

A Comparative Study of Artificial Intelligence-Generated and Human-Produced Content in Text, Image, and Video Domains

Sanwar Bendsure¹, Yash Bhandari², Deven Katariya³, Rakhi Meshram⁴

¹ Sanwar Bendsure, Dept. of Artificial Intelligence & Data Science, AISSMS, Pune | sanwarbendsure60@gmail.com

² Yash Bhandari, Dept. of Artificial Intelligence & Data Science, AISSMS, Pune | yashhhbhandari4@gmail.com

³ Deven Katariya, Dept. of Artificial Intelligence & Data Science, AISSMS, Pune | devenkatariya2005@gmail.com

⁴ Rakhi Meshram, Dept. of Artificial Intelligence & Data Science, AISSMS, Pune | rakhi.meshram@aissmsioit.org

Abstract – The proliferation of generative artificial intelligence has fundamentally altered the content creation landscape, enabling mass production of synthetic text, images, and video at unprecedented scale and fidelity. This paper presents a comprehensive, structured comparative analysis of AI-Generated Content (AIGC) and Human-Produced Content (HPC) across three primary modalities: text, image, and video. Each modality is evaluated along seven dimensions — Veracity, Engagement, Realism, Intent Coherence, Temporal/Structural Integrity, Authenticity Perception, and Societal Impact — forming the novel VERITAS evaluation framework. Key findings reveal that AI-generated text achieves near-parity in surface fluency yet lags in long-form reasoning and factual grounding, with hallucination rates ranging from 0.7% to over 10% across leading models. AI-generated images, driven by diffusion architectures such as FLUX.1 and Stable Diffusion 3.5, achieve photorealism sufficient to reduce human detection accuracy to 63.7%. AI-generated video, exemplified by Google Veo 3 and OpenAI Sora, struggles with physical plausibility and long-range temporal coherence despite achieving short-form realism. Deepfake videos grew 900% between 2023 and 2025, while consumer preference for human-created content rebounded to 74% in 2026. We introduce the TRACE detection taxonomy, propose policy and technical recommendations, and identify open research challenges in cross-modal forensics, adversarial robustness, and the preservation of human creative value.

Key Words: AI-generated content, human-produced content, generative AI, text generation, image synthesis, video generation, deepfakes, content authenticity, hallucination, diffusion models, VERITAS framework, TRACE taxonomy.

1. INTRODUCTION

The emergence of large-scale generative artificial intelligence has initiated one of the most consequential transformations in the history of human communication and creative production. For millennia, generating meaningful content in any modality — written text, visual imagery, moving image — was an exclusively human enterprise constrained by biological limits of time, skill, and creative bandwidth. The deployment of foundation models trained on internet-scale corpora has shattered these constraints, enabling the automated production of synthetic content that is, in many measurable dimensions, indistinguishable from human authorship.

The global generative AI content creation market, valued at \$14.8 billion in 2024, is projected to reach \$80.12 billion by 2030 at a compound annual growth rate (CAGR) of 32.5%. The AI video generator market

alone reached \$665.33 million in 2024 and is expected to exceed \$2.84 billion by 2034 [1]. Three content modalities — text, image, and video — have been most profoundly affected, each presenting a distinct profile of capability, limitation, and societal consequence. ChatGPT reached 100 million users within two months of launch, while FLUX.1, Stable Diffusion 3.5, and DALL·E 3 democratised photorealistic image synthesis. OpenAI Sora, Google Veo 3, and Runway ML Gen-3 extended this to temporally coherent video generation.

The pace of technological development has outpaced our collective understanding of the meaningful differences between AIGC and HPC — differences that matter for quality assurance, public trust, democratic discourse, intellectual property, and the economic viability of human creative labour. This paper addresses this gap with the following contributions:

- The VERITAS framework: a seven-axis evaluation model applicable uniformly across all three content modalities
- A cross-modal comparative analysis grounded in 2024–2026 empirical studies and human perception experiments.
- The TRACE detection taxonomy: a structured five-category classification of AIGC detection approaches and their limitations.
- Identification of open research challenges and a forward-looking agenda for the field.

2. THE VERITAS EVALUATION FRAMEWORK

A central problem in comparing AIGC and HPC is the absence of a unified, multi-dimensional evaluation lens. Prior work typically reduces quality to single scalar metrics — FID score for images, BLEU/ROUGE for text, FVD for video — that are necessary but insufficient for capturing the full spectrum of meaningful differences. We propose the VERITAS framework comprising seven evaluation axes:

V — Veracity: The factual accuracy and logical consistency of generated content relative to ground truth. For text, this encompasses hallucination rate and citation reliability; for image, physical realism; for video, temporal physics consistency and causal coherence.

E — Engagement: Measurable audience interaction including click-through rate, dwell time, emotional arousal, and social sharing behaviour. Engagement is modality-specific: AI-generated image advertisements may achieve higher CTR than human-made equivalents while AI documentary video fails to sustain viewer attention.

R — Realism: Perceptual authenticity as judged by human evaluators and automated forensic classifiers under controlled experimental conditions. Realism is distinct from quality: content can be technically excellent yet clearly AI-generated.

I — Intent Coherence: Alignment between the creator’s communicative intent — whether expressed as a prompt, brief, or creative directive — and the final output. Failures manifest as prompt misinterpretation, attribute confusion, or narrative drift.

T — Temporal/Structural Integrity: Consistency over time in video, or over long-form structure in text and multi-scene imagery. This axis captures hallucination cascades in long-form LLM text, identity drift across video frames, and spatial inconsistencies in composite imagery.

A — Authenticity Perception: The audience’s belief in the human origin of content and the trust assigned as a consequence. Authenticity Perception is empirically distinct from Realism: audiences may correctly identify AI content yet still find it credible, or misclassify it as human-made yet assign lower credibility upon disclosure.

S — Societal Impact: Aggregate downstream effects on misinformation ecosystems, creative labour markets, intellectual property regimes, cultural homogenisation, and democratic processes at population scale.

3. COMPARATIVE ANALYSIS: TEXT DOMAIN

3.1 Architectures and Capabilities

Contemporary large language models — GPT-4o, Claude 3.7 Sonnet, Gemini 2.0 Ultra, and Llama 3.3 — represent the state of the art in AI text generation. Built on transformer architectures with hundreds of billions of parameters and trained via unsupervised pre-training plus Reinforcement Learning from Human Feedback (RLHF), these models exhibit remarkable surface-level linguistic competence across virtually any requested style, domain, or format. Automated evaluation metrics consistently place top-tier LLMs within the upper quartile of human writing performance for short-form, structured content categories including product descriptions, news summaries, legal briefs, and email drafting [2].

3.2 Quality Ceiling and Failure Modes

Surface-level equivalence degrades sharply across three conditions: increasing document length, increasing domain specificity, and increasing adversarial scrutiny. Hallucination remains the most clinically significant failure mode. LLMs are fundamentally prediction engines that generate the most statistically likely next token conditioned on context — they do not understand truth but predict

plausibility [3]. A critical MIT finding from 2025 established that when AI models hallucinate, they use more confident language: models were 34% more likely to use affirmative phrases like “definitely” and “certainly” when generating incorrect information [3]. This confidence-accuracy inversion represents a fundamental epistemic hazard for high-stakes applications.

Quantitative benchmarking using the Vectara Hughes Hallucination Evaluation Model (HHEM) leaderboard reveals significant variation across model families. Google Gemini-2.0-Flash achieves the lowest rate at 0.7%, GPT-4 variants range from 0.8% to 2.0%, while some open-weight models exceed 10% on the refreshed 7,700-article benchmark [3]. Sycophantic drift — where RLHF-trained models adjust factual claims to match perceived user expectations — and structural flattening in long-form text represent additional systematic failure modes that distinguish AI from human writing at the architectural level.

3.3 Detection and Volume Parity

Automated detection of AI-generated text has achieved 94–99% classification accuracy via cross-linguistic transfer-learning models employing mBERT and XLM-RoBERTa variants [4]. Yet human readers without specialised training cannot reliably identify AI-generated text at better than chance. An analysis of 65,000 web URLs published between 2020 and 2025 found that AI-generated articles surpassed human-written articles in volume in November 2024, remaining at parity through mid-2025 [5]. A 2022 Europol report had forecast 90% AI content saturation by 2026; actual trajectory suggests a more gradual but structurally significant shift [1].

4. COMPARATIVE ANALYSIS: IMAGE DOMAIN

4.1 Generative Architectures

AI image generation has progressed through three paradigmatic phases. Generative Adversarial Networks (GANs), pioneered in 2014, enabled photorealistic face synthesis but suffered from training instability and mode collapse. Variational Autoencoders (VAEs) offered cleaner probabilistic latent spaces but produced blurrier outputs. Denoising Diffusion Probabilistic Models (DDPMs) now represent the dominant paradigm, transforming random noise into coherent images through iterative

refinement with superior semantic alignment and visual consistency [6]. The current state of the art — FLUX.1 (Black Forest Labs), Stable Diffusion 3.5 (Stability AI), and DALL·E 3 (OpenAI) — employs Diffusion Transformer (DiT) architectures replacing convolutional U-Net layers with self-attention mechanisms, achieving the highest FID, CLIP Score, and Human Preference Score (HPS) benchmarks of 2025.

4.2 Human Perception Studies

A controlled perception experiment using 32 stimulus images across four categories (human portraits, landscapes, everyday scenes, detailed objects) generated via Midjourney, DALL·E 3, and Adobe Firefly found a statistically significant effect of origin on realism perception specifically for human portrait stimuli ($p < 0.001$), with human-made portraits rated as more realistic [7]. A Frontiers in AI study employing more sophisticated models — FLUX.1-dev, Stable Diffusion 3.5, and Midjourney v6 — found an overall human recognition accuracy of 63.7% for landscapes, architecture, and interiors [8]. While statistically above the 50% chance baseline, this modest detection advantage indicates that the perceptual gap is closing rapidly. A larger study involving 749,828 observations from 50,444 participants found symmetrical recognition accuracy of 76% for synthetic and 74% for authentic photographs [8], suggesting the task has become a general test of perceptual attention rather than a specific AI-detection skill.

4.3 Engagement vs. Authenticity: Format-Dependent Reversal

A 2026 Springer e-commerce study directly comparing AI-generated versus human-made advertising content found a striking format-specific result: AI-generated image advertisements outperformed human-made image ads on customer engagement metrics, while human-produced video advertisements outperformed AI equivalents [9]. This reversal is theoretically significant: AI systems optimised for visual composition achieve higher engagement-maximising scores for static imagery, while the temporal, relational, and emotional dimensions of video continue to favour human creative direction.

5. COMPARATIVE ANALYSIS: VIDEO DOMAIN

5.1 Architectural Evolution and State of the Art

AI video generation has progressed through three generations. First-generation systems (2020–2022) employed RNNs and GAN-based frame generation, producing low-resolution clips with severe temporal inconsistencies. Second-generation systems (2022–2024) introduced latent diffusion applied to video, exemplified by ModelScope Text2Video. Third-generation systems (2024–2026) — OpenAI Sora, Google Veo 3, Runway ML Gen-3, and Kling — apply diffusion transformer architectures with world-model training objectives, achieving photorealistic fidelity in clips of 30 seconds to several minutes [10]. Google Veo 3 (DeepMind, 2025) represents the current apex, generating long-form high-definition clips with consistent subject identity, dynamic camera motion, audio-visual alignment, and strong physical realism. OpenAI Sora employs spacetime diffusion transformers treating video as sequences of spatiotemporal patches for coherent long-range temporal modelling. Runway ML Gen-3 and Pika 1.5 provide commercially deployed browser-accessible systems that have democratised AI video creation for non-technical creators [10].

5.2 Persistent Limitations

Despite remarkable progress, three fundamental limitation categories persist. Physical plausibility failures are common: fluids exhibit impossible dynamics, rigid bodies interpenetrate, lighting changes direction inconsistently across cuts. Quality assessment research has introduced specific evaluation methods for geometric consistency across frames, noting that standard metrics such as FID and FVD fail to penalise these violations [11]. Long-range temporal coherence breakdown degrades character identity, environmental continuity, and narrative logic as clip duration exceeds 60 seconds in most current systems. Finally, the absence of semantic world modelling — grounded knowledge of physical causality, social context, and spatial logic — produces scenes that violate common sense in ways that human directors routinely avoid.

5.3 The Deepfake Crisis

The most consequential application of AI video generation is synthetic face-swap and identity impersonation. Cybersecurity firm DeepStrike

estimates deepfake videos grew from 500,000 in 2023 to 8 million in 2025 — a 900% increase — while human detection accuracy for high-quality deepfakes has been measured at approximately 25%, barely above chance [1]. Studies document that repeated exposure to deepfake content increases its perceived credibility through familiarity effects, constituting a structural threat to video as an evidentiary medium in legal, journalistic, and electoral contexts.

6. SUMMARY COMPARISON TABLE

Table-1: VERITAS Framework — AIGC vs. Human-Produced Content (2025–2026)

VERITAS Axis	Text
Veracity	AI < Human
Engagement	AI $\approx$ Human
Realism	AI < Human
Struct. Integrity	AI < Human
Detection Acc.	94–99%
Consumer Pref.	Human: 74%
Deepfakes	N/A

* Human portraits: statistically significant ($p < 0.001$)

7. THE TRACE DETECTION TAXONOMY

We propose the TRACE taxonomy to classify AIGC detection approaches across five categories, each with distinct applicability, accuracy profile, and adversarial robustness:

T — Technical Artifact Detection: Exploits low-level statistical signatures of generative models — GAN spectral artifacts, diffusion noise patterns, LLM token probability distributions. Achieves 94–99% accuracy in text, 85–92% in image, and 78–88% in video under standard benchmark conditions, but is sensitive to post-processing obfuscation [12].

R — Reasoning Consistency Analysis: Evaluates whether content maintains logical, causal, and factual consistency across long-form sequences. Identifies hallucination cascades, causal contradictions, and world-state inconsistencies characteristic of autoregressive generation without world modelling [3].

A — Adversarial Probing: Queries suspected AI-generated text with follow-up questions requiring grounded recall, specific personal experience, or domain-expert reasoning. AI systems

characteristically fail to provide consistent, detailed, verifiable answers that a genuine human author could provide.

C — Cross-Modal Consistency Verification:

Detects inconsistencies between simultaneously AI-generated text, image, and audio: lip-sync errors in video, mismatch between described scene and generated visual, or audio not matching visual environmental cues. Uniquely applicable to multimodal AIGC.

E — Embedding Forensics: Applies machine learning classifiers trained on distributional properties of human versus AI-generated content in high-dimensional embedding spaces. Achieves higher robustness to surface-level paraphrasing than token-probability approaches but requires large labelled training datasets.

8. ETHICAL AND SOCIETAL DIMENSIONS

8.1 The Authenticity Crisis

Consumer attitudes towards AI-generated content have undergone a dramatic reversal. Consumer preference for AI-generated content peaked at 60% in 2023, driven by novelty and early adopter enthusiasm, collapsing to 26% by 2026 as audiences explicitly prioritised human-created content as a signal of authenticity and brand trustworthiness [1]. This inversion was catalysed by the proliferation of “AI slop” — low-quality, high-volume AI-generated content optimised for search engine manipulation. The term was named Word of the Year 2025 by both Merriam-Webster and the Australian National Dictionary [1]. Research estimates that 21–33% of YouTube’s content feed may consist of AI slop, generating approximately \$117 million annually in advertising revenue from content culturally empty by design.

8.2 Intellectual Property and Labour Displacement

Generative AI models trained on human-created content raise fundamental intellectual property questions existing legal frameworks were not designed to address. Training data provenance, copyright ownership of AI-generated outputs, and liability for reproducing copyrighted training material are under active litigation in multiple jurisdictions as of early 2026. The GDC 2026 survey found that 52% of game industry professionals believe generative AI is having a negative impact, up from 30% in 2025

[13], driven primarily by concerns about creative labour displacement and undisclosed AI content. The displacement is asymmetric: entry-level and freelance creative workers face maximum income compression while senior creative professionals retain their irreducible strategic value.

8.3 Misinformation and Democratic Integrity

AI-generated content poses risks to the epistemic foundations of democratic public discourse. Coordinated inauthentic behaviour — the automated generation of politically motivated content by AI-powered account networks — has been documented across major social media platforms in multiple electoral contexts since 2024. The integration of AI text generation with AI image synthesis and AI voice cloning enables end-to-end synthetic media production pipelines requiring minimal human creative input while producing maximally persuasive political content [1]. Proposed legislation in multiple U.S. states seeks to establish criminal penalties for identity theft and extortion conducted via synthetic media, acknowledging that existing legal frameworks were not designed for AI-generated impersonation at scale.

9. CROSS-MODAL SYNTHESIS

The comparative analysis reveals three structural convergences across all three modalities. First, surface quality has largely converged while deep quality retains significant human advantage. AI systems have achieved near-parity in the dimensions most easily quantified — fluency, resolution, visual sharpness — while maintaining substantial deficits in reasoning depth, emotional authenticity, physical plausibility, and sustained structural coherence. This divergence between measurable and meaningful quality is the central epistemological challenge for evaluation methodology.

Second, detection difficulty scales inversely with modality computational cost. Text generation is least expensive and most detectable (94–99%); image generation is moderate in both dimensions (85–92% detection, 63.7% human accuracy); video generation is most expensive and least reliably detected (78–88% automated, 25% human accuracy for deepfakes). This scaling implies that as computational costs fall, detection challenges in all modalities will intensify simultaneously.

Third, the most theoretically significant finding is the systematic divergence between AIGC quality (technical merit) and AIGC authenticity perception (audience trust). AI content can be technically superior — higher resolution, more grammatically correct, more visually polished — while simultaneously perceived as less credible and valuable by audiences informed of or correctly inferring its AI origin. Quality metrics alone are therefore fundamentally insufficient for evaluating the social impact of AIGC, motivating the VERITAS framework's explicit inclusion of Authenticity Perception (A) as a first-class evaluation axis distinct from Realism (R).

10. OPEN CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Cross-Modal Detection and Unified Forensics:

Current AIGC detection systems are modality-specific, with no shared architecture or cross-modal reasoning. As multimodal generative systems become standard, unified cross-modal forensic architectures — with shared embedding spaces, multimodal consistency verification, and watermarking protocols persisting across transformations — are urgently required [12].

Adversarial Robustness: Current detection systems fail under post-processing obfuscation: compression defeats image forensics, paraphrasing defeats text detectors, re-encoding damages video fingerprints. Detection systems with formal robustness guarantees against adversarial obfuscation represent a critical open problem.

Standardised Disclosure and Provenance Tracking: Technical solutions for reliable, tamper-resistant provenance tracking — including Content Credentials (C2PA), cryptographic watermarking, and model fingerprinting — are under active development but not yet universally deployed. Provenance systems must survive the content transformation pipeline across screenshot, crop, compress, and reshare operations.

Quantum-Accelerated Generation: Quantum-classical hybrid computing architectures may substantially relax the computational constraints limiting AI video generation to short clips. If quantum acceleration achieves practical advantage for video generation, current physical constraints on deepfake quality and duration will be eliminated, dramatically

compressing the timeline to arbitrary-content long-form video synthesis.

Preserving Human Creative Value: The deepest long-term question is economic and cultural: as AI achieves functional parity across more content categories, what institutional mechanisms will preserve the viability and cultural centrality of human creative labour? Proposed interventions span technical (biometric content attestation), economic (AI content taxes), and regulatory (platform-level human content quotas) dimensions, with no consensus yet emerging [13].

11. CONCLUSIONS

This paper has presented a comprehensive comparative analysis of AI-generated and human-produced content across text, image, and video domains, evaluated through the VERITAS seven-axis framework and organised by the TRACE detection taxonomy. The evidence reveals a nuanced, modality-specific landscape: AI has achieved functional superiority or near-parity with human creators in several specific format categories and measurable quality dimensions while retaining systematic deficits in factual accuracy under adversarial conditions, long-form structural coherence, physical plausibility in video, and the perceptual authenticity that audiences associate with human intentionality.

The most urgent challenge is not at the frontier of model capability but in the social infrastructure surrounding AI-generated content: the detection systems that identify it, the disclosure regimes that label it, the legal frameworks that govern it, and the economic structures that determine whether human creative production remains viable in its presence. The VERITAS and TRACE frameworks offered here are contributions to the operational vocabulary this interdisciplinary challenge requires. As generative AI systems continue their rapid capability development, the research community faces an obligation not only to document the expanding frontier of AIGC quality but to ensure this expansion occurs within a social architecture that preserves the trust, authenticity, and diversity that make communication meaningful.

. ACKNOWLEDGEMENT

The authors thank the open-access research communities at IEEE, Springer Nature, Frontiers in AI, PMC, and ACM whose empirical studies formed the empirical backbone of this analysis. The authors also acknowledge the guidance and institutional support provided by the Department of Artificial Intelligence and Data Science, AISSMS College of Engineering, Pune.

. REFERENCES

- [1] Future Center UAE, "Digital pollution: Trends in AI-generated content in 2026," Future Center for Advanced Research and Studies, Mar. 2026. Available: <https://futureuae.com/en-US/Mainpage/Item/10785>
- [2] H. Al-Khassawneh et al., "Human vs. machine: A comparative study on the detection of AI-generated content," *ACM Trans. on Asian and Low-Resource Language Information Processing*, 2025. DOI: 10.1145/3708889.
- [3] Suprmind Research, "AI hallucination statistics: Research report 2026," Suprmind, Mar. 2026. Available: <https://suprmind.ai/hub/insights/ai-hallucination-statistics-research-report-2026/>
- [4] S. Liu et al., "A hallucination detection and mitigation framework for faithful text summarization using LLMs," *Scientific Reports*, vol. 16, Art. no. 1374, Jan. 2026. DOI: 10.1038/s41598-025-31075-1.
- [5] J. Pinho, "AI-written web pages haven't overwhelmed human-authored content, study finds," *Axios*, Oct. 2025. Available: <https://www.axios.com/2025/10/14/ai-generated-writing-humans>
- [6] L. Zhang et al., "Comprehensive exploration of diffusion models in image generation: A survey," *Artificial Intelligence Review*, Springer, Jan. 2025. DOI: 10.1007/s10462-025-11110-3.
- [7] C. Andreu-Sánchez et al., "Interpretation of AI-generated vs. human-made images," *Journal of Imaging*, vol. 11, p. 227, Jul. 2025. DOI: 10.3390/jimaging11070227.
- [8] M. Fischer et al., "What you see is not what you get anymore: A mixed-methods approach on human perception of AI-generated images," *Frontiers in Artificial Intelligence*, Nov. 2025. DOI: 10.3389/frai.2025.1707336.
- [9] A. El Assadi, "AI vs. human creativity: The impact of text-to-image and text-to-video ads on customer engagement," *Electronic Commerce Research*, Springer, Feb. 2026. DOI: 10.1007/s10660-026-10103-w.
- [10] Y. Chen et al., "Bibliometric analysis and review of AI-based video generation: Research dynamics and application trends (2020–2025)," *Discover Computing*, Springer, Jun. 2025. DOI: 10.1007/s10791-025-09628-9.
- [11] T. Katsenou et al., "Quality prediction of AI generated images and videos: Emerging trends and opportunities," *arXiv*, Oct. 2024. Available: <https://arxiv.org/html/2410.08534v1>
- [12] Z. Huang et al., "A sanity check for AI-generated image detection," in *Proc. ICLR 2025*, 2025. Available: https://proceedings.iclr.cc/paper_files/paper/2025/file/b0303773962ea1b5394c3a83cc7dd066
- [13] S. Hemraj, "AI and the future of creative development: Redefining digital media production," *AI and Ethics*, Springer, vol. 5, pp. 5105–5119, Jun. 2025. DOI: 10.1007/s43681-025-00765-x.