2020). Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. *JMIR Public Health and Surveillance, 6*, e21978.

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine, 9*, 48-57.

Chang, H., Li, L., Huang, J., Zhang, Q., & Chin, K.-S. (2022). Tracking traffic congestion and accidents using social media data: A case study of Shanghai. *Accident Analysis & Prevention, 169*, 106618.

Chassagnon, G., Vakalopolou, M., Paragios, N., & Revel, M.-P. (2020). Deep learning: definition and perspectives for thoracic imaging. *European radiology, 30*, 2021-2030.

Chen, J., Lalor, J., Liu, W., Druhl, E., Granillo, E., Vimalananda, V. G., & Yu, H. (2019). Detecting hypoglycemia incidents reported in patients' secure messages: using cost-sensitive learning and oversampling to reduce data imbalance. *Journal of Medical Internet Research, 21*, e11990.

Chiang, Y.-C., Chu, M., Lin, S., Cai, X., Chen, Q., Wang, H., Li, A., Rui, J., Zhang, X., & Xie, F. (2021). Capturing the trajectory of psychological status and analyzing online public reactions during the coronavirus disease 2019 pandemic through Weibo posts in China. *Frontiers in psychology, 12*.

Cotfas, L.-A., Delcea, C., & Gherai, R. (2021). COVID-19 vaccine hesitancy in the month following the start of the vaccination process. *International Journal of Environmental Research and Public Health, 18*, 10438.

de Sousa, W. G., de Melo, E. R. P., Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly, 36*, 101392.

Deng, L., & Liu, Y. (2018). A joint introduction to natural language processing and to deep learning. In L. Deng & Y. Liu (Eds.), *Deep Learning in Natural Language Processing* (pp. 1-22). Singapore: Springer.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. https://arxiv.org/abs/1810.04805

Duan, H. K., Vasarhelyi, M. A., Codesso, M., & Alzamil Z. (2023). Enhancing the government accounting information systems using social media information: An application of text mining and machine learning. *International Journal of Accounting Information Systems*, 48, 100600.

Essam, N., Moussa, A. M., Elsayed, K. M., Abdou, S., Rashwan, M., Khatoon, S., Hasan, M. M., Asif, A., & Alshamari, M. A. (2021). Location Analysis for Arabic COVID-19 Twitter Data Using Enhanced Dialect Identification Models. *Applied Sciences, 11*, 11328.

Evans, J. A., & Aceves, P. (2016). Machine translation: Mining text for social theory. *Annual Review of Sociology, 42*, 21-50.

Fersini, E., Messina, E., & Pozzi, F. A. (2017). Earthquake management: a decision support system based on natural language processing. *Journal of Ambient Intelligence and Humanized Computing, 8*, 37-45.

Ghahramani, M., Galle, N. J., Ratti, C., & Pilla, F. (2021). Tales of a city: Sentiment analysis of urban green space in Dublin. *Cities, 119*, 1-7.

Hammer, M., & Mangurian, G. E. (1987). SMR FORUM The Changing Value of Communications Technology. *Sloan Management Review (1986-1998), 28*, 65.

Hasan, A., Levene, M., Weston, D., Fromson, R., Koslover, N., & Levene, T. (2022). Monitoring COVID-19 on social media: development of an end-to-end natural language processing pipeline using a novel triage and diagnosis approach. *Journal of Medical Internet Research, 24*, e30397.

He, L., He, C., Reynolds, T. L., Bai, Q., Huang, Y., Li, C., Zheng, K., & Chen, Y. (2021). Why do people oppose mask wearing? A comprehensive analysis of US tweets during the COVID-19 pandemic. *Journal of the American Medical Informatics Association, 28*, 1564-1573.

Hu, G., Han, X., Zhou, H., & Liu, Y. (2019). Public perception on healthcare services: evidence from social media platforms in China. *International journal of environmental research and public health, 16*, 1273.

Hu, Y., Deng, C., & Zhou, Z. (2019). A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments. *Annals of the American Association of Geographers, 109*, 1052-1073.

Hussain, A., Tahir, A., Hussain, Z., Sheikh, Z., Gogate, M., Dashtipour, K., Ali, A., & Sheikh, A. (2021). Artificial intelligence–enabled analysis of public attitudes on facebook and twitter toward covid-19 vaccines in the united kingdom and the united states: Observational study. *Journal ofMmedical Internet Research, 23*, e26627.

Jelodar, H., Wang, Y., Orji, R., & Huang, S. (2020). Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE Journal of Biomedical and Health Informatics, 24*, 2733-2742.

Kabbani, O., Klumpenhouwer, W., El-Diraby, T., & Shalaby, A. (2022). What do riders say and where? The detection and analysis of eyewitness transit tweets. *Journal of Intelligent Transportation Systems*, 1-17.

Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics, 7*, 139-172.

Kar, A. K., & Dwivedi, Y. K. (2020). Theory building with big data-driven research–Moving away from the "What" towards the "Why". *International Journal of Information Management, 54*, 102205.

Khatoon, S., Alshamari, M. A., Asif, A., Hasan, M. M., Abdou, S., Elsayed, K. M., & Rashwan, M. (2021). Development of social media analytics system for emergency event detection and crisismanagement. *Computers, Materials and Continua, 68*, 3079-3100.

Kim, A. Y., Ha, J. G., Choi, H., & Moon, H. (2018). Automated Text Analysis Based on Skip-Gram Model for Food Evaluation in Predicting Consumer Acceptance. *Computational Intelligence and Neuroscience, 2018*, 9293437.

Kim, N., & Hong, S. (2021). Automatic classification of citizen requests for transportation using deep learning: Case study from Boston city. *Information Processing & Management, 58*, 102410.

Kim, S. Y., Ganesan, K., Dickens, P., & Panda, S. (2021). Public sentiment toward solar energy—opinion mining of twitter using a transformer-based language model. *Sustainability, 13*, 1-19.

Kumar, A., & Sharma, A. (2017). Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence. *Webology, 14*.

Kumar, A., Singh, S., & Kaur, G. (2019). Fake news detection of Indian and United States election data using machine learning algorithm. *International Journal of Innovative Technology and Exploring Engineering, 8*, 1559-1563.

Kumar, R., Mukherjee, S., Choi, T.-M., & Dhamotharan, L. (2022). Mining voices from self-expressed messages on social-media: Diagnostics of mental distress during COVID-19. *Decision Support Systems*, 1-13.

Li, L., Bensi, M., Cui, Q., Baecher, G. B., & Huang, Y. (2021). Social media crowdsourcing for rapid damage assessment following a sudden-onset natural hazard event. *International Journal of Information Management, 60*, 1-18.

Li, Y., Thomas, M. a., & Liu, D. (2020). From semantics to pragmatics: where IS can lead in Natural Language Processing (NLP) research. *European Journal of Information Systems*, 1-22.

Liang, W., Chen, X., Huang, S., Xiong, G., Yan, K., & Zhou, X. (2023). Federal learning edge network based sentiment analysis combating global COVID-19. *Computer Communications*, *204*, 33-42.

Lichtenthaler, E. (2004). Technological change and the technology intelligence process: a case study. *Journal of Engineering and Technology Management, 21*, 331-348.

Liew, T. M., & Lee, C. S. (2021). Examining the utility of social media in COVID-19 vaccination: unsupervised learning of 672,133 twitter posts. *JMIR public health and surveillance, 7*, e29789.

Lindquist, J., Thomas, D. M., Turner, D., Blankenship, J., & Kyle, T. K. (2021). Food for thought: A natural language processing analysis of the 2020 Dietary Guidelines publice comments. *The American Journal of Clinical Nutrition, 114*, 713-720.

Lock, O., & Pettit, C. (2020). Social media as passive geo-participation in transportation planning–how effective are topic modeling & sentiment analysis

in comparison with citizen surveys? *Geo-spatial Information Science, 23*, 275-292.

Luo, L., Wang, Y., & Liu, H. (2022). COVID-19 personal health mention detection from tweets using dual convolutional neural network. *Expert Systems with Applications, 200*, 1-9.

Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law, 28*, 237-266.

Miao, L., Last, M., & Litvak, M. (2022). Tracking social media during the COVID-19 pandemic: The case study of lockdown in New York State. *Expert Systems with Applications, 187*.

Misuraca, G., van Noordt, C., & Boukli, A. (2020). The use of AI in public services: results from a preliminary mapping across the EU. In *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance* (pp. 90-99).

Ngai, E. W. T., Choi, Y. S., & Chai, Y. F. (2017). Method and System for Linguistics Query-based Text Analysis (LQTA). In (Vol. Chinese Patent for Invention No. CN104346382B).

Ngai, E. W. T., Lee, M. C., Choi, Y., & Chai, P. (2018). Multiple-Domain Sentiment Classification for Cantonese Using a Combined Approach. In *PACIS* (pp. 297).

Ngai, E. W. T., & Wat, F. (2002). A literature review and classification of electronic commerce research. *Information & Management, 39*, 415-429.

Ngai, E. W. T, Moon, K.K. L., Lam, SS, Chin, E.S. K., Tao, S. S., C. (2015), Social Media Models, Technologies, Applications: An Academic Review and Case Study, *Industrial Management & Data Systems,* 115, 5, 759 - 802.

Nord, J. H., & Nord, G. D. (1995). MIS research: Journal status assessment and analysis. *Information & Management, 29*, 29-42.

Nunamaker Jr, J. F., Chen, M., & Purdin, T. D. (1990). Systems development in information systems research. *Journal of Management Information Systems, 7*, 89-106.

Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020). Big Data and AI–A transformational shift for government: So, what next for research? *Public Policy and Administration, 35*, 24-44.

Pérez-Pérez, M., Pérez-Rodríguez, G., Fdez-Riverola, F., & Lourenço, A. (2019). Using twitter to understand the human bowel disease community: Exploratory analysis of key topics. *Journal of Medical Internet Research, 21*, e12610.

Pu, X., Jiang, Q., & Fan, B. (2022). Chinese public opinion on Japan's nuclear wastewater discharge: A case study of Weibo comments based on a thematic model. *Ocean & Coastal Management, 225*, 106188.

Ragini, J. R., Anand, P. R., & Bhaskar, V. (2018). Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management, 42*, 13-24.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Reynard, D., & Shirgaokar, M. (2019). Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster? *Transportation research part D: Transport and environment, 77*, 449-463.

Riggins, F. J. (1999). A framework for identifying web-based electronic commerce opportunities. *Journal of Organizational Computing and Electronic Commerce, 9*, 297-310.

Roy, S., & Ghosh, P. (2021). A comparative study on distancing, mask and vaccine adoption rates from global twitter trends. *Healthcare, 9*, 488.

Saleh, S. N., Lehmann, C. U., McDonald, S. A., Basit, M. A., & Medford, R. J. (2021). Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing on Twitter. *Infection Control & Hospital Epidemiology, 42*, 131-138.

Saleh, S. N., Lehmann, C. U., & Medford, R. J. (2021). Early crowdfunding response to the COVID-19 pandemic: cross-sectional study. *Journal of Medical Internet Research, 23*.

Scott Morton, M. S. (1984). State of the art of research in management support systems. In F. W. McFarlan (Ed.), *The Information Systems Research Challenge. Harvard Business School Press* (pp. 13-41). Boston, MA.

Shah, A. M., Yan, X., Tariq, S., & Ali, M. (2021). What patients like or dislike in physicians: Analyzing drivers of patient satisfaction and dissatisfaction using a digital topic modeling approach. *Information Processing & Management, 58*, 102516.

Shahrezaye, M., Suter, V., & Meckel, M. COVID-19 induced misinformation on YouTube: An analysis of user commentary. *Frontiers in Political Science*, 26.

Son, J., Lee, H. K., Jin, S., & Lee, J. (2019). Content features of tweets for effective communication during disasters: A media synchronicity theory perspective. *International Journal of Information Management, 45*, 56-68.

Stutzer, R., Rinscheid, A., Oliveira, T. D., Loureiro, P. M., Kachi, A., & Duygan, M. (2021). Black coal, thin ice: the discursive legitimisation of Australian coal in the age of climate change. *Humanities and Social Sciences Communications, 8*, 1-9.

Sun, Y., Wang, Z., Zhang, B., Zhao, W., Xu, F., Liu, J., & Wang, B. (2020). Residents' sentiments towards electricity price policy: Evidence from text mining in social media. *Resources, Conservation and Recycling, 160*, 1-9.

Tan, H., Peng, S.-L., Zhu, C.-P., You, Z., Miao, M.-C., & Kuai, S.-G. (2021). Long-term effects of the COVID-19 pandemic on public sentiments in mainland China: sentiment analysis of social media posts. *Journal of Medical Internet Research, 23*, e29150.

Tang, L., Liu, W., Thomas, B., Tran, H. T. N., Zou, W., Zhang, X., & Zhi, D. (2021). Texas public agencies' tweets and public engagement during the COVID-19 pandemic: Natural language processing approach. *JMIR public health and surveillance, 7*, e26720.

Tao, D., Yang, P., & Feng, H. (2020). Utilization of text mining as a big data analysis tool for food science and nutrition. *Comprehensive Reviews in Food Science and Food Safety, 19*, 875-894.

Tavory, I., & Timmermans, S. (2014). *Abductive analysis: Theorizing qualitative research*: University of Chicago Press.

Teh, P. L., Piao, S., Almansour, M., Ong, H. F., & Ahad, A. (2022). Analysis of Popular Social Media Topics Regarding Plastic Pollution. *Sustainability, 14*, 1-23.

Umamaheswaran, S., Dar, V., Sharma, E., & Kurian, J. S. (2023). Mapping Climate Themes From 2008-2021—An Analysis of Business News Using Topic Models. *IEEE Access*, 11, 26554-26565.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (Vol. 30). Long Beach, CA, USA. https://doi.org/10.5555/3295222.3295349

Wahid, J. A., Shi, L., Gao, Y., Yang, B., Wei, L., Tao, Y., Hussain, S., Ayoub, M., & Yagoub, I. (2022). Topic2Labels: A framework to annotate and classify the social media data through LDA topics and deep learning models for crisis response. *Expert Systems with Applications, 195*, 1-15.

Wang, Y., & Taylor, J. E. (2019). DUET: Data-driven approach based on latent Dirichlet allocation topic modeling. *Journal of Computing in Civil Engineering, 33*, 1-8.

Wu, P., Li, X., Shen, S., & He, D. (2020). Social media opinion summarization using emotion cognition and convolutional neural networks. *International Journal of Information Management, 51*, 1-15.

Xia, H., Wang, Y., Zhang, J. Z., Zheng, L. J., Kamal, M. M., & Arya, V. (2023). COVID-19 fake news detection: A hybrid CNN-BiLSTM-AM model. *Technological Forecasting and Social Change*, *195*, 122746.

Xiong, J., Hswen, Y., & Naslund, J. A. (2020). Digital surveillance for monitoring environmental health threats: A case study capturing public opinion from Twitter about the 2019 Chennai water crisis. *International Journal of Environmental Research and Public Health, 17*, 5077.

Yang, Y.-C., Al-Garadi, M. A., Bremer, W., Zhu, J. M., Grande, D., & Sarker, A. (2021). Developing an Automatic System for Classifying Chatter About Health Services on Twitter: Case Study for Medicaid. *Journal of Medical Internet Research, 23*, e26616.

Yin, H., Song, X., Yang, S., & Li, J. (2022). Sentiment analysis and topic modeling for COVID-19 vaccine discussions. *World Wide Web, 25*, 1067-1083.

Zaeem, R. N., Manoharan, M., Yang, Y., & Barber, K. S. (2017). Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Computers & Security, 65*, 50-63.

Zavattaro, S. M., French, P. E., & Mohanty, S. D. (2015). A sentiment analysis of US local government tweets: The connection between tone and citizen involvement. *Government Information Quarterly, 32*, 333-341.

Zhai, C., & Massung, S. (2016). *Text data management and analysis: a practical introduction to information retrieval and text mining*: Morgan & Claypool.

Zhang, C., Xu, S., Li, Z., & Hu, S. (2021). Understanding concerns, sentiments, and disparities among population groups during the COVID-19 pandemic via Twitter data mining: large-scale cross-sectional study. *Journal of Medical Internet Research, 23*, e26482.

Zhang, D., Qiang, M., Jiang, H., Wen, Q., An, N., & Xia, B. (2018). Social sensing system for water conservation project: a case study of the South-to-North Water Transfer Project in China. *Water Policy, 20*, 667-691.

Zhang, Y., Chen, K., Weng, Y., Chen, Z., Zhang, J., & Hubbard, R. (2022). An intelligent early warning system of analyzing Twitter data using machine learning on COVID-19 surveillance in the US. *Expert Systems with Applications, 198*, 116882.

Zheng, C., Xue, J., Sun, Y., & Zhu, T. (2021). Public Opinions and Concerns Regarding the Canadian Prime Minister's Daily COVID-19 Briefing: Longitudinal Study of YouTube Comments Using Machine Learning Techniques. *Journal of Medical Internet Research, 23*, e23957.

Zhou, S., Kan, P., Huang, Q., & Silbernagel, J. (2021). A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura. *Journal of Information Science*, 01655515211007724.