Explainable Generative AI: An Incremental Transparency Perspective on Techniques, Evaluation and Future Directions

Bekhzod Norboev[1], Nodirbek Islomov[1], Gulmira Pardaeva[1], Golib Berdiev[1], Alisher Rustamov[1], Shamiljon Rustamov[1], Adylkhan Kibishov[2] and Andokulov Tulkin[1]
[1]*University of Information Technology and Management, Saroy Str. 3, 180212 Qarshi, Uzbekistan*
[2]*Khoja Akhmet Yassawi International Kazakh-Turkish University, B. Sattarkhanov Ave. 29, 161200 Turkistan, Kazakhstan*
norboyevbexzod98@gmail.com, nodirbek.islomov.dev@gmail.com, mis.gulish@gmail.com, golibberdiev@gmail.com, alirus900@gmail.com, shamiljon1988@gmail.com, adylkhan.kibishov@ayu.edu.kz, mr.warner@mail.ru

Keywords: Generative Artificial Intelligence, Explainable Artificial Intelligence, Incremental Transparency, Trustworthy AI, Multimodal Explainability, Model Interpretability.

Abstract: Generative Artificial Intelligence (GenAI) enables machine learning systems to generate realistic text, images, audio, and video, driving widespread adoption across multiple industries. However, the internal decision-making processes of generative models remain largely opaque, limiting interpretability and undermining user trust. Although outputs may appear accurate at a surface level, they can reflect biased or misaligned information, raising concerns about ethical, medical, and social acceptability. Existing explainability techniques for GenAI are still immature and largely adapted from methods designed for deterministic predictive models, such as those used in finance and risk analysis. While effective in those domains, these approaches fail to capture generative-specific phenomena, including latent-space dynamics and probabilistic generation, resulting in predominantly surface-level explanations. To address these limitations, this study conducts a systematic analytical review of GenAI explainability techniques and introduces an Incremental Transparency Perspective, which frames explainability as a continuous and cumulative optimization process across the GenAI lifecycle. Empirical trend analysis reveals structural gaps and architecture-dependent limitations in current approaches, underscoring the need for measurable, stakeholder-centered, and architecture-aware explainability frameworks. Together, these contributions establish a foundation for developing trustworthy, accountable, and deployable GenAI systems.

1 INTRODUCTION

Since the emergence of Generative Artificial Intelligence (GenAI), a paradigm shift has occurred in how machines produce human-level, context-aware content across domains such as healthcare, finance, cybersecurity, and creative industries. Advances in generative architectures including transformer-based large language models (LLMs), diffusion models, Generative Adversarial Networks (GANs), and Variational Autoencoders (VAEs) have enabled outputs with increased realism, diversity, and scalability, driving the transition of GenAI systems from experimental prototypes to real-world deployment in decision-support and human-AI collaborative workflows [1]-[5].

Unlike traditional predictive models, GenAI systems rely on stochastic sampling in high-dimensional latent spaces and iterative decoding, which significantly complicates interpretability. The internal reasoning processes of these models are often opaque, making it difficult to trace how inputs, latent variables, or architectural components influence outputs. Traditional explainability methods such as feature attribution, saliency mapping, and surrogate models are ill-suited to this setting, as they fail to capture generative specific phenomena including latent-space dynamics, probabilistic branching, and multi-path decoding [2], [6]-[8]. As GenAI systems are increasingly deployed in high-stakes and regulated environments, transparency, accountability, and auditability become essential. This paper argues that explainability in GenAI should be treated as a continuous optimization process rather than a binary property. Systematically integrating incremental transparency across the model lifecycle design, training, inference, and evaluation offers a practical pathway for embedding explainability into GenAI

system development and deployment [1], [6], [9]-[11].

Traditional explainable artificial intelligence (XAI) methodologies are primarily grounded in the assumption of a deterministic or near-deterministic relationship between inputs and outputs, typically expressed as:

$$y = f(x). \quad (1)$$

Under this formulation, explanations aim to identify which input features most strongly influence a given prediction. While effective for classification and regression tasks, this paradigm does not adequately reflect the operational logic of generative models [6], [7].

In contrast, GenAI systems generate outputs by sampling from probabilistic distributions, often mediated through latent variables and complex parameter spaces:

$$y \mid Z, \theta \sim P(Y \mid Z, \theta), Z \mid \theta \sim P(Z \mid \theta). \quad (2)$$

Where Z represents latent variables and θ denotes model parameters. This probabilistic formulation highlights that a single input may correspond to multiple plausible outputs, each arising from different sampling trajectories within the latent space. As a result, explainability in GenAI must extend beyond surface-level output attribution to encompass latent-space transformations, stochastic sampling decisions, and iterative decoding dynamics [2], [12].

As a result of their inherent complexity, these processes also add a new layer of difficulty when it comes to being transparent. The latent variables produced by generative models are frequently high-dimensional and abstract in nature; hence a direct semantic interpretation cannot be made. In addition to the high-dimensional aspect of generative models, decoding methods utilized to extract an output from the latent variable may involve multiple branches, leading to numerous output predictions, each having an equal likelihood of occurrence. Furthermore, based on architectural style, there is much variability between types of generative models (e.g., diffusion-based systems vs autoregressive transformers) which further increase the diversity of the approaches used to construct and evaluate reasoning.

Because of these differences, it is necessary to develop generative-native frameworks for explainability that incorporate probabilistic behaviour, sensitivity to model structure and the temporal nature associated with generative processes. Thus, to construct an effective framework for explainable GenAI it is important to develop conceptual and evaluation strategies tailored to build upon the fundamental tenets of generative modelling, instead of piecing together different types of explainable XAI techniques from a disparate selection of methodologies. The gap in this area must be closed if we are to be able to provide users with a meaningful degree of transparency; the ability to build trust with users and provide responsible deployment of GenAI systems in practical situations.

2 METHODOLOGY AND EMPIRICAL ANALYSIS

This section outlines the methodological framework and empirical procedures employed to examine explainability in Generative Artificial Intelligence systems. It describes the systematic literature review process, the analytical criteria used to evaluate existing explainability techniques, and the empirical approaches applied to assess trends, methodological limitations, and evaluation practices. By combining structured review protocols with comparative and architecture-aware analysis, this section establishes the empirical foundation for identifying gaps in current GenAI explainability methodologies and for motivating the proposed Incremental Transparency Perspective.

2.1 Systematic Review Protocol (PRISMA)

The systematic review protocol used in this research follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, which specify how to identify, screen, and analyse relevant studies related to Generative Artificial Intelligence. This protocol ensures a transparent, reproducible, and methodologically rigorous approach that is more robust than a narrative or ad hoc review.

The literature search was conducted across Scopus, Web of Science, IEEE Xplore, and the ACM Digital Library for the period 2020-2026 using the query: (“generative AI” OR “diffusion model” OR “large language model”) AND (“explainability” OR “interpretability” OR “transparency”).

All studies were identified through a comprehensive search of major academic databases. Duplicate records were removed prior to screening, and the remaining studies were assessed for relevance based on predefined inclusion and exclusion criteria. The screening process included expert review, where domain specialists evaluated methodological relevance and explainability focus.

Studies were included if they satisfied the following criteria:

1) focus on explainable AI methods applied to generative models;
2) provide empirical or experimental validation;
3) report sufficient methodological detail for analysis.

Methodological coding was performed by the authors to classify architectures, XAI techniques, and evaluation strategies. Only peer-reviewed articles and high-quality academic preprints were retained in the final corpus.

As shown in Figure 1, the PRISMA diagram provides a transparent and reproducible overview of the study selection process, reducing selection bias and improving analytical reliability.

2.2 Visualization-Based Explainability

Techniques based on visualizations for explain- ing generative artificial intelligence (GenAI) are frequently used techniques in the body of knowl- edge related to explainable GenAI research. Ex- amples of frequently used visualization tech- niques are Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM has been used to represent the importance of specific regions of an image or vision-language model output for a model's decision-making process.

However, such visualization-based explanations remain limited in capturing generative causality and latent-space dynamics [2], [6].

Although these types of visualizations provide some level of human interpretability at the local level by indicating where to look within the generative system's output, they offer limited explanatory capability. They focus primarily on the final output of the model and do not consider the transformations carried out by the model within the latent space, the stochastic nature of decision-making processes when performing sampling, and the iterative generation pathways for the final output of the model. As a result, using such visualizations to explain the decision-making process may lead to incomplete or misleading interpretations of generative reasoning, as they do not capture latent-space dynamics and stochastic decoding pathways.

As shown in Figure 2, Grad-CAM visualizations provide localized interpretability but do not capture latent-space dynamics and stochastic decoding pathways.

The figure contrasts real and AI-generated samples using Grad-CAM heatmaps, guided backpropagation, and occlusion maps. It demonstrates that these methods highlight salient regions in the final output but do not reveal latent transformations or probabilistic generation paths, thereby illustrating the limits of visualization-based explanations for generative models.

As shown in Figure 2, Grad-CAM visualizations provide localized interpretability for both real and AI-generated samples. The heatmaps highlight salient regions that influence the model's output; however, they do not reveal latent-space transformations, stochastic sampling trajectories, or multi-step decoding processes. This highlights the limitations of visualization-based explainability methods in capturing latent generative dynamics and stochastic decoding behavior.

2.3 Global Feature Attribution

Global feature attribution techniques, particularly SHAP (SHapley Additive exPlanations), have been widely applied to evaluate the relative importance of input features when using GenAI systems; these techniques aggregate the feature contributions of each input feature across multiple outputs or samples from the model to derive a global explanation for the output generated by the model based on the set of input features.

As shown in Figure 3, SHAP provides a global feature attribution view but does not capture latent generative dynamics.

As a result, global attribution methods provide only partial transparency for generative inference processes [7], [12].

Although there is widespread interest in approaches to explaining generative artificial intelligence (GAI), much of the support for these methods is based on their reliance upon surrogate models with simple mapping functions that approximate the creation of generative content. While statistical transparency is provided by surrogate models, true insight into how generative models operate internally is limited. In particular, global attribution methods don't represent the relationships between latent variables, reduce the influence of random sampling to branch points and don't provide an explanation that encompasses the entirety of the decoding process that a generative inference represents. Thus, the ability of a global attribution model to explain the generative model's reasoning is limited.

2.4 Distribution of Explainability Techniques

An aggregated analysis of the reviewed literature reveals a pronounced reliance on explainability

techniques originally developed for discriminative and predictive models.

As shown in Figure 4, the distribution highlights the dominance of non-native explainability techniques adapted from discriminative models.

This imbalance highlights the need for generative-native explainability research agendas [1], [12].

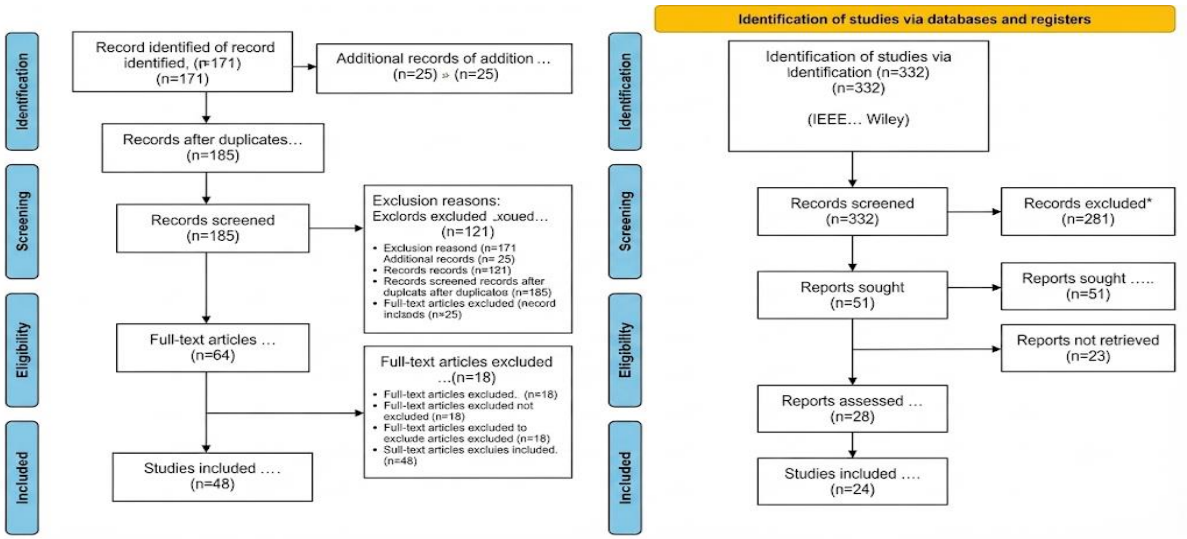

Figure 1: PRISMA flow diagram summarizing study identification, screening, eligibility, and inclusion stages.

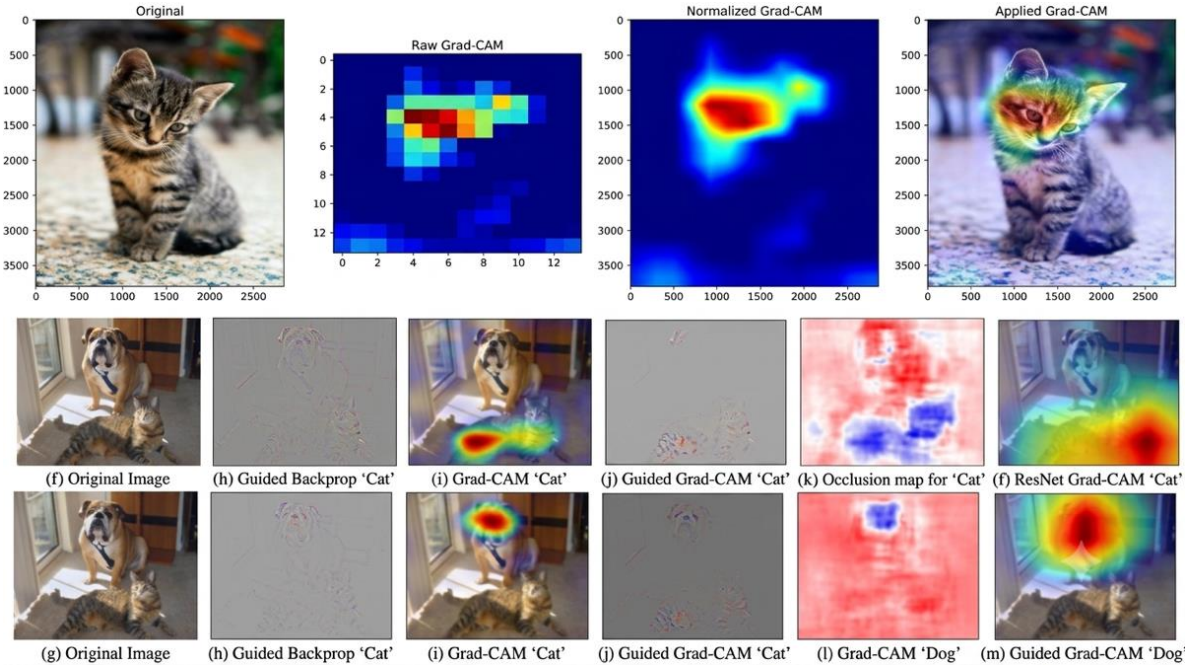

Figure 2: Grad-CAM visualizations for real and AI-generated samples.

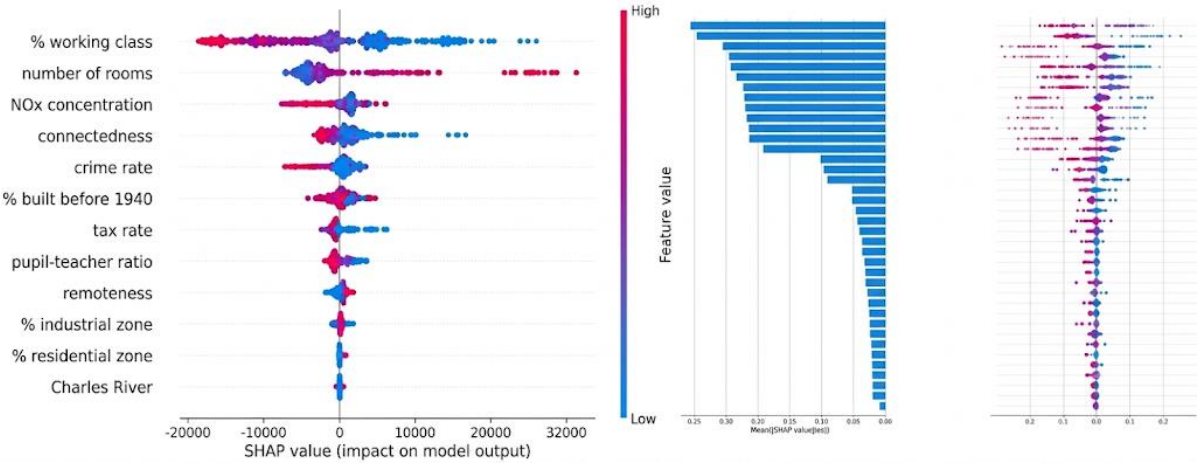

Figure 3: Conceptual SHAP summary plot for generative model outputs.

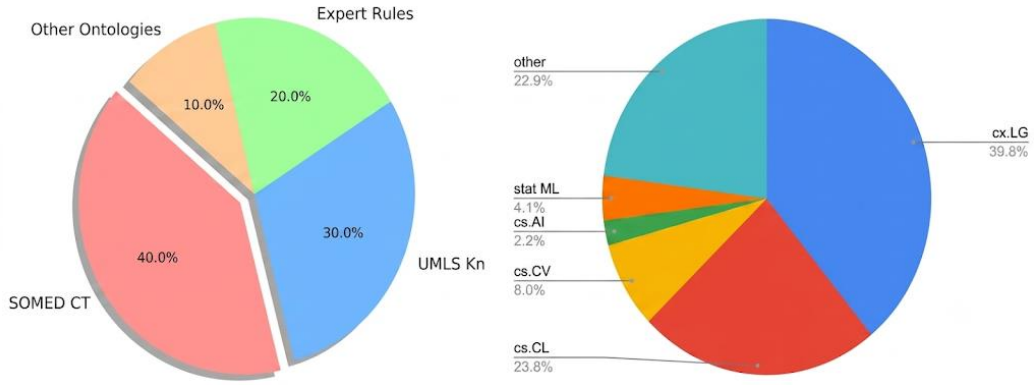

Figure 4: Distribution of explainability techniques across common AI and GenAI-specific methods.

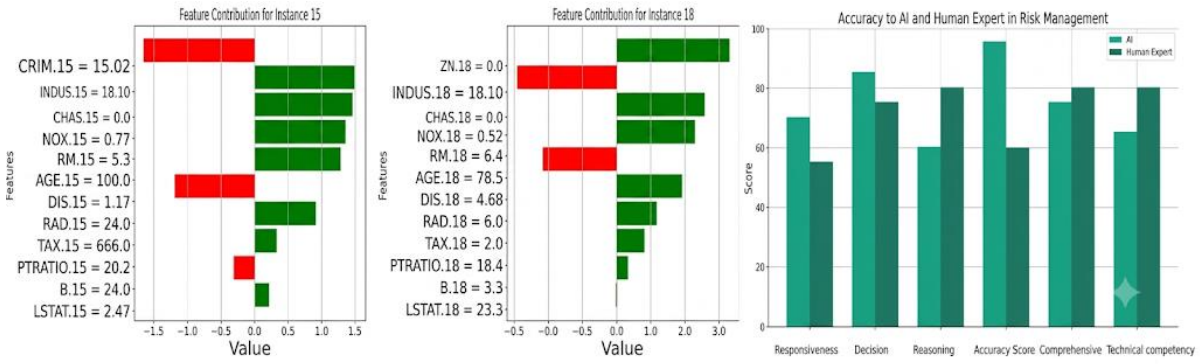

Figure 5: Distribution of evaluation methods used in explainable GenAI studies.

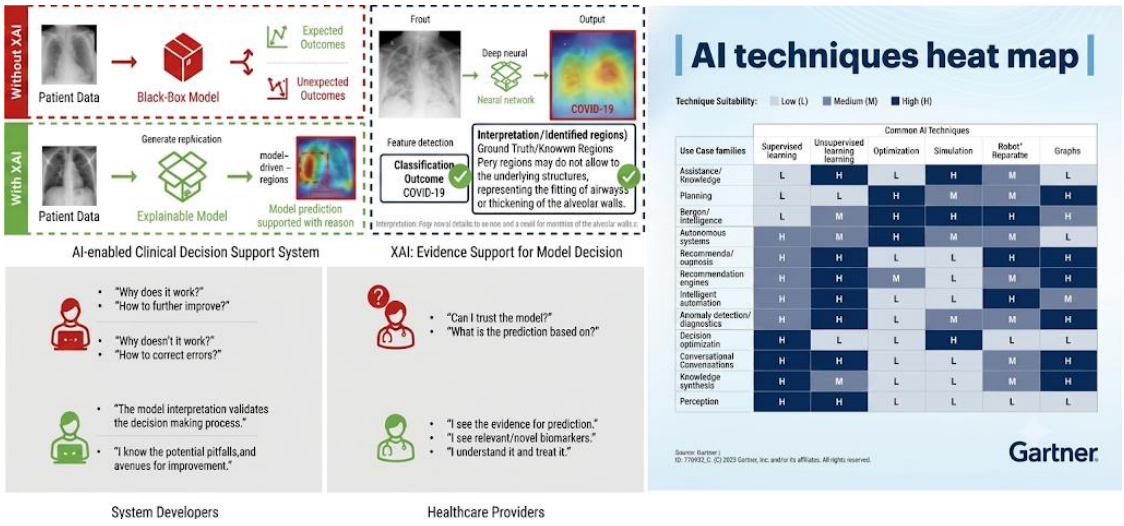

Figure 6: Relationship between explainability technique categories and evaluation rigor.

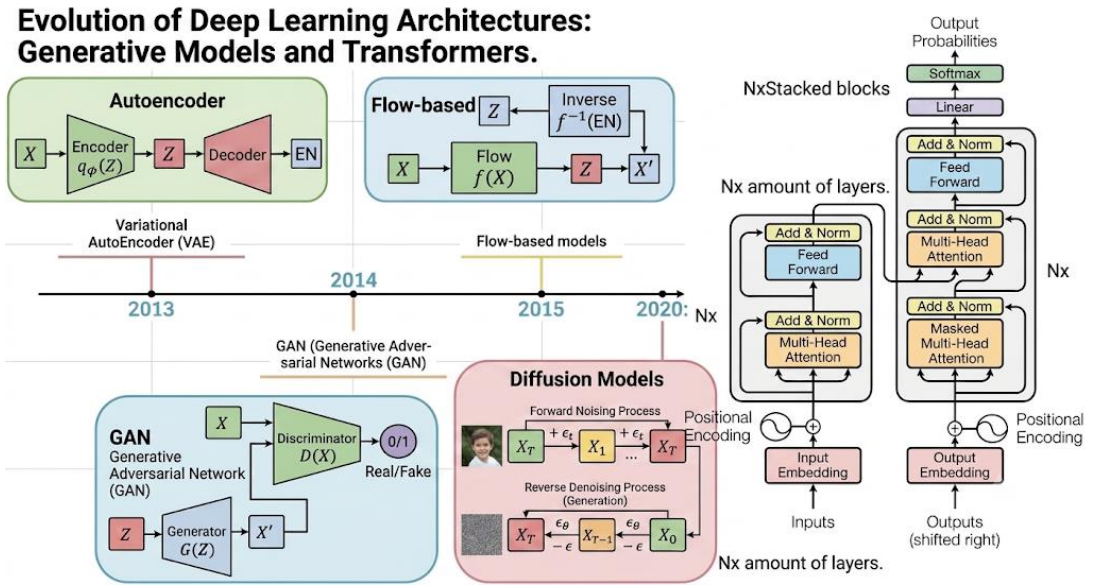

Figure 7: Architecture-sensitive explainability mapping.

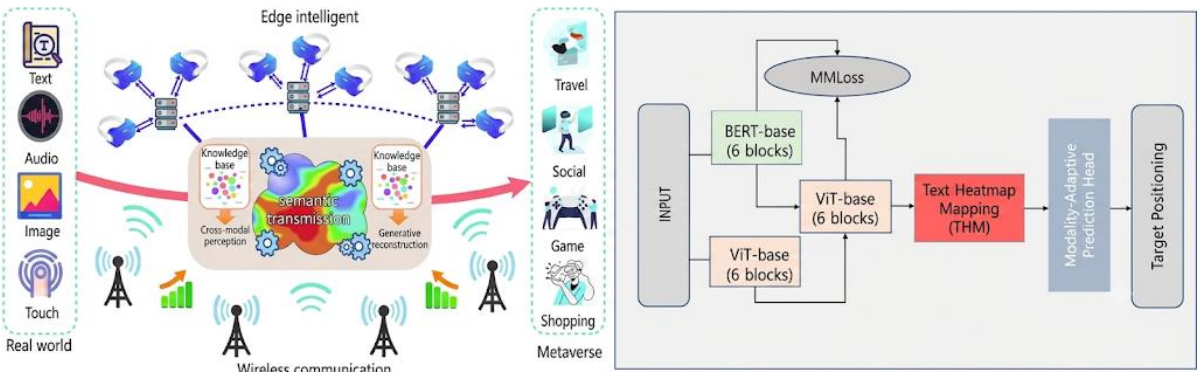

Figure 8: Cross-modal semantic alignment challenges in multimodal GenAI systems.

The widespread and dominant use of adapted non-native methods for providing explanation for GAI highlights a structural inequity that exists within current research on explainability in GAI. Although there are many reasons to use these adapted non-native approaches, they do not fully account for the specific characteristics and features associated with GAI. More importantly, the structural inequity of the current state of explainability in GAI highlights an area where further research is necessary, as more research is needed on developing explainability techniques specifically around the generative architecture of GAI and the probabilistic modeling paradigm that underlies GAI.

2.5 Evaluation Methodologies

Evaluation practices in explainable GenAI research are predominantly skewed toward performance-oriented metrics, such as accuracy, realism, or output quality, rather than explanation quality.

As shown in Figure 5, evaluation practices are dominated by performance-oriented metrics rather than explanation quality measures.

Without standardized evaluation metrics, the reliability of explainability claims remains limited [7], [13].

Due to the unequal representation of adapted non-native methods, many individuals have fallen into the trap of believing that the quality of the output provided by the generative model represents the quality of the explanation, when in fact there is no objective means of establishing a correlation between the quality of the output produced by a generative model and the quality of the explanation provided by the global attribution method. Due to the lack of standardized metrics for determining the fidelity, completeness and usefulness of the explanations provided by various global attribution methods, the systematic evaluation and comparison of various global attribution techniques for assessing explainability in GAI systems is severely hindered.

3 ADVANCED ANALYSIS AND CONCEPTUAL CONTRIBUTIONS

This section builds upon the empirical findings by providing deeper analytical insights and introducing the core conceptual contributions of the study. It examines the relationship between explainability quality and methodological choices, analyzes the architecture-dependent nature of explainability in generative models, and explores the challenges introduced by multimodal GenAI systems. Together, these analyses support the formulation of an Incremental Transparency Perspective, highlighting structural limitations in existing approaches and outlining principled directions for advancing explainability in Generative Artificial Intelligence.

3.1 Explainability Quality vs. Technique Category

As shown in Figure 6, generative-native methods exhibit higher conceptual alignment with GenAI architectures but lack mature empirical validation.

Generatives and Novel-Type methods have high conceptual congruency with Generation AI methods but have a lack of substantial empirical validation. This impedes their use in practice and their acceptance in regulated and/or safety-critical areas. Conversely, established/existing methods benefit from mature evaluation capabilities but provide limited explainability of Generative processes.

3.2 Architecture-Sensitive Explainability

As shown in Figure 7, explainability effectiveness varies significantly across generative architectures, reflecting differences in internal representations, inference mechanisms, and generation dynamics [2].

This study finds that explainability techniques cannot be directly transferred across architectures (e.g., autoregressive transformers, diffusion models, and GANs). Therefore, the use of explainability techniques that do not account for the underlying architecture may lead to misleading or incomplete explanations, reinforcing the need for architecture-specific design and evaluation of explainability methods.

3.3 Multimodal Explainability Challenges

As shown in Figure 8, these challenges underscore the limitations of single-modality explainability approaches in multimodal GenAI systems [4], [5], [14].

A primary area of research that is still not solved is to trace the semantic intent and causative effect across modalities. Misalignment of modalities creates obfuscation within the explanation chain and reduces user trust in situations where consistent and interpretable cross-modal reasoning is needed.

Table 1: Incremental transparency scorecard (0-5 scale).

Indicator	Description	Score Range
Traceability	Ability to track output to input tokens/latents	0-5
Faithfulness	Explanation reflects true model behavior	0-5
Completeness	Coverage of full generation pipeline	0-5
Uncertainty disclosure	Sampling randomness exposed	0-5
Architecture awareness	Method aligned with model type	0-5
Latent interpretability	Insight into latent space dynamics	0-5
Decoding transparency	Visibility of generation steps	0-5
Stakeholder relevance	Human-understandable explanations	0-5

Table 2: Transparency score comparison across representative genai architectures.

Indicator	LLM	Diffusion
Traceability	2	3
Faithfulness	3	3
Completeness	2	4
Uncertainty disclosure	2	5
Architecture awareness	3	4
Latent interpretability	1	3
Decoding transparency	2	4
Stakeholder rel.	4	3
Total	19	29

Table 3: Summary of included studies (prisma corpus).

ID	Ref.	Architecture	XAI Method	Evaluation Metric	Domain
S1	[6]	General XAI / GenAI	Conceptual framework	Fidelity, human evaluation	Cross-domain
S2	[1]	Foundation models	Survey taxonomy	Transparency criteria	Cross-domain
S3	[7]	Generative models	Interpretability evaluation	Faithfulness, completeness	Methodology
S4	[2]	Diffusion	Latent trajectory analysis	Sampling traceability	Computer vision
S5	[9]	Foundation models	Trustworthiness assessment	Risk & transparency metrics	Cross-domain
S6	[15]	Vision foundation models	Saliency / feature attribution	Localization accuracy	Vision
S7	[16]	LLM + knowledge systems	Attention analysis	Reasoning faithfulness	NLP
S8	[17]	LLM	Mechanistic interpretability	Circuit fidelity	NLP
S9	[18]	Generative AI	Bias auditing framework	Fairness metrics	Healthcare
S10	[3]	LLM	Security analysis	Robustness, safety metrics	Cybersecurity
S11	[10]	LLM	Transparency compliance	Auditability score	Governance
S12	[11]	LLM (clinical)	Expert evaluation	Human trust score	Healthcare
S13	[13]	LLM	Bias evaluation	Bias audit metrics	Education
S14	[19]	LLM	Benchmarking analysis	Performance vs explainability	NLP
S15	[8]	Generative AI	Mechanistic interpretability	Circuit fidelity	Methodology
S16	[20]	LLM	Architecture survey	Structural analysis	NLP
S17	[4]	Multimodal LLM	Cross-modal attention	Alignment score	Multimodal AI
S18	[5]	Multimodal models	Evaluation taxonomy	Multimodal reasoning metrics	Multimodal AI
S19	[12]	Explainable GenAI	Two-stage XAI review	Taxonomy completeness	Cross-domain
S20	[14]	Vision-language models	Fusion interpretability	Cross-modal consistency	Robotics / Vision

Current explainability techniques have no mechanisms in place to account for the full range of these interactions, so this area is a critical part of future research efforts.

3.4 Incremental Transparency Scorecard

Each indicator was scored on a 0-5 ordinal scale based on observable evidence in model documentation, available interpretability tools, and reported evaluation protocols. Scores were assigned independently by two reviewers and averaged to reduce subjectivity. The transparency indicators, their descriptions, and corresponding scoring ranges are summarized in Table 1.

$$TS = \sum_{i=1}^{8} S_i. \quad (3)$$

where $S_i \in [0,5]$ and Max TS = 40.

TS ≥ 30: high transparency; TS = 20-29: moderate; TS < 20: low.

As shown in Table 2, diffusion models achieve higher transparency due to explicit sampling trajectories and step-wise generation processes: the aggregate transparency score increased from 19/40 for LLMs to 29/40 for diffusion models.

In contrast, LLMs exhibit limited latent traceability, resulting in lower overall transparency scores.

Table 3 summarizes the PRISMA corpus of the 20 included studies, mapping each reference to its architecture, XAI method, evaluation metric, and application domain. This table ensures alignment between the PRISMA flow, in-text citations, and the reference list.

4 CONCLUSIONS

The current state of explainability in Generative Artificial Intelligence (GenAI) remains inconsistent and weakly standardized, despite the rapid advancement and deployment of generative systems across text, vision, and multimodal domains. While GenAI models increasingly demonstrate high performance and practical utility, their internal decision-making processes remain opaque, limiting public trust, organizational reliance, and regulatory confidence particularly as these systems become embedded in high-stakes decision-support infrastructures such as healthcare, finance, law, and policy-making.

This study demonstrates that conceptualizing explainability as an incremental optimization process, rather than a binary property, enables systematic and sustainable progress toward trustworthy and accountable GenAI systems. By framing transparency as cumulative, the proposed approach aligns with the inherently probabilistic nature of generative models, allowing incremental gains in transparency to compound over time and improve interpretability, accountability, and governance readiness.

Empirical analysis reveals that current explainability approaches rely heavily on techniques adapted from discriminative models and are insufficient for capturing generative-specific behaviors. The findings highlight the importance of architecture-sensitive explainability, as generative architectures such as transformers and diffusion models require fundamentally different explanatory strategies. In addition, evaluation practices are shown to overemphasize output quality, often conflating model accuracy with explanation quality, which obscures meaningful assessment of transparency.

To address these challenges, this work introduces a framework integrating architecture awareness, multimodal semantic alignment, and stakeholder-centered requirements. The analysis further demonstrates that multimodal GenAI systems introduce interpretability challenges that cannot be addressed through single-modality explanations, underscoring the need for cross-modal transparency mechanisms.

Finally, the study identifies key directions for future research, including the development of generative-native explainability metrics, standardized evaluation protocols, and tools for inspecting latent-space dynamics and sampling behavior. By establishing an analytical foundation for incremental transparency, this work supports the responsible deployment of GenAI systems and aligns technical innovation with societal trust, ethical oversight, and regulatory governance. These findings align with recent advances in explainable and generative AI research [1], [7]-[9], [12].

REFERENCES

[1] G. Vilone and L. Longo, “A survey of explainable artificial intelligence in the era of foundation and generative models,” ACM Comput. Surv., early access, 2025.

[2] H. Liu, Y. Zhang, and M. Chen, “Explainability in diffusion models: Interpreting latent trajectories and sampling dynamics,” IEEE Trans. Pattern Anal. Mach. Intell., early access, 2024.

[3] J. Ren et al., "Large language models for generative artificial intelligence: A survey on security of usage," IEEE/CAA J. Autom. Sin., vol. 12, no. 11, pp. 2557-2576, 2025, [Online]. Available: https://doi.org/10.1109/JAS.2025.126332.

[4] Y. Jin et al., "Efficient multimodal large language models: A survey," Int. J. Data Sci. Anal., early access, 2025, [Online]. Available: https://doi.org/10.1007/s44267-025-00099-6.

[5] Z.-C. Zhang et al., "Large multimodal models evaluation: A survey," Sci. China Inf. Sci., early access, 2025.

[6] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Müller, "Explainable artificial intelligence: Concepts, methods, and evaluation," Nat. Mach. Intell., vol. 6, no. 1, pp. 1-15, 2024.

[7] F. Doshi-Velez, B. Kim, and A. Sudjianto, "Towards rigorous evaluation of interpretability methods for generative models," IEEE Trans. Artif. Intell., vol. 5, no. 2, pp. 123-137, 2024.

[8] L. Ranaldi, "Survey on the role of mechanistic interpretability in generative AI," Big Data Cogn. Comput., vol. 9, no. 8, art. no. 193, 2025, [Online]. Available: https://doi.org/10.3390/bdcc9080193.

[9] G. O. S. Olabiyisi, O. A. Egwutuoha, and G. R. D. C. Choo, "Trustworthy requirements and evaluation of foundation models: A survey and future directions," Eng. Appl. Artif. Intell., vol. 143, art. no. 110222, 2026, [Online]. Available: https://doi.org/10.1016/j.engappai.2025.110222.

[10] W. Jiang, "Enhancing transparency in large language models to achieve compliance and safe deployment," in Proc. ACM Conf. Inf. Technol. Social Good (GoodIT), 2025, pp. 1-5, [Online]. Available: https://doi.org/10.1145/3716554.3716597.

[11] Q. Chang et al., "2025 expert consensus on retrospective evaluation of large language model applications in clinical scenarios," Intell. Med., vol. 5, no. 4, pp. 318-330, 2025, [Online]. Available: https://doi.org/10.1016/j.imed.2025.09.001.

[12] N. Davis et al., "Explainable generative AI: A two-stage review of existing techniques and future research directions," AI, vol. 7, no. 1, art. no. 31, 2026, [Online]. Available: https://doi.org/10.3390/ai7010031.

[13] S. Jones, "Beyond fairness metrics: A practical audit-style framework for evaluating large language model bias," Front. Educ., vol. 10, art. no. 1574993, 2025, [Online]. Available: https://doi.org/10.3389/feduc.2025.1574993.

[14] X. Han et al., "Multimodal fusion and vision-language models: A survey for robot vision," Inf. Fusion, vol. 126, art. no. 103652, 2026, [Online]. Available: https://doi.org/10.1016/j.inffus.2025.103652.

[15] R. Kazmierczak, E. Berthier, G. Frehse, and G. Franchi, "Explainability and vision foundation models: A survey," Inf. Fusion, vol. 126, art. no. 103184, 2025, [Online]. Available: https://doi.org/10.1016/j.inffus.2025.103184.

[16] W. Yang et al., "A comprehensive survey on integrating large language models with structured knowledge-based systems," Knowl.-Based Syst., vol. 319, art. no. 113503, 2025, [Online]. Available: https://doi.org/10.1016/j.knosys.2025.113503.

[17] N. Jain, "Large language models: From black box to white box," Commun. Eng., vol. 4, art. no. 144, 2025, [Online]. Available: https://doi.org/10.1038/s44172-025-00327-4.

[18] N. Templin et al., "A framework for evaluating bias in generative AI," npj Digit. Med., vol. 8, art. no. 114, 2025, [Online]. Available: https://doi.org/10.1038/s41746-025-01400-5.

[19] Y. Saleh, M. Abu Talib, Q. Nasir, and F. Dakalbab, "Evaluating large language models: A systematic review of efficiency, applications, and future directions," Front. Comput. Sci., vol. 7, art. no. 1523699, 2025, [Online]. Available: https://doi.org/10.3389/fcomp.2025.1523699.

[20] S. M. Sajjadi Mohammadabadi et al., "A survey of large language models: Evolution, architectures, adaptation, benchmarking, applications, challenges, and societal implications," Electronics, vol. 14, no. 18, art. no. 3580, 2025, [Online]. Available: https://doi.org/10.3390/electronics14183580.