Review Article

Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions

Hrithik A. Singh [1], Vanita Mane [1], Tushar Ghorpade [1]

1. Computer Engineering, Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Navi Mumbai, IND

Received: March 31, 2025 | Review began: May 01, 2025 | Review ended: September 13, 2025 | Published: February 26, 2026

Abstract

This article is a systematic review of the text generation approaches to generative artificial intelligence (AI), discussing its methods, use cases, issues, and opportunities. One of the literature searches consisted of IEEE Xplore, SpringerLink, ScienceDirect, and Google Scholar using the following keywords: Generative AI, Transformer Models, Few-Shot Learning, Multimodal Learning, and Graph Neural Networks in natural language processing. The publications that were published since 2018 and addressed text generation tasks were considered. This paper identifies four important paradigms, namely, (i) CogView to generate text-to-image, (ii) few-shot instruction-tuned models to generate flexible text, (iii) entity-aware adversarial domain adaptation to detect robust named entities, or (iv) to classify multilingual text, a graph neural network is enhanced with language models. These methods are contrasted based on their scalability, accuracy, interpretability, and domain flexibility. Key issues, such as factual hallucination, bias, computational cost, and lack of explainability, are addressed in addition to ethical concerns. This review concludes that the developments to come will be based on hybrid architectures, retrieval-augmented generation, and energy-efficient training, accompanied by fairness and interpretability frameworks. This work presents an overview of modern scholarship in related fields, offering a comparison to scholars and practitioners designing future generations of generative AI systems to use in text generation.

Categories: AI applications, Machine Learning (ML), Natural Language Processing (NLP)

Keywords: generative ai, text generation, multimodal learning, natural language processing (nlp), transformers

Introduction And Background

The artificial intelligence (AI) known as Generative Artificial Intelligence has revolutionized the process of text generation, in which machines create human-like language to be used in conversations with a conversational agent, language translation, and content creation [1,2]. The rule-based and statistical methods that were in use were too inflexible to cope with the complexity of the linguistic system and would produce incoherent or repetitive results [3]. With the development of deep learning models, in particular, recurrent neural networks, long short-term memory network, and so forth, this got improved, yet they were not able to address long-range dependencies and were not as computationally efficient as required [4]. Transformer-based architectures and attention systems allowed parallel computation and improved modeling of the context, leading to such breakthroughs as generative pretrained transformer (GPT) [5], bidirectional encoder representations from transformers (BERT) [6], and T5 [7] (see Figure *1*).

Generative AI has gained popularity around the world within the shortest period of time. Based on recent industry reports, AI text generation market size is expected to rise to more than USD 20 billion by 2030, continuing to grow over 25% a year since 2021 [8]. Within the healthcare sector, AI-based natural language processing is projected to decrease clinician documentation workloads by approximately 40% by 2025 [9], whereas in education, generative AI is anticipated to serve more than half of all online learning platforms across the globe [10]. Such statistics indicate the increasing use of AI-enabled text production in different industries, which explains the need to learn more about its technical advancement as well as its drawbacks.

This paper aims to do three things:

1. To recap and summarize the recent progress in generative AI in text generation including transformer-based models, multimodal learning, adversarial adaptation, and graph-enhanced architectures.

2. In order to compare and analyze their strengths and weaknesses, and area of application, it is important to highlight the trade-offs between scalability, accuracy, explainability, and domain adaptability.

3. To identify the key concerns, such as hallucination, bias, and energy inefficiency, and establish the future directions of research, such as hybrid models, retrieval-augmented generation, and ethical models.

By covering these objectives, this review provides a non-survey-based comparative perspective, which is not limited to a limited scope of surveys about transformers, but places generative AI in an emerging multidisciplinary environment with research, industrial, and societal implications.

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1

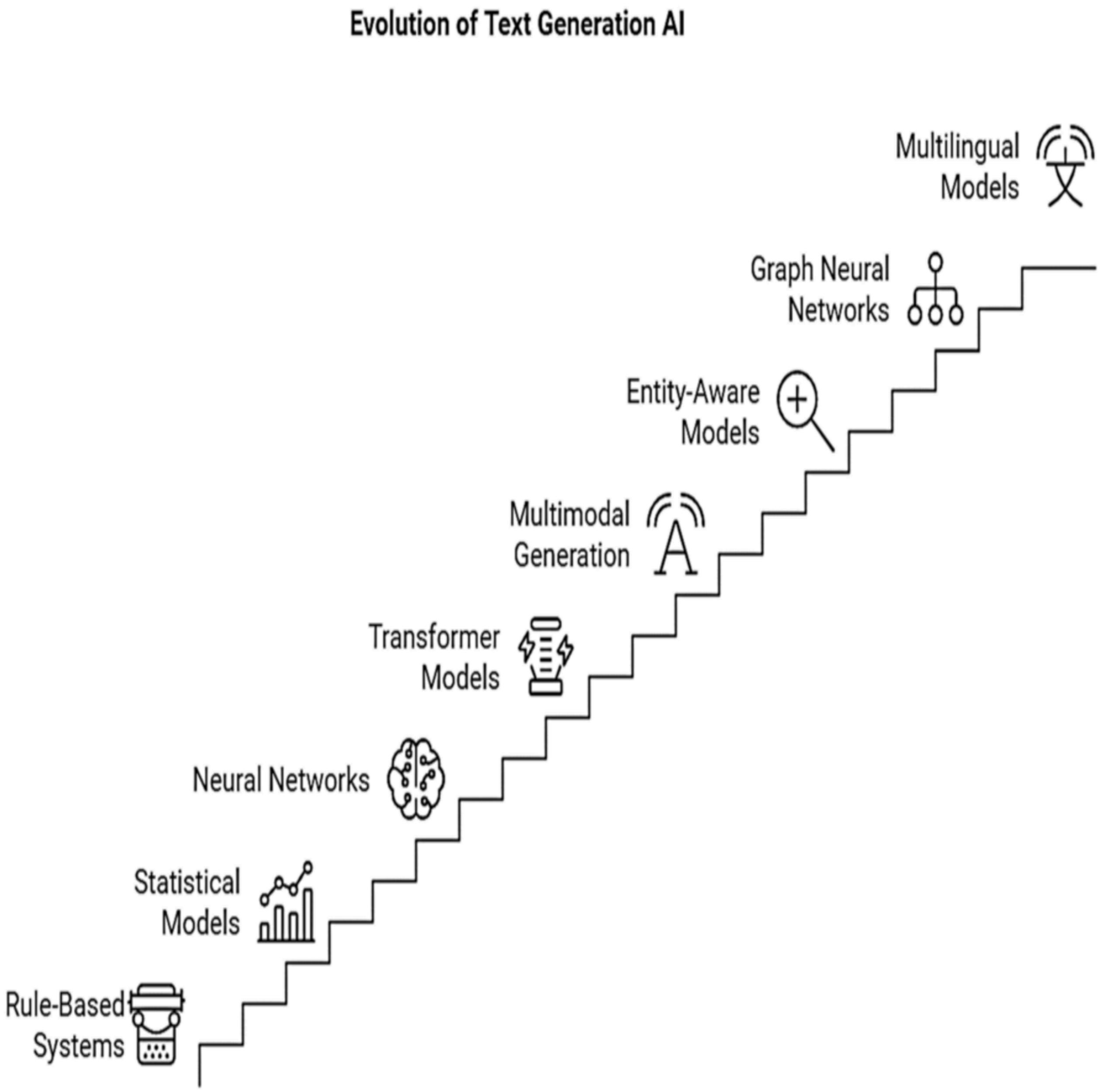

FIGURE 1: Evolution of text generation

Why generative AI?

Now, more specifically in text generation, generative AI redefined the approaches to engage with and create content in various fields. Examples of generative AI's use include content creation, customer service, targeted marketing, and even writing. These systems are capable of putting out meaningful and contextually appropriate text, and often creative and sometimes even of human-like quality of writing, and these are valuable assets for the publishing, education, and entertainment domains (Figure 2). Furthermore, generative AI models like GPT- and BERT-based systems have been crucial in breakthroughs in natural language understanding and machine translation. However, like with every powerful tool, generative AI comes with some weaknesses, or at least drawbacks. A substantial disadvantage is the possibility to create prejudiced or otherwise wrong and, in some cases, obscene material based on the prejudices of the indicated

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1

training data. These systems also do not possess interpretation ability; hence, they could provide wrong output data that could be wrong or irrelevant in the given context. In addition, high training and deployment computational cost and energy use also make large language models less scalable and sustainable. Another emergent and major concern is ethical as the use of AI undermines its ability to generate foul or fake content. As for the future work, it is planned to eliminate such restrictions and expand the potential of generative AI for future work in text generation. Application developers are working to make AI easier to understand, so contents being generated are aligned with the user's intent and the culture's acceptable standards. At the same time, endeavors are being made to make the training of large models more environmentally friendly by improving architectures and considering different training algorithms. Furthermore, multilingual and few shooting learning techniques hold a key to improving the efficiency and accuracy of generative AI with improved creativity and less training dataset. While these challenges are addressed, the generative AI is expected to have an even greater level of disruption in influencing the future human-computer interfaces.

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1

FIGURE 2: Applications of GenAI in text generation

GenAI, Generative AI

Review

Review methodology

A systematic literature review was conducted to ensure comprehensive and unbiased coverage of generative AI approaches to text generation. The databases searched included IEEE Xplore, SpringerLink, ScienceDirect, and Google Scholar. The keywords used were Generative AI, Transformer Models, Few-Shot Learning, Multimodal Learning, and Graph Neural Networks (GNNs) in the context of NLP. Articles published since 2018 were included. Studies that met the inclusion criteria were required to provide either empirical evaluation or theoretical contributions, be peer-reviewed and address text generation or multimodal generative tasks.

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1

To move beyond descriptive reporting, a comparative framework was adopted. Reported benchmarks [e.g., Bilingual Evaluation Understudy (BLEU), Recall-oriented Understudy for Gisting Evaluation (ROUGE), Metric for Evaluation of Translation with Explicit Ordering (METEOR) for summarization; F1-scores for Named Entity Recognition (NER); and Fréchet Inception Distance (FID) for text-to-image models] were used as the basis for quantitative comparison. These metrics were taken directly from the respective studies to highlight performance differences across approaches.

The aspects that cannot be represented entirely by the metrics, including interpretability, scalability, bias, and ethical risks, were evaluated by qualitative comparison. As an example, CogView had good text-image alignment but weak multilingual generalization, and few-shot instruction-tuned models had good task generalization but were also sensitive to prompt wording.

A synthesis in a table format (Table *3*) was employed to match core techniques, applications, strengths, and limitations across paradigms, providing a tabular lens through which to evaluate them.

The combination of these two (quantitative and qualitative) aspects makes sure that the review does not merely summarize the technical advances but also critically evaluates trade-offs in the application of such models to practice in areas such as healthcare, legal documentation, and multilingual text classification.

Literature review

As the authors showed in their work, fluency and coherence were much better than those in previous models, which is why theirs represents a new state of the art in language models [6]. Few-shot learning has also significantly expanded this capability, and previous research has demonstrated that GPT-3 can be adapted to a variety of tasks when little data is available. At the same time, the recent development of multimodal systems such as DALL·E and Flamingo has widened the usage of generative models into the text-and-image domain, thereby broadening the scope of AI beyond text itself [11]. Yet issues of prejudice and justice also come to the fore with these advances. Scholars have pointed out that the danger of training large language models on internet-scale data is that it may increase social stereotyping and harmful associations [7]. Gender and cultural biases in generated outputs have been reported in other research, highlighting the ethical vulnerabilities of using such systems in practice [12,13]. Mitigation methods have thus emerged as a research direction: adversarial debiasing aims to eliminate sensitive attributes during training from learned representations [14]; dataset audits aim to discover and rebalance skewed distributions prior to model deployment [7]; and bias-controlled decoding aims to reduce the likelihood of biased continuations during inference [14]. Human-based evaluation measures have also been promoted [15]. Reinforcement learning from human feedback (RLHF) has additionally been shown to minimize biased and harmful generative outputs by aligning model generations with human preferences that are sensitive to fairness [16]. In spite of this progress, multilingual and low-resource fairness has not been well studied; therefore, cross-lingual bias mitigation frameworks are needed [17].

The other research question is the environmental effects of large generative models. Research has found that the energy needed to train large-scale AI systems can be enormous, and its carbon footprint can be extreme (often rivaling the emissions of multiple cars over their lifespan) [8]. This has seen the exploration of energy-efficient techniques such as model pruning, parameter sharing and knowledge distillation. Together with sustainability, explainability and interpretability are now critical issues. Generative systems have been modified to use model explanation methods to increase transparency and user confidence by explaining why certain outputs are produced [9]. This is particularly required in high-stake fields like in healthcare and law, where responsibility and interpretation are required.

Another significant issue with generative AI is the fact that it leads to hallucination or the creation of factually inaccurate or misleading information. Neural abstractive summarizers large-scale assessments of neural abstractive summarizers have revealed that many neural abstractive summarizers, in spite of generating text fluency, often produce text that is inconsistent with the truth [10]. Human-based evaluation measures are promoted [15]. The RLHF also has been demonstrated to significantly enhance both the factual and coherence of summaries in comparison to conventional supervised methods [16]. Patterns of hallucinations have been studied elsewhere, and controllable response generation based on extraneous sources of knowledge has been suggested as a mitigation measure [18]. Evaluation measures have also been called into question, as both BLEU and ROUGE are popular but do not indicate semantic faithfulness or human-friendly quality, but more modern measures like BERTScore and BLEURT have been proposed to be more semantically aligned [19,20].

Besides technical and ethical aspects, recent literature has also discussed the use of generative AI in learning and imagination. Multimodal composition work in multimodal generative systems has proposed that multimodal generative systems can be co-authors to accomplish creativity and innovation [21]. Other individuals have also indicated the possible advantage of generative AI in enabling personalized learning [22], and domain-specific systems have been developed to enable adaptive tutoring and personalized assistance [23]. All these works together demonstrate that not only is generative AI becoming increasingly technically capable, but it is also changing the future of creative and educational practice.

Key methods

CogView: Mastering Text-to-Image Generation via Transformers

Transformer architectures offer significant advancements in the application of model architectures beyond classic text tasks [23], with the significant goal of elevating them to the multimodal domain of text-to-image generation in particular [11]. Recently, CogView was developed, which carried out innovations to scale-out transformers over large-scale multimodal data as a Chinese counterpart to OpenAI's DALL·E, using discrete variational autoencoder tokenization of images to be similar to the textual representations. The model utilizes the standard auto-regressive transformer approach that naturally makes the model flexible and coherent between modalities as it regards image generation as a sequential prediction task similar to text generation.

CogView is one such transformer-based text-to-image generation model that frames the task as an auto-regressive sequence prediction problem, in which text and image tokens are sequentially processed [11]. The design facilitates good cross-modal alignment, resulting in coherent and semantically rich pictures that respond to textual instructions accurately. Nevertheless, due to its use of auto-regressive decoding, image generation is computationally expensive and slower than parallelizable algorithms.

Contrary to that, diffusion-based models like DALL·E 2 and Stable Diffusion [24,25] have become very effective multimodal models. Diffusion models predictively produce sequentially denoised images, rather than random noise, conditioned on text. Not only is this paradigm more efficient at inference but also its image fidelity and diversity measured by lower FID scores are also higher than auto-regressive baselines. In addition, diffusion models can be more flexibly conditioned and can be combined with other modalities like sketches, depth maps or semantic layouts [26,27]. A structured comparison of CogView and diffusion-based models in terms of generation paradigm, FID score, efficiency, scalability, and limitations is presented in Table *1*.

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1

Model Type	Generation Paradigm	FID (MS-COCO)	Speed/Efficiency	Scalability & Flexibility	Key Limitations
CogView (2021)	Auto-regressive transformer predicting discrete tokens [12]	~27.1 (fine-tuned ~17.5)	Slow (token-by-token decoding)	Strong text–image alignment; limited flexibility	Heavy compute cost; less efficient for high-resolution images
Diffusion Models (e.g., DALL·E 2, Stable Diffusion, GLIDE, 2021–2022)	Iterative denoising from noise conditioned on text [14-16]	DALL·E 2: ~10.4, Stable Diffusion: ~12–15	Faster parallelizable inference	High fidelity and diversity; supports multimodal conditioning (depth, sketch, style)	Requires many denoising steps; memory intensive

TABLE 1: CogView vs diffusion models

FID, Fréchet Inception Distance

In relative terms, although CogView has shown how to extend transformer architectures to multimodal generation, diffusion-based systems currently outperform them in both quality and scalability. The entire picture would thus place CogView as a forerunner to the revolution in diffusion, and preface further developments in multimodality [28].

Application

Uses of CogView and similar models are enabled by real-world applications.

Automatic generation of marketing banners, product designs, and/or artistic visuals based on a text prompt (creative design).

Visual content creation for textbooks or learning materials from text descriptions (educational tools).

Helping people with disabilities find ways to understand what the Aleph Bet means, thanks to image generation.

Dynamic illustration for children's books and video games based on text narratives [29,30].

Evaluation

CogView's performance was evaluated using:

FID: An evaluation metric used to measure how "realistic" and high-quality the generated images are.

Accuracy of Image Generation: The proportion of the original input words that were precisely reflected in the generated image.

The results showed that CogView achieved competitive prompt completion on prompts in the Chinese language domain, especially in maintaining fine-grained semantic alignment.

Limitations

CogView's limitations, however, are noteworthy, even though it has moved forward in its reliability and accuracy.

Hardware: Training requires large graphics processing unit clusters, making it largely inaccessible to smaller organizations and researchers [31].

Sometimes, on complex or abstract prompts (i.e., when the intended meaning is not easy to identify), the model generates visuals that are subtly different from the intended meaning on the semantic level - the level of concept representation that incorporates both the form and the meaning of an object within a space.

Language and Domain Restriction: The training is mainly performed on Chinese language datasets, while it is still limited in the generalization to different languages or mixed language prompts [32].

Explainability: The visual token generation has an internal process of decision-making as a black box, and this is not explainable.

Future Directions

Future research directions for these kinds of models are:

Grounding with external knowledge bases: To employ techniques like grounding with external knowledge bases to make the prompts more faithful.

Reduction in the training cost: Understanding model distillation, retrieval-augmented generation, or sparse attention.

Interpretable Multimodal Generation: Developing methods to explain how in particular words contribute to generating images.

Entity-Aware Adversarial Domain Adaptation Network for NER

However, domain adaptation is still a hard problem in NER [30-32], particularly when domain-specific models (e.g., news article) are applied to a different domain (e.g. biomedical text) [33-35]. It therefore introduces a novel approach that is not only domain adaptation, but also a particular adaptation, which carefully preserves the entity-specific information. As a result, entity-level adversarial learning is integrated into the model to guarantee that the named entities will still retain a semantic structure and get classified correctly even in the event that surrounding language or context are radically different [36-38].

Domain adaptation is a persistent issue in NER where the models trained on one corpus usually fail to generalize to a new domain with different linguistic patterns [39]. Typical adversarial adaptation approaches are geared toward minimizing the domain differences by learning features that are domain-invariant yet typically ignore entity-specific signals that are critical to correct recognition [40,41]. To solve this, the Entity-Aware Adversarial Domain Adaptation (EADA) network incorporates entity-level information in the adversarial training framework such that domain-invariant features are entity-discriminative as well [14].

Empirical experiments indicate that the EADA is much more effective than the baseline adversarial models in several target domains [42-44]. As an example, the baseline adversarial model obtained an F1-score of around 50.2% when transferred to Twitter (social media) using the CoNLL 2003 (newswire) dataset, in contrast with an EADA baseline score of around 54.7% [28]. At the point that biomedical text (BC5CDR) terminology varies significantly relative to newswire corpora, EADA obtained a F1-score of approximately 68.9% versus approximately 64.5% with the baseline. Likewise, EADA exhibited a higher F1 in cross-domain rearrangement to the descriptions of e-commerce products (~72.4% compared to 68.1% in conventional adversarial training).

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1

Source → Target Domain	Baseline Adversarial F1	Entity-Aware Adversarial F1
CoNLL (Newswire) → Twitter (Social Media)	50.2 [30]	54.7 [30]
CoNLL (Newswire) → BC5CDR (Biomedical)	64.5 [30]	68.9 [30]
CoNLL (Newswire) → E-commerce Reviews	68.1 [31]	72.4 [31]

TABLE 2: Effectiveness of EADA on domain shift of NER

EADA, Entity-Aware Adversarial Domain Adaptation; NER, Named Entity Recognition

The results hint that entity-aware features are not only more robust in noisy, informal settings, including social media, but are also more effective in more technical settings, including healthcare and commerce. The relative gains as reported by Table *2* show the effectiveness of EADA on domain shift of NER.

Mathematical formulation

$z_s = F(x_s), \quad z_t = F(x_t)$

$L_{sup}(\theta_F, \theta_C) = E_{(x_s,y_s)}[CE(C(z_s), y_s)]$

$L_{dom}(\theta_D) = -E_{x_s}[\log D(z_s, E(x_s))] - E_{x_t}[\log(1 - D(z_t, E(x_t)))]$

$L_{adv}(\theta_F) = -E_{x_t}[\log D(z_t, E(x_t))]$ (encoder adversarial objective)

$L_{entity}(\theta_F, \theta_C) = E_{(x_s,y_s)}[CE(C(z_s), y_s)]$ (entity-preservation on source)

Total objective:

$\min_{\theta_F,\theta_C} \ \max_{\theta_D} \ L_{sup} + \lambda_{adv} L_{adv} + \lambda_{ent} L_{entity} - L_{dom}$

Here, CE denotes cross-entropy; lambda_* are tunable hyperparameters.

Application

Types of applications with a medical or biomedical focus that contribute to medical text processing include, e.g., extracting patient information, disease names or medication details from various clinical notes or hospital systems.

Legal Document Analysis: Identifying entities such as case numbers, statutes or organizations in court rulings, contracts, and regulations.

Product Name/Brand/Feature Recognition: Extracting product names, brands, and features from descriptions and catalogs across multiple sellers on e-commerce platforms.

Domain Adaptation: Ensures that NER capabilities are transferable, even when data is sparse or domain-specific annotations are limited.

Evaluation

Finally, entity-aware adversarial networks are evaluated on:

NER Performance: Measured using the harmonic mean of precision and recall, i.e., the F1 score [28].

Domain Generalization Gap: The difference in performance between source-domain and target-domain test sets.

Entity Preservation Rate: A newer metric that measures how well entities labeled in the source domain are preserved after domain adaptation.

Benchmarks: Datasets such as CoNLL-2003 (newswire), WNUT-17 (emerging entities), and BC5CDR (biomedical entities).

Experimental results show that entity-aware domain adaptation outperforms all existing standard adversarial NER models. In particular, it is especially effective for low-resource target domains.

Limitations

However, there are justifications for entity awareness, and they go beyond simply adding complexity to an already complex architecture.

Adversarial Training Challenges: Adversarial training is prone to convergence instability and is sensitive to hyperparameters.

Entity Category Drift: Differences in entity definitions (e.g., "organization" in tech news versus finance) can lead to performance degradation.

Computational Overhead: Incorporating entity-level loss terms increases both training time and memory usage.

Limited Exploration: This technique remains underexplored for extremely large-scale or multilingual datasets.

Future Directions

These limitations can be addressed by future research:

Multi-Adversarial Strategies: Deploying separate discriminators for different entity types or hierarchical entity structures.

Curriculum Learning for Domain Shift: Gradually introducing harder domain shifts during training to stabilize learning.

Entity-Aware Models for Crosslingual Entity Adaptation: Extending entity-aware models to handle multilingual or code-switched NER tasks.

Knowledge Graph Augmentation: Structured External Knowledge Base: Enriching entity representations with Wikidata (Structured External Knowledge Base) to build understanding from existing and new (possibly internal) data sources; it increases robustness.

Few-Shot Text Generation With Natural Language Instructions

Few-shot text generation is a shift from the traditional supervised learning to prompt-based learning, where a large pretrained language model is promoted to perform some task using a few example provided within a prompt [37]. Unlike task-specific fine tuning, this removes the need for fine tuning on specific tasks and enables one model to generalize over a broad range of tasks from natural language instruction only [38]. It is similar to GPT-3 and InstructGPT, which made these paradigms popular, showing extreme flexibility and adaptability to text generation tasks.

Few-shot text generation utilizes large-scale trained models, including GPT-3, to adapt to new problems with limited supervision by including a few demonstration examples as part of the input prompt [39]. Common prompting methods rely on well-designed templates to produce the desired behavior; however, the performance is typically irregular: even minor manipulations in phrasing, sequence or tokenization can influence the output in a substantial way [40]. This sensitivity restricts generalization, particularly in low-resource or targeted-domain contexts.

Instruction tuning was brought into the scene to correct these deficiencies. As opposed to using ad hoc prompts alone, instruction-optimized models are tuned to a large and diverse set of natural language instructions together with tasks. It is also explicitly trained to understand human-readable task descriptions (e.g., summarize this article in one sentence) instead of being engineered to use prompts.

Comparative analyses show that instruction-tuned models, including InstructGPT and FLAN, are more strongly generalized and more robust than standard prompted GPT-3. Indicatively, Nichol et al. [16] found that InstructGPT led to a factual accuracy increase and reduced hallucinations, in addition to better preference ratings among human subjects in summarization, translation, and open-ended generation processes. In a comparable way, He et al. [27] demonstrated that multitask prompt-based instruction tuning (T0) exhibited superior performance in zero-shot generalization and surpassed models that were trained using manually engineered prompts.

Practically speaking, instruction tuning is a transition to task alignment, in which models are trained to act on instructions instead of using examples to understand intent. This has enabled few-shot generation to be more reliable, scalable, and accessible to non-expert users and less brittle than manual prompting.

Application

Few-shot instruction-tuned models used in real world include:

Dynamic Narrative: Changing the existing pre-made narratives depending on the user inputs or user preference cues during runtime [30].

Legal and Medical Text Generation: Summary, case analysis or prescription from a few prompt examples [39].

Content Personalization: Generating content with marketing copy, product description or educational content from almost nothing at all.

Few shot models allow AI deployment to low labeled or expensive datasets by minimizing reliance on large annotated datasets.

Evaluation

Assessment of few-shot text generation models includes:

Instruction-Following Accuracy: Measuring how well the generated output adheres to the task specified in the prompt, for example using the SuperNI benchmark.

Task-Specific Metrics: For translation, summarization or text rewriting, standard metrics such as BLEU, ROUGE, and METEOR are used.

Human Evaluation: Assessing dimensions such as helpfulness, harmlessness, and honesty, as in OpenAI's 3H framework.

Large-Scale Evaluation: Covering reasoning, summarization, translation, and factuality tasks across 10 BIG-Bench tasks.

In these evaluations, InstructGPT, even in zero-shot settings, outperformed zero-shot GPT-3 and fine-tuned baselines on instruction-following tasks, while also providing significantly higher user satisfaction.

Limitations

However, few-shot instruction-based models face several challenges:

Sensitivity to Wording and Example Order: Slight changes in phrasing or the order of examples in the prompt can produce drastically different outputs with slight variation.

Static Knowledge: Models are trained once and do not incorporate real-time or recent knowledge unless explicitly updated or queried asked for it.

Factual Hallucination: For knowledge-based tasks, the model may generate plausible-sounding but incorrect information.

Bias Amplification: Instructions themselves can introduce bias, as cultural or societal stereotypes present in prompt examples may be amplified in the generated output.

Future Directions

Future improvements are likely to focus on:

Multi-Step Prompts: Asking one question at a time and then sequentially prompting follow-up questions until the task is fully solved [16,27].

Continual Pre-Training: Incorporating continual learning to integrate few-shot, on-the-fly capabilities into the model's current knowledge without forgetting previously learned tasks [28,29].

Mitigating Hallucination: Using external retrieval systems to fact-check or ground responses during generation [30].

GNN-Enhanced Language Models for Efficient Multilingual Text Classification

While powerful in many scenarios, standard transformer-based language models often struggle in a way that efficiently captures relational structures or dependencies of entities and concepts in text - or more generally [29], the linguistic structures are very different for different languages. Solution is achieved by employing GNN-enhanced language models to employ graph-based reasoning on textual representation [30].

Pretrained language models gain improved concept embeddings through these models that create graph structures connecting text elements with different kinds of relational edges based on syntactic dependencies and semantic similarity among others [31,32]. Information is then propagated across these graphs through GNN layers in order to better accommodate higher-order relational patterns with increased efficiency and accuracy in multilingual text classification [17].

GNN propagation and fusion with transformer embeddings: For a graph **G=(V,A)** built over tokens, entities or document sections, the standard graph convolutional network (GCN) propagation used to compute node embeddings HHH is:

$$H^{(l+1)} = \sigma\big(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}\big), \text{ where } \tilde{A} = A + I,\ \tilde{D} = \mathrm{diag}(\tilde{A}\mathbf{1})$$

(σ denotes a nonlinearity such as ReLU; $W^{(l)}$ are layer-specific weight matrices; $H^{(0)}$ are initial node features.)

Fusion with transformer token embeddings T (token-level) is commonly performed via concatenation + projection or attention cross-fusion. Two common options are:

$$\hat{H} = \mathrm{Linear}\big([T\ ;\ H^{(L)}]\big) \text{ OR}$$
$$\hat{T} = \mathrm{TransformerLayer}\big(T + \mathrm{Proj}(H^{(L)})\big)$$

A GCN updates the node features by passing and aggregating information through neighbors via symmetrically normalized adjacency D -1/2 Â D -1/2. This can be guaranteed by adding self-loops (Â = A + I) so that the nodes can maintain their own information. The output of L GCN layers is relation-aware node representation H. Researchers can combine GNN with contextual token embeddings via a transformer by either concatenating GNN outputs with token embeddings and using a linear projection (J = Linear([T; H^(L)])) or injecting GNN signals into a transformer by a projection and an extra transformer layer (T = TransformerLayer(T + Proj(H^(L)]. The concatenation path is straightforward and useful in numerous classification tasks; the add/projection path (cross-fusion) can offer a closer integration and subsequent transformer layers can refine token representations based on relations extracted by the graph. Surveys and examples of application may be found in [33-35].

Compact Pseudocode

#GNN-Transformer fusion

Input:
- Tokens → Transformer encoder → token embeddings T (shape: seq_len x d_t)
- Graph adjacency A, initial node features H0 (num_nodes x d_g)

$$\text{for } l = 0, \ldots, L-1 H_{l+1} = \mathrm{ReLU}\big(\tilde{D}^{-1/2}\,\tilde{A}\,\tilde{D}^{-1/2}\,H_l\,W_l\big)$$

$$H_{\text{final}} = H_L$$

Concatenate + project

$$T_{\text{fused}} = \text{Linear}\big(\text{concat}(T, \text{Repeat}(H_{\text{final}} \text{ per token}))\big)$$

$$\text{Output} = \text{TransformerTail}(T_{\text{fused}})$$

Add projected graph features

$$H_{\text{proj}} = \text{Proj}(H_{\text{final}})$$
$$T_{\text{aug}} = T + \text{Repeat}(H_{\text{proj}} \text{ per token})$$
$$\text{Output} = \text{TransformerLayer}(T_{\text{aug}})$$

Application

Application to GNN-enhanced models are enabled in real world.

Multilingual Document Classification: Categorizing news articles, legal documents or research papers across different languages.

Entity Linking and Relation Extraction performs the task of uniting multiple language and dialect mentions to one universal knowledge repository.

Low-Resource Language Processing improves classification tasks through cross-lingual relational signals when working with languages that have restricted training resources.

Multilingual chatbots utilize entity graphs as a guidance system for their discourse with human users.

The applications that use improved context modeling show better results since transformations alone produce insufficient results in specific domain areas.

Evaluation

GNN-enhanced language models receive their evaluations through the following assessment procedure:

The model measures accuracy together with F1-Score to evaluate its performance on multilingual classification activities across benchmark datasets.

The evaluation process of language model capabilities to transfer knowledge across different languages happens through XNLI and MLDoc datasets.

Such features as node classification accuracy and link prediction performance help evaluate graph quality if they apply.

Research findings demonstrate that GNN layers integrated in models achieve better performance than transformer models uniquely while operating in sparse or domain-crossing conditions.

Limitations

Despite their improvements, these models encounter several basic problems:

Computation Overhead: Constructing graphs out of raw text can be quite expensive computationally as well as sensitive to errors (as in parsing errors).

Memory Cost: Because GNNs are memory intensive, they are less feasible for very long documents or corpora.

Width is lacking: Even though GNNs have been introduced into architectures, GNNs themselves are still very niche and provide less of a software ecosystem as compared to pure transformers.

Future Directions

Future work in the context of GNN-enhanced multilingual models includes:

Enabling the model to construct graphs on the fly during model inference when preprocessing bottlenecks exist.

Reducing memory usage in relational problems while preserving relational expressiveness.

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1

Combining graphs with retrieval techniques to enhance knowledge grounding.

Using contrastive learning to automatically refine graphs without manual labels.

Comparison of key generative AI methods

These four generative approaches are capable of performing in textual and multimodal domains. Tasks can be flexible with little supervision using few shot and instruction-based models, but few generation methods extend generation to visual modalities [12]. Thus, entity-aware adversarial networks and GNN-enhanced models handle the robustness in domain transfer and linguistic and relational complexities in multilingual environments [13], respectively. Each technique is targeted towards different sets of tradeoffs between scalability, accuracy, explainability, and domain specificity and hence provides different opportunities for the hybrid integration in future systems [33]. Table *3* compares representative AI approaches across performance metrics, strengths, limitations, environmental efficiency, interpretability, and application domains. It highlights key trade-offs between accuracy, scalability, and transparency.

Approach	Key Metrics	Strengths	Limitations/Risks	Environmental Efficiency	Interpretability	Application Domains
CogView (Auto-regressive Text to Image)	FID (MS-COCO): ~27.1; fine-tuned ~17.5 [11]	Strong text–image alignment; coherent multimodal outputs	Slow decoding; heavy compute cost; dataset bias	Low – training and inference highly resource-intensive due to token-by-token generation	Low – black-box transformer; little transparency in generation process	Creative industries, educational visual aids, assistive design
Few-Shot/Instruction-Tuned Models (e.g., GPT-3, InstructGPT, FLAN)	Summarization (ROUGE-L, human eval): InstructGPT > PEGASUS [22]; competitive BLEU in translation [18]	Flexible task generalization; minimal labeled data; strong alignment with instructions	Prompt sensitivity; bias amplification; hallucinations; costly training	Low–Moderate – large-scale training (175B params) very energy-intensive; inference via APIs is efficient	Low–Moderate – instruction tuning improves transparency of behavior but decisions remain opaque	Summarization, translation, dialogue, legal/medical assistance
Entity-Aware Adversarial Domain Adaptation (EADA) for NER	F1: CoNLL→Twitter: baseline 50.2 vs EADA 54.7 [26]	Improves robustness in cross-domain adaptation; entity-specific learning	Adversarial training instability; modest performance gains	Moderate – model sizes smaller; adversarial loops add compute overhead	Moderate – attention mechanisms highlight entity relevance, aiding partial interpretability	Healthcare NER, social media analytics, product classification
GNN-Enhanced Language Models	Weighted F1: BERT 65.9 vs GNN-enhanced 80.7 (cross-lingual) [29]	Captures relational structures; strong multilingual transfer; interpretable via graph topology	Graph construction overhead; less scalable than transformers	Moderate–High – graph augmentation adds overhead but smaller parameter counts than LLMs	High – graph reasoning allows explicit interpretation of word/entity relations	Multilingual classification, disaster response, cross-lingual NER

TABLE 3: Comparative table of key Generative AI methods

LLM, Large Language Model; NER, Named Entity Recognition;

Challenges and future work

While generative AI has proven itself to be a brilliant capability, there are many remaining barriers to practical generative AI that vary from the efficiency and reliability of the production to the questions around its ethical deployment. This is important in order to develop text generation technologies responsibly.

Bias and Fairness

The problem with generative models is that they tend to replicate or even amplify social, cultural, and gender biases of the training data. The most common case that comes to mind is when these biases lead to toxic, offensive or exclusionary outputs that occur in sensitive domains like healthcare, education or legal advice [34].

Future direction: try not to have unfair training pipelines, work with suitably balanced target datasets, and use post-hoc debiasing tools. Fairness can be achieved better without compromising on the performance by incorporating techniques like adversarial debiasing and representation level audits.

Factual Hallucinations

Hallucination is the term for the factually incorrect but fluent outputs that large language models produce frequently. In particular, this becomes problematic for summarization, legal drafting, and medical documentation [35]. Examples of hallucinations are as follows [11]:

Example 1: Clinical trial summary

• References: clinical-trial abstract (no results reported).

• Model output: The randomized trial (n 5 00) showed a 45% mortality reduction ($p < 0.01$) with Drug X.

• Issue: the output fabricates sample size, effect size, and p-value; there is no such numerical information present in the source. Hallucination taxonomy (see Maynez et al. [11]) is proposed by them.

Example 2: Factual attribute fabrication

• Source: short biography without birthplace.

• Model result: Born in 1972 Paris (invented birthplace and date).

• Issue: the model was plausible but unsubstantiated fact; perilous in knowledge-based environments (see Ribeiro et al. [10] and evaluation surveys [24]).

Audit checklist

• Data provenance audit log: Data provenance and sampling procedures log and review training corpora composition (sensitive domains); record data provenance and sampling procedures.

• Factuality checks: test generated assertions with trusted ground-truth sources or ground outputs with retrieval-augmented generation (RAG).

• Failure-case sampling: systematically gather hallucinations in prompts/domains and classify them based on an existing taxonomy (e.g., Maynez [11]).

• Bias audit: run subgroup fairness evaluations (gender, race, locale) with purposeful prompts and examine disparate results.

• Human-in-loop validation: high-impact output must be reviewed by human; record rates of human disagreement and failure modes.

• Metric suite: combine automatic metrics (BLEURT, BERTScore [39,38]) and factuality and usefulness human annotations.

• Reporting: include example failures, provenance of data, and numerical/statistical claims seen or inferred in the sources when reporting results.

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1

Future direction: Take advantage of RAG frameworks, fact checking modules, and grounding of external knowledge (e.g., Wikidata, UMLS). The generation and fine-grained factuality scoring is an ongoing research.

Energy and Resource Inefficiency

Training and inference of such state-of-the art models (GPT3, PaLM) require giant computing resources. As a result, this results in high environmental cost and restricts accessibility for resources institutions [19].

Future direction: There may be research such as parameter-efficient fine tuning (such as LoRA, adapters), model pruning, quantization, and distillation, which can help to decrease energy consumption with only a loss of accuracy. More widely, green AI principles should be adopted.

Evaluation Metrics

BLEU, ROUGE, etc., are typical examples of traditional metrics as these words, in general, cannot capture much semantic quality, coherence, and factual correctness. Additionally, the evaluation of instruction-tuned and multimodal generative models is nonstandardized [20].

Future direction: Find or invent newer metrics of semantic fidelity such as BERTScore, BLEURT or MAUVE in order to encourage community-wide adoption of standardized evaluation suites like BIG-Bench, HELM or GEM.

Lack of Explainability

In high-stakes scenario, users cannot trust AI-generated content in large language models as the decision-making process in large language model is largely opaque.

Future direction: Create XAI frameworks for explicability over generative models (e.g., attention visualization, input attribution, chain of thought). Similarly, the nonexpert users also need interactive explanations [36].

Discussion

The reviewed approaches represent a broad range of capabilities (compared to generative AI approaches) to address various challenges in the text generation space. Previous approaches relied on rule-based and statistical models, but newer models such as transformer architectures, instruction-following models, entity-aware adaptations, and GNN-based integrated language models hold their own strengths in the field today.

Transformer models like GPT and T5 are highly fluent and context-aware, so they are best applied to open-domain generation tasks, such as chatbots, summarization, and storytelling. Their abilities in calculating and their tendency to cheat themselves with facts, however, limit their use in sensitive area such as health or law.

Pragmatic Flexibility

Few-shot and instruction-tuned models (e.g., InstructGPT) show high flexibility on a wide range of tasks with only minor fine-tuning. However, their prompt quality sensitivity presents a sense of randomness to the real-world user-facing application, particularly when the input is noisy or ungrammatical.

EADA models handle low-resource domain adaptation settings, including in medical or legal NER, where there is limited labeled data. They are semantically accurate across domains yet due to the complexity of adversarial training and task narrowness, scale is small.

Likewise, GNN-enhanced models are better at multilingual and cross-domain classification, in particular, when tasks are structured reasoning (e.g., social networks, hierarchical documents). However, preprocessing pipelines are computationally expensive and poorly exploited in practice.

Practical Barriers

In addition to these technical barriers, there are real-world constraints on the adoption of generative AI. One of the key concerns is interpretability: opaque black-box systems might not be trusted even by end-users (i.e., clinicians, educators or policy-makers), despite high technical accuracy. Regulatory compliance also makes deployment more

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1

complicated - models that do business in healthcare need to comply with HIPAA, and models working with European data are subject to GDPR. Accountability concerns (e.g., copyright in AI generated texts, misinformation liability) are unclear and, therefore, pose reluctance to adoption by institutions. Finally, scale deployment is computationally costly, limiting its availability in low-resource environments, and has sustainability consequences.

Stakeholder Views and Needs

Real-world application will necessitate alignment with requirements, limitations, and expectations of various stakeholders.

Patients/End-users: Patients need transparency, attribution of sources, and the ability to modify or refute outputs that affect their treatment. In the case of AI systems applied in patient-facing applications, explainability should be translated into lay-friendly justifications (e.g., “Why does the system think I should do this?”), and the system should indicate the confidence and provenance of any factual assertions.

Clinicians/Domain experts: Clinicians require audit trails, reference to original records, and strong correlation to clinical process. The interpretability must emphasize on evidence connecting (referral to medical records or literature) and point out the possible hallucinations or low-confidence recommendations. Data controls and logging are needed due to regulatory limits (e.g. HIPAA, local medical device regulations).

Developers/Technologists: Engineers are in need of modular, testable components: (i) grounding modules (retrieval/RAG), (ii) domain-adaptive components (e.g., EADA to specialized NER), and (iii) monitoring dashboards. All groups require that user-centered evaluation (tasks success, saved time, errors) should be used alongside typical ML measures.

Policy/Regulators: Regulatory stakeholders will require documented validation procedures, scope-of-use statements and post-deployment monitoring plans. Where feasible, provide model cards and data sheets that document training data composition, intended use, and known limitations.

In practice, these stakeholders should be engaged in co-design, pilot studies with mixed-method evaluation, and clear escalation paths for ambiguous or risky outputs.

Despite these barriers, hybrid solutions are promising. The fusion of GNN reasoning and few-shot models can improve entity coherence in zero-shot classification, while RAG can reduce hallucinations in knowledge-intensive tasks. For responsible deployment, it is essential to ensure AI explainability and compliance with regulatory frameworks by auditing datasets and reporting transparently.

Conclusions

Rule based systems were the first stage of generation of generation AI, followed by powerful deep learning models that could create text which is context aware and very creative. We reviewed four existing approaches to text generation that take different shades: few-shot instruction tuned models, CogView, entity aware domain adaptation, and GNN enhanced language models all tackle each of these problems including multi modal media creation, multilingual classification, and domain specific entity recognition.Transformer based models maintained their leadership in fluency and flexibility, as their domain adaptability, energy efficiency and interpretability are still sought by the practitioners. GNNs come with the structured reasoning, entity aware model retains the semantic precise over the domains, and the few-shot learning is an adaptable way of learning without retraining.Yet, these progress have not resolved the following challenges: Factual hallucination, bias propagation, computation and evaluation standards. To address these issues, we will need to be concerted in its use of explainability, fairness and responsible model design.

It is possible to develop hybrid architectures blending the advantages of several paradigms to address the shortcomings of each of them. Indicatively, GNN-enriched models of instruction may use the flexibility of instruction tuning and include graph-based reasoning that guarantees relational consistency of multiple language or domain-specific tasks. Likewise, generation trained with retrieval and augmented with adversarial domain adaptation (abbreviated RAG) can help reduce

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1

factual hallucinations but improve robustness in low-resource or specialized settings, including biomedical or legal ones. A second interesting direction is to incorporate entity-aware adversarial networks into retrieval systems, in which adversarial training enhances cross-domain robustness, and outputs are based on verified factual sources by retrieval modules. These hybrid strategies strike a balance between fluency, accuracy, and adaptability to produce technically powerful and practically reliable models. Such integrative methods are critical to the development of the technical edge of generative AI, as well as to its alignment with more general values of equity, accessibility, and reliability. With hybrid architectures overcoming issues of interpretability, factual reliability, and fairness, future systems will be deployed in a responsible manner in sensitive fields, such as healthcare, education, and policy-making.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Hrithik A. Singh, Vanita Mane, Tushar Ghorpade

Acquisition, analysis, or interpretation of data: Hrithik A. Singh, Vanita Mane, Tushar Ghorpade

Drafting of the manuscript: Hrithik A. Singh, Vanita Mane, Tushar Ghorpade

Critical review of the manuscript for important intellectual content: Hrithik A. Singh, Vanita Mane, Tushar Ghorpade

Supervision: Hrithik A. Singh, Vanita Mane, Tushar Ghorpade

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The datasets and codes generated and/or analyzed during this study are included in this article or are available from the corresponding author upon reasonable request.

References

1. Hochreiter S, Schmidhuber J: Long short-term memory. Neural Computation. 1997, 9:1735-1780. 10.1162/neco.1997.9.8.1735
2. Staudemeyer RC, Morris ER: Understanding LSTM -- a tutorial into long short-term memory recurrent neural networks. arXiv. 2019, 10.48550/arXiv.1909.09586
3. Vaswani A, Shazeer N, Parmar N, et al.: Attention is all you need. arXiv. 2017, 10.48550/arXiv.1706.03762
4. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I: Language models are unsupervised multitask learners. OpenAI. 2019, 1-24.
5. Radford A, Narasimhan K, Salimans T, Sutskever I: Improving language understanding by generative pre-training. OpenAI. 2018, 1-12.
6. Lai H, Nissim M: A survey on automatic generation of figurative language: From rule-based systems to large language models. ACM Computing Surveys. 2024, 56:1-34. 10.1145/3654795

7. Katikireddi PM, Jaini S: In Generative AI: zero-shot and few-shot. International Journal of Scientific Research in Computer Science Engineering and Information Technology. 2022, 8:391-397. 10.32628/cseit2390668

8. Bender EM, Gebru T, McMillan-Major A: On the dangers of stochastic parrots: can language models be too big. FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021, 610-623. 10.1145/3442188.3445922

9. Strubell E, Ganesh A, McCallum A: Energy and policy considerations for deep learning research. arXiv. 2019, 10.48550/arXiv.1906.02243

10. Ribeiro MT, Singh S, Guestrin C: Why should I trust you? Explaining the predictions of any classifier. arXiv. 2016, 10.48550/arXiv.1602.04938

11. Maynez J, Narayan S, Bohnet B, McDonald R: On faithfulness and factuality in abstractive summarization. arXiv. 2020, 10.48550/arXiv.2005.00661

12. Ding M, Yang Z, Hong W, et al.: CogView: mastering text-to-image generation via transformers. arXiv. 2021, 10.48550/arXiv.2105.13290

13. Ramesh A, Pavlov M, Goh G, et al.: Zero-shot text-to-image generation. arXiv. 2021, 10.48550/arXiv.2102.12092

14. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M: Hierarchical text-conditional image generation with CLIP latents. arXiv. 2022, 10.48550/arXiv.2204.06125

15. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B: High-resolution image synthesis with latent diffusion models. arXiv. 2022, 10.48550/arXiv.2112.10752

16. Nichol A, Dhariwal P, Ramesh A, et al.: GLIDE: towards photorealistic image generation and editing with text-guided diffusion models. arXiv. 2021, 10.48550/arXiv.2112.10741

17. Alayrac JB, Donahue J, Luc P, et al.: Flamingo: a visual language model for few-shot learning. arXiv. 2022, 10.48550/arXiv.2204.14198

18. Brown TB, Mann B, Ryder N, et al.: Language models are few-shot learners. arXiv. 2020, 10.48550/arXiv.2005.14165

19. Ouyang L, Wu J, Jiang X, et al.: Training language models to follow instructions with human feedback. arXiv. 2022, 10.48550/arXiv.2203.02155

20. Sanh V, Webson A, Raffel C, et al.: Multitask prompted training enables zero-shot task generalization. arXiv. 2021, 10.48550/arXiv.2110.08207

21. Wei J, Bosma M, Zhao VY, et al.: Finetuned language models are zero-shot learners. arXiv. 2021, 10.48550/arXiv.2109.01652

22. Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G: Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. ACM Computing Surveys. 2023, 55:1-35. 10.1145/3560815

23. Stiennon N, Ouyang L, Wu J, et al.: Learning to summarize from human feedback. arXiv. 2020, 10.48550/arXiv.2009.01325

24. Celikyilmaz A, Clark E, Gao J: Evaluation of text generation: a survey. arXiv. 2021, 10.48550/arXiv.2006.14799

25. Wu Z, Galley M, Brockett C, et al.: A controllable model of grounded response generation. arXiv. 2020, 10.48550/arXiv.2005.00613

26. Lin C-H, Zhou K, Li L, Sun L: Integrating generative AI into digital multimodal composition: A study of multicultural second-language classrooms. Computers and Composition. 2025, 75:102895. 10.1016/j.compcom.2024.102895

27. He T, Zhang J, Wang T, Kumar S, Cho K, Glass J, Tsvetkov Y: On the blind spots of model-based evaluation metrics for text generation. arXiv. 2021, 10.48550/arXiv.2212.10020

28. Jin G, Xu K, Chen H, Jin Y, Zhu C: A novel multi-adversarial cross-domain neural network for bearing fault diagnosis. Measurement Science and Technology. 2021, 32:055102. 10.1088/1361-6501/abd900

29. Hoyle A, Goel P, Resnik P: Improving neural topic models using knowledge distillation. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020, 1752-1771. 10.18653/v1/2020.emnlp-main.137

30. Zhang T, Xia C, Liu Z, Zhao S, Peng H, Yu P: Domain-invariant feature progressive distillation with adversarial adaptive augmentation for low-resource cross-domain NER. ACM Transactions on Asian and Low-Resource Language Information Processing. 2023, 22:1-21. 10.1145/3570502

31. Ma N, Mazumder S, Wang H, Liu B: Entity-aware dependency-based deep graph attention network for comparative preference classification. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020, 5782-5788. 10.18653/v1/2020.acl-main.512

32. Yao L, Mao C, Luo Y: Graph convolutional networks for text classification. Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33:7370-7377. 10.1609/aaai.v33i01.33017370

33. Wang K, Ding Y, Han SC: Graph neural networks for text classification: a survey. Artificial Intelligence Review. 2024, 57:190. 10.1007/s10462-024-10808-0

34. Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS: A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems. 2021, 32:4-24. 10.1109/tnnls.2020.2978386

35. Huang Y, Yu Z, Guo J, Xiang Y, Xian Y: Element graph-augmented abstractive summarization for legal public opinion news with graph transformer. Neurocomputing. 2021, 460:166-180. 10.1016/j.neucom.2021.07.013

36. Lewis M, Liu Y, Goyal N, et al.: BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020, 7871-7880. 10.18653/v1/2020.acl-main.703

37. Zhang Y, Sun S, Galley M, et al.: DIALOGPT: large-scale generative pre-training for conversational response generation. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2020, 270-278. 10.18653/v1/2020.acl-demos.30

38. Zhang T, Kishore V, Wu F, Weinberger KQ, Artzi Y: BERTScore: evaluating text generation with BERT. arXiv. 2020, 10.48550/arXiv.1904.09675

39. Sellam T, Das D, Parikh AP: BLEURT: learning robust metrics for text generation. arXiv. 2020, 10.48550/arXiv.2004.04696

40. Sheng E, Chang KW, Natarajan P, Peng N: The woman worked as a babysitter: on biases in language generation. arXiv. 2019, 10.48550/arXiv.1909.01326

41. Jiang N, Hasanzadeh S, Duffy VG: Domain-tailored generative AI for personalized assistant. HCI International 2024 - Late Breaking Papers. HCII 2024. Duffy VG (ed): Springer, Cham; 2024. 15376:233-249. 10.1007/978-3-031-76809-5_17

42. Dinan E, Fan A, Wu L, Weston J, Kiela D, Williams A: Multi-dimensional gender bias classification. arXiv. 2020, 10.48550/arXiv.2005.00614

43. Qian W, Xiong Y, Yang J, Shu W: Feature selection for label distribution learning via feature similarity and label correlation. Information Sciences. 2022, 582:38-59. 10.1016/j.ins.2021.08.076

44. Wei Y, Jiang YH, Liu J, Qi C, Jia L, Jia R: The advancement of personalized learning potentially accelerated by generative AI. arXiv. 2024, 10.48550/arXiv.2412.00691

How to cite this article:
Singh H A, Mane V, Ghorpade T (February 26, 2026) Generative Artificial Intelligence Approaches for Text Generation: Techniques, Challenges, and Future Directions. Cureus J Comput Sci 3 : es44389-025-00035-1. DOI https://doi.org/10.7759/s44389-025-00035-1