	To Consider	Don't Consider	paperId	url	title	venue	year	reference	citationC	influent	abstract	externalIds.PubMed	publicationVenue.u	subquery	journal	externalIds.MAG	
0	☐	☑	39ba6d541d941	https://www.semanticscholar.	Training-Free Consistent Text-to-Image Generation	ACM Transactions o	2024.0	55.0	113.0	27.0	Text-to-image models offer a new level of creative flexibility by allowing users to guide the image generation process through natural language. However, using these models to consistently portray the same subject across diverse prompts remains challenging. Existing approaches fine-tune the model to teach it new words that describe specific user-provided subjects or add image conditioning to the model. These methods require lengthy persubject optimization or large-scale pre-training. Moreover, they struggle to align generated images with text prompts and face difficulties in portraying multiple subjects. Here, we present ConsiStory, a training-free approach that enables consistent subject generation by sharing the internal activations of the pretrained model. We introduce a subject-driven shared attention block and correspondence-based feature injection to promote subject consistency between images. Additionally, we develop strategies to encourage layout diversity while maintaining subject consistency. We compare ConsiStory to a range of baselines, and demonstrate state-of-the-art performance on subject consistency and text alignment, without requiring a single optimization step. Finally, ConsiStory can naturally extend to multi-subject scenarios, and even enable training-free personalization for common objects.		http://www.acm.org	Paraphrase | NLP | Natural Language Generation | Language Model | Bias | Privacy | Controllable | Creative | Machine | Automated | Tra			
1	☐	☑	2078ff863ad473	https://www.semanticscholar.	MIGC: Multi-Instance Generation Controller for Text-to-Image Synthesis	Computer Vision an	2024.0	69.0	103.0	18.0	We present a Multi-Instance Generation (MIG) task, si-multaneously generating multiple instances with diverse controls in one image. Given a set of predefined coordinates and their corresponding descriptions, the task is to ensure that generated instances are accurately at the designated locations and that all instances' attributes adhere to their corresponding description. This broadens the scope of current research on Single-instance generation, elevating it to a more versatile and practical dimension. Inspired by the idea of divide and conquer, we introduce an innovative approach named Multi-Instance Generation Controller (MIGC) to address the challenges of the MIG task. Ini-tially, we break down the MIG task into several subtasks, each involving the shading of a single instance. To ensure precise shading for each instance, we introduce an instance enhancement attention mechanism. Lastly, we aggregate all the shaded instances to provide the necessary information for accurately generating multiple instances in stable diffusion (SD). To evaluate how well generation models per-form on the MIG task, we provide a COCO-MIG bench-mark along with an evaluation pipeline. Extensive experiments were conducted on the proposed COCO-MIG bench-mark, as well as on various commonly used benchmarks. The evaluation results illustrate the exceptional control ca-pabilities of our model in terms of quantity, position, at-tribute, and interaction. Code and demos will be released at https://migcproject.github.io/.		https://ieeexplore.ie	Paraphrase | NLP | Evaluation Metrics | Bias | Privacy | Controllable | Creative | Machine | Automated | Task | Cost | Detection | Text Ger			
2	☐	☑	3f19484f941f209	https://www.semanticscholar.	GPT-4V(ision) is a Human-Aligned Evaluator for Text-to-3D Generation	Computer Vision an	2024.0	73.0	135.0	16.0	Despite recent advances in text-to-3D generative methods, there is a notable absence of reliable evaluation metrics. Existing metrics usually focus on a single criterion each, such as how well the asset aligned with the input text. These metrics lack the flexibility to generalize to different evaluation criteria and might not align well with human preferences. Conducting user preference studies is an alternative that offers both adaptability and human-aligned results. User studies, however, can be very ex-pensive to scale. This paper presents an automatic, ver-satile, and human-aligned evaluation metric for text-to-3D generative models. To this end, we first develop a prompt generator using GPT-4V to generate evaluating prompts, which serve as input to compare text-to-3D models. We further design a method instructing GPT-4V to compare two 3D assets according to user-defined crite-ria. Finally, we use these pairwise comparison results to assign these models Elo ratings. Experimental results suggest our metric strongly aligns with human preference across different evaluation criteria. Our code is available at https://github.com/3DTopia/GPTEval3D.		https://ieeexplore.ie	Paraphrase | NLP | Evaluation Metrics | Bias | Privacy | Controllable | Creative | Machine | Automated | Cost | Detection | Text Generatio			
3	☐	☑	ba08ea91272491	https://www.semanticscholar.	Taming Stable Diffusion for Text to 360° Panorama Image Generation	Computer Vision an	2024.0	66.0	43.0	15.0			https://ieeexplore.ie	Paraphrase | NLP | Bias | Privacy | Controllable | Creative | Automated | Cost | Detection | Text Generation (None)			
4	☐	☑	c4b83e361b417	https://www.semanticscholar.	Text2HOI: Text-Guided 3D Motion Generation for Hand-Object Interaction	Computer Vision an	2024.0	36.0	57.0	10.0	This paper introduces the first text-guided work for generating the sequence of hand-object interaction in 3D. The main challenge arises from the lack of labeled data where existing ground-truth datasets are nowhere near generalizable in interaction type and object category, which inhibits the modeling of diverse 3D hand-object interaction with the correct physical implication (e.g., contacts and semantics) from text prompts. To address this challenge, we propose to decompose the interaction generation task into two subtasks: hand-object contact generation; and hand-object motion generation. For contact generation, a VAE-based network takes as input a text and an object mesh, and generates the probability of contacts between the surfaces of hands and the object during the interaction. The network learns a variety of local geometry structure of diverse objects that is independent of the objects' category, and thus, it is applicable to general objects. For motion generation, a Transformer-based diffusion model utilizes this 3D contact map as a strong prior for generating physically plausible hand-object motion as a function of text prompts by learning from the augmented labeled dataset, where we annotate text labels from many existing 3D hand and object motion data. Finally, we further introduce a hand refiner module that minimizes the distance between the object surface and hand joints to improve the temporal stability of the object-hand contacts and to suppress the penetration artifacts. In the experiments, we demonstrate that our method can generate more realistic and diverse interactions compared to other baseline methods. We also show that our method is applicable to unseen objects. We will release our model and newly labeled data as a strong foundation for future research. Codes and data are available in: https://github.com/JunukCha/Text2HOI.		https://ieeexplore.ie	Paraphrase | NLP | Bias | Privacy | Controllable | Creative | Automated | Task | Cost | Detection | Text Generation (None)			
5	☑	☐	5d4a6ce39cfe7b	https://www.semanticscholar.	A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges	IEEE Access	2024.0	181.0	476.0	20.0	Large Language Models (LLMs) recently demonstrated extraordinary capability in various natural language processing (NLP) tasks including language translation, text generation, question answering, etc. Moreover, LLMs are new and essential part of computerized language processing, having the ability to understand complex verbal patterns and generate coherent and appropriate replies in a given context. Though this success of LLMs has prompted a substantial increase in research contributions, rapid growth has made it difficult to understand the overall impact of these improvements. Since a plethora of research on LLMs have been appeared within a short time, it is quite impossible to track all of these and get an overview of the current state of research in this area. Consequently, the research community would benefit from a short but thorough review of the recent changes in this area. This article thoroughly overviews LLMs, including their history, architectures, transformers, resources, training methods, applications, impacts, challenges, etc. This paper begins by discussing the fundamental concepts of LLMs with its traditional pipeline of the LLMs training phase. Then the paper provides an overview of the existing works, the history of LLMs, their evolution over time, the architecture of transformers in LLMs, the different resources of LLMs, and the different training methods that have been used to train them. The paper also demonstrates the datasets utilized in the studies. After that, the paper discusses the wide range of applications of LLMs, including biomedical and healthcare, education, social, business, and agriculture. The study also illustrates how LLMs create an impact on society and shape the future of AI and how they can be used to solve real-world problems. Finally, the paper also explores open issues and challenges to deploy LLMs in real-world scenario. Our review paper aims to help practitioners, researchers, and experts thoroughly understand the evolution of LLMs, pre-trained architectures, applications, challenges, and future goals.		http://www.ieee.org	NLP | Natural Language Generation | Training			
6	☐	☑	978040c55f9094	https://www.semanticscholar.	CodeS: Towards Building Open-source Language Models for Text-to-SQL	Proc. ACM Manag. I	2024.0	87.0	210.0	43.0	Language models have shown promising performance on the task of translating natural language questions into SQL queries (Text-to-SQL). However, most of the state-of-the-art (SOTA) approaches rely on powerful yet closed-source large language models (LLMs), such as ChatGPT and GPT-4, which may have the limitations of unclear model architectures, data privacy risks, and expensive inference overheads. To address the limitations, we introduce CodeS, a series of pre-trained language models with parameters ranging from 1B to 15B, specifically designed for the text-to-SQL task. CodeS is a fully open-source language model, which achieves superior accuracy with much smaller parameter sizes. This paper studies the research challenges in building CodeS. To enhance the SQL generation abilities of CodeS, we adopt an incremental pre-training approach using a specifically curated SQL-centric corpus. Based on this, we address the challenges of schema linking and rapid domain adaptation through strategic prompt construction and a bi-directional data augmentation technique. We conduct comprehensive evaluations on multiple datasets, including the widely used Spider benchmark, the newly released BIRD benchmark, robustness-diagnostic benchmarks such as Spider-DK, Spider-Syn, Spider-Realistic, and Dr.Spider, as well as two real-world datasets created for financial and academic applications. The experimental results show that our CodeS achieves new SOTA accuracy and robustness on nearly all challenging text-to-SQL benchmarks.			Natural Language Generation | Language Model | Evaluation Metrics | Privacy | Task | Training			
7	☐	☑	587547493b5cf2	https://www.semanticscholar.	Cell2Sentence: Teaching Large Language Models the Language of Biology	bioRxiv	2024.0	79.0	48.0	10.0	We introduce Cell2Sentence (C2S), a novel method to directly adapt large language models to a biological context, specifically single-cell transcriptomics. By transforming gene expression data into "cell sentences," C2S bridges the gap between natural language processing and biology. We demonstrate cell sentences enable the fine-tuning of language models for diverse tasks in biology, including cell generation, complex cell-type annotation, and direct data-driven text generation. Our experiments reveal that GPT-2, when fine-tuned with C2S, can generate biologically valid cells based on cell type inputs, and accurately predict cell types from cell sentences. This illustrates that language models, through C2S fine-tuning, can acquire a significant understanding of single-cell biology while maintaining robust text generation capabilities. C2S offers a flexible, accessible framework to integrate natural language processing with transcriptomics, utilizing existing models and libraries for a wide range of biological applications.	39554079.0	http://biorxiv.org/	Natural Language Generation			
8	☐	☑	969e53e9e8714	https://www.semanticscholar.	Auffusion: Leveraging the Power of Diffusion and Large Language Models for Text-to-Audio Generation	IEEE/ACM Transact	2024.0	62.0	61.0	9.0	Recent advancements in diffusion models and large language models (LLMs) have significantly propelled the field of generation tasks. Text-to-Audio (TTA), a burgeoning generation application designed to generate audio from natural language prompts, is attracting increasing attention. However, existing TTA studies often struggle with generation quality and text-audio alignment, especially for complex textual inputs. Drawing inspiration from state-of-the-art Text-to-Image (T2I) diffusion models, we introduce Auffusion, a TTA system adapting T2I model frameworks to TTA task, by effectively leveraging their inherent generative strengths and precise cross-modal alignment. Our objective and subjective evaluations demonstrate that Auffusion surpasses previous TTA approaches using limited data and computational resources. Furthermore, the text encoder serves as a critical bridge between text and audio, since it acts as an instruction for the diffusion model to generate coherent content. Previous studies in T2I recognize the significant impact of encoder choice on cross-modal alignment, like fine-grained details and object bindings, while similar evaluation is lacking in prior TTA works. Through comprehensive ablation studies and innovative cross-attention map visualizations, we provide insightful assessments, being the first to reveal the internal mechanisms in the TTA field and intuitively explain how different text encoders influence the diffusion process. Our findings reveal Auffusion's superior capability in generating audios that accurately match textual descriptions, which is further demonstrated in several related tasks, such as audio style transfer, inpainting, and other manipulations.		https://ieeexplore.ie	Paraphrase | NLP | Natural Language Generation | Language Model | Evaluation Metrics | Bias | Privacy | Controllable | Creative | Auton			
9	☐	☑	75d61f9286e298	https://www.semanticscholar.	Sniffer: Multimodal Large Language Model for Explainable Out-of-Context Misinformation Detection	Computer Vision an	2024.0	59.0	84.0	16.0	Misinformation is a prevalent societal issue due to its potential high risks. Out-Of-Context (OOC) misinformation, where authentic images are repurposed with false text, is one of the easiest and most effective ways to mislead audiences. Current methods focus on assessing image- text consistency but lack convincing explanations for their judgments, which are essential for debunking misinformation. While Multimodal Large Language Models (MLLMs) have rich knowledge and innate capability for visual rea- soning and explanation generation, they still lack sophisti- cation in understanding and discovering the subtle cross- modal differences. In this paper, we introduce Sniffer,a novel multimodal large language model specifically engi- neered for OOC misinformation detection and explanation. Snifferemploys two-stage instruction tuning on Instruct- BLIP. The first stage refines the model's concept alignment of generic objects with news-domain entities and the sec- ond stage leverages OOC-specific instruction data gener- ated by language-only GPT-4 to fine-tune the model's dis- criminatory powers. Enhanced by external tools and re- trieval, Sniffernot only detects inconsistencies between text and image but also utilizes external knowledge for con- textual verification. Our experiments show that Sniffersurpasses the original MLLM by over 40% and outperforms state-of-the-art methods in detection accuracy. Snifferalso provides accurate and persuasive explanations as val- idated by quantitative and human evaluations.		https://ieeexplore.ie	Language Model | Detection			

10	☑	☐	4fcd49afa8d0a9f	https://www.semanticscholar.org/paper/4fcd49afa8d0a9f	Large Language Model guided Protocol Fuzzing	Network and Distribu	2024.0	45.0	206.0	16.0	—How to find security flaws in a protocol implementation without a machine-readable specification of the protocol? Facing the internet, protocol implementations are particularly security-critical software systems where inputs must adhere to a specific structure and order that is often informally specified in hundreds of pages in natural language (RFC). Without some machine-readable version of that protocol, it is difficult to automatically generate valid test inputs for its implementation that follow the required structure and order. It is possible to partially alleviate this challenge using mutational fuzzing on a set of recorded message sequences as seed inputs. However, the set of available seeds is often quite limited and will hardly cover the great diversity of protocol states and input structures. In this paper, we explore the opportunities of systematic interaction with pre-trained large language models (LLMs), which have ingested millions of pages of human-readable protocol specifications, to draw out machine-readable information about the protocol that can be used during protocol fuzzing. We use the knowledge of the LLMs about protocol message types for well-known protocols. We also checked the LLM's capability in detecting "states" for stateful protocol implementations by generating sequences of messages and predicting response codes. Based on these observations, we have developed an LLM-guided protocol implementation fuzzing engine. Our protocol fuzzer C HAT AFL constructs grammars for each message type in a protocol, and then mutates messages or predicts the next messages in a message sequence via interactions with LLMs. Experiments on a wide range of real-world protocols from P RO F UZZ B ENCH show significant efficacy in state and code coverage. Our LLM-guided stateful fuzzer was compared with state-of-the-art fuzzers AFLN ET and NSF UZZ . C HAT AFL covers 47.60% and 42.69% more state transitions, 29.55% and 25.75% more				http://www.isoc.org/	Language Model
11	☑	☐	da54bda4ef0ba0	https://www.semanticscholar.org/paper/da54bda4ef0ba0	Structured information extraction from scientific text with large language models	Nature Communicat	2024.0	64.0	396.0	13.0	Extracting structured knowledge from scientific text remains a challenging task for machine learning models. Here, we present a simple approach to joint named entity recognition and relation extraction and demonstrate how pretrained large language models (GPT-3, Llama-2) can be fine-tuned to extract useful records of complex scientific knowledge. We test three representative tasks in materials chemistry: linking dopants and host materials, cataloging metal-organic frameworks, and general composition/phase/morphology/application information extraction. Records are extracted from single sentences or entire paragraphs, and the output can be returned as simple English sentences or a more structured format such as a list of JSON objects. This approach represents a simple, accessible, and highly flexible route to obtaining large databases of structured specialized scientific knowledge extracted from research papers.		38360817.0	https://www.nature.	Language Model | Task
12	☑	☐	037b4345769b0(	https://www.semanticscholar.org/paper/037b4345769b0(	On the Evaluation of Large Language Models in Unit Test Generation	International Confer	2024.0	91.0	64.0	8.0	Unit testing is an essential activity in software development for verifying the correctness of software components. However, manually writing unit tests is challenging and time-consuming. The emergence of Large Language Models (LLMs) offers a new direction for automating unit test generation. Existing research primarily focuses on closed-source LLMs (e.g., ChatGPT and CodeX) with fixed prompting strategies, leaving the capabilities of advanced open-source LLMs with various prompting settings unexplored. Particularly, open-source LLMs offer advantages in data privacy protection and have demonstrated superior performance in some tasks. Moreover, effective prompting is crucial for maximizing LLMs' capabilities. In this paper, we conduct the first empirical study to fill this gap, based on 17 Java projects, five widely-used open-source LLMs with different structures and parameter sizes, and comprehensive evaluation metrics. Our findings highlight the significant influence of various prompt factors, show the performance of open-source LLMs compared to the commercial GPT-4 and the traditional Evosuite, and identify limitations in LLM-based unit test generation. We then derive a series of implications from our study to guide future research and practical use of LLM-based unit test generation.CCS CONCEPTS • Software and its engineering → Software testing and debugging.				http://ase.informatik	Evaluation Metrics
13	☐	☑	450011c7e08967	https://www.semanticscholar.org/paper/450011c7e08967	Motion-I2V: Consistent and Controllable Image-to-Video Generation with Explicit Motion Modeling	International Confer	2024.0	88.0	165.0	12.0	We introduce Motion-I2V, a novel framework for consistent and controllable text-guided image-to-video generation (I2V). In contrast to previous methods that directly learn the complicated image-to-video mapping, Motion-I2V factorizes I2V into two stages with explicit motion modeling. For the first stage, we propose a diffusion-based motion field predictor, which focuses on deducing the trajectories of the reference image's pixels. For the second stage, we propose motion-augmented temporal attention to enhance the limited 1-D temporal attention in video latent diffusion models. This module can effectively propagate reference image features to synthesized frames with the guidance of predicted trajectories from the first stage. Compared with existing methods, Motion-I2V can generate more consistent videos even in the presence of large motion and viewpoint variation. By training a sparse trajectory ControlNet for the first stage, Motion-I2V can support users to precisely control motion trajectories and motion regions with sparse trajectory and region. This offers more controllability of the I2V process than solely relying on textual instructions. Additionally, Motion-I2V's second stage naturally supports zero-shot video-to-video translation. Both qualitative and quantitative comparisons demonstrate the advantages of Motion-I2V over prior approaches in consistent and controllable image-to-video generation. Please see our project page at https://xiaoyushi97.github.io/Motion-I2V/.				http://www.siggraph.org	Controllable | Training
14	☐	☑	a008cc8a48467(	https://www.semanticscholar.org/paper/a008cc8a48467(	Towards Understanding Cross and Self-Attention in Stable Diffusion for Text-Guided Image Editing	Computer Vision an	2024.0	41.0	116.0	13.0	Deep Text-to-Image Synthesis (TIS) models such as Stable Diffusion have recently gained significant popularity for creative text-to-image generation. However, for domain-specific scenarios, tuning-free Text-guided Image Editing (TIE) is of greater importance for application developers. This approach modifies objects or object properties in images by manipulating feature components in attention layers during the generation process. Nevertheless, little is known about the semantic meanings that these attention layers have learned and which parts of the attention maps contribute to the success of image editing. In this paper, we conduct an in-depth probing analysis and demonstrate that cross-attention maps in Stable Diffusion often contain object attribution information, which can result in editing failures. In contrast, self-attention maps play a crucial role in preserving the geometric and shape details of the source image during the transformation to the target image. Our analysis offers valuable insights into understanding cross and self-attention mechanisms in diffusion models. Furthermore, based on our findings, we propose a simplified, yet more stable and efficient, tuning-free procedure that modifies only the self-attention maps of specified attention layers during the denoising process. Experimental results show that our simplified method consistently surpasses the performance of popular approaches on multiple datasets. 11Source code and datasets are available at https://github.com/ailibaba/EasyNLP/tree/master/diffusion/FreePromptEditing.				https://ieeexplore.ie	Creative
15	☐	☑	9761bcf4989260	https://www.semanticscholar.org/paper/9761bcf4989260	Diffusion Model-Based Image Editing: A Survey	IEEE Transactions o	2024.0	329.0	189.0	7.0	Denoising diffusion models have emerged as a powerful tool for various image generation and editing tasks, facilitating the synthesis of visual content in an unconditional or input-conditional manner. The core idea behind them is learning to reverse the process of gradually adding noise to images, allowing them to generate high-quality samples from a complex distribution. In this survey, we provide an exhaustive overview of existing methods using diffusion models for image editing, covering both theoretical and practical aspects in the field. We delve into a thorough analysis and categorization of these works from multiple perspectives, including learning strategies, user-input conditions, and the array of specific editing tasks that can be accomplished. In addition, we pay special attention to image inpainting and outpainting, and explore both earlier traditional context-driven and current multimodal conditional methods, offering a comprehensive analysis of their methodologies. To further evaluate the performance of text-guided image editing algorithms, we propose a systematic benchmark, EditEval, featuring an innovative metric, LMM Score. Finally, we address current limitations and envision some potential directions for future research.		40031849.0	http://www.computer	Machine
16	☑	☐	29fed0dcbf101c3	https://www.semanticscholar.org/paper/29fed0dcbf101c3	ChatGPT Label: Comparing the Quality of Human-Generated and LLM-Generated Annotations in Low-Resource Language NLP Tasks	IEEE Access	2024.0	42.0	76.0	6.0	This research paper presents a comprehensive comparative study assessing the quality of annotations in Turkish, Indonesian, and Minangkabau Natural Language Processing (NLP) tasks, with a specific focus on the contrast between annotations generated by human annotators and those produced by Large Language Models (LLMs). In the context of NLP, high-quality annotations play a pivotal role in training and evaluating machine-learning models. The study encompasses three core NLP tasks: topic classification, tweet sentiment analysis, and emotion classification, each reflecting a distinct aspect of text analysis. The research methodology incorporates a meticulously curated dataset sourced from a variety of text data, spanning diverse topics and emotions. Human annotators, proficient in the Turkish, Indonesian, and Minangkabau language, were tasked with producing high-quality annotations, adhering to comprehensive annotation guidelines. Additionally, fine-tuned Turkish LLMs were employed to generate annotations for the same tasks. The evaluation process employed precision, recall, and F1-score metrics, tailored to each specific NLP task. The findings of this study underscore the nuanced nature of annotation quality. While LLM-generated annotations demonstrated competitive quality, particularly in sentiment analysis, human-generated annotations consistently outperformed LLM-generated ones in more intricate NLP tasks. The observed differences highlight LLM limitations in understanding context and addressing ambiguity. This research contributes to the ongoing discourse on annotation sources in Turkish, Indonesian, and Minangkabau NLP, emphasizing the importance of judicious selection between human and LLM-generated annotations. It also underscores the necessity for continued advancements in LLM capabilities, as they continue to reshape the landscape of data annotation in NLP and machine learning.			http://www.ieee.org	Paraphrase | NLP | Natural Language Generation | Language Model | Evaluation Metrics | Bias | Privacy | Controllable | Creative | Mach
17	☑	☐	a8a42eaf78d66b	https://www.semanticscholar.org/paper/a8a42eaf78d66b	Exploring and Unleashing the Power of Large Language Models in Automated Code Translation	Proc. ACM Softw. Er	2024.0	69.0	125.0	19.0	Code translation tools, namely transpilers, are developed for automatic source-to-source translation. Latest learning-based transpilers have shown impressive enhancement against rule-based counterparts in both translation accuracy and readability, owing to their task-specific pre-training on extensive monolingual corpora. Nevertheless, their current performance still remains unsatisfactory for practical deployment, and the associated training resources are also prohibitively expensive. Large Language Models (LLMs), pre-trained on huge amounts of human-written code/text, have shown remarkable performance in many code intelligence tasks due to their powerful generality, even without task-specific re-training/fine-tuning. Thus, LLMs can potentially circumvent the above limitations, but they have not been exhaustively explored yet. This paper investigates diverse LLMs and learning-based transpilers for automated code translation tasks, finding that: although certain LLMs have outperformed current transpilers, they still have some accuracy issues, where most of the failures are induced by a lack of comprehension of source programs (38.51%), missing clear instructions on I/O types in translation (14.94%), and ignoring discrepancies between source and target programs (41.38%). Enlightened by the above findings, we further propose UniTrans, a Unified code Translation framework, applicable to various LLMs, for unleashing their power in this field. Specifically, UniTrans first crafts a series of test cases for target programs with the assistance of source programs. Next, it harnesses the above auto-generated test cases to augment the code translation and then evaluate their correctness via execution. Afterward, UniTrans further (iteratively) repairs incorrectly translated programs prompted by test case execution results. Extensive experiments are conducted on six settings of translation datasets between Python, Java, and C++. Three recent LLMs of diverse sizes, including GPT-3.5 and LLaMA-13B/7B, are tested with UniTrans, and all achieve substantial improvements in terms of computational accuracy and exact match accuracy among almost all translation settings, showing the universal effectiveness of UniTrans in practice.				Automated
18	☑	☐	603d3f90fc40f79	https://www.semanticscholar.org/paper/603d3f90fc40f79	AI models collapse when trained on recursively generated data	Nature	2024.0	11.0	515.0	37.0	Stable diffusion revolutionized image creation from descriptive text. GPT-2 (ref. 1), GPT-3(.5) (ref. 2) and GPT-4 (ref. 3) demonstrated high performance across a variety of language tasks. ChatGPT introduced such language models to the public. It is now clear that generative artificial intelligence (AI) such as large language models (LLMs) is here to stay and will substantially change the ecosystem of online text and images. Here we consider what may happen to GPT-{n} once LLMs contribute much of the text found online. We find that indiscriminate use of model-generated content in training causes irreversible defects in the resulting models, in which tails of the original content distribution disappear. We refer to this effect as 'model collapse' and show that it can occur in LLMs as well as in variational autoencoders (VAEs) and Gaussian mixture models (GMMs). We build theoretical intuition behind the phenomenon and portray its ubiquity among all learned generative models. We demonstrate that it must be taken seriously if we are to sustain the benefits of training from large-scale data scraped from the web. Indeed, the value of data collected about genuine human interactions with systems will be increasingly valuable in the presence of LLM-generated content in data crawled from the Internet. Analysis shows that indiscriminately training generative artificial intelligence on real and generated content, usually done by scraping data from the Internet, can lead to a collapse in the ability of the models to generate diverse high-quality output.		39048682.0	https://www.nature.	Task | Training

19	☐	☑	492bc8339d8aac	https://www.semanticscholar.	VideoCrafter2: Overcoming Data Limitations for High-Quality Video Diffusion Models	Computer Vision an	2024.0	54.0	484.0	63.0	Text-to-video generation aims to produce a video based on a given prompt. Recently, several commercial video models have been able to generate plausible videos with mini-mal noise, excellent details, and high aesthetic scores. However, these models rely on large-scale, well-filtered, high-quality videos that are not accessible to the community. Many existing research works, which train models using the low-quality WebVid-10M dataset, struggle to generate high-quality videos because the models are optimized to fit WebVid-10M. In this work, we explore the training scheme of video models extended from Stable Diffusion and investigate the feasibility of leveraging low-quality videos and synthesized high-quality images to obtain a high-quality video model. We first analyze the connection between the spatial and temporal modules of video models and the distribution shift to low-quality videos. We observe that full training of all modules results in a stronger coupling between spatial and temporal modules than only training temporal modules. Based on this stronger coupling, we shift the distribution to higher quality without motion degradation by finetuning spatial modules with high-quality images, resulting in a generic high-quality video model. Evaluations are conducted to demonstrate the superiority of the proposed method, particularly in picture quality, motion, and concept composition.			https://ieeexplore.ie	Training
20	☐	☑	39b270b9cb8d4f	https://www.semanticscholar.	Integrating Forecasting Service and Gen2 Blockchain Into a Local Energy Trading Platform to Promote Sustainability Goals	IEEE Access	2024.0	79.0	11.0	1.0	Peer-to-peer (P2P) trading in a local energy market (LEM) offers various participants the opportunity to negotiate and strike energy deals among themselves using a distributed ledger technology called blockchain. In this paper, a new local model is presented using a layer-2 scalability solution for second-generation (Gen2) blockchain technology to enable P2P trading among four types of participants: consumers, prosumers with solar photovoltaic (PV) systems, prosumers with solar PV systems and battery energy storage systems (BESSs), and electric vehicles (EVs). The proposed LEM trading platform involves several critical steps, including the creation of typical forecasting profiles for load consumption, solar generation, and battery state-of-charge (SOC) through a forecasting solution. Next, the LEM participants place their pricing bids using a trading agent service, and the trading engine collects the profiles data and bid prices, which performs matchmaking in a forward-facing market. The output of the trading engine consists of dispatch signals for prices and energy values that are sent to each participant to execute actual trading. Furthermore, the trading engines store the accepted and past bidding data and energy values of P2P trades for each participant in blockchain technology, which can be retrieved and displayed on the LEM user interface screens of participants and administrators using their blockchain addresses at any time during the trading process. This study focuses on simulating proposed LEM models, incorporating functional limitations and market rules. These rules aim to reduce energy costs, enhance margins for utilities and retailers, and mitigate grid congestion through BESSs, resulting in reduced operational and capital expenditure. LEM outcomes are analysed and compared with a Business-as-usual (BAU) model. Participants' energy trading behaviour, cost-revenue dynamics, grid impact, and blockchain implementation costs are explored. The study highlights LEM benefits in terms of reduced CO2 emissions by 984 kg CO2, increased self-sufficiency by 2.2%, and improved financial benefits of all participants by 21.6%. The use of modern blockchain technology guarantees secure data storage and rapid, cost-effective energy trading, thereby making the proposed LEM platform a viable solution in the distribution market.			http://www.ieee.org	CO2 Emission
21	☐	☑	04002818dbb33	https://www.semanticscholar.	Index Based Techno-Economic Assessment of FACTS Devices Installed With Wind Farms	IEEE Access	2024.0	49.0	6.0	1.0	With the rising demand for electricity and limited fossil fuel reserves, wind energy conversion systems are nowadays considered a pivotal and indispensable resource within the realm of alternative power generation. Flexible AC Transmission Systems (FACTS) devices can be utilized to improve the voltage profile and stability in the presence of large wind farms. This study presents a comprehensive techno-economic assessment for different FACTS devices used with wind farms. In the technical assessment, the improved voltage (IV) index is presented to quantify the contribution of the Static Synchronous Compensator (STATCOM) and Unified Power Flow Controller (UPFC) devices with the focus on the improvement of the voltage levels during different types of faults. In the economic assessment, a cost comparison between the chosen FACTS devices is presented, where a cost related (CR) index is proposed which considers the investment cost in the FACTS device and the minimum revenue that could be obtained as a result of FACTS installation to keep the wind farm connected as a result of the voltage improvement during faults. In addition, guidelines to calculate the threshold for annual additional energy from the wind farm and its revenue needed to compensate for the FACTS device total cost are presented. To consider the environmental impact, the role of the FACTS devices on reducing CO2 emissions, and the associated carbon credit certificates, is considered. Finally, to combine both assessments, a new weighted techno-economic (TEI) index is presented which combines the IV and the CR indices obtained from the technical and economic assessments, respectively. This index will assist system operators in selecting the optimal device, where a breakeven point serving as the criterion for selecting the best device based on the proposed approach. The results show that the TEI index for the proposed system indicates that the STATCOM is considered a preferable solution when the focus is towards the economic weight. On the other hand, the UPFC becomes a preferable solution when the technical aspect is more important to the system operator. In this study, a technical weight of 0.8364 (economic weight of 0.1636) is the turning point between the two devices. Below this weight, the STATCOM is more suitable for installation and above this weight, the UPFC is a preferred option.			http://www.ieee.org	CO2 Emission
22	☐	☑	0f0a794e8d3ac4	https://www.semanticscholar.	Multimodal Data Fusion for Few-shot Named Entity Recognition Method	International Journal	2024.0	50.0	2.0	0.0	As a crucial subtask in Natural Language Processing (NLP), Named Entity Recognition (NER) aims to extract import information from text, which can help many downstream tasks such as machine translation, text generation, knowledge graph construction, and multimodal data fusion to deeply understand the complex semantic information of the text and effectively complete these tasks. In practice, due to time and labor costs, NER suffers from annotated data scarcity, known as few-shot NER. Although few-shot NER methods based on text have achieved good generalization performance, the semantic information that the model can extract is still limited due to the few samples, which leads to the poor prediction effect of the model. To this end, in this paper we propose a few-shot NER model based on multimodal data fusion, which provides additional semantic information with multimodal data for the first time, to help the model prediction and can further effectively improve the effect of multimodal data fusion and modeling. This method converts image information into text information as auxiliary modality information, which effectively solves the problem of poor modality alignment caused by the inconsistent granularity of semantic information contained in text and images. In order to effectively consider the label dependencies in few-shot NER, we use the CRF framework and introduce the state-of-the-art meta-learning methods as the emission module and the transition module. To alleviate the negative impact of noise samples in the auxiliary modal samples, we propose a general denoising network based on the idea of meta-learning. The denoising network can measure the variability of the samples and evaluate the beneficial extent of each sample to the model. Finally, we conduct extensive experiments on real unimodal and multimodal datasets. The experimental results show the outstanding generalization performance of the proposed method, where our method outperforms the state-of-the-art methods by 10% F 1 scores in the 1-shot scenario.			http://www.jisi.org/	NLP | Machine | CO2 Emission
23	☐	☑	ae3d214ed450bf	https://www.semanticscholar.	AN OVERVIEW OF COMPONENTS FOR ENERGY EFFICIENT MULTIMEDIA NETWORKS BASED ON 5G RADIO ACCESS TECHNOLOGIES	Measuring and com	2024.0	2.0	0.0	0.0	Modern society is actively transitioning into an information-based one, and multimedia technologies have become an integral part of this process. Thanks to the proliferation of wireless access networks, users are becoming more mobile, and the development of the fifth generation of mobile networks, known as 5G, is a significant step in the advancement of information and communication technologies. 5G networks offer low latency and reliable connectivity, expanding the capabilities of mobile internet and machine communication. However, along with the opportunities provided by multimedia communication, there is a responsibility to consider the impact of associated technologies on the environment, as well as to address new challenges and the need for prudent resource utilization. The article defines the concept of "multimedia" and discusses various aspects of this concept, including digital storage and processing of information, components (text, photos, audio, and video), interactivity, and hypertextuality. It is noted that the transmission of multimedia data and the use of information technologies are closely linked to wireless access networks. The authors discuss challenges and solutions in the field of energy efficiency for networks. They provide statistics indicating that the information and communication technology (ICT) industry is responsible for a significant portion of global energy consumption and CO2 emissions, with radio access networks being a major contributor. Various components and technologies that can contribute to the development of energy-efficient multimedia networks based on 5G radio access technologies are examined. Specifically, heterogeneous networks, non-orthogonal multiple access (NOMA) technologies, and multiple-input multiple-output (MIMO) technologies are highlighted as key components for achieving energy efficiency. The importance of using heterogeneous networks to reduce the distance between transmitters and receivers is emphasized, along with the possibility of putting small base stations into sleep mode when there is no network load. Technologies like NOMA and MIMO are discussed as crucial components for achieving spectral efficiency and energy efficiency. Additionally, the article focuses on wireless sensor networks (WSNs) and suggests ways to optimize them for energy efficiency. This includes operating sensor units only when necessary, implementing wireless charging, using energy-efficient optimization methods, and applying efficient routing schemes. The authors also highlight the role of green data centers in reducing CO2 emissions and optimizing the use of green energy in high-performance networks. Methods such as using renewable energy sources, increasing the energy efficiency of hardware, and implementing energy-efficient routing are discussed. In conclusion, the article underscores the importance of energy efficiency and reduced CO2 emissions in modern multimedia networks, particularly in the context of 5G networks. It calls for interdisciplinary efforts to address these critical challenges in the field of information and communication technologies.				Machine | CO2 Emission
24	☐	☑	0bf0e85a7d3d6f	https://www.semanticscholar.	Short-term forecasting of German generation-based CO2 emission factors using parametric and non-parametric time series models	Energy Informatics	2024.0	72.0	6.0	0.0	This study focuses on forecasting German generation-based CO2 emission factors to develop accurate prediction models, which help to shift flexible loads in time with low emissions. While most existing research relies on point forecasts to predict CO2 emission factors, the presented methods are utilized to perform interval forecasts. In addition, compared to other studies, recent data that extends over a long period is used. The study describes the used data and discusses the concept of walk-forward validation. Further, various models are employed and tuned to forecast the emission factors, including benchmark, parametric (e.g., SARIMAX), and non-parametric (bagging, random forest, gradient boosting, CNN, LSTM, MLP) models. The study reveals that all applied parametric and non-parametric models yield better results than the benchmark models, while the gradient boosting model has the lowest mean absolute error with 40.66 gCO2/kWh, the lowest mean absolute percentage error 8.17%, and the random forest has the lowest root mean square error with 57.61 gCO2/kWh. However, the potential of the deep learning models was not fully exploited. In a live application, the implementation effort should be evaluated against the benefit of better prediction.			http://link.springer.c	CO2 Emission
25	☑	☐	a97ee862a3fa5a	https://www.semanticscholar.	Detection and Classification of ChatGPT Generated Contents Using Deep Transformer Models	International Resear	2024.0	8.0	11.0	4.0	AI advancements, particularly in neural networks, have brought about groundbreaking tools like text generators and chatbots. While these technologies offer tremendous benefits, they also pose serious risks such as privacy breaches, spread of misinformation, and challenges to academic integrity. Previous efforts to distinguish between human and AI-generated text have been limited, especially with models like ChatGPT. To tackle this, we created a dataset containing both human and ChatGPT-generated text, using it to train and test various machine and deep learning models. Your results, particularly the high F1-score and accuracy achieved by the RoBERTa-based custom deep learning model and Distill BERT, indicate promising progress in this area. By establishing a robust baseline for detecting and classifying AI-generated content, your work contributes significantly to mitigating potential misuse of AI-powered text generation tools.				Privacy | Machine | Detection | Paraphrase | NLP | Bias | Controllable | Creative | Automated | Task | Training | Cost | Machine Generate

26	☑	☐	ef7567b87b2706	https://www.semanticscholar.	Exploring the boundaries of authorship: a comparative analysis of AI-generated text and human academic writing in English literature	Frontiers in Educatic	2024.0	34.0	50.0	4.0	As artificial intelligence (AI) increasingly permeates educational landscapes, its impact on academic writing has become a subject of intense scrutiny. This research delved into the nuanced dimensions of authorship and voice in academic writing, specifically focusing on the application of OpenAI's ChatGPT. In this study, the research team compared and contrasted an essay written by one second-year English student for a course on English literature with a similar essay produced by ChatGPT. The current research also, tried to clarify whether artificial intelligence can satisfy the formal requirements of academic writing and maintain the distinctive voice inherent in human-authored content. The examination hinges on parameters such as assertiveness, self-identification, and authorial presence. Additionally, the researchers shed light on the challenges inherent in producing AI-generated academic text. While ChatGPT presented an ability to generate contextually relevant content, the results highlighted its need for support in guaranteeing factual accuracy and capturing the complex aspects of authorship that are common in human writing. Notably, when compared to human-generated text, the AI-generated text was deficient in terms of specificity, depth, and accurate source referencing. While AI has potential as an additional tool for academic writing, this study's findings indicated that its current capabilities—particularly in producing academic text are limited, and remain constrained. This study emphasizes upon the imperative for continued refinement and augmentation of AI models to bridge the existing gaps in achieving a more seamless integration into the academic writing landscape.			https://epub.uni-reg	Machine Generated Text (None)
27	☐	☑	2f64a7cf3a02291	https://www.semanticscholar.	Ten Words Only Still Help: Improving Black-Box AI-Generated Text Detection via Proxy-Guided Efficient Re-Sampling	International Joint C	2024.0	51.0	22.0	3.0	With the rapidly increasing application of large language models (LLMs), their abuse has caused many undesirable societal problems such as fake news, academic dishonesty, and information pollution. This makes AI-generated text (AIGT) detection of great importance. Among existing methods, white-box methods are generally superior to black-box methods in terms of performance and generalizability, but they require access to LLMs' internal states and are not applicable to black-box settings. In this paper, we propose to estimate word generation probabilities as pseudo white-box features via multiple re-sampling to help improve AIGT detection under the black-box setting. Specifically, we design POGER, a proxy-guided efficient re-sampling method, which selects a small subset of representative words (e. g., 10 words) for performing multiple re-sampling in black-box AIGT detection. Experiments on datasets containing texts from humans and seven LLMs show that POGER outperforms all baselines in macro F1 under black-box, partial white-box, and out-of-distribution settings and maintains lower re-sampling costs than its existing counterparts.			http://www.ijcai.org/	Detection | Natural Language Generation | Language Model | Machine Generated Text (None)
28	☑	☐	ef4f93ddff902e7	https://www.semanticscholar.	Enhancing Text Authenticity: A Novel Hybrid Approach for AI-Generated Text Detection	2024 IEEE 4th Inter	2024.0	34.0	19.0	3.0	The rapid advancement of Large Language Models (LLMs) has ushered in an era where AI-generated text is increasingly indistinguishable from human-generated content. Detecting AI-generated text has become imperative to combat misinformation, ensure content authenticity, and safeguard against malicious uses of AI. We introduce an innovative mixed methodology that integrates conventional TF-IDF strategies with sophisticated machine learning algorithms, including Bayesian classifiers, Stochastic Gradient Descent (SGD), Categorical Gradient Boosting (CatBoost), and 12 instances of Deberta-v3-large models. Our method tackles the difficulties of identifying AI-produced text by combining the advantages of conventional feature extraction techniques with the latest advancements in deep learning models. Through extensive experiments on a comprehensive dataset, we demonstrate the effectiveness of our proposed method in accurately distinguishing between human and AI-generated text. Our approach achieves superior performance compared to existing methods. This research contributes to the advancement of AI-generated text detection techniques and lays the foundation for developing robust solutions to mitigate the challenges posed by AI-generated content.				Paraphrase | NLP | Natural Language Generation | Language Model | Bias | Privacy | Controllable | Creative | Machine | Automated | Ta
29	☐	☑	f20dc28114c4e3	https://www.semanticscholar.	Towards Multimodal Empathetic Response Generation: A Rich Text-Speech-Vision Avatar-based Benchmark	The Web Conferenc	2025.0	64.0	17.0	3.0	Empathetic Response Generation (ERG) is one of the key tasks of the affective computing area, which aims to produce emotionally nuanced and compassionate responses to user's queries. However, existing ERG research is predominantly confined to the singleton text modality, limiting its effectiveness since human emotions are inherently conveyed through multiple modalities. To combat this, we introduce an avatar-based Multimodal ERG (MERG) task, entailing rich text, speech, and facial vision information. We first present a large-scale high-quality benchmark dataset, AvaMERG, which extends traditional text ERG by incorporating authentic human speech audio and dynamic talking-face avatar videos, encompassing a diverse range of avatar profiles and broadly covering various topics of real-world scenarios. Further, we deliberately tailor a system, named Empatheia, for MERG. Built upon a Multimodal Large Language Model (MLLM) with multimodal encoder, speech and avatar generators, Empatheia performs end-to-end MERG, with Chain-of-Empathetic reasoning mechanism integrated for enhanced empathy understanding and reasoning.Finally, we devise a list of empathetic-enhanced tuning strategies, strengthening the capabilities of emotional accuracy and content, avatar-profile consistency across modalities. Experimental results on AvaMERG data demonstrate that Empatheia consistently shows superior performance than baseline methods on both textual ERG and MERG. All data and code are open at https://AvaMERG.github.io/.			http://www.iw3c2.org/	Paraphrase | NLP | Natural Language Generation | Language Model | Bias | Privacy | Controllable | Creative | Machine | Automated | Ta
30	☐	☑	efc70732194f21e	https://www.semanticscholar.	Text2Earth: Unlocking text-driven remote sensing image generation with a global-scale dataset and a foundation model	IEEE Geoscience ar	2025.0	86.0	36.0	2.0	Recently, generative foundation models (GFMs) have significantly advanced large-scale text-driven natural image generation and become a prominent research trend across various vertical domains. However, in the remote sensing field, there is still a lack of research on large-scale text-to-image (text2image) generation technology. Existing remote sensing image–text datasets are small in scale and confined to specific geographic areas and scene types. Besides, existing text2image methods have struggled to achieve global-scale, multiresolution controllability, and unbounded image generation. To address these challenges, this article presents two key contributions: the Git-10M dataset and the Text2Earth foundation model. Git-10M is a global-scale image–text dataset consisting of 10.5 million image–text pairs, five times larger than the previous largest one. The dataset covers a wide range of geographic scenes and contains essential geospatial metadata, significantly surpassing existing datasets in both size and diversity. Building on Git-10M, we propose Text2Earth, a 1.3 billion-parameter GFM based on the diffusion framework to model global-scale remote sensing scenes. Text2Earth integrates a resolution guidance mechanism, enabling users to specify image resolutions. A dynamic condition adaptation (DCA) strategy is proposed for training and inference to improve image generation quality. Text2Earth not only excels in zero-shot text2image generation but also demonstrates robust generalization and flexibility across multiple tasks, including unbounded scene construction, image editing, and cross-modal image generation. This robust capability surpasses previous models restricted to basic fixed sizes and limited scene types. On the previous text2image benchmark dataset, Text2Earth outperforms previous models, with a significantly improved +26.23 Fréchet inception distance (FID) score and +20.95% zero-shot classification overall accuracy (OS-OA) metric. Our project page is https://chen-yang-liu.github.io/Text2Earth/.			http://ieeexplore.iee	Paraphrase | NLP | Natural Language Generation | Language Model | Evaluation Metrics | Bias | Privacy | Controllable | Creative | Mach
31	☐	☑	d1da580e72e907	https://www.semanticscholar.	UAV-VLA: Vision-Language-Action System for Large Scale Aerial Mission Generation	IEEE/ACM Internati	2025.0	21.0	25.0	2.0	The UAV-VLA (Visual-Language-Action) system is a tool designed to facilitate communication with aerial robots. By integrating satellite imagery processing with the Visual Language Model (VLM) and the powerful capabilities of GPT, UAV-VLA enables users to generate general flight paths-and-action plans through simple text requests. This system leverages the rich contextual information provided by satellite images, allowing for enhanced decision-making and mission planning. The combination of visual analysis by VLM and natural language processing by GPT can provide the user with the path-and-action set, making aerial operations more efficient and accessible. The newly developed method showed the difference in the length of the created trajectory in 22% and the mean error in finding the objects of interest on a map in 34.22 m by Euclidean distance in the K-Nearest Neighbors (KNN) approach. Additionally, the UAV-VLA system generates all flight plans in just 5 minutes and 24 seconds, making it 6.5 times faster than an experienced human operator. The code is available here: https://github.com/sautenich/uav-vla			http://www.wikicfp.c	Paraphrase | NLP | Natural Language Generation | Language Model | Bias | Privacy | Controllable | Creative | Machine | Automated | Ta
32	☑	☐	1db86e01300e2f	https://www.semanticscholar.	TGEA: An Error-Annotated Dataset and Benchmark Tasks for TextGeneration from Pretrained Language Models	Annual Meeting of t	2025.0	65.0	8.0	2.0	In order to deeply understand the capability of pretrained language models in text generation and conduct a diagnostic evaluation, we propose TGEA, an error-annotated dataset with multiple benchmark tasks for text generation from pretrained language models (PLMs). We use carefully selected prompt words to guide GPT-2 to generate candidate sentences, from which we select 47K for error annotation. Crowdsourced workers manually check each of these sentences and detect 12k erroneous sentences. We create an error taxonomy to cover 24 types of errors occurring in these erroneous sentences according to the nature of errors with respect to linguistics and knowledge (e.g., common sense). For each erroneous span in PLM-generated sentences, we also detect another span that is closely associated with it. Each error is hence manually labeled with comprehensive annotations, including the span of the error, the associated span, minimal correction to the error, the type of the error, and rationale behind the error. Apart from the fully annotated dataset, we also present a detailed description of the data collection procedure, statistics and analysis of the dataset. This is the first dataset with comprehensive annotations for PLM-generated texts, which facilitates the diagnostic evaluation of PLM-based text generation. Furthermore, we use TGEA as a benchmark dataset and propose a series of automatic diagnosis tasks, including error detection, error type classification, associated span detection, error rationale generation, to further promote future study on the automatic error detection and correction on texts generated by pretrained language models.			https://www.aclweb.c	Paraphrase | NLP | Natural Language Generation | Language Model | Evaluation Metrics | Bias | Privacy | Controllable | Creative | Mach
33	☐	☑	2a9e3dda0b504	https://www.semanticscholar.	Fg-T2M++: LLMs-Augmented Fine-Grained Text Driven Human Motion Generation	International Journal	2025.0	51.0	33.0	1.0	We address the challenging problem of fine-grained text-driven human motion generation. Existing works generate imprecise motions that fail to accurately capture relationships specified in text due to: (1) lack of effective text parsing for detailed semantic cues regarding body parts, (2) not fully modeling linguistic structures between words to comprehend text comprehensively. To tackle these limitations, we propose a novel fine-grained framework Fg-T2M++ that consists of: (1) an LLMs semantic parsing module to extract body part descriptions and semantics from text, (2) a hyperbolic text representation module to encode relational information between text units by embedding the syntactic dependency graph into hyperbolic space, and (3) a multi-modal fusion module to hierarchically fuse text and motion features. Extensive experiments on HumanML3D and KIT-ML datasets demonstrate that Fg-T2M++ outperforms SOTA methods, validating its ability to accurately generate motions adhering to comprehensive text semantics.			https://www.springe	Paraphrase | NLP | Language Model | Bias | Privacy | Controllable | Creative | Machine | Automated | Task | Training | Cost | Detection |
34	☐	☑	d06d44484fc79e	https://www.semanticscholar.	LLMER: Crafting Interactive Extended Reality Worlds with JSON Data Generated by Large Language Models	IEEE Transactions o	2025.0	35.0	12.0	1.0	The integration of Large Language Models (LLMs) like GPT-4 with Extended Reality (XR) technologies offers the potential to build truly immersive XR environments that interact with human users through natural language, e.g., generating and animating 3D scenes from audio inputs. However, the complexity of XR environments makes it difficult to accurately extract relevant contextual data and scene/object parameters from an overwhelming volume of XR artifacts. It leads to not only increased costs with pay-per-use models, but also elevated levels of generation errors. Moreover, existing approaches focusing on coding script generation are often prone to generation errors, resulting in flawed or invalid scripts, application crashes, and ultimately a degraded user experience. To overcome these challenges, we introduce LLMER, a novel framework that creates interactive XR worlds using JSON data generated by LLMs. Unlike prior approaches focusing on coding script generation, LLMER translates natural language inputs into JSON data, significantly reducing the likelihood of application crashes and processing latency. It employs a multi-stage strategy to supply only the essential contextual information adapted to the user's request and features multiple modules designed for various XR tasks. Our preliminary user study reveals the effectiveness of the proposed system, with over 90% reduction in consumed tokens and around 60% reduction in task completion time compared to state-of-the-art approaches. The analysis of users' feedback also illuminates a series of directions for further optimization.	40063485.0		http://ieeexplore.iee	Natural Language Generation | Language Model | Task
35	☑	☐	007e7d6cb64a7c	https://www.semanticscholar.	EdgeShard: Efficient LLM Inference via Collaborative Edge Computing	IEEE Internet of Thi	2025.0	46.0	135.0	16.0	Large language models (LLMs) have shown great success in content generation and intelligent intelligent decision making for IoT systems. Traditionally, LLMs are deployed on the cloud, incurring prolonged latency, high bandwidth costs, and privacy concerns. More recently, edge computing has been considered promising in addressing such concerns because the edge devices are closer to data sources. However, edge devices are cursed by their limited resources and can hardly afford LLMs. Existing studies address such a limitation by offloading heavy workloads from edge to cloud or compressing LLMs via model quantization. These methods either still rely heavily on the remote cloud or suffer substantial accuracy loss. This work is the first to deploy LLMs on a collaborative edge computing environment, in which edge devices and cloud servers share resources and collaborate to infer LLMs with high efficiency and no accuracy loss. We design EdgeShard, a novel approach to partition a computation-intensive LLM into affordable shards and deploy them on distributed devices. The partition and distribution are nontrivial, considering device heterogeneity, bandwidth limitations, and model complexity. To this end, we formulate an adaptive joint device selection and model partition problem and design an efficient dynamic programming algorithm to optimize the inference latency and throughput. Extensive experiments of the popular Llama2 serial models on a real-world testbed reveal that EdgeShard achieves up to 50% latency reduction and $2 \times$ throughput improvement over the state-of-the-art.			https://www.ieee.or	Language Model | Privacy | Text Generation (None)

36	☐	☑	aa2728836e242	https://www.semanticscholar.	Toward De Novo Protein Design from Natural Language	bioRxiv	2025.0	84.0	19.0	3.0				http://biorxiv.org/	Language Model | Task	
37	☐	☑	5c2f9dc8596ff55	https://www.semanticscholar.	A benchmark for large language models in bioinformatics	bioRxiv	2025.0	0.0	10.0	2.0	The rapid advancements in artificial intelligence, particularly in Large Language Models (LLMs) such as GPT-4, Gemini, and LLaMA, have opened new avenues for computational biology and bioinformatics. We report the development of BioLLMBench, a novel framework designed to evaluate LLMs in bioinformatics tasks. This study assessed GPT-4, Gemini, and LLaMA through 2,160 experimental runs, focusing on 24 distinct tasks across six key areas: domain expertise, mathematical problem-solving, coding proficiency, data visualization, research paper summarization, and machine learning model development. Tasks ranged from fundamental to expert-level challenges, and each area was evaluated using seven specific metrics. A Contextual Response Variability Analysis was implemented to understand how model responses varied under different conditions. Results showed diverse performance: GPT-4 led in most tasks, achieving a 91.3% proficiency in domain knowledge, while Gemini excelled in mathematical problem-solving with a 97.5% proficiency score. GPT-4 also outperformed in machine learning model development, though Gemini and LLaMA struggled to generate executable code. All models faced challenges in research paper summarization, scoring below 40% using the ROUGE metric. Model performance variance increased when using a new chat window, though average scores remained similar. The study also discusses the limitations and potential misuse risks of these models in bioinformatics.				http://biorxiv.org/	Language Model | Evaluation Metrics | Machine | Task | Machine Generated Text (None)	
38	☑	☐	4e3a1d0d5fc64b	https://www.semanticscholar.	Toward Immersive Computational Storytelling: Card-Framework for Enhanced Persona-Driven Dialogues	IEEE Transactions c	2025.0	47.0	4.0	1.0	In the realm of role-playing games (RPGs), creating immersive, persona-driven dialogues remains a challenge, especially in intricate settings, such as Call of Cthulhu. Existing methodologies often falter in portraying character personas within complex conversations accurately. To address this, we introduce a novel card-based framework, utilizing the advanced 7B language model for tailored dialogue generation. Guided by detailed scene settings and character personas, 7B language model exhibited a striking ability to craft context-aware dialogues for even unseen characters and scenarios. To assess the quality of these dialogues, we present an innovative metric, circumventing the traditional hurdles of human evaluations. Furthermore, insights into the attention mechanism shed light on the dynamics of information flow during dialogue creation. Collectively, our findings underscore the transformative potential of large language models in computational storytelling, particularly in RPG settings.				https://ieeexplore.ie	Language Model | Evaluation Metrics	
39	☐	☑	bbef7c969845a6	https://www.semanticscholar.	Quality Assurance for LLM-RAG Systems: Empirical Insights from Tourism Application Testing	International Confer	2025.0	11.0	2.0	1.0	This paper presents a comprehensive framework for testing and evaluating quality characteristics of Large Language Model (LLM) systems enhanced with Retrieval-Augmented Generation (RAG) in tourism applications. Through systematic empirical evaluation of three different LLM variants across multiple parameter configurations, we demonstrate the effectiveness of our testing methodology in assessing both functional correctness and extra-functional properties. Our framework implements 17 distinct metrics that encompass syntactic analysis, semantic evaluation, and behavioral evaluation through LLM judges. The study reveals significant information about how different architectural choices and parameter configurations affect system performance, particularly highlighting the impact of temperature and top-p parameters on response quality. The tests were carried out on a tourism recommendation system for the Varmland region in Sweden, utilizing standard and RAG-enhanced configurations. The results indicate that the newer LLM versions show modest improvements in performance metrics, though the differences are more pronounced in response length and complexity rather than in semantic quality. The research contributes practical insights for implementing robust testing practices in LLM-RAG systems, providing valuable guidance to organizations deploying these architectures in production environments.					Language Model | Evaluation Metrics	
40	☐	☑	522dfb2e42cbb0	https://www.semanticscholar.	ProxDDP: Proximal Constrained Trajectory Optimization	IEEE Transactions c	2025.0	98.0	27.0	3.0	Trajectory optimization has been a popular choice for motion generation and control in robotics for at least a decade. Several numerical approaches have exhibited the required speed to enable online computation of trajectories for real-time of various systems, including complex robots. Many of these said are based on the differential dynamic programming (DDP) algorithm—initially designed for unconstrained trajectory optimization problems—and its variants, which are relatively easy to implement and provide good runtime performance. However, several problems in robot control call for using constrained formulations (e.g., torque limits, obstacle avoidance), from which several difficulties arise when trying to adapt DDP-type methods: numerical stability, computational efficiency, and constraint satisfaction. In this article, we leverage proximal methods for constrained optimization and introduce a DDP-type method for fast, constrained trajectory optimization suited for model-predictive control (MPC) applications with easy warm-starting. Compared to earlier solvers, our approach effectively manages hard constraints without warm-start limitations and exhibits good convergence behavior. We provide a complete implementation as part of an open-source and flexible C++ trajectory optimization library called aligator. These algorithmic contributions are validated through several trajectory planning scenarios from the robotics literature and the real-time whole-body MPC of a quadruped robot.				http://ieeexplore.iee	Controllable	
41	☑	☐	12811af37a00d1	https://www.semanticscholar.	Division-of-Thoughts: Harnessing Hybrid Language Model Synergy for Efficient On-Device Agents	The Web Conferenc	2025.0	40.0	22.0	2.0	The rapid expansion of web content has made on-device AI assistants indispensable for helping users manage the increasing complexity of online tasks. The emergent reasoning ability in large language models offer a promising path for next-generation on-device AI agents. However, deploying full-scale Large Language Models (LLMs) on resource-limited local devices is challenging. In this paper, we propose Division-of-Thoughts (DoT), a collaborative reasoning framework leveraging the synergy between locally deployed Smaller-scale Language Models (SLMs) and cloud-based LLMs. DoT leverages a Task Decomposer to elicit the inherent planning abilities in language models to decompose user queries into smaller sub-tasks, which allows hybrid language models to fully exploit their respective strengths. Besides, DoT employs a Task Scheduler to analyze the pair-wise dependency of sub-tasks and create a dependency graph, facilitating parallel reasoning of sub-tasks and the identification of key steps. To allocate the appropriate model based on the difficulty of sub-tasks, DoT leverages a Plug-and-Play Adapter, which is an additional task head attached to the SLM that does not alter the SLM's parameters. To boost adapter's task allocation capability, we propose a self-reinforced training method that relies solely on task execution feedback. Extensive experiments on various benchmarks demonstrate that our DoT significantly reduces LLM costs while maintaining competitive reasoning accuracy. Specifically, DoT reduces the average reasoning time and API costs by 66.12% and 83.57%, while achieving comparable reasoning accuracy with the best baseline methods.				http://www.iw3c2.org/	Language Model | Task | Training	
42	☑	☐	93a7c1cb29c05d	https://www.semanticscholar.	MVAN: Multi-View Attention Networks for Fake News Detection on Social Media	IEEE Access	2025.0	48.0	78.0	6.0	Fake news on social media is a widespread and serious problem in today's society. Existing fake news detection methods focus on finding clues from Long text content, such as original news articles and user comments. This paper solves the problem of fake news detection in more realistic scenarios. Only source shot-text tweet and its retweet users are provided without user comments. We develop a novel neural network based model, Multi-View Attention Networks (MVAN) to detect fake news and provide explanations on social media. The MVAN model includes text semantic attention and propagation structure attention, which ensures that our model can capture information and clues both of source tweet content and propagation structure. In addition, the two attention mechanisms in the model can find key clue words in fake news texts and suspicious users in the propagation structure. We conduct experiments on two real-world datasets, and the results demonstrate that MVAN can significantly outperform state-of-the-art methods by 2.5% in accuracy on average, and produce a reasonable explanation.				http://www.ieee.org	Detection | Text Generation (None)	
43	☐	☑	2048c84458f5c2	https://www.semanticscholar.	Nucleotide dependency analysis of DNA language models reveals genomic functional elements	bioRxiv	2025.0	56.0	14.0	2.0	Deciphering how nucleotides in genomes encode regulatory instructions and molecular machines is a long-standing goal in biology. DNA language models (LMs) implicitly capture functional elements and their organization from genomic sequences alone by modeling probabilities of each nucleotide given its sequence context. However, using DNA LMs for discovering functional genomic elements has been challenging due to the lack of interpretable methods. Here, we introduce nucleotide dependencies which quantify how nucleotide substitutions at one genomic position affect the probabilities of nucleotides at other positions. We generated genome-wide maps of pairwise nucleotide dependencies within kilobase ranges for animal, fungal, and bacterial species. We show that nucleotide dependencies indicate deleteriousness of human genetic variants more effectively than sequence alignment and DNA LM reconstruction. Regulatory elements appear as dense blocks in dependency maps, enabling the systematic identification of transcription factor binding sites as accurately as models trained on experimental binding data. Nucleotide dependencies also highlight bases in contact within RNA structures, including pseudoknots and tertiary structure contacts, with remarkable accuracy. This led to the discovery of four novel, experimentally validated RNA structures in Escherichia coli. Finally, using dependency maps, we reveal critical limitations of several DNA LM architectures and training sequence selection strategies by benchmarking and visual diagnosis. Altogether, nucleotide dependency analysis opens a new avenue for discovering and studying functional elements and their interactions in genomes.				http://biorxiv.org/	Language Model | Machine | Training | Machine Generated Text (None)