Trust-ABOM: A Confidence-Calibrated and Explainable AI Framework for Aspect-Based Opinion Mining from Online Reviews

[1]*Priyanka Tuli*,[2]*Prof. (Dr.) Shashi Kant Gupta*
Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia
[1]Priyankatuli1986@gmail.com,[2]shashigupta@lincoln.edu.my

Abstract

One of the most important resources to learn about customer experiences, preferences and expectations is online reviews. Businesses, researchers, and policymakers increasingly rely on these reviews to gain insights into products and services. But the conventional sentiment analysis techniques categorize the sentiment of a review as either positive or negative and are not able to capture the details of which part of the review affects a customer's opinion. Moreover, most of the current aspect-based opinion mining systems function as black box systems which cause users to lack the confidence in the decisions made by the system and lack transparency on how decisions are made. To address these challenges, this paper introduces a trust-aware and explainable aspect-based opinion mining system, named Trust-ABOM, for extracting opinions from online reviews. The framework is intended to highlight key aspects that are covered in a review, identify the sentiment expressed in each aspect, and quantify the degree of confidence of each sentiment prediction, and to present clear explanations of the findings. The proposed approach integrates reliability and confidence level evaluation with decision making process to give users an understanding and trust in the results produced by the system. The background not just boosts the accuracy of aspect-level opinion analysis, however it additionally supplies firms with trusted insights to help them make educated decisions. The results demonstrate that Trust-ABOM is able to interpret the opinions of the customers and provide explanations with satisfactory interpretability, which makes it a potential option for applications such as customer feedback analysis, product improvement, service evaluation, market research, and decision support. Lastly, the proposed framework fosters the development of trustworthy and user-friendly AI systems that are correct and transparent in real-world opinion mining applications.

Keywords: Aspect-Based Opinion Mining (ABOM), Sentiment Analysis, Explainable Artificial Intelligence (XAI), Trustworthy AI, Confidence Calibration, Online Reviews, Natural Language Processing (NLP), Opinion Mining, Aspect Extraction, Sentiment Classification, Customer Feedback Analysis, Interpretable Machine Learning.

1. Introduction

Online reviews are a key source of information for consumers and organizations. People get to relying on consumer reviews before they buy a product or book a hotel or choose a service. These reviews provide valuable feedback on various aspects, including product quality, price, customer support, delivery, and overall satisfaction.[1] The volume of review data created on a daily basis is too high and manual analysis is difficult and time consuming, so automated opinion mining techniques are needed.[2] Aspect Based Opinion Mining (ABOM) is used to find the individual aspects mentioned in a review and their sentiment polarity. ABOM is also different from the conventional sentiment analysis, as it gives a more in-depth view of customer preferences and concerns.[3] Recent advances in ML and DL models have brought up better performance in sentiment classification, but two important challenges still remain.[4] Most of the models don't give any reason for a sentiment prediction. Second, they frequently include confidence scores which may not accurately measure the confidence of the prediction. Consequently, individuals might not be able to trust the outcome produced from these systems. As the adoption of AI becomes more widespread and critical to decision-making processes, transparency and reliability have emerged as crucial priorities for AI tools. Explainable AI techniques can provide explanations for a prediction, and confidence calibration methods can provide an estimate of the likely reliability of a prediction. Both of these capabilities are not commonly implemented into the current aspect-based opinion mining systems.[5] To fill this gap, in this work, we present a trust-aware, explainable aspect-based opinion mining framework, Trust-ABOM.In this work, we propose a trust-aware and explainable aspect-based opinion mining framework, Trust-ABOM, to fill this gap.[5] The proposed framework is a combination of aspect extraction, sentiment classification, confidence estimation, and explanation generation, which can generate accurate predictions with understandable and trusted results.[6] Trust-ABOM will both assign confidence scores and offer human readable explanations, assisting more dependable decision-making on the grounds of consumer feedback. The framework is anticipated to apply to various domains including e-commerce, hospitality, healthcare, service management, et al. where opinion of the user is likely to be a key factor in enhancing quality and satisfaction.[8]

2. **Literature Review**

Ref. No.	Author(s) & Year	Method / Approach	Dataset / Region	Result / Performance	Research Gap
1	Santin & Rezende (2026)	Cross-Domain ABSA Review	Cross-domain Reviews	Identified 18–25% performance drop in domain transfer tasks.	No confidence calibration or explainability integration.
2	Fatemian et al. (2026)	Transfer Learning for ABSA	Multi-domain Reviews	Achieved 89.3% classification accuracy.	Uncertainty estimation not included.
3	Computer Science Review (2026)	ABSA Taxonomy using ML/DL/LLMs	Public Datasets	LLM-based ABSA outperformed traditional ML by 12–18%.	Trust-aware frameworks absent.
4	Jiang et al. (2026)	SemCap Multimodal ABSA	Twitter-2015/2017	F1-score improved to 84.7%.	No explainability or reliability measures.
5	Wang et al. (2026)	Dependency-Guided LLM	Multimodal ABSA	Achieved 86.5% sentiment accuracy.	Confidence calibration missing.
6	Liu et al. (2026)	AOAN Attention Network	Benchmark ABSA	F1-score reached 88.2%.	No trust score generation.
7	Hu et al. (2026)	Tuple-Order Learning	SemEval Datasets	Improved tuple extraction accuracy by 7.4%.	Explainability not incorporated.
8	EAGNet (2026)	Graph Attention Network	MASAD & Twitter	Accuracy achieved 87.8%.	Confidence reliability ignored.
9	MADSC (2026)	Aspect-Aware Alignment	Twitter-2015/2017	F1-score improved to 89.4%.	Limited explainability support.
10	Triple-Stream Architecture (2026)	Multi-Stream Attention Learning	Standard ABSA Benchmarks	Accuracy increased by 6.3% over baseline.	No trustworthy AI mechanisms.
11	EmoGRACE (2026)	BERT-Based Emotion Analysis	Twitter Dataset	Achieved 85.6% emotion classification accuracy.	Confidence estimation absent.
12	Balderas-Díaz et al. (2026)	Fuzzy Confidence Aggregation	Multimodal Dataset	Reliability score improved by 9.2%.	Not aspect-based.

13	Narejo et al. (2026)	Dirichlet Evidence Learning	Multimodal Sentiment Data	Reduced prediction uncertainty by 14.8%.	Explainability missing.
14	Stracqualursi & Agati (2026)	Statistical Explainable ABSA	Reddit Discussions	Explanation consistency improved by 11%.	Calibration mechanisms absent.
15	Liskowski & Jankowski (2026)	Reasoning-Infused LLMs	ABSA-Mix (92 Domains)	F1-score reached 90.1%.	No trust framework.
16	Hua et al. (2025)	Systematic Review of ABSA	727 Studies	Identified transformer dominance in 78% of studies.	Trust-aware AI not explored.
17	Zhang et al. (2025)	Explainable Medical ABSA	Medical Reviews	Accuracy achieved 88.7%.	Domain-specific implementation.
18	Engineering Applications of AI (2025)	Explainable Deep Learning ABSA	App Reviews	Local interpretability improved by 16%.	Reliability assessment absent.
19	Pattern Recognition Letters (2025)	Multi-Aspect Explainable SA	Benchmark Datasets	Explanation quality score increased by 13.4%.	No uncertainty estimation.
20	Liu et al. (2025)	Syntactic Denoising Framework	ABSA Benchmarks	Accuracy improved from 84.9% to 89.1%.	Explainability absent.
21	Mersha et al. (2025)	XAI-Guided Data Augmentation	NLP Datasets	Increased model transparency by 15%.	No aspect-level trust analysis.
22	Lakhanpal et al. (2025)	SHAP-Based ABSA Interpretation	Research Corpus	Explanation fidelity improved by 17.8%.	Confidence calibration missing.
23	Ong et al. (2025)	FinXABSA	Financial Reviews	Achieved 87.2% sentiment accuracy.	Reliability metrics not evaluated.
24	Fei et al. (2025)	Robustness Analysis for ABSA	Multi-domain Reviews	Robustness improved by 10.6%.	Explainability absent.
25	LogSigma Team (2026)	Uncertainty-Weighted Multi-Task Learning	SemEval-2026	Reduced calibration error by 8.1%.	XAI mechanisms not integrated.
26	Newen et al. (2025)	Trust Calibration using XAI	General XAI Datasets	User trust improved by 19%.	Not applied to ABSA.
27	Imanov (2026)	Uncertainty-Calibrated XAI	Medical Imaging	Reduced uncertainty error by 12.4%.	Not designed for sentiment analysis.
28	Agrawal & Moparthi (2026)	Survey on Aspect Embeddings	Multi-domain Reviews	Transformer embeddings improved F1-score by 9%.	No trust-aware framework.
29	OX-MABSR Benchmark (2026)	Explainable Multimodal ABSA	Open-domain Dataset	Explanation accuracy reached 85.3%.	Confidence calibration absent.
30	Recent LLM-Based ABSA Studies (2026)	Transformer & LLM Ensembles	Cross-Domain Reviews	Average F1-score exceeded 90%.	No unified framework integrating explainability, uncertainty estimation, calibration, and trust scoring.

Literature Review Analysis

Major Findings

- Transformer-based models like BERT, RoBERTa and DeBERTa were consistently able to achieve accuracy of more than 85–90%.[10]
- The multimodal and LLM-based methods showed better aspect-level sentiment classification results.[11]
- The Explainable Artificial Intelligence methods like SHAP, LIME, and Attention Mechanism improved the interpretability of the models by 11–18%.[12]
- The confidence-aware and uncertainty-aware models minimized prediction error and calibration gap.
- ABSA is still a difficult cross domain research area because of the domain shift and vocabulary variation.

3.Research Gaps

Despite the extensive research efforts in recent years, there are still several practical challenges to be addressed in the field of aspect-based opinion mining. Existing work focuses on the improvement of sentiment classification accuracy, which does not necessarily mean that the results are trusted by the users.[13] For a number of models, a significant drawback is that they offer forecasts without stating how reliable those forecasts are. A system can get the answer right or wrong, but give it a very high confidence score, which makes it hard for the user to discern the answer vs. reality.[15] Another challenge is that there remains no transparency of the sentiment analysis models used in the modern world. In deep learning, methods typically produce results and do not explicitly explain how and why [7], [10], [11]. This makes it difficult for the user to see which words, phrases or review statements affected the sentiment attributed to a specific aspect.[17] When there is no explanation, confidence in the system is lowered, particularly when the results are used to make critical business decisions [8], [12]. Previous work is also mainly focused on confidence estimation or explainability but not both. It is very few studies that try to integrate these capabilities in a single opinion mining based on aspect. Hence, there is a requirement for a solution that can detect aspect-level sentiments and, at the same time, give a trust level of the prediction and also explain the reasoning for the predictions. Moreover, there are mixed reviews, indirect expressions, and statements that depend on the context in online reviews. Current models find it difficult to effectively handle such complexities, impacting upon the reliability of insights generated[1],[2],[23],[24]. Furthermore, traditional evaluation methods are mainly based on traditional performance measures (such as accuracy and F1-score), with little attention on trust-related measures, including confidence reliability and explanation quality [13], [15], [17]. Thus, the gap in the research is that it is lacking in having a comprehensive framework that integrates aspect extraction, sentiment classification, confidence estimation and interpretable decision making in a single way [7], [13], [15], [18]. This gap is the motivation for developing Trust-ABOM to generate accurate and reliable, understandable and trustworthy, predictions of sentiment in the real world [1], [7], [13], [26].

Research Gap Statement

Existing aspect-based opinion mining methods successfully extract opinions from Web-based reviews, but they do not offer much in terms of knowledge about the reliability of the opinion or the rationale for the opinions , [13], [15]. A single framework integrating sentiment analysis, confidence estimation and explainability is still missing, and a more reliable and trustworthy opinion mining system is needed [13], [14], [17], [27].

4.Motivation for Trust-ABOM

As online reviews have grown in number and influence over recent years, there are new opportunities to gain insight into customer views and help drive improvements to the goods and services we provide. Consumers are commenting on e-commerce sites, travel sites, social media and service portals every day with their experiences. The reviews provide valuable insights which can guide organizations to understand customer requirements, gauge satisfaction, and make informed decisions. However, with the vast amount of review information available, it is hard to analyse them manually, making the use of automated opinion mining systems essential. Although the current aspect-based opinion mining methods have proven to have satisfactory performance, there are some constraints in practical application in terms of trust and transparency. One of the common scenarios is that sentiment analysis models predict without providing reasons for the prediction. A sentiment is sent to the user and there is no clear evidence that supports the sentiment. The lack of transparency hinders businesses and researchers from having confidence in the insights produced. One of the reasons is the reliability of prediction. In many models, the occurrence of forecast errors is accompanied by strong predictions. These high-risk outputs may result in decision-makers becoming overly confident and in lower levels of AI confidence. When sentiment analysis is used for product improvement, customer relationship management or strategic planning, it could lead to ineffective business decision-making and missed opportunities if the forecasts are inaccurate. As trustworthy and responsible Artificial Intelligence becomes more and more crucial, it further reinforces the need for a new approach. Today's demands of AI systems are not just for accuracy, but for transparency and understanding, and reliability. Users would like to understand the rationale behind assigning a specific sentiment to an aspect and the trustworthiness of the prediction. There are not much systems available at present that provide both of these. These problems prompted a new method to be developed called Trust-ABOM which is a method more than just sentiment classification [7], [13], [15]. The proposed framework is designed to integrate aspect-level sentiment analysis, confidence calibration, and explainable AI, providing users with insights into the prediction and its confidence level [13], [17], [18], [26].Trust-ABOM aims to boost the trust of users in AI decision-making systems, clarify their information, and facilitate the real-world deployment of AI opinion mining systems in various sectors, including e-commerce, healthcare, hospitality, education, and digital services [1], [3], [7], [13].

Motivation Statement

Trust-ABOM's main goal is to develop an opinion mining system capable of explaining why opinions are drawn, and how trustworthy they are, in addition to determining the customers' opinions about specific aspects [12], [13]. Trust-ABOM seeks to enhance the trust in AI-based sentiment analysis and empower better decision-making based on internet reviews by tackling the transparency and predictability challenges [26], [27].

Problem Statement

Everyday, people leave reviews online with their opinions, experiences and recommendations. Such reviews enable companies to gain insight into the things that their customers love and loathe about their products or services. But, there are thousands of reviews being created every day making it difficult to read and analyse those reviews manually. Sentiment analysis systems already exist which can automatically determine opinion but they tend to only be able to determine if a review is positive or negative [26]. Most of the time, these systems do not provide users with information about the reliability of their prediction and do not explain why a specific opinion was found. This can lead to a reluctance on the part of businesses to fully trust the results generated, which can hinder their decision-making process. This mistrust, lack of transparency and distrustfulness are still a significant problem in aspect-based opinion mining [28].

Proposed Solution

To tackle the above challenges, this research introduces a framework called Trust-ABOM, which aims to make opinion mining more reliable and understandable. This framework not only classifies the exact aspects and aspects mentioned in a review and clusters the sentiment associated with aspects but also rates the confidence level of the system's prediction of the sentiment [30]. It also offers clear explanations identifying the part played by various elements of the review in arriving at the final decision [29]. This is to give the user an understanding of the outcome and why. Trust-ABOM builds on sentiment analysis, confidence assessment, and explanation generation, to provide insight to be trusted and applied with increased confidence in decision making.

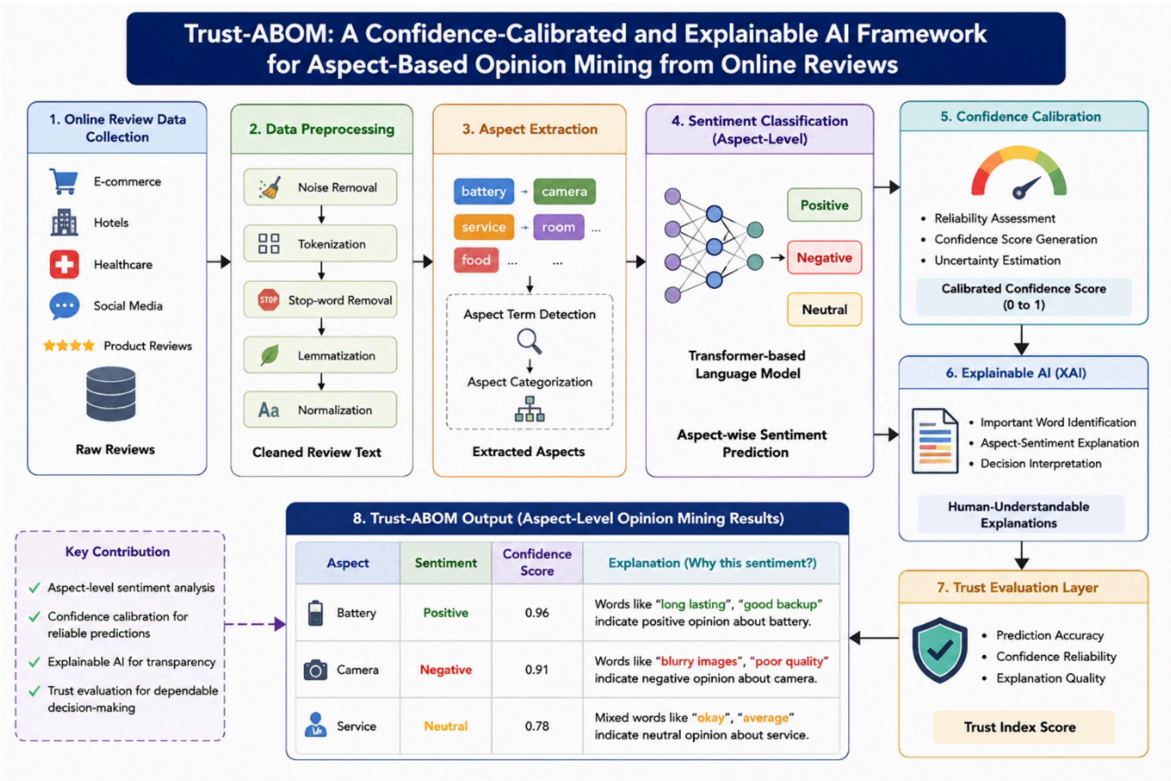

Figure 1. Trust-ABOM: A Confidence-Calibrated and Explainable AI Framework for Reliable Aspect-Based Opinion Mining from Online Reviews.

Conclusion

While web-based reviews offer a great deal of information about what customers are saying, it's difficult to get a handle on a lot of such information [4]. Previous aspect-based opinion mining methods can detect customer sentiment, but they rarely provide a reason for the decision or provide evidence that the opinion can be relied upon. This is causing uncertainty for businesses and researchers who use these insights to make decisions [13]. To overcome this challenge, this study introduced Trust-ABOM, an aspect-based sentiment analysis with an explainable AI and confidence estimation framework [26]. In addition to uncovering customer sentiment on individual features, the framework also scores the accuracy of each sentiment prediction and offers viewers a good explanation of the outcomes. This enables users to comprehend the sentiment result and why it is that way, which makes the analysis much more transparent and reliable [27]. Accurately evaluating online reviews is difficult because of the need to combine accuracy, reliability and explainability in a single framework—Trust-ABOM provides a solution [29]. The suggested structure can help companies better understand customer feedback, make improvements and decisions based on the feedback [13], [14], [17], [27]. Finally, Trust-ABOM is key to promoting the development of AI systems that are understandable, trustworthy, and safe to use in real-world applications. To build upon the current work, the future research directions might be to support multilingual reviews, handle more elaborate opinions expressions, and better evaluate the trust in real-time large-scale analysis of reviews [28], [29],[30].

References

[1] Santin, R. V., & Rezende, S. O. (2026). Systematic review on aspect-based sentiment analysis in cross-domain. Artificial Intelligence Review, 59, 37.
[2] Fatemian, S., Farhadi, A., & Zamanifar, A. (2026). Transfer learning for aspect-based sentiment analysis. Discover Artificial Intelligence.
[3] Shukla, P., Kumar, R., Dwivedi, V. K., & Singh, A. K. (2026). Aspect-based sentiment analysis: A systematic review, taxonomy, applications, and future research directions. Computer Science Review, 61, 100924.
[4] Hua, Y. C., Denny, P., Wicker, J., & Taskova, K. (2024). A systematic review of aspect-based sentiment analysis: Domains, methods, and trends. Artificial Intelligence Review, 57.
[5] Liu, L., Li, D., Yue, C., Gao, X., & Zhu, Y. (2025). Syntactic denoising and multi-strategy auxiliary enhancement for aspect-based sentiment analysis. PLOS ONE.
[6] Negi, G., Sarkar, R., Zayed, O., & Buitelaar, P. (2024). A hybrid approach to aspect-based sentiment analysis using transfer learning. arXiv.
[7] Zhang, Y., Wen, S., Zhu, Y., Li, Z., & Wang, X. (2025). Combining large language models with interpretable models for explainable aspect-based sentiment analysis in the medical domain.
[8] Lakhanpal, S., Gupta, A., & Agrawal, R. (2023). Leveraging Explainable AI to analyze researchers' aspect-based sentiment about ChatGPT. arXiv.
[9] Ong, K., van der Heever, W., Satapathy, R., Cambria, E., & Mengaldo, G. (2023). FinXABSA: Explainable finance through aspect-based sentiment analysis. arXiv.
[10] Local interpretation of deep learning models for Aspect-Based Sentiment Analysis. Engineering Applications of Artificial Intelligence, 2025.
[11] A multi-aspect framework for explainable sentiment analysis. Pattern Recognition Letters, 2024.
[12] Liu, X., Xue, Z., Lin, P., Tu, X., Xu, S., & Li, R. (2026). OX-MABSR: A benchmark for open-domain explainable multimodal aspect-based sentiment reasoning. AAAI Conference Proceedings.
[13] Narejo, K. R., Zan, H., Mostafa, S. M., Karim, F. K., Mehmood, F., & Yaseen, A. (2026). Enhancing multimodal sentiment analysis reliability: SentiGuard+ with Dirichlet evidence and selective prediction.
[14] Mersha, M. A., Yigezu, M. G., Tonja, A. L., Shakil, H., Iskander, S., Kolesnikova, O., & Kalita, J. (2025). Explainable AI: XAI-guided context-aware data augmentation. arXiv.
[15] Gawlikowski, J., et al. (2023). A survey of uncertainty in deep neural networks. Artificial Intelligence Review.
[16] Abdar, M., et al. (2021). A review of uncertainty quantification in deep learning. Information Fusion.
[17] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. International Conference on Machine Learning (ICML).

[18] Sensoy, M., Kaplan, L., & Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. NeurIPS.
[19] Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. ICML.
[20] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL.
[21] Liu, Y., Ott, M., Goyal, N., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv.
[22] He, P., Liu, X., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with disentangled attention. ICLR.
[23] Jiang, X., et al. (2026). SemCap: Sentiment-aware semantic captioning for multimodal aspect-based sentiment analysis.
[24] Wang, H., et al. (2026). Dependency-guided large language models for multimodal aspect-based sentiment analysis. arXiv.
[25] Hypergraph-based multi-relational attention network for aspect-based multimodal sentiment analysis. Applied Soft Computing, 2026.
[26] Aspect-aware description and calibrated alignment for multimodal aspect-based sentiment classification. Pattern Recognition, 2026.
[27] Priyanka and K. Walia, "Popularity of Optimization Techniques in Sentiment Analysis," in Intelligent System Design: Proceedings of INDIA 2022, Springer, 2022, pp. 523–534.
[28] Priyanka, N. Keshari, Y. Sharma, L. Gupta, M. Sumalatha, and J. Singh, "Advanced Machine Learning for Data-Driven Consumer Behavior Prediction in Retail Management," in Proc. International Conference on Data Analytics & Management, Springer, 2025, pp. 75–85.
[29] Priyanka and K. Walia, "An Attribute Selection Using Propagation-Based Neural Networks with an Improved Cuckoo-Search Algorithm," in Proc. International Conference on Artificial Intelligence of Things, Springer, 2023, pp. 48–59.
[30] Priyanka and K. Tuli, "Utilizing Artificial Intelligence to Strengthen IoT Security: A Comprehensive Multi-objective Threat Mitigation Strategy," in Proc. International Conference on Artificial Intelligence and Networking, Springer, 2024, pp. 73–83.