LUCE: A Dynamic Framework and Interactive Dashboard for Opinionated Text Analysis

Omnia Zayed, Gaurav Negi, Sampritha Manjunath, Devishree Pillai, and Paul Buitelaar
Insight SFI Research Centre for Data Analytics
Data Science Institute
University of Galway
IDA Business Park, Lower Dangan, Galway, Ireland
{firstname.lastname}@universityofgalway.ie

Abstract

We introduce *LUCE*, an advanced dynamic framework with an interactive dashboard for analysing opinionated text aiming to understand people-centred communication. The framework features computational modules of text classification and extraction explicitly designed for analysing different elements of opinions, e.g., sentiment/emotion, suggestion, figurative language, hate/toxic speech, and topics. We designed the framework using a modular architecture, allowing scalability and extensibility with the aim of supporting other NLP tasks in subsequent versions. *LUCE* comprises trained models, python-based APIs, and a user-friendly dashboard, ensuring an intuitive user experience. *LUCE* has been validated in a relevant environment, and its capabilities and performance have been demonstrated through initial prototypes and pilot studies.

1 Introduction

In an era where user-generated content on social media, forums, surveys, and review sites plays a pivotal role in shaping public opinion and influencing decision-making, understanding these opinionated discourse becomes essential. This sheer amount of opinionated content opened a wide range of possibilities for businesses, researchers, policymakers, and other stakeholders. At the same time, it creates a pressing need for automated language analysis to understand public discourse and visualise patterns through statistical analysis and results aggregation. However, analysing this data accurately and efficiently remains a challenge due to its sheer volume, diversity, and rapid pace of evolution. The preliminary research we conducted identified several critical gaps in existing opinion analysis applications: 1) tailored for commercial brand monitoring, making them difficult to adapt to other sectors (Nanda and Kumar, 2021); 2) limited to the analysis of particular tasks (e.g., sentiment analysis) instead of offering a holistic approach to opinion analysis; 3) high language-dependence (mostly English); 4) high domain-dependence (Purnat et al., 2021; White et al., 2023), requiring significant manual development for their adaptation to new domains.

We introduce *LUCE*, Italian for *light* and short for ***L**isten, **U**nderstand, **C**onnect, **E**ngage*, an advanced dynamic AI-powered framework for analysing opinionated text that aims to study people-centred public communication. The framework features an interactive dashboard to ensure a comfortable and intuitive user experience. This framework addresses the aforementioned gaps by leveraging state-of-the-art natural language processing (NLP) techniques to analyse public communication on multiple interconnected levels, including sentiment, emotion, suggestions, hate speech, sarcasm, figurative language, and topics. *LUCE* is designed to be domain, sector, and language-independent, which means that it is not limited to social media text or specific language. *LUCE*'s modular design ensures scalability and adaptability to various use cases and applications.

The initial version of *LUCE* focuses on text classification and extraction through three main modules to identify 1) opinion dimensions, 2) topics (aspect terms and categories), and 3) shorter text spans of suggestions. We employ dynamic transfer learning-based computational models for domain adaptation (Negi et al., 2024). Additionally, the introduced modules are designed to support domain and language independence through utilising language-agnostic embeddings (Feng et al., 2022) and cross-lingual transfer learning (Singla et al., 2018). One of the core objectives of *LUCE* is to allow end-users to analyse opinionated text automatically based on their needs through a user-friendly dashboard supporting data integration, results aggregation, and output visualisation. The visual analytics components offered by the dashboard allow

Proceedings of the 31st International Conference on Computational Linguistics: System Demonstrations, pages 104–116
January 19–24, 2025.

end-users to understand the discourse and visualise patterns/relationships.

In this paper, we present the first prototype of *LUCE*[1], which has been validated in a relevant environment, and its capabilities and performance have been demonstrated through various use cases and pilot studies. Examples include the EU-funded PANDEM-2 project[2], which included a social media analysis (SMA) component that was a forerunner of *LUCE* to support two-way communication during pandemics where reactions from the general public to government measures during the pandemic were successfully analysed. Similarly, the SFI-National Challenge-funded project Platform Urbanism[3], in collaboration with Galway City Council, employed the initial forerunner of LUCE to analyse public communication around urban development. Currently, the *LUCE* prototype is being used in the University of Galway Research Process Improvement project to analyse staff survey responses to university support of excellent and impactful research. The main contributions of this paper are summarised as follows:

1. We introduce the *LUCE* dynamic framework for opinionated text analysis developed with state-of-the-art performance.
2. We validated the proposed technology in relevant environments through pilot studies on various domains.
3. We designed the framework using a modular architecture, allowing scalability and extensibility to support other NLP tasks in subsequent versions.
4. We introduce a user-friendly web-based dashboard encompassing the framework's pretrained models and Python-based APIs, ensuring an intuitive user experience.

2 *LUCE* Framework and Interactive Dashboard

2.1 Architecture

LUCE follows a modular architecture, as shown in Figure 1, to allow scalability and extensibility. The framework comprises three main modules, which currently enable 1) the classification of opinion dimensions in a given text (e.g., survey responses, social media posts, etc.); 2) the identification of core aspect terms and categories related to people's perceptions; 3) the extraction of shorter spans (e.g., of suggestions or hate speech) from a given text.

The development process of *LUCE* is based on a structured pipeline divided into four stages, as depicted in Figure 2. The data preparation step focuses on collecting, preparing, and preprocessing benchmark datasets for each task the framework supports. These datasets are used for training and evaluating the developed models. The development stage of each module is designed to allow the ease of modification and extensibility by utilising state-of-the-art approaches for text classification (e.g., sentiment classification, emotion classification, suggestion mining, etc.), aspect terms and categories extraction (e.g., term extraction, keyphrase extraction, and clustering), and span extraction. Quantitative and qualitative evaluation schemes were followed to assess the performance of each developed model. The quantitative analysis employed the traditional metrics of measuring the performance of each task. For example, the performance of the classification models is evaluated in terms of precision, recall, F1-score and accuracy (Rijsbergen, 1979) and the performance of the extraction models is evaluated in terms of BLEU (Papineni et al., 2002) and ROUGE (Lin, 2004) scores.

2.2 System Components

LUCE has a Python-based back-end and an interactive front-end dashboard. The architecture is designed to provide seamless interaction between data processing, opinion analysis, and output visualisation, with each component integrated through well-defined workflows.

Python-based Back-end. The back end is responsible for processing the input data, running opinion analysis models, and managing interactions with the front end. It utilises several components to ensure efficient and scalable operations. Communication between the back-end and front-end is facilitated through REST API endpoints. These endpoints handle requests, such as data submission, analysis initiation, and result retrieval. The opinion analysis modules encompass trained deep models served using TorchServe[4]. The entire back-end is encapsulated within a self-contained Docker image, ensuring consistency and ease of deployment

[1] This first prototype corresponds to *LUCE* Beta v0.1.

[2] https://pandem-2.eu

[3] https://www.sfi.ie/challenges/future-digital/cathair-shamhlu/

[4] https://pytorch.org/serve/

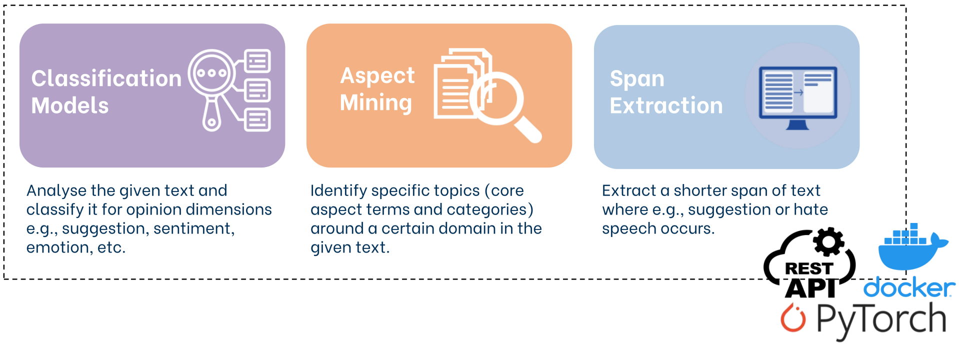


Figure 1: Main Modules in the current prototype of *LUCE*.

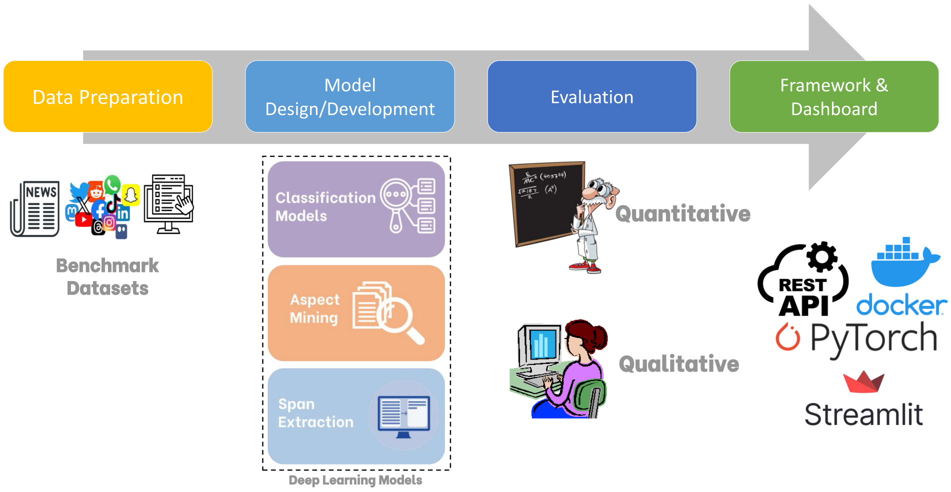


Figure 2: *LUCE*'s Structured Development Process.

across different environments. Docker enables the system to run with all necessary dependencies, facilitating portability and simplifying the installation process. The back-end is hosted locally on a Linux-based server.

Interactive Front-end Dashboard. The front end is implemented using Streamlit[5], an open-source Python library that simplifies the creation of interactive web-based dashboards. *LUCE*'s dashboard allows users to interact with the framework, view the results of the opinion analysis, and explore the data visually. Users can upload text data through the dashboard (in two formats, as discussed in Section 2.3), initiate the analysis process, and view outputs such as sentiment scores, emotion classes, top terms, and detailed visualisations. The front end is designed to be simple and easy to use, ensuring that both technical and non-technical users can easily navigate the interface.

Workflow. The following steps summarise the workflow between the back and front end.

- Data Input: The user uploads opinionated text data through the dashboard, which triggers a request to the back-end via a REST API.
- Model Processing: The back-end receives the data (as JSON objects), processes it, and passes it to the TorchServe model for inference.
- Opinion Analysis: The trained models analyse the text data, identifying opinion dimensions and aspect terms/categories. The results are then packaged (as JSON objects) and sent to the front end.
- Data Visualisation: The results are then displayed on the Streamlit-based dashboard, where users can interact with the various visual components and extract meaningful insights.
- Real-Time Feedback: The system ensures real-time feedback by updating the dashboard with results as soon as they are available, pro-

[5]https://streamlit.io

viding an intuitive and responsive user experience.

The following section demonstrates this workflow in detail, and Section 3 navigates through real-world use cases where *LUCE* has been deployed within various projects.

2.3 System Demonstration

LUCE's dashboard allows an interactive user experience through multiple visualisation components to facilitate the analysis of opinionated text. This section provides a guided tour of the interactive dashboard, highlighting *LUCE*'s main features.

Figure 7 shows a snapshot of the Home page showing the main modules on the left-hand menu. This page serves as the entry point to the framework. It briefly introduces users to *LUCE* and explains how to upload and analyse their data.

Each module page has its upload page, which allows users to submit their text data for analysis. The dashboard accepts two input formats, raw text and text file (comma-separated dataset), making it adaptable for different needs. Figure 3 shows both options side by side. Once the data is uploaded, the user can choose various options to proceed with the analysis through drop-down menus, such as the type of opinion classification, e.g., sentiment, emotion, suggestion, etc., the language[6], the trained model (e.g., Attention-based (Baziotis et al., 2017; Chronopoulou et al., 2018), Transformer-based (Devlin et al., 2019; Liu et al., 2019), LLM-based (Negi et al., 2024) models), the aspect term extraction techniques (e.g., term extraction (Frantzi et al., 2000; Zhang et al., 2016), keyphrase extraction (Boudin, 2018; Campos et al., 2020)), and aspect clustering mechanisms (e.g., centroid-based (MacQueen, 1967) or hierarchical density-based (McInnes et al., 2017)). Furthermore, when uploading a comma-separated file, the user can specify the file format in terms of text column, delimiter type, quotation type, etc.

Once the analysis is complete, the results will automatically appear through a number of interactive visual components in a dedicated section on each module's page. Generally, the result section comprises the following:

- Opinion Distribution: The distribution of each opinion dimension is represented by pie charts that break down the opinion categories within the dataset. Below these pie charts, users can find a bar chart highlighting the top 10 bigrams and a word cloud highlighting the most frequently used words. Figure 4 shows an example of the results section of opinion classification on a University Survey data.
- Text-level Analysis: A data frame view showing individual text entries with their corresponding identified opinion dimensions and probabilities (e.g., sentiment polarity, emotions categories, suggestion class, etc.) as shown in Figure 8.
- Interactive Filters: Users can filter the results based on specific criteria, such as opinion class or keyword.

2.4 Supported Functionality

The current prototype provides language analysis in terms of 1) opinion dimensions and 2) topics (aspect terms and categories)[7]. For opinion dimensions, the current version of *LUCE* supports sentiment analysis, emotion analysis, and suggestion classification. Sentiment analysis is concerned with identifying whether a sentence holds a positive, negative, or neutral sentiment (Liu and Zhang, 2012; Rosenthal et al., 2017). Emotion analysis further identifies emotional expressions conveyed in the text by utilising various psychological classification schemes. We employ an extension of the Plutchik (1980) model to identify 11 expressions (Mohammad et al., 2018): *anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise,* and *trust*. The task of suggestion detection focuses on classifying a given post as to whether it contains a suggestion or not (Negi et al., 2019).

The uploaded data can revolve around a specific theme or topic. However, it is essential to analyse it on a fine-grained level to understand the subtopics (aspects) discussed. This analysis is referred to as aspect mining, where aspect terms and categories are extracted from opinionated text through a multi-stage process that involves the extraction of aspect terms (features) of an entity or object in a particular domain and figuring out opinions about those aspects. The process can also involve assigning

[6] The dashboard is currently hosting the English-based models, we will include the implemented multilingual models in subsequent versions.

[7] The framework already supports suggestion span extraction, but it was not included in the dashboard when this manuscript was submitted.

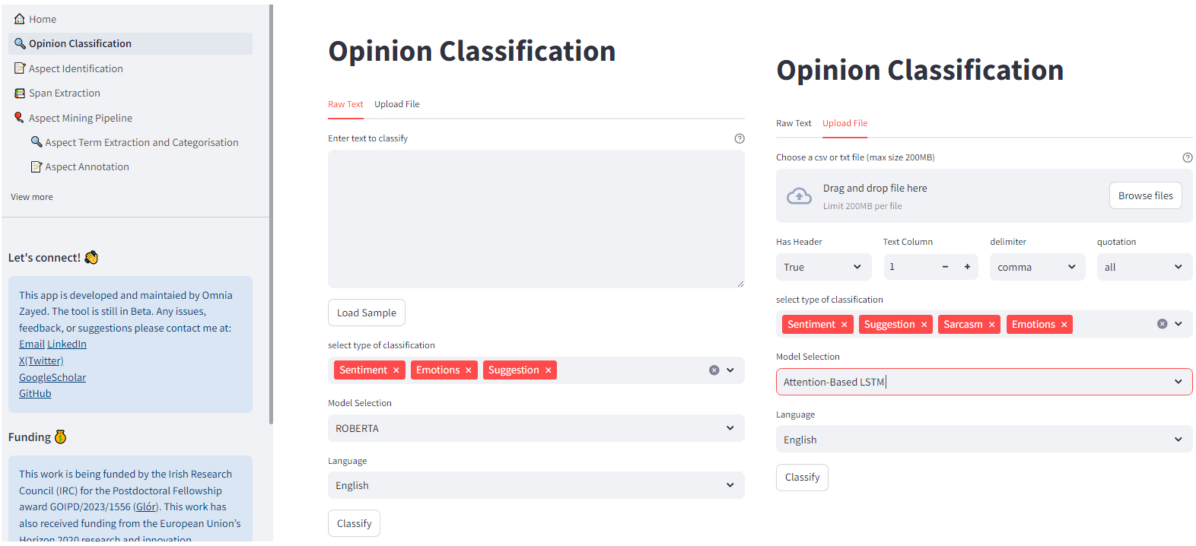


Figure 3: A snapshot of the input formats, raw text and text file, which *LUCE*'s dashboard offers.

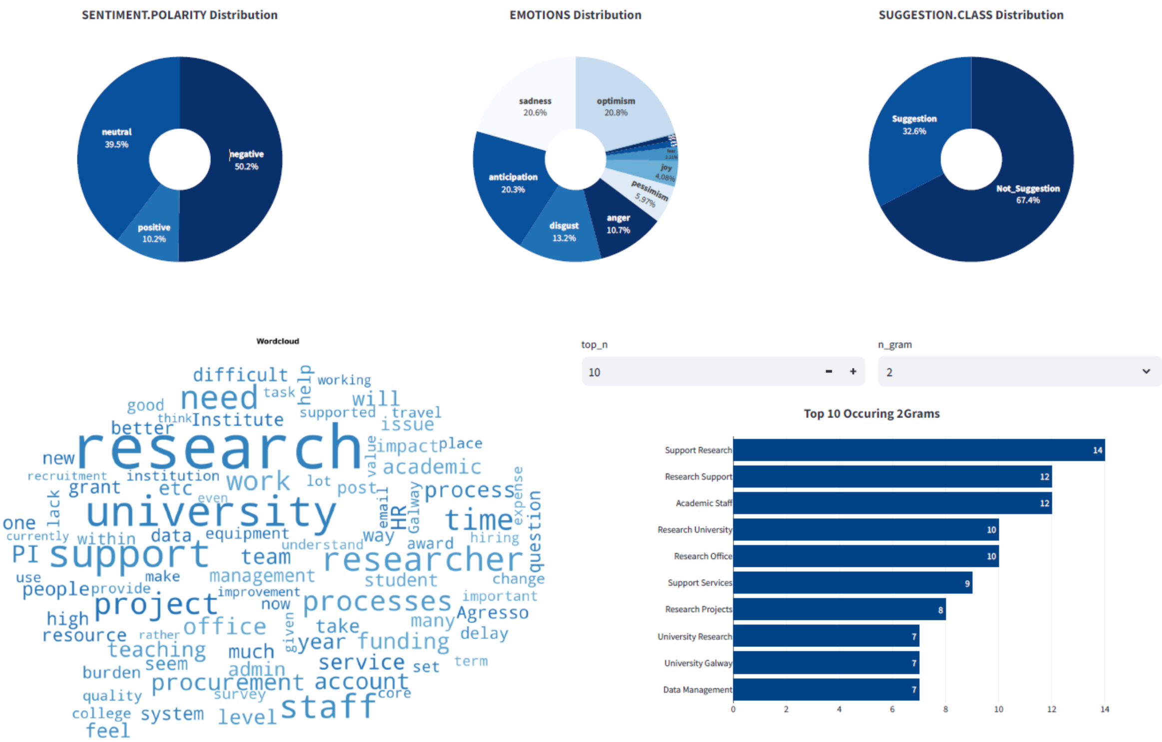


Figure 4: A snapshot of the pie charts in the opinion dimension analysis result.

the identified aspect terms to predefined higher-level categories (Pang and Lee, 2008; Liu, 2020). Aspect term identification is hosted on a separate page on the dashboard where the user can upload text dataset as shown in Figure 5. Once the analysis is done, aspect terms and their corresponding categories will be visualised in a tabular format. Additionally, the end-user can inspect a visualisation of the clustered aspects, as shown in Figure 6.

Furthermore, the aspect categories identified for a given dataset can be reused to annotate new unseen text from the same domain (see Figure 9 in Appendix A).

3 Use Cases

An initial prototype of *LUCE* has been deployed within multiple projects to address a wide range of use cases across various domains and types of text.

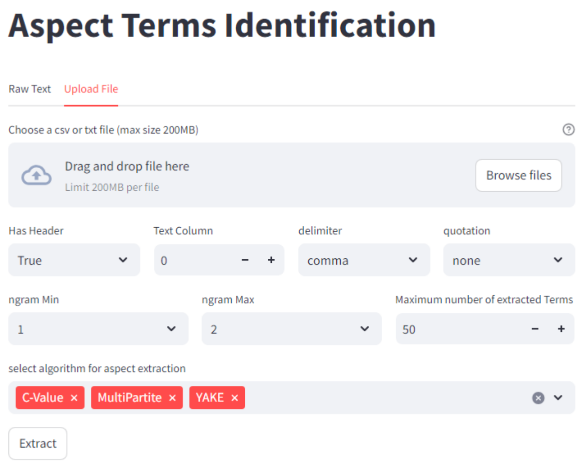


Figure 5: A snapshot of the data loading option on the Aspect Identification page. The user can upload raw text or a text file.

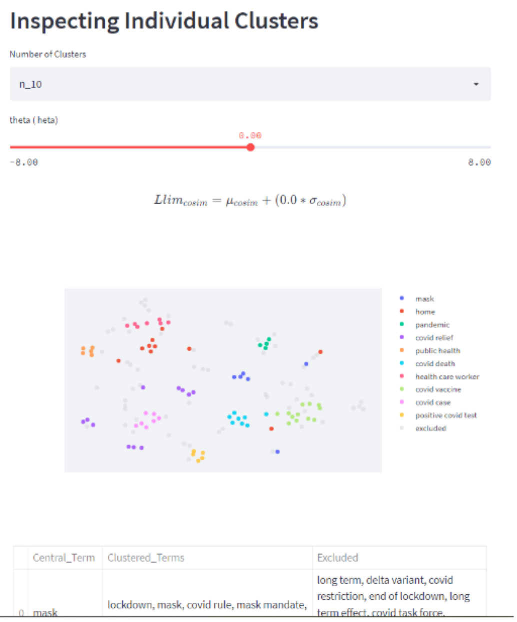


Figure 6: A screenshot of the aspect clustering result of the COVID dataset.

Public Health. The forerunner prototype of *LUCE* has been deployed in the EU-funded PANDEM-2 project. The project focused on pandemic preparedness and response and included an initial prototype of the language analysis framework to support two-way communication[8] on social media during pandemics. Public health agencies within the project's consortium have extensively used the proposed solution and validated the need for insight extraction from social media with the framework's demonstrated capabilities. In this project, tweets (rebranded as X posts) related to COVID-19 were analysed in real-time. The live streaming of tweets was done using the ECDC-developed tool Epitweetr (Espinosa et al., 2022) before discontinuing the free API access. Figure 10[9] shows a snapshot of the social media analysis page on the PANDEM-2 dashboard showing trending topics at a specific period during the COVID-19 pandemic, along with sentiment/emotion analysis. The interactive dashboard permitted the comparison of opinion dimensions across countries based on the end-user choice (Figure 11). Moreover, the conducted suggestion analysis gave public health managers the ability to sift through the communicated suggestions on social media by the general public based on dynamically identified categories (as shown in Figure 12). The implemented filtering mechanisms in the user interface facilitated analysing public communication in a particular country, in a particular language, or time period according to the needs of end-users. In addition to analysing real-time tweets, the system was used to analyse a stratified hydrated random sample of around 500K tweets of a publicly available large-scale COVID dataset of tweets (Lamsal, 2021).

Urban Development. Following the successful implementation of the advanced opinion analysis technology in the PANDEM-2 project, the SFI-National Challenge-funded Platform Urbanism project, in collaboration with Galway City Council, was encouraged to use it to analyse public communication. The project specifically concentrated on analysing Reddit posts around Galway City[10] to identify various opinion dimensions and aspects of concern to the Irish citizens regarding the city's urban development.

University Research Process Improvement. Currently, the *LUCE* prototype is being used in the University of Galway Research Process Improvement project to analyse staff survey responses to university support of excellent and impactful research. The language analysis is conducted on the textual responses to open-ended questions from the survey. The management and consultants use the outcome to understand the barriers within the current processes and operating models to enable

[8]Two-way communication ensures the mutual flow of information between multiple parties, e.g., the public and public health agencies.

[9]Some figures are moved to Appendix A due to space limitations.

[10]https://www.reddit.com/r/galway/

research and innovation.

4 Conclusion and Future Work

We introduced *LUCE*, an advanced dynamic framework for opinionated text analysis to study public communication. The current beta prototype of *LUCE* features computational state-of-the-art neural-based modules of text classification and extraction explicitly designed for analysing different elements of opinions, e.g., sentiment/emotion, suggestion, and topics. The framework is designed with a modular architecture to ensure scalability and extensibility for future versions, supporting a wide range of NLP tasks. *LUCE* comprises pre-trained models, python-based APIs, and a user-friendly dashboard, ensuring an intuitive user experience. We have validated the technology in relevant environments and demonstrated its capabilities and performance across diverse use cases in different domains and applications. Currently, we are expanding the opinion dimensions module to encompass additional identification tasks such as sarcasm and hate speech and enhancing the dashboard design to include an extraction module developed for analysing suggestion text.

5 Broader Impact

The *LUCE* framework and the proposed technology will have substantial societal, political, and academic impacts due to the universal need of organisations, such as public service agencies, government entities, and corporations, to understand opinionated data produced in massive volumes daily. Such understanding permits 1) the adjustment of communication approaches, 2) the support of decision-making and policies, and 3) the building of public trust, ultimately improving public/private service delivery. The deployment of this prototype in various projects, such as the EU-funded PANDEM-2, proved its timeliness and necessity. We have engaged with potential stakeholders in the public health sectors and beyond, gathering valuable feedback to refine the proposed technology. Prospective national and international stakeholders showed interest in the prototype and are seeking to beta-test future versions of *LUCE*. Other prospective stakeholders include national public agencies, research institutions, and companies. In all cases, the stakeholders have emphasised the need for insight extraction from user-generated content with the capabilities demonstrated using the *LUCE* prototype.

6 Ethical Considerations

Privacy/Copyright. One of the ethical considerations we are tackling is the privacy and copyright of the uploaded datasets by end-users. Since the framework relies on collecting and analysing text data, it is crucial to ensure that 1) the framework is secure enough to handle and protect private data and 2) the uploaded data is collected without infringing copyright or redistribution policies.

Bias. The computational models implemented in the *LUCE* framework are trained on benchmark datasets that may reflect inherent bias, which, in turn, can affect the fairness and accuracy of the analysis. To mitigate this, we conduct regular qualitative analysis[11] on the data to understand the framework's performance to improve fairness and reduce bias.

Acknowledgments

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland for the Postdoctoral Fellowship award GOIPD/2023/1556 (Glór). This work has also received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883285 (PANDEM-2). Additionally, this work has been partially funded by the Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2 (Insight_2).

References

Christos Baziotis, Nikos Pelekis, and Christos Doulkeridis. 2017. DataStories at SemEval-2017 task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 747–754, Vancouver, Canada. Association for Computational Linguistics.

Florian Boudin. 2018. Unsupervised keyphrase extraction with multipartite graphs. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 667–672, New Orleans, Louisiana. Association for Computational Linguistics.

[11]The results of this qualitative analysis are out of the scope of this demo paper.

Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Jorge, Célia Nunes, and Adam Jatowt. 2020. Yake! keyword extraction from single documents using multiple local features. *Information Sciences*, 509:257–289.

Alexandra Chronopoulou, Aikaterini Margatina, Christos Baziotis, and Alexandros Potamianos. 2018. NTUA-SLP at IEST 2018: Ensemble of neural transfer methods for implicit emotion classification. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 57–64, Brussels, Belgium. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Laura Espinosa, Ariana Wijermans, Francisco Orchard, Michael Höhle, Thomas Czernichow, Pietro Coletti, Lisa Hermans, Christel Faes, Esther Kissling, and Thomas Mollet. 2022. Epitweetr: Early warning of public health threats using Twitter data. *Eurosurveillance*, 27(39).

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics.

Katerina Frantzi, Sophia Ananiadou, and Hideki Mima. 2000. Automatic recognition of multi-word terms: The c-value/nc-value method. *International Journal on Digital Libraries*, 3(2):115–130.

Rabindra Lamsal. 2021. Design and analysis of a large-scale covid-19 tweets dataset. *applied intelligence*, 51:2790–2804.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Bing Liu. 2020. *Aspect and Entity Extraction*, page 168–226. Studies in Natural Language Processing. Cambridge University Press.

Bing Liu and Lei Zhang. 2012. *A Survey of Opinion Mining and Sentiment Analysis*, pages 415–463. Springer US, Boston, MA.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *ArXiv*, abs/1907.11692.

J. B. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press.

Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205.

Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. SemEval-2018 Task 1: Affect in tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, SemEval '18, pages 1–17, New Orleans, LA, USA.

Pooja Nanda and Vikas Kumar. 2021. Social media analytics: tools, techniques and present day practices. *International Journal of Services Operations and Informatics*, 11(4):422.

Gaurav Negi, Rajdeep Sarkar, Omnia Zayed, and Paul Buitelaar. 2024. A hybrid approach to aspect based sentiment analysis using transfer learning. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 647–658, Torino, Italia. ELRA and ICCL.

Sapna Negi, Tobias Daudert, and Paul Buitelaar. 2019. SemEval-2019 task 9: Suggestion mining from online reviews and forums. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 877–887, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1–2):1–135.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, ACL '02, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Robert Plutchik. 1980. Chapter 1 - a general psycho-evolutionary theory of emotion. In Robert Plutchik and Henry Kellerman, editors, *Theories of Emotion*, pages 3–33. Academic Press.

Tina D. Purnat, Paolo Vacca, Stefano Burzo, Tim Zecchin, Amy Wright, Sylvie Briand, and Tim Nguyen. 2021. WHO Digital Intelligence Analysis for Tracking Narratives and Information Voids in the COVID-19 Infodemic. In John Mantas, Lăcrămioara Stoicu-Tivadar, Catherine Chronaki, Arie Hasman, Patrick Weber, Parisis Gallos, Mihaela Crişan-Vida, Emmanouil Zoulias, and Oana Sorina Chirila, editors, *Studies in Health Technology and Informatics*. IOS Press.

C. J. Van Rijsbergen. 1979. *Information Retrieval*, 2nd edition. Butterworth-Heinemann.

Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518, Vancouver, Canada. Association for Computational Linguistics.

Karan Singla, Dogan Can, and Shrikanth Narayanan. 2018. A multi-task approach to learning multilingual representations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 214–220, Melbourne, Australia. Association for Computational Linguistics.

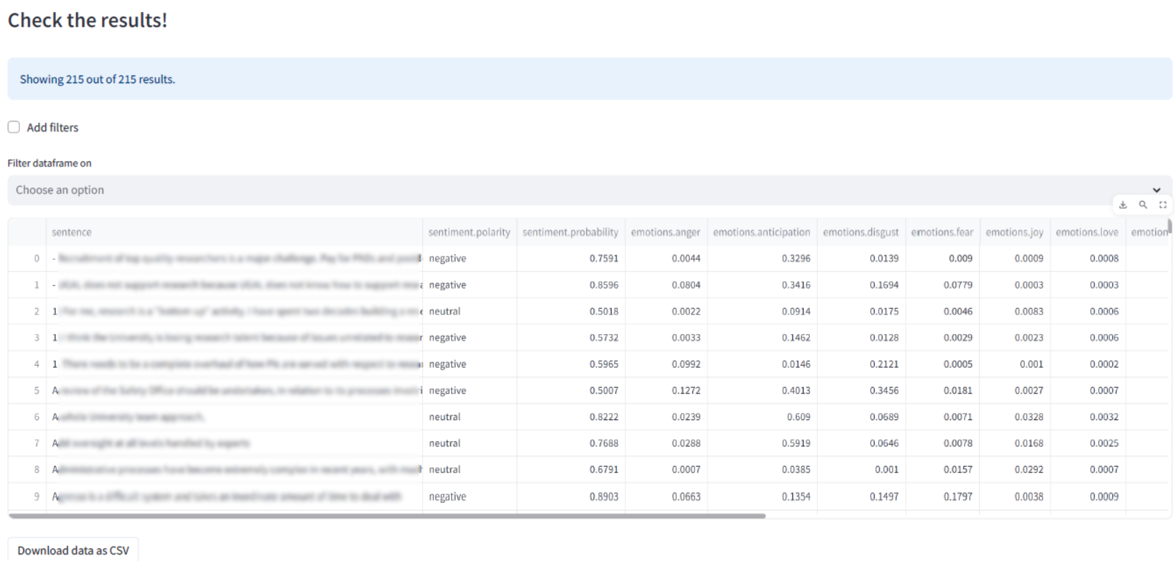
Becky K White, Arnault Gombert, Tim Nguyen, Brian Yau, Atsuyoshi Ishizumi, Laura Kirchner, Alicia León, Harry Wilson, Giovanna Jaramillo-Gutierrez, Jesus Cerquides, Marcelo D'Agostino, Cristiana Salvi, Ravi Shankar Sreenath, Kimberly Rambaud, Dalia Samhouri, Sylvie Briand, and Tina D Purnat. 2023. Using Machine Learning Technology (Early Artificial Intelligence–Supported Response With Social Listening Platform) to Enhance Digital Social Understanding for the COVID-19 Infodemic: Development and Implementation Study. *JMIR Infodemiology*, 3:e47317.

Ziqi Zhang, Jie Gao, and Fabio Ciravegna. 2016. JATE 2.0: Java automatic term extraction with Apache Solr. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2262–2269, Portorož, Slovenia. European Language Resources Association (ELRA).

A Appendix

This Appendix includes additional figures showing snapshots from the *LUCE* dashboard and snapshots from the social media analysis component that was a forerunner of *LUCE* deployed as part of the PANDEM-2 project.

Figure 7: *LUCE*'s Structured Development Process.



	sentence	sentiment.polarity	sentiment.probability	emotions.anger	emotions.anticipation	emotions.disgust	emotions.fear	emotions.joy	emotions.love	emotion
0	-	negative	0.7591	0.0044	0.3296	0.0139	0.009	0.0009	0.0008	
1	-	negative	0.8596	0.0804	0.3416	0.1694	0.0779	0.0003	0.0003	
2	1	neutral	0.5018	0.0022	0.0914	0.0175	0.0046	0.0083	0.0006	
3	1	negative	0.5732	0.0033	0.1462	0.0128	0.0029	0.0023	0.0006	
4	1	negative	0.5965	0.0992	0.0146	0.2121	0.0005	0.001	0.0002	
5	A	negative	0.5007	0.1272	0.4013	0.3456	0.0181	0.0027	0.0007	
6	A	neutral	0.8222	0.0239	0.609	0.0689	0.0071	0.0328	0.0032	
7	A	neutral	0.7688	0.0288	0.5919	0.0646	0.0078	0.0168	0.0025	
8	A	neutral	0.6791	0.0007	0.0385	0.001	0.0157	0.0292	0.0007	
9	A	negative	0.8903	0.0663	0.1354	0.1497	0.1797	0.0038	0.0009	

Figure 8: A snapshot of the tabular output of the opinion dimension analysis on the University Survey Data. The text is blurred due to copyright restrictions.

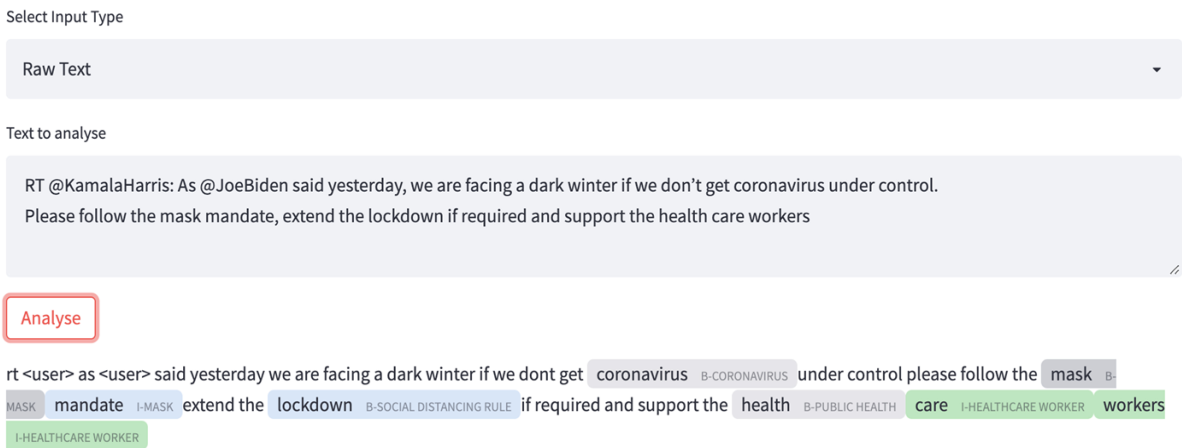


Figure 9: A snapshot of aspect annotation of a new given text.

Figure 10: A snapshot of the social media analysis page of the forerunner prototype of *LUCE* deployed within the PANDEM-2 project.

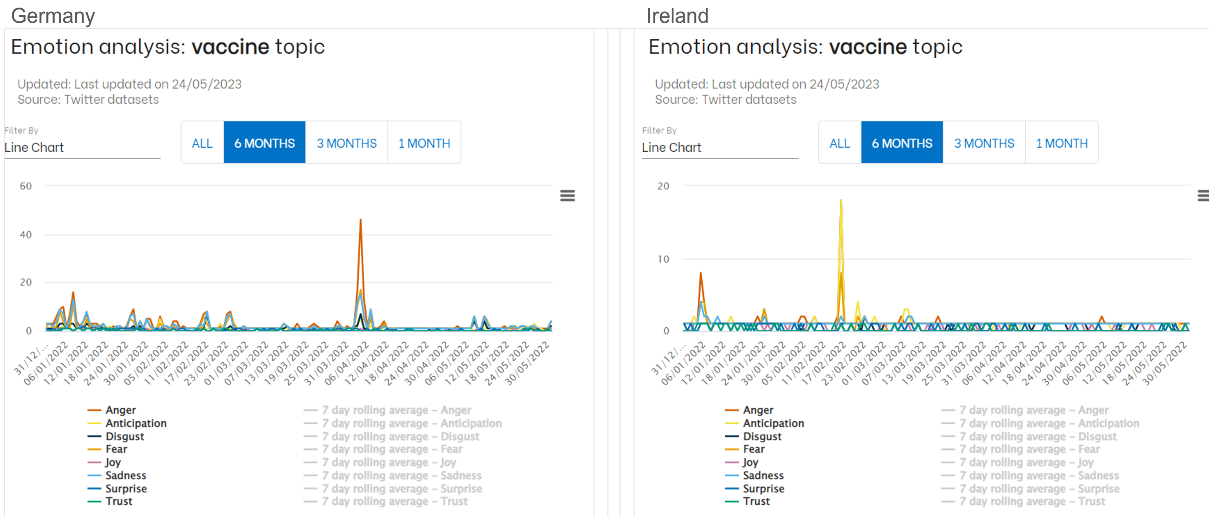


Figure 11: A snapshot comparing the emotions extracted around vaccine topic in Germany versus Ireland using the forerunner prototype of *LUCE* deployed within the PANDEM-2 project.

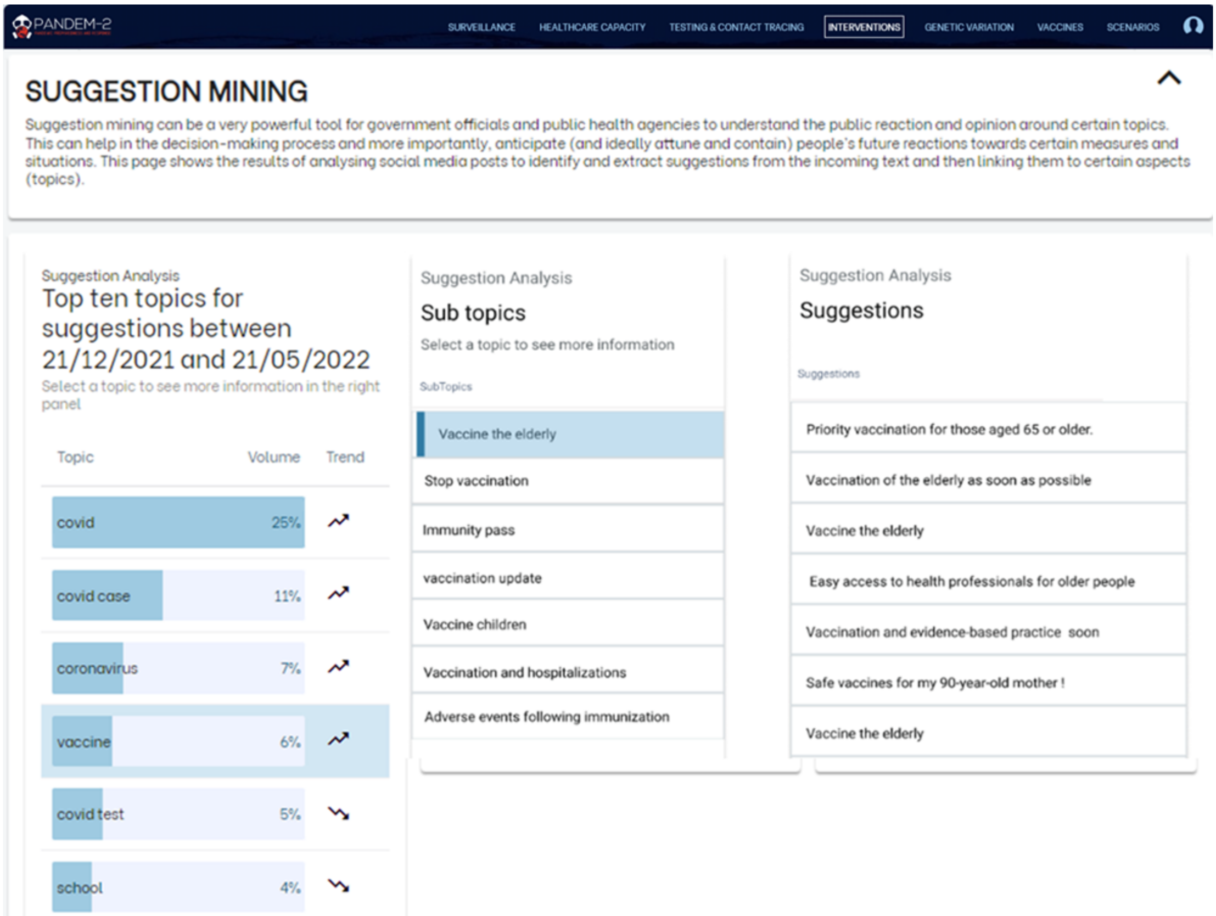


Figure 12: A snapshot of the results of the suggestion analysis module in the forerunner prototype of *LUCE* deployed within the PANDEM-2 project.