Interdisciplinary Systems for Global Management
Vol. 1 (2025), No. 1, pp. 52-58
DOI: https://doi.org/10.55578/isgm.2508.004

ARTICLE

Managing Data Saturation in AI Systems: A Cross-Domain Framework Integrating Human Insights and Algorithmic Verification

Michael Mncedisi Willie [1,*]

[1] *Council for Medical Schemes, Pretoria, South Africa*

[*]Corresponding author. Email: m.willie@medicalschemes.co.za

Received: 13 August 2025, Accepted: 20 August 2025, Published: 25 August 2025

Abstract

Artificial Intelligence (AI) systems are increasingly relied upon for real-time decision-making, yet high data volumes often introduce noise, inefficiencies, and performance degradation. Drawing inspiration from data saturation concepts in qualitative research, this study examined how excessive or unfiltered data compromises AI system effectiveness. The study aimed to develop a cross-domain conceptual framework that integrated human insights with algorithmic verification to manage data saturation and ensure timely, accurate, and trustworthy AI outputs. A conceptual and analytical approach was employed, combining an extensive literature review with case analyses of high-volume navigation platforms, including Waze and Google Maps. The study investigated adaptive filtering, contextual relevance scoring, hybrid verification, and temporal decay mechanisms to identify generalisable strategies for managing data saturation in AI systems. Findings show that adaptive filtering, prioritisation of contextually relevant information, hybrid verification combining human input and automated data, and temporal decay mechanisms can potentially mitigate data saturation. Lessons from Waze and Google Maps highlighted the importance of balancing human participation with computerised processes, enhancing system performance and user trust. Managing data saturation proves to be a critical operational challenge for AI systems, not merely a research consideration. The proposed framework provides a structured approach for integrating human insights and algorithmic verification, improving decision accuracy, responsiveness, and reliability across domains. The study recommends implementing adaptive data management strategies that combine human and algorithmic inputs, regularly evaluate system performance under high data loads, and adopt context-aware prioritisation techniques. Further research is suggested to explore empirical validation of the framework across diverse AI applications.

Keywords: Adaptive filtering; Contextual relevance; Crowdsourced verification; Data saturation; Feedback loops; Hybrid data integration; Real-time AI; Temporal decay

1. INTRODUCTION

Artificial Intelligence (AI) systems are inherently data-driven, relying on continuous information streams to refine decision-making processes and improve adaptability. In organisational contexts, AI-driven decision support systems have expanded managerial capabilities by enhancing analytic depth, enabling faster competitive responses, and supporting long-term strategic planning. While these benefits are notable, they are accompanied by concerns regarding transparency, accountability, and trust in automated decision-making [1].

The evolution of AI has also transformed information technology (IT) strategies, advancing business alignment, predictive analytics, operational efficiency, and sector-specific automation. However, this transformation is not without challenges. Ethical considerations, privacy concerns, and workforce adaptation issues persist, raising questions about the responsible integration of AI into complex socio-technical environments [2].

https://doi.org/10.55578/isgm.2508.004

A persistent assumption underpinning AI development is that larger volumes of data inevitably yield better performance. Empirical evidence and theoretical insights suggest this is not always true. Drawing parallels from qualitative research, data saturation, the point at which additional data no longer produces new insights, provides a compelling lens for examining AI performance limitations. In thematic analysis, reaching saturation strengthens the validity of findings [3]. Conversely, excessive or poorly curated data in AI systems can generate bottlenecks, increase noise, and degrade system outputs [4].

This paper advances the debate by introducing a novel cross-domain conceptual framework for managing data saturation in AI systems, a perspective not yet systematically articulated in current literature. Unlike existing approaches that primarily emphasise scaling computational power or refining algorithms, this framework draws inspiration from real-time navigation platforms. It introduces four integrative mechanisms: adaptive filtering, contextual relevance scoring, hybrid verification, and temporal decay to mitigate saturation while preserving performance integrity. Its originality lies in repositioning data saturation as a strategic operational challenge rather than a mere technical constraint, offering a transferable model that can guide AI researchers and practitioners in designing more sustainable, context-aware, and resilient AI systems.

2. LITERATURE REVIEW

2.1. Data Saturation and AI Performance

The Information Saturation Theorem posits that additional data introduces noise and complexity beyond a certain threshold, resulting in diminishing or even negative returns [5]. This phenomenon can compromise decision accuracy, slow processing, and increase computational costs within AI. While qualitative research uses saturation to signal an endpoint in data collection, AI faces a more dynamic and continuous form of the problem, especially in real-time applications. Scholars have highlighted that saturation is inconsistently defined and operationalised, leading to alternative frameworks such as information power and conceptual depth for determining optimal data collection limits [6-8]. These frameworks underscore the importance of balancing data richness with relevance; a balance AI must achieve to remain accurate and efficient.

2.2. Risks of Information Overload

Information overload occurs when the volume of incoming data surpasses the processing capacity of either human operators or automated systems. In AI-mediated communication and decision-making, over-digitalisation and pervasive ICT usage can amplify cognitive strain, affecting both human collaborators and machine systems [9,10]. These risks are magnified in high-stakes environments where delayed or inaccurate decisions can have severe consequences.

2.3. Real-Time Data Management in Practice

Real-time navigation platforms such as Waze and Google Maps exemplify how high-volume, dynamic data streams can be effectively managed. Waze's community-driven reporting system employs hybrid verification by cross-referencing user inputs with sensor data, reducing inaccuracies and building user trust [11,12]. Google Maps applies contextual relevance scoring and AI-driven updates to its expansive database, integrating multimodal data, including Street View imagery and user contributions, to maintain accuracy and timeliness [13].

Both platforms employ adaptive filtering mechanisms that continuously refine data processing strategies through feedback loops. These mechanisms help mitigate saturation risks and align information delivery with user context and intent. Outside such specialised domains, many AI systems rely on static filtering methods that lack the adaptability to manage evolving and complex data environments.

2.4. Generative AI and Emerging Challenges

Generative AI models like Google's Gemini illustrate the potential for responsive, context-aware data processing. Evaluations highlight Gemini's strengths in multimodal interaction, integration capabilities, and ethical safeguards [14,15]. However, even advanced models remain vulnerable to saturation effects if input data is not carefully filtered and prioritised. This study suggests that an integrated approach drawing from best practices in navigation systems can offer a pathway to mitigating data saturation in broader AI applications.

3. METHODS

This study employed a conceptual and analytical approach to investigate data saturation in AI input feeds, drawing lessons from high-data platforms such as Waze, Google Maps, and Google Search. As Furner [16] highlighted, conceptual analysis provides a structured lens for understanding complex archival concepts, clarifying the meaning of terms like "evidence," identifying necessary conditions for evidentiary validity, and revealing parallels between archival science and social epistemology. Building on this, Naeem et al. [3] demonstrated that systematic thematic analysis enables the development of conceptual models from qualitative data while enhancing methodological rigour, transparency, and replicability across diverse qualitative traditions.

The choice of Waze and Google Maps as primary cases is deliberate, as these platforms epitomise real-time, high-velocity, and high-volume AI systems. They continuously integrate geospatial, temporal, and crowdsourced behavioural data, making them ideal exemplars for examining saturation management. Unlike many AI applications that process relatively static or pre-curated datasets, navigation systems must adapt dynamically to shifting conditions such as traffic congestion, accidents, and road closures. Their reliance on adaptive filtering, contextual relevance scoring, and hybrid verification creates a microcosm of the broader challenges data-intensive AI environments face.

While Waze and Google Maps do not represent all high-data AI systems, they provide an accessible and analytically rich case for deriving generalisable principles. Other domains such as social media moderation, where platforms filter billions of posts daily; predictive maintenance in industrial IoT, where sensor streams must be contextualised and verified in real time; and healthcare diagnostics, where data overload can obscure clinically relevant patterns face parallel saturation challenges. However, these domains often operate under additional ethical or domain-specific constraints (e.g., privacy in healthcare, bias in content moderation).

3.1. Lessons from Waze, Google Maps, and Google

3.1.1. Adaptive Filtering in Data-Intensive Navigation Systems

High-volume navigation platforms like Waze and Google Maps illustrate the practical application of adaptive filtering strategies in environments saturated with user-generated data. These platforms must continuously manage vast crowdsourced inputs and prioritise relevant information while minimising noise. Empirical evidence highlights the efficiency of such filtering mechanisms: Gu et al. [17] found that Waze users reported crashes an average of 2.2 minutes earlier than official state logs, with location accuracy within six feet. This demonstrates the potential for crowdsourced systems to achieve both timeliness and precision, underscoring the importance of adaptive data handling in real-time applications.

3.1.2. Crowdsourced Data Integrity and Verification

While crowdsourced inputs are valuable, their reliability requires rigorous verification. Hybrid mechanisms that cross-check user reports against automated sensor data are critical for maintaining accuracy. Fan, Liu, Wang, and Liu 18] demonstrated that integrating crowdsourced and sensor-driven data enhances real-time traffic awareness and route-planning efficiency. Kim, Jeon, and Kim [19] further show that combining crowdsourced feedback with historical traffic and accident data enables adaptive identification of high-risk road segments, improving road safety. These findings underscore a broader lesson for AI and data-driven systems: combining human insight with automated validation strengthens reliability and scalability.

3.1.3. User Behaviour and System Design Influence

The design of navigation systems strongly shapes user trust and engagement. Trapsilawati, Wijayanto, and Septiawan Jourdy [11] observed that Waze's flexible information-sharing features enhanced user confidence, prompting some Google Maps users to switch platforms after experiencing Waze's advantages. Laor and Galily [20] further note that Waze's functional utility can induce behaviours akin to technological dependency, including mood modification, conflict, relapse, and withdrawal. These findings suggest that system design facilitates navigation and actively shapes behavioural patterns, implying that user experience considerations are critical to sustaining platform adoption and loyalty.

3.1.4. Prioritisation of Information for Contextual Relevance

Waze's dynamic filtering prioritises reports most relevant to a driver's immediate route, such as nearby traffic congestion, while deprioritising distant or less critical events. This selective presentation of information enhances usability and engagement, guiding drivers toward optimal decisions without overwhelming them. In contrast, Google Maps relies more heavily on local traffic patterns to optimise predictive routing. Research by Khoo and Asitha [21] highlights that navigation applications aligned with user needs through prioritised, actionable information are more likely to influence route choice and encourage long-term use. This demonstrates that adaptive information prioritisation is central to system efficiency and user behavioural outcomes.

3.2. Implications for AI and System Design

Integrating human-sourced insights with algorithmic verification provides a model for adaptive, trustworthy AI system design. Crowdsourced navigation platforms illustrate that effective performance depends on data volume and strategic filtering, prioritisation, and validation. Lessons from Waze and Google Maps reveal that timeliness, contextual relevance, user experience, and data integrity must be jointly considered to optimise system outcomes. This approach to system design highlights the importance of balancing human participation and automated processing, offering key insights for broader applications in AI and real-time decision support environments.

The proposed framework (**Figure 1**) integrates human-sourced insights with algorithmic verification to create adaptive and trustworthy AI systems, emphasising the balance between human participation and automated processing. Drawing lessons from crowdsourced navigation platforms like Waze and Google Maps, it incorporates adaptive filtering, contextual relevance scoring, hybrid verification, and temporal decay to manage data saturation while ensuring timeliness, accuracy, and overall system performance.

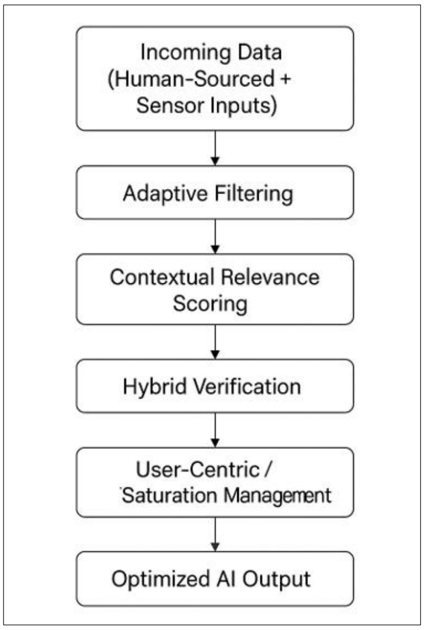

Figure 1. A Cross-Domain Framework for Balancing Human Insights and Algorithmic Verification in AI Systems

4. CONCLUSION

This study highlighted that managing data saturation is a critical operational challenge for AI systems, extending beyond its traditional role as a qualitative research concern. Drawing lessons from high-volume navigation platforms such as Waze and Google Maps, the study demonstrated that integrating human insights with algorithmic verification through adaptive filtering, contextual relevance scoring, hybrid verification, and temporal decay effectively mitigates the risks of excessive or uncurated data. The proposed cross-domain framework offers a structured approach for balancing human participation and automated processing, enhancing system accuracy, timeliness, and trustworthiness.

For practitioners, several actionable recommendations emerge. First, adaptive filtering mechanisms that evolve with contextual shifts should be designed, rather than relying on static data rules. This requires embedding feedback loops that continuously recalibrate thresholds for relevance as operational environments change. Second, contextual relevance scoring should be incorporated to prioritise information that aligns with domain-specific objectives, ensuring that data noise does not obscure critical insights. Third, algorithmic outputs should be combined with human-in-the-loop verification, particularly in domains where ethical considerations, safety, or public trust are paramount. This hybrid approach leverages the efficiency of automation while retaining the interpretive depth of human judgment. Fourth, temporal decay functions are applied to data streams, ensuring that outdated or redundant inputs are systematically phased out, thereby preserving system agility and responsiveness.

Finally, practitioners should view data saturation management as a strategic capability rather than a technical afterthought. This involves institutionalising governance structures for data curation, investing in explainable AI tools that improve accountability, and training teams to critically evaluate the quantity and quality of incoming data. By operationalising these recommendations, organisations can build AI systems that are not only scalable but also sustainable, trustworthy, and resilient across diverse application domains.

Acknowledgements and Use of AI: While preparing this manuscript, the author employed AI-assisted tools, including Grammarly and QuillBot, to support language refinement, editing, and plagiarism checking. All outputs generated by these tools were carefully reviewed and revised by the author, who assumes full responsibility for the accuracy and integrity of the final content presented in this publication.

Funding Statement: This research and manuscript preparation were conducted without external funding or financial support.

Author Contributions: The author solely contributed to the conception, research, writing, and editing of this manuscript.

Informed Consent Statement: Not applicable.

Ethical Approval: This study did not involve human participants or animal subjects and thus did not require ethical approval.

Data Availability Statement: This study exclusively used data obtained from secondary sources through a comprehensive literature review. All referenced data are publicly accessible and have been appropriately cited within the manuscript.

Conflict of Interest: The authors declare no conflict of interest.

REFERENCES

1. Dodda, S., Narne, S., Adedoja, T., & Ayyalasomayajula, M. M. T. (2024). AI-driven decision support systems in management: Enhancing strategic planning and execution. *International Journal on Recent and Innovation Trends in Computing and Communication, 12*(1), 1.
2. Ugwa, S. (2023). Artificial intelligence for IT strategy: Driving data-driven decision-making and business alignment. *World Journal of Advanced Research and Reviews, 18*(2), 1455–1474. https://doi.org/10.30574/wjarr.2023.18.2.0922

3. Naeem, M., Ozuem, W., & Ranfagni, S. (2024). Demystification and actualisation of data saturation in qualitative research through thematic analysis. *International Journal of Qualitative Methods.* Advance online publication. https://doi.org/10.1177/16094069241229777
4. Walton, P. (2018). Artificial intelligence and the limitations of information. *Information, 9*(12), 332. https://doi.org/10.3390/info9120332
5. Nelson, J. (2016). Using conceptual depth criteria: Addressing the challenge of reaching saturation in qualitative research. *Qualitative Research, 17*(5), 554–570. https://doi.org/10.1177/1468794116679873
6. Sims, I., & Cilliers, F. (2023). Qualitatively speaking: Deciding how much data and analysis is enough. *African Journal of Health Professions Education, 15*(1). https://doi.org/10.7196/AJHPE.2023.v15i1.1657
7. Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2017). Saturation in qualitative research: Exploring its conceptualisation and operationalisation. *Qualitative & Quantitative Research, 52*(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8
8. Tight, M. (2023). Saturation: An overworked and misunderstood concept? *Qualitative Inquiry, 30*(7). https://doi.org/10.1177/10778004231183948
9. Moko, A. B., Victor-Ikoh, M., & Okardi, B. (2023). Information overload: A conceptual model. *European Journal of Computer Science and Information Technology, 11*(5), 19–29. https://doi.org/10.37745/ejcsit.2013/vol11n51929
10. Arnold, M., Goldschmitt, M., & Rigotti, T. (2023). Dealing with information overload: A comprehensive review. *Frontiers in Psychology, 14*, 1122200. https://doi.org/10.3389/fpsyg.2023.1122200
11. Trapsilawati, F., Wijayanto, T., & Septiawan Jourdy, E. (2019). Human-computer trust in navigation systems: Google Maps vs Waze. *Communications in Science and Technology, 4*(1), 38–43. https://doi.org/10.21924/cst.4.1.2019.112
12. Amin-Naseri, M., Chakraborty, P., Sharma, A., & Hong, M. (2018). Evaluating the reliability, coverage, and added value of crowdsourced traffic incident reports from Waze. *Transportation Research Record: Journal of the Transportation Research Board, 2672*(4), 1–12. https://doi.org/10.1177/0361198118790619
13. Pericolosi, B. (2022, December 14). How Google Maps Platform uses AI/ML and community contributions to keep its points of interest up to date. *Google Maps Platform Blog.* https://www.blog.google/products/maps/how-google-maps-platform-uses-ai-ml-and-community-contributions/
14. McIntosh, T. R., Susnjak, T., Liu, T., Watters, P., Xu, D., Liu, D., & Halgamuge, M. N. (2025). From Google Gemini to OpenAI Q (Q-Star): A survey on reshaping the generative artificial intelligence (AI) research landscape. *Technologies, 13*(2), 51. https://doi.org/10.3390/technologies13020051
15. Akhtar, Z. B. (2024). From Bard to Gemini: An investigative exploration journey through Google's evolution in conversational AI and generative AI. *Computing and Artificial Intelligence, 2*(1), 1378. https://doi.org/10.59400/cai.v2i1.1378
16. Furner, J. (2004). Conceptual analysis: A method for understanding information as evidence, and evidence as information. *Archival Science, 4*(3–4), 233–265. https://doi.org/10.1007/s10502-005-2594-8
17. Gu, Y., Zhang, H., Brakewood, C., & Han, L. D. (2022). WAZE data reporting (Final Report No. RES2019-11). Center for Transportation Research, The University of Tennessee. https://rosap.ntl.bts.gov/view/dot/66932
18. Fan, X., Liu, J., Wang, Z., Jiang, Y., & Liu, X. (2017). Crowdsourced road navigation: Concept, design, and implementation. IEEE Communications Magazine, 55(6), 126–128. https://doi.org/10.1109/MCOM.2017.1600738
19. Kim, J., Jeon, W., & Kim, S. (2024). Analysing the relationship between user feedback and traffic accidents through crowdsourced data. Sustainability, 16(22), 9867. https://doi.org/10.3390/su16229867

20. Laor, T., & Galily, Y. (2022). In WAZE we trust? GPS-based navigation application users' behavior and patterns of dependency. *PLoS ONE, 17*(11), e0276449. https://doi.org/10.1371/journal.pone.0276449
21. Khoo, H. L., & Asitha, K. S. (2015). User requirements and route choice response to smartphone traffic applications (apps). *Travel Behaviour and Society, 3*(3), 87–100. https://doi.org/10.1016/j.tbs.2015.08.004