

Article

Not peer-reviewed version

Cost-Efficient Secure Hybrid RAG Assessment

[jitender singh](#)*

Posted Date: 3 December 2024

doi: 10.20944/preprints202412.0196.v1

Keywords: Semantic Relevance; Factual Accuracy; Context Consistency; Hallucination Detection; Responsible AI Evaluations; Conversational Systems; Healthcare Informatics; Finance Analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Cost-Efficient Secure Hybrid RAG Assessment

Jitender Singh

Manager Data Science, Enterprise Architect, Publicis Sapient, jitender.singh1@publicissapient.com

Abstract: This paper presents a cost-effective and scalable hybrid methodology to evaluate the generator in retrieval-augmented generation (RAG) systems, tailored for offline use in critical fields like healthcare and finance. Current generator evaluations rely on online LLMs, which can hallucinate and bias scores, leading to cost-inefficient, non-reproducible, and non-scalable outcomes. Meanwhile, traditional NLP approaches lack a multi-faceted view and are overly dependent on entity or phrase matching mechanisms, missing deeper semantic understanding. Our approach leverages sophisticated NLP metrics to comprehensively evaluate AI system performance across semantic relevance, factual accuracy, context consistency, semantic coherence, and hallucination detection through specialized tools, offering a more robust approach than current market solutions. The methodology demonstrates superior performance over LLM-based judges by providing stable, scalable, and cost-effective offline assessment across both RAG systems and natural language generation tasks. We implement a bounded scoring system (0-1) using harmonic means with PCA, adaptive, and entropy-weighting techniques for trustworthy scoring. The framework's modular architecture enables continuous metric adaptation and domain-specific customization while maintaining evaluation consistency, making it particularly valuable for high-stakes applications requiring rigorous quality assessment without external API dependencies.

Keywords: Semantic Relevance; Factual Accuracy; Context Consistency; Hallucination Detection; Responsible AI Evaluations; Conversational Systems

1. Introduction

In the rapidly evolving field of retrieval-augmented generation (RAG) systems, evaluating the generator component using online large language models (LLMs) presents significant challenges (Kau, 2024). LLMs are prone to hallucinations, where they generate information that is factually incorrect or irrelevant, leading to unreliable evaluations. Additionally, these models can introduce scoring biases, affecting the accuracy and fairness of the assessment (Reddy et al., 2024). This reliance on online models results in non-reproducible outcomes, as the same inputs may yield different results depending on external factors such as updates or model changes (Mehmet Yusuf, 2019). Furthermore, online assessments are often cost-inefficient, requiring continuous API calls and computational resources for large-scale evaluations, making them impractical for many organizations (Ramírez et al., 2024).

Traditional NLP evaluation mechanisms, such as precision, recall, BLEU, METEOR, and ROUGE, rely heavily on surface-level entity or phrase matching to assess the quality of generated text (et al., 2021). While these metrics are widely used, they fail to capture deeper semantic understanding, which is crucial for tasks like contextual relevance, factual accuracy, and coherence (Linguistic and Philosophical Investigations 2024). Precision and recall, for instance, focus primarily on how well entities or keywords match between the predicted and reference text, but they do not account for the meaning or relationships between those entities (Hickling & Hanley, 2005). This can lead to misleading evaluations, especially in cases where the generated content is semantically correct but does not match the reference text exactly (Deutsch et al., 2022).

Metrics like BLEU and ROUGE, which measure n-gram overlap, are also limited in their ability to assess the fluency and coherence of generated text, particularly in more complex or abstract tasks

(Kim et al., 2018). They might reward models for generating common phrases or sequences without evaluating the contextual appropriateness or factual accuracy of the content. METEOR attempts to improve upon these by considering synonyms and stemming, but it still struggles to evaluate deeper levels of semantic meaning and the overall quality of the response (Fay et al., 2023). Consequently, these traditional methods do not provide a holistic view of a model's performance, especially in real-world applications where nuance and accuracy are critical.

Referenced, advanced semantic metrics are crucial for a holistic evaluation of generated text in RAG systems. *Query relevance* ensures the response directly addresses the user's intent, improving the alignment of generated content with the actual query (BAI et al., 2014). *Factual accuracy* is important to verify that the information generated is correct, especially in high-stakes domains like healthcare and finance (Ni & Huang, 2024). *Context consistency* ensures the generated content aligns with the broader discourse, preventing contradictions (Carnielli & Malinowski, 2018). *Semantic Coherence* measures the logical flow of the output, ensuring readability and comprehensibility (Li, 2022). *Semantic relevance* captures the deeper meaning of text, moving beyond surface-level word matching, while *hallucination detection* identifies and mitigates the generation of inaccurate or fabricated information (Li et al., 2024). Together, these techniques provide a comprehensive assessment of both relevance and quality in generated text, offering a more nuanced view compared to traditional metrics.

Offline assessment solutions are critical in healthcare and finance due to security, compliance, and operational concerns (Gonzalez-Granadillo et al., 2021). They ensure data privacy by keeping sensitive information within controlled environments, addressing regulatory requirements like HIPAA and GDPR. Offline solutions also mitigate data sovereignty issues by processing data within local jurisdictions (Tzanou, M. (2020). Additionally, they offer cost-efficiency and scalability, avoiding the high costs and limitations of online services. With better reliability and reproducibility, offline evaluations provide consistent, secure assessments, essential for high-stakes decisions in both sectors, such as patient care in healthcare and risk management in finance Rani, S., Kataria, A., Bhambri, P., Pareek, P. K., & Puri, V. (2024).

Our approach employs established metrics such as relevance, coherence, contextual consistency, factual accuracy, and hallucination detection, but refines them with advanced techniques, thereby enhancing their precision and effectiveness for accurate target metric assessment (Kaifi, 2024). For instance, our semantic similarity computation integrates multi-dimensional embedding analysis, cosine similarity, dot product similarity, and context-aware TF-IDF weights (Chiny et al., 2021). By applying a weighted scoring method, this framework delivers a robust and precise measure of semantic relevance, surpassing the performance of traditional single-metric techniques (Wu, 2019). Additionally, each metric is computed by evaluating 2-3 sub-metrics, which are then aggregated into a single score. This aggregation process, inspired by ensemble machine learning techniques such as voting, ensures a holistic and accurate assessment aligned with the target intent (Phuoc et al., 2010).

2. Methods

2.1. The Query Relevance Assessment Method Integrates Three Distinct Techniques to Provide a Comprehensive Evaluation Score by Capturing Semantic, Factual, Structural, and Probabilistic Aspects of the Query-Response Relationship and Is Designed to Be Adaptable to Various Query Types and Response Lengths

- a) Semantic similarity analysis utilizes sentence embedding models to compute sentence-level semantic alignment which captures detailed meaning and context correspondence between query and response.
- b) Knowledge graph analysis identifies concepts by extracting key entities from both query and response using NLP techniques then relationship mapping is done by constructing graph

representations depicting connections between identified entities where graph comparison analyses similarities between query and response graphs based on: Node overlap (shared entities), Edge overlap (shared relationships), Graph centrality and Additional contextual information in the response.

- c) Facts consistency analysis extract and match similarities between numerical, location and temporal facts between query, context and response.

2.2. The Factual Accuracy Assessment Method Combines Four Techniques to Evaluate Response Correctness Through Entity and Numerical Extraction Matched Against Provided Context

- a) Query-response semantic alignment as described in 2.1.a
- b) Context-response semantic alignment using equivalent methodology as 2.1.a, applied to context-response pairs
- c) Knowledge graph analysis following methodology in 2.1.b, applied to context-response entity relationships
- d) Facts consistency analysis as detailed in 2.1.c

2.3. The Context Consistency Measures How Well the Response Aligns with the Context by Integrating Three Distinct Techniques to Provide a Comprehensive Evaluation Score

- a) Coherence analysis employs natural language inference to assess context-response relationships through probabilistic computation of entailment, neutral, and contradiction states which quantifies the logical consistency between the input context and generated response.
- b) Context-response semantic alignment using equivalent methodology as 2.1.a, applied to context-response pairs
- c) Knowledge graph analysis following methodology in 2.1.b, applied to context-response entity relationships

2.4. The Semantic Relevance Measures Overall Coverage, Contextual Coherence and Contextual Relevance by Integrating Four Distinct Techniques to Provide a Comprehensive Evaluation Score

- a) Coverage score computes aspect embeddings and segment embeddings by identifying main topic and aspects from question and segments/tokens from answer. Alignment cosine scores are computed using both embeddings and coverage by taking their mean difference from one.
- b) Coherence score computes taking mean of cosine similarities of previous step identified segment embeddings.
- c) Utilization score uses context embeddings to compute cosines mean against segment and contextual embeddings.
- d) Relevance score computes by non-linear modeling sigmoid over contextual embeddings.

2.5. The Semantic Coherence Measures Overall Coherence and Relevance of the Response to the Given Query, Context by Integrating four Distinct Techniques to Provide a comprehensive Evaluation Score: NLI Coherence and Linguistic Acceptability (CoLA)

- a) Computes the coherence between a query-response sentences and context-response sentences pair using a Natural Language Inference (NLI) model which tokenizes the inputs, passes them through the model, and calculates probabilities for entailment and neutral classifications. The final coherence score is derived from these probabilities, with entailment given full weight and neutral given half weight.

- b) CoLA method computes linguistic acceptability coherence scores for query-response and context-response pairs using a pre-trained CoLA model to assess the grammaticality and coherence of the text. The method tokenizes each text, passes it through the model, and obtains probabilities of coherence. It then calculates a weighted average of these scores, with higher weights given to query-response and context-response pairs measuring how well-formed and contextually appropriate the response is.

2.6. The Hallucination Score Assess the Likelihood of Hallucinations in Generated Text, Providing a Quantitative Measure of Semantic Consistency Between the Context and the Response Using Two Techniques to Provide a Comprehensive Score

- a) Natural Language Inference (NLI) uses a pre-trained NLI model to compute a contradiction score between the context and each sentence in the response. Then applies SoftMax to model outputs to obtain probabilities for contradiction, neutral, and entailment classes.
- b) Question generation and answering technique employs a question generation model to create questions from each sentence in the response utilizing a question-answering model to generate answers based on the context then compares the generated answer with the original sentence using semantic similarity.

2.7. The Answer Correctness Measures the Generated Response Alignment Against the Ground Truth in Terms of Relevance, Factual Accuracy, and Consistency Using Three Techniques to Provide a Comprehensive Score

- a) Query-relevance, ref pt 2.1, applied using ground truth replacing input query against the response.
- b) Factual-accuracy, ref pt 2.2, applied using ground truth replacing context against input response.
- c) Context-consistency, ref pt 2.3, applied using ground-truth replace input context against the response.

2.8. The Generator Competence Module Computes Weighted Scores Using Three Strategies

- a) PCA (Principal Component Analysis): Analyses variance
- b) Entropy weighting: Assesses information content
- c) Adaptive weighting: Evaluates relative variability among scores

The final score combines these weighted approaches with a harmonic mean, penalizing low-performing metrics. This robust mechanism reduces bias from any single metric, providing a comprehensive measure of overall system effectiveness in text generation tasks.

Pseudocode explanation of the algorithm

Algorithm 1 Compute Semantic Metrics

Input: query, response, groundTruth, context**Output:** Various semantic metric scores

```

1: procedure COMPUTESEMANTRICMETRICS(query, response, groundTruth, context)
2:   Initialize all variables to None
3:   queryRelevance, querySemantic  $\leftarrow$  COMPUTEQUERYRELEVANCE(query, response, context,
   sents)
4:   if context  $\neq$  None then
5:     factualAccuracy, kgScore, kiScore, contextSemantic  $\leftarrow$  COMPUTEFACTUALACCU-
   RACY(query, context, response, sents, querySemantic)
6:     contextConsistency  $\leftarrow$  COMPUTECONTEXTCONSISTENCY(query, context, response, sents,
   contextSemantic, kgScore, kiScore)
7:     hallucinationScore  $\leftarrow$  COMPUTEHALLUCINATIONSCORE(context, sents[response])
8:   end if
9:   semanticRelevance, contextEntailment, responseCoherence  $\leftarrow$  COMPUTESEMANTRICRELE-
   VANCE(sents[query], sents[response], sents[context], contextEntailment)
10:  semanticCoherence  $\leftarrow$  COMPUTESEMANTRICCOHERENCE(query, context, response,
   contextEntailment)
11:  if groundTruth  $\neq$  None then
12:    answerRelevance, aquerySemantic  $\leftarrow$  COMPUTEQUERYRELEVANCE(groundTruth, response,
   context, sents)
13:    answerAccuracy, akgScore, akiScore, acontextSemantic  $\leftarrow$  COMPUTEFACTUALACCU-
   RACY(query, groundTruth, response, sents, aquerySemantic)
14:    answerConsistency  $\leftarrow$  COMPUTECONTEXTCONSISTENCY(query, groundTruth, response,
   sents, acontextSemantic, akgScore, akiScore)
15:    if answerRelevance  $\neq$  None and answerConsistency  $\neq$  None then
16:      scoreHistory  $\leftarrow$  { answerRelevance, answerAccuracy, answerConsistency,
   responseCoherence }
17:      answerCorrectness  $\leftarrow$  WEIGHTEDSCORE(scoreHistory)
18:    end if
19:  end if
20:  return queryRelevance, factualAccuracy, contextConsistency, semanticRelevance,
   semanticCoherence, hallucinationScore, answerCorrectness
21: end procedure

```

Algorithm 2 GeneratorCompetence**Input:** List of scores**Output:** Weighted Final Score

```

1: procedure ROUNDWEIGHTS(weights)
2:   return {k : round(v, 4) for k, v ∈ weights.items()}
3: end procedure
4: procedure PCAWEIGHTING(scores)
5:   if len(scores) < 2 then
6:     return uniform weights.
7:   end if
8:    $X \leftarrow$  matrix from score values.
9:   Apply PCA, normalize first component weights.
10:  return rounded weights.
11: end procedure
12: procedure ENTROPYWEIGHTING(scores)
13:  if len(scores) < 2 then
14:    return uniform weights.
15:  end if
16:  Compute entropy and scale weights as  $1 - \frac{\text{entropy}}{\sum \text{entropy}}$ .
17:  return rounded weights.
18: end procedure
19: procedure ADAPTIVEWEIGHTING(scores)
20:  if len(scores) < 2 then
21:    return uniform weights.
22:  end if
23:  Compute coefficient of variation  $\frac{\sigma}{\mu}$ .
24:  Normalize and round weights.
25:  return rounded weights.
26: end procedure
27: procedure COMBINESCORES(scores, weights)
28:  return  $\sum_{k \in \text{scores}} \text{scores}[k] \cdot \text{weights}[k]$ 
29: end procedure
30: procedure WEIGHTEDSCORE(scores, scoreWeights)
31:  if Only one score with single value then
32:    return that value.
33:  end if
34:  if scoreWeights provided then
35:    Apply weights, combine PCA, entropy, and adaptive scores.
36:  end if
37:  Compute harmonic mean and final weighted score.
38:  return final score.
39: end procedure
40: procedure COMPUTEHARMONIC(scoresList)
41:  Compute harmonic mean of scoresList.
42: end procedure

```

3. Architecture**Layer by Layer explanation of the architectural diagram:**

1. FastAPI Layer: Serves as the API interface for reference queries and generated outcomes along with evaluation metrics.
2. Metrics Aggregator Layer: Collects and combines metrics from various evaluation modules.
3. Semantic Metrics Layer: Measures various dimensions of the output generated by the model.
The key metrics:

- 3.1. Query Relevance: How relevant the response is to the input query.
- 3.2. Factual Accuracy: Whether the information provided is factually correct.
- 3.3. Context Consistency: Checks if the response is consistent within the given context.

- 3.4. Semantic Relevance: Measures how closely the response aligns with the topic.
- 3.5. Semantic Coherence: Ensures that the generated text flows logically.
- 3.6. Hallucinations Score: Detects fabricated or incorrect information in the response.
- 3.7. Answer Correctness: Assesses whether the response is appropriate for provided answer/ground truth.
- 3.8. Generator Competence: Evaluates how well the text-generation module functions overall.
4. Models Layer: Contains various models responsible for both understanding and generating responses:
 - 4.1. **SpaCy**: Likely used for named entity recognition (NER), part-of-speech tagging, and other language processing tasks.
 - 4.2. **Pretrained NLI (Natural Language Inference)**: Used to check the logical consistency and infer relationships between the query and the generated response.
 - 4.3. **Pretrained CoLA (Corpus of Linguistic Acceptability)**: Used to measure linguistic correctness and fluency in the response.
 - 4.4. **Pretrained QA (Question Answering)**: A pre-trained model for question-answering tasks.
 - 4.5. **Text-Generation Pipeline**: Likely the core LLM pipeline that generates the actual response text from the input query.
5. Cloud Platform Layer: Models hosting and running platforms like:
 - 5.1. GKE (Google Kubernetes Engine): A managed Kubernetes service that handles deployment, scaling, and management of the models and services.
 - 5.2. Proprietary/Open-Source LLM: The system can work with either proprietary or open-source large language models (LLMs) for text generation.

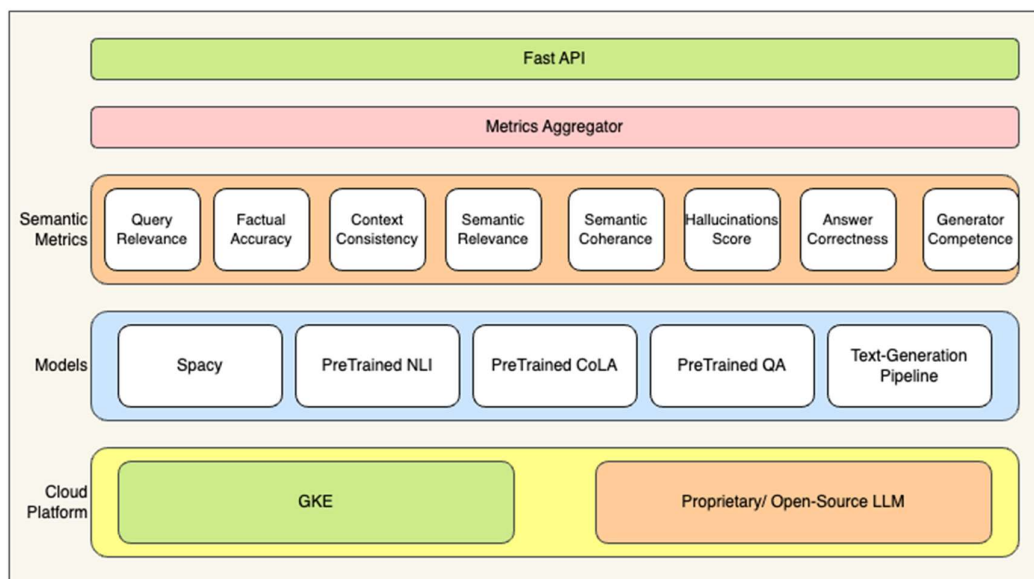


Figure 1. diagram represents a system architecture for evaluating and generating responses using various semantic metrics and language models (LLMs).

4. Results and Discussions

Following are the main contributions of this paper:

- 1) **Cost-Effective, Scalable Offline Evaluation:** Provides a scalable, offline solution tailored for high-stakes sectors like healthcare and finance, reducing external dependencies and ensuring compliance with data privacy regulations.
- 2) **Security and Compliance:** Safeguards sensitive data by keeping evaluation processes within secure internal systems, mitigating risks associated with online processing and ensuring regulatory adherence.
- 3) **Enhanced Reliability and Reproducibility:** Ensures stable, reproducible results free from the biases and hallucinations common in online LLM-based evaluations.
- 4) **Holistic, Multi-Dimensional Assessment:** Utilizes advanced semantic metrics (e.g., query relevance, context consistency) to offer a comprehensive evaluation of generated content, surpassing traditional entity-based methods.
- 5) **Robust Scoring System:** Leverages **harmonic means**, **PCA**, and **entropy-Weighting** techniques to provide a balanced, trustworthy, and accurate scoring mechanism that reflects multiple evaluation dimensions.
- 6) **Modular and Customizable Framework:** Allows for continuous adaptation and domain-specific customization, ensuring consistent quality assessments across various NLP tasks and specialized industries.

Sr. No	Metric (against Generated Response)	Semantic Evaluation	LLM-as-Judge Evaluation	Percentage Change
1	Query Relevance	0.8825	0.7624	13.61%
2	Factual Accuracy	0.9432	0.8824	6.45%
3	Context Consistency	0.8825	0.7055	20.06%
4	Semantic Coherence	0.8945	0.7704	13.87%
5	Semantic Relevance	0.8025	0.7830	2.43%
6	Hallucination Detection	1.0000	0.9601	3.99%
7	Answer Correctness	0.9324	0.8253	11.49%
8	Generator Competence	0.9053	0.8127	10.27%

Figure 2. reflects offline metrics under 'semantic-evaluation' column against llm-as-judge online approach between reference and generated texts.

5. Conclusion

Our approach significantly advances the evaluation of retrieval-augmented generation (RAG) systems by integrating diverse assessment aspects, including semantic relevance, factual accuracy, coherence, consistency, and hallucination detection. This holistic framework offers a more comprehensive evaluation compared to traditional metrics, providing a deeper, multi-dimensional analysis of generated content. Beyond RAG systems, the framework has broader applicability across natural language generation tasks, enhancing reliability and performance assessment in AI-driven models. By addressing challenges such as hallucinations, bias, and context consistency, it offers researchers and developers a robust tool for improving system quality. This methodology ensures more accurate, scalable, and cost-effective assessments, making it indispensable for high-stakes fields like healthcare and finance.

References

1. Kau, A. (2024) Understanding retrieval pitfalls: Challenges faced by retrieval augmented generation (RAG) models [Preprint]. doi:10.59350/wek8b-sbj84.
2. Reddy, G.P., Pavan Kumar, Y.V. and Prakash, K.P. (2024) 'Hallucinations in large language models (llms)', 2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), pp. 1–6. doi:10.1109/estream61684.2024.10542617.
3. Mehmet Yusuf, E. (2019) 'Analysis of different data sets of the same clinical trial may yield different results', *Annals of Advanced Biomedical Sciences*, 2(2). doi:10.23880/aabsc-16000135.
4. Ramírez, G. et al. (2024) 'Cache & Distil: Optimising API calls to large language models', *Findings of the Association for Computational Linguistics ACL 2024*, pp. 11838–11853. doi:10.18653/v1/2024.findings-acl.704.
5. Saadany, H. and Orăsan, C. (2021) 'Bleu, Meteor, bertscore: Evaluation of metrics performance in assessing critical translation errors in sentiment-oriented text', Proceedings of the Translation and Interpreting Technology Online Conference TRITON 2021, pp. 48–56. doi:10.26615/978-954-452-071-7_006.
6. 'Enhancing natural language understanding with deep learning: Contextual and semantic implementation' (2024) Linguistic and Philosophical Investigations [Preprint]. doi:10.52783/lpi.64.
7. Hickling, T. and Hanley, W. (2005) Methodologies and metrics for assessing the strength of relationships between entities within semantic graphs [Preprint]. doi:10.2172/883760.
8. Deutsch, D., Dror, R. and Roth, D. (2022) 'On the limitations of reference-free evaluations of generated text', Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing [Preprint]. doi:10.18653/v1/2022.emnlp-main.753.
9. Kim, G., Fukui, K. and Shimodaira, H. (2018) 'Word-like character N-Gram embedding', Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text, pp. 148–152. doi:10.18653/v1/w18-6120.
10. Fay, K., Moritz, C. and Whaley, S. (2023) 'Considering meaning: What's it about? what might it really be about?', *Powerful Book Introductions*, pp. 41–67. doi:10.4324/9781032682099-3.
11. BAI, L. et al. (2014) 'Long tail query recommendation based on query intent', *Chinese Journal of Computers*, 36(3), pp. 636–642. doi:10.3724/sp.j.1016.2013.00636.
12. Ni, B. and Huang, Q. (2024) Verifying through involvement: Exploring how anthropomorphism enhances users' intentions to verify AI-generated information [Preprint]. doi:10.2139/ssrn.4963144.
13. Carnielli, W. and Malinowski, J. (2018) 'Contradictions, from consistency to inconsistency', *Trends in Logic*, pp. 1–9. doi:10.1007/978-3-319-98797-2_1.
14. Li, W. (2022) 'Text genres, readability and readers' comprehensibility', *European Journal of Computer Science and Information Technology*, 10(4), pp. 52–62. doi:10.37745/ejsit.2013/vol10n45262.
15. Li, X. et al. (2024) 'A text generation hallucination detection frame-work based on fact and semantic consistency', 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), pp. 970–974. doi:10.1109/ainit61980.2024.10581604.
16. cyber-physical solutions for real-time detection, analysis and visualization at operational level in water CIs. Cyber-Physical Threat Intelligence for Critical Infrastructures Security: Securing Critical Infrastructures in Air Transport, Water, Gas, Healthcare, Finance and Industry. <https://doi.org/10.1561/9781680838237.ch8>
17. Addressing big data and AI challenges. Health Data Privacy under the GDPR, 106-132. <https://doi.org/10.4324/9780429022241-9>
18. Artificial intelligence in personalized health services for better patient care. *Information Systems Engineering and Management*, 89-108. https://doi.org/10.1007/978-3-031-65022-2_6
19. Kaifi, R. (2024) 'Enhancing brain tumor detection: A novel CNN approach with advanced activation functions for accurate medical imaging analysis', *Frontiers in Oncology*, 14. doi:10.3389/fonc.2024.1437185.
20. Chiny, M. et al. (2021) 'Netflix recommendation system based on TF-IDF and cosine similarity algorithms', Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning, pp. 15–20. doi:10.5220/0010727500003101.
21. Wu, D. (2019) 'Robust face recognition method based on kernel regularized relevance weighted discriminant analysis and deterministic approach', *Sensing and Imaging*, 20(1). doi:10.1007/s11220-019-0258-7.
22. Phuoc, T. et al. (2010) 'Network intrusion detection using machine learning and voting techniques', *Machine Learning* [Preprint]. doi:10.5772/9150.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.