

	To Consider	Don't Consider	paperId	corpusUrl	title	abstract	year	referenceCount	citationCount	influentialCita	s2FieldsOfStudy	publicationType	openAccessPdfUrl	subquery
89	<input checked="" type="checkbox"/>	<input type="checkbox"/>	02033482048045	2303131	<a href="https://www.arxiv.org/abs/2303.13131">https://www.arxiv.org/abs/2303.13131</a>	DExperts: Decoding-Time Controlled Text Generation with Experts and Anti-Experts	Despite recent advances in natural language generation, it remains challenging to control attributes of generated text. We propose DExperts: Decoding-time Experts, a decoding-time method for controlled text generation that combines a pretrained language model with "expert" LMs and/or "test-experts" LMs as a product of experts. Intuitively, under the ensemble, tokens only get high probability if they are considered likely by the experts, and unlikely by the anti-experts. We apply DExperts to language detoxification and sentiment-controlled generation, where we outperform existing controlled-generation methods on both automatic and human evaluations. Moreover, because DExperts operates only on the output of the pretrained LM, it is effective with (anti-)experts of smaller size, including when operating on GPT-3. Our work highlights the promise of tuning small LMs on text with (un)desirable attributes for efficient decoding-time steering.	2021	48	90	26	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2303.13131v1.pdf">https://arxiv.org/pdf/2303.13131v1.pdf</a>	Paraphrase   Natural Language Generation   Language Model   Bias   Priv
90	<input checked="" type="checkbox"/>	<input type="checkbox"/>	b6c496e06bf1	2332107	<a href="https://www.arxiv.org/abs/2303.12107">https://www.arxiv.org/abs/2303.12107</a>	FUDGE: Controlled Text Generation With Future Discriminators	We propose Future Discriminators for Generation (FUDGE), a flexible and modular method for controlled text generation. Given a pre-existing model G for generating text from a distribution of interest, FUDGE enables conditioning on a desired attribute a (for example, formality) while requiring access only to G's output tokens. FUDGE learns an attribute predictor operating on a partial prefix of tokens and uses the predictor's outputs to adjust G's original probabilities. We show that FUDGE models terms corresponding to a Bayesian decomposition of the conditional distribution of G given attribute a. Moreover, FUDGE can easily compose predictors for multiple desired attributes. We evaluate FUDGE on three tasks — co-speech counting in poetry, topic control in language generation, and formality change in machine translation — and observe gains in all three tasks.	2021	48	91	22	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2303.12107v1.pdf">https://arxiv.org/pdf/2303.12107v1.pdf</a>	Paraphrase   Language Model   Bias   Privacy   Controllable   Creative   M
91	<input checked="" type="checkbox"/>	<input type="checkbox"/>	a16ae7070de11	235894	<a href="https://www.arxiv.org/abs/2305.08944">https://www.arxiv.org/abs/2305.08944</a>	All That's Human Is Not God: Evaluating Human Evaluation of Generated Text	Human evaluations are typically considered the gold standard in natural language generation, but as models' fluency improves, how well can evaluators detect and judge machine-generated text? We run a study distinguishing between GPT-3 and human-authored text at random chance level. We explore three approaches for quickly training evaluators to better identify GPT-3-authored text (detailed instructions, annotated examples, and paired examples) and find the triple approach to improve up to 55%. It did not significantly improve across the three domains. Given the inconsistent results across test domains and the often contradictory reasons evaluators gave for their judgments, we examine the role untrained human evaluators play in NLG evaluation and provide recommendations to NLG researchers for improving human evaluations of text generated from state-of-the-art models.	2021	40	114	19	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2305.08944v1.pdf">https://arxiv.org/pdf/2305.08944v1.pdf</a>	Machine   Paraphrase   NLP   Natural Language Generation   Language M
92	<input checked="" type="checkbox"/>	<input type="checkbox"/>	85241d5942966f	2382598	<a href="https://www.arxiv.org/abs/2308.25988">https://www.arxiv.org/abs/2308.25988</a>	Aspect Sentiment Quad Prediction as Paraphrase Learning	Aspect-based sentiment analysis (ASBA) has been extensively studied in recent years, which typically involves four fundamental sentiment elements, including the aspect category, aspect term, opinion term, and sentiment polarity. Existing studies usually consider the detection of partial sentiment elements, instead of predicting the four elements in one shot. In this work, we introduce the Aspect Sentiment Quad Prediction (ASQP) task, aiming to jointly detect all sentiment elements in quads for a given opinionated sentence, which can reveal a more comprehensive and control of single sentiment structure. We further propose a novel Paraphrase modeling paradigm to cast the ASQP task to a paraphrase generation process. On one hand, the generation formulation allows solving ASQP in an end-to-end manner, alleviating the potential error propagation in the pipeline solution. On the other hand, the semantics of the sentiment elements can be fully exploited by learning to generate them in the natural language form. Extensive experiments on benchmark datasets show the superiority of our proposed method and the capacity of cross-task transfer with the proposed unified Paraphrase modeling framework.	2021	42	43	16	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2308.25988v1.pdf">https://arxiv.org/pdf/2308.25988v1.pdf</a>	Paraphrase
93	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ab047321118de1	247315	<a href="https://www.arxiv.org/abs/2307.03155">https://www.arxiv.org/abs/2307.03155</a>	Is GPT-3 Text Indistinguishable from Human Text? Scarecrow: A Framework for Scrutinizing Machine Text	Modern neural language models can produce remarkably fluent and grammatical text. So much, in fact, that recent work by Clark et al. (2021) has reported that conventional crowdsourcing can no longer reliably distinguish between machine-authored (GPT-3) and human-authored writing. As errors in machine-generated documents become ever subtler and harder to spot, it poses a new challenge to the research community for robust machine text evaluation. We propose a new framework called Scarecrow for scrutinizing machine text via crowd annotation. To support the broad range of real machine errors that can be identified by laypeople, we let our error categories of Scarecrow—such as redundancy, common-sense error, and incoherence—are identified through several rounds of crowd annotation experiments without a predefined ontology. We then use Scarecrow to collect over 41k error spans in human-written and machine-generated paragraphs of English language news text. We isolate factors for detailed analysis, including parameter count, training data, and various decoding time configurations. Our approach successfully quantifies measurable gaps between human authored text and generations from models of several sizes, including fourteen configurations of GPT-3. In addition, our analysis unveils new insights, with detailed rationales provided by laypeople, e.g., that the common-sense capabilities have been improving with larger models while math capabilities have not, and that the choices of simple decoding hyperparameters can make remarkable differences on the perceived quality of machine text. We release our training material, annotation toolkit and dataset at <a href="https://yo-yo-dub.github.io/scarecrow/">https://yo-yo-dub.github.io/scarecrow/</a> .	2021	39	49	9	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2307.03155v1.pdf">https://arxiv.org/pdf/2307.03155v1.pdf</a>	Paraphrase   NLP   Language Model   Bias   Privacy   Controllable   Creati
94	<input type="checkbox"/>	<input checked="" type="checkbox"/>	96e565a0067c1	2320927	<a href="https://www.arxiv.org/abs/2302.20927">https://www.arxiv.org/abs/2302.20927</a>	WIT: Wikipedia-based Image Text Dataset for Multimodal Multilingual Machine Learning	The milestone improvements brought about by deep representation learning and pre-training techniques have led to large performance gains across downstream NLP, IR and Vision tasks. Multimodal multimodal techniques aim to leverage large high-quality video-linguistic datasets for learning complementary information across image and text modalities. In this paper, we introduce the Wikipedia-based Image Text (WIT) dataset to better facilitate multimodal machine learning. WIT is composed of a curated set of 37.1 million entity-rich image-text examples with 11.5 million unique images across 108 Wikipedia languages. WIT also enables WIT to be used as a pretraining dataset for multimodal models, as it shows when applied to downstream tasks such as image-text retrieval. WIT has four main and unique advantages. First, WIT is the largest multimodal dataset by the number of image-text pairs (2.5 billion). Second, WIT is the first of its kind to be available in 108 languages. Third, WIT is the first of its kind to be available in 108 languages and provides cross-lingual tests for many images. Third, WIT represents a more diverse set of concepts and real-world entities relative to what previous datasets cover. Lastly, WIT provides a very challenging real-world test set, as we empirically illustrate using an image-text retrieval task as an example. WIT Dataset is available for download and use via a Creative Commons license here: <a href="https://github.com/compnlp-research/wikipedia-image-text-dataset/">https://github.com/compnlp-research/wikipedia-image-text-dataset/</a> .	2021	44	117	30	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2302.20927v1.pdf">https://arxiv.org/pdf/2302.20927v1.pdf</a>	NLP   Creative   Training
95	<input type="checkbox"/>	<input checked="" type="checkbox"/>	1adadbf9d643a	2375681	<a href="https://www.arxiv.org/abs/2307.56811">https://www.arxiv.org/abs/2307.56811</a>	Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus	Large language models have led to remarkable progress on many NLP tasks, and researchers are turning to ever-larger text corpora to train them. Some of the largest corpora available are made by scraping and archiving the contents of the internet and web, such as the Common Crawl Corpus (CC) (Raffel et al., 2020), a dataset created by applying a set of filters to a snapshot of OpenWebText. We begin by investigating where the data came from, and find a significant amount of text from unexpected sources like patents and US military websites. Then we explore the content of the text itself, and find machine-generated text (e.g., from machine translation systems) and evaluation examples from other benchmark NLP datasets. To understand the impact of the filters applied to create the dataset, we evaluate the text that was removed, and show that blocked filtering disproportionately removes text from and about minority individuals. Finally, we conclude with some recommendations for how to create and document web-text datasets from a scrape of the internet.	2021	79	107	7	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2307.56811v1.pdf">https://arxiv.org/pdf/2307.56811v1.pdf</a>	NLP
96	<input checked="" type="checkbox"/>	<input type="checkbox"/>	d7a7ebd1565c31	2361258	<a href="https://www.arxiv.org/abs/2306.12588">https://www.arxiv.org/abs/2306.12588</a>	Pretrained Language Models for Text Generation: A Survey	Text generation has become one of the most important yet challenging tasks in natural language processing (NLP). The resurgence of deep learning has greatly advanced this field by neural generation models, especially the paradigm of pretrained language models (PLMs). In this paper, we present an overview of the major advances achieved in the topic of PLMs for text generation. As the preliminaries, we present the general task definition and briefly describe the mainstream architectures of PLMs for text generation. As the core content, we discuss how to adapt existing PLMs to model different input data and satisfy special properties in the generated text. We further summarize several important fine-tuning strategies for text generation.	2021	244	68	6	[[Category: "Computer [JournalArticle",	<a href="https://www.icml.org/proceedings/2023/papers/p244d7a7ebd1565c312361258.pdf">https://www.icml.org/proceedings/2023/papers/p244d7a7ebd1565c312361258.pdf</a>	Paraphrase   NLP   Natural Language Generation   Language Model   Bias
97	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ce3b304bf6e38d	2317119	<a href="https://www.arxiv.org/abs/2301.17119">https://www.arxiv.org/abs/2301.17119</a>	BOLD: Dataset and Metrics for Measuring Biases in Open-Ended Language Generation	Recent advances in deep learning techniques have enabled machines to generate cohesive open-ended text when prompted with a sequence of words as context. While these models now empower many downstream applications from conversation bots to automatic storywriting, they have been shown to generate texts that exhibit social biases. To systematically study and benchmark social biases in open-ended language generation, we introduce the Bias in Open-Ended Language Generation Dataset (BOLD). BOLD is a dataset of 25,673 text generation prompts that has been benchmarking across five domains: profession, gender, race, religion, and political ideology. We also propose new automated metrics for toxicity, psycholinguistic norms, and text gender polarity to measure social biases in open-ended text generation from multiple angles. An examination of text generated from three popular language models reveals that the majority of these models exhibit a larger social bias than human-written Wikipedia text across all domains. With these results we highlight the need to benchmark biases in open-ended language generation and caution users of language generation models on downstream tasks to be cognizant of these embedded prejudices.	2021	45	81	16	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2301.17119v1.pdf">https://arxiv.org/pdf/2301.17119v1.pdf</a>	Bias   Automated   Natural Language Generation   Evaluation Metrics
98	<input checked="" type="checkbox"/>	<input type="checkbox"/>	431ae0a7a7211d	2375077	<a href="https://www.arxiv.org/abs/2307.50777">https://www.arxiv.org/abs/2307.50777</a>	Compression, Transduction, and Creation: A Unified Framework for Evaluating Natural Language Generation	Natural language generation (NLG) spans a broad range of tasks, each of which serves for specific objectives and desires different properties of generated text. The complexity makes automatic evaluation of NLG particularly challenging. Previous work has typically focused on a single task and developed individual evaluation metrics based on specific intuitions. In this paper, we propose a unifying perspective based on the nature of information change in NLG tasks, including compression (e.g., summarization), transduction (e.g., text rewriting), and creation (e.g., dialog). Information alignment, between input, context, and output text, plays a common central role in characterizing the generation. With automatic alignment predictions, we develop a family of interpretable metrics that are suitable for evaluating any aspects of different NLG tasks, often without need of gold reference text. Empirical results show the unified metrics achieve stronger or comparable correlations with human judgement compared to state-of-the-art metrics in each of diverse tasks, including text summarization, style transfer, and knowledge-grounded dialog.	2021	60	38	15	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2307.50777v1.pdf">https://arxiv.org/pdf/2307.50777v1.pdf</a>	Natural Language Generation   Evaluation Metrics
99	<input type="checkbox"/>	<input checked="" type="checkbox"/>	1c30fe0a4394f3c	2343388	<a href="https://www.arxiv.org/abs/2304.33888">https://www.arxiv.org/abs/2304.33888</a>	e-ViL: A Dataset and Benchmark for Natural Language Explanations in Commonsense ViL	Recently, there has been an increasing number of efforts to introduce models capable of generating natural language explanations (NLEs) for their predictions on vision language (VL) tasks. Such models are appealing, because they can provide human-interpretable and comprehensible explanations. However, there is a lack of comparison between existing methods, which is due to a lack of re-usable evaluation frameworks and a scarcity of datasets. In this work, we introduce e-ViL and e-SNLiVE: e-ViL is a benchmark for explainable vision-language tasks that establishes a unified evaluation framework and provides the first comprehensive comparison of existing approaches that evaluate NLEs in VL tasks. It spans four models and three datasets and both automatic metrics and human evaluation are used to assess model-generated explanations. e-SNLiVE is currently the largest existing VL dataset with NLEs (over 430k instances). We also propose a new model that combines UNITER [15], which learns joint embeddings of images and text, and GPT2 [38], a pre-trained language model that is well-suited for explanation. It surpasses the previous state of the art by a large margin across all datasets. Code and data are available here: <a href="https://github.com/compnlp-research/e-vil">https://github.com/compnlp-research/e-vil</a> .	2021	51	43	13	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2304.33888v1.pdf">https://arxiv.org/pdf/2304.33888v1.pdf</a>	Natural Language Generation
100	<input type="checkbox"/>	<input checked="" type="checkbox"/>	03e1c3bdf0a9b6	2447994	<a href="https://www.arxiv.org/abs/2304.79944">https://www.arxiv.org/abs/2304.79944</a>	Zero-Shot Text-Guided Object Generation with Dream Fields	We combine neural rendering with multi-modal image and text representations to synthesize diverse 3D objects solely from natural language descriptions. Our method, Dream Fields, can generate the geometry and color of a wide range of objects without 3D supervision. Due to the scarcity of diverse, captioned 3D data, prior methods only generate objects from a handful of categories, such as ShapeNet. Instead, we guide generation with image-text models pre-trained on large datasets of captioned images from the web. Our method optimizes a Neural Radiance Field from many camera views so that rendered images score highly with a target caption according to a pre-trained CLIP model. To improve fidelity and visual quality, we introduce simple geometric priors, including sparsely-inducing transmittance regularization, scene bounds, and new MLP architectures. In experiments, Dream Fields produce realistic, multi-view consistent object geometry and color from a variety of natural language captions.	2021	64	140	28	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2304.79944v1.pdf">https://arxiv.org/pdf/2304.79944v1.pdf</a>	Language Model
101	<input checked="" type="checkbox"/>	<input type="checkbox"/>	B00b37a2117f1c	2361967	<a href="https://www.arxiv.org/abs/2306.19677">https://www.arxiv.org/abs/2306.19677</a>	Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization	Current dialogue summarization systems usually encode the text with a number of general semantic features (e.g., keywords and topics) to gain more powerful dialogue modeling capabilities. However, these features are obtained via open-domain toolkits that are dialog-agnostic or heavily relied on human annotations. In this paper, we show how DialoGPT, a pre-trained model for conversational response generation, can be developed as an unsupervised dialogue annotator, which takes advantage of dialogue background knowledge encoded in DialoGPT. We apply DialoGPT to label three types of features on two dialogue summarization datasets, SAMSum and AMI, and employ pre-trained and non-pre-trained models as our summarizers. Experimental results show that our proposed method can obtain remarkable improvements on both datasets and achieve new state-of-the-art performance on the SAMSum dataset.	2021	47	42	13	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2306.19677v1.pdf">https://arxiv.org/pdf/2306.19677v1.pdf</a>	Language Model
102	<input type="checkbox"/>	<input checked="" type="checkbox"/>	38e0567e0b3386	2323296	<a href="https://www.arxiv.org/abs/2302.23296">https://www.arxiv.org/abs/2302.23296</a>	CLIPScore: A Reference-Free Evaluation Metric for Image Captioning	Image captioning has conventionally relied on reference-based automatic evaluations, where machine captions are compared against captions written by humans. This is in contrast to the reference-free manner in which humans assess caption quality. In this paper, we report the surprising empirical finding that CLIP (Radford et al., 2021), a cross-modal model pretrained on 409M image-caption pairs from the web, can be used for robust automatic evaluation of image captioning without the need for references. Experiments spanning several corpora demonstrate that our new reference-free metric, CLIPScore, achieves the highest correlation with human judgements, outperforming existing reference-based metrics like CIDEr and SPICE. Information gain experiments demonstrate that CLIPScore, with its light focus on image-text compatibility, is complementary to existing reference-based metrics that emphasize text-text similarities. Thus, we also present a reference-augmented version, RefCLIPScore, which achieves even higher correlation. Beyond literal description tasks, several case studies reveal domains where CLIPScore performs well (clip-art images, all text rating), but where also it is relatively weaker in comparison to reference-based metrics, e.g., news captions that require richer contextual knowledge.	2021	75	190	56	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2302.23296v1.pdf">https://arxiv.org/pdf/2302.23296v1.pdf</a>	Evaluation Metrics
103	<input checked="" type="checkbox"/>	<input type="checkbox"/>	73401891d2101e	2462198	<a href="https://www.arxiv.org/abs/2306.21988">https://www.arxiv.org/abs/2306.21988</a>	QAFactEval: Improved QA-Based Factuality Consistency Evaluation for Summarization	Factuality consistency is an essential quality of text summarization models in practical settings. Existing work in evaluating this dimension can be broadly categorized into two lines of research, entailment-based and question answering (QA)-based metrics, and different experimental setups often lead to contrasting conclusions as to which paradigm performs the best. In this work, we conduct an extensive comparison of entailment and QA-based metrics, demonstrating that carefully choosing the components of a QA-based metric, especially question generation and answerability classification, is critical to performance. Building on those insights, we propose an optimized metric, which we call QAFactEval, that leads to a 14% average improvement over previous QA-based metrics on the SummaC factuality consistency benchmark, and also outperforms the best-performing entailment-based metric. Moreover, we find that QA-based and entailment-based metrics can offer complementary signals and be combined into a single metric for a better performance boost.	2021	56	49	18	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2306.21988v1.pdf">https://arxiv.org/pdf/2306.21988v1.pdf</a>	Evaluation Metrics
104	<input type="checkbox"/>	<input checked="" type="checkbox"/>	194ae47df37ee	244714	<a href="https://www.arxiv.org/abs/2304.07144">https://www.arxiv.org/abs/2304.07144</a>	Vector Quantized Diffusion Model for Text-to-Image Synthesis	We present the vector quantized diffusion (VQ-Diffusion) model for text-to-image generation. This method is based on a vector quantized variational autoencoder (VQ-VAE) whose latent space is modified by a conditional variant of the recently developed Denoising Diffusion Probabilistic Model (DDPM). We find that this latent-space method is well-suited for text-to-image generation tasks because it not only eliminates the unconditional bias with existing methods but also allows us to incorporate a mask-and-replace diffusion strategy to avoid the accumulation of errors, which is a serious problem with existing methods. Our experiments show that the VQ-Diffusion provides significant results when compared with conventional autoregressive (AR) models and generative models of parameter count. We compare our method with previous QA-based text-to-image methods. Our VQ-Diffusion can handle more diverse scenes and improve the synthesized image quality by a large margin. Finally, we show that the image generation performance can be made highly competitive with the current AR methods. With its superior AR performance, the VQ-Diffusion model can be used as a high-quality image generator and hence to quite time consuming even for normal size images. The VQ-Diffusion allows us to achieve a better trade-off between quality and speed. Our experiments indicate that the VQ-Diffusion model with the reparameterization is 15 times faster than traditional AR methods while achieving a better image quality. The code and models are available at <a href="https://github.com/compnlp-research/vq-diffusion">https://github.com/compnlp-research/vq-diffusion</a> .	2021	73	195	18	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2304.07144v1.pdf">https://arxiv.org/pdf/2304.07144v1.pdf</a>	Bias
105	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ce0ca360036037	2320795	<a href="https://www.arxiv.org/abs/2302.20795">https://www.arxiv.org/abs/2302.20795</a>	Self-Diagnosis and Self-Debiasing: A Proposal for Reducing Corpus-Based Bias in NLP	This paper contains prompts and model outputs from the internet, language models pick up and reproduce all kinds of undesirable biases that can be found in the data. They often generate racist, sexist, violent, or otherwise toxic language. As large models require millions of training examples to achieve good performance, it is difficult to absolutely prevent them from being exposed to such content. In this paper, we first demonstrate a surprising finding: Pretrained language models can recognize, to a considerable degree, their undesirable biases and the toxicity of the content they produce. We refer to this capability as self-diagnosis. Based on this finding, we then propose a debiasing algorithm that, given only a few hundred examples of the undesirable behavior, reduces the probability of a language model producing problematic text. We refer to this approach as self-debiasing. Self-debiasing does not rely on manually curated word lists, nor does it require any training data changes to the model's parameters. While we try to measure the model's self-diagnosis and self-debiasing capabilities, we believe our approach to be an important step in this direction.	2021	55	134	17	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2302.20795v1.pdf">https://arxiv.org/pdf/2302.20795v1.pdf</a>	Bias
106	<input type="checkbox"/>	<input checked="" type="checkbox"/>	a250224cf2144	2447144	<a href="https://www.arxiv.org/abs/2304.07144">https://www.arxiv.org/abs/2304.07144</a>	ZeroCap: Zero-Shot Image-to-Text Generation for Visual-Semantic Arithmetic	Recent text-to-image matching models apply contrastive learning to large corpora of uncensored pairs of images and sentences. While such models can provide a powerful score for matching and summarizing zero-shot tasks, they are not capable of generating caption given an image. In this work, we reurpose such models to generate a descriptive text given an image at inference time, without any further training or tuning step. This is done by combining the visual-semantic model, benefiting from the knowledge in both web-scale models. The resulting captions are much less restrictive than style captions obtained by supervised captioning models. Moreover, as a zero-shot learning method, it is extremely flexible and we demonstrate its ability to perform image arithmetic in which the inputs can be either images or text and the output is a sentence. This enables novel high-level visual tasks such as comparing two images or performing visual analogy tests. Our code and pre-trained models are available at <a href="https://github.com/compnlp-research/zerocap">https://github.com/compnlp-research/zerocap</a> .	2021	80	48	8	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2304.07144v1.pdf">https://arxiv.org/pdf/2304.07144v1.pdf</a>	Paraphrase   Language Model   Bias   Privacy   Controllable   Creative   Tr
107	<input type="checkbox"/>	<input checked="" type="checkbox"/>	a0c78801e45d235458	2324586	<a href="https://www.arxiv.org/abs/2302.24586">https://www.arxiv.org/abs/2302.24586</a>	Text2Event: Controllable Sequence-to-Structure Generation for End-to-end Event Extraction	Event extraction is challenging due to the complex structure of event records and the semantic gap between text and event. Traditional methods usually extract event records by decomposing the complex structure prediction task into multiple subtasks. In this paper, we propose Text2Event, a sequence-to-structure generation paradigm that can directly extract events from the text in an end-to-end manner. We design a sequence-to-structure network for unified event extraction, a constrained decoding algorithm for event knowledge injection during inference, and a curriculum learning algorithm for efficient model learning. Empirical results show that, by unifying different tasks in a single model and universally predicting different task labels in a single method, we can achieve competitive performance with state-of-the-art models that have both supervised learning and transfer learning settings.	2021	57	112	39	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2302.24586v1.pdf">https://arxiv.org/pdf/2302.24586v1.pdf</a>	Controllable
108	<input type="checkbox"/>	<input checked="" type="checkbox"/>	a0219974290a1c	2358988	<a href="https://www.arxiv.org/abs/2305.08988">https://www.arxiv.org/abs/2305.08988</a>	Increasing Faithfulness in Knowledge-Grounded Dialogue with Controllable Features	Knowledge-grounded dialogue systems are intended to convey information that is based on evidence provided in a given source text. We discuss the challenges of training a generative neural dialogue model for such systems that is controlled to stay faithful to the evidence. Existing datasets contain a mix of conversational responses that are faithful to selected evidence as well as more subjective, off-the-charts responses. We propose different evaluation measures to disentangle these different styles of responses by quantifying the informativeness and objectivity. At training time, additional inputs based on these evaluation measures are given to the dialogue model. At generation time, these additional inputs act as stylistic controls that encourage the model to generate responses that are faithful to the evidence. We also investigate the usage of additional controls at decoding time using resampling techniques. In addition to automatic metrics, we perform a human evaluation study where raters judge the output of these controlled generation models to be generally more objective and faithful to the evidence compared to baseline dialogue systems.	2021	28	45	13	[[Category: "Computer [JournalArticle",	<a href="https://arxiv.org/pdf/2305.08988v1.pdf">https://arxiv.org/pdf/2305.08988v1.pdf</a>	Evaluation Metrics   Controllable

				Online misinformation is a prevalent societal issue, with adversaries relying on tools ranging from cheap fakes to sophisticated deep fakes. We are motivated by the threat scenario where an image is used out of context to support a certain narrative. While some prior datasets for detecting image-text inconsistency generate template text manipulation, we propose a dataset where both image and text are unmanipulated but mismatched. We introduce several strategies for automatically retrieving convincing images for a given caption, capturing cases with inconsistent entities or semantic context. Our large-scale automatically generated the NewsCLIPings Dataset: (1) demonstrates that machine-driven image repurposing is now a realistic threat, and (2) provides samples that represent challenging instances of mismatch between text and image in news that are able to mislead humans. We benchmark several state-of-the-art models on our dataset and analyze their performance across different pretraining domains and visual backbones.	2021	27	23	9	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2012.088">https://arxiv.org/pdf/2012.088</a>	Machine   Natural Language Generation
109				The dominant paradigm for neural text generation is left-to-right decoding from autoregressive language models. Constrained or controllable generation under complex constraints, however, requires foresight to plan ahead feasible future paths. Drawing inspiration from the A* search algorithm, we propose NeuroLogic-A*, a decoding algorithm that incorporates heuristic estimates of future cost. We develop lookahead heuristics that are efficient for large-scale language models, making our method a drop-in replacement for common techniques such as beam search and top-k sampling. To enable constrained generation, we build on NeuroLogic decoding (Lu et al., 2021), combining its flexibility in incorporating logical constraints with A*’s heuristic estimates of future constraint violations. Our approach outperforms competitive baselines on five generation tasks, and achieves new state-of-the-art performance on table-to-text generation, constrained machine translation, and keyword-constrained generation. The improvements are particularly notable on tasks that require complex constraints satisfaction or in low-shot or zero-shot settings. NeuroLogic-A*’s success illustrates the power of decoding for improving and enabling new capabilities of large language models.	2021	65	44	6	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2002">https://arxiv.org/pdf/2002</a>	Paraphrase   Bias   Privacy   Controllable   Creative   Machine   Cost   Text
110				Document grounded generation is the task of using the information provided in the document to improve text generation. This work focuses on two different document grounded generation tasks: Wikipedia Update or verification task and Dialogue response generation. Our work introduces two novel abstractive decoder models focusing on building context driven representation of the document and enabling specific attention to the information in the document. Additionally, we provide a stronger BART baseline for these tasks. Our proposed techniques outperform existing methods on both automated (at least 48% increase in BLEU-4 points) and human evaluation for closeness to evaluation and relevance to the document. Furthermore, we perform comprehensive manual inspection of the generated output and categorize errors to provide insights into future directions in modeling these tasks.	2021	55	23	8	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2001">https://arxiv.org/pdf/2001</a>	Automated
111				Generating code-switched text is a problem of growing interest, especially given the scarcity of corpora containing large volumes of real code-switched text. In this work, we adapt a state-of-the-art neural machine translation model to generate Hindi-English code-switched sentences starting from monolingual Hindi sentences. We outline a carefully designed curriculum of training steps, including the use of synthetic code-switched text, that enable the model to generate high-quality code-switched text. Using text generated from our model as data augmentation, we show significant reductions in perplexity on a language modeling task, compared to using text from other generative models of CS text. We also show improvements using our text for a downstream code-switched natural language inference task. Our generated text is further subjected to a rigorous evaluation using a human evaluation study and a range of objective metrics, where our performance comparable (and sometimes even superior) to code-switched text obtained via crowd workers who are native Hindi speakers.	2021	50	21	5	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2002">https://arxiv.org/pdf/2002</a>	Paraphrase   NLP   Language Model   Evaluation Metrics   Bias   Privacy
112				Fine-tuning is the de facto way of leveraging large pretrained language models for downstream tasks. However, fine-tuning modifies all the language model parameters and therefore necessitates storing a full copy for each task. In this paper, we propose prefix-tuning, a lightweight alternative to fine-tuning for natural language generation tasks, which keeps language model parameters frozen and instead optimizes a sequence of continuous task-specific vectors, which we call the prefix. Prefix-tuning draws inspiration from prompting for language models, allowing subsequent tokens to attend to this prefix as if it were “virtual tokens”. We apply prefix-tuning to GPT-2 for table-to-text generation and to BART for summarization. We show that by learning only 0.1% of the parameters, prefix-tuning obtains comparable performance in the full data settings, outperforms fine-tuning in low-data settings, and extrapolates better to examples with topics that are unseen during training.	2021	53	1139	185	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2002">https://arxiv.org/pdf/2002</a>	Task
113				We present CoToT, a pre-trained, transformer-based encoder-decoder model that learns the representative content from natural language (NL) and programming language (PL) using self-supervision. CoToT is pre-trained on large programming language corpora to learn a general understanding of language and code. CoToT supports downstream NL-PL tasks such as code summarization/documentation, code generation, defect detection, and code debugging. We train CoToT on different combinations of PL corpora including both “benign” and “unbenign” data. Here, benign data is the combination of text and corresponding code snippets, whereas unbenign data is merely code snippets. We first evaluate CoToT with multi-task learning: we perform Code Summarization on 6 different programming languages and Code Refinement on both small and medium size datasets featured in the CodeXGLUE dataset. We further conduct extensive experiments to investigate CoToT on other tasks within the CodeXGlue dataset, including Code Generation and Defect Detection. We consistently achieve SOTA results in these tasks, demonstrating the versatility of our model.	2021	25	63	13	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2001">https://arxiv.org/pdf/2001</a>	Task   Detection
114				Generating long and coherent text is an important but challenging task, particularly for open-ended language generation tasks such as story generation. Despite the successes in modeling intra-sentence coherence, existing generation models (e.g., BART) struggle to maintain coherence throughout the generated text. We conjecture that this is because of the difficulty to decide to update the high-level semantics and discourse structures in the context beyond token-level co-occurrence. In this paper, we propose a long text generation model, which can represent the prefix semantics at sentence level and discourse level in the decoding process. To this end, we propose two pretraining objectives to learn the representations by predicting intra-sentence semantic similarity and distinguishing between normal and shuffled sentence orders. Extensive experiments show that our model can generate more coherent texts than state-of-the-art baselines.	2021	36	32	7	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2001">https://arxiv.org/pdf/2001</a>	Paraphrase   Language Model   Bias   Privacy   Creative   Task   Text Gen
115				Pretrained language models (PLMs) have recently advanced graph-to-text generation, where the input graph is linearized into a sequence and fed into the PLM to obtain its representation. However, efficiently encoding graphs in PLMs is challenging because of the combinatorial explosion of graph representations. In this paper, we propose StructuAdapt, an adapter module to encode graph structure into PLMs. Contrary to prior work, StructuAdapt effectively models interactions among the nodes based on the graph connectivity, only training graph structure-aware adapter parameters. In this way, we incorporate task-specific knowledge while maintaining the topological structure of the graph. We empirically show the benefits of explicitly encoding graph structure into PLMs using StructuAdapt, outperforming the state of the art on two AMR-to-text datasets, training only 5.1% of the PLM parameters.	2021	66	37	6	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2001">https://arxiv.org/pdf/2001</a>	Paraphrase   Natural Language Generation   Bias   Privacy   Controllable
116				The availability of large-scale image captioning and visual question answering datasets has contributed significantly to recent successes in vision-and-language pretraining. However, these datasets are often collected with overreductive requirements irrelevant from their original target tasks (e.g., image caption corpora are collected with the resulting dataset size and diversity. We take a step further in pushing the limits of vision-and-language pretraining, able by relaxing the data collection pipeline to introduce in Conceptual 12M (CC1M) [54], a dataset with 12 million image-text pairs specifically meant to be used for vision-and-language pre-training. We perform an analysis of this dataset and benchmark its effectiveness against CC3M on multiple downstream tasks with an emphasis on long tail image recognition. Our results clearly illustrate the benefit of scaling up the size of the dataset, as indicated by the new state-of-the-art results on both the image and language tasks, surpassing previous benchmarks. 1	2021	96	330	72	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2012.088">https://arxiv.org/pdf/2012.088</a>	Training
117				Code summarization, code generation and program conversion between programming language (PL) and natural language (NL), while code translation aware the migration of legacy code from one PL to another. This paper introduces PLBART, a sequence-to-sequence model capable of performing a broad spectrum of program and language understanding and generation tasks. PLBART is pre-trained on an extensive collection of Java and Python functions and associated NL text via denoising autoencoders. Experiments on code summarization in the English language, code generation, and code translation in seven programming languages show that PLBART outperforms or rivals state-of-the-art models. Moreover, experiments on discriminative tasks, e.g., program repair, clone detection, and variable code detection, demonstrate that PLBART’s effectiveness in program understanding. Furthermore, analysis reveals that PLBART learns program syntax, style, etc., (i.e., identifier naming convention), logical flow (e.g., “if” block inside an “else” block is equivalent to “else if” block) that are crucial to program semantics with limited data access.	2021	61	274	63	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2001">https://arxiv.org/pdf/2001</a>	Training   Detection
118				Abstract Text-to-image generation is a novel, yet challenging task. We parameterize classical natural language-oriented food logic systems using a Transformer-based auto-regressive language model, which subsumes different dialog modules into a single neural network. We pre-train, on heterogeneous dialog corpora, a task-grounded response generation model, which can generate dialog responses grounded in user goals and real-world knowledge for task completion. The pre-trained model can be efficiently adapted to accomplish new tasks with a handful of task-specific dialog data, where training samples are generated by human guides interacting with the system. Experiments show that (i) Solosol creates new state-of-the-art on well-studied task-oriented dialog benchmarks, including CamBART and MultiWOZ (v2), (ii) in the few-shot fine-tune settings, Solosol significantly outperforms existing methods, and (iii) the use of machine teaching substantially reduces the labeling cost of fine-tuning. The pre-trained models and codes are available at <a href="https://github.com/ysy1993">https://github.com/ysy1993</a> .	2021	84	104	19	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2012.088">https://arxiv.org/pdf/2012.088</a>	Cost
119				Scaling Visual Question Answering (VQA) to the open-domain and multi-hop nature of web searches, requires fundamental advances in visual representation learning, knowledge aggregation, and language understanding. In this work, we introduce WebQA, a novel visual question-answering benchmark that proves difficult for large-scale visual question-answering models which lack language groundable visual representations for novel objects and the ability to reason, yet vital for humans. WebQA mirrors the way humans use the web: 1) Ask a question, 2) Choose sources to aggregate, and 3) Produce a fluent language response. This is the behavior we would be expecting from IoT devices and digital assistants. We argue that a model can better reason about knowledge in images if it can text. WebQA includes a new web-querying VQA task to ensure improved visual understanding does not come at the cost of language understanding. Our challenge for the community is to create unified multimodal reasoning models that answer questions regardless of the source modality, moving us closer to digital assistants that not only query language web search, but also the richer visual online world.	2021	40	30	10	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2012.088">https://arxiv.org/pdf/2012.088</a>	Cost
120				Vision-language pre-training (VLP) on large-scale image-text pairs has achieved great success for the cross-modal downstream tasks. The most existing pre-training methods mainly adopt a two-step training procedure, which firstly employs a pre-trained object detector to extract region-based visual features, then concatenates the image representation and text embedding as the input of Transformer to train. However, these methods are inefficient in learning long-range visual dependencies and the computation of large-scale patch-pairwise dot-product is prohibitive. In this paper, we propose the first end-to-end vision-language pre-trained model for both VLP understanding and generation, namely E2E-VLP, where we build a unified Transformer framework to jointly learn visual representation, and semantic alignment between image and text. We incorporate the tasks of object detection and image captioning into pre-training with a unified Transformer encoder-decoder architecture for enhancing visual learning. An extensive set of experiments have been conducted on well-established vision-language downstream tasks to demonstrate the effectiveness of this novel VLP paradigm.	2021	42	70	13	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2001">https://arxiv.org/pdf/2001</a>	Detection
126				In this work, we propose TextGAN, a novel framework for multi-modal image generation and manipulation with textual descriptions. The proposed method consists of three components: StyleGAN inversion module, linguistic similarity learning, and instance-level optimization. The StyleGAN module learns the latent space of a well-trained StyleGAN. The linguistic similarity module learns the text-image matching by mapping the image and text into a common embedding space. The instance-level optimization is for identity preservation in manipulation. Our model can produce diverse and high-quality images with an unprecedented resolution of 1024x2. Using a control mechanism based on style-mixing, our TextGAN inherently supports image synthesis with multi-modal inputs, such as sketches or semantic labels, with or without instance guidance. To facilitate text-guided multi-modal synthesis, we propose the Multi-Modal ControlNet-HQ, a large-scale dataset consisting of real face images and corresponding semantic labels, segmentation map, sketch, and textual descriptions. Extensive experiments on the introduced dataset demonstrate the superior performance of our proposed method. Code and data are available at <a href="https://github.com">https://github.com</a> .	2021	46	163	43	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2012.033">https://arxiv.org/pdf/2012.033</a>	Text Generation (None)
127				Recent progress in generative language models has enabled machines to generate astonishingly realistic texts. While there are many legitimate applications of such models, there is also a rising need to distinguish machine-generated texts from human-written ones (e.g., fake news detection). However, to our best knowledge, there is currently no benchmark environment with datasets and tasks to systematically study the so-called Turing Test problem for neural text generation methods. In this work, we present the TuringBench benchmark environment, which is comprised of (1) a dataset with 200K human- or machine-generated samples across 20 labels (Human, GPT-1, GPT-2, small, GPT-2, medium, GPT-2, large, GPT-2, x, GPT-2, PyTorch, GPT-3, GROVER, base, GROVER, large, GROVER, mega, CTRL, XLNet, base, XLNet, large, FARE, sent1), (FAIR, word2vec, TRANSFORMER, XL, Pytorch, dist, Pytorch, dist, GPT-2) two benchmark tasks, i.e., Turing Test (TT) and Authorship Attribution (AA), and (2) a series of leaderboards. Our preliminary empirical results using TuringBench show that FARE, word2vec and GPT-3 are the current winners, among all language models tested, in generating the most human-like indistinguishable texts with the lowest F1 score by five state-of-the-art TT detection models. The TuringBench is available at <a href="https://turingbench.bnu.edu.cn">https://turingbench.bnu.edu.cn</a> .	2021	75	25	3	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2001">https://arxiv.org/pdf/2001</a>	Paraphrase   Bias   Privacy   Creative   Machine   Detection   Text Generat
128				Given the potential misuse of recent advances in synthetic text generation by language models (LMs), it is important to have the capacity to attribute authorship of synthetic text. While stylistic organic (i.e., human written) authorship attribution has been quite successful, it is unclear whether similar approaches can be used to attribute a synthetic text to its source LM. We address this question with the key insight that synthetic texts carry subtle distinguishable marks inherited from their source LM and that these marks can be leveraged by machine learning (ML) algorithms for attribution. We propose and test several ML-based attribution methods. Our best attribute built using a fine-tuned version of XLNet (XLNet-FIT) consistently achieves excellent accuracy scores (81% to near perfect 88%) in terms of attributing the parent pre-trained LM behind a synthetic text. Our experiments show promising results across a range of experiments where the synthetic text may be generated using pre-trained LMs, fine-tuned LMs, or by varying text generation parameters.	2021	37	9	3	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2001">https://arxiv.org/pdf/2001</a>	Paraphrase   Language Model   Bias   Privacy   Controllable   Creative   M
129				Text generation has made significant advances in the last few years. Yet, evaluation metrics have lagged behind, as the most popular choices (e.g., BLEU and ROUGE) may correlate poorly with human judgment. We propose BLEURT, a learned evaluation metric for English based on BLEURT. BLEURT can model human judgment with a few thousand possibly biased training examples. A key aspect of our approach is a novel pre-training scheme that uses millions of synthetic examples to help the model generalize. BLEURT provides state-of-the-art results on the last three years of the WMT Metrics shared task and the WebNLG data set. In contrast to a vanilla BERT-based approach, it yields superior results even when the training data is scarce and out-of-distribution.	2020	50	618	184	[[Category: "Computer [JournalArticle", <a href="https://www.activy.org/anth">https://www.activy.org/anth</a>	Paraphrase   Evaluation Metrics   Bias   Privacy   Controllable   Creative
41				We present ToTTo, an open-domain English table-to-text dataset where human-generated text is given a Wikipedia table, text and a set of highlighted table cells, produce a one-sentence description. To obtain generated targets that are natural but also faithful to the source table, we introduce a dataset construction process where annotators directly review existing candidate sentences from Wikipedia. We evaluate human and model performance on the dataset and analyze their generalization ability as well as results achieved by several state-of-the-art baselines. Our results show that while hallucinate phrases that are not supported by the table, suggesting that this dataset can serve as a useful research benchmark for high-precision conditional text generation.	2020	51	194	49	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2001">https://arxiv.org/pdf/2001</a>	Paraphrase   Natural Language Generation   Bias   Privacy   Controllable
42				Recently, large-scale pre-trained language models have demonstrated impressive performance on several commonsense-reasoning benchmark datasets. However, building machines with commonsense to respond realistically plausible sentences remains a challenge. In this paper, we present a constrained text generation task, CommonGen associated with a benchmark dataset, to explicitly test machines for the ability of generative commonsense reasoning. Given a set of common concepts (e.g., dog, fridge, catch, throw), the task is to generate a coherent sentence describing an everyday scenario using these concepts. To make this task a frigate and fairer, we have 1) The CommonGen task is 1) relational reasoning with background knowledge and 2) commonsense reasoning. Commonsense reasoning is a key component of human intelligence, and the ability to work on unseen concept combinations. Our dataset, constructed through a combination of crowdsourced and existing corpora, consists of 77k commonsense descriptions over 35k unique concept sets. Experiments show that there is a large gap between state-of-the-art text generation models (e.g., 1%) and human performance (61.6% vs. 63.5% in SPICE metric). Furthermore, we demonstrate that the learned generative commonsense reasoning capability can be transferred to improve generalization tasks such as CommonsenseQA (78.9% to 74.4 in dev accuracy) by generating additional context.	2020	93	195	47	[[Category: "Computer [JournalArticle", <a href="https://www.activy.org/anth">https://www.activy.org/anth</a>	Paraphrase   Bias   Privacy   Controllable   Creative   Task   Text Generat
43				Despite significant progress in text generation models, a serious limitation is their tendency to produce text that is factually inconsistent with information in the input. Recent work has studied whether textual entailment systems can be used to identify factual errors, however, these sentence-level entailment models are trained to solve a different problem than generation filtering and they do not localize which part of a generation is non-factual. In this paper, we propose a new formulation of entailment that decomposes it at the level of dependency arcs. Rather than focusing on aggregate dependencies, we instead ask whether the semantic relationship manifested by individual dependency arcs in the generated output is supported by the input. Human judgments on this task are difficult to obtain; we therefore propose a method to automatically create data based on existing entailment or paraphrase corpora. Experiments show that our dependency arc entailment model trained on this data can identify factual inconsistencies in paraphrasing and summarization better than those based on out-of-distribution or those based on out-of-distribution or those based on out-of-distribution data.	2020	40	69	15	[[Category: "Computer [JournalArticle", <a href="https://www.activy.org/anth">https://www.activy.org/anth</a>	Paraphrase
44				Systems for story generation are asked to produce plausible and enjoyable stories given an input context. This task is underspecified, as a vast number of diverse stories can originate from a single input. The large output space makes it difficult to build and evaluate story generation models, as (1) existing datasets lack rich enough contexts to meaningfully guide models, and (2) existing evaluations (both crowdsourced and automatic) are unreliable for assessing long-form creative text. To address these issues, we introduce a dataset and evaluation platform built from STORIUM, an online collaborative platform where human-generated data contains 8K lengthy stories (125K tokens) with fine-grained natural language annotations (e.g., character goals and attributes) interspersed throughout each narrative, forming a robust resource for story models. We evaluate language models fine-tuned on our dataset by integrating them into STORIUM, where real authors can query model-generated story continuations and then rate them. Automatic metrics computed over these texts correlate well with both user ratings of generated stories and qualitative feedback from semi-structured user interviews. We release both the STORIUM dataset and evaluation platform as open-source resources to support story generation research.	2020	36	58	9	[[Category: "Computer [JournalArticle", <a href="https://www.activy.org/anth">https://www.activy.org/anth</a>	Creative   Machine   Paraphrase   Natural Language Generation   Language
45				The goal of text-to-text generation is to make machines express like a human in many applications such as conversation, summarization, and translation. It is one of the most important yet challenging tasks in natural language processing (NLP). Various neural encoder-decoder models have been proposed to achieve the goal by learning to map input text to output text. However, the input text alone often provides limited knowledge to generate the desired output of text generation. In this paper, we propose a new framework to generate text by leveraging the knowledge from external knowledge, including internal knowledge embedded in the input text and (ii) external knowledge from outside sources such as knowledge base and knowledge graph into the text generation system. This research topic is known as knowledge-enhanced text generation. In this survey, we present the state-of-the-art research in this field over the past few years. The main content of this two parts: (i) the input text and (ii) the external knowledge.	2020	235	102	13	[[Category: "Computer [JournalArticle", <a href="https://arxiv.org/pdf/2016">https://arxiv.org/pdf/2016</a>	NLP   Natural Language Generation   Text Generation (None)
46				We present the Language Interpretability Toolkit (LIT) an open-source platform for visualization and understanding of NLP models. We focus on core questions about model behavior: Why did my model make this prediction? When does it perform poorly? What happens under a controlled change in the input? LIT integrates local explanations, aggregate analysis, and counterfactual generation into a streamlined, browser-based interface to enable rapid exploration and error analysis. We use case studies for a diverse set of word-level tasks, including exploring counterfactuals for sentiment analysis, generating gender bias in conference systems, and exploring local behavior in text generation. LIT supports a wide range of models—including classification, sequence, and structured prediction—and is highly extensible through a declarative, framework-agnostic API. LIT is under active development, with code and full documentation available at <a href="https://github.com/paijia-code/lit">https://github.com/paijia-code/lit</a> .	2020	49	114	13	[[Category: "Computer [JournalArticle", <a href="https://www.activy.org/anth">https://www.activy.org/anth</a>	NLP   Bias

48			48	275a2e17056d229031	<a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a>	Learning Contextual Representations for Semantic Parsing with Generation-Augmented Pre-Training	Most recently, this field has been significant interest in learning contextual representations for various NLP tasks, by leveraging large scale text corpora to train powerful language models with self-supervised learning objectives, such as Masked Language Model (MLM). Based on a pilot study, we observe three issues of existing general purpose language models when they are applied in the text-to-SQL semantic parsers: fail to select the column mentions of all differences, to infer the column mentions of the cell values, and to integrate these values, to generate the target SQL queries when they are complex. To mitigate these issues, we present a novel framework, Generation-Augmented Pre-Training (GAP-PT), that jointly learns representations of natural language and logical schemas, by leveraging generation model to generate high-quality pre-train data. The GAP Model is trained on 2M natural language and 20K database schema GPT-2. We observe that the model is significantly improved by generation models. Based on experimental results, neural semantic parsers that leverage GAP Model as a representation encoder obtain new state-of-the-art results on both Spider and Criteria-to-SQL benchmarks.	2020	48	64	12	[[category: 'Computer', 'JournalArticle', ''], <a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a> ]	NLP
49			49	03031624049b7c228063	<a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a>	Generate Your Counterfactuals: Towards Controlled Counterfactual Machine Checkers for Text (Mac2hamp) in NLP for Multi-task Learning and Toolkits	Machine Learning has been tremendous growth recently, which has led to its larger adoption of ML systems for educational assessments, credit risk, healthcare, employment, criminal justice, to name a few. The advantages of ML and NLP systems is a crucial aspect and requires a parallel with the field. In this paper, we propose a novel framework GYC to generate of exhaustive counterfactual text, which are crucial for testing these ML systems. Our main contributions include: 1) We introduce GYC, a framework to generate counterfactual samples such that the generation is feasible, diverse, goal-oriented, and effective. 2) We generate counterfactuals with a corresponding text-to-text model, which can directly generate counterfactual text with the generated or sentiment. Our experimental results on various domains show that GYC generates counterfactual samples exhibiting the above four properties. GYC generates counterfactuals that can act as test cases to evaluate a model and any text debasing algorithm.	2020	44	48	9	[[category: 'Computer', 'JournalArticle', ''], <a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a> ]	NLP / Machine
50			50	11ab6c2f11963219124	<a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a>	Transfer Learning, Particular Approaches That Combine Multi-task Learning with Pre-trained Contextualized Embeddings and Fine-tuning, have advanced the field of Natural Language Processing tremendously in recent years. In this paper we present Mac2hamp, a toolkit for easy fine-tuning of contextualized embeddings in multi-task settings. The benefits of Mac2hamp are its flexible configuration options, and the support of a variety of natural language processing tasks. We demonstrate the effectiveness of Mac2hamp on a wide range of natural language processing tasks, including sentiment analysis, named entity recognition, and question answering. We also provide a detailed analysis of the model's performance on these tasks, and discuss the implications of our findings for future research.	2020	197	60	8	[[category: 'Computer', 'JournalArticle', ''], <a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a> ]	NLP	
51			51	342e21fc1032c210605d	<a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a>	Natural Language Generation (NLG) models have recently shown remarkable progress in fluency and coherence. However, existing studies on natural NLG are primarily focused on surface-level realizations with limited emphasis on logical inference, an important aspect of human thinking and language. In this paper, we suggest a new NLG task where a model is tasked with generating natural language statements that are logically entailed by the facts in an open-domain question-answer dataset. To facilitate the study of the proposed logical NLG problem, we use the existing Text2Fact dataset (CTATON) instead with a wide range of logical/symbolic inferences as the test set, and propose new automatic metrics to evaluate the quality of generated NLG. In logical inference, the new task poses challenges to the existing monolithic generation frameworks due to the interplay between sequential and logical inference. In our experiments, we comprehensively survey different generation architectures (LSTM, Seq2Seq, Transformer, Trained ML) trained with different algorithms (RL, Adversarial Training, Coarse-to-Fine) on the dataset and made following observations. 1) Pre-Trained ML can significantly boost both the fluency and logical fidelity metrics, 2) RL and Adversarial Training are trading fluency for fidelity, 3) Coarse-to-Fine generation can help partially alleviate the fidelity issue while maintaining high fluency. The code and data are available at <a href="https://github.com/mehmetoglu/nlg">https://github.com/mehmetoglu/nlg</a> .	2020	48	97	39	[[category: 'Computer', 'JournalArticle', ''], <a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a> ]	Natural Language Generation	
52			52	6f3304da2959f222141	<a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a>	Data-to-text generation has recently attracted substantial interests due to its wide applications. Existing methods have shown impressive performance on an array of tasks. However, they rely on a sequence of labeled data for each task, which is costly to acquire and thus limits their application to new tasks and domains. In this paper, we propose to leverage pre-training and transfer learning to learn a general model for data-to-text generation. We propose a knowledge-grounded pre-training (KGPT), which consists of two parts. 1) a general knowledge-grounded generation model to generate knowledge-enriched text. 2) a pre-training paradigm on a large-scale knowledge-grounded text corpus to learn a general model for data-to-text generation. We present specific text-to-text pre-training and transfer learning. We present specific text-to-text pre-training and transfer learning.	2020	56	80	26	[[category: 'Computer', 'JournalArticle', ''], <a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a> ]	Natural Language Generation / Language Model / Task / Training / Text G	
53			53	06b339f1887022606a	<a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a>	Large-scale pre-trained language models, such as BERT and GPT-2, have achieved excellent performance in language representation learning and free-form text generation. However, these models cannot be directly used for text-to-text generation tasks. To address this challenge, we present GPT2-2T (Two-Stage Transformer), a simple yet novel iterative-based approach for hard-constrained text generation. The proposed method works by progressively inserting new tokens between existing tokens in a parallel manner. This procedure is recursively applied until a sequence is completed. The resulting coarse-to-fine hierarchy makes the generation process intuitive and interpretable. We pre-train our model with the proposed progressive insertion-based approach on a 12GB Wikipedia dataset, and fine-tune it on downstream hard-constrained generation tasks. Non-autoregressive decoding yields an empirically logarithmic time complexity during inference. Experimental results on both News and Yelp datasets demonstrate that PONTER achieves state-of-the-art performance on constrained text generation. We released the pre-trained models and the source code to facilitate future research (this <a href="https://github.com/zhengzhang1990/ponter">https://github.com/zhengzhang1990/ponter</a> ).	2020	56	49	14	[[category: 'Computer', 'JournalArticle', ''], <a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a> ]	Natural Language Generation / Language Model / Training / Text Generation	
54			54	20e8b6c8ed126218487	<a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a>	Text generation from a knowledge base aims to transcend knowledge triplets to natural language descriptions. Most existing methods ignore the faithfulness between a generated text description and the original table, leading to generated information that goes beyond the content of the table. In this paper, for the first time, we propose a novel Transformer-based generation framework to achieve the goal. The core techniques in our method to enforce faithfulness include a new table-to-text optimal transport matching loss and a table-text embedding similarity loss based on the Transformer model. Furthermore, to evaluate faithfulness, we propose a new automatic metric specialized to the table-to-text generation problem. We also provide detailed analysis on each component of our model in our experiments. Automatic and human evaluations show that our framework can generate more faithful and diverse text descriptions than the baseline methods.	2020	40	53	17	[[category: 'Computer', 'JournalArticle', ''], <a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a> ]	Language Model / Text Generation (None)	
55			55	b30642760911125067f	<a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a>	Prior work on Data-to-Text Generation, the task of converting knowledge graph (KG) triplets into natural text, focused on domain-specific benchmark datasets. In this paper, however, we verbalize the entire Entity Wikidata KG, and discuss the unique challenges associated with a broad, open-domain, large-scale verbalization. We further study that verbalizing a comprehensive, encyclopedic KG like Wikidata can be used to integrate structured KGs and natural language corpora, contrast to the traditional approach of integrating these two sources, our approach covers the KG into natural text, allowing it to be seamlessly integrated into existing language models. It carries the further advantages of improved factual accuracy and reduced toxicity in the resulting language model. We evaluate this approach by comparing the performance of the generated text with the baseline methods on a retrieval language model and showing significant improvement on open-domain QA and the LAMA knowledge probe.	2020	51	98	16	[[category: 'Computer', 'JournalArticle', ''], <a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a> ]	Language Model	
56			56	72c0c3001d31218688f	<a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a>	Abstract Meaning Representations (AMRs) are broad-coverage sentence-level semantic graphs. Existing approaches to generating text from AMR have focused on training sequence-to-sequence or graph-to-text models on AMR annotated data only. In this paper, we propose an alternative approach that combines a strong pre-trained language model with cycle-consistent based scoring. Despite the simplicity of the approach, our experimental results show that our model outperforms all previous generation methods on AMR-to-text generation tasks, including the recent of transformer architectures. In addition to the state-of-the-art performance, we provide human evaluation experiments that further substantiate the strength of our approach.	2020	35	73	15	[[category: 'Computer', 'JournalArticle', ''], <a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a> ]	Language Model / Evaluation Metrics / Training / Text Generation (None)	
57			57	0beeac866e21128487f	<a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.00147v1</a>	It is well known that the standard likelihood training and approximation methods in natural language processing are not suitable for text-to-text generation tasks such as language modeling and story generation. In this paper we have analyzed limitations of these models for abstractive document summarization and found that these models are highly prone to hallucinate content that is unfaithful to the input document. We argue that the standard likelihood training and approximation methods are not suitable for text-to-text generation tasks such as language modeling and story generation.	2020	59	450	90	[[category: 'Computer', 'JournalArticle', ''], <a href="https://arxiv.org/abs/2005.00147v1">https://arxiv.org/abs/2005.0014</a>		



69	<input checked="" type="checkbox"/>	<input type="checkbox"/>	39a33093b78ed1215754 <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	End-to-end neural data-to-text (D2T) generation has recently emerged as an alternative to pipeline-based architectures. However, it has faced challenges generalizing to new domains and generating semantically consistent text. In this work, we present DataTune, a neural, end-to-end data-to-text generation system that makes data representation assumptions about the data representations used in the target domain. We take a two-stage generation-re-ranking approach, combining a fine-tuned language model with a semantic fidelity classifier. Each component is learnt end-to-end without needing dataset-specific heuristics, entity delexicalization, or post-processing. We show that DataTune achieves state-of-the-art results on automatic fidelity across four major D2T datasets (LDC2017T10, WN18G, VGGO, and CoNLL-E2), with fluency assessed by human annotators as meeting or exceeding the human-written reference texts. Our generated text has better semantic fidelity than the state-of-the-art on these datasets. We further demonstrate that our model-based semantic fidelity score is a better assessment tool compared to traditional heuristic-based measures of semantic accuracy.	2020	56	58	13	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Language Model   Automated   Text Generation (None)
70	<input type="checkbox"/>	<input checked="" type="checkbox"/>	581c3d28d7fca0218596f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	COVID-Classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images	2020	32	110	9	[[category: 'Medicine', 'source: 'xdrma', 'https://www.nature.com/articles/2020.03.10.330000/]]	Automated
71	<input checked="" type="checkbox"/>	<input type="checkbox"/>	169866ab0b0841218673b <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	Syntax-Guided Controlled Generation of Paraphrases	2020	54	55	8	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Paraphrase   Automated
72	<input checked="" type="checkbox"/>	<input type="checkbox"/>	94551d326b5e11218595f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	Enabling Language Models to Fill in the Blanks	2020	28	123	28	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Task
73	<input type="checkbox"/>	<input checked="" type="checkbox"/>	9678afaf6b501214623f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	ScrabbleGAN: Semi-Supervised Varying-Length Handwritten Text Generation	2020	39	65	18	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Task   Training   Text Generation (None)
74	<input type="checkbox"/>	<input checked="" type="checkbox"/>	a60bce1e6b7d221970 <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	A Diagnostic Study of Explainability Techniques for Text Classification	2020	33	100	23	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Cost
75	<input checked="" type="checkbox"/>	<input type="checkbox"/>	4ef0c182e097f122225 <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	Few-Shot Learning for Opinion Summarization	2020	43	41	14	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Cost
76	<input type="checkbox"/>	<input checked="" type="checkbox"/>	46031f78563693227840 <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	Multilingual Speech Translation from Efficient Fine-tuning of Pretrained Models	2020	55	81	10	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Cost
77	<input type="checkbox"/>	<input checked="" type="checkbox"/>	14ee16280426b1220647f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	What Affects Usage Satisfaction in Mobile Payments? Modeling User Generated Content to Develop the Digital Service Usage Satisfaction Model	2020	128	104	7	[[category: 'Medicine', 'JournalArticle', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Cost
78	<input type="checkbox"/>	<input checked="" type="checkbox"/>	f1c2ebf46c0cb3121297f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	Optimal Design and Model Predictive Control of Stochastic HRES: A Real Case Study for Residential Demand Side Management	2020	96	44	2	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	CO2 Emission
83	<input type="checkbox"/>	<input checked="" type="checkbox"/>	52d181218121374219737f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	ContourNet: Taking a Further Step Toward Accurate Arbitrary-Shaped Scene Text Detection	2020	51	118	17	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Detection
84	<input type="checkbox"/>	<input checked="" type="checkbox"/>	ea1df28226253a221095f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	TextRay: Contour-based Geometric Modeling for Arbitrary-shaped Scene Text Detection	2020	40	48	11	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Detection
85	<input checked="" type="checkbox"/>	<input type="checkbox"/>	01e85322e7f11226254f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	Detecting Hallucinated Content in Conditional Neural Sequence Generation	2020	62	72	10	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Detection
86	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2734b6464b3571220390f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	TwepFake: About detecting deepfake tweets	2020	66	55	5	[[category: 'Medicine', 'JournalArticle', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Language Model   Detection
87	<input checked="" type="checkbox"/>	<input type="checkbox"/>	5ea29444b065122377f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	Neural Deepfake Detection with Textual Features of Text	2020	27	22	4	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Language Model   Controllable   Creative   Machine   Task   Detection   Ma
88	<input type="checkbox"/>	<input checked="" type="checkbox"/>	e8e8af6a3dab1a21827f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	PHINC: A Parallel Hinglish Social Media Code-Mixed Corpus for Machine Translation	2020	15	44	3	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Controllable   Creative   Machine   Machine Generated Text (None)
130	<input checked="" type="checkbox"/>	<input type="checkbox"/>	8b027b6594b1e150041f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	HellaGag: Can a Machine Really Finish Your Sentence?	2019	22	397	66	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Paraphrase   Bias   Privacy   Controllable   Creative   Machine   Task   Cos
131	<input type="checkbox"/>	<input checked="" type="checkbox"/>	ac0998a671b1592763f <a href="https://www.aclweb.org/anthology/2020.acl-main.100/">https://www.aclweb.org/anthology/2020.acl-main.100/</a>	Knowledge-driven Encode, Retrieve, Paraphrase for Medical Image Report Generation	2019	42	130	26	[[category: 'Computer [JournalArticle]', 'https://www.aclweb.org/anthology/2020.acl-main.100/']	Paraphrase

[illegible]

Enabled by the pull-based development model, developers can easily contribute to a project through pull requests (PRs). When creating a PR, developers can add a free-form description to describe what changes are made in the PR and/or why. Such a description is helpful for reviewers to get a high-level overview of the PR without browsing the details and may reduce the possibility of the PR being ignored or rejected. However, developers sometimes neglect to write descriptions for PRs. For example, in our collected dataset with over 333K PRs, more than 34% of the PR descriptions are empty. To alleviate this problem, we propose an approach to automatically generate PR descriptions based on the commit messages and the added source code comments in the PRs. We regard this problem as a text summarization problem and solve it using a novel sequence-to-sequence model. To cope with out-of-vocabulary words in software artifacts and bridge the gap between the training loss function of the sequence-to-sequence model and the evaluation metric ROUGE, which has been shown to correspond to human evaluation, we integrate the pointer generator and develop optimization for ROUGE using reinforcement learning and a special loss function. We build a dataset with over 41K PRs and evaluate our approach on this dataset through ROUGE and a human evaluation. Our evaluation results show that our approach outperforms two baselines by significant margins.												
157	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Self2Doc2Vec239 2025777	Automatic Generation of Pull Request Descriptions	form: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R	2019	61	60	7	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1909.0060	Automated
158	<input type="checkbox"/>	<input checked="" type="checkbox"/>	220ea16cf578f4 9118877	Time Sensitive Network (TSN) is a set of standards enabling a performance deterministic communication using different scheduling mechanisms. Due to the size of industrial networks, configuring TSN networks is challenging to be done manually. We present TSNeched, a tool for automatic generation of schedules for TSN. TSNeched takes as input the logical topology of a network, expressed as flows, and outputs schedules for TSN switches using an SAT solver. The generated schedule guarantees the desired network performance (specified in terms of latency and jitter) if such schedules exist. TSNeched can synthesize IEEE 802.1Qbv schedules and supports unicast and multicast flows, such as in Publish/Subscribe networks. TSNeched can be run as a standalone tool and also allows rapid prototyping with the available JAVAR. We evaluate TSNeched on a number of realistic size network topologies. TSNeched can generate high performance schedules, with average latency less than 51000ns, jitter@99, and average jitter less than 2200ns/multi-flow, for TSN networks, with up to 73 subscribers and up to 10 multicast flows.	Generating neural network representations for the structured tables which consist of multiple attribute-value tuples in a convenient way to help people to understand the tables. Most neural table-to-text models are based on the encoder-decoder framework. However, it is hard for a vanilla encoder to learn the accurate semantic representation of a complete table. The challenges are two-fold: firstly, the table-to-text dataset often contain a great number of attributes across different domains, thus it is hard for the encoder to incorporate these heterogeneous resources. Secondly, the single encoder also has difficulties in modeling the attribute-value structure of the tables. To this end, we first propose a two-level hierarchical encoder with coarse-to-fine attention to handle the attribute-value structure of the tables. Furthermore, to capture the accurate semantic representations of the tables, we propose 3 joint tasks apart from the prime encoder-decoder learning, namely auxiliary sequence labeling task, slot autoencoder and multi-labeling classification, as the auxiliary supervisors for the table encoder. We test our models on the widely used dataset Wikitables which contains Wikidata infoboxes and related descriptions. The dataset contains complex tables as well as large number of attributes. We achieve the state-of-the-art performance on both automatic and human evaluation metrics.	2019	32	134	5	[[category: "Computer [JournalArticle]",	https://link.springer.com/book	Automated
159	<input type="checkbox"/>	<input checked="" type="checkbox"/>	c7d5a77e299f1 2029737	TSNeched: Automated Schedule Generation for Time Sensitive Networking		2019	13	25	5	[[category: "Computer [JournalArticle]",	https://posinfo.info/ubp/bjwp	Automated
160	<input type="checkbox"/>	<input checked="" type="checkbox"/>	c3b9991fa4d38c 7022757	Hierarchical Encoder with Auxiliary Supervision for Neural Table-to-Text Generation: Learning Better Representation for Tables		2019	40	33	10	[[category: "Computer [JournalArticle]",	https://ojs.aaai.org/index.php	Task
161	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ce07f95837a0e1 2025411	Enhancing AMR-to-Text Generation with Dual Graph Representations		2019	48	52	9	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1909.0003	Language Model   Task
162	<input checked="" type="checkbox"/>	<input type="checkbox"/>	c48925d511c0e1 1023511	Jointly Measuring Diversity and Quality in Text Generation Models		2019	23	37	7	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1904.0393	Task
163	<input checked="" type="checkbox"/>	<input type="checkbox"/>	5022ba19d36e1 2027847	Reducing Sentiment Bias in Language Models via Counterfactual Evaluation	Generating Sentiment Preserving Fake Online Reviews Using Neural Language Models and Their Human- and Machine-based Detection	2019	56	103	13	[[category: "Computer [JournalArticle]",	https://www.aclweb.org/anth	Evaluation Metrics   Bias   Language Model   Training
164	<input checked="" type="checkbox"/>	<input type="checkbox"/>	432eb11c27584f 1981477	Text Generation with Exemplar-based Adaptive Decoding		2019	38	54	9	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1907.0911	Detection   Language Model   Machine   Training   Machine Generated Text
165	<input checked="" type="checkbox"/>	<input type="checkbox"/>	7505bdc8de4c 9298667	Denouncing based Sequence-to-Sequence Pre-training for Text Generation		2019	72	43	8	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1904.0446	Training
166	<input checked="" type="checkbox"/>	<input type="checkbox"/>	44de9be3ad67 2013067	A Scalable Handwritten Text Recognition System		2019	64	32	6	[[category: "Computer [JournalArticle]",	https://www.aclweb.org/anth	Task   Training
167	<input type="checkbox"/>	<input checked="" type="checkbox"/>	512b23554570e1 1261807	TS-RNN: Text Steganalysis Based on Recurrent Neural Networks		2019	33	50	4	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1904.0891	Cost
174	<input type="checkbox"/>	<input checked="" type="checkbox"/>	be3980abe07c1 1700797	Data-to-text Generation with Entity Modeling		2019	32	43	8	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1905.1309	Detection
175	<input checked="" type="checkbox"/>	<input type="checkbox"/>	416e39ff26243 1748017	Neural data-to-text generation: A comparison between pipeline and end-to-end architectures		2019	47	87	25	[[category: "Computer [JournalArticle]",	https://www.aclweb.org/anth	Text Generation (None)
176	<input checked="" type="checkbox"/>	<input type="checkbox"/>	33bf4346f019e 2016667	Improving Robustness of Machine Translation with Synthetic Noise		2019	34	109	15	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1908.0500	Text Generation (None)
177	<input type="checkbox"/>	<input checked="" type="checkbox"/>	bd0835470647 6785677	Query and Output: Generating Words by Querying Distributed Word Representations for Paragraph Generation		2019	21	66	7	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1902.0060	Machine   Machine Generated Text (None)
178	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0e07f9a2d06eb1 3708647	Toward Diverse Text Generation with Inverse Reinforcement Learning		2018	41	55	3	[[category: "Computer [JournalArticle]",	https://www.aclweb.org/anth	Paraphrase
183	<input checked="" type="checkbox"/>	<input type="checkbox"/>	6e187de0998a 44690507	Towards Explainable NLP-A Generative Explanation Framework for Text Classification		2018	27	70	12	[[category: "Computer [JournalArticle]",	https://www.jair.org/proceed	NLP   Bias   Task
184	<input type="checkbox"/>	<input checked="" type="checkbox"/>	b59646dddc10d1 5315367	Automatic question generation (QG) is a useful yet challenging task in NLP. Recent neural network-based approaches represent the state-of-the-art in this task. In this work, we attempt to strengthen them significantly by adopting a holistic and novel generator-evaluator framework that directly optimizes objectives that reward semantics and structure. The generator is a sequence-to-sequence model that incorporates the structure and semantics of the question being generated. The generator predicts an answer to the passage that the question can pivot on. Employing the copy and coverage mechanisms, it also addresses other contentually important (and possibly rare) keywords in the passage that the question needs to conform to, while not redundantly repeating words. The evaluator model evaluates and assigns a reward to each predicted question based on the structure and semantics of the question. We propose two novel QG-specific metrics for text content and answer quality. We compare our model with existing QG models and find that our model achieves state-of-the-art performances on these three benchmark datasets.	Text generation is a crucial task in NLP. Recently, several adversarial generative models have been proposed to improve the question bias problem in text generation. Though these models gain great success, they still suffer from the problems of reward sparsity and mode collapse. In order to address these two problems, in this paper, we employ inverse reinforcement learning (IRL) for text generation. Specifically, the IRL framework learns a reward function from the output of the model on training data, and then an optimal policy to maximize the expected total reward. Similar to the adversarial generative models, our generator generates questions. Our method has two advantages. (1) The reward function can produce more dense reward signals. (2) the generation policy, trained by "entropy regularized" policy gradient, encourages to generate more diversified results. Experimental results demonstrate that our proposed method outperforms the previous methods on the question bias problem in text generation.	2018	32	96	10	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1811.0011	NLP
185	<input checked="" type="checkbox"/>	<input type="checkbox"/>	56880b6b1017c1 2025777	Putting the Horse before the Cart: A Generator-Evaluator Framework for Question Generation from Text		2018	31	31	4	[[category: "Computer [JournalArticle]",	https://www.aclweb.org/anth	NLP
186	<input type="checkbox"/>	<input checked="" type="checkbox"/>	08a1c2d8c0c20 5445747	Generation of Synthetic Electronic Medical Record Text		2018	24	39	2	[[category: "Computer [JournalArticle]",	https://arxiv.org/pdf/1812.0227	NLP   Machine
188	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0d4509644689 5324527	E2E NLG Challenge Submission: Towards Controllable Generation of Information in neural network NLG systems to be a promising research direction		2018	37	14	4	[[category: "Computer [JournalArticle]",	https://www.aclweb.org/anth	Natural Language Generation   Controllable
189	<input checked="" type="checkbox"/>	<input type="checkbox"/>	5678f704d0f38c1 6707627	Text Simplification without Simplified Corpora		2018	126	8	3	[[category: "Computer Science", source	https://www.jstatsoft.org/	Natural Language Generation   Task
190	<input checked="" type="checkbox"/>	<input type="checkbox"/>	36332171cdce7e1 5296507	Sequence-to-Sequence Models for Data-to-Text Natural Language Generation: Word vs. Character-based Processing and Output Diversity		2018	59	22	2	[[category: "Computer [JournalArticle]",	https://www.aclweb.org/anth	Natural Language Generation



[illegible]

220	<input type="checkbox"/>	<input checked="" type="checkbox"/>	569c9d0164ae04 501968f <a href="https://www.aclweb.org/anthology/2018.acl-main.1968/">https://www.aclweb.org/anthology/2018.acl-main.1968/</a>	Tracking State Changes in Procedural Text: A Challenge Dataset and Models for Process Paragraph Comprehension	We present a new dataset and models for comprehending paragraphs about processes (e.g., photosynthesis), an important genre of text describing a dynamic world. The new dataset, ProPara, is the first to contain natural (rather than machine-generated) text about a changing world (static location and existence) during those changes (81k datapoints). The end task, tracking the location and existence of entities through the text, is challenging because the causal effects of actions are often implicit and need to be inferred. We find that previous models that have worked well on synthetic data achieve only mediocre performance on ProPara, and introduce two new neural models that exploit alternative mechanisms for state prediction, in particular using LSTM input encoding and span prediction. The new models improve accuracy by up to 19%. We are releasing the ProPara dataset and our models to the community.	2018	27	95	20	[{"category": "Computer", "JournalArticle": "https://www.aclweb.org/anthology/2018.acl-main.1968/"}]	Machine   Task   (Machine Generated Text (None))
221	<input checked="" type="checkbox"/>	<input type="checkbox"/>	f6a1f95f16b206 150041f <a href="https://www.aclweb.org/anthology/2018.acl-main.150041f/">https://www.aclweb.org/anthology/2018.acl-main.150041f/</a>	Auditing Data Provenance in Text-Generation Models	To help enforce data-protection regulations such as GDPR and detect unauthorized uses of personal data, we develop a new model auditing technique that helps users check if their data was used to train a machine learning model. We focus on auditing deep-learning models that generate natural language text, including auto-regressive and autoregressive models. These models are at the core of popular online services and are often trained on personal data such as users' messages, searches, chats, and comments. We design and evaluate a black-box auditing method that can detect, with very few queries to a model, if a particular user's texts were used to train (among thousands of other users). We empirically show that our method can successfully audit well-generalized models that are not overfitted to the training data. We also analyze how first-generation models memorize word sequences and explain why this memorization makes them amenable to auditing.	2018	47	152	18	[{"category": "Computer", "Book": "JournalArticle", "https://doi.org/10.1145/3206197.3206210"}]	Machine   Training   Text Generation (None)
222	<input type="checkbox"/>	<input checked="" type="checkbox"/>	5317bc1f01c1b3 490738f <a href="https://www.aclweb.org/anthology/2018.acl-main.490738f/">https://www.aclweb.org/anthology/2018.acl-main.490738f/</a>	Using machine learning for sequence-level automated MRI protocol selection in neurology	Incorrect imaging protocol selection can lead to important clinical findings being missed, contributing to both wasted health care resources and patient harm. We present a machine learning method for analyzing the unstructured text of clinical indications and patient demographics from magnetic resonance imaging (MRI) orders to automatically propose MRI procedures at the sequence level. We compare 3 machine learning models: support vector machine, gradient boosting, and random forest. To a baseline model, we add a feature matrix that includes the most common protocol for all observations in our test set. The model that predicted the most accurate protocol for all observations in our test set, with very few false positives, was the random forest model. This model significantly outperformed the baseline and demonstrated the best performance of the 3 models in terms of accuracy (90%), precision (86%), recall (80%), and F1 score (0.847). This demonstrates the feasibility of automating sequence selection by applying machine learning to MRI orders. Automated sequence selection has important safety, quality, and financial implications and may facilitate improvements in the quality and safety of medical imaging service delivery.	2018	22	47	3	[{"category": "Computer", "JournalArticle": "https://academic.oup.com/ajnr/article/39/1/1/5077777"}]	Automated
223	<input type="checkbox"/>	<input type="checkbox"/>	c7bea706094d1 240251f <a href="https://www.aclweb.org/anthology/2018.acl-main.240251f/">https://www.aclweb.org/anthology/2018.acl-main.240251f/</a>	Recent advances in data-to-text generation have led to the use of large-scale datasets and neural network models which are trained end-to-end, without explicitly modeling what to say and in what order. In this work, we present a neural network architecture which incorporates content selection and planning without sacrificing end-to-end training. We decompose the generation task into two stages. Given a corpus of data records (paired with descriptive documents), we first generate a content plan highlighting which information should be mentioned and in which order and then generate the document while taking the content plan into account. Automatic and human-based evaluation experiments show that our model outperforms strong baselines improving the state-of-the-art on the recently released RatWriteV2 dataset.	2018	0	24	2	[{"category": "Computer", "Review": "https://link.springer.com/book/10.1007/978-1-4939-9999-9_1"}]	Automated	
224	<input checked="" type="checkbox"/>	<input type="checkbox"/>	f24c6419f5364c 521539f <a href="https://www.aclweb.org/anthology/2018.acl-main.521539f/">https://www.aclweb.org/anthology/2018.acl-main.521539f/</a>	Data-to-Text Generation with Content Selection and Planning	Background With the development of artificial intelligence (AI) technology centered on deep-learning, the computer has evolved to a point where it can read a given text and answer a question based on the content of the text. Such a specific task is known as the task of machine comprehension. Existing machine comprehension tasks mostly use datasets of general texts, such as news articles or elementary school-level storybooks. However, no attempt has been made to determine whether an up-to-date deep learning-based machine comprehension model can process biomedical articles as well as general texts. Since there is no dataset for the biomedical literature comprehension task, our work includes generating a large-scale question answering dataset using PubMed and manually evaluating the generated dataset. Methods We present an attention-based deep neural model tailored to the biomedical domain. To further enhance the performance of our model, we used a pretrained word vector and biomedical entity type embedding. We also developed an ensemble method of combining the results of several independent models to reduce the variance of the answers from the models. Results The experimental results showed that our proposed deep neural network model outperformed the baseline model by more than 7% on the new dataset. We also evaluated human performance on the new dataset. The human evaluation result showed that our deep neural model outperformed humans in comprehension by 22% on average. Conclusions In this work, we introduced a new task of machine comprehension in the biomedical domain using a deep neural model. Since there was no large-scale dataset for training deep neural models in the biomedical domain, we created a new close-style dataset, the Biomedical Knowledge Comprehension Test (BKCT), and the Biomedical Knowledge Comprehension Last Sentence (BKML-S) together referred to as Biomedical Knowledge Comprehension) using the PubMed corpus. The experimental results showed that the performance of our model is much higher than that of humans. We observed that our model performed consistently better regardless of the degree of difficulty of a text, whereas humans have difficulty when performing biomedical literature comprehension tasks that require expert-level knowledge.	2018	44	224	41	[{"category": "Computer", "JournalArticle": "https://doi.org/10.1145/3206197.3206210"}]	Task   Training   Text Generation (None)
225	<input type="checkbox"/>	<input checked="" type="checkbox"/>	d54004c0c08b356124 <a href="https://www.aclweb.org/anthology/2018.acl-main.356124f/">https://www.aclweb.org/anthology/2018.acl-main.356124f/</a>	A Pilot Study of Biomedical Text Comprehension using an Attention-Based Deep Neural Reader: Design and Experimental Analysis	Neural text generation, including neural machine translation, image captioning, and summarization, has been quite successful recently. However, during training time, we typically only one reference is considered for each output, even though there are often multiple references available, e.g., 4 references in NIST MT evaluations, and 5 references in image captioning data. We first investigate several different ways of utilizing multiple human references during training. But more importantly, we propose a neural graph-to-sequence model, using a novel LSTM structure for directly encoding graph-like sentences. On a standard benchmark, our model is much higher than that of humans. We observed that our model performed consistently better regardless of the degree of difficulty of a text, whereas humans have difficulty when performing biomedical literature comprehension tasks that require expert-level knowledge.	2018	30	18	6	[{"category": "Medicine", "JournalArticle": "https://medinform.jmir.org/2018/11/e320000/"}]	Task
226	<input checked="" type="checkbox"/>	<input type="checkbox"/>	9a716b5f64557 521249f <a href="https://www.aclweb.org/anthology/2018.acl-main.521249f/">https://www.aclweb.org/anthology/2018.acl-main.521249f/</a>	Multi-Reference Training with Pseudo-References for Neural Translation and Text Generation	Data scarcity is one of the main obstacles of domain adaptation in spoken language understanding (SLU) due to the high cost of creating manually tagged SLU datasets. Recent works in neural text generative models, particularly latent variable models such as variational autoencoder (VAE), have shown promising results in regards to generating plausible and natural sentences. In this paper, we propose a novel generative architecture which leverages the generative power of latent variable models to jointly synthesize and generate utterances. Our experiments show that existing SLU models trained with the additional synthetic examples achieve performance gains. Our approach not only helps alleviate the data scarcity issue in the SLU task for many datasets but also inductively improves language understanding performance for various SLU models, supported by extensive experiments and rigorous statistical testing.	2018	20	21	2	[{"category": "Computer", "JournalArticle": "https://www.aclweb.org/anthology/2018.acl-main.521249f/"}]	Machine   Training
227	<input type="checkbox"/>	<input checked="" type="checkbox"/>	b21850c88960a 521779f <a href="https://www.aclweb.org/anthology/2018.acl-main.521779f/">https://www.aclweb.org/anthology/2018.acl-main.521779f/</a>	Data Augmentation for Spoken Language Understanding via Joint Variational Generation	Previous deep learning based state-of-the-art scene text detection methods can be roughly classified into two categories. The first category treats scene text as a type of general objects and follows general object detection paradigms to localize scene text by regressing the text box locations, but troubled by the arbitrary-orientation and large aspect ratios of scene text. The second one segments text regions directly, but mostly needs complex post processing. In this paper, we present a method that combines the ideas of the two types of methods while avoiding their shortcomings. We propose to detect scene text by localizing corner points of text bounding boxes and segmenting text regions in relative positions. In inference stage, candidate boxes are generated by sampling and grouping corner points, which are further sorted by segmentation mass and suppressed by NMS. Compared with previous methods, our method can handle long and oriented text naturally and doesn't need complex post processing. The experiments on ICDAR2015, ICDAR2018, MSRA-TD500, MLT and COCO-Text demonstrate that the proposed algorithm achieves better or comparable results in both accuracy and efficiency. Based on VGG16, it achieves an F-measure of 84.3% on ICDAR2015 and 81.5% on MSRA-TD500.	2018	38	71	5	[{"category": "Computer", "JournalArticle": "https://doi.org/10.1145/3206197.3206210"}]	Cost
236	<input type="checkbox"/>	<input checked="" type="checkbox"/>	d2881259016b 356786f <a href="https://www.aclweb.org/anthology/2018.acl-main.356786f/">https://www.aclweb.org/anthology/2018.acl-main.356786f/</a>	Multi-oriented Scene Text Detection via Corner Localization and Region Segmentation	The problem of AMR-to-text generation is to recover a text representing the same meaning as an input AMR graph. The current state-of-the-art method uses a sequence-to-sequence model, leveraging LSTM for encoding a linearized AMR structure. Although being able to model non-local semantic information, a sequence LSTM can lose information from the AMR graph structure, and thus facing challenges with large-scope graphs, which result in long sequences. We introduce a neural graph-to-sequence model, using a novel LSTM structure for directly encoding graph-like semantics. On a standard benchmark, our model shows superior results to existing methods in the literature.	2018	56	291	22	[{"category": "Computer", "JournalArticle": "http://arxiv.org/pdf/1802.08808v1.pdf"}]	Detection
237	<input type="checkbox"/>	<input checked="" type="checkbox"/>	573b44ba4c58f 137476f <a href="https://www.aclweb.org/anthology/2018.acl-main.137476f/">https://www.aclweb.org/anthology/2018.acl-main.137476f/</a>	An anchor-free region proposal network for Faster R-CNN-based text detection approaches	The problem of AMR-to-text generation is to recover a text representing the same meaning as an input AMR graph. The current state-of-the-art method uses a sequence-to-sequence model, leveraging LSTM for encoding a linearized AMR structure. Although being able to model non-local semantic information, a sequence LSTM can lose information from the AMR graph structure, and thus facing challenges with large-scope graphs, which result in long sequences. We introduce a neural graph-to-sequence model, using a novel LSTM structure for directly encoding graph-like semantics. On a standard benchmark, our model shows superior results to existing methods in the literature.	2018	57	100	5	[{"category": "Computer", "JournalArticle": "http://arxiv.org/pdf/1804.09000v1.pdf"}]	Detection
240	<input checked="" type="checkbox"/>	<input type="checkbox"/>	b4812702a1c1c1 251116f <a href="https://www.aclweb.org/anthology/2018.acl-main.251116f/">https://www.aclweb.org/anthology/2018.acl-main.251116f/</a>	A Graph-to-Sequence Model for AMR-to-Text Generation	The problem of AMR-to-text generation is to recover a text representing the same meaning as an input AMR graph. The current state-of-the-art method uses a sequence-to-sequence model, leveraging LSTM for encoding a linearized AMR structure. Although being able to model non-local semantic information, a sequence LSTM can lose information from the AMR graph structure, and thus facing challenges with large-scope graphs, which result in long sequences. We introduce a neural graph-to-sequence model, using a novel LSTM structure for directly encoding graph-like semantics. On a standard benchmark, our model shows superior results to existing methods in the literature.	2018	34	223	29	[{"category": "Computer", "JournalArticle": "https://www.aclweb.org/anthology/2018.acl-main.251116f/"}]	Text Generation (None)
241	<input type="checkbox"/>	<input checked="" type="checkbox"/>	b62a0510a810f 125550f <a href="https://www.aclweb.org/anthology/2018.acl-main.125550f/">https://www.aclweb.org/anthology/2018.acl-main.125550f/</a>	No Longer Lost in Translation: Evidence that Google Translate Works for Comparative Bag-of-Words Text Applications	Automated text analysis allows researchers to analyze large quantities of text. Yet, comparative researchers are presented with a big challenge: across countries people speak different languages. To address this issue, some analysts have suggested using Google Translate to convert all texts into English before starting the analysis (Lucas et al. 2015). But in doing so, do we get lost in translation? This paper evaluates the usefulness of machine translation for bag-of-words models—such as topic models. We use the europarl dataset and compare term-document matrices (TDMs) as well as topic model results from gold standard translated texts and machine-translated text. We evaluate results at both the document and the corpus level. We first find TDMs for both text corpora to be highly similar, with minor differences across languages. What is more, we find considerable overlap in the set of features generated from human-translated and machine-translated texts. With regard to LDA topic models, we find word-topic prevalence and topical context to be highly similar with again only small differences across languages. We conclude that Google Translate is a useful tool for comparative researchers when using bag-of-words text applications.	2018	28	131	2	[{"category": "Computer Science", "source": "https://www.cambridge.org/core"}]	Machine   Automated   (Machine Generated Text (None))
242	<input checked="" type="checkbox"/>	<input type="checkbox"/>	5a5a1076c3a3d 127372f <a href="https://www.aclweb.org/anthology/2018.acl-main.127372f/">https://www.aclweb.org/anthology/2018.acl-main.127372f/</a>	A Deep Generative Framework for Paraphrase Generation	Paraphrase generation is an important problem in NLP, especially in question answering, information retrieval, information extraction, conversation systems, to name a few. In this paper, we address the problem of generating paraphrases automatically. Our proposed method is based on a combination of deep generative models (VAE) with sequence-to-sequence models (LSTM) to generate paraphrases, given an input sentence. Traditional VAEs when combined with recurrent neural networks can generate text but they are not suitable for paraphrase generation for a given text. We address this problem by conditioning the both, encoder and decoder sides of VAE, on the original sentence, so that it can generate the given sentence's paraphrases. Unlike most existing models, our model is simple, modular and can generate multiple paraphrases for a given sentence. Quantitative evaluation on a benchmark paraphrase dataset shows that the proposed method outperforms the state-of-the-art methods by a significant margin, whereas qualitative human evaluation indicate that the generated paraphrases are well-formed, grammatically correct, and are relevant to the input sentence. Furthermore, we evaluate our method on a new released question paraphrase dataset, and establish a new baseline for future research.	2017	36	216	41	[{"category": "Computer", "JournalArticle": "https://doi.org/10.1145/3206197.3206210"}]	Paraphrase   NLP
243	<input checked="" type="checkbox"/>	<input type="checkbox"/>	bcdad001c395ef 756214f <a href="https://www.aclweb.org/anthology/2018.acl-main.756214f/">https://www.aclweb.org/anthology/2018.acl-main.756214f/</a>	Machine Comprehension by Text-to-Text Neural Question Generation	We propose a recurrent neural model that generates natural language questions from documents, conditioned on answers. We show how to train the model using a combination of supervised and reinforcement learning. After teacher forcing for standard maximum likelihood training, we fine-tune the model using policy gradient techniques to maximize several rewards that measure question quality. Most notably, one of these rewards is the performance of a question-answering system. We motivate question generation as a means to improve the performance of question answering systems. Our model is trained and evaluated on the recent question-answering dataset SQuAD.	2017	52	167	13	[{"category": "Computer", "JournalArticle": "https://www.aclweb.org/anthology/2018.acl-main.756214f/"}]	Machine   Paraphrase   Bias   Controllable   Creative   Cost
244	<input checked="" type="checkbox"/>	<input type="checkbox"/>	563a451d3ac4c3 521184f <a href="https://www.aclweb.org/anthology/2018.acl-main.521184f/">https://www.aclweb.org/anthology/2018.acl-main.521184f/</a>	Learning Context-Sensitive Convolutional Filters for Text Processing	Convolutional neural networks (CNNs) have recently emerged as a popular building block for natural language processing (NLP). Despite their success, most existing CNN models employed in NLP share the same learned (and static) set of filters for all input sentences. In this paper, we consider an approach of using a small meta-network to learn context-sensitive convolutional filters for text processing. The role of meta-network is to abstract the contextual information of a sentence or document into a set of input-sensitive filters. We further generalize this framework to model sentence pairs. When a bidirectional filter generation mechanism is introduced to encapsulate co-dependent sentence representations, our benchmarks on four different tasks, including ontology classification, sentiment analysis, answer sentence selection, and paraphrase identification, our proposed model, a modified CNN with context-sensitive filters, consistently outperforms the standard CNN and attention-based CNN baselines. By visualizing the learned context-sensitive filters, we further validate and rationalize the effectiveness of proposed framework.	2017	43	19	2	[{"category": "Computer", "JournalArticle": "https://www.aclweb.org/anthology/2018.acl-main.521184f/"}]	Paraphrase
245	<input checked="" type="checkbox"/>	<input type="checkbox"/>	a6881c7b71c0f83 64553f <a href="https://www.aclweb.org/anthology/2018.acl-main.64553f/">https://www.aclweb.org/anthology/2018.acl-main.64553f/</a>	Identifying computer-generated text using statistical analysis	Computer-based automatically generated text is used in various applications (e.g., text summarization, machine translation) and has come to play an important role in daily life. However, computer-generated text may produce confusing information due to translation errors and inappropriate wording caused by faulty language processing, which could be a critical issue in presidential elections and product advertisements. Previous methods for detecting computer-generated text typically estimate text fluency, but this may not be useful in the near future due to the development of neural-network-based natural language generation that produces wording close to human-crafted writing. A different approach to detecting computer-generated text is thus needed. We hypothesize that human-crafted wording is more consistent than that of a computer. For instance, Zipf's law states that the most frequent word in human-written text has approximately twice the frequency of the second most frequent word, nearly three times that of the third most frequent word, and so on. We found that this is not true in the case of computer-generated text. We hence propose a method to identify computer-generated text on the basis of statistics. First, the word distribution frequencies are compared with the corresponding Zipfian distributions to extract the frequency features. Next, complex phrase features are extracted because human-generated text contains more complex phrases than computer-generated text. Finally, the higher consistency of the human-generated text is quantified at both the sentence level using phrase verbs and at the paragraph level using word co-occurrence relationships, which are integrated into consistency features. The combination of the frequencies, the complex phrases, and the consistency features was evaluated for 100 English books written originally in English and 100 English books translated from Finnish. The results show that our method achieves better performance (accuracy = 98.0%, equal error rate = 2.9%) compared with the most suitable method for books using parsing tree feature extraction. Evaluation using two other languages (French and Dutch) showed similar results. The proposed method thus works consistently in various languages.	2017	15	15	2	[{"category": "Computer", "JournalArticle": "https://www.purp.ac.uk/uk/"}]	Paraphrase   NLP   Natural Language Generation   Evaluation Metrics   Bi
247	<input checked="" type="checkbox"/>	<input type="checkbox"/>	9c3472849c20a 217212f <a href="https://www.aclweb.org/anthology/2018.acl-main.217212f/">https://www.aclweb.org/anthology/2018.acl-main.217212f/</a>	Learning to Ask: Neural Question Generation for Reading Comprehension	We study automatic question generation for sentences from text passages in reading comprehension. We introduce an attention-based sequence learning model for the task and investigate the effect of encoding sentence via paragraph-level information. In contrast to all previous work, our model does not rely on hand-crafted rules or a sophisticated NLP pipeline. It is instead trainable end-to-end via sequence-to-sequence learning. Automatic evaluation results show that our system significantly outperforms the state-of-the-art rule-based system. In human evaluations, questions generated by our system are rated as being more natural (i.e., grammatically, fluency) and as more difficult to answer (in terms of syntactic and logical divergence from the original text and reasoning needed to answer).	2017	39	523	133	[{"category": "Computer", "JournalArticle": "https://www.aclweb.org/anthology/2018.acl-main.217212f/"}]	NLP
248	<input checked="" type="checkbox"/>	<input type="checkbox"/>	d13bb17f7870f 160463f <a href="https://www.aclweb.org/anthology/2018.acl-main.160463f/">https://www.aclweb.org/anthology/2018.acl-main.160463f/</a>	Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation	This paper surveys the current state of the art in Natural Language Generation (NLG), defined as the task of generating text or speech from non-linguistic input. A survey of NLG is timely in view of the changes that the field has undergone over the past two decades, especially in relation to new (statistical data-driven) methods, as well as new applications of NLG technology. This survey therefore aims to (a) give an up-to-date synthesis of research on the core tasks and NLG and the architectures adopted in which such tasks are organized; (b) highlight a number of recent research topics that have arisen partly as a result of growing interest in NLG; and (c) draw attention to the challenges in NLG evaluation, relating them to similar challenges faced in other areas of NLP, with an emphasis on different evaluation methods and the relationships between them.	2017	556	634	53	[{"category": "Computer", "JournalArticle": "https://doi.org/10.1145/3206197.3206210"}]	NLP   Natural Language Generation
249	<input checked="" type="checkbox"/>	<input type="checkbox"/>	8284f496b987f 769991f <a href="https://www.aclweb.org/anthology/2018.acl-main.769991f/">https://www.aclweb.org/anthology/2018.acl-main.769991f/</a>	Multilingual CALL Framework for Automatic Language Exercise Generation from Free Text	This paper describes a web-based application to design and answer exercises for language learning. It is available in Basque, Spanish, English, and French. Based on open-source Natural Language Processing (NLP) technology such as word embedding models and word sense disambiguation, the application enables users to automatically create easily and in real time three types of exercises, namely: Fill-in-the-Gaps, Multiple Choice, and Shuffled Sentences questionnaires. These are generated from texts of the user's own choice, so they can train their language skills with content of their particular interest.	2017	6	8	2	[{"category": "Computer", "JournalArticle": "https://www.aclweb.org/anthology/2018.acl-main.769991f/"}]	NLP
250	<input checked="" type="checkbox"/>	<input type="checkbox"/>	4a67539e254dc 110304f <a href="https://www.aclweb.org/anthology/2018.acl-main.110304f/">https://www.aclweb.org/anthology/2018.acl-main.110304f/</a>	Neural Response Generation via GAN with an Approximate Embedding Layer	This paper presents a Generative Adversarial Network (GAN) to model word-to-sentence and sentence-to-sentence generation. The GAN is trained to generate sentences that are as similar as possible to human-generated sentences, but with a discriminative classifier that measures the differences between human-produced responses and machine-generated ones. In addition, the proposed method introduces an approximate embedding layer to solve the non-differentiable problem caused by the sampling in the Softmax operation in the GAN. The GAN setup provides an effective way to avoid noninformative responses (aka a "safe responses"), which are frequently observed in traditional neural response generators. The experimental results show that the proposed approach significantly outperforms existing neural response generation models in diversity metrics, with slight increases in relevance scores as well, when evaluated on both a Mandarin corpus and an English corpus.	2017	26	80	7	[{"category": "Computer", "JournalArticle": "https://www.aclweb.org/anthology/2018.acl-main.110304f/"}]	Natural Language Generation   Language Model
251	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2ce08ff8bfadef 368994f <a href="https://www.aclweb.org/anthology/2018.acl-main.368994f/">https://www.aclweb.org/anthology/2018.acl-main.368994f/</a>	Automatic Generation of Natural Language Explanations	An interesting challenge for explainable recommender systems is to provide successful interpretation of recommendations using structured sentences. It is well known that user-generated reviews have strong influence on the user's decision. Recent techniques exploit user reviews to generate natural language explanations. In this paper, we propose a character-level attention-enhanced long short-term memory model to generate natural language explanations. We integrate explanations with the neural network using two real-world review datasets. The generated text is present readable and similar to a real user's writing, due to the ability of reproducing negation, misspellings, and domain-specific vocabulary.	2017	30	79	2	[{"category": "Computer", "JournalArticle": "http://arxiv.org/pdf/1707.01521v1.pdf"}]	Natural Language Generation
253	<input checked="" type="checkbox"/>	<input type="checkbox"/>	6ae02c05e2b0e 189994f <a href="https://www.aclweb.org/anthology/2018.acl-main.189994f/">https://www.aclweb.org/anthology/2018.acl-main.189994f/</a>	Affect-LM: A Neural Language Model for Customizable Affective Text Generation	Human verbal communication includes affective messages which are conveyed through use of emotionally colored words. There has been a lot of research effort in this direction but the problem of integrating state-of-the-art neural language models with affective information remains an area ripe for exploration. In this paper, we propose an extension to an LSTM (Long Short-Term Memory) language model for generation of conversational text, conditioned on affect categories. Our proposed model, Affect-LM enables us to customize the degree of emotional content in generated sentences through an additional design parameter. Perceptual studies conducted using Amazon Mechanical Turk show that Affect-LM can generate naturally looking emotional sentences without sacrificing grammatical correctness. Affect-LM also learns affect-discriminative word representations, and perplexity experiments show that additional affective information in conversational text can improve language model prediction.	2017	29	163	15	[{"category": "Computer", "JournalArticle": "https://www.aclweb.org/anthology/2018.acl-main.189994f/"}]	Language Model
257	<input type="checkbox"/>	<input checked="" type="checkbox"/>	0651918350ef 648400f <a href="https://www.aclweb.org/anthology/2018.acl-main.648400f/">https://www.aclweb.org/anthology/2018.acl-main.648400f/</a>	Style Transfer in Text: Exploration and Evaluation	The ability to transfer styles of texts or images, is an important measurement of the advancement of artificial intelligence (AI). However, the progress in language style transfer is lagged behind other domains, such as computer vision, mainly because of lack of parallel data and reliable evaluation metrics. In response to the challenge of lacking parallel data, we explore learning style transfer from non-parallel data. We propose two models to achieve this goal. The key idea behind the proposed models is to learn separate content representations and style representations using adversarial networks. Considering the problem of lacking principle evaluation metrics, we propose two novel evaluation metrics that measure two aspects of style transfer: transfer strength and content preservation. We benchmark our models and the evaluation metrics on two style transfer tasks: paper-news title transfer, and positive-negative review transfer. Results show that the proposed content preservation metric is highly correlate to human judgments, and the proposed models are able to generate sentences with similar content preservation score but higher style transfer strength comparing to auto-encoder.	2017	33	428	104	[{"category": "Computer", "JournalArticle": "https://doi.org/10.1145/3206197.3206210"}]	Evaluation Metrics



