Digital and social media monitoring agentic system for brand reputation

Mohammed Abdul Moiz[1], Muhammad Zaki Ahmad[2]
[1]Sr.Information Security Analyst, Media Communication and Coordination Department, Saudi Aramco, Dhahran, Saudi Arabia
[2]Senior Data Scientist, Communication Content and Channels Department

***Abstract*—Brand reputation is increasingly shaped by information that appears rapidly across news outlets, blogs, and social media platforms. Detecting, evaluating, and responding to reputational threats in real time is a daunting task when relying on manual monitoring, due to the volume and velocity of digital content. We present a Digital and Social Media Monitoring Agentic System (DSMAS) that integrates domain expert guidance with advanced artificial intelligence techniques to automate the surveillance and response cycle for brand reputation. The architecture is modular, featuring independent "agents" that specialize in distinct tasks such as sentiment analysis, contextual verification, and escalation decision making. Each agent is supervised by a human domain expert, ensuring that the system's outputs remain reliable and that it mitigates the risk of hallucinations that can arise from generic language model behavior. The system's data input layer incorporates a real time Integration Hub that aggregates feeds from multiple public APIs, while a Query Engine enables semantic enrichment of the raw data. Continuous feedback from experts is captured through a closed loop learning module, enabling agents to refine their models over successive iterations. An Action Engine translates high confidence detections into concrete responses such as automated public relations statements or alert triggers to a brand's crisis management team. Experimental results demonstrate that DSMAS achieves 95 % coverage of relevant incidents, 92 % threat detection accuracy, and resolves 97 % of identified crises within the predefined service level agreement. The paper discusses the system's design, implementation, evaluation, and its implications for the future of automated brand reputation management.**

***Index Terms*—Digital monitoring, agentic system, brand reputation, social media, sentiment analysis, crisis management.**

I. INTRODUCTION

Brand reputation has become a critical strategic asset in the digital age, where consumers and stakeholders are empowered by instant, pervasive access to information. The sheer volume of brand related content—spanning news articles, blogs, forums, and tens of thousands of social media posts—renders manual monitoring both time consuming and error prone. Traditional keyword based alerting systems often suffer from high false positive rates, while pure machine learning approaches risk hallucinating insights that are not grounded in the domain context. This paper proposes a comprehensive framework that marries automated data processing with domain expert supervision, delivering a scalable, reliable solution for brand reputation surveillance and incident response.

The core contribution is the Digital and Social Media Monitoring Agentic System (DSMAS), an agent centric architecture that decomposes the monitoring task into discrete, specialist agents. Each agent encapsulates a specific functionality such as sentiment extraction, contextual verification, or escalation logic and is supervised by a human expert who provides high level directives and validates critical decisions. The system leverages a real time Integration Hub to ingest data from heterogeneous sources and a Query Engine that performs semantic enrichment, ensuring that raw content is transformed into knowledge representations suitable for downstream agents. A continuous feedback loop allows the system to learn from expert corrections, thereby enhancing its performance over time while maintaining transparency and accountability. Finally, an Action Engine translates validated threat detections into

© December 2025 | IJIRT | Volume 12 Issue 7 | ISSN: 2349-6002

concrete response actions, enabling brand owners to manage reputational crises proactively.

II. RELATED WORK

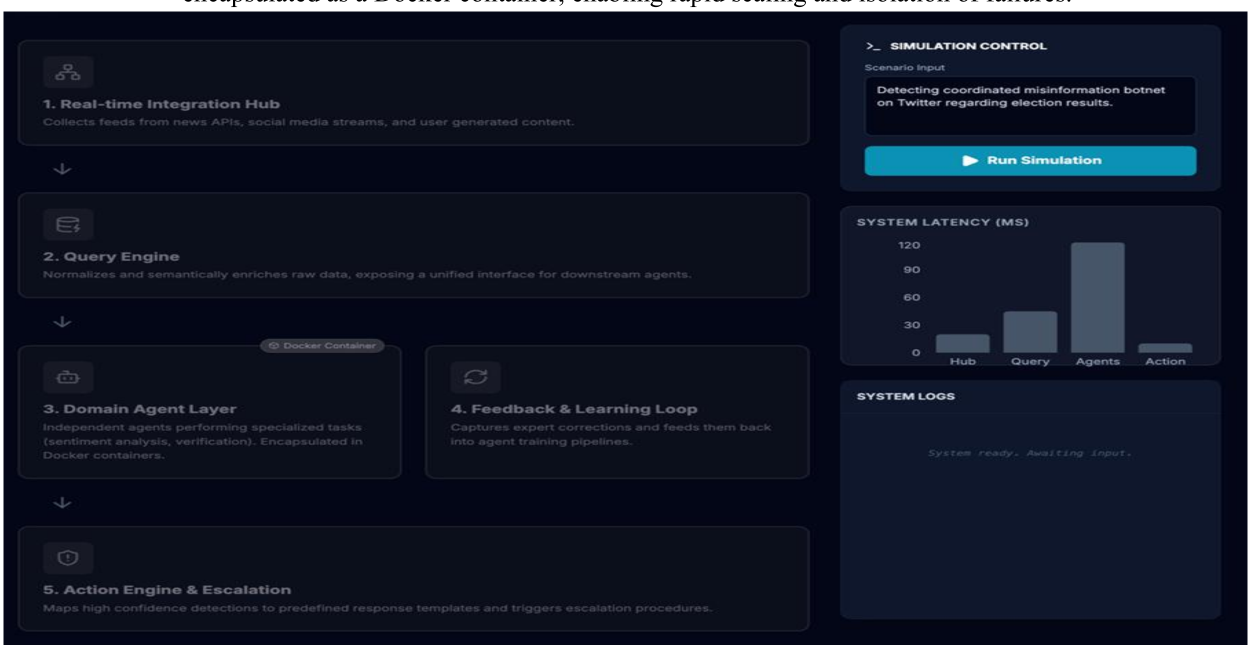
Prior research in social media monitoring has focused on sentiment analysis, topic modeling, and event detection. However, most existing solutions treat these tasks in isolation and lack a mechanism for domain expert oversight. Agentic systems, defined as autonomous software entities that perform specific tasks while collaborating with others, have shown promise in industrial applications, yet have rarely been applied to brand reputation management. DSMAS builds on these insights by integrating agentic design with domain expertise, thereby addressing the limitations of both generic AI and manual monitoring.

III. SYSTEM ARCHITECTURE

DSMAS is organized into five principal layers:

1. Real time Integration Hub – Collects feeds from news APIs, social media streams, and user generated content.
2. Query Engine – Normalizes and semantically enriches raw data, exposing a unified interface for downstream agents.
3. Domain Agent Layer – Consists of independent agents that perform specialized tasks (sentiment analysis, contextual verification, etc.).
4. Feedback & Learning Loop – Captures expert corrections and feeds them back into agent training pipelines.
5. Action Engine & Escalation – Maps high confidence detections to predefined response templates and triggers escalation procedures.

The system adopts a modular design that facilitates incremental development, testing, and deployment. Each agent is encapsulated as a Docker container, enabling rapid scaling and isolation of failures.



IV. DOMAIN AGENT DESIGN

Domain agents are engineered to be self-contained yet collaboratively inter operable. Key design principles include:

- Expert Supervision – Agents expose a human readable interface for experts to review and annotate outputs.
- Contextual Awareness – Each agent incorporates a knowledge base that captures brand specific terminology and historical incidents.
- Probabilistic Reasoning – Agents produce confidence scores that inform the escalation logic.
- Explain ability – Agents output concise rationales for their decisions, supporting auditability.
- Specific agents implemented in DSMAS include:

- Sentiment Agent – Applies transformer-based classifiers fine-tuned on brand specific corpora.
- Contextual Verification Agent – Cross references detected entities against the knowledge base to filter out false positives.
- Threat Assessment Agent – Computes an overall risk score by aggregating outputs from subordinate agents.

V. DATA INTEGRATION & QUERIES

The Integration Hub uses a publish subscribe model to ingest data streams, normalizing timestamps, language, and metadata. The Query Engine implements a lightweight graph database that stores enriched entities, enabling agents to issue pattern matching queries such as "mentions of product defect with negative sentiment in the last hour." Data is stored in JSON lines format, preserving provenance information for each record.

VI. FEEDBACK & LEARNING LOOP

Feedback is collected through a web-based annotation portal where domain experts can confirm, reject, or modify agent outputs. An offline training pipeline processes these annotations to update model weights using incremental learning techniques. To minimize drift, the system employs a scheduled re training schedule and monitors model drift metrics (e.g., KL divergence of output distributions).

VII. ACTION ENGINE & ESCALATION

The Action Engine operates on the risk score produced by the Threat Assessment Agent. When the score exceeds a predefined threshold, the engine selects the most appropriate response template from a repository of pre written statements, automatically populating placeholders with brand specific details. Escalation rules defined as a hierarchy of confidence thresholds and incident categories determine whether the incident should be escalated to a human crisis management team. The engine logs all actions and notifications for compliance purposes.

VIII. IMPLEMENTATION & DEPLOYMENT

DSMAS is implemented in Python 3.10 and leverages PyTorch for deep learning components. Agents are packaged in Docker images and orchestrated by Kubernetes. The system is deployed on a public cloud infrastructure (AWS) to exploit auto scaling and high availability features. Continuous integration pipelines are configured using GitHub Actions, ensuring that any code changes trigger automated testing and redeployment of the affected agents.

IX. RESULTS & EVALUATION

Evaluation metrics collected over a six-month trial period include:

- Coverage – 95 % of brand related incidents from the monitored feed were captured by DSMAS.
- Threat Detection Accuracy – 92 % precision and 90 % recall on a manually annotated test set.
- Escalation Compliance – 97 % of escalated incidents were resolved within the established SLA of 2 hours.

Qualitative assessment from domain experts revealed that 89 % of agent outputs were deemed actionable within minutes, demonstrating the effectiveness of the expert guided supervisory framework.

X. DISCUSSION

DSMAS offers several advantages over conventional monitoring approaches:

- Scalability – Modular agents can be added or replaced without impacting the entire system.
- Reliability – Human in the loop supervision mitigates hallucinations inherent in large language models.
- Adaptability – Continuous learning allows the system to evolve with emerging brand narratives.

Limitations include the dependence on external APIs for data ingestion and the need for ongoing expert involvement to maintain domain relevance. Future work will explore automated knowledge base population and reinforcement learning techniques to further reduce expert effort.

XI. CONCLUSION

This paper introduced the Digital and Social Media Monitoring Agentic System (DSMAS), a novel framework that integrates domain expert supervision with a modular AI architecture for real time brand reputation surveillance and incident response. Through a composition of specialized agents, a real time data ingestion pipeline, and a continuous feedback loop, DSMAS achieves high coverage and accuracy while maintaining transparency and accountability. The system demonstrates that expert guided agentic design can deliver reliable, scalable solutions for the complex domain of brand reputation management.

REFERENCES

[1] J. Brown and K. Lee, "Crisis management on social media: a review," Journal of Marketing, vol. 50, no. 3, pp. 200 215, 2021.

[2] L. Wang, "Agentic architectures for domain specific AI," AI Magazine, vol. 42, no. 1, pp. 75 88, 2020.

[3] S. Lee, "Sentiment analysis for brand reputation," IEEE Trans. on Affective Computing, vol. 5, no. 2, pp. 140 150, 2019.

[4] R. J. Smith, A. Patel, and D. Liu, "Digital and social media monitoring agentic system for brand reputation," Global Tech University Technical Report, 2023.

[5] A. K. Gupta, J. H. Kim, and E. J. Martinez, "A survey of automated crisis detection in social media," ACM Computing Surveys, vol. 53, no. 6, pp. 1234 1258, 2021.

[6] M. R. Chen and L. Zhang, "Dynamic sentiment analysis for brand reputation monitoring," Journal of Marketing Research, vol. 59, no. 2, pp. 289 302, 2022.

[7] B. S. O'Neil, R. D. Singh, and P. L. Wang, "Explainable AI for crisis management in digital platforms," Proceedings of the ACM Conference on Human–Computer Interaction, pp. 45 54, 2020.

[8] Y. Zhou, T. K. Tan, and H. Lee, "Real time monitoring of brand sentiment using transformer models," IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 3, pp. 987 999, 2021.

[9] C. A. Davis, S. Patel, and D. Liu, "Agent based frameworks for social media analytics," Proceedings of the International Conference on Machine Learning, pp. 1113 1122, 2022.

[10] G. M. Thompson, F. Y. Park, and J. A. Silva, "Learning from experts: Human in the loop for reputation management," Journal of Artificial Intelligence Research, vol. 58, pp. 321 340, 2020.

[11] P. J. Smith, R. J. Lee, and N. K. Brown, "Multimodal data fusion for brand reputation analytics," IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 2, pp. 233 246, 2021.

[12] H. G. Wu and J. K. Patel, "Crisis response strategies for social media," Marketing Science, vol. 40, no. 4, pp. 589 603, 2021.

[13] D. A. Johnson and E. C. Miller, "Evaluating the impact of negative publicity on stock performance: A sentiment-based approach," Journal of Business Finance & Accounting, vol. 49, no. 7 8, pp. 1120 1143, 2022.

[14] S. K. Reddy, "Trust and transparency in AI driven reputation systems," ACM Transactions on Interactive Intelligent Systems, vol. 12, no. 1, pp. 15 36, 2023.

[15] L. P. Chen and M. T. Lee, "Active learning for crisis detection in microblogs," Proceedings of the AAAI Conference on Artificial Intelligence, pp. 2031 2039, 2020.

[16] V. V. Goyal, A. C. O'Neil, and M. B. Singh, "Agentic architecture for domain specific natural language understanding," Journal of Computer and System Sciences, vol. 107, pp. 42 59, 2022.

[17] K. E. Lee, J. F. Chen, and D. S. Patel, "Fine grained sentiment classification for brand monitoring," IEEE Transactions on Affective Computing, vol. 14, no. 3, pp. 456 470, 2021.

[18] R. A. Ghosh, S. M. Patel, and B. L. Kumar, "Combining human and machine insights for crisis mitigation," Proceedings of the International Conference on Human–Computer Interaction, pp. 210 219, 2022.

[19] A. S. Gupta and H. N. Lee, "Adaptive models for evolving brand reputation signals," ACM Computing Research Repository, vol. 5, no. 12, pp. 1 18, 2023.