

Association for Information Systems

## AIS Electronic Library (AISeL)

---

ACIS 2024 Proceedings

Australasian (ACIS)

---

12-10-2024

# Explainable AI through Thematic Clustering and Contextual Visualization: Advancing Macro-Level Explainability in AFV Systems

Manju Vallayil

*Auckland University of Technology*, manju.vallayil.vijayalekshmi@autuni.ac.nz

Parma Nand

*Auckland University of Technology*, parma.nand@aut.ac.nz

Wei Qi Yan

*Auckland University of Technology*, weiqi.yan@aut.ac.nz

Follow this and additional works at: <https://aisel.aisnet.org/acis2024>

---

### Recommended Citation

Vallayil, Manju; Nand, Parma; and Yan, Wei Qi, "Explainable AI through Thematic Clustering and Contextual Visualization: Advancing Macro-Level Explainability in AFV Systems" (2024). *ACIS 2024 Proceedings*. 101. <https://aisel.aisnet.org/acis2024/101>

This material is brought to you by the Australasian (ACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ACIS 2024 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Explainable AI through Thematic Clustering and Contextual Visualization: Advancing Macro-Level Explainability in AFV Systems

Indicate submission type: Research-in-progress

## Manju Vallayil

School of Engineering, Computer and Mathematical Sciences  
Auckland University of Technology  
Auckland 1010, New Zealand  
Email: manju.vallayil.vijayalekshmi@autuni.ac.nz

## Parma Nand

School of Engineering, Computer and Mathematical Sciences  
Auckland University of Technology  
Auckland 1010, New Zealand  
Email: [parma.nand@aut.ac.nz](mailto:parma.nand@aut.ac.nz)

## Wei Qi Yan

School of Engineering, Computer and Mathematical Sciences  
Auckland University of Technology  
Auckland 1010, New Zealand  
Email: weiqi.yan@aut.ac.nz

## Abstract

In this paper, we present a new framework for computing explainability in Automated Fact Verification models. Our approach uses a graph-based visualization mechanism based on thematic clustering, integrating Gaussian Mixture Models and an Expectation-Maximization algorithm, which are statistical methods for identifying patterns within data. These methods work together in our framework to deliver both local and global explainability. We present initial insights into applying our framework using a fact verification dataset consisting of claim evidence pairs, aiming to compute explainability for verification classifications at both the datapoint level (local) and across datapoints (global). We anticipate that integrating local and global perspectives will yield deeper insights into how individual claims fit within a cluster's context, providing richer, context-sensitive explanations for fact verification tasks, where the context greatly influences the interpretation of individual facts.

**Keywords:** Explainable AI, Automated Fact Verification, Thematic Clustering, Graph-Based Visualization

## 1 Introduction

The “AI/ML Black Box Paradox” highlights the challenges of understanding the decision-making processes of advanced AI and machine learning systems, where the complexity of these methods makes their predictions challenging for humans to comprehend. This lack of clarity raises concerns about the reliability and trustworthiness of these systems, especially in sensitive applications. As a result, there has been a growing research interest in the field of explainable AI (XAI) to make AI models more transparent and understandable (Ali et al. 2023; Gunning 2021, Vallayil et al. 2023).

XAI encompasses a range of strategies and techniques designed to clarify the decision rationale of AI models by explaining how they generate their results. These efforts may leverage the intrinsic explainability of simpler, interpretable models like statistical models or employ post-hoc explainability methods for more complex models that lack intrinsic explainability, such as deep neural networks (DNNs). In some cases, XAI combines these approaches by using interpretable machine learning models for the post-hoc analysis of black-box models, as exemplified by LIME (Ribeiro et al. 2016). However, despite being less interpretable, DNNs such as transformers, which excel at managing long dependencies and understanding contextual relationships in sequential data, are widely adopted due to their high accuracy across various real-life applications, including Automated Fact Verification (AFV) systems.

AFV systems are designed to assess the veracity of claims by leveraging advanced AI techniques. Typically, these systems follow a three-stage pipeline: document retrieval (gathering relevant documents), evidence selection (identifying pertinent evidence sentences), and Recognizing Textual Entailment (RTE, which predicts the relationship between the claim and the evidence). Given the reliance on complex models like transformers, as discussed earlier, these AI-based systems often face a trade-off between accuracy and transparency (Ali et al. 2023). However, both these aspects are critical, especially in the case of AFV. To address this, post-hoc explainability techniques are recommended in the literature and are generally integrated into the RTE phase of AFV systems (Atanasova et al. 2020; Kotonya and Toni 2020b).

Post-hoc methods work by analysing trained AI models to approximate their internal workings and decision logic. These methods reveal relationships between feature values and model predictions without requiring access to the model’s internal mechanisms. This enables users to identify crucial features, quantify their importance, reproduce the decisions made by the black-box model, and detect potential biases in the data or model itself. Post-hoc explainability can be categorized into local and global perspectives, which have been discussed with slight variations in the literature (Atanasova et al. 2020; Shu et al. 2019). A widely accepted interpretation is that local explainability pertains to the analysis of individual data points, whereas global explainability focuses on understanding the broader patterns and associations in the model’s decisions across the entire dataset (Doshi-Velez and Kim 2017; Murdoch et al. 2019; Ribeiro et al., 2016).

However, despite the recognized importance of global explainability in AFV, its implementation remains limited, as existing systems primarily focus on local explainability, which only provides insights into individual predictions without offering a holistic view of the system’s overall behaviour (Kotonya and Toni 2020a). This narrow focus on local explainability fails to capture the broader decision-making patterns of AFV models, indicating the need for a more comprehensive approach. To address this gap, we conducted and published a comprehensive literature review (Vallayil et al. 2023). This current study builds on the findings and gaps identified in that review, focusing on a research direction that advocates integrating both global and local perspectives to provide a balanced approach to explainability in AFV systems. We employ methods that offer insights into AFV model behaviour and reasoning patterns on both global and local levels, aiming for a holistic understanding of explainability in AFV systems. Through this approach, we seek to understand the broader relationships and patterns that inform AFV model decisions while addressing specific concerns related to individual instances.

## 2 Methodology

This section outlines our structured approach to enhancing explainability in AFV systems, focusing on the use of clustering techniques to achieve global explainability. Our methodology encompasses generation of embedded vectors, application of thematic clustering, and the identification of evidence clusters. These steps form the foundation for the subsequent case study and visualizations discussed in Section 3.

## 2.1 Dataset Overview

The dataset we employed in this study is novel and specifically curated for XAI research in AFV. Each claim in the dataset is associated with a unique claim ID, indicating its thematic context, such as climate change, COVID-19, and electric vehicles, facilitating structured analysis of these themes and their impact on fact verification. Claims are labelled as True, False, or Not Enough Info, similar to the structure used in standard fact verification datasets like FEVER (Thorne et al. 2018). Additionally, each claim is paired with six annotated pieces of evidence. By structuring the dataset with multiple pieces of evidence for each claim, we support detailed exploration for XAI research. At the same time, the decision to store these as separate entries, along with the three-class labelling format, aligns with conventional fact verification practices. Thereby, we anticipate that this dataset will support research initiatives in both traditional fact verification as well as explainability-focused studies in AFV systems. Moreover, the dataset’s structure is also tailored to facilitate both local and global explainability, which is a key focus of our study. On a local level, the multiple pieces of evidence for each claim allow for a granular examination of the relationships between a claim and its supporting evidences, providing detailed insights into the rationale behind individual claim evaluations. On a global level, the thematic organization and clustering techniques applied to the dataset help uncover broader patterns and relationships, enhancing the overall understanding of the AFV system’s decision-making process. A detailed description of the dataset, including its creation, annotation processes, and potential applications in AFV systems, will be provided in an upcoming journal publication, with further empirical analysis to be discussed in subsequent work as outlined in our future research section.

## 2.2 Data Preprocessing and Embedded Vector Generation

We begin by loading and preprocessing the dataset, denoted as  $D$ , and then extract claims and their associated evidences within a specific theme, such as climate change. This thematic subset, represented as  $D_r$ , serves as the basis for subsequent analyses. The preprocessing steps ensure that the data is clean, consistent, and ready for embedding generation and clustering. This process involves grouping the data by claims and their corresponding evidences, as well as standardizing text to maintain uniformity.

In our research, selecting an effective model for generating sentence vectors is crucial, as accurately capturing the semantic nuances and contextual relationships within the text is vital for deducing the interrelations between claims and other elements in the dataset. Traditional models like Word2Vec (Mikolov et al. 2013) and GloVe (Pennington et al. 2014), which provide static embeddings, often fall short in capturing the depth needed for nuanced text understanding in NLP tasks like AFV. To address this, we considered BERT (Devlin et al. 2019) and its variants, which generate context-aware vectors and offer a significant improvement over traditional model. However, despite BERT’s success in many NLP tasks, its sentence vectors without fine-tuning have been found to be inadequate for tasks requiring detailed semantic textual similarity, often even underperforming compared to static embeddings like GloVe (Li et al. 2020; Reimers and Gurevych 2019).

Therefore, we selected Sentence-BERT (SBERT), which modifies the pre-trained BERT network by incorporating siamese and triplet network structures (Reimers and Gurevych 2019). This approach enables SBERT to produce semantically meaningful sentence vectors that can be effectively compared using cosine similarity. Among the pre-trained SBERT models available, we chose the open-source *all-mpnet-base-v2* for our research, which is fine-tuned on over 1 billion textual pairs for textual similarity tasks. This model has demonstrated superior performance in capturing both semantic and syntactic correspondences among the fine-tuned models for general-purpose textual similarity tasks, specifically those based on pre-trained transformer models like BERT and MPNet, as well as parameter-reduced models like DistilBERT (Jayanthi et al. 2021), making it particularly effective for AFV tasks that require a nuanced understanding of textual similarity.

The high-quality vectors produced by *all-mpnet-base-v2* are not only critical for accurately assessing textual similarity between claims and evidence in AFV but also support our approach to clustering and visualization. These vectors serve as the foundation for exploring the interconnections within thematic clusters, which is central to enhancing the explainability of AFV systems in our approach, as discussed in the upcoming sections.

## 2.3 Thematic Clustering with GMM-EM

Utilizing the embeddings generated in the previous step, we apply Gaussian Mixture Models (GMM) with the Expectation-Maximization (EM) algorithm to identify thematic clusters in the dataset. Although GMM and EM have been widely used in fields such as speaker identification, emotion recognition, and brain image segmentation (Al-Dujaili and Ebrahimi-Moghadam 2024; Barai et al.

2022; Jiao et al. 2023; Kasim et al. 2021; Li et al. 2020; Moondra and Chahal 2023), our approach uniquely adapts these techniques for AFV, modelling the data as a combination of multiple underlying patterns. Each pattern is represented by a Gaussian distribution defined by its mean and spread, helping us to identify distinct thematic clusters within the dataset.

While GMM supports soft clustering, where data points can belong to multiple clusters with varying degrees of membership, our methodology focuses solely on the primary cluster assignment for each data point, bypassing the probabilistic flexibility that GMM offers. This approach ensures that each claim and evidence is clearly associated with a single cluster, simplifying the subsequent steps of inference and visualization. This is particularly advantageous in the context of AFV as it allows for a clear understanding and representation of the relationships between claims and evidences, facilitating more precise analysis and interpretation.

The clusters are identified as follows:

$$EM(T) = \arg \max_{\theta} \sum_{i=1}^n \log p(e_{T,i} | \theta) \rightarrow \{C_i\}_{i=1}^k \quad (1)$$

where:

- $EM(T)$  denotes the EM algorithm applied to the embeddings of data within the theme  $T$ .
- $e_{T,i}$  is the embedding of the  $i$ -th text input (claim or evidence) from the thematic subset  $D_T$ .
- $\theta$  represents the set of parameters of the GMM, including the means, covariances, and mixture weights of the Gaussian distributions.
- $p(e_{T,i} | \theta)$  is the probability density function of the embedding  $e_{T,i}$  given the model parameters  $\theta$ , as defined by the GMM.
- $\{C_i\}_{i=1}^k$  represents the set of clusters formed as a result of the EM algorithm, where each  $C_i$  is a cluster and  $k$  is the total number of clusters.

## 2.4 In-Depth Analysis of Claims Within Thematic Clusters

To conduct a focused analysis on a specific claim  $c_i$  within a theme, we programmatically determine the cluster  $C_i$  that contains the claim, among the  $k$  clusters identified by the EM algorithm (Equation 1). Within this cluster, we further refine the embeddings to identify a subset of pertinent evidences and claims based on their semantic relevance to the claim  $c_i$  using a similarity measure; this subset hereafter referred to as the Subset of Interest (SOI).

The SOI for  $c_i$  is defined as follows:

$$S(C_i) = \{e_{i,j} | \text{sim}(e_{i,j}, c_i) > \delta \text{ and } e_{i,j} \in C_i\} \cup \{c_k | \text{sim}(c_k, c_i) > \delta \text{ and } c_k \in C_i\} \quad (2)$$

where:

- $S(C_i)$  represents the set of relevant evidences and related claims associated with the claim  $c_i$ .
- $e_{i,j}$  denotes the  $j$ -th evidence associated with the claim  $c_i$ .
- $\text{sim}(e_{i,j}, c_i)$  is the similarity measure between the evidence  $e_{i,j}$  and the claim  $c_i$ .
- $\text{sim}(c_k, c_i)$  denotes the similarity measure between a claim  $c_k$  and the claim  $c_i$ .
- $\delta$  is the similarity threshold above which an evidence  $e_{i,j}$  or a claim  $c_k$  is considered relevant to the claim  $c_i$ .

This analysis provides the foundational data for the graph-based visualizations described in Section 3.2, where we visualize the interconnections and contextual relationships that are specifically influential in the context of the selected claim.

## 3 Case Study Preliminary Findings: Graph-Based Visualizations and Interpretive Insights

In this section, we present preliminary findings from a case study conducted using our novel dataset, demonstrating the implementation of the proposed methodology described in Section 2. These interim results demonstrate the effectiveness of thematic clustering in providing meaningful insights through visualization, demonstrating the relationships and interconnections among claims and their associated evidences (direct and indirect) within the dataset. These visualizations provide a macro-

level view of the thematic clusters (see Appendix 1) and a micro-level analysis of specific claims and their related evidences (see Appendix 2).

### 3.1 Thematic Cluster Visualization

We start by visualizing the thematic clusters identified through the GMM-EM (Section 2.3). For this case study, we focus on the ‘Climate Change’ theme, which has been clustered into three distinct groups ( $k=3$  in Equation 1). This resulted in the total number of samples in this theme ( $n=378$ ) to be segregated to, Cluster 0 with 140 samples, Cluster 1 with 138 samples, and Cluster 2 with 100 samples.

We generated graph visualizations for these clusters, using distinct colours to differentiate between claims and evidence pieces, making the relationships visually clear (see Appendix 1 for the visualization of Cluster 0). The edges between nodes represent the semantic relation between claims and evidence sentences, indicating their relevance. The graph is complex and challenging to interpret in detail due to the dense interconnections. However, it serves as an essential overview of the thematic cluster, providing context for more focused analyses in the following subsection. For example, the claim marked as ‘Claim\_21,’ visible in Cluster 0’s visualization at the top-right corner, is explored in detail for in-depth analysis in the next subsection as part of the case study.

### 3.2 In-Depth Visualization of Claims Within Thematic Clusters

To gain insights into the relationships and dependencies of a selected claim from a post-hoc explainability perspective (Moradi and Samwald 2021), we apply the focused analysis conceptualized in Section 2.4, visualizing the SOI identified through Equation 2 for Claim\_21. This graph highlights the thematic interconnections the claim has with other participants in its cluster, offering a more refined view of its relationships within the SOI (see Appendix 2).

As mentioned in Section 3.1, Cluster 0 has 140 samples and their associated evidences from its theme. Before refining the SOI for Claim\_21, its interconnections graph reveals 168 nodes and 80 edges, which shows the web of relationships within this specific thematic cluster. The large number of nodes initially represents all possible relationships. After applying a similarity threshold ( $\delta=0.7$  in Equation 2), the graph is streamlined to 81 nodes and 80 edges, highlighting the most relevant connections and with more refined and meaningful relationships. This reduction in the number of nodes suggests that many of the initial nodes were either isolated or did not meet the similarity threshold for meaningful connections. The number of edges remaining constant indicates that the relevant relationships were preserved, ensuring that the most pertinent connections between claims and evidences are emphasized. By focusing on these significant relationships, the visualization provides a clear and interpretable representation of both direct associations, such as the annotated evidence of the claim, and related or indirect associations, including connections with other claims and evidence sentences, within the thematic interconnections of the AFV system.

## 4 Discussion

### 4.1 Implications for Strategic Decision-Making through Integrating Global and Local Perspectives for Comprehensive Explainability

As mentioned earlier, our research efforts aim to enhance explainability in AFV systems by integrating both global and local perspectives. The annotated evidences from our novel dataset, typically comprising six evidence datapoints, which offer a detailed analysis of individual claims, addressing the specific concerns and nuances of each instance, which enables us to compute local post-hoc explainability. Concurrently, our proposed approach in this paper, helps to identify a subset of semantically relevant evidences and claims, offering a macro-level view that highlights broader patterns across diverse factual claims. By combining these perspectives, our methodology seeks to develop a comprehensive framework that not only captures macrolevel thematic patterns but also elucidates micro-level insights into individual instances. This ongoing work is guided by the principles of XAI, focusing on achieving key XAI objectives such as interpretability, accuracy, and fidelity (Ali et al. 2023; Guidotti et al. 2018; Doshi-Velez and Kim 2017). By adhering to these objectives, we aim to create AFV systems that are both transparent and interpretable, facilitating more reliable and understandable model behaviour.

We also acknowledge that the interim insights derived in this paper, are preliminary and part of an ongoing investigation. Further empirical validation particularly that defined in the Section 4.2 and consequent refinement of the methodology are planned for future research.

## 4.2 Future Research Directions

Our future research plans focus on two primary areas: objectives for an upcoming publication and key research directions.

**Upcoming Publication Objectives:** As mentioned in Section 2.1, we will be publishing our novel dataset, which we believe will significantly aid and support other research efforts in AFV and XAI. The publication will also feature a comparative analysis of in-depth visualization graphs of different claims, highlighting how our methodology does not necessarily incorporate all manually annotated evidences from the dataset. For example, in the case study we examined, Evidence\_21\_3 was notably absent from the visualization despite being one of the direct pieces of evidence of Claim\_21 (see Appendix 2). This suggests the SOI identifies alternative and potentially more relevant evidence pieces. This insight underscores the value of our approach in providing a deeper and more nuanced understanding of the data, which is critical for high-stakes decision-making scenarios. Additionally, the detailed algorithmic framework of our methodology will also be included in that publication.

**Key Research Directions:** To build upon the initial work detailed in this paper and further advance our contributions to AFV and XAI, we have outlined a detailed plan for continued research. While the SOI of a claim is currently employed for its visualization, we plan to conduct empirical experiments using the SOI for inference. The envisioned plan is to employ an open-source LLM like Llama2 (Touvron et al. 2023) to generate post-hoc explanations in two scenarios: with annotated evidences (the six direct evidences from the dataset) and with the aggregate cluster embeddings of the SOI of the selected individual claim; introducing both local and global perspectives to the claim introspection as discussed. This contextual embedding aggregation (Iliadis et al. 2024; Tang et al. 2023; Zhao et al. 2024) is calculated by averaging the embeddings of all claims and evidences within an SOI of the claim for a cluster. Additionally, we will benchmark our method by conducting a comprehensive performance evaluation using Retrieval Augmented Generation (RAG) (Lewis et al. 2020) to generate explanations for the claim's inference in the context of the entire dataset. This will allow us to assess the effectiveness of our method, as RAG has emerged as a prominent approach in NLP, combining retrieval and generation models to generate more cohesive answers and decrease the occurrence of hallucinations (Salemi and Zamani 2024). Moreover, the claim-based in-depth graph visualization we proposed in this paper (see Appendix 2), can serve as a visual aid for global explainability, similar to how LIME (Ribeiro et al. 2016) visualization provides instance-level local explainability. While this holistic approach is expected to provide more comprehensive insights, it requires empirical validation through our ongoing research.

Additionally, we intend to record the results of experiments involving various clustering techniques and similarity thresholds, examining their effects across different themes and clusters. This will allow us to refine our methodology and better understand its robustness and applicability. We also plan to open-source our project on GitHub, ensuring transparency and fostering collaboration by maintaining and version-controlling the code. Furthermore, we will host a prototype space on Hugging Face, enabling practitioners and researchers to interactively test our results and validate the effectiveness of our approach.

## 5 Conclusion

Our ongoing research leverages thematic clustering and contextual visualization to enhance macro-level explainability in AFV systems. By identifying additional relevant evidences, beyond those manually annotated, our methodology enhances decision-making processes through a more comprehensive understanding of the information landscape. This refinement process, is crucial for making informed and accurate decisions, as it highlights the most pertinent connections, providing decision-makers with a clearer view of the evidence base, ultimately supporting more reliable and informed decisions. Guided by XAI objectives, we aim to develop AFV systems that are both transparent and interpretable, facilitating more reliable and deeper understandable model behaviour. As we continue to refine and expand this methodology, we anticipate further insights that will advance the explainability in AFV systems.

## References

- Al-Dujaili Al-Khazraji, M. J., and Ebrahimi-Moghadam, A. 2024. "An Innovative Method for Speech Signal Emotion Recognition Based on Spectral Features Using GMM and HMM Techniques." *Wireless Personal Communications* 134(2): 735–753.
- Ali, Sajid, Rashid, Uzma, Qamar, Umer, Siddique, Muhammad Asim, and Mustafa, Rizwan. 2023. "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence." *Information Fusion*, p. 101805.
- Atanasova, P., Simonsen, J.G., Lioma, C., and Augenstein, I. 2020. "Generating Fact Checking Explanations." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, 5–10 July 2020, pp. 7352–7364.
- Barai, B., Chakraborty, T., Das, N., Basu, S., and Nasipuri, M. 2022. "Closed-Set Speaker Identification Using VQ and GMM Based Models." *International Journal of Speech Technology* 25(1): 173-196.
- Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, and Toutanova, Kristina. 2019. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*.
- Doshi-Velez, Finale, and Kim, Been. 2017. "Towards A Rigorous Science of Interpretable Machine Learning." *arXiv: Machine Learning*. url: <https://api.semanticscholar.org/>.
- Guidotti, Riccardo, Monreale, Anna, Ruggieri, Salvatore, Turini, Franco, Giannotti, Fosca, and Pedreschi, Dino. 2018. "A survey of methods for explaining black box models." *ACM Computing Surveys (CSUR)* 51(5), pp. 1–42.
- Gunning, David, Vorm, Eric, Wang, J.Y., and Turek, Matt. 2021. "DARPA's Explainable AI (XAI) Program: A Retrospective." *Applied AI Letters* 2, e61. <https://doi.org/10.1002/AIL2.61>.
- Guo, Z., Schlichtkrull, M., and Vlachos, A. 2022. "A Survey on Automated Fact-Checking." *Transactions of the Association for Computational Linguistics*, 10, 178–206.
- Iliadis, Dimitrios, Peikou, Maria, Adamidou, Chrysoula, and Kyriakopoulou, Antonia. 2024. "A comparison of embedding aggregation strategies in drug–target interaction prediction." *BMC Bioinformatics* 25(1), p. 59.
- Jayanthi, Sai Muralidhar, Embar, Varsha, and Raghunathan, Karthik. Dec. 2021. "Evaluating Pretrained Transformer Models for Entity Linking in Task-Oriented Dialog." *Proceedings of the 18th International Conference on Natural Language Processing (ICON)*. Ed. by Sivaji Bandyopadhyay, Sobha Lalitha Devi, and Pushpak Bhattacharyya. National Institute of Technology Silchar, Silchar, India: NLP Association of India (NLP AI), pp. 537–543. url: <https://aclanthology.org/2021.icon-main.65>.
- Jiao, Z., Ji, Y., Gao, P., and Wang, S. H. 2023. "Extraction and Analysis of Brain Functional Statuses for Early Mild Cognitive Impairment Using Variational Auto-Encoder." *Journal of Ambient Intelligence and Humanized Computing*, 1–12.
- Kasim, Fatin Amelia Binti, Yaacob, Siti Nurulhuda Mohd, Mohamad, Haslina, and Md Nor, Anis Nurashikin. 2021. "Gaussian mixture model-expectation maximization algorithm for brain images." 2021 2nd *International Conference on Artificial Intelligence and Data Sciences (AiDAS)*. *IEEE*, pp. 1–5.
- Kotonya, Neema, and Toni, Francesca. 2020a. "Explainable Automated Fact-Checking: A Survey." *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*. Online, pp. 5430–5443. isbn: 9781952148279. doi: 10.18653/v1/2020.coling-main.474. arXiv: 2011.03870.

- Kotonya, Neema, and Toni, Francesca. 2020b. “Explainable Automated Fact-Checking for Public Health Claims.” In *Proceedings of the EMNLP 2020—2020 Conference on Empirical Methods in Natural Language Processing*, Online, 19–20 November 2020, pp. 7740–7754.
- Lewis, Patrick, Perez, Ethan, Piktus, Aleksandra, Petroni, Fabio, Karpukhin, Vladimir, Goyal, Naman, Kwiatkowski, Tom, Parikh, Ankur, Boyd-Graber, Jordan, and Riedel, Sebastian. 2020. “Retrieval-augmented generation for knowledge-intensive NLP tasks.” *Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS ’20*. Vancouver, BC, Canada: Curran Associates Inc. isbn: 9781713829546.
- Li, Bohan, Zhou, Hao, He, Junxian, Wang, Mingxuan, Yang, Yiming, and Li, Lei. 2020. “On the Sentence Embeddings from Pre-trained Language Models.” In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, pp. 9119–9130. Association for Computational Linguistics.
- Li, X., Zhong, J., Wu, X., Yu, J., Liu, X., and Meng, H. 2020. “Adversarial Attacks on GMM i-Vector Based Speaker Verification Systems.” In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6579-6583. IEEE.
- Mikolov, Tomas, Chen, Kai, Corrado, Gregory S., and Dean, Jeffrey. 2013. “Efficient Estimation of Word Representations in Vector Space.” In *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Moondra, A., and Chahal, P. 2023. “Speaker Recognition Improvement for Degraded Human Voice Using Modified-MFCC with GMM.” *International Journal of Advanced Computer Science and Applications* 14(6).
- Moradi, Milad, and Samwald, Matthias. 2021. “Post-hoc explanation of black-box classifiers using confident itemsets.” *Expert Systems with Applications* 165, p. 113941.
- Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. 2019. “Definitions, Methods, and Applications in Interpretable Machine Learning.” *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 116, pp. 22071–22080.
- Pennington, J., Socher, R., and Manning, C.D. 2014. “GloVe: Global Vectors for Word Representation.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Reimers, Nils, and Gurevych, Iryna. 2019. “Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks.” In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.
- Ribeiro, Marco Tulio, Singh, Sameer, and Guestrin, Carlos. 2016. ““Why should I trust you?” Explaining the predictions of any classifier.” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144.
- Salemi, Alireza, and Zamani, Hamed. 2024. “Evaluating Retrieval Quality in Retrieval-Augmented Generation.” *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR ’24*. Washington DC, USA: Association for Computing Machinery, pp. 2395–2400. isbn: 9798400704314. doi: 10.1145/3626772.3657957. url: <https://doi.org/10.1145/3626772.3657957>.
- Shu, K., Cui, L., Wang, S., Lee, D., and Liu, H. 2019. “dDEFEND: Explainable Fake News Detection.” In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA*, 4–8 August 2019, pp. 395–405.
- Tang, Jiaxiang, Yin, Dawei, Zhang, Jinhui, and Yin, Jianwei. 2023. “Secure Embedding Aggregation for Federated Representation Learning.” *2023 IEEE International Symposium on Information Theory (ISIT)*. IEEE, pp. 2392–2397.

Thorne, James, Andreas Vlachos, Oana Cocarascu, Christos Christodoulopoulos, and Arpit Mittal. 2018. "The Fact Extraction and VERification (FEVER) Shared Task." *In Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, Brussels, Belgium, 1 November 2018, pp. 1–9.

Touvron, Hugo, Lavril, Thibaut, Izacard, Gautier, Martinet, Xavier, Lachaux, Marie-Anne, Lacroix, Timothée, Rozière, Baptiste, Goyal, Naman, Hambro, Eric, Azhar, Faisal, Rodriguez, Aurelien, Joulin, Armand, Grave, Edouard, and Lample, Guillaume. 2023. "Llama 2: Open Foundation and Fine-Tuned Chat Models." *ArXiv abs/2307.09288*. url: <https://api.semanticscholar.org/CorpusID:259950998>.

Vallayil, Manju, Parma Nand, Wei Qi Yan, and Héctor Allende-Cid. 2023. "Explainability of Automated Fact Verification Systems: A Comprehensive Review." *Applied Sciences* 13(23), p. 12608. <https://doi.org/10.3390/app132312608>.

Zhao, Wenbin, Zheng, Wenjie, Wang, Yibo, and Wang, Tianyi. 2024. "Fake News Detection Based on Knowledge-Guided Semantic Analysis." *Electronics* 13(2), p. 259.

## Appendix 1

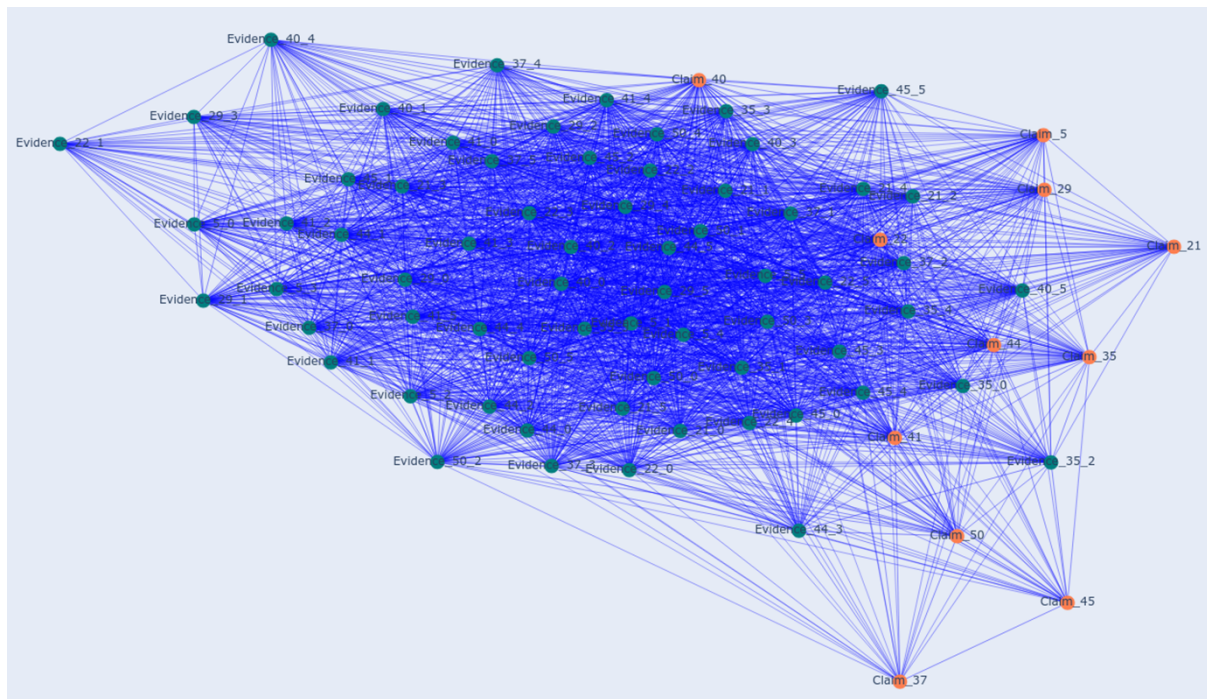


Figure 1: Graph visualization of Cluster o

This thematic cluster visualization shows the relationships between claims (coral nodes) and evidences (teal nodes) within a specific cluster of a theme. The edges represent the connections between claims and evidences, demonstrating the complexity and density of the thematic cluster. This visualization provides a comprehensive overview of the thematic context, setting the stage for more focused analyses.

