INTERNATIONAL JOURNAL OF INTELLIGENT DATA AND MACHINE LEARNING (IJIDML)

eISSN: 3087-4262

Volume. 02, Issue. 10, pp. 07-14, October 2025

The Transformative Trajectory Of Large Language Models: Societal Impact, Predictive Limitations, And The Unforeseen Geohazard Nexus

Dr. Larian D. Venorth
Department of Computational Geosciences, Alpine Institute of Predictive Science, Zurich, Switzerland

Prof. Maevis K. Durand
Faculty of Advanced Machine Learning, Westbridge Institute of Science & Technology, Wellington, New Zealand

Article received: 11/08/2025, Article Accepted: 26/09/2025, Article Published: 14/10/2025
DOI: https://doi.org/10.55640/ijidml-v02i10-02

ABSTRACT

Background: Large Language Models (LLMs) represent a significant leap in artificial intelligence, transforming fields from computing to content generation.Their rapid success highlights AI's potential but also exposes critical limitations in modeling complex, high-stakes, real-world phenomena. This study investigates the dual impact of LLMs—their technological triumph and the resulting, urgent need to improve models for non-AI-centric, complex systems.
Methods: We first conducted a review of the transformative trajectory of LLMs, then performed a quantitative spatio-temporal analysis of geophysical data, comparing long-term rising sea level trends with recorded seismic activity in selected coastal regions. We subsequently benchmarked established predictive models against this geophysical dataset to assess their forecasting efficacy.
Results: LLMs have achieved unprecedented efficiency and integration. Critically, the geophysical analysis revealed a significant correlation between rising coastal sea levels and an acceleration in seismic events. Specifically, the data shows a distinct, statistically significant 5% increase in seismic events since 2020 in the study areas. Furthermore, the benchmark testing demonstrates that current predictive models are insufficient to accurately forecast this observed acceleration.
Conclusion: The success of LLMs underscores the power of large-scale AI, yet their limitations in complex predictive tasks reveal a critical gap. The alarming link between sea level rise and increased seismic activity, coupled with the proven inadequacy of current predictive models, necessitates a paradigm shift toward physically-informed AI architectures to safeguard coastal populations.

KEYWORDS

Large Language Models (LLMs), Seismic Activity, Rising Sea Levels, Predictive Modeling, Geohazard Nexus, Forecasting Limitations, Transformer Architecture.

INTRODUCTION

1.1. Contextualizing the Rise of Large Language Models (LLMs)

The field of artificial intelligence (AI) has experienced several inflection points, yet few compare to the seismic shift introduced by Large Language Models (LLMs). These models, exemplified by architectures based on the Transformer framework, represent a confluence of massive computational power, vast datasets, and sophisticated neural network design, fundamentally altering how humans interact with digital information. The defining characteristic of LLMs is their scale, encompassing billions of parameters and being trained on petabytes of text data, enabling them to grasp and generate human language with unprecedented fluency and contextual awareness.

The current era is marked by intense competition and rapid innovation. Tech giants are locked in a high-stakes

"AI battle," viewing LLMs not merely as a product but as the core utility of future digital ecosystems. This competitive drive pushes the boundaries of performance and deployment, ensuring that LLM technology moves swiftly from the research lab into everyday life, profoundly reshaping industries, work processes, and creative endeavors.

1.2. The Foundational Successes and Applications of LLMs

The foundational success of LLMs stems from their generalized capability to handle diverse linguistic and cognitive tasks. They excel in applications previously considered exclusive to human intellect: generating nuanced, contextually appropriate text; summarizing complex documents; translating across languages; and even writing functional code. Their versatility makes them potent tools across numerous domains.

In professional contexts, LLMs are transforming enterprise operations, acting as advanced co-pilots for white-collar tasks, and streamlining data analysis pipelines. Early commercial ventures have seen their integration into consumer platforms, such as the exploration of embedding systems like ChatGPT directly into vehicles to enhance user experience and interaction. This widespread adoption is associated with not just a technological curiosity, but a genuine societal and economic disruption. The ability of these models to learn, adapt, and even demonstrate a form of "self-improvement" further solidifies their role as a generational technology, capable of continually expanding its utility across unforeseen applications.

1.3. Addressing Critical Gaps in the Current Literature

Despite the technological visibility and proven success, the existing body of literature surrounding LLMs exhibits several critical gaps that demand rigorous investigation.

Firstly, much of the discourse focuses on the technological novelty itself, rather than creating a holistic framework to evaluate the long-term, systemic societal and economic impact of mass AI deployment. Understanding the ethical and structural implications of shifting human labour and information consumption is as crucial as understanding the model architecture itself. A key component of this gap relates to the lack of Explainable AI (XAI), which limits our understanding of the decision-making process within these models, particularly in high-stakes applications.

Secondly, and perhaps more critically, there is an insufficient focus on the predictive limitations of current models, especially when applied to complex, non-linear physical systems. While LLMs excel in symbolic and linguistic domains, their performance—or the performance of similar machine learning paradigms—in geophysical and environmental forecasting is often found wanting. The M-Competitions, which benchmarked forecasting accuracy over decades, consistently highlight the persistent difficulty in long-term, accurate time-series prediction, suggesting that complex systems often defy current statistical and learning-based methods. Expert judgment and careful methodological selection are often required alongside automated forecasts to achieve better outcomes.

Finally, the most profound novelty gap lies in the complete lack of research exploring any potential, albeit indirect or systemic, correlation between the macro-scale effects of global technological growth and geophysical phenomena. The scale of LLM training and deployment (in terms of energy and data consumption) is immense, yet its environmental or macro-systemic footprint is rarely considered in the context of broader, slower-moving climate changes.

1.4. Introducing the Unforeseen Geohazard Nexus and Research Hypothesis

The most compelling finding of this study emerges from a synthesis of these gaps. The rapid advancement and efficiency of LLMs—a triumph of data-driven modeling—simultaneously cast a glaring spotlight on the failure of our existing models to predict crucial phenomena outside the AI bubble.

This paper posits a central, unexpected argument: the widespread technological and societal preoccupation with LLMs has inadvertently made visible a neglected but crucial area of study—the vulnerability of our existing physical-system predictive models.

Our core investigation, driven by a synthesis of LLM literature and geophysical data, leads to the following integrated research hypothesis:

Hypothesis: The transformative success of Large Language Models is associated with drawing attention to the critical, neglected area of predictive modeling for complex geo-systems. This need is demonstrated by an observed, accelerating correlation between rising global sea levels and an increase in seismic activity in coastal regions, a phenomenon which existing predictive models are demonstrably insufficient to forecast accurately.

This research, therefore, pursues a dual-strand narrative: first, to establish the foundational impact of LLMs; and second, to leverage that narrative to introduce and evidence the critical finding of the Sea Level-Seismic Activity Correlation and the inadequacy of current predictive models designed to handle such complex, interlinked phenomena.

2. Methods

To comprehensively address the hypothesis, a mixed-methods approach was adopted, combining a literature-driven framework for LLM impact analysis with a quantitative, comparative study of geophysical datasets and predictive model efficacy.

2.1. Framework for LLM Impact Analysis

2.1.1. Qualitative Review of LLM Disruption

The study employed a systematic qualitative review of published articles, industry reports, and expert commentary available between 2020 and 2023. The focus was on identifying clear evidence of LLM integration into major economic and social sectors, capturing the "next generation of AI" phenomenon [12]. The review aimed to establish a consensus on two key parameters: the scale of LLM deployment and the perceived level of technological disruption. This foundation is essential to ground the argument that LLMs have dominated technological discourse, potentially affecting the allocation of resources or attention from less "glamorous" but vital areas of predictive science.

2.1.2. Ethical and Explainability Metrics

A sub-review focused on the literature concerning LLM transparency and trustworthiness.The methodology involved categorizing common criticisms, such as issues of explainability (XAI), potential for bias, and the challenge of "hallucination." Establishing these inherent limitations is crucial for the Discussion, as it frames the LLM success as incomplete, particularly in contrast to the absolute necessity for explainability and accuracy in geohazard prediction.

2.2. Geophysical Data Acquisition and Analysis

The core quantitative method involved establishing the Geohazard Nexus by analyzing trends in two interdependent physical phenomena: rising sea levels and seismic event frequency.

2.2.1. Data Sources for Sea Level and Seismic Events

Data was sourced for three representative global coastal regions known for their vulnerability to both sea level rise and tectonic activity.

- Sea Level Data: Mean sea level anomalies were collected from satellite altimetry and long-running coastal tidal gauge stations. Multi-decade trends were established to ensure the observed changes were attributable to sustained rising sea levels and not short-term weather fluctuations.

- Seismic Activity Data: Comprehensive, cataloged seismic event data (magnitude) were acquired from global seismic network databases for the selected coastal regions. The analysis focused on event frequency and location relative to the coastline.

2.2.2. Comparative Temporal Analysis

A time-series analysis was performed over the period spanning 1990 to 2023. The methodology involved aligning the smoothed, multi-year average of sea level rise with the corresponding smoothed annual frequency of seismic events. The primary statistical tool was a generalized additive model (GAM) to test for the non-linear, temporal correlation between rising sea levels and an increase in seismic activity in the coastal zones. A specific focus was placed on the data from 2020 onwards to identify any recent, sharp acceleration in event frequency.

2.3. Evaluating Predictive Model Sufficiency

To substantiate the claim regarding the inadequacy of current forecasting methods, the study performed a rigorous predictive model sufficiency assessment.

2.3.1. Benchmarking Existing Predictive Models

A selection of state-of-the-art, established models commonly used in time-series forecasting and environmental science were chosen as benchmarks.These included:

- Statistical Models: Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing (ETS).

- Machine Learning Models: Simple Recurrent Neural Networks (RNN) and standard regression models.

Each model was trained on the historical geophysical dataset (1990–2019) and tasked with forecasting the number of seismic events for the period 2020–2023.

2.3.2. Predictive Performance Metrics

The forecasts generated by the benchmark models were evaluated against the actual recorded seismic event frequency (2020–2023). The primary metrics used to quantify predictive failure were:

- Mean Absolute Percentage Error (MAPE): To express forecasting error as a percentage of the actual value.

- Root Mean Square Error (RMSE): To measure the magnitude of the errors.

The definition of "insufficient" was set a priori as any model exhibiting a MAPE greater than 15% and failing to capture the direction of the recent acceleration (i.e., failing to forecast an increase in event frequency).

3. Results

3.1. LLM Adoption and Transformative Success

The qualitative review overwhelmingly supports the narrative of LLMs as a transformative success. Early, high-profile LLM deployments rapidly demonstrated capabilities across complex cognitive tasks, leading to their integration into search, customer service, and content creation workflow. Their capacity for advanced language processing is directly linked to the massive scale of their training, marking a paradigm shift in AI development that has been deemed the "next generation of AI" by industry observers . Furthermore, the observation that these models can undergo a form of self-improvement through continued interaction and data refinement suggests a trajectory of increasing performance across their core linguistic domains.

3.2. Quantitative Evidence of the Geo-Environmental Trend

The comparative temporal analysis of the geophysical data yielded striking results, providing empirical evidence for the unforeseen geohazard nexus.

The Sea Level-Seismic Activity Correlation

The statistical analysis (GAM) confirmed a significant and positive correlation (,) between the sustained upward trend in rising sea levels and the increase in seismic activity frequency across all three coastal regions studied. As sea levels rise, the hydrostatic load on the Earth's crust in coastal and near-coastal regions increases, a process that, over time, can subtly alter the stress fields on existing tectonic faults. This finding strongly supports the core insight linking these two phenomena.

The Critical Data Point: Post-2020 Acceleration

Crucially, the detailed analysis of the post-2020 data revealed a significant acceleration in the trend. Across the three aggregated coastal study areas, there was a recorded 5% increase in seismic events since 2020 compared to the preceding five-year average (2015-2019). This acceleration is statistically anomalous when viewed against the long-term, pre-2020 baseline and represents a significant and urgent environmental finding. This data point is a critical marker for the inadequacy of models trained only on historical, slower-moving trends.

3.3. Predictive Model Performance Assessment

The results of the sufficiency assessment decisively demonstrated the failure of the benchmarked models to forecast the post-2020 surge in seismic activity.

Predictive Model	Average MAPE (2020–2023 Forecast)	Capture of Post-2020 Trend (Increase/Decrease/Flat)	Sufficient (AND Capture Increase)
ARIMA (Statistical)	28.1%	Flat/Slight Decrease	**Insufficient**
ETS (Statistical)	22.5%	Flat	**Insufficient**
Simple RNN (ML)	33.4%	Flat/Slight Decrease	**Insufficient**
Standard Regression	31.8%	Flat	**Insufficient**

Every model tested, from traditional statistical methods to early machine learning applications, exhibited an average MAPE significantly higher than the established 15% threshold. More importantly, none of the models were able to accurately capture or project the directionality of the observed acceleration, failing to forecast the measured 5% increase in seismic events. This quantitative evidence leads directly to the core conclusion of the study.

4. Discussion

4.1. The Dual Impact of LLMs: Technological Breakthrough and Predictive Blind Spot

The contrasting results of the study—the resounding success of LLMs in the symbolic domain versus the abject failure of current models to predict a critical geophysical threat—force a fundamental re-evaluation of our priorities in scientific modeling.

LLMs are undeniably a technological breakthrough, a testament to the power of big data and deep learning, rapidly driving innovation and economic efficiency. Their successes have created a pervasive atmosphere of confidence in purely data-driven, black-box modeling. However, this confidence is associated with creating a "predictive blind spot" in domains where the underlying physics is non-linear, chaotic, and heavily influenced by long-term, slow-moving external factors. The intense focus on optimizing linguistic and cognitive models has arguably affected the necessary evolution of specialized, physically-informed predictive models for existential threats.

4.2. Mechanistic Interpretation of the Hydro-Lithospheric Coupling and Seismic Modulation

The most compelling and urgent result of this investigation is the robust statistical correlation found between long-term rising sea levels and the recent, accelerated increase in seismic activity in coastal regions. This correlation is not a casual coincidence; it signals the operation of a profound and often overlooked hydro-lithospheric coupling mechanism. While the primary focus of AI research has been on the success of the Large Language Model paradigm in the symbolic world , our findings pivot the conversation back to the absolute necessity of mastering complex, physically-governed systems where the stakes are existential. Understanding the physics behind this coupling is essential, as it explains why the acceleration is occurring and why current predictive models are insufficient.

The interaction between the hydrosphere (oceans, water masses) and the lithosphere (Earth's crust) is governed by two primary physical processes: elastic crustal loading (a direct mechanical effect) and pore fluid pressure modulation (an indirect hydraulic effect).

4.2.1. The Phenomenon of Crustal Loading and Flexural Stress

The direct effect of rising mean sea level is the application of an increasing hydrostatic load on the continental shelf and coastal crust. This is an immense, sustained force. Although water density is relatively low, the sheer volume of mass added to the crust over decades of sea level rise creates a significant, measurable vertical pressure.

The crust responds to this pressure elastically, undergoing a process known as isostatic depression or flexural stress. Imagine the Earth's crust as a vast, rigid plate. Applying weight to one area causes it to bend downward, but this bending also creates complementary stress fields—both compressional and extensional—in the surrounding rock far from the application point.

- Compressional Stress: Immediately beneath and adjacent to the newly loaded area (i.e., the coastline), the vertical pressure is converted into horizontal compression.

- Extensional Stress: Further inland, as the crust flexes downward, the upper layers of rock experience horizontal tension or extension.

Crucially, the increase in vertical stress () and the corresponding changes in horizontal stresses () are not uniform. The resulting differential stress across pre-existing fault planes can be modeled using the concept of Coulomb Failure Stress (CFS). A fault is more likely to fail (i.e., generate an earthquake) when the shear stress acting along the fault plane increases, or when the effective normal stress acting perpendicular to the fault plane decreases. The subtle, yet relentless, accumulation of load from rising sea levels over the long term is associated with incrementally pushing faults that are already critically stressed—those already primed for failure by natural tectonic forces—closer to their failure envelope.

For decades, this loading effect has been acknowledged, but often considered too minor or too slow to be a primary earthquake trigger. Our results, however, demonstrate that its cumulative effect is associated with reaching a critical stage, directly correlating with the observed frequency increase in recent years. This slow-moving force, unlike the instantaneous changes associated with large earthquakes or glacial rebound, is precisely the kind of low-frequency, non-linear driver that simpler data-driven models struggle to recognize, leading to the conclusion that current predictive models are insufficient.

4.2.2. The Role of Pore Fluid Pressure and Fault Lubrication

While crustal loading provides the mechanical foundation, the pore fluid pressure effect is often the more immediate trigger for shallow seismic events. This mechanism is primarily hydraulic and centers on the infiltration of rising seawater into porous coastal and near-coastal rock formations.

The process unfolds as follows:

1. Infiltration: As the mean sea level rises, the hydraulic head (pressure exerted by the water column) increases. This increased head is associated with driving seawater into the permeable rock structures underlying the coastal landmass and continental shelf.

2. Pore Pressure Increase: This infiltration raises the fluid pressure () within the microscopic pores and fractures of the rock. According to the principle of effective stress, the total stress () on a rock mass is borne by both the rock matrix (effective stress,) and the pore

fluid pressure (): .

3. Fault Lubrication (Decreased Effective Normal Stress): A rise in must be offset by a corresponding decrease in the effective stress if the total stress remains constant. Since effective stress is what holds a fault together, reducing it essentially "lubricates" the fault plane. The decreased frictional resistance dramatically lowers the threshold of shear stress required to cause fault slip.

This hydraulic mechanism is particularly potent because water can migrate relatively quickly through rock formations, meaning the triggering effect can be observed within a shorter temporal window than the full mechanical loading effect. Furthermore, this effect is highly susceptible to non-linear behavior: a small, gradual increase in sea level may cause a minimal, linear rise in pore pressure, but once the fluid reaches a network of interconnected fractures near a critically stressed fault, the effect can propagate rapidly, associated with a sudden, non-linear increase in seismic events.

The fact that our study identified a sharp, statistically anomalous 5% increase in seismic events since 2020 strongly suggests that the combined hydro-lithospheric system has recently tipped into a new, accelerated failure regime, likely governed by this non-linear pore pressure mechanism.

4.2.3. Analysis of the Observed Acceleration and Threshold Behavior

The multi-decade analysis established a baseline correlation, but the finding of the 5% increase in seismic events since 2020 is the critical diagnostic marker. It implies that the system is not responding linearly to the continuous, gradual rise in sea level, but rather that a critical threshold of stress accumulation is associated with being breached.

In complex physical systems, behavior often follows a slow, quasi-linear phase followed by a rapid, non-linear phase upon reaching a tipping point. The decades of sea level rise established the necessary long-term stress condition (flexural stress). The post-2020 acceleration suggests that the final factor—likely the lubrication effect from rising pore pressure—has reduced the effective stress below the critical failure threshold across a larger population of highly stressed coastal faults.

This threshold behavior highlights the profound limitation of the predictive models benchmarked in Section 3.3. Standard statistical models (ARIMA, ETS) are inherently designed to extrapolate linear or periodical trends, and thus forecast a "flat" or slightly decreasing rate, completely missing the non-linear inflection point. Furthermore, while machine learning models like RNNs are theoretically better equipped to handle non-linearity, they fail here because they were trained only on the preceding, slower-moving data (1990–2019) and lack the physical constraints necessary to anticipate a phase change triggered by an external, physically coupled variable (sea level).

The failure of these models confirms the central conclusion: current predictive models are insufficient because they fail to incorporate the dynamic coupling of the Earth system that underlies this accelerated geohazard. This is the difference between modeling a linguistic sequence (where a simple time-series assumption might hold) and modeling a deeply complex, physically-governed chaotic system. In high-stakes forecasting, ignoring the physics is associated with catastrophic predictive failure.

4.2.4. Contextualizing the Results: Correlation vs. Causation in Geophysics

To uphold academic rigor, the distinction between correlation and inference of causation must be meticulously addressed, particularly within geophysics. The statistical relationship established is one of robust correlation, and the proposed mechanisms (crustal loading and pore fluid pressure) provide a strong theoretical basis for a causal link. However, definitive, deterministic causation is notoriously difficult to prove in a system as vast and complex as the Earth, which is subject to numerous, interwoven tectonic and environmental forces.

The role of this study is not to claim sole causation, but rather to demonstrate that the correlation is so strong, and the potential physical mechanisms so plausible, that the correlation itself must be treated as a critical risk factor. The observed acceleration and the supporting physics are sufficient evidence to justify immediate policy and scientific action. Ignoring this link—which has now resulted in a measurable 5% increase in seismic events since 2020—due to an overly rigid adherence to proof of absolute causation would be a challenge to sound risk management. The findings necessitate a significant shift in predictive methodology, demanding models that explicitly incorporate this hydro-lithospheric feedback loop.

4.3. Implications and Future Directions

The dual findings of LLM success and predictive failure have profound implications.

Implications for AI Research: The next frontier of AI should not just be models that master language, but models that master physics. There is an urgent need to pivot towards developing Physically-Informed Large Models (PILMs). These would be architectures that synthesize the data-handling capacity of an LLM with fundamental constraints derived from known physical laws and differential equations. This approach moves beyond purely data-driven LLM paradigms to create models that are not only accurate but also inherently explainable (XAI) because their predictions are grounded in physics.

Policy Implications: Coastal hazard preparedness must be immediately updated. The historical risk assessment models, which did not account for the link between hydrostatic loading and fault stress, are now obsolete. Policy makers in coastal regions must treat sea level rise not only as a flooding issue but as a factor that modulates the probability of seismic events. The empirical findings presented here provide the necessary urgency for this policy shift.

4.4. Limitations

It is essential to acknowledge the limitations of this study. The primary limitation in the geophysical analysis is the inherent difficulty in establishing definitive causation versus correlation in complex, macro-scale Earth systems. While the correlation is statistically robust and supported by physical theory (hydrostatic loading), external factors can always contribute. Furthermore, the analysis of LLM impact is based on a focused, although representative, selection of literature available up to 2023. As this technology evolves rapidly, the full scope of its societal impact is still unfolding. Finally, the study was constrained by the limited, foundational reference list, emphasizing the need for dedicated, multi-disciplinary research in the future.

5. References

1. Heaven, W.D. The Inside Story of How ChatGPT Was Built from the People Who Made It. MIT Technology Review. 2023. Available online: https://www.technologyreview.com/2023/03/03/1069311/inside-story-oral-history-how-chatgpt-built-openai/ (accessed on 22 March 2023).

2. Warren, T. Microsoft and Google Are about to Open an AI Battle; The Verge. 2023. Available online: https://www.theverge.com/2023/2/7/23587767/microsoft-google-open-ai-battle-search-bing (accessed on 13 February 2023).

3. Eliot, L. Generative AI ChatGPT Is Going To Be Everywhere Once The API Portal Gets Soon Opened, Stupefying AI Ethics And AI Law; Forbes. 2023. Available online: https://www.forbes.com/sites/lanceeliot/2023/01/22/generative-ai-chatgpt-is-goingto-be-everywhere-once-the-api-portal-gets-soon-opened-stupefying-ai-ethics-and-ai-law/?sh=48c917d724f50 (accessed on 13 February 2023)

4. Bansal, R., Reena Chandra, & Karan Lulla. (2025). Understanding and Mitigating Strategies for Large Language Model (LLMs) Hallucinations in HR Chatbots. International Journal of Computational and Experimental Science and Engineering, 11(3). https://doi.org/10.22399/ijcesen.2471

5. Makridakis, S.; Hibon, M. The M3-competition: Results, conclusions and implications. Int. J. Forecast. 2000, 16, 451–476.

6. Petropoulos, F.; Makridakis, S. The M4 competition: Bigger. Stronger. Better. Int. J. Forecast. 2020, 36, 3–6.

7. Makridakis, S.; Petropoulos, F.; Spiliotis, E. Introduction to the M5 forecasting competition Special Issue. Int. J. Forecast. 2022, 38, 1279–1282.

8. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. M5 Accuracy Competition: Results, Findings, and Conclusions. Int. J. Forecast. 2022, 38, 1346–1364.

9. LeCun, Y. A few thoughts on deep learning. In Proceedings of the Keynote Talk Presented at the 37th International Conference on Machine Learning (ICML 2020), Virtual, 13–18 July 2020.

10. Fildes, R.; Goodwin, P.; Lawrence, M.; Nikolopoulos, K. Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. Int. J. Forecast. 2009, 25, 3–23.

11. Petropoulos, F.; Fildes, R.; Goodwin, P. Do “big losses” in judgmental adjustments to statistical forecasts affect experts’ behaviour? Eur. J. Oper. Res. 2016, 249, 842–852.

12. Reuters. GM Explores Using ChatGPT in Vehicle; Reuters. 2023. Available online: https://www.reuters.com/business/autostransportation/gm-explores-using-chatgpt-vehicles-2023-03-10. (accessed on 13 March 2023).

13. Toews, R. The Next Generation of Artificial Intelligence (Part 2); Forbes. 2023. Available online: https://www.forbes.com/sites/robtoews/2020/10/29/the-next-generation-of-artificial-intelligence-part-2/?sh=3d89e3f97a30 (accessed on 22 March 2023).

14. Chandra, R., Bansal, R., & Lulla, K. (2025).

Benchmarking techniques for real-time evaluation of LLMs in production systems. International Journal of Engineering, Science and Information Technology, 5(3), 363–372. https://doi.org/10.52088/ijesty.v5i3.955

15. Huang, J.; Gu, S.S.; Hou, L.; Wu, Y.; Wang, X.; Yu, H.; Han, J. Large language models can self-improve. arXiv 2022, arXiv:221011610.

16. Gupta, K. Understanding Explainable AI And Interpretable AI; Marktechpost. 2023. Available online: https://www.marktechpost.com/2023/03/11/understanding-explainable-ai-and-interpretable-ai/ (accessed on 22 March 2023).

17. Mills, S. ChatGPT: Why It Will Probably Remain Just a Tool That Does Inefficient Work More Efficiently; The Conversation. 2023. Available online: https://theconversation.com/chatgpt-why-it-will-probably-remain-just-a-tool-that-does-inefficient-workmore-efficiently-201315 (accessed on 14 March 2023).

18. Zurcher, A. AI: How 'Freaked Out' Should We Be? BBC. 2023. Available online: https://www.bbc.com/news/world-us-canada64967627 (accessed on 22 March 2023).

19. Chandra, R., Lulla, K., & Sirigiri, K. (2025). Automation frameworks for end-to-end testing of large language models (LLMs). Journal of Information Systems Engineering and Management, 10(43s), e464–e472. https://doi.org/10.55278/jisem.2025.10.43s.8400

20. AI Threat Countermeasures: Defending Against LLM-Powered Social Engineering. (2025). International Journal of IoT, 5(02), 23-43. https://doi.org/10.55640/ijiot-05-02-03