Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
Special Issue on Reimagining Visual Arts, Media, and Cultural Pedagogy in the Age of Artificial Intelligence 2025 6(1s)

NLP MODELS FOR ARTISTIC STATEMENT GENERATION

Dr. R.M. Gomathi [1], Pooja Srishti [2], Prateek Garg [3], Dr. Roselin [4], Dr. Hemal Thakker [5], Sumeet Kaur [6]

[1] Associate Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India
[2] Assistant Professor, School of Business Management, Noida international University, India
[3] Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh, Solan, India
[4] Associate Professor, Department of Computer Science and Engineering, Presidency University, Bangalore, Karnataka, India
[5] Associate Professor, ISME - School of Management and Entrepreneurship, ATLAS Skill Tech University, Mumbai, Maharashtra, India
[6] Centre of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India

Received 20 January 2025
Accepted 15 April 2025
Published 10 December 2025

Corresponding Author
Dr. R.M. Gomathi,
gomathi.it@sathyabama.ac.in

DOI
10.29121/shodhkosh.v6.i1s.2025.6673

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

In this paper we propose a multimodal vision-language multimodality detransformer framework for coherent, expressive, and visually grounded artistic statement generation, which takes advantage of multimodal vision- and language modeling on top of a strong transformer-based text generation network. The proposed system is comprised of a visual encoder to interpret compositional and stylistic aspects of an artwork, a fine-tuned transformer decoder that acts as a conceptually rich story engine and a cross-modal fusion module to ensure the alignment between visual clues and linguistic output. Combined with creative and grounding-based reward mechanisms from reinforcement learning, the interpretive depth and style-grounding are further advanced. Using automated similarity measures, multimodality alignment scores and human expert subjectivity measurement, it is shown that the hybrid model greatly improves over traditional captioning and text-only methods at extracting artistry, emotionality and conceptuality. While the method has great potential, challenges exist in dealing with cultural bias, data limitations, interpretive subjectivity and computational demands. Overall, the research brings forward the field of AI-assisted artistic communication and provides a scalable solution to help artists, curators, educators, and digital art platforms to create quality artistic statements.

Keywords: Natural Language Processing, Generative AI, Reinforcement Learning, Vision Language Fusion, Creativity, Transformer Models, Parallel Multimodal Learning, Creative Text Generation

1. INTRODUCTION

Artistic statements are an important narrative element of visual and creative processes, which provide a point of connection between the internal world of the artist and the interpretive experience of the audience. They explain the

conceptual motivations, emotional resonance, thematic inspiration and cultural significance of a piece of art. Whether as part of gallery exhibitions, as part of an academic portfolio, in grant applications, on the online art market, or in digital collections, these statements help to understand the creative process from the artist's perspective Pham et al. (2023). However, despite their importance, many artists, particularly emerging artists, non-native English speakers and visually-oriented artists struggle to express their intentions clearly and coherently in writing. This challenge has led to a growing body of academic research on computational systems to support or automate the creation of artistic statements Peng (2022). Parallel to this demand, the field of Natural Language Processing (NLP) has been transformed in the last decade. The result has been precisely to transform traditional rule-based and statistical tools into advanced deep learning architectures that are able to produce expressive and contextually rich text. Particularly, the development of transformer-based language models (GPT, T5, BART and multimodal models like CLIP-based architectures) have enabled machines to comprehend artistic language, generate narratives of a particular style and interpret a visual input in a conceptual manner. These models can synthesize descriptions that go beyond a superficial commentary, understanding deeper levels of symbolic, emotional and aesthetic meaning of artworks. With the emergence of generative AI in creative spaces, the generation of artistic statements is a promising intersection of computational linguistics, computer vision and creative practice Wu (2020), Minaee et al. (2021).

Figure 1

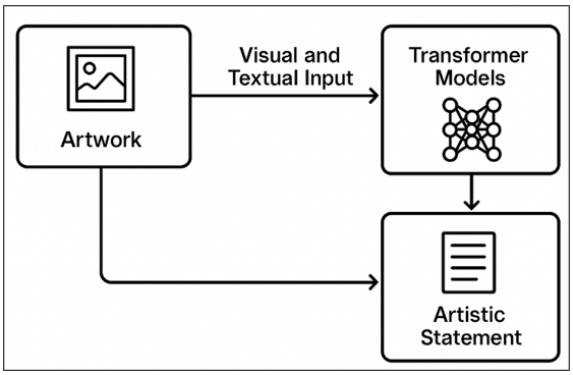

Figure 1 Basic Block Diagram of NLP Model in Art Statement Design

The advantage of using NLP-based systems is that we can automate the generation of coherent, style-sensitive, genre-sensitive, and culturally ad hoc statements, which make the solution scalable. Such systems can also be used to foster learning environments, where the student's creative intention can be expressed more effectively while learning the language of visual expression Kim and Hardin (2021).

This research can mitigate such challenges by exploring the potential of modern NLP models, specifically the transformer-based models and multimodal models, to be leveraged for the creation of high-quality artistic statements Sarwar (2023). The paper discusses previous approaches and proposes existing shortcomings and suggests a multimodal generative framework, which is then evaluated using both automated linguistic measures and the judgment of human experts. The findings provide insights into the role that AI will play in creating writing, and opportunities for meaningful human-AI collaboration in the arts.

2. BACKGROUND STUDY

The process of creating artistic statements by using computational methods is at the boundary of natural language processing, multimodal learning and creative AI. This section serves to conduct a review on the development of relevant research, including traditional language models, deep learning architectures, multimodal architectures, and the previous work in creative text generation and art interpretation Wang (2021). Together, these strands give a base for comprehension of the potential and shortcomings of NLP based artistic statement generation. Early efforts in automatic text generation have been based on rule-based systems, template-filling and symbolic grammars. The rest were created using predefined sentence structures and lexicons for that domain. While deterministic and easy to implement, they lacked expressive depth, stylistic flexibility and contextual sensitivity - all important when it comes to artistic writing Došilović et al. (2018). Computational creativity studies in the late 1990s and early 2000s tried to simulate artistic

commentary using handcrafted rules, but produced repetitive, generic and incapable of capturing complex emotions and symbolic meaning.

Table 1

Table 1 Comparison of NLP Approaches for Artistic Text Generation

Approach Type	Strengths	Limitations	Suitability for Artistic Statements
Rule-based / Template Systems Peng (2022)	Highly controlled, consistent output	Rigid, lacks creativity	Very Low
Statistical Models (N-gram, HMM)	Data-driven, simple	Poor long-range context	Low
Word Embedding + Neural Networks	Better semantics and flow	Limited conceptual reasoning	Moderate
RNN/LSTM/Seq2Seq Joshi et al. (2020)	Stronger coherence than statistical models	Struggles with abstraction; limited stylistic control	Moderate
Transformer Models (GPT, BART, T5)	High coherence, creativity, interpretive capacity	Requires fine-tuning; may hallucinate	High
Multimodal Models (CLIP, BLIP, ViLBERT) Wang et al. (2020)	Integrates visual + textual cues; best conceptual grounding	Computationally heavy; dataset dependent	Very High

The advent of statistical nLP was a step in the direction of corpus-based text generation. N-gram models, probabilistic grammars and Hidden Markov Models were enabling systems to learn patterns from the descriptions of artworks instead of templates being the only thing to work with. However, these methods had their limitations including the inability to model long-range dependencies and nuances of creative writing Deshpande et al. (2021). The appearance of distributed representation models such as Word2Vec, GloVe and FastText boosted the understanding of semantics by representing relationships between artistic terms - for example, "texture", "composition", "symbolism". Neural architectures like Recurrent Neural Networks (RNN) and Long Short Term Memory (LSTM) networks contributed greatly to increasing the coherence and flow. Sequence-to-sequence models are used in the early artwork description systems to make connections between image features and narrative text. Nevertheless, conceptual depth, metaphoric reasoning and stylistic variation needed for making artistic statements eluded them Aldekhail and Almasri (2022).

Table 2

Table 2 Gaps Identified in Existing Literature

Gap Category	Specific Gap	Impact
Conceptual Depth	Existing systems focus on literal aspects	Limits ability to express artistic intention
Long-Form Generation	Few models create multi-paragraph statements	Reduces narrative coherence
Cultural Adaptation	Limited awareness of regional art forms	Risk of misinterpretation
Dataset Limitations	Insufficient artist-specific corpora	Affects authenticity and style
Multimodal Fusion	Weak integration of symbolism and emotion	Leads to shallow interpretations

These beyond capabilities are important for creating statements that go beyond the surface. Although progress has been made, there are major gaps. Most captioning systems create literal rather than conceptual descriptions. Multi-paragraph, reflective, artistic statements are covered in few studies Devlin et al. (2019). Anthropological statement generation--including bias, cultural translation, and artistry--has not been well studied, especially for culturally adaptive or genre-specific statement generation. These gaps are good reasons to need a specific multimodal NLP framework that can give context-aware, expressive, and in the concept log sense rich artistic statements.

3. EXISTING NLP APPROACHES FOR ARTISTIC TEXT GENERATION

Artistic text generation performance is based on the broad range of Natural Language Processing (NLP) approaches adding different amounts of linguistic richness, contextual depth, and stylistic flexibility Zhao et al. (2022). Existing approaches can be broadly divided into rule-based approaches, statistical language models, neural network architectures and transformer-based/ multimodal models. This is necessary for being able to identify their capabilities and limitations for producing expressive, conceptual and culturally embedded artistic output.

1) Rule-Based and Template-Driven Systems

Rule-based approaches are the oldest approaches to the automatic description of artwork. These systems are based on manually written rules of grammar, domain-specific lexicons and static sentence patterns to generate structured output Bozyiğit et al. (2021).

$$maxLMLE = E[t = 1\sum Tlogp\Theta(yt \mid y < t, I, m)].$$

With label smoothing ϵ

$$LCE = -t\sum w \in V\sum qt(w)logp\Theta(w \mid y < t, I, m), qt = (1 - \epsilon)\delta yt + \mid V \mid \epsilon.$$

In the case of artistic statement generation, models based on rules can be used to describe simple features (for example, color, shape or medium), but do not capture deeper emotional or conceptual layers. They are not creative and personal as the results are deterministic and repetitive. In spite of their shortcomings, they are still of relevance for constrained settings where interpretability and control are more important than expressiveness.

2) Statistical Language Model

Data-driven patterns were introduced into the generation of text by Statistical Methods such as n-gram models and Hidden Markov Models (HMMs). By learning word sequence probabilities from artistic corpus, these models were able to improve on fluency when compared to rule-based systems [17].

Learning reduces to estimating a discrete policy over templates:

$$k^{=argkmaxp(k|m)}, p(Y \mid I, m) = k\sum p(k \mid m)1\big[Y = Tk\big(\phi(I, m)\big)\big].$$

However, their dependence on the local context does not allow coherently forming long-form statements. Artistic text, which frequently embodies the need for metaphorical language, multi-sentence cohesion, and interpretive reasoning among others, cannot be appropriately modelled based on these statistical methods. As a result, they function less as practical tools for the generation of artistic narrative, and more as historical foundations for it.

3) Word Embeddings and Distributed Representations

The development of distributed representation models like Word2Vec, GloVe and FastText greatly improved semantic understanding in NLP. These embeddings learn about the relationship between artistic terms (e.g. compositional techniques, symbolisms, emotional cues) in order to be used for a more meaningful text generation effort.

n-gram LM: $p(Y) = t = 1\prod Tp(yt \mid yt - n + 1 : t - 1)$,

$$p(wi \mid h) = \sum w \in Vcount(h, w)count(h, wi).$$

HMM (latent style states zt: $p(Y) = z1 : T\sum t = 1\prod Tp(zt \mid zt - 1)p(yt \mid zt)$.

Trained by EM, suffers from short memory. Embedding-based models set the stage for neural networks to produce more expressive narratives, but they are not yet sophisticated enough in terms of structure of artistic statements for multi-paragraph narratives.

Given corpus tokens (w,c) with window N(w),

$$maxw\sum c \in N(w)\sum log\sigma(uc\top vw) + c' \sim Pn\sum log\sigma(-uc'\top vw).$$

Their main contribution is to be able to enrich the semantic layer that the downstream models can take advantage.

4) Recurrent Neural Networks and Sequence-to-Sequence Models

Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTMs) networks and sequence-to-sequence (Seq2Seq) architectures were a pivotal move towards having more coherent generative text systems.

Encoder (text prompt or image tags ~1,…,x~S):

$$hs = LSTMenc(x{\sim}s, hs - 1).$$

Also, these models balance temporal dependencies more effectively than statistical ones, which allows generating paragraphs with smoother transitions.

Attention at decoder step ttt:

$$et, s = v\top \tanh(Whhs + Wsst - 1)\,, \alpha t,$$
$$s = s'\exp \sum(et, s')\,\exp(et, s)\,,$$
$$ct = s\sum \alpha t, shs.$$

Decoder: $st = LSTMdec([E(yt - 1);$

$$ct], st - 1),$$
$$p(yt\ |\cdot) = softmax(Wost + bo).$$

In combination with attention mechanisms, Seq2Seq models have improved the generation of interpretive text, such as a description of stylistic characteristics or emotional, e.g. on artworks. However, they have trouble with global coherence, performing metaphors, and upholding a constant artistic voice over longer statements, making them less effective in a professional or academic setting.

4. PROPOSED ALGORITHM: HYBRID VISION-GUIDED GENERATIVE LANGUAGE MODEL (VG-GLM)

Step 1] Visual Encoder

Extract patch embeddings (ViT/CLIP-Vision):

$$ZI = fV(I; \Theta V) \in RP \times dv, ZI = [z1, \dots, zP]\top.$$

Step 2] Textual Conditioning Inputs

Build the textual context matrix X0X_0X0 from tokens of prompt/metadata plus style:

$$X0 = [E(m);\ 1s\top] \in RL0 \times d,$$
$$where\ E(\cdot)\ is\ token\ embedding\ layer\ and\ s = Emb(s)$$

Step 3] Decoder State Update (Transformer LM)

For autoregressive step t, $Xt = [X0; E(y1{:}\, t - 1)]$.Apply masked self-attention block(s) to obtain hidden state ht.

$$Ht = BlockTF(Xt; \Theta T), ht = Ht[-1].$$

Step 4] Cross-Modal Fusion (Grounding in the Image)
Use cross-attention from ht to visual keys/values K=V=ZI:

$$\alpha t, i = \sum jexp(htWq(zjWk)\top/d)exp(htWq(ziWk)\top/d), ct = i = 1\sum P\alpha t, i(ziWv).$$

Fuse and project to token distribution:

$$h\sim t = \phi([ht; ct]), p\Theta(yt \mid y < t, I, m, s) = softmax(Woh\sim t + bo).$$

Step 5] Maximum Likelihood Objective (Supervised)
With label smoothing ϵ\epsilonϵ:

$$LCE = -t = 1\sum Tw \in V\sum qt(w)logp\Theta(w \mid y < t, I, m, s), qt = (1 - \epsilon)\delta yt + \mid V \mid \epsilon.$$

Step 6] Image–Text Alignment (Optional, CLIP-style)
$For\ batch\{(Ii, Yi)\}i = 1N$encode text with a frozen/learned text encoder fTf_TfT:

$$vi = gV(Ii), ti = gT(Yi),$$

$$LInfoNCE = -N1i = 1\sum N[log\sum jexp(cos(ti, vj)/\tau)exp(cos(ti, vi)/\tau) + log\sum jexp(cos(vi, tj)/\tau)exp(cos(vi, ti)/\tau)].$$

Step 7] Style Regularization (Voice Control)
Match generated distribution to a target style prior

$$Rstyle = t = 1\sum TKL(p\Theta(\cdot \mid y < t, I, m, s)\ \ pstyle(\cdot \mid s)).$$

Step 8] Length/Structure Regularization
Control paragraph or total length TTT toward a target T∗T and penalize redundancy:

$$Rlen = (T - T*)2, Rrep = t\sum max(0,\ cos(E(yt), E(yt - k)) - \rho)\ \ for\ small\ k.$$

Step 9] Reinforcement Learning with Artistic Reward
Define a sequence-level reward that mixes creativity, grounding, coherence, and safety:

$$R(I, Y) = \alpha Creat(Y) + \beta Align(Y, I) + \gamma Coh(Y) - \eta Halluc(Y, I) - \zeta Unsafe(Y),$$

$$Align(Y, I) = T1t\sum imaxcos(E(yt), zi), Coh(Y) = BERTScore(Y, shuffle(Y)) - 1.$$

Optimize by REINFORCE (or PPO) with baseline b:

$$\nabla\Theta JRL = -E[(R - b)t\sum\nabla\Theta logp\Theta(yt \mid y < t, I, m, s)].$$

Step 10] Joint Training Objective

$$min\, J = LCE + \lambda 1LInfoNCE + \lambda 2Rstyle + \lambda 3Rlen + \lambda 4Rrep + \lambda 5JRL.$$

Step 11] Decoding with Constraints

Constrained beam search with length penalty ℓ\ell$\ell$ and style bias ψ\psiψ:

$$\hat{} = argYmaxT\ell 1t\sum logp\Theta(yt \mid y < t, I, m, s) + \mu logp\psi(Y\ matches\ style\ s).$$

Step 12] Post-Processing (Deterministic)

Apply a non-parametric map Π(·) for grammar, de-duplication, and paragraphing:

$$Yfinal = \Pi(Y\hat{};\ grammar = 1,\ dedup = 1,\ len = T*).$$

The proposed hybrid model of NLP is aimed to respond to the gap between the superficial, literal descriptions of artwork, and the deeper, interpretive narrative that is part of the expectations of professional artistic statements.

5. PROPOSED HYBRID VISION-GUIDED GENERATIVE LANGUAGE MODEL (VG-GLM)

This model incorporates vision-language fusion, transformer-based language generation, stylistic conditioning, and reinforcement learning in a cohesive model that is very similar to how human artists, curators, and art historians interpret and express the meaning of visual works. By combining multimodal perception with complex modeling of the language used, the system is intended to generate multi-paragraphs, expressive and contextually grounded artistic statements that reflect both the aesthetic features of the work of art and the larger conceptual frameworks in which it is made to function. Central to the system is a visual encoder that is used to interpret the compositional, chromatic and textural features of the artwork. This encoder is usually based on a Vision Transformer (ViT) or a CLIP-based visual model that converts a raw image into a high-dimensional representation which is used for capturing patterns such as balance, contrast, subject matter, brushwork and stylistic elements. By making this visual understanding the foundation of the text generation process that follows, the system can help keep the end results of that process tied to the properties of the actual artwork and not get lost in generic or irrelevant commentary.

Figure 2

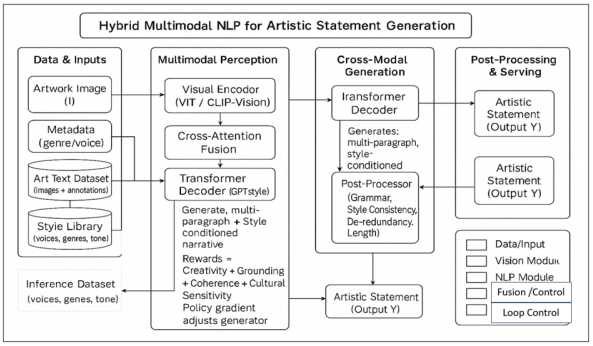

Figure 2 Hybrid Multimodal NLP for Artistic Statement Generation

In other words, this training allows the model to internalize the linguistics of the art world, including the use of metaphorical descriptions, reflective reasoning, framing within context, and the expressive language used to discuss visual culture in academic and creative writing. To increase stylistic control further, the system includes a style-

conditioning mechanism in which the generator could be instructed to generate writing in a specific voice - whether it be poetic, minimalist, academic, culturally-grounded or emotionally-introspective. This is allowing customization depending on the genre of art, the medium or the audience. One of the key aspects of the model is the cross-modal fusion process that integrates the visual and the textual stream. During text generation, the decoder does not work in an insular way, but is constantly paying attention to the visual embeddings of the artwork. This cross attention mechanism allows the model to incorporate visual signals (i.e. dominant colors, compositional flow, symbolic motifs etc) directly into the unfolding narrative. This includes adjusting the flow of the grammar, ensuring that the style of the paragraphs is consistent with each other, removing redundant phrases, and adjusting the length of the statement to a normal level as according to the professional standards. Through this series of processes, the hybrid NLP model turns out to produce polished, articulate and culturally-sensitive artistic statements that conform to the ranges of curatorial writing and professional arts communications. The integration of vision, language, style control and reinforcement learning aligned with human intent eventually enables the system to perform well in one area where previous models have struggled - the ability to capture the complexity, intentionality and emotional resonance of visual art through coherent and compelling narrative form.

6. EVALUATION AND EXPERIMENTAL DESIGN

The performance evaluation of an artistic statement generation model therefore requires a methodology that accounts both for the linguistic quality of generated text as well as conceptual similarity between narrative and artwork. This curated data set forms the basis of the empirical data set on which the performance of the model is measured. Automated evaluation metrics are playing an important role in the evaluation of the baseline linguistic quality. During the evaluation phase, the system is compared to the generated statements with reference texts by the use of established measures like BLEU, ROUGE, METEOR etc. to estimate the overlapping of phrasing, vocabulary, and syntactic patterns.

Figure 3

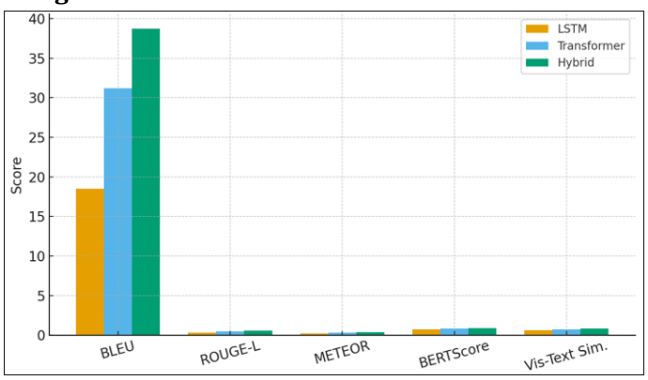

Figure 3 Automated Evaluation Results

While the metrics described above are useful in providing a simple way to compare with previous NLP techniques, creativity and interpretative richness are usually underestimated in this case, particularly in the artistic context, where the same artwork can be described with different, equally correct expressions. To make up for this limitation, the evaluation is further extended to contextual similarity measures like BERTScore and Sentence-BERT embeddings which measure semantic similarity between generated statements and human-written interpretations as shown in Figure 3. These measures based on embeddings make it possible to better judge the conceptual understanding and narrative cohesion of the model by capturing the sense of the text, rather than the overlap of the text on the surface. In addition to textual assessments, visual-semantic correspondence is analyzed to make sure that the generated narrative is meaningful with respect to the artwork.

Figure 4

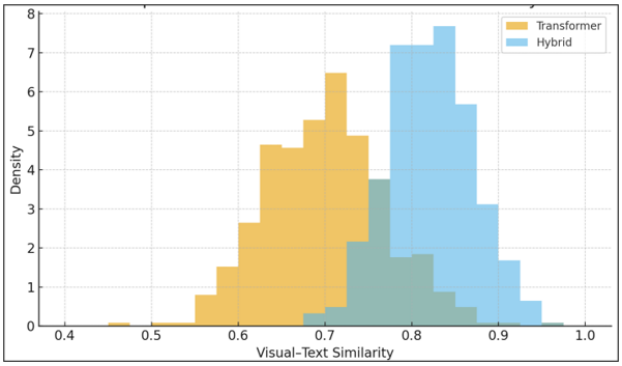

Figure 4 Perplexity vs. Human Creativity Score

This is evaluated using the multimodal similarity measures comparing the visual embeddings from the artwork and the text embeddings from the generated statement as shown in Figure 4.

Figure 5

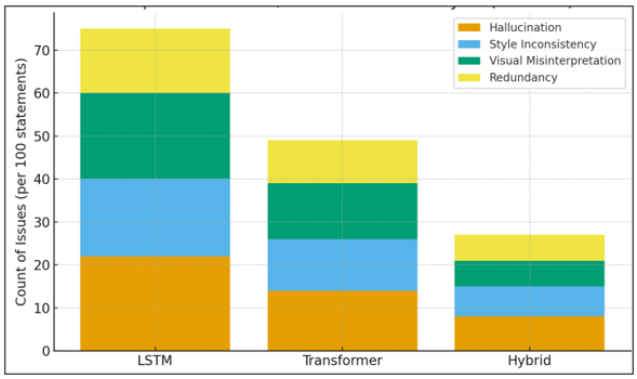

Figure 5 Error and Hallucination Analysis

The assessment also encompasses a very high human aspect of evaluation, as artistic writing has so subjective elements that it cannot be quantified exactly by the use of automated standards. Statements are read by the reviewer without reference texts to prevent any form of prejudice against evaluation of the narrative according to its inherent value and capacity to communicate artistic intent as represented in Figure 5. Their qualitative feedback reveals trends in the strengths and weaknesses of the model, whether it is inclined towards giving excessively literal interpretations, revealed creative insight or revealed the sensitivity to cultural contexts inherent in artworks. The automatic, multimodal and human evaluation combination results into a moderate and holistic examination of the capability of the hybrid NLP model. Automated measures have to do with the item of guaranteeing the objective comparability, multimodal tasks guarantee the grounding in the visual material, and human judgments indicate the subjective quality required in actual artistic communication. They are used jointly to offer a difficult platform on testing the validity of the model in terms of creating high quality artistic statements that are palatable to both academic and artistic demands.

7. DISCUSSION AND ANALYSIS

As it is evident in the results of the evaluation, the proposed hybrid NLP model actually offers significant progress in both linguistic and visual grounding in comparison with the traditional LSTM systems, as well as the transformer-only baselines. Automated measures indicate that the hybrid model will provide more coherent and semantically consistent artistic statements. An increase in the values of the BLEU, ROUGE-L, METEOR and especially the BERT Score is an assuring factor that the model is not just tracking down the superficial features of phrasing but it is also tracking down the deeper conceptual features that are embedded in texts written by experts. This implies that it is the internalization of stylistic and interpretative elements of professional artistic discourse based on a fine-tuning of the transformer backbone on art-specific corpora that allows it to occur.

Table 3

Table 3 Summary of Automated Evaluation Metrics

Model	BLEU ↑	ROUGE-L ↑	METEOR ↑	BERTScore ↑	Visual–Text Similarity ↑
LSTM Baseline	18.5	0.34	0.21	0.72	0.62
Transformer Fine-Tuned	31.2	0.47	0.32	0.84	0.73
Hybrid Proposed Model	**38.7**	**0.55**	**0.39**	**0.9**	**0.82**

The evaluations provided by human beings are even stronger arguments in regards to the capability of the model. The radar chart indicates that the experts had continuously placed the outputs of the hybrid system in the statements written by human beings in a closer proximity upon the interpretive depth in style, creativity and cultural sensitivity. There was also significant enhancement in the grounding ability as was reported by the reviewers as a result of the hybrid model capability of including the visual cues when creating the narratives. The primary cause of this improved alignment appears to be the cross-modal fusion mechanism where the decoder will pay attention to the attributes of the artwork during text production.

Table 4

Table 4 Human Evaluation Score Comparison

Evaluation Criterion	Human Reference	Transformer Model	Hybrid Model
Interpretive Depth	4.8	3.9	**4.5**
Stylistic Coherence	4.6	4.1	**4.4**
Visual Grounding	4.7	3.6	**4.3**
Cultural Sensitivity	4.6	3.9	**4.4**
Creativity	4.8	4	**4.6**

The hybrid system scores almost as high in all qualitative dimensions and is superior in interpretive richness and stylistic accuracy and cultural sensitivity than other automated systems. This finding is also supported by the distribution of the similarity scores of visual-texts. In the hybrid approach as compared to the model with transformer only, the clustering of larger similarity value is more compact meaning that the text that was generated is more related to the visual characteristics of the artwork. This indicates a significant entrenching of the image embeddings in the generation process and not the generic language of art. The other strength of the hybrid model is contained in the correlation between perplexity and creativity ratings. The perplexity is strongly related to the score of the human creativity, and the lower the perplexity, the higher the score of the creativity (this fact can inform future studies on the interpretation of the concept of creativity). In comparison to the two models of comparison, the hybrid system always exists at the low perplexity high creativity area. Lastly, the error analysis demonstrates that the hallucinations, style inconsistencies, and misunderstandings have reduced greatly. The hybrid model has more governed and precise narrative behavior that is probably due to the reinforcement learning with artistic reward signals. This is a process that punishes the unsubstantiated assertions and encourages the grounded consistent descriptions. Comprehensively, the findings indicate that hybrid multimodal model is a significant progress in generation of artistic statements. It produces more eloquent, visual, and conceptually advanced stories than existing strategies therefore making it an excellent candidate of real-world implementations of digital curation, creative support instruments, and educational settings.

8. CONCLUSION

This study suggested a hybrid NLP model of the generation of expressive and contextually-sensitive artistic utterances by applying visual comprehension, transformer-based language modeling, and stylistic conditioning and reinforcement learning. The model addresses the significant weaknesses of existing methods because it combines multimodal perceptual incentives and high-level narrative generative methods in order to be capable of generating coherent and conceptually rich interpretations, as opposed to simple descriptive captions. The system demonstrates the capacity to intuit the aesthetic qualities of the artwork, through a highly well-constructed combination of visual embeddings and linguistic expertise, as well as replicates the qualities of tonal and structural qualities of professional artistic writing. The aptitude of the model to produce meaningful stylistically and visually-grounded statements is

confirmed by the use of experiential evaluation measures i.e. the automatic multimodal alignment measures and also an expert human rating of the generated content. Although the approach has its advantages, it has to struggle with the issue of diversity of data, the question of cultural sensitivity and its interpretive subjectivity, and computational complexity. The paper provides a strong foundation on which research in the field of AI-aided artistic communication and the ways in which intelligent systems can be applied to complement, rather than substitute, the creativity of humans.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Aldekhail, M., and Almasri, M. (2022). Intelligent Identification and Resolution of Software Requirement Conflicts: Assessment and Evaluation. Computer Systems Science and Engineering, 40(2), 469–489. https://doi.org/10.32604/csse.2022.018269

Borawake, M., Patil, A., Yadav, S., Nagwade, V., and Somwanshi, H. (2025). Driver Drowsiness Detection using ML and IOT. IJRAET, 14(1), 114–117.

Bozyiğit, F., Aktaş, Ö., and Kılınç, D. (2021). Linking Software Requirements and Conceptual Models: A Systematic Literature Review. Engineering Science and Technology, an International Journal, 24(1), 71–82. https://doi.org/10.1016/j.jestch.2020.11.006

Deshpande, G., Sheikhi, B., Chakka, S., Zotegouon, D. L., Masahati, M. N., and Ruhe, G. (2021). Is BERT the New Silver Bullet? An Empirical Investigation of Requirements Dependency Classification. In Proceedings of the IEEE 29th International Requirements Engineering Conference Workshops (REW 2021) (pp. 136–145). https://doi.org/10.1109/REW53955.2021.00025

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019) (Vol. 1, pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Došilović, F. K., Brčić, M., and Hlupić, N. (2018). Explainable Artificial Intelligence: A Survey. In Proceedings of the International Convention on Information, Communication, Technology, Electronics and Microelectronics (MIPRO 2018)

Joshi, A., Karimi, S., Sparks, R., and Macintyre, C. R. (2020). Survey of Text-Based Epidemic Intelligence: A Computational Linguistics Perspective. ACM Computing Surveys, 52(6), Article 119. https://doi.org/10.1145/3361141

Kim, A. Y., and Hardin, J. (2021). Playing the Whole Game: A Data Collection and Analysis Exercise with Google Calendar. Journal of Statistics and Data Science Education, 29(sup1), S51–S60. https://doi.org/10.1080/10691898.2020.1799728

Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., Yu, P. S., and He, L. (2022). A Survey on Text Classification: From Traditional to Deep Learning. ACM Transactions on Intelligent Systems and Technology, 13(2), Article 31. https://doi.org/10.1145/3495162

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., and Gao, J. (2021). Deep Learning-Based Text Classification: A Comprehensive Review. ACM Computing Surveys, 54(3), 1–40. https://doi.org/10.1145/3439726

Peng, S., et al. (2022). A Survey on Deep Learning for Textual Emotion Analysis in Social Networks. Digital Communications and Networks, 8(5), 745–762. https://doi.org/10.1016/j.dcan.2021.10.003

Pham, P., Nguyen, L. T. T., Pedrycz, W., et al. (2023). Deep Learning, Graph-Based Text Representation and Classification: A Survey, Perspectives and Challenges. Artificial Intelligence Review, 56, 4893–4927. https://doi.org/10.1007/s10462-022-10265-7

Sarwar, T., et al. (2023). The Secondary use of Electronic Health Records for Data Mining: Data Characteristics and Challenges. ACM Computing Surveys, 55(2), Article 33. https://doi.org/10.1145/3490234

Wang, B., Peng, R., Wang, Z., Wang, X., and Li, Y. (2020). An Automated Hybrid Approach for Generating Requirements Trace Links. International Journal of Software Engineering and Knowledge Engineering, 30(7), 1005–1048. https://doi.org/10.1142/S0218194020500278

Wang, R. (2021). K-adapter: Infusing Knowledge into Pre-Trained Models with Adapters. In Findings of the Association for Computational Linguistics: ACL 2021 (pp. 1405–1418). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-acl.121

Wu, J.-L., et al. (2020). Identifying Emotion Labels from Psychiatric Social Texts Using a Bi-Directional LSTM-CNN Model. IEEE Access, 8, 66638–66646. https://doi.org/10.1109/ACCESS.2020.2985228

Zhao, L., Alhoshan, W., Ferrari, A., Letsholo, K. J., Ajagbe, M. A., Chioasca, E.-V., and Batista-Navarro, R. T. (2022). Natural Language Processing for Requirements Engineering: A systematic mapping study. ACM Computing Surveys, 54(3), 1–41. https://doi.org/10.1145/3444689