Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Hallucination Detection and Mitigation in Large Language Models: A Comprehensive Review

Dr. Sanjay Nakharu Prasad Kumar
Researcher, USA

ARTICLE INFO

Received: 14 Dec 2024

Revised: 15 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Large Language Models have achieved unprecedented capabilities in natural language generation but remain vulnerable to hallucinations—outputs that are fluent and plausible yet factually incorrect or ungrounded. This comprehensive review examines the current landscape of hallucination detection and mitigation in LLMs, analyzing the theoretical foundations, detection methodologies, and mitigation strategies that have emerged to address this critical challenge. To explore the fundamental taxonomy distinguishing intrinsic hallucinations that deviate from input content from extrinsic hallucinations that contradict real-world facts, while examining how these manifestations vary across different natural language generation tasks. The review synthesizes five primary detection approaches including uncertainty estimation, attention pattern analysis, self-consistency checks, external fact verification, and trained evaluators, each offering unique advantages for identifying hallucinated content. To analyze mitigation strategies at both architectural levels through techniques like retrieval-augmented generation, tool integration, and factual fine-tuning, and systemic levels through guardrails, fallback policies, and human oversight. The evaluation landscape is examined through diverse benchmarks ranging from general-purpose frameworks to domain-specific assessments, with particular emphasis on the growing importance of multilingual and multimodal evaluation. The analysis reveals that while complete elimination of hallucinations is theoretically impossible in sufficiently complex models, a layered combination of improved architectures, rigorous detection methods, and systemic defenses offers the most effective path toward safe and trustworthy LLM deployment across critical applications.

Keywords: Hallucination detection, large language models, retrieval-augmented generation, attention mechanism analysis, multilingual evaluation

INTRODUCTION

Large Language Models (LLMs) have transformed natural language processing with their unparalleled ability to produce fluent, contextually appropriate text for a wide range of applications. These strong systems, though, have one weakness of utmost importance: hallucinations—responses that sound credible but are factually wrong or not grounded in the world. According to Zhang et al.'s comprehensive survey, hallucinations manifest across multiple dimensions, from fabricated facts and logical inconsistencies to contextual deviations, presenting significant challenges for real-world deployment

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

[1]. This phenomenon poses significant risks in high-stakes domains, from healthcare to legal services, where fabricated information could have severe consequences.

Hallucinations in LLMs can be categorized into two primary types: intrinsic hallucinations that deviate from provided input, and extrinsic hallucinations that contradict real-world facts. These categories align with two fundamental quality dimensions, faithfulness to source content and factuality in relation to verifiable truths. The survey by Zhang et al. emphasizes that hallucinations occur across various natural language generation tasks, including summarization, dialogue generation, and question answering, with each task presenting unique challenges for detection and mitigation [1]. Recent theoretical findings suggest that hallucinations are mathematically inevitable in sufficiently complex, Turing-complete language models, shifting the research focus from elimination to effective detection and mitigation strategies.

This comprehensive review examines the current landscape of hallucination management in LLMs, analyzing detection techniques, mitigation approaches at both architectural and systemic levels and evaluation methodologies. As detailed in recent research, the field has developed sophisticated approaches ranging from uncertainty-based detection methods to retrieval-augmented generation systems, each offering unique advantages in addressing different aspects of the hallucination problem [2]. We provide a critical synthesis of recent research while identifying persistent challenges and future directions for developing more trustworthy AI systems that can maintain high levels of accuracy and reliability across diverse applications and domains.

THEORETICAL FOUNDATIONS AND TAXONOMY OF HALLUCINATIONS

Understanding hallucinations in LLMs requires a robust theoretical framework and clear taxonomy. The foundational work by Ji et al. introduced the distinction between intrinsic and extrinsic hallucinations, which has since evolved into more nuanced categorizations encompassing logical inconsistencies, temporal errors, and ethical violations. According to Ji et al.'s comprehensive survey, hallucinations in natural language generation systems can be systematically categorized based on their relationship to source content and real-world knowledge, providing a structured framework for understanding these phenomena across different NLG tasks, including summarization, dialogue generation, and data-to-text generation [3].

Intrinsic hallucinations manifest as deviations from input content, such as unsupported details in summaries or contradictions to provided context. These errors often arise from the model's tendency to generate plausible-sounding information that extends beyond the source material given. The survey by Ji et al. highlights that intrinsic hallucinations are particularly prevalent in abstractive summarization tasks, where models must balance faithfulness to source content with generating coherent, fluent text [3]. Extrinsic hallucinations, conversely, involve statements that conflict with established facts or trusted external sources, reflecting the model's limitations in maintaining accurate world knowledge.

The theoretical underpinnings of hallucinations reveal their fundamental nature in complex language models. Recent research has introduced novel approaches to understanding hallucinations through attention mechanism analysis, where topological patterns in attention graphs can reveal when models are likely to generate hallucinated content [4]. This topological approach provides insights into the internal mechanisms that lead to hallucinations, suggesting that certain attention patterns correlate with factually incorrect generations. Mathematical proofs demonstrate that for any sufficiently expressive model capable of universal computation, there exist inputs that will inevitably produce hallucinated outputs. This insight has profound implications for mitigation strategies, suggesting that complete elimination is theoretically impossible and that practical approaches must focus on detection, reduction, and harm minimization through understanding the underlying attention mechanisms and topological structures that contribute to hallucination generation [4].

Hallucination Type	Category	Primary Characteristics	Common Occurrence Context	Detection Method
Intrinsic Hallucinations	Source-based	Deviations from input content	Abstractive summarization	Attention pattern analysis
Extrinsic Hallucinations	Knowledge-based	Conflicts with real-world facts	Open-domain QA	External fact-checking
Logical Inconsistencies	Reasoning-based	Contradictory statements	Multi-step reasoning	Self-consistency checks
Temporal Errors	Time-based	Incorrect temporal relationships	Historical narratives	Temporal validation
Ethical Violations	Value-based	Breaches of ethical guidelines	Content generation	Rule-based filters

Table 1: Ethical Risk Assessment Across AI Application Domains [3, 4]

METHODS

The detection of hallucinations in LLM outputs has evolved from simple heuristics to sophisticated multi-modal approaches. Current detection techniques can be organized into five primary categories, each offering unique advantages and limitations. According to Rawte et al.'s comprehensive survey, the field has witnessed rapid advancement in developing methods that can identify hallucinations across various modalities and tasks, with particular emphasis on addressing the challenges posed by large foundation models [5].

Uncertainty and confidence estimation methods, particularly semantic entropy, provide zero-resource approaches to identify potentially hallucinated content. By measuring uncertainty in meaning space rather than surface text variations, these techniques can flag low-confidence outputs without requiring task-specific training data. The survey by Rawte et al. emphasizes that these uncertainty-based approaches have become increasingly sophisticated, leveraging the internal confidence signals of foundation models to detect when outputs may be unreliable [5]. Complementary approaches utilize token-level log probabilities, Monte Carlo dropout, and ensemble variance to quantify model uncertainty.

Attention pattern analysis, exemplified by the TOHA (Topology of Hallucination Attention) framework, examines the internal attention mechanisms linking inputs to outputs. Strong, focused attention patterns typically indicate proper grounding in source context, while diffuse or weak attention often correlates with hallucinated content. This approach provides interpretable insights into the model's reasoning process. Research on multimodal large language models has shown that hallucination patterns can manifest differently across modalities, with visual-language models exhibiting unique challenges in maintaining consistency between textual and visual information [6].

Self-consistency checks, implemented through frameworks like SelfCheckGPT, generate multiple responses to original and perturbed prompts. High divergence among responses suggests unstable factual grounding, effectively identifying areas where the model lacks reliable knowledge. External fact-checking and retrieval-based verification leverage Natural Language Inference (NLI) models and authoritative sources to validate generated claims, while trained evaluator models like FactCC specialize in detecting contradictions between outputs and reference texts. The survey on multimodal hallucinations highlights that detection methods must evolve to handle the increased complexity of cross-modal verification, where hallucinations can arise from misalignment between different input modalities [6].

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Context/Application	Uncertainty Methods	Attention Analysis	Consistency Checks	External Verification	Trained Evaluators
Foundation Models	High	Medium	High	Medium	Medium
Visual-Language Models	Low	Medium	Medium	High	Low
Real-time Applications	High	Low	Low	Low	Medium
High-stakes Domains	Medium	High	High	High	High
Resource-constrained	High	Medium	Low	Low	Medium

Table 2: Detection Method Effectiveness Across Application Contexts in Large Language Models [5, 6]

RESULTS

Evaluation Frameworks and Benchmarks

Comprehensive evaluation of hallucination detection and mitigation systems requires diverse benchmarks, automated metrics, and human assessment protocols. The evaluation landscape has matured significantly, with specialized datasets addressing different aspects of the hallucination problem. Recent research emphasizes that truthfulness evaluation must extend beyond English-centric assessments to capture the global deployment of language models across diverse linguistic contexts [9].

General-purpose benchmarks like TruthfulQA test models' susceptibility to common misconceptions through adversarially crafted questions, while HalluLens provides taxonomy-aware evaluation across multiple task types. The importance of multilingual evaluation has become increasingly apparent, as research shows that truthfulness patterns vary significantly across languages, with models exhibiting different hallucination tendencies when generating content in non-English languages [9]. Domain-specific benchmarks have emerged to address field-specific challenges: MedHallu and Med-HALT focus on medical hallucinations, CodeHaluEval targets programming-related errors, and HALLUCINOGen evaluates vision-language models.

Automated metrics range from simple overlap-based measures like ROUGE and BERTScore to sophisticated learned classifiers and NLI-based systems. The M3Exam benchmark represents a significant advancement in evaluation methodology by providing a multilingual, multimodal, and multilevel framework that assesses LLMs across diverse tasks and languages simultaneously [10]. Knowledge-based verification through frameworks like KILT and retrieval-based evaluation via RAE provide complementary approaches to assess factual accuracy. However, human evaluation remains the gold standard, with annotators assessing correctness, faithfulness, and hallucination types at fine-grained levels.

The evaluation ecosystem continues to evolve, with recent efforts focusing on continuous post-deployment monitoring and standardized reporting metrics. The M3Exam framework demonstrates how comprehensive evaluation must consider multiple dimensions simultaneously, incorporating various difficulty levels, subject domains, and modalities to provide a holistic assessment of model capabilities and limitations [10]. These developments are crucial for maintaining trust as LLMs are deployed in increasingly critical applications across diverse cultural and linguistic contexts.

Benchmark Type	Example	Primary Focus	Language Support
General-purpose	TruthfulQA	Common misconceptions	English
Taxonomy-aware	HalluLens	Multiple task types	English
Medical	MedHallu	Clinical accuracy	Multi-language

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Programming	CodeHaluEval	Code errors	Multi-language
Multimodal	HALLUCINOGen	Vision-language	Multi-language

Table 4: Language Support Coverage Across Different Hallucination Benchmark Categories [9, 10]

Mitigation Strategies: Architectural and Systemic Approaches

Mitigation of hallucinations requires a multi-layered approach combining architectural innovations with systemic safeguards. At the architectural level, four primary strategies have emerged as particularly effective. Recent research on longchain approaches demonstrates that structured reasoning chains can significantly reduce hallucinations by breaking down complex queries into verifiable sub-components, allowing models to validate each step before proceeding to the next [7].

The implementation of these mitigation strategies increasingly relies on sophisticated cloud infrastructure to handle the computational demands of real-time hallucination detection and correction. Recent work on scalable cloud architectures for AI-driven decision systems demonstrates how serverless computing paradigms enable complex AI pipelines to be decomposed into independently scalable components, providing the flexibility needed for dynamic hallucination detection workloads [11]. This architectural approach allows hallucination detection systems to scale elastically based on demand, ensuring consistent performance even under varying workloads. The integration of container orchestration through Kubernetes provides a unified control plane that manages the complex interdependencies between different components of hallucination detection pipelines, from initial uncertainty estimation to final verification stages.

Tool use and external module integration, exemplified by systems like Toolformer, enable LLMs to autonomously invoke APIs for factual sub-tasks such as calculations or database queries. This approach offloads specific knowledge requirements to specialized systems, reducing the burden on the language model to maintain perfect factual accuracy. The longchain methodology extends this concept by creating systematic verification chains that can identify and correct potential hallucinations before they propagate through the generation process [7]. Multi-cloud architectures further enhance these capabilities by distributing verification workloads across providers, improving geographical distribution and ensuring compliance with data residency requirements while maintaining low-latency access to fact-checking resources. The implementation of feature stores and model registries within these cloud environments enables consistent tracking of model performance metrics related to hallucination rates, facilitating continuous improvement through disciplined MLOps practices.

Factual fine-tuning involves training models on curated datasets emphasizing accuracy, often incorporating adversarially generated hallucination corrections. This approach directly addresses the model's tendency to generate plausible but incorrect information. Research by Varshney et al. introduces a proactive detection mechanism that validates low-confidence generations before they manifest as hallucinations, implementing a "stitch in time" philosophy that prevents errors rather than correcting them post-hoc [8]. The deployment of these fine-tuned models benefits significantly from cloud-native architectures that support real-time model serving and A/B testing, allowing organizations to continuously evaluate and improve hallucination mitigation strategies while maintaining production stability.

Systemic strategies operate at the deployment level, implementing guardrails and rule-based filters to validate outputs against domain-specific knowledge bases. The low-confidence validation approach demonstrates that by identifying and addressing uncertain generations early in the process, systems can prevent the cascade of errors that often lead to severe hallucinations in downstream tasks [8]. Fallback and refusal policies enable systems to abstain from answering when confidence is low, while human-in-the-loop oversight remains critical for high-stakes applications. Transparency features in user interfaces, including confidence indicators and citation displays, empower users to verify information independently.

Detection Methodologies

The detection of hallucinations in LLM outputs has evolved from simple heuristics to sophisticated multi-modal approaches. Current detection techniques can be organized into five primary categories, each offering unique advantages and limitations. According to Rawte et al.'s comprehensive survey, the field has witnessed rapid advancement in developing methods that can identify hallucinations across various modalities and tasks, with particular emphasis on addressing the challenges posed by large foundation models [5].

Uncertainty and confidence estimation methods, particularly semantic entropy, provide zero-resource approaches to identify potentially hallucinated content. By measuring uncertainty in meaning space rather than surface text variations, these techniques can flag low-confidence outputs without requiring task-specific training data. The survey by Rawte et al. emphasizes that these uncertainty-based approaches have become increasingly sophisticated, leveraging the internal confidence signals of foundation models to detect when outputs may be unreliable [5]. Complementary approaches utilize token-level log probabilities, Monte Carlo dropout, and ensemble variance to quantify model uncertainty.

The application of anomaly detection principles from other domains provides valuable insights for hallucination detection. Research on fraud detection in credit card transactions demonstrates how autoencoders and deep neural networks can identify anomalous patterns in high-dimensional data streams [12]. These techniques, originally developed to detect fraudulent transactions in real-time across financial institutions, offer parallels to hallucination detection where the goal is identifying outputs that deviate from expected patterns. The autoencoder architecture learns normal patterns during training and flags significant deviations during inference, a principle that translates effectively to detecting when language models generate content outside their reliable knowledge boundaries. This cross-domain insight suggests that hallucination detection can benefit from established anomaly detection methodologies, particularly in scenarios requiring real-time identification of problematic outputs.

Attention pattern analysis, exemplified by the TOHA (Topology of Hallucination Attention) framework, examines the internal attention mechanisms linking inputs to outputs. Strong, focused attention patterns typically indicate proper grounding in source context, while diffuse or weak attention often correlates with hallucinated content. This approach provides interpretable insights into the model's reasoning process. Research on multimodal large language models has shown that hallucination patterns can manifest differently across modalities, with visual-language models exhibiting unique challenges in maintaining consistency between textual and visual information [6].

Self-consistency checks, implemented through frameworks like SelfCheckGPT, generate multiple responses to original and perturbed prompts. High divergence among responses suggests unstable factual grounding, effectively identifying areas where the model lacks reliable knowledge. External fact-checking and retrieval-based verification leverage Natural Language Inference (NLI) models and authoritative sources to validate generated claims, while trained evaluator models like FactCC specialize in detecting contradictions between outputs and reference texts. The survey on multimodal hallucinations highlights that detection methods must evolve to handle the increased complexity of cross-modal verification, where hallucinations can arise from misalignment between different input modalities [6]. The integration of these diverse detection approaches, from uncertainty estimation to anomaly detection patterns borrowed from fraud detection systems, creates a robust framework for identifying hallucinated content across various contexts and applications.

CONCLUSION

The challenge of hallucinations in Large Language Models represents a fundamental tension between the generative capabilities that make these systems valuable and the accuracy requirements essential for real-world applications. This comprehensive review has demonstrated that while theoretical results confirm hallucinations cannot be completely eliminated from sufficiently complex models, effective management strategies can significantly reduce their occurrence and impact through multi-layered

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

approaches. The progress from naive heuristic detection to state-of-the-art approaches using semantic entropy, topological attention analysis, and cross-modal verification demonstrates the growth of the field in reasoning about and tackling this challenge. The state-of-the-art best practices now include combining architectural advances like retrieval-augmented generation, structured reasoning chains, and tool use with systemic mitigations such as domain-specific guardrails and human monitoring, building solid defense mechanisms that develop for various application domains. The evaluation environment has also adapted to include multilingual, multimodal, and domain-specific test sets that more accurately capture the varying deployment environments of contemporary LLMs. Challenges remain, though, such as non-standard definitions and measures, the challenge of identifying subtle hallucinations, the lack of complete explainability of detection mechanisms, and ongoing post-deployment monitoring requirements. Future research directions include integrated evaluation models, uncertainty-sensitive generation paradigms, and a hybrid neuro-symbolic architecture that brings together the fluency of neural models and the accuracy of symbolic reasoning. As LLMs move deeper into high-stakes applications from medicine to law, building strong hallucination control systems extends beyond technical prowess to being an ethical necessity for safe AI deployment, demanding ongoing collaboration among researchers, practitioners, and domain specialists to ensure such powerful instruments can be relied upon where precision is most critical.

REFRENCES

[1] Yue Zhang et al., "Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models." ResearchGate, September 2023. https://www.researchgate.net/publication/373686208_Siren's_Song_in_the_AI_Ocean_A_Survey_on_Hallucination_in_Large_Language_Models

[2] Sanjay Nakharu Prasad Kumar, "Scalable Cloud Architectures for AI-Driven Decision Systems." Journal of Computer Science and Technology Studies, Al-Kindi Publishers, August 2025.

https://al-kindipublishers.org/index.php/jcsts/article/view/10545

[3] Xiaorui Ma et al., "Efficiently Integrate Large Language Models with Visual Perception: A Survey from the Training Paradigm Perspective," ScienceDirect, January 2026. https://www.sciencedirect.com/science/article/pii/S1566253525004920

[4] Sanjay Nakharu Prasad Kumar, "RMHAN: Random Multi-Hierarchical Attention Network with RAG-LLM-Based Sentiment Analysis Using Text Reviews." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, World Scientific, 2025.

https://www.worldscientific.com/doi/10.1142/S1469026825500075

[5] Sanjay Nakharu Prasad Kumar, "Improving Fraud Detection in Credit Card Transactions Using Autoencoders and Deep Neural Networks." The George Washington University, 2022.

https://scholarspace.library.gwu.edu/concern/gw_etds/cv43nx607

[6] Sanjay Nakharu Prasad Kumar, "AI and Cloud Data Engineering Transforming Healthcare Decisions." SAR Council, August 2025.

https://sarcouncil.com/2025/08/ai-and-cloud-data-engineering-transforming-healthcare-decisions

[7] Ziwei Ji et al., "Survey of Hallucination in Natural Language Generation." ResearchGate, February 2022.
https://www.researchgate.net/publication/358458381_Survey_of_Hallucination_in_Natural_Language_Generation

[8] Sanjay Nakharu Prasad Kumar, "Recent Innovations in Cloud-Optimized Retrieval-Augmented Generation Architectures for AI-Driven Decision Systems." Engineering Management Science Journal, Vol. 9, No. 4, 2025.

https://doi.org/10.59573/emsj.9(4).2025.81

[9] Alexandra Bazarova et al., "Hallucination Detection in LLMs via Topological Divergence on Attention Graphs." ResearchGate, April 2025. https://www.researchgate.net/publication/390772361_Hallucination_Detection_in_LLMs_via_Topological_Divergence_on_Attention_Graphs

[10] Sanjay Nakharu Prasad Kumar, "Quantum-Enhanced AI Decision Systems: Architectural Approaches for Cloud-Based Machine Learning Applications." SAR Council, August 2025.

https://sarcouncil.com/2025/08/quantum-enhanced-ai-decision-systems-architectural-approaches-for-cloud-based-machine-learning-applications

[11] Vipula Rawte et al., "A Survey of Hallucination in Large Foundation Models." ResearchGate, September 2024. https://www.researchgate.net/publication/386380732_A_Survey_of_Hallucination_in_Large_Foundation_Models

[12] Sanjay Nakharu Prasad Kumar, "Optimized Convolutional Neural Network for Land Cover Classification via Improved Lion Algorithm." Transactions in GIS, Wiley, March 2024.

https://onlinelibrary.wiley.com/doi/10.1111/tgis.13150

[13] Sanjay Nakharu Prasad Kumar, "Deep Embedded Clustering with Matrix Factorization Based User Rating Prediction for Collaborative Recommendation." Microprocessors and Microsystems, SAGE, 2022.

https://journals.sagepub.com/doi/abs/10.3233/MGS-230039

[14] Zechen Bai et al., "Hallucination of Multimodal Large Language Models: A Survey." ResearchGate, April 2024. https://www.researchgate.net/publication/380263724_Hallucination_of_Multimodal_Large_Language_Models_A_Survey

[15] Sanjay Nakharu Prasad Kumar, "SCSLnO-SqueezeNet: Sine Cosine–Sea Lion Optimization Enabled SqueezeNet for Intrusion Detection in IoT." Information and Computer Security, Taylor & Francis, 2023.

https://www.tandfonline.com/doi/abs/10.1080/0954898X.2023.2261531

[16] Jinchao Li & Quan Hong. "A Longchain Approach to Reduce Hallucinations in Large Language Models." ResearchGate, June 2024. https://www.researchgate.net/publication/381186640_A_Longchain_Approach_to_Reduce_Hallucinations_in_Large_Language_Models

[17] Sanjay Nakharu Prasad Kumar, "An Approach for DoS Attack Detection in Cloud Computing Using Sine Cosine Anti-Coronavirus Optimized Deep Maxout Network." International Journal of Pervasive Computing and Communications, Emerald, 2023.

https://doi.org/10.1108/IJPCC-05-2022-0197

[18] Neeraj Varshney et al., "A Stitch in Time Saves Nine: Detecting and Mitigating Hallucinations of LLMs by Validating Low-Confidence Generation." ResearchGate, July 2023. https://www.researchgate.net/publication/372247807_A_Stitch_in_Time_Saves_Nine_Detecting_and_Mitigating_Hallucinations_of_LLMs_by_Validating_Low-Confidence_Generation

[19] Sanjay Nakharu Prasad Kumar, "Optimized Attention-Driven Bidirectional Convolutional Neural Network: Recurrent Neural Network for Facebook Sentiment Classification." International Journal of Intelligent Information Technologies, IGI Global, 2023.

https://www.igi-global.com/article/optimized-attention-driven-bidirectional-convolutional-neural-network/349572

[20] Sanjay Nakharu Prasad Kumar, “Analyzing the Impact of Corporate Social Responsibility on the Profitability of Multinational Companies: A Descriptive Study.” International Journal of Interdisciplinary Management Studies, 2022.

https://ijims.org/index.php/home/article/view/56

[21] Blanca Calvo Figueras et al., "Truth Knows No Language: Evaluating Truthfulness Beyond English." ResearchGate, February 2025. https://www.researchgate.net/publication/388963717_Truth_Knows_No_Language_Evaluating_Truthfulness_Beyond_English

[22] Sanjay Nakharu Prasad Kumar, “Optimal Weighted GAN and U-Net Based Segmentation for Phenotypic Trait Estimation of Crops Using Taylor Coot Algorithm.” Applied Soft Computing, Elsevier, 2023.

https://www.sciencedirect.com/science/article/abs/pii/S1568494623004143

[23] Wenxuan Zhang et al., "M3Exam: A Multilingual, Multimodal, Multilevel Benchmark for Examining Large Language Models." ResearchGate, June 2023. https://www.researchgate.net/publication/371414283_M3Exam_A_Multilingual_Multimodal_Multilevel_Benchmark_for_Examining_Large_Language_Models

[24] Sanjay Nakharu Prasad Kumar, “ECG-Based Heartbeat Classification Using Exponential-Political Optimizer Trained Deep Learning for Arrhythmia Detection.” Biomedical Signal Processing and Control, Elsevier, 2023.

https://www.sciencedirect.com/science/article/abs/pii/S1746809423002495

[25] Sanjay Nakharu Prasad Kumar, “PSSO: Political Squirrel Search Optimizer–Driven Deep Learning for Severity Level Detection and Classification of Lung Cancer.” International Journal of Information Technology & Decision Making, World Scientific, 2023.

https://www.worldscientific.com/doi/abs/10.1142/S0219622023500189