Multimedia Tools and Applications (2023) 82:43089–43144
https://doi.org/10.1007/s11042-023-15224-0

A Systematic survey on automated text generation tools and techniques: application, evaluation, and challenges

Rupali Goyal[1] · Parteek Kumar[1] · V. P. Singh[1]

Received: 5 April 2021 / Revised: 2 March 2023 / Accepted: 30 March 2023 /
Published online: 19 April 2023

Abstract

Automatic text generation is the generation of natural language text by machines. Enabling machines to generate readable and coherent text is one of the most vital yet challenging tasks. Traditionally, text generation has been implemented either by using production rules of a predefined grammar or performing statistical analysis of existing human-written texts to predict sequences of words. Recently a paradigm change has emerged in text generation, induced by technological advancements, including deep learning methods and pre-trained transformers. However, many open challenges in text generation need to be addressed, including the generation of fluent, coherent, diverse, controllable, and consistent human-like text. This survey aims to provide a comprehensive overview of current advancements in automated text generation and introduce the topic to researchers by offering pointers and synthesis to pertinent studies. This paper studied the relevant twelve years of articles from 2011 onwards in the field of text generation and observed a total of 146 prime studies relevant to the objective of this survey that has been thoroughly reviewed and discussed. It covers core text generation applications, including text summarization, question–answer generation, story generation, machine translation, dialogue response generation, paraphrase generation, and image/video captioning. The most commonly used datasets for text generation and existing tools with their application domain have also been mentioned. Various text decoding and optimization methods have been provided with their strengths and weaknesses. For evaluating the effectiveness of the generated text, automatic evaluation metrices have been discussed. Finally, the article discusses the main challenges and notable future directions in the field of automated text generation for potential researchers.

Keywords Automatic Text Generation (ATG) · Text Generation Methods · Text Generation Tools · Application-Specific Standard Datasets · Text Decoding Techniques · Automatic Text Evaluation Methods

✉ Rupali Goyal
rrupali20_phd17@thapar.edu

[1] Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala, Punjab, India

1 Introduction

Text Generation is a sub-discipline of Natural Language Processing used to fulfill specific communicative requirements by automatically generating natural language texts that leverage computational linguistics and artificial intelligence abilities [168]. Text generation has many real-world [51] applications depending on the input (data, text, or multimodal); however, the output is always a natural language text. Thus, based on the type of input, the text generation has been categorized mainly into three categories: text-to-text generation (T2T), data-to-text generation (D2T), and multimodality-to-text generation (M2T), as shown in Fig. 1. The text-to-text generation tasks take existing text as input and automatically generate a new, coherent text as output. For T2T generation, the most common applications include summarising the input document [13, 101], generating questions and answers from a text document [4, 48, 193], translating a sentence from one language to another [1, 11], creating or completing a story outline [66, 171, 211]. The data-to-text generation tasks automatically generate text from numerical or structured data such as key-value lists and tables. For D2T generation, the example applications include reports generation from numerical data [148, 154], and generating text from the meaning representations to represent the meaning of natural language [113, 180]. The multimodality-to-text generation tasks transfer the semantics in multimodal input, such as videos or images, into natural language texts. For M2T generation, the example applications include generating captions from images or videos [71, 126], video summarization, and visual storytelling.

The research on text generation has a long history. The earliest text generation systems used template and rule based methods to capture linguistic knowledge of vocabulary, syntax, and grammar. The next-generation models encode the dependency between vocabulary and context in conditional probability. These methods also couple with template-based methods for text generation. Then with the development of deep learning technologies, neural-based models gradually occupy a dominant position. Deep Learning belongs to a class of machine learning algorithms that identifies patterns in text and identifies features that assist in solving several text generation tasks [97]. The capacity of deep neural networks to learn representations of varying degrees of complexity has aided in achieving state-of-the-art performances across different text generation tasks, such as machine translation, text summarization, storytelling, and dialogue systems [62]. The availability and accessibility of a vast number of corpora and massive computational resources are other factors supporting deep learning growth. Most recently, the pre-trained text generation models based on the Transformer architecture have the ability to better capture the linguistic knowledge of vocabulary, syntax, and grammar. However, while these models generate fluent and grammatical text, they are prone to making factual errors that contradict the input text. Generating fluent, informative, well-structured, and coherent text is pivotal

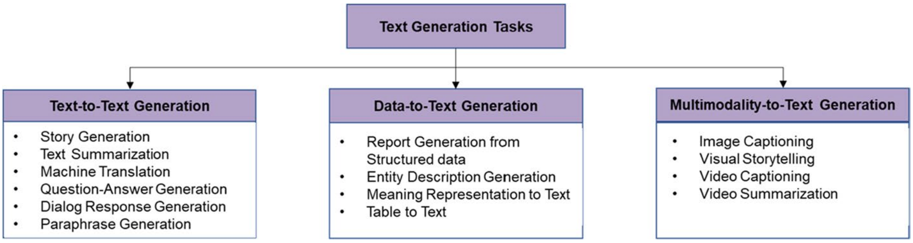


Fig. 1 Categories of common text generation tasks

for many text generation tasks. It takes significant effort by humans to generate text that is consistent and model long-term dependencies. And it is an equally challenging to do it automatically because of the discrete nature of textual data. Traditional template-based methods generate reliable texts but lack diversity, fluency, and informativeness. The deep learning models generate fluent and informative texts but are limited by the faithfulness and controllability of the neural-based models. And the transformer based models generate fluent, informative, and controllable text, but the inconsistency with the input information persists. The motivation for conducting this survey and the primary contributions of the paper are presented in the following subsection.

1.1 Motivation and our contribution

Automated text generation has been gaining attention with the advances in deep learning. In the last ten years, the text generation field has evolved significantly. Numerous appealing surveys [51, 59, 120, 123, 173, 216] have been introduced, summarizing the work done in this field. However, there is no proper survey on ATG in terms of prominent benchmark datasets, real-world tools, decoding methods, evaluation metrices, and challenges of automated text generation applications. This motivates us to perform a Systematic Literature Review for automatic text generation using deep learning techniques. Another motivation is the gaining interest in this research field of text generation. The analysis of the twelve-year articles published in this field has consistently increased, indicating that automated text generation is gaining interest each year. Keeping this in mind, this paper studies all the relevant articles from 2011 onwards to find methods for automated generation of text in different application domains, different existing tools and datasets used with their application domain, and evaluation metrices for evaluating the effectiveness of the generated text. The purpose of this survey is to provide a comprehensive overview of current advancements in automated text generation and introduce the topic to researchers by providing pointers and synthesis to pertinent studies.

The main contribution of this paper comprises the following key points:

I. This survey provides an up-to-date synthesis of automated text generation along with its core applications, including text summarization, question–answer generation, dialogue generation, machine translation, story generation, paraphrasing, and image captioning, and the key techniques behind them.

II. A comprehensive outline of techniques and methods employed to generate text automatically, including traditional statistical methods, deep learning, and pre-trained transformer based models, has been discussed.

III. This paper enlists standard datasets required to train, test, and validate the text generation models for the automated generation of fluent and coherent texts.

IV. The text decoding strategies and optimization techniques significantly impact the quality of the generated text. This paper discussed these decoding techniques and optimization methods with their strengths and limitations.

V. In this article, real-time task-specific tools for automated text generation have also been provided with their features and URLs.

VI. For the effectiveness of the generated text, various metrices/approaches have also been summarized to evaluate text generation models automatically that depict different text attributes such as fluency, grammaticality, coherence, readability, and diversity.

VII. This paper also identifies various challenges of text generation applications, including the generation of human-like text that is fluent, diverse, controllable, and consistent. This survey also outlines potential future research directions in the area of automated text generation.

1.2 Comparison with other surveys

There have been related attempts and literature surveys on automated text generation. This subsection overviews such attempts and highlights the contrast between existing and current surveys with their strengths and limitations. For example, Gatt et al. [51] surveyed natural language generation emphasizing image-to-text tasks. Liu et al. [120] reviewed deep learning architectures and limited text generation applications. Xie [216] describes techniques for training and dealing with natural language generation models using neural networks. Santhanam et al. Lu et al. [123] surveyed only neural text generation models. [173] review language generation with a focus on dialogue systems. Garbacea et al. [59] present an overview of natural language generation methods, tasks, and assessments. Yu et al. [232] reviewed knowledge-enhanced text generation. However, there are many existing surveys on text generation but are limited in terms of standard datasets, existing real-time tools, optimization methods, evaluation metrices, and challenges of automated text generation applications. This motivates us to perform a Systematic Literature Review on automatic text generation. This survey captures the comprehensive study and up-to-date review of current advancements in the field of text generation and also studies various methods for automated generation of text in different application domains, different existing tools, and datasets used with their application domain, text decoding, and optimization techniques, and evaluation metrices for evaluating the effectiveness of the generated text. Table 1 summarizes these aspects in comparison to the surveys mentioned above reports and contributions in literature.

The rest of this paper is organized as follows. A detailed review strategy and various research questions with significance are provided in Sect. 2. This section also mentions the search criteria and the research parameters for writing this survey paper. The extraction of studies and discussion is presented in Sects. 3 to 10. In Sect. 3, the core applications of text generation are reviewed. Section 4 mentions the methods and techniques employed for

Table 1 Comparative analysis of the proposed survey with existing surveys

Authors [Ref.]	1	2	3	4	5	6	7
Gatt et al. [51]	✓	✓	⋆	⋆	✗	⋆	✗
Liu et al. [120]	⋆	✓	✗	✗	✗	✗	✗
Xie [216]	✗	✓	✗	✗	✗	✗	✗
Lu et al. [123]	✗	✓	✗	✗	✗	✗	✗
Santhanam et al. [173]	✓	✓	✗	⋆	⋆	✗	✗
Garbacea et al. [59]	✓	✓	✗	✗	✗	✓	✗
Yu et al. [233]	✓	✓	✗	✗	✗	✗	✗
Our Survey	✓	✓	✓	✓	✓	✓	✓

where 1: Text Generation Applications, 2: Text Generation Approaches, 3: Real-Time Tools, 4: Standard Datasets, 5: Text Decoding and Optimization Methods, 6: Automatic Evaluation Metrices, 7: Text Generation Application Challenges; and✓—Detailed Study, ⋆—Limited Consideration, ✗—No Discussion.

generating text. Section 5 enlists application-specific standard datasets required to train, test, and validate the models. Text decoding and optimization techniques significantly impact the generated text, which are mentioned in Sect. 6. There are many real-time task-specific tools for text generation, which are provided in Sect. 7, along with their access URLs. Section 8 reviewed the approaches to evaluate the effectiveness of the generated text. Various open challenges to automate the text generation task is mentioned in Sect. 9. Finally, Sect. 10 concludes this paper and outlines potential directions for future research.

2 Research methodology

The research methodology is a process of systematically researching. It includes an empirical analysis of all concepts relevant to the field of research. Generally, it includes the concepts of phases, models, and quantitative as well as qualitative techniques. This paper follows the review process suggested by Kitchenham and Charters [92], which includes planning, conducting, and reporting the review, as shown in Fig. 2.

2.1 Planning review

The planning process included identifying the need for a Systematic Literature Review (SLR) and concluding with the formulation and validation of the review procedure. A systematic review is needed to identify, compare, and classify the existing text generation work. The studies published on text generation in the last twelve years are observed, but none are robust. This paper comprehensively analyzes emerging models, methods, tools, and deep-learning-based text generation application techniques to identify and compare them systematically.

The research questions (RQs) are prepared to facilitate the review process to be more focused, clear, and consistent. Eight research questions (RQ 1 to RQ 8) have been framed, which help to perform SLR. The research questions and their significance in this literature review are mentioned in Table 2.

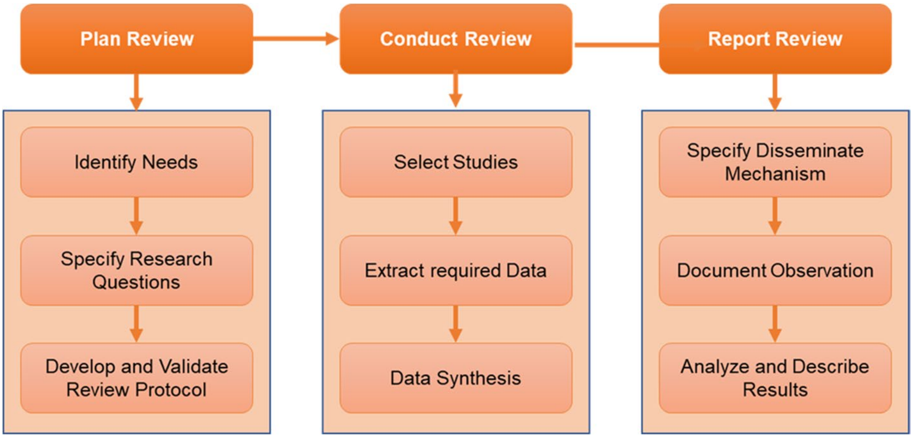


Fig. 2 Overview of research methodology

Table 2 Research Questions and their Significance

Research questions	Significance
RQ 1: How the automated text generation study evolves with advancements in deep learning?	To identify the rate of adaptation of automated text generation studies and the yearly analysis of research in text generation with advancements in the technology models
RQ 2: What are the main text generation applications that have arisen with advancements in the field?	To classify text generation tasks and to describe various text generation applications based on input and usage
RQ 3: What are the various approaches and associated architectures in the field of text generation?	To identify and analyze various methods and techniques for automatic text generation. This helps to identify strengths and limitations associated with the approaches
RQ 4: What are the available datasets adopted in which the stated applications are organized?	To identify the available standard datasets according to application to test, train and validate the model
RQ 5: What are various text decoding and optimization techniques are used to generate fluent text automatically?	To identify the task specific text decoding strategies and optimizers for automated text generation
RQ 6: What real-time tools are available for automatic text generation tasks?	To have a perception of the usage of text generation applications in the real world and to analyze the strengths and weaknesses of the existing text generation tools
RQ 7: Which metrics or indicators are used to evaluate the generated text?	To evaluate the effectiveness of the generated text, identification of the evaluation metrices used for text generation applications
RQ 8: What challenges are faced in automated text generation tasks?	To identify the issues related to the applications of automatic text generation tasks

2.2 Conducting review

This phase involves selecting studies, extracting required resources, and synthesizing knowledge. This SLR includes research papers from different publications and the various online electronic databases selected, such as IEEE Xplore, ACM digital library, Science Direct-Elsevier, Springer link, Web of Science, and Wiley online library. The search string includes keywords: “Text Generation” OR “Natural Language Generation” OR “Text Generation using Deep Learning” OR “Neural Text Generation” OR “Neural Language Generation” AND “Applications” OR “Text Generation Applications.” The sources contain documents of several types, such as book chapters, research articles, reviews, and proceeding papers, published in the last twelve years, i.e., from 2011 to 2022. It discusses the research papers from journals, magazines, conferences, workshops, and symposiums. The studies were explored and based on inclusion–exclusion criteria, and a total of 146 research papers were obtained, as shown in Fig. 3.

These 146 research papers from the ‘2011–2022’ time frame are thoroughly reviewed and discussed in this survey paper. The number of extracted research papers based on their year of publication is shown in Fig. 4. It can be observed that before 2011 there was limited work in the research area of text generation using deep neural networks. And there has been a gradual increase in the number of research papers from 2011 onwards, showing growth in the field of automated text generation with the developments in deep neural models.

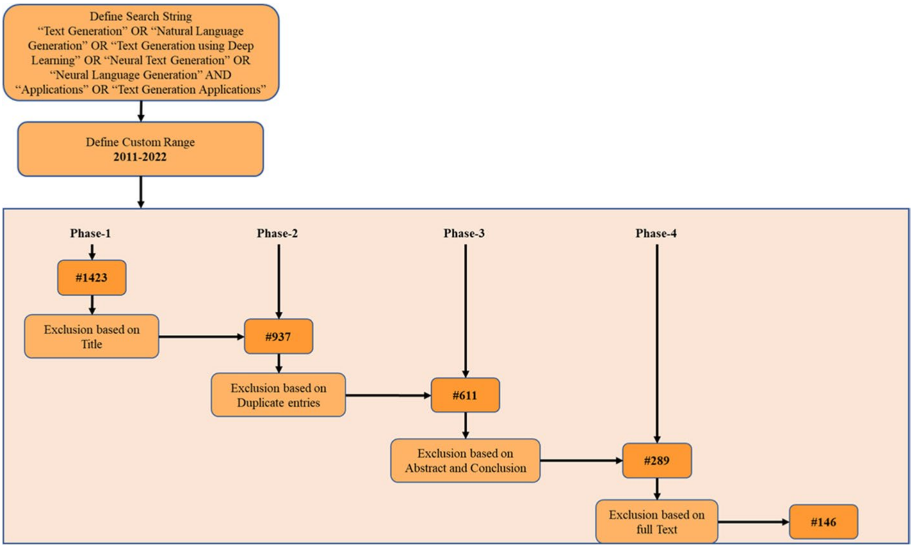

Define Search String
"Text Generation" OR "Natural Language Generation" OR "Text Generation using Deep Learning" OR "Neural Text Generation" OR "Neural Language Generation" AND "Applications" OR "Text Generation Applications"
Define Custom Range
2011-2022
Phase-1
Phase-2
Phase-3
Phase-4
#1423
Exclusion based on Title
#937
Exclusion based on Duplicate entries
#611
Exclusion based on Abstract and Conclusion
#289
Exclusion based on full Text
#146

Fig. 3 Inclusion/Exclusion technique used in the systematic review

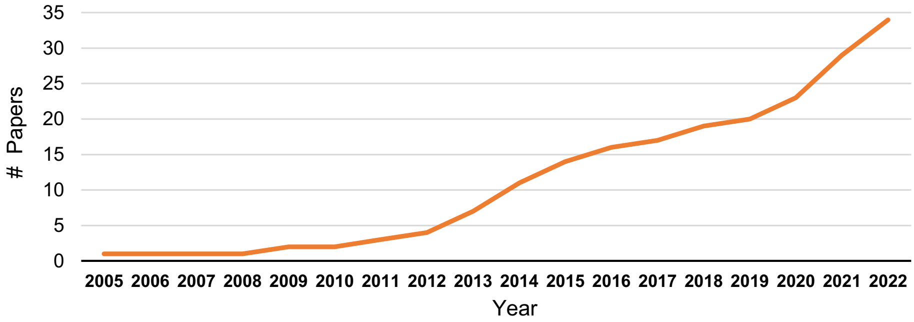


Fig. 4 Yearly analysis of the papers in the text generation research area with development in technology

Discussion on research question 1

This yearly analysis of papers helps answer RQ 1- "How the automated text generation study evolves with advancements in deep learning." It has been observed that limited work has been done on text generation before 2011, and there is a gradual increase in the number of research papers from the year 2011 onwards. This growth results from advancements in text generation methods, from traditional rule-based methods to deep neural networks and pre-trained transformer models. The traditional template or rule based methods were used for text generation usually before 2013, but these rules/templates are difficult to design and are very time consuming. These shortcomings of traditional approaches were overcome in the years with the developments in deep learning methods. The research in the field of text generation increases gradually thereafter. The availability of powerful deep neural models and computationally intensive architecture results in the incredible adoption of a variety of text generation applications, including text summarization, machine translation,

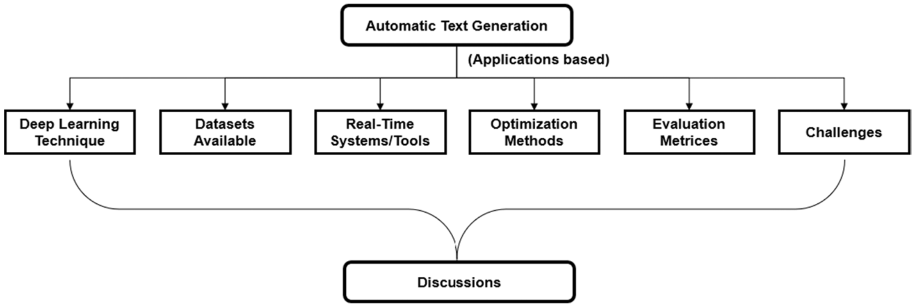


Fig. 5 Reviewing parameters for each text generation application

creative applications such as story generation, and dialogue generation. Now, with pre-trained transformers models, the automated text generation work has immensely accelerated. Many sectors have started using automated text to improve user experience as the recent advancements in technology is capable of generating human-like texts. The content generated by the automated tools is fast and cheap. The analysis of the published articles in this field indicates consistent growth and adaptation to the research area of automated text generation from 2011. Keeping this in mind, this paper studies all the relevant articles from 2011 onwards to find methods for automated generation of text in different application domains, different existing tools and datasets used to achieve the task, text decoding and optimization techniques, and evaluation metrices for the effectiveness of the generated text.

2.3 Reporting review

For reporting the review, this phase provides the research parameters. The research parameters that have been followed for SLR include the core text generation applications. For each text generation application, the deep learning technique used, standard datasets, existing real-time tools for that application, optimization methods, evaluation metrics, and challenges for each application are presented. The different parameters used in this paper are shown in Fig. 5. The extraction of studies and discussions is presented in this survey.

These research parameters are discussed in-depth in the following sections, and the results or analysis of these parameters are also reviewed and mentioned in Sects. 3 to 10.

3 Applications of automated text generation

The field of artificial intelligence has developed techniques that generate text automatically in seconds. Automatic text generation is one such application that is the need for the hour. Many applications of text generation are crucial and very significant for smart systems and enable better communication between humans and machines, for example, machine translation, summarization, and simplification of long or complex texts, grammar, and spelling correction, generating peer reviews for scientific papers, questionnaire generation, automatic documentation systems for large software, question–answer generation, business letter writing, chatbots and much more. The core text generation applications are shown in Fig. 6, and the details about these applications are discussed below.

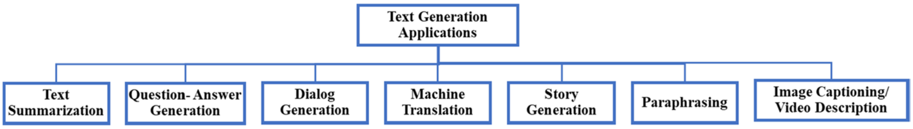

Text Generation Applications
Text Summarization
Question- Answer Generation
Dialog Generation
Machine Translation
Story Generation
Paraphrasing
Image Captioning/ Video Description

Fig. 6 Text Generation Applications

3.1 Text summarization

With each day, enormous amounts of data from diversified sources have evolved. This massive volume of data incorporates crucial facts, information, and knowledge that needs to be effectively summarized to be helpful. Thus, automatic text summarization came into the picture to tackle the problem of information overloading [142]. A text summarization method generates an abbreviated version of a document by filtering the significant information from the original document [57]. A strong summary consists of all aspects, such as coverage, non-redundancy, cohesion, relevancy, and readability, in addition to relevant key points [145]. There are two prominent types of summarization techniques. First, extractive summarization techniques form summaries by copying parts of the input sentences [134], and second, the abstractive summarization technique [5, 132] generates a summary by including words and phrases not present in the source [135]. Nallapati et al. [135] propose recurrent neural network-based encoder-decoder models for abstractive text summarization. In follow-up work [134], extractive summarization techniques using recurrent neural networks are presented. Rush et al. [172] propose an attention-based network for the abstractive summarization of sentences, and Cheng et al. [28] proposed an attentional encoder-decoder for extractive single-document summarization. See et al. [177] used a pointer generator network for abstractive summarization. Paulus et al. [147] use the reinforcement learning model for abstractive summarization, while others use reinforcement learning for extractive single-document summarization [136, 212]. Mehta et al. [129] use Long Short Term Memory (LSTM) and attention model to summarize scientific papers. Liu et al. [118] focus on multi-document summarization by generating fluent, coherent multi-sentence Wikipedia articles using extractive summarization. Modified BERT transformer [40] for extractive summarization is capable of extracting automatically the features in the internal layers [116] Multi-document summarization using abstractive methods has also been used [15, 239]. Xu et al. [221] propose a multi-task framework with a hybrid of the extractive and abstractive models. Transformer architecture also performs great in many NLG tasks [195]. Tan et al. [20] used a pretrained model, GPT-2 for the summarization task with the idea that the model will start generating a summary based on the delimiter. More recent works leverage pre-trained transformer based networks, such as GPT [162], BART [102], T5 [163], and PEGASUS [240], for summary generation [63, 119, 213].

3.2 Question answer generation

Automatic question generation (QG) aims to generate questions from some form of input, such as raw text or a database, whereas Question Answering (QA) is the task of automatically providing precise responses to questions in the natural language given corresponding document. In the last years, the widespread use of QA-based personal assistants has been observed, including Microsoft's Cortana, Apple's Siri, Samsung's Bixby, Amazon's Alexa, and Google Assistant, which have answered a wide variety of questions. QG systems proposed by [46, 185] automatically generate answer-unaware questions from within the given document, whereas [88, 180, 186] generate answer-agnostic questions. Du et al. [47] initiated a neural question generation model using an attention sequence-to-sequence model [11]; subsequently, [48, 69, 245] also adopted an attention mechanism. Zhao et al. [241] proposed a gated self-attention encoder. Most neural QG models [69, 95, 204, 245] employ the copying mechanism for question generation. Weston et al. [206] proposed the use of a Memory networks model in the system to answer the questions effectively. The Dynamic Memory Networks model [94] overcomes the shortcomings of the memory networks by combining the paradigms of memory networks and attention mechanisms. This work was later extended by Xiong et al. [219] for visual question answering. Other works, including visual question answering [2, 9, 58, 122] have generated natural and engaging questions for an image. [233] have adopted policy gradient methods to diversify the generated question. [40, 99, 225] uses pre-trained models for the question-answering task, and [100] uses transformer-based models to generate answer aware questions. [203, 204] propose a neural model for question generation and answering that jointly asks and answers questions given a document. Most of the earlier work focuses on using a single QA dataset, such as SQuAD [165]. While working on the generation of multi-hop [30], open ended and controllable [23], or cause-effect [183] questions have gained attention, each direction is studied in isolation as it usually requires a separate question–answer dataset. More recent works leverage pre-trained transformer based networks, such as BART [102], T5 [163], and PEGASUS [240], for question generation, which have been successful in many applications [6, 98, 107, 164, 194].

3.3 Dialogue response generation

Dialogue systems or conversational agents are computer programs capable of replying with natural, coherent, meaningful, and engaging responses. A good dialogue model generates dialogues with high human similarity [104]. [131, 181] work on building end-to-end dialog generation systems using neural networks, whereas [178, 191] use hierarchical encoder-decoder to generate responses. [218, 228] use the attention model, while [103, 110] use reinforcement learning, and Li et al. [111] use generative adversarial networks for dialog generation. Niu et al. [141] also use a reinforcement learning model focusing on polite dialogue responses. The use of pre-trained models for conversational agents is also observed. [10, 242] use embeddings; [41, 209] use transformers for response generation [98, 107]. These conversational models have enabled robots to interact with humans in natural languages; for example, Window’s Cortana, Google’s assistant, Apple’s Siri, and Amazon’s Alexa are the software and devices that follow Dialogue systems. [76, 153] proposes dialogue generation with recognition of emotions, and [56, 167] also generates empathetic dialogues. [181, 243, 244] generates single-turn dialogue responses while [159, 167, 235] generates multi-turn dialogues. [184] has used text style transfer and GPT for the creation of a dialog generation system over gender-specific, emotion-specific, and sentiment-specific dialogue datasets.

3.4 Neural machine translation

The data accessible to everyone is a challenge because language becomes the barrier, and machine translation came into the picture to overcome this. Machine translation is the task of automatic translation of written text from one natural language into another. Neural machine translation (NMT) uses neural nets to transform the source sentence into the target sentence [29, 82, 188]. [11] introduced attention mechanism in NMT models, which was later extended [124]. Luong et al. [124] have used a unidirectional recurrent neural network model, while [214] used bidirectional recurrent neural networks (BRNN). [3] and [79] does the translation for low-resource language pairs. [29, 31] uses gated recurrent units and achieves better performance on NMT. Tu et al. [192] use a copy and coverage mechanism; Wang et al. [200] use a pre-computed word embedding layer, GlOVe (Global Vectors for Word Representation). Park et al. [144] proposed a mobile device-based sign language translation system. [1] uses an attention-based multi-layer neural network. Transformer architecture [195] also performs well in NMT. [12, 200] proposed deep Transformer models for translation. Transformers like BART [102], BERT [35, 40] and GPT [20] have also been used for the NMT task. More recent works leverage pre-trained transformer based networks for machine translation [54, 64]. Recent works on non-autoregressive neural machine translation [65, 72, 161] improve model efficiency by decoding in parallel as compared to sequential decoding in traditional autoregressive machine translation methods [215].

3.5 Story generation

Automated story generation is the task of automatically identifying a series of actions, events, or words that have been told as a story. Li [106] attempts to automatically generate a story about any domain without prior knowledge. To encode the context, recurrent networks, and convolutional networks successfully model sentences [38, 81]. A fusion mechanism [182] is introduced to support sequence-to-sequence models build dependencies between their input and output. Pawade et al. [149] have implemented a recurrent neural network-based story system to generate a new story based on a series of inputted stories. Vaswani et al. [195] use multi-head attention. [108, 170] use LSTM networks to learn the text hierarchically. Jain et al. [78] chain a series of variable length independent descriptions together into a well-formed comprehensive story. Clark et al. [34] model entities in story generation. Martin et al. [128] present an event-based end-to-end story generation pipeline. Similarly, [68] generates summaries of movies as sequences of events using a recurrent neural network (RNN) and sample event representations. [53, 227] propose a hierarchical story generation framework that first plans a storyline and then generates a story based on the storyline. [151, 208] propose a framework that enables controllable story generation. [7, 189] uses policy gradient deep reinforcement learners to perform an event-to-event task. [26] uses the BERT language model for story plot generation.

3.6 Paraphrase generation

Texts that convey a similar meaning but different expressions are referred to as paraphrases. Paraphrase generation refers to an activity in which, given a sentence, the system creates paraphrases of it. Bowman et al. [18] use a variational autoencoder (VAE) to model holistic properties of sentences such as style, topic, and other features. Gupta et al. [67] use VAE-LSTM to generate more diverse paraphrases. Prakash et al. [155] employ a stacked residual LSTM network in the Sequence-to-sequence model. [105, 166] propose deep reinforcement learning (RL) to guide Sequence-to-sequence training. Cao et al. [22] utilize a novel sequence-to-sequence model to join copying, and restricted generation [237] tackle a comparable task with the Sequence-to-sequence model coupled with deep reinforcement learning. See et al. [177] use a pointer-generator while [125] utilizes an attention layer. Iyyer et al. [77] utilize syntactic information for controllable paraphrase generation. Yang et al. [224] propose an end-to-end conditional generative architecture for generating paraphrases. Qian et al. [160] propose an approach that generates a diverse variety of different paraphrases. [21, 43] tackle the problem of QA-specific paraphrasing while [223] help diversifies the response of chatbot systems. [117] first uses abstract rules and then leverage neural networks to generate paraphrases by refining the transformed sentences.

3.7 Image/Video caption generation

The generation of semantically and syntactically correct description sentences of an image is called image captioning. The recognition of vital objects, their properties, and their relationships in an image is required for image captioning. Kiros [90] propose the initial work for extracting image features with the use of a convolutional neural network (CNN) in generating image captions. Then, with the use of LSTM [90] extended their work [91]. Mao et al. [127] proposed a multimodal recurrent neural network (m-RNN) and [229] used hierarchical recurrent neural networks for generating image descriptions. [201] proposed a deep Bi-LSTM based method for image captions. [80, 220] proposed an attention-based image captioning method. [169, 238] introduced a reinforcement learning-based image captioning method. [37, 179] proposed an image captioning method based on Generative Adversarial Networks (GAN). Vinyals et al. [199] proposed a neural image caption generator method. Donahue et al. [42] propose long-term recurrent convolutional networks that have been processing variable-length inputs. [150, 230] propose an attention-based image captioning model. Some method uses a CNN for image representations and an LSTM for generating image captions. Yao et al. [226] proposed a copying mechanism to generate a description for novel objects. [55, 85] use pre-computed word embedding layers and thus generate better image captions. [202] has proposed a framework that unifies a diverse set of cross-modal and unimodal tasks, including image captioning, and language modelling.

Discussion on research question 2

Text generation research consists of various tasks, topics, or trends. This section helped to answer RQ 2- “What are the main core text generation applications that have arisen with advancements in the field.” It has been observed that there are many real-world applications depending on the input (data, text, or multimodal); however, the output is always a natural language text. Thus, based on the type of input, the text generation has been categorized mainly into three categories: text-to-text generation (T2T), data-to-text generation (D2T), and multimodality-to-text generation (M2T), as discussed earlier. For T2T generation, the most common applications include summarising the input document, generating questions and answers from a text document, translating a sentence from one language to another, and creating or completing a story outline. For D2T generation, the example applications include reports generation from numerical data and generating text from the meaning representations to represent the meaning of natural language. For M2T generation, the example applications include generating captions from images or videos, video summarization, and visual storytelling. The summarized description of the above-mentioned text generation applications is mentioned in Table 3.

Table 3 Summarized description of text generation applications

Application [Ref.]	Description	Input	Output
Machine Translation [11]	Translate a text document from a source language into a target language	A text document in language X (e.g., Punjabi)	Document in language Y (e.g., English)
Text Summarization [135]	Summarize the text document into a concise and shorter text	A long text document like research work	A summary
Headline Generation [121]	Summarize an article into a brief headline	A long document like a news article	A headline
Question Answering [50]	Given a question and a text document, generate the answer to the question	A textual question	A short text as an answer from the document
Question Generation [52]	Generate questions from a textual document, article, or an image	A text document, article, or image	A wide range of questions related to the inputted document
Dialogue Generation [10]	Generate dialogue among agents, e.g., between a human and a robot	A dialogue from the audibles or the first agent	A dialogue from the audibles or the second agent
Story Telling [170]	Given an image or story outline, generate a textual story	A story outline or an image	A story in text form
Smart Reply/email answer suggestion [84]	Given an original message, suggest the most likely responses	An incoming email/message	Most probable responses
Paraphrase Generation [155]	Given a sentence, generate multiple paraphrases	An original sentence	Multiple paraphrases for the given sentence
Image Captioning [201]	Generate a caption that explains the image's content	An image	A caption describing that image
Video Description [112]	Generate a video description that explains the content of the video	A video clip	A sentence describing that video

The availability of powerful deep neural models, computationally intensive architecture, and pre-trained transformer models results in the incredible adoption of a variety of text generation applications. The applications use different methods including recurrent neural networks, long short-term memory networks, gated recurrent units for learning language representations, and later sequence-to-sequence learning, which opens a new chapter characterized by the wide application of the encoder-decoder architecture. However, these sequence-to-sequence models cannot capture long term dependencies, motivated the development of pointer networks and attention networks. Then, the transformer architecture incorporates an encoder and a decoder with self-attention mechanism, which is now widely used by text generation tasks. Applying these models to different text generation tasks can result in different levels of performance due to differences in task-specific requirements, training data availability, model architecture, hyperparameters, and evaluation metrics. Even if the same or similar models are used for different tasks, the architecture of the model may need to be modified or fine-tuned based on the requirements of the specific task.

The availability and quality of training data can significantly impact the performance of a text generation model. Models that are trained on large, diverse, and high-quality datasets specific to a given task or domain tend to perform better than those trained on more general datasets or limited data. Different text generation tasks may require different data preprocessing steps, such as tokenization, normalization, stemming, and stop-word removal. The choice of the model architecture and hyperparameters can also impact the performance of a text generation model. For example, transformer-based models such as GPT tend to perform well on a variety of text generation tasks due to their ability to capture long-term dependencies, but different hyperparameters, such as the number of layers or attention heads, can affect the model's performance. Different text generation tasks have different requirements and constraints that affect the effectiveness of the model. Models that are optimized for a specific task may perform better than those that are more general-purpose. Thus, it is important to carefully consider these factors when selecting a model for a particular task.

In this section, advancements in text generation applications have been seen with the rise of deep neural network approaches. The text generation approaches are discussed in the next section.

4 Text generation approaches

Text Generation is an emerging area of research. Recently, deep learning approaches have made remarkable success in various text generation tasks [138], including text summarization, machine translation, question answering, story generation, short-dialog generation, and paraphrasing. This section presents the traditional approaches for text generation, deep learning techniques, and pre-trained transformer-based approaches to text generation.

4.1 Traditional approaches

Traditionally, text generation was done either by using templates or production rules of a predefined grammar or performing statistical analysis of existing human-written texts to predict sequences of words [17, 60, 139, 222]. The template-based text generation systems adopted rules and templates to design different modules for text generation that reflect the linguistic knowledge of vocabulary, syntax, and grammar. This approach decomposes the text-generation task into several interacting subtasks depending on the task-specific text generation application. The template-based approaches usually consist of several components, including content planning (deciding the input data, selecting and structuring content), sentence planning (choosing words, syntactic structures, choosing appropriate referring expressions to describe input entities), and text realization (converting specifications to a real text), each performing a specific function [137]. The statistical-based text generation systems encode the dependency between vocabulary and context in conditional probability [93]. The most popular statistical text generation model is the n-gram language model, which is usually coupled with the template-based approach for re-ordering and selecting fluent generated texts. With the traditional approaches, it is very time-consuming to automatically generate text like those generated by humans. Deep learning techniques have overcome these shortcomings of traditional approaches. With the development of deep learning approaches, the neural-based text generation models have gradually occupied a dominant position that better models the statistical relationship between vocabulary and context, thus significantly improving the performance of text generation, as discussed in the subsequent section.

4.2 Deep learning techniques

Deep learning architectures and algorithms have recently achieved state-of-the-art results in question–answer generation, machine translation, text summarization, dialogue response generation, and other text generation tasks. Deep learning supports automated multi-level attribute representation learning. The deep neural networks provide a uniform end-to-end framework for text generation. First, a neural network creates a representation of the user input. Then, this representation is used as input to a decoder which generates the system response. Representation learning often happens in a continuous space, such that different modalities of text (words, sentences, and even paragraphs) are represented by dense vectors.

A variety of architectures based on deep neural networks have been developed for the different application tasks of text generation. This section introduces deep learning techniques that are commonly used in text generation application tasks.

4.2.1 The encoder-decoder framework

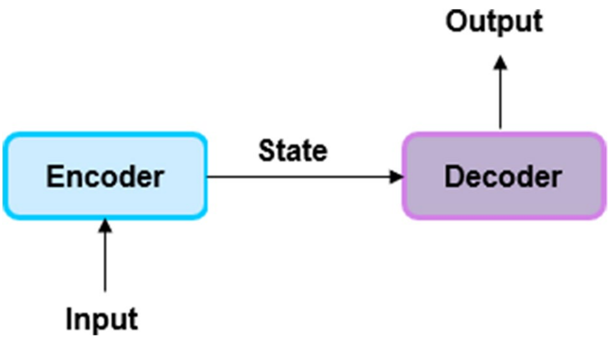
Much of the work on neural text generation adopts the encoder-decoder approach that was first advocated and shown to be successful for machine translation [32, 188]. First, the input is encoded into a continuous representation using an encoder. Then the text is produced using the decoder. Figure 7 illustrates this encoder-decoder framework for text generation. This network is often referred to as a sequence-to-sequence model as it takes as input a sequence, one element at a time, and then outputs a sequence, one element at a time.

This encoder and decoder are neural networks. The encoder depicts the input sequence as a hidden state vector and then transfers it to the decoder. The decoder then produces the output sequence.

4.2.2 Convolutional neural networks

Convolutional neural networks (CNNs) are specialized for processing data with a known grid-like topology. These networks have succeeded in computer vision tasks, which have been represented as 2-dimensional grids of image pixels [201, 217]. In recent years, CNNs have also been applied to natural language. In particular, they have been used to learn word

Fig. 7 Encoder-Decoder framework



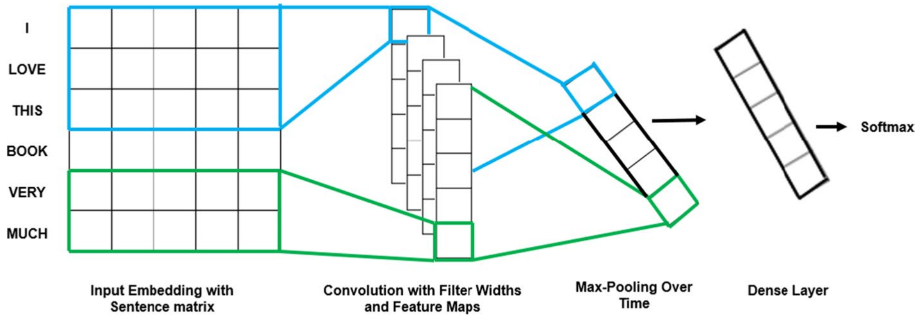

I
LOVE
THIS
BOOK
VERY
MUCH
Softmax
Input Embedding with Sentence matrix
Convolution with Filter Widths and Feature Maps
Max-Pooling Over Time
Dense Layer

Fig. 8 Convolutional Neural Network

representations for language modelling effectively [177] and summarization [28, 136, 138]. CNNs employ a specialized kind of linear operation called convolution (filter which extracts a specific pattern), followed by a pooling operation (subsamples the input on each filter to a fixed dimension of output), to build a representation that is aware of spatial interactions among input data points as shown in Fig. 8.

There are many variants of CNN that have different application areas, as mentioned in Table 4.

4.2.3 Recurrent neural networks

Recurrent Neural Networks are based on the concept of processing sequential data. They are termed “recurrent” since they perform the same computation over each token in the sequence, and each step depends on the results of previous computations, as shown in Fig. 9.

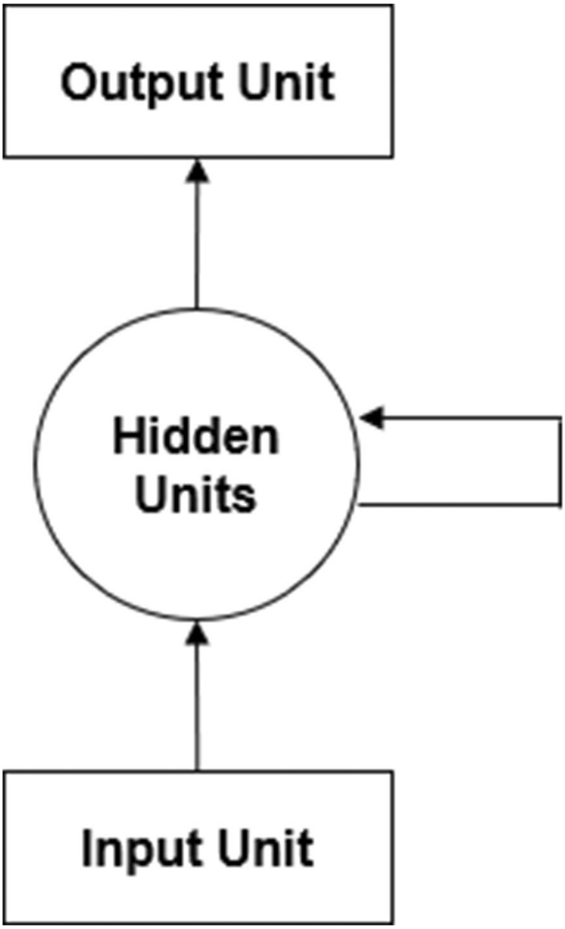
Most work on neural text generation has used RNNs due to their ability to capture the sequential nature of the text naturally and to process inputs and outputs of arbitrary length. There are various variants of RNN, including Bi-directional RNN, Parallel-RNN, Quasi RNN, RNN with external memory, and Convolutional RNN. Their features and application areas are provided in Table 5.

However, as the length of the input sequence grows, RNNs are prone to losing information from the beginning of the sequences due to vanishing and exploding gradients issues

Table 4 Variants of CNN with their application areas

Variants of CNN [Ref.]	Features	Application
CNN [61]	It encodes the entire source sentence simultaneously	Neural Machine Translation
Multichannel CNN [130]	It makes several predictions using multi-task learning	Question Identification, Sentence Classification
Dynamic CNN [39]	It maps the meanings of words to a sentence	Document Summarization
Multi-column CNN [44]	It uses multiple column networks to extract information from the input context	Question Answering
Dynamic multi-pool CNN [27]	It incorporates a dynamic multi-pooling layer that employs event triggers and arguments to retain critical data from the pooling layer	Event Extraction
Hybrid CNN-RNN [87]	Along with the frequency axis, it provides frequency shift-invariance	Image Captioning

Fig. 9 Recurrent Neural Network



[16, 146]. RNNs fail to model the long-range dependencies of natural languages. Consequently, Long short-term memory [74] and gated recurrent unit [32] have been proposed as alternative recurrent networks that are better prepared to learn long-distance dependencies. These units are better at learning to memorize only the part that is relevant for the future. At each time step, they dynamically update their states, deciding on what to memorize and what to forget from the previous input. The LSTM cell has separate input and forget gates as shown in Fig. 10, while the GRU cell performs both of these operations together using its reset gate.

The forget gate decides which information of the long-term memory is useful and which to forget. The next input gate determines which new information to be added to the network, and the final output gate decides the new hidden state. In a vanilla RNN, the entire cell state is updated with the current activation, whereas both LSTMs and GRUs have the mechanism to keep the memory from previous activations. This allows recurrent networks with LSTM or GRU cells to remember features for a long time and reduces the vanishing gradient problems as the gradient back propagates through multiple bounded nonlinearities. LSTMs and GRUs have been very successful in modelling natural languages in recent years

Table 5 Variants in RNN with their application areas

Variants of RNN [Ref.]	Features	Application
Simple RNN [29]	It uses the idea of processing sequential information	Machine translation, Sentence classification, Question Answering, language modelling
Bi-directional RNN (BRNN) [175, 187]	It allows training the network in both time directions simultaneously	Machine Translation, Speech Recognition
RNN-EM (RNN with External Memory)[152]	The use of external memory improves the memorization capability of RNNs	Language Modelling
GF-RNN (Gated Feedback RNN) [31]	It controls the strength of the temporal connection adaptively and allows multiple adaptive timescales	Character-level language modelling
p-RNN (Parallel-RNN) [73]	It leverages the added value of multiple item representations	Session-based recommendations
Q-RNN (Quasi RNN) [19]	It allows for parallel computation, thus enabling high throughput and good scaling to long sequences	Language Modelling, Document-level sentiment classification, Character-level machine translation
MV-RNN (Multi-View RNN) [36]	It monitors changes in user's preferences over time	Sequential Recommendation
CRNN (Convolutional RNN)[174]	It combines features of CNN and RNN	Text Recognition

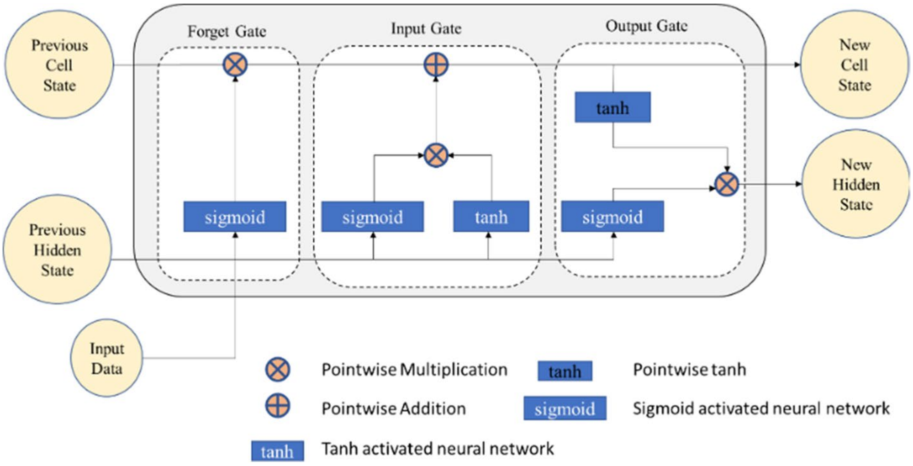
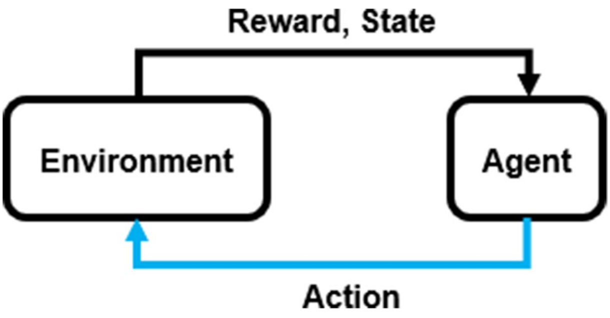


Fig. 10 Long Short Term Memory (LSTM)

Fig. 11 Reinforcement Learning



4.2.4 SeqGANs and reinforcement learning

SeqGAN (Sequential Generative Adversarial Network) is a variant of GAN (Generative Adversarial Network), used to generate text. This SeqGAN combines reinforcement learning and GANs for learning from discrete sequence data [231]. In SeqGANs, the generator is treated as an RL agent. The tokens generated till a particular time become the state. The token to be generated next is the action and the reward is the feedback given by the discriminator to guide the generator in evaluating the generated sequence.

Reinforcement learning (RL) is a gradual stamping of behaviour [86, 166] where an agent learns how to act in an environment by performing actions and analyzing the outcomes, as shown in Fig. 11. The performance is maximized by allowing software agents and machines to determine the ideal behaviour within a specific context automatically. The agents are required to learn their behaviour using simple reward feedback, known as the reinforcement signal.

When using reinforcement learning for automated text generation, the actions are writing words, and the states are the words already written by the algorithm. The actions, rewards, and policies corresponding to text generation tasks are mentioned in Table 6.

4.2.5 Transformer models

The Transformer models introduced by [195] have facilitated the enhancement of a wide range of text generation tasks. Transformer models are based on global dependencies between the input and output, use attention mechanisms, and have the capability to capture the linguistic knowledge of vocabulary, syntax, and semantics. The transformer is an encoder-decoder architecture. RNNs and LSTM architectures have significant difficulties with longer sequences as a result of the vanishing gradient problem [16, 146]. The probability of keeping context from a word that is further away from the word that is being processed diminishes exponentially as the sentence grows longer. Parallelization is a practical approach for training on larger datasets. The transformers expand with data and architectural size, capture longer sequence features, and enables parallel training. As a result, more effective and coherent language models are feasible. Prior to this, most of the ATG models were trained on supervised learning. However, supervised models need a large amount of annotated data for learning a particular task which is often not easily available, and they fail to generalize for other tasks.

The Transformer is a sequence-to-sequence model and consists of an encoder and decoder. Both encoder and decoder are multiple identical blocks layered on top of each other. The overall architecture of the Transformer is shown in Fig. 12. Each encoder block consists of a multi-head self-attention module followed by a position-wise feed-forward network (FFN). Around each module, a residual connection is employed, followed by the layer Normalization. Compared to the encoder blocks, decoder blocks additionally insert a third module, known as encoder-decoder attention, between the multi-head self-attention and FFN module. Furthermore, the masked attention module preserves the auto-regressive property, ensuring to prevent each position from attending to subsequent positions.

Transformer models such as BERT [40], and GPT-3 [20] are pre-trained on large corpora and use unsupervised learning for text generation. These pre-trained transformers are classified into three categories, namely: encoder-only (like BERT), decoder-only (like GPT-n), and encoder-decoder (like T5). Bidirectional Encoder Representations from Transformers (BERT) is the first deep bidirectional, unsupervised language representation [40] model. It is built upon work in contextual representations. BERT uses an attention mechanism along with a Transformer that learns contextual relations in text to generate a text [40]. Generative Pre-Trained Transformer (GPT) is an autoregressive language model that was trained with 175 billion parameters to generate text automatically. It uses unlabeled data and then fine-tuning for the specific downstream task. GPT-n series (GPT-1 [162], GPT-2 [20], GPT-3 [20]) shows significant performance on various ATG tasks even without finetuning or gradient updates. Transformer-based models, such as GPT, T5, XLNet, and BERT [20, 40, 163], showed impressive results on several text generation tasks such as question answering, language modelling, machine translation, sentiment analysis, and summarization, as shown in Table 7.

Discussion on research question 3

With the advancements in deep neural networks, text generation models are capable of generating realistic, fluent, and coherent natural language. This section helped to answer RQ 3- “What are the various approaches and associated architectures in the field of text generation.” It has been observed that the field of text generation has undergone significant

Table 6 Policy, Action and Reward function for different Applications

Application	Policy	Action	Reward
Text Summarization, Question Generation	Attention-based models, pointer-generators,	Selecting the next word for a headline, translation, and summary	ROUGE, BLEU
Question Answering	Sequence-to-sequence model	Selecting the answer from input document or selecting the start and end index of the answer	F1 score
Image/Video Captioning	Sequence-to-sequence model	Selecting the next token for the caption	CIDEr, SPICE, METEOR
Dialog Generation	Sequence-to-sequence model	Dialogue utterance to generate	BLEU, Length, and Diversity of dialogue

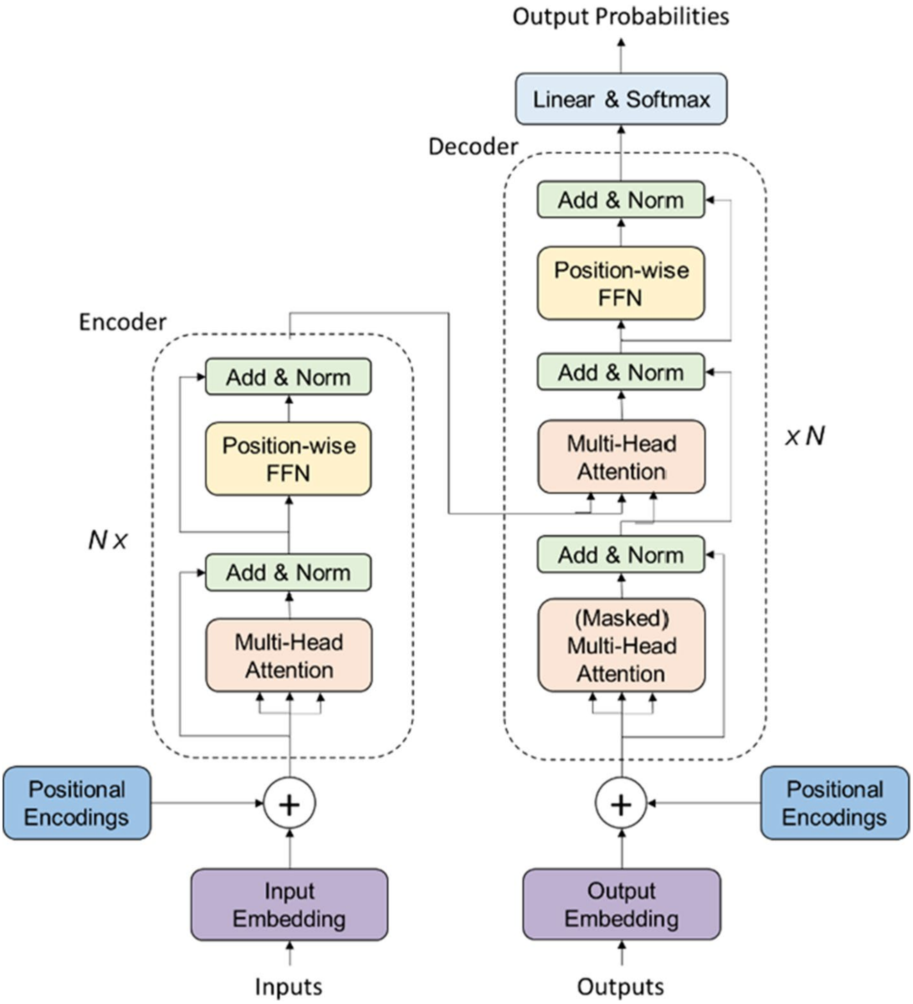


Fig. 12 Transformer architecture

changes from template-based and statistical-based traditional approaches to, most recently, pre-trained transformer based models. As a result of these advancements in techniques, the text generation research field has witnessed remarkable progress and a surge in interest for the study. The shift starts with recurrent neural networks, long short-term memory networks, gated recurrent units for learning language representations, and later sequence-to-sequence learning, which opens a new chapter characterized by the wide application of the encoder-decoder architecture. However, these sequence-to-sequence models cannot capture long term dependencies, which motivated the development of pointer networks and attention networks. Then, the transformer architecture incorporates an encoder and a decoder with self-attention mechanism, which is now widely used by text generation tasks. The availability of powerful deep neural models, computationally intensive architecture and pre-trained transformer models results in the incredible adoption of a variety of text generation applications, including text summarization, machine translation, creative

Table 7 Transformer Models with parameters and datasets

Transformer Model	Variations with parameters	Dataset used	Application
BERT (Bidirectional Encoder Representations from Transformers) [40]	BERT_Base (110 million parameters), BERT_Large (345 million parameters)	16 GB (Books Corpus + Wikipedia) + 3.3 Billion Words	Question Answering, Text Summarization, Machine Translation
DistilBERT (distilled version of BERT)	Base (66 million parameters)	16 GB (Books Corpus + Wikipedia) + 3.3 Billion Words	Question Answering, Language Modelling
RoBERTa (Robustly Optimized BERT Pretraining Approach)	Base (110 million parameters), Large (345 million parameters)	16 GB (Books Corpus, Wikipedia) + 144 GB (38 GB Web text corpus, 76 GB of Common Crawl News dataset)	Language Modelling
XLNet (combines the bidirectional capability of BERT with the autoregressive technology of Transformer-X)	Base (110 million parameters), Large (345 million parameters)	130 GB (Books Corpus + Wikipedia + additional) + 33 Billion Words	Question Answering, Sentiment analysis, Document Ranking
GPT-1 (Generative Pre-trained Transformer) [162]	117 M parameters	BooksCorpus dataset	NER, Language Modelling
GPT-2 [20]	1.5 billion parameters	WebText corpus (8 million documents), Children's Book dataset, LAMBADA dataset	Comprehension reading, Summarization, Translation
GPT-3 [20]	175 billion parameters	Common Crawl, WebText2, Books1, Books2, Wikipedia	Machine translation, transfer learning
T5 (Text-to-Text Transfer Transformer) [163]	T5-Small (60 million parameters), T5-Base (220 million parameters), T5-Large (770 million parameters), T5-3B (3 billion parameters), T5-11B (11 billion parameters)	C4 (Colossal Clean Crawled Corpus)	Summarization, Question Answering, Machine Translation, News Article generation

applications such as story generation, and dialogue generation. However, applying these models to different neural text generation tasks can depend on various factors, including the type of task, the architecture of the model, the size and quality of the training data, and the evaluation metrics used to measure the performance of the model.

The deep learning approaches arise due to the availability of a large number of corpora and significant computational resources. The standard datasets for text generation applications are mentioned in the next section.

5 Datasets for text generation tasks

In research, the datasets have been used to assess the performance of a proposed method. The deep learning models trained on large-scale datasets demonstrate unrivalled abilities to understand patterns in the data, opening a whole slew of new possibilities for creating realistic and coherent texts. Several datasets were recently created to support the training of text generation models. The datasets vary in terms of output lengths, generation tasks, and domain specificity. This section describes some of the datasets that are commonly used in text generation tasks. In Table 8, a shortlist of some of the task-specific standard datasets is provided, which is organized by the text generation applications.

Each dataset has many files, including training, testing, and validation files, in various formats. Few datasets have files in json format, text format, and excel format, while others are in csv format. For a better understanding of the above-mentioned datasets, screenshots of few datasets have been provided in the Fig. 13 below.

Discussion on research question 4

The availability of large and diverse datasets has also benefited the recent progress in text generation. This section helped to answer RQ 4- "What are the available datasets adopted in which the stated applications are organized." It has been observed that there are many datasets that vary in terms of text generation tasks and domain specificity. The datasets help to train, test and validate the text generation models. Nowadays, there is a trend to train models on massive datasets. However, training text generation models on diverse datasets provide the opportunity to improve their robustness. The models trained on massive datasets show an unmatched ability to automate the generation of fluent and coherent texts. Thus, while training a text generation model for a particular task, it is critical to choose the dataset carefully. For a specific text generation task, the most commonly used datasets are shown in Fig. 14. It also specifies the percentage of articles in which a given dataset is used for a specific text generation task.

After training the model, text decoding plays a vital role in the generation of text. The following sections discuss text decoding techniques and optimization methods for text generation.

6 Text decoding techniques and optimization algorithms

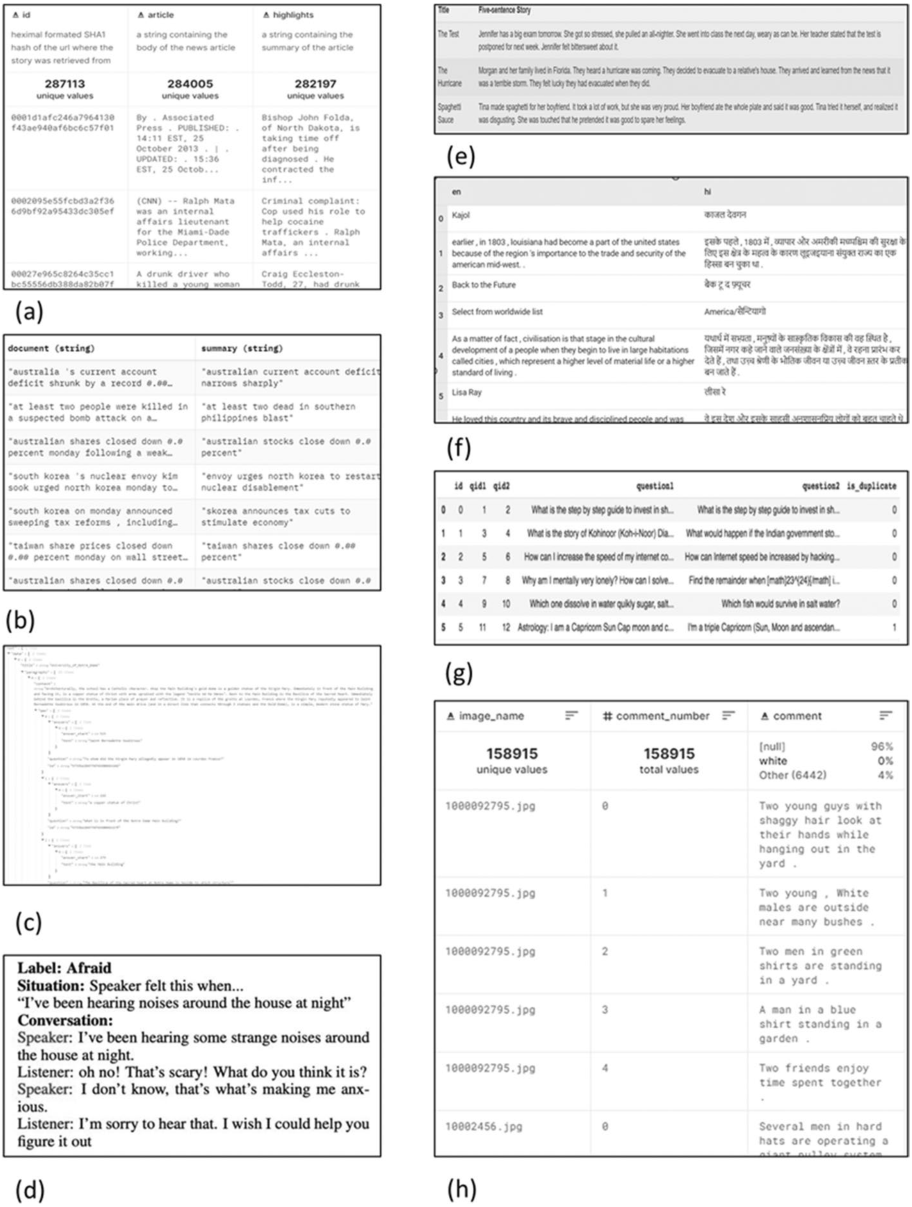
The automatic text generation model aims to generate text that is as good as human-written text. And after training the model, the quality of the generated text has a significant impact on the decoding strategy and optimization technique that one employs. In this section,

Table 8 Application-specific Text Generation Datasets

Application	Dataset [Year]	Description	URL
Text Summarization	CNN-Daily Mail dataset [2017]	Around 287 K news articles, each with upto four summaries	https://github.com/abisee/cnn-dailymail
Text Summarization	Newsroom dataset [2018]	1.3 M news articles and metadata	https://summari.es/
Text Summarization	DUC-2003 and DUC-2004 (Document Summarization Challenge) [2003, 2004]	500 news articles, each with four different human-generated reference summaries	https://duc.nist.gov/data.html
News Summary	XSum [2018]	One sentence summary of the input news article, 3,99,147 documents with 81,092 summaries	https://github.com/shashiongithub/XSum
Headline Generation	Gigaword [2015]	It contains 8 M + news articles	https://github.com/harvardnlp/sent-summary
News Summary	NY Times articles [2008]	1,399,358 documents, 294,011 summaries	https://www.kaggle.com/nzalake52/new-york-times-articles?select=nytimes_news_articles.txt
Question Answering and Question Generation	Stanford Question Answering data set (SQuAD) (1.0 and 2.0) [2016, 2018]	100 k + question–answer pairs collected by crowdsourcing through compilation of Wikipedia articles	https://rajpurkar.github.io/SQuAD-explorer/
Question Answering and Question Generation	TriviaQA [2017]	650 k question–answer-and-evidence triples	http://nlp.cs.washington.edu/triviaqa/
Yes/No Questions	BoolQ [2019]	16 k Questions with answers	https://github.com/google-research-datasets/boolean-questions
Multiple Choice Questions	ARC Corpus [2018]	Around 8 k science questions with multiple choice answers	https://allenai.org/data/arc
Question Answering	QuAC [2018]	100 k QA pairs	https://quac.ai/
Question Answering	HotPotQA [2018]	113 k Wikipedia-based question–answer pairs	https://hotpotqa.github.io/
Reading Comprehension with Commonsense Reasoning	ReCoRD [2019]	70,000 + news articles with 1,20,000 + queries, each query question has answer from corresponding news	https://sheng-z.github.io/ReCoRD-explorer/

Table 8 (continued)

Application	Dataset [Year]	Description	URL
Conversational Question Answering	CoQA dataset [2019]	127 k conversational questions with free-form text answers	https://stanfordnlp.github.io/coqa/
Dialogue Generation	OpenSubtitles [2016]	It contains conversations for 20 k + movies between movie characters	http://opus.nlpl.eu/OpenSubtitles.php
Dialogue Generation	Cornell Movie Dialogues Corpus [2011]	20 k dialogues between more than 10 k movie characters	https://www.cs.cornell.edu/~cristian/Cornell_Movie-Dialogs_Corpus.html
Machine Translation	WMT' 14 [2014]	850 M words from English–French parallel corpora of UN (421 M words), Europarl (61 M words), news commentary (5.5 M words), and two crawled corpora of 90 M and 272.5 M words	http://www.statmt.org/wmt14/translation-task.html
Story Telling	ROCStories [2016]	Around 100 k five-sentence commonsense stories	https://www.cs.rochester.edu/nlp/rocstories/
Paraphrase Generation	Microsoft Paraphrase corpus [2005]	5800 pairs of semantic equivalent sentences taken from online news sources with human annotations	https://www.microsoft.com/en-us/download/details.aspx?id=52398
Paraphrase Generation	ACL's Semantic Text Similarity competition [2017]	It includes text from news headlines, image captions, and user forums for 2012–2017	http://ixa2.si.ehu.es/stswiki/index.php/STSbenchmark
Paraphrase Generation	Quora Duplicate Questions [2017]	Around 400 k question pairs (of these around 111 k questions occur across multiple pairs)	http://qim.fs.quoracdn.net/quora_duplicate_questions.tsv
Image Captioning	COCO data set [2021]	330 k images	http://cocodataset.org/
Video Captioning	MSR-VTT [2017]	10 k videos and each containing 20 human-annotated captions	http://ms-multimedia-challenge.com/2017/challenge
Video Captioning	YouTube2Text/MSVD [2011]	1970 videos and each 40 human-annotated captions	http://www.cs.utexas.edu/users/ml/clamp/videoDescription/



id	article	highlights
heximal formated SHA1 hash of the url where the story was retrieved from	a string containing the body of the news article	a string containing the summary of the article
287113 unique values	284005 unique values	282197 unique values
0001d1afc246a7964130 f43ae940af6bc6c57f01	By . Associated Press . PUBLISHED: . 14:11 EST, 25 October 2013 . | . UPDATED: . 15:36 EST, 25 Octob...	Bishop John Folda, of North Dakota, is taking time off after being diagnosed . He contracted the inf...
0002095e55fcbd3a2f36 6d9bf92a95433dc305ef	(CNN) -- Ralph Mata was an internal affairs lieutenant for the Miami-Dade Police Department, working...	Criminal complaint: Cop used his role to help cocaine traffickers . Ralph Mata, an internal affairs ...
00027e965c8264c35cc1 bc55556db388da82b07f	A drunk driver who killed a young woman	Craig Eccleston-Todd, 27, had drunk

(a)

document (string)	summary (string)
"australia 's current account deficit shrunk by a record 0.00…	"australian current account deficit narrows sharply"
"at least two people were killed in a suspected bomb attack on a…	"at least two dead in southern philippines blast"
"australian shares closed down 0.0 percent monday following a weak…	"australian stocks close down 0.0 percent"
"south korea 's nuclear envoy kim sook urged north korea monday to…	"envoy urges north korea to restart nuclear disablement"
"south korea on monday announced sweeping tax reforms , including…	"skorea announces tax cuts to stimulate economy"
"taiwan share prices closed down 0.00 percent monday on wall street…	"taiwan shares close down 0.00 percent"
"australian shares closed down 0.0	"australian stocks close down 0.0

(b)

[illegible]

(c)

Label: Afraid
Situation: Speaker felt this when...
"I've been hearing noises around the house at night"
Conversation:
Speaker: I've been hearing some strange noises around the house at night.
Listener: oh no! That's scary! What do you think it is?
Speaker: I don't know, that's what's making me anxious.
Listener: I'm sorry to hear that. I wish I could help you figure it out

(d)

Title	Five-sentence Story
The Test	Jennifer has a big exam tomorrow. She got so stressed, she pulled an all-nighter. She went into class the next day, weary as can be. Her teacher stated that the test is postponed for next week. Jennifer felt bittersweet about it.
The Hurricane	Morgan and her family lived in Florida. They heard a hurricane was coming. They decided to evacuate to a relative's house. They arrived and learned from the news that it was a terrible storm. They felt lucky they had evacuated when they did.
Spaghetti Sauce	Tina made spaghetti for her boyfriend. It took a lot of work, but she was very proud. Her boyfriend ate the whole plate and said it was good. Tina tried it herself, and realized it was disgusting. She was touched that he pretended it was good to spare her feelings.

(e)

	en	hi
0	Kajol	काजल देवगन
1	earlier , in 1803 , louisiana had become a part of the united states because of the region 's importance to the trade and security of the american mid-west . .	इसके पहले , 1803 में , व्यापार और अमरीकी मध्यपश्चिम की सुरक्षा के लिए इस क्षेत्र के महत्व के कारण लुइज़ियाना संयुक्त राज्य का एक हिस्सा बन चुका था .
2	Back to the Future	बैक टू द फ्यूचर
3	Select from worldwide list	America/सैन्टियागो
4	As a matter of fact , civilisation is that stage in the cultural development of a people when they begin to live in large habitations called cities , which represent a higher level of material life or a higher standard of living .	यथार्थ में सभ्यता , मनुष्यों के सांस्कृतिक विकास की वह स्थित है , जिसमें नगर कहे जाने वाले जनसंख्या के क्षेत्रों में , वे रहना प्रारंभ कर देते हैं , तथा उच्च श्रेणी के भौतिक जीवन या उच्च जीवन स्तर के प्रतीक बन जाते हैं .
5	Lisa Ray	लीसा रे
	He loved this country and its brave and disciplined people and was	वे इस देश और इसके साहसी अनुशासनप्रिय लोगों को बहुत चाहते थे

(f)

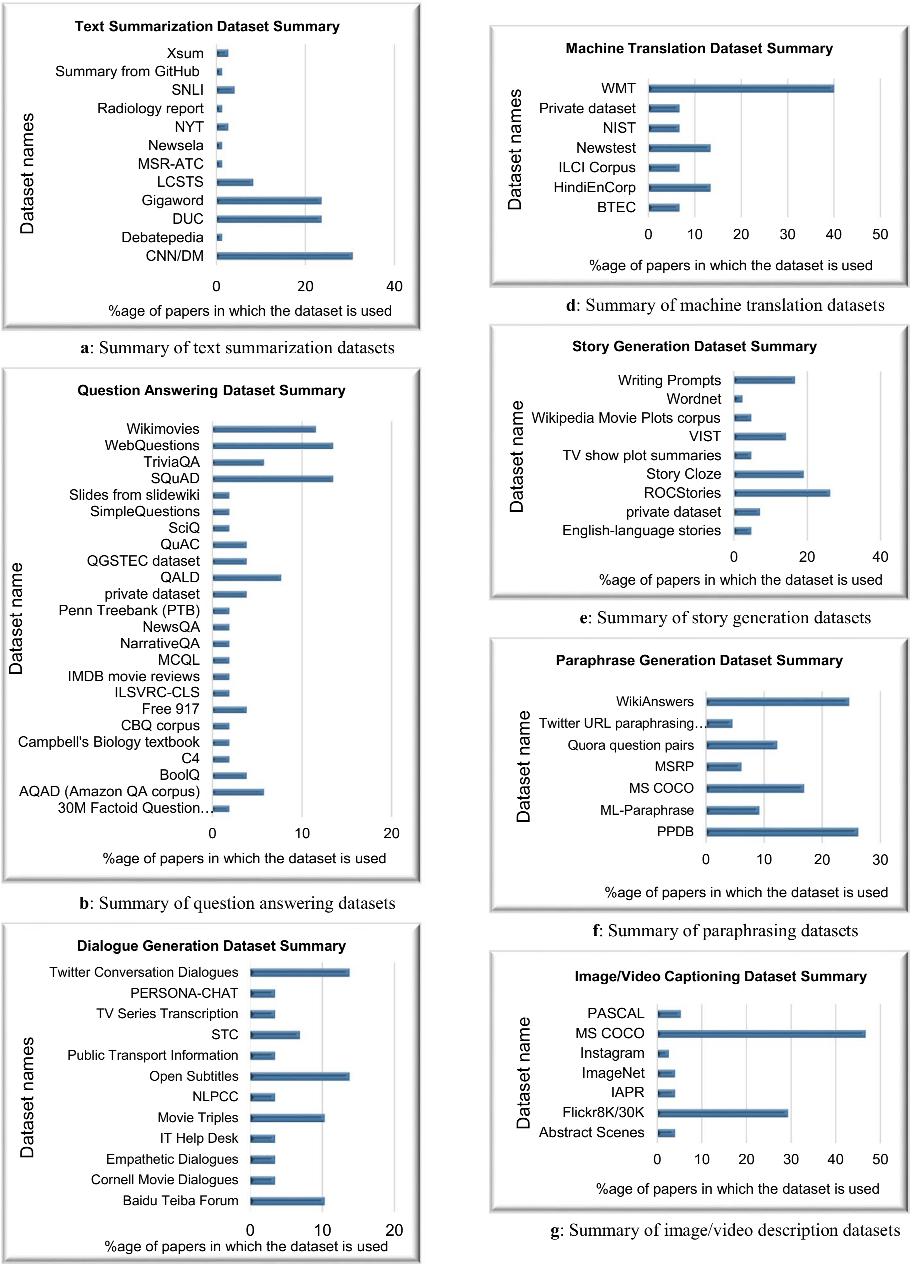
	id	qid1	qid2	question1	question2	is_duplicate
0	0	1	2	What is the step by step guide to invest in sh...	What is the step by step guide to invest in sh...	0
1	1	3	4	What is the story of Kohinoor (Koh-i-Noor) Dia...	What would happen if the Indian government sto...	0
2	2	5	6	How can I increase the speed of my internet co...	How can Internet speed be increased by hacking...	0
3	3	7	8	Why am I mentally very lonely? How can I solve...	Find the remainder when [math]23^{24}[/math] i...	0
4	4	9	10	Which one dissolve in water quikly sugar, salt...	Which fish would survive in salt water?	0
5	5	11	12	Astrology: I am a Capricorn Sun Cap moon and c...	I'm a triple Capricorn (Sun, Moon and ascendan...	1

(g)

image_name	# comment_number	comment
158915 unique values	158915 total values	[null] 96% white 0% Other (6442) 4%
1000092795.jpg	0	Two young guys with shaggy hair look at their hands while hanging out in the yard .
1000092795.jpg	1	Two young , White males are outside near many bushes .
1000092795.jpg	2	Two men in green shirts are standing in a yard .
1000092795.jpg	3	A man in a blue shirt standing in a garden .
1000092795.jpg	4	Two friends enjoy time spent together .
10002456.jpg	0	Several men in hard hats are operating a giant pulley system

(h)

Fig. 13 Screenshots of datasets [**a** CNN/DM dataset for summarization task **b** Gigaword dataset for summarization task **c** SQuAD dataset for question-answering task **d** Empathetic dialogues dataset for dialogue generation task **e** ROC story dataset for story generation and question-answering task **f** WMT dataset for translation task **g** Quora question pairs dataset for paraphrasing task **h** Flickr30k dataset for captioning task]



a: Summary of text summarization datasets

b: Summary of question answering datasets

c: Summary of dialogue generation datasets

d: Summary of machine translation datasets

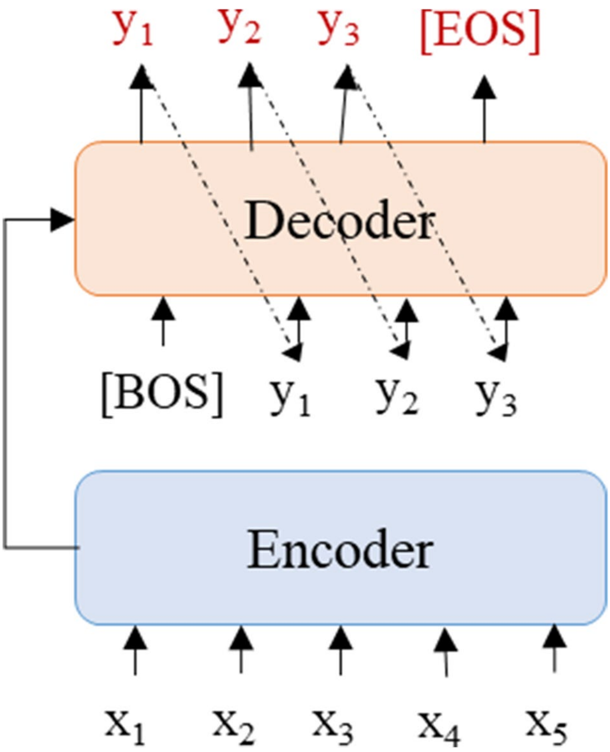
e: Summary of story generation datasets

f: Summary of paraphrasing datasets

g: Summary of image/video description datasets

Fig. 14 Summarized percentage of datasets usage in particular text generation task [**a** Summary of text summarization datasets **b** Summary of question answering datasets **c** Summary of dialogue generation datasets **d** Summary of machine translation datasets **e** Summary of story generation datasets **f** Summary of paraphrasing datasets **g** Summary of image/video description datasets]

Fig. 15 Autoregressive method



decoding methods and optimization techniques important for text generation have been discussed.

6.1 Decoding techniques

Decoding is the process of generating natural language texts from a model. As there is a need for one-to-one correspondence between input and output time steps of generation, which leads to a crucial key aspect named decoding. The decoding approach in a neural text generation system describes how the system searches for potential output utterances when generating a sequence. It specifies how the words are combined to form text and sentences. Without an appropriate decoding technique, the generated text results in vague and dull text.

Primarily, decoding can be categorized as autoregressive and non-autoregressive [215]. In an autoregressive generation, the target tokens are generated one by one in a sequential manner, as shown in Fig. 15. The beginning and end of decoding are controlled by special tokens, including [BOS] (beginning of a sentence) and [EOS] (end of a sentence), which implicitly determine the target length during decoding.

Traditional models capture the true distributions of words using this strategy. The fundamental reason is the conditional dependence property from left to right. The transformer-based models that cannot replicate the training benefits as training can be nonsequential, and inference holds to be sequential with autoregressive decoding is among one issue with this technique. Another issue with the autoregressive approach is that it is time-consuming,

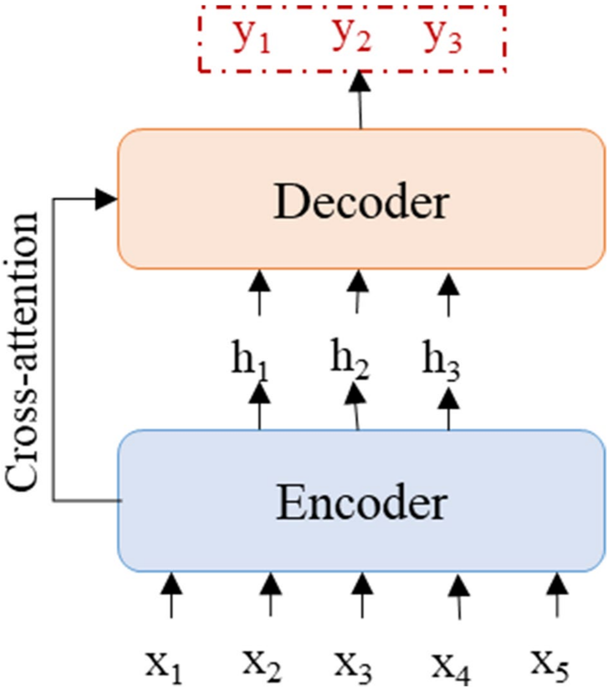


Fig. 16 Non-autoregressive method

especially for generating long target sentences. To alleviate this problem and accelerate decoding, non-autoregressive generation is proposed [65, 72]. The non-autoregressive decoding technique can generate all the target tokens in parallel, as shown in Fig. 16.

In this method, as all the target tokens are generated in parallel, there is no need for special token or target information to guide the termination of decoding. Using this technique, inference speed is hugely increased. [115, 157] proposes non-autoregressive techniques for summarization. The autoregressive techniques can be further viewed as sampling and search techniques. There are many different decoding strategies, including greedy, beam search, random sampling, top-k sampling, and nucleus sampling, as discussed.

A greedy search selects the most probability word from the language model. It uses this word as the next word and feeds it as input on the next step till it reaches maximum length [207]. However, greedy search is bound in a loop of the same words, resolved by random sampling [14]. Greedy search also lacks backtracking, which results in unnatural and meaningless sentences. The greedy search is not optimal for generating high-probability sentences [109], and this problem has been addressed by the beam search decoding method.

Random sampling picks the word randomly according to the conditional word probability extracted from the text generation model [14]. However, directly using the probabilities extracted from the text generation models often leads to incoherent text. Moreover, this text decoding method is not deterministic. Nevertheless, applying a softmax over the probability distribution and varying its parameter makes it smoother.

Beam search keeps the most probable words by tracking multiple possible sequences at once. It keeps track of the k-most probable partial sequences at each step, where k is the beam size. Beam search chooses the words to obtain an overall highest probability sentence [109, 198]. The text generated with beam search is more fluent as compared to greedy search. But when beam size is equal to one, beam search behaves as greedy search. Beam search produces a list of nearly identical sequences that fail to capture the inherent ambiguity of complex text generation tasks. Diverse Beam Search overcomes this problem by describing beam search as an optimization problem and augments the objective with diversity [197].

In top-k sampling, the 'k' most likely next words are selected, and then the next predicted word is sampled only from these 'k' words [53]. As'k' is fixed in top-k sampling, the number of words filtered from the next word probability distribution is not dynamically modified [205]. As a result, unlikely words may be selected among these 'k' words if the next word probability distribution is very sharp.

Nucleus sampling (or top-p sampling) selects words from the smallest possible set with a cumulative probability greater than some probability p. As a consequence, the number of words in the set can dynamically decrease and increase according to the next word probability distribution [75]. It is the best available decoding strategy for generating long-form text that is both high-quality as measured by human evaluation and as diverse as human-written text.

Other decoding techniques include semi-autoregressive, iterative, and mixed decoding [215]. The semi-autoregressive decoding generates multiple target tokens at one decoding step. The iterative decoding provides target information on each decoding step. Some works aim to combine these decoding strategies into a unified model. Tian et al. [190] propose a unified approach for machine translation that supports autoregressive, iterative, and autoregressive decoding methods.

Thus, based on the strengths and limitations of the text decoding techniques, the choice of decoding method has a significant impact on the linguistic features and the quality of the generated text.

6.2 Optimization techniques

With the tremendous growth in the amount of data, optimization has become an essential part of deep learning. The goal of the optimization algorithm is to minimize the loss function by reaching global semi-minima. Deep Learning models are becoming efficient and achieve better results with the use of optimization techniques. This section describes the commonly used optimization methods from a neural text generation perspective. There has been much interest in modifying the stochastic gradient descent algorithm with an adaptive learning rate for more stable training, e.g., AdaGrad, AdaDelta, and Adam, as shown in Fig. 17.

Further, these optimization algorithms are reviewed in a summarized manner based on their properties, pros, and cons, as mentioned in Table 9.

Discussion on research question 5

The text decoding methods and optimization techniques significantly impact the quality of the generated text This section helps to answer RQ 5- "What are various text decoding and optimization techniques used to automatically generate text." These techniques can be applied to different text generation tasks, i.e., text summarization, story generation, paraphrasing task, translation, and image captioning. Based on the text generation task, the autoregressive and non-autoregressive decoding technique can be utilized. Using a good model with bad decoding strategies lead to repetitive loop problem and inconsistent,

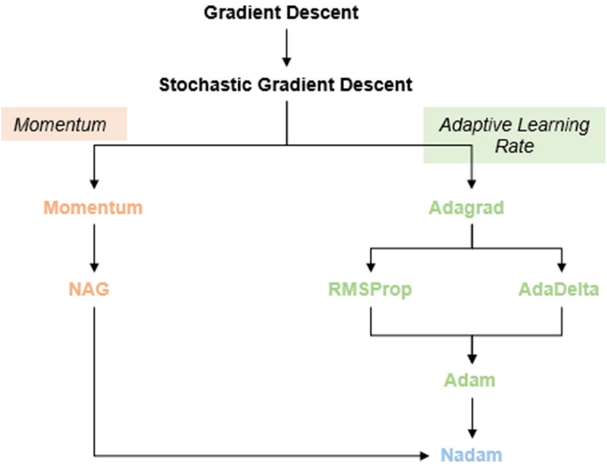


Fig. 17 Optimization algorithms for Text Generation

incoherent text generation problems. As a result, it is recommended to choose a decoding strategy carefully. Decoding methods like, top-k sampling and nucleus sampling produce more fluent text than beam search and greedy search. However, top-k sampling has suffered from generating repetitive word sequences recently. There has also been observed that greedy and beam search perform better if a different training objective is used by the text generation model.

In an end-to-end neural framework, all kinds of inputs, including target generated text are firstly mapped into numeric embeddings, and then neural modules feed-forward information layer by layer. Finally, the last output of the neural framework is used to generate the target tokens with a decoding strategy and calculate the losses to optimize parameters. Optimization algorithms are among those parameters and play an important part to infer the losses between neural networks. Thus, optimization algorithms are equally important for text generation models. It has been observed that the training performance of the model is influenced by the selection of an optimization algorithm. After understanding the concepts of various optimization algorithms and the function of their parameters, text generation models perform better. For a particular text generation task, the most commonly used optimizers are shown in Fig. 18. It also specifies the percentage of articles in which a given optimizer is used for a specific text generation task. There are many real-time text generation tools which are discussed in next section.

7 Real-time text generation tools

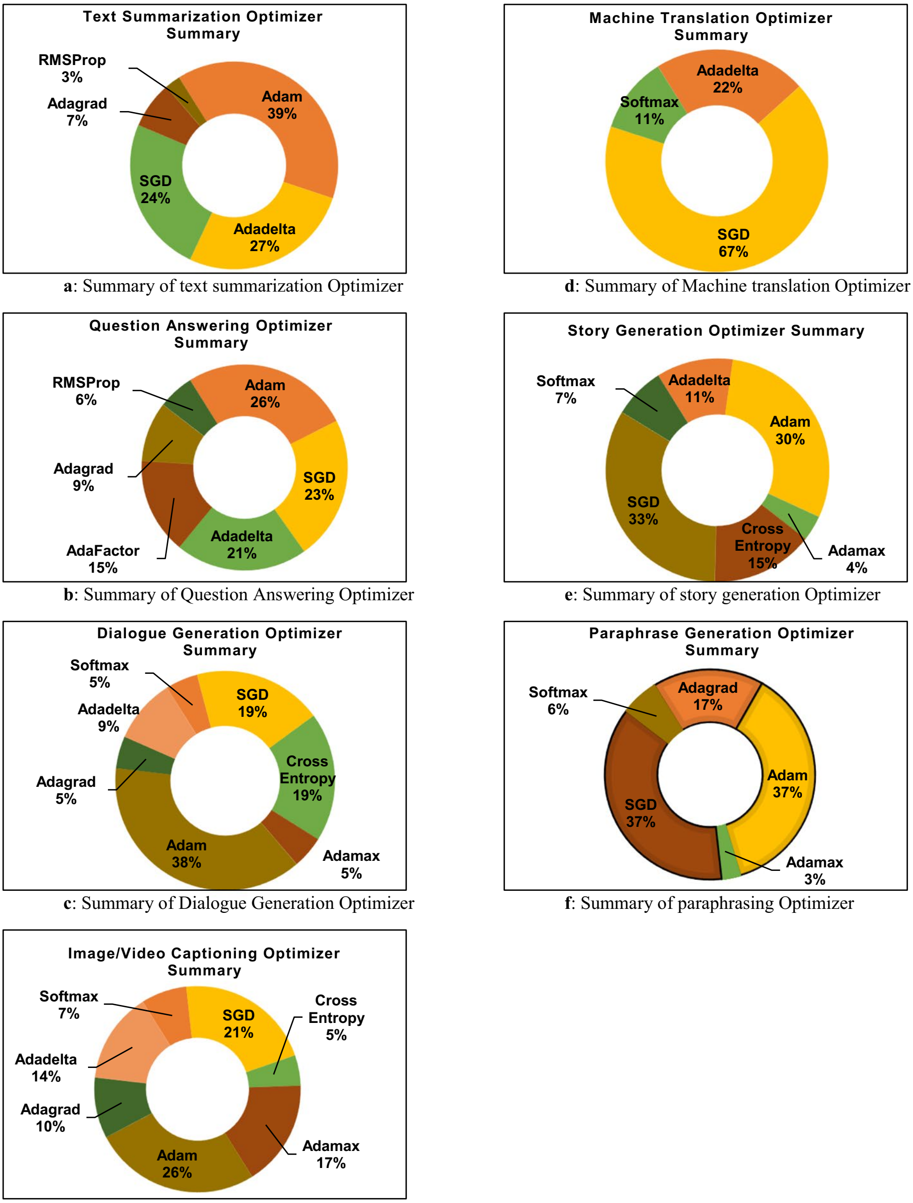
Text generation has played an essential role in various applications of text generation, such as paraphrasing, question generation, summarization, and dialogue systems. Text generation systems assist human writers and make the writing process more effective and time-saving. This section describes several real-time tools for text generation. The tools for text generation applications are mentioned in Table 10.

Discussion on research question 6

The existing automatic text generation application tools have been able to generate interesting text but are limited in terms of consistency, fluency, controllability, and diversity of

Table 9 Summarized review of Optimization algorithms

Method [Ref.]	Properties	Pros	Cons
Gradient Descent	It minimizes a given function, computes the gradient of the cost function	It is locally optimal and globally convergent	It performs redundant computations for large datasets
Stochastic Gradient Descent (SGD)	An iterative method that optimizes an objective function	It is simple and fast convergence for tasks having an extensive training set	It is implicitly sequential, parallelizing it with GPUs is difficult
AdaGrad [49]	Adapts the learning rate to the parameters, shrinks the learning rate adaptively	Infrequent parameters have larger updates, and periodic parameters have smaller updates	The learning rate shrinks as the accumulated sum grows, gradually being infinitely small
AdaDelta [234]	Adapts learning rates based on a moving window of gradient updates	Even after several updates, it continues learning	Not suitable for stationary problems
Root-Mean-Square propagation (RMSprop)	It collects the gradient for the learning rate as an exponentially weighted average	It performs well in a non-convex setting with differences between global and local structures	The upgrade process may be replicated around the local minimum in the late training stage
Adaptive Moments (Adam) [89]	It combines several properties from Momentum, RMSprop, AdaGrad	It has adaptive learning rates for each parameter	The method may not converge in some cases
Nesterov Momentum to Adam (Nadam) [45]	It combines NAG and ADAM	Achieves best results with Nesterov momentum	Not always the best algorithm to choose
SGD with momentum [158]	It makes use of past gradients to update the trainable parameters of the neural network	It accelerates SGD in the relevant direction, reduces oscillations, and faster convergence	It makes use of only simple momentum
Nesterov Accelerated Gradient Descent (NAG) [45]	It evaluates the gradient with the momentum	Momentum helps to jump out of locally optimal solutions	It is challenging to choose a suitable learning rate



a: Summary of text summarization Optimizer

d: Summary of Machine translation Optimizer

b: Summary of Question Answering Optimizer

e: Summary of story generation Optimizer

c: Summary of Dialogue Generation Optimizer

f: Summary of paraphrasing Optimizer

g: Summary of image/video description Optimizer

Fig. 18 Summarized percentage of optimizers used in particular text generation task [**a** Summary of text summarization Optimizer **b** Summary of Question Answering Optimizer **c** Summary of Dialogue Generation Optimizer **d** Summary of Machine translation Optimizer **e** Summary of story generation Optimizer **f** Summary of paraphrasing Optimizer **g** Summary of image/video description Optimizer]

Table 10 Real-time text generation application tools

Tool Name	Application area	Features	Access Link
Sassbook AI Summarizer	Text Summarizer	It understands the context and generates summaries in its own words, which captures the essence of the topic	https://sassbook.com/ai-summarizer
Summarize Bot	Summary Generator	It extracts and summarizes information from weblinks, news articles, scientific articles, books, e-mails, lectures, patents, and legal documents	https://www.summarizebot.com/
Resoomer	Summary Generator	It summarizes argumentative texts and articles, scientific texts, and educational documents	https://resoomer.com/
Automatic Summarizer	Text Summarizer	It generates multi-language summarization	https://autosummarizer.com/
SMMRY	Summary Generator	It creates a summary of articles and text in the specified number of lines	https://smmry.com/
Text Compactor	Summary Generator	It is a free online summarization tool mainly for textbooks	https://www.textcompactor.com/
Flexudy	Question generation, Summary generator	It creates WH-question, fill-in-the-blanks, and summaries out of user text	https://flexudy.com/
Lumos Learning	Question–Answer generator, Summarization	It provides automated assessment questions and summarization of test reports	https://www.lumoslearning.com/llwp/
GeneratorQ	Question–Answer generator	It generates a variety of questions and even answers based on inputted text	https://generatorq.azurewebsites.net/
Questo	Question generation	It generates questions from text and images of text as well	https://questo.ai/
Quillionz	Question/Answer Generation	It generates a wide range of questions, including multiple-choice	https://www.quillionz.com/
RevUp	Question Generation	It generates quality tests for self-motivated learners and teachers	https://paperswithcode.com/paper/revup-automatic-gap-fill-question-generation
AceBot	Conversation generator	It guides survey respondents to complete a survey through a fun and interactive conversation through chat messages	https://acebot.ai/
Pana	Conversational Agent	It is a virtual travel agent that suggests places to visit and where to eat	-
Sephora	Conversational Agent	It assists in the selection and purchase of cosmetics as well as the acquisition of beauty advice	-

Table 10 (continued)

Tool Name	Application area	Features	Access Link
HealthTap	Conversational agent	It allows people to ask questions, get test results evaluated, and receive feedback from physicians, making healthcare accessible	-
DeepL	Translation	It understands and translates text and helps in overcoming language barriers	https://www.deepl.com/home
Anusaaraka	Translation	It does the translation of English text to Hindi text	http://sampark.iiit.ac.in/anusaaraka/
Plot Generator	Story Generation	It creates customized short stories	https://www.plot-generator.org.uk
Storyline Creator	Story-telling	It allows for drafting ideas and organizing scenes and characters, allowing users to focus on the story plot and the creativity	https://www.storylinecreator.com/
Creative Help	Story Generation	It creates stories where users interact with the generated text	http://www.get-creative.help
Converse Smartly	Speech to text	It automatically transcribes audio from the English language in real-time, and has been detecting multiple speakers as well	https://www.folio3.ai/converse-smartly/
QuillBot	Paraphrasing tool	It helps in rewriting and enhancing any sentence, paragraph, or article	https://www.quillbot.com/
Paraphrase Online	Paraphrasing tool	It helps in automatic online paraphrasing of the article	https://www.paraphrase-online.com/
Wide Eyes	Image Captioning	It saves time and effort by generating tags for objects in the images for the fashion and e-commerce industry	https://wideeyes.ai/
Imagga	Image Captioning	It identifies the content of visuals, analyzes, and extracts their features	https://imagga.com/

the generated text. This section helps to answer RQ 6- "What real-time tools are available for automatic text generation tasks." It has been observed that there are diverse categories of automated text generation tools for different applications in the real world, such as some excel in generating short texts like headlines or tweets, and others excel at generating long texts like articles or blog entries. However, the authenticity of the content is missing with automated text generation tools, thus the fake/inaccurate content is roaming around. Many sectors have started using automated text to improve user experience as the algorithms are capable of generating human-like texts. But the automated tools can be exploited negatively. These tools can be abused by students who want to cheat on school work and hampers the student's ability. The content generated by these tools is fast and cheap but lacks the artistry involved in expressing thoughts. However, some of the observed tools are average in text generation, while others generate fluent text but are not freely accessible. Sometimes the text generated by these tools is superficial and repetitive. Thus, there is still much research being done and many problems to be solved, including long-term dependencies, redundancy while generating text, word sense ambiguity, incorrect grammar, consistency, and many more. These tools are limited by the data they were trained on and may not have a deep understanding of the topic one is writing about. The automated text generators cannot provide original and creative ideas, and makes people lazy and dependent on automation. For the effective text generation and the reliable assessment of the text generation models, there are many task-specific evaluation measures, as described in the next section.

8 Evaluation metrics for text generation

This section discusses the automatic evaluation measures that are frequently used to assess the advancements in the text generation system. Without proper evaluation, it is difficult to measure a system's competitiveness, which hinders the development of advanced algorithms for text generation. The goal of evaluation metrics is to evaluate the effectiveness of text generation tasks, and for this, a robust and unbiased evaluation metric is important. An automatic metric that correlates well with human assessments is ideal. It is desirable to employ a variety of metrics to assess the efficiency of the system over multiple aspects. The most popular automated evaluation methods for evaluating machine-generated text are mentioned in Table 11, with the pros and cons of the metric.

Automated text evaluation metrics are used to assess the text generation models, such as question–answer generation, text summarization, or machine translation. These evaluation measures provide a score that reflects the similarity between a human written reference text and an automatically generated text. There are many criteria based on which one decides which metric to use for which text generation task, as shown in Table 12.

Discussion on research question 7

As the field of text generation is continually advancing, evaluation is becoming critical for assessing progress in the area and performing comparisons between text generation models. This section answered the RQ 7- "Which metrics or indicators are used to evaluate automated text generation." It has been observed that traditionally language models have been evaluated based on perplexity, which concerns with the probability of a sentence being produced by the model. There are many well established automated evaluation metrics for assessing specific text generation tasks, such as METEOR and ROUGE for text

Table 11 Summarized review of automatic evaluation metrics for text generation

Metric [Ref.]	Description	Pros	Cons
BLEU [143]	It measures the similarity between two sentences	It is simple to use	It does not correlate well with human judgment
NIST [156]	It evaluates the quality of the text	Less frequent n-grams have heavy weights	It uses heavier weights for rarer words
ROUGE [114]	It evaluates the adequacy of the output text generated	It includes median or mean score from individual output text	It does not provide information of the grammar, and the narrative flow of the generated text
METEOR [96]	It counts the overlap of word or word units between candidate sentence and reference sentence	At sentence level, it yields a good correlation with human judgments	It treats all words equally in the reference set
CIDEr [196]	It is a consensus-based protocol that analyzes a sentence's similarity to a set of sentences in human form	It reflects a high agreement with the human consensus	It sometimes causes unimportant details of a sentence to be weighted more
WER [25]	It calculates the edit distance between the hypothesis sentence and references	The metric is computed efficiently and is reproducible	There is a dependency on the reference sentences
SPICE [8]	It employs scene graphs to calculate the correlation between reference and hypothesis text	It reflects a strong correlation to human assessments	It neglects the fluency of the captions generated
RIBES [210]	It relies on the order of words in the generated text	It is independent of word boundaries	It shows a lower correlation to human assessment
WMD [133]	It calculates the distance between two sequences represented by relative word frequencies	It is hyperparameter free and performs well in information retrieval	As the length of the document becomes large, the relation between sentences is lost
SMD [33]	It utilizes sentence embeddings to evaluate text in a continuous space	It correlates well with human evaluation	It considers equal weight for each sentence in a document
TER [210]	It computes the number of edits to the best corresponding reference	It assigns equal cost to all edits	It considers only identical matches between the reference and the hypothesis
BERTScore [236]	It computes the token similarity of two sentences dousing contextual embeddings	At system-level and sentence-level evaluations, it correlates well with human judgments	It computes only token similarity and not an exact match
HUSE [70]	It measures both the diversity and the quality of the text generated	It detects the low-diverse generations that humans fail to detect	For training the model, it relies on human judgments

Table 12 Summary of evaluation metrics with their application usage

Criteria	Evaluation Metric	SUM	QAG	DG	MT	SG	PRG	IC
n-gram based metrics	BLEU	✓	✓	✓	✓	✓	✓	✓
	NIST	-	-	✓	✓	-	-	-
	ROUGE	✓	✓	-	✓	✓	-	✓
	METEOR	✓	✓	✓	✓	✓	✓	✓
	CIDEr	✓	-	-	-	-	-	✓
	WER	-	-	✓	-	-	-	✓
	SPICE	-	-	-	-	-	-	✓
	RIBES		-	-	✓	-	-	-
	F-Score	✓	✓	✓	✓	✓	✓	✓
Distance-based Metrics	TER	-	-	-	-	✓	✓	-
	EED	✓	-	-	-	-	-	-
	WER	-	-	-	✓	-	-	-
Diversity-based Metric	TTR	-	-	✓	-	-	-	-
	SELF-BLEU	-	-	-	-	-	-	✓
	Distinct-k	-	-	✓	-	-	-	-
Embedding-based Metrics	WMD	-	-	-	-	-	-	✓
	SMD	✓	-	-	✓	-	-	-
	RUBER	-	-	✓	-	-	-	-
Learned Evaluation Metrics	BERTScore	✓	✓	-	✓	-	✓	✓
	MoverScore	✓	-	-	-	✓	-	-
	HUSE	-	-	✓	-	✓	-	-
Human-Centric Evaluation	Human Judgement	✓	✓	✓	✓	✓	✓	✓

where MT- Machine translation; QAG- Question–Answer Generation; SUM- Summarization; SG- Story Generation; DG- Dialog Response Generation; PRG- Paraphrase generation; IC- Image Captioning

summarization, BLEU for machine translation, SPICE and CIDEr for image captioning. However, there is no universal metric that suits all text generation tasks and reflects all relevant features of text. These work well to judge the quality of the generated text that the model has generated natural, human-like and grammatically correct sentences. However, with open-ended generation tasks such as story telling or dialogue generation, the model is expected to not only produce high quality text but also to be creative and diverse. Another important aspect to open-ended text generation is commonsense reasoning, which is referred as consistency. Since the models are expected to produce much longer text, they are more prone to generating illogical or factually incorrect sentences. BEAMetrics [176], a Benchmark to Evaluate Automatic Metrics help in better understanding the strengths and limitations of current metrics across a broad spectrum of tasks. Fast and reliable evaluation metrics are key to progress in research. While traditional natural language generation metrics are fast, they are not very reliable. Conversely, new metrics based on large pretrained language models are much more reliable, but require significant computational resources [83]. It is important that language models are evaluated in all dimensions of open-ended text generation—quality, diversity and consistency [140]. When evaluating language models on open-ended text generation task, it has been observed that Corpus-BLEU is the best metric to evaluate the quality of generated text due to its similarity with human judgement. As for diversity, Self-BLEU appears to be the best metric to use due to its simplicity to calculate. To evaluate the consistency of the generated text, using selection accuracy on the MultiNLI dataset is good enough for most cases. For specific task such as story generation, other dataset can be considered such as StoryCloze. Thus, the evaluation of large language models such as GPT is task specific [24, 63, 119]. While evaluating the text generation models for efficiency, it is necessary to rely on multiple metrics that reflect different text attributes such as fluency, grammaticality, coherence, readability, diversity, etc., Though human evaluations represent the gold standard for assessing the quality of machine-generated texts, but it is costly and time-consuming. As a result, automated measures for evaluation are to be used. But these automated evaluation measures should only be used as a supplement to human judgments and not as a replacement. Also, the automated metrices of evaluation are to be used when they present a reasonable correlation with human judgments.

However, there are still many issues or open problems in the generation of automated text which are discussed in the next section.

9 Challenges of automated text generation tasks

Generating fluent, meaningful, well-structured, and coherent text is pivotal for many text generation tasks. It takes significant effort by humans to model long-term dependencies while generating consistent text. It is an equally challenging task to do it automatically due to the discrete nature of textual data. This section identifies the main difficulties or challenges for the effective generation of text as below.

9.1 Text summarization

Text summarization is a challenging task since it requires thorough text analysis to provide a reliable summary [57]. A good summary must include relevant details and must be precise, but it must also consider aspects such as non-redundancy, significance, coverage, coherence, and readability [145]. To achieve all these things, in summary, is a major challenge. While many text summarization models provide tangible results, several issues are being suppressed. They often tend to repeat factually inaccurate information, struggle with Out of Vocabulary (OOV) words, emphasize a word/phrase several times, and are also a bit repetitive [142]. Another challenge is to develop a system that summarizes multi-lingual texts and generates a summary whose quality matches that of a human generated summary [57]. For multi-document summarizing, redundancy is the biggest problem [15, 239]. The so far proposed systems strive to identify important sentences in groups of different themes and hence suffer from the problem of sentence ordering. There is a need for a richer dataset and computation power. Thus, pre-trained models came into the picture [195]. The hybrid approach has gained attention recently [221]. By combining extractive and abstractive techniques, developing efficient hybrid approach methods to generate good quality summaries so that they match closely to human-written ones is another major challenge. The automatic text summarization evaluation metric such as ROUGE [114] is not considered complete [147]. The challenge with summary evaluation is to determine how adequate or useful a summary is relative to its source. Thus, methods for generating and evaluating summaries should complement each other.

9.2 Question-answering

Automatic question–answer generation is a significant advancement still many challenging issues are yet to be resolved. One such issue is to precisely understand the natural language questions and deduce the exact meaning to retrieve specific responses [50]. Another challenge is the selection of a question that has good coverage of the content and is of appropriate difficulty, and in the case of multiple-choice questions, distractor generation is the big challenge. As input texts grow longer, sequence-to-sequence models struggle to effectively utilize relevant contexts while avoiding unnecessary information [47]. The models do not pay much attention to the answers that are critical to question word generation. As a result, the generated question words do not match the answer form [186]. Previous neural question generation models suffer from a problem where a large percentage of the generated questions contain words from the question target. As a result, they generate unintended questions [88]. The models are not aware of the positions of the context words. Instead of considering the close and relevant words to the answer, they copy the context words that are far apart and irrelevant to the answer [186]. The one most frequent problem is the Lexical gap between questions. It concerns variation in the formation of questions in natural language. Users formulate the question in different ways and ask for the same information. This results in questions that differ lexically but are semantically equivalent. The problem of word sense ambiguity is still a challenge in the QA field [180]. To leverage the wide range of the available datasets for the question–answer generation is not trivial. The task of selecting an appropriate dataset is still an open problem [50]. Challenges also arise due to the limited size of the user's utterance, ambiguous, and missing information while interpreting a question.

9.3 Dialog response generation

Usually, conversational systems rely on RNN models, and RNNs are not able to model high-level variability [178]. The end-to-end conversational agents are prone to generating dull, generic, and boring responses [173]. To elicit a coherent, novel, and insightful response that is in line with the conversation spectrum, The conversational agents require adequate, accessible data [178, 181]. However, the conversational models, even with the powerful performance of neural networks, lack style, which possesses to be an issue as users may not be entirely satisfied with the interaction. Another problem is to encode contextual data such as world facts from knowledge bases or prior conversations. The response generated has to be contextually relevant to the conversation and also convey accurate paralinguistic functionality. Generating personalized dialogues is another challenging task [141].

9.4 Machine translation

Although neural machine translation (NMT) has been witnessing fast-growing research progresses, there are still many challenges. The major neural MT challenges are listed here. A major limitation of NMT is that it is not able to incorporable larger contextual information efficiently due to the learning ability of the model itself. The problem of reordering has not been addressed much so far [124]. The problems of alignment mechanism and vocabulary coverage always affect most of the NMT models [192]. NMT also struggles to deal with the translation of idioms [3]. Low-resource language MT is another hot spot, owing to multiple reasons, including morphological complexity and diversity, in addition to a lack of resources for many languages [79].

9.5 Story generation

Automatic story generation is challenging since it requires the generation of long-range dependencies and coherent natural language to describe a sensible sequence of events [227]. Another challenge in story generation is to create interactive narration along a certain story path so that the interactor has been provided the ability to modify the space or even the plot [208]. The commonly observed issues in generated stories are repeated plots, conflicting logic, and inter-sentence incoherence [20, 34, 53]. Another challenge is to use constraints to generate a creative story within the structure of the plot [151]. In most systems, evaluating the topicality, fluency, and overall quality of the stories generated poses a unique challenge [53].

9.6 Paraphrase generation

The ability to automatically generate alternative phrases of the same content has been demonstrated to be useful in several NLG areas, such as text summarization and question generation [155]. Automatically generating diverse and accurate paraphrases continues to be a difficult challenge due to the complexity of natural language [224]. Evaluation of the paraphrases generated is the most difficult aspect [105]. Another issue to be addressed is the generation of multiple diverse paraphrases of high quality to enhance generalization and robustness [160]. The issue of model holistic properties of sentences such as topic, style, and other features is still challenging.

9.7 Image captioning/ Video description

The major challenge in describing visual information to text is to learn the intermediate representation between the natural language domain and the visual domain [127]. Another challenge is the fine-grained natural descriptions of images or videos [42]. For instance, occlusions of interactive objects and unclear unit boundaries present additional challenges in effectively decoding the intent of the human behavior in a video. There are challenges associated with automatically generating textual reports for medical images and helping medical professionals produce reports more accurately and efficiently. The first is to generate prolonged texts with several sentences or even paragraphs, and the other is to generate captions with a wide range of heterogeneous forms [229].

Discussion on research question 8

The automated text generation applications have various challenges, including the generation of human-like text that is fluent, unambiguous, and diverse. This section helped to answer the RQ 8- “What challenges are faced in automated text generation tasks.” As the textual data is discrete in nature, it takes time and effort to model long-term dependencies while generating consistent text. Thus, applying neural models to different neural text generation tasks can depend on various factors, including the type of task, the architecture of the model, the size and quality of the training data, and the evaluation metrics used to measure the performance of the model. It has also been observed that the automated text generation is challenging, since it is not always possible to reproduce the reported results. Since the datasets used for text generation models are not always available publicly, it is difficult to conduct comparisons among various approaches to text generation. The size of training data and other key hyperparameters have a substantial impact on the quality of the generation text. However, there are still many open challenges in text generation that need to be addressed, including the generation of fluent, coherent, diverse, controllable, and consistent human-like text. Inspired by these challenges, the future aspects of this research area are presented in the next section.

10 Conclusion and future aspects

This survey captures a comprehensive study and up-to-date systematic review of current advancements in the field of text generation. Text generation applications has been categorized mainly into three categories: text-to-text generation (T2T), data-to-text generation (D2T), and multimodality-to-text generation (M2T), depending on the input (data, text, or multimodal). A variety of text-to-text generation applications, including text summarization, question–answer generation, story generation, machine translation, dialogue response generation, and paraphrase generation, have been discussed and analyzed. The main focus of this survey is on text-to-text generation applications, and it is beyond the scope of this survey to include all the recent developments in the various data or multimodality-to-text applications. This paper also mentions various models for text generation, including traditional and statistical models, deep learning based models, and pre-trained transformer architectures, and observed that deep learning approaches and transformer-based architectures have been generally achieving better performance than traditional methods. However, applying these models to different neural text generation tasks can depend on various factors, including the type of task, the architecture of the model, the size and quality of the training data, and the evaluation metrics used to measure the performance of the model. The quality of the generated text has a huge impact on the decoding strategy and optimization technique that one employs. This paper discussed the decoding methods and optimization techniques important for generating human-like fluent text. A diverse text generation task-specific standard datasets that are required to train, test, and validate the systems have also been provided in this article, along with their URLs. This field has made much progress in recent years. As a result, various text generation application-specific tools are available in the real world, which is also reviewed in this paper along with their strengths and limitations. Many sectors have started using automated text to improve user experience as the recent advancements in the technology is capable of generating human-like texts. Though, the content generated by the automated tools is fast and cheap but lacks artistry involved in expressing thoughts. The automatic text generation system's goal is to generate text as good as human-written text. The assessment of the generated text is essential to improve the performance of text generative models. However, human evaluation remains the gold standard for assessing the quality of automated generated texts, but it is time-consuming and expensive.

For effective text evaluation, various automatic evaluation metrices have been analyzed and reviewed in this paper. Nevertheless, many open challenges in text generation need to be addressed, including the generation of fluent, coherent, diverse, controllable, and consistent human-like text. Inspired by these challenges, the future aspects in this research direction include generating long-term fluent and coherent text.The advancements in the generic models should be done so that they can learn from some low resources and can handlemultiple languages without large quantities of training data. The generation of diverse texts conditioned by specific attributesand characteristics is another research direction. Practical applications for real-time text generation should be developed thatensure responsible usage of the generated text. The need of the hour is to create a universal evaluation metric that suits all textgeneration tasks and reflects all desired properties of text that correlate with human judgments.

This survey aims to provide a comprehensive overview of current advancements in automated text generation and to introducethe topic to researchers by providing pointers and synthesis to pertinent studies. This paper is believed to serve as a valuablereference for those concerned with learning and advancing this interesting research area.

Declarations I hereby confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere. No potential conflict of interest was reported.

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval The author declares that this article complies with the ethical standard.

References

1. Abrishami M, Rashti MJ, Naderan M (2020) Machine Translation Using Improved Attention-based Transformer with Hybrid Input. In: 2020 6th International Conference on Web Research (ICWR). IEEE, pp 52–57
2. Acharya M, Kafle K, Kanan C (2018) TallyQA: Answering complex counting questions. arXiv. https://doi.org/10.1609/aaai.v33i01.33018076
3. Agrawal R, Sharma DM (2017) Building an Effective MT System for English-Hindi Using RNN's. Int J Artif Intell Appl 8:45–58. https://doi.org/10.5121/ijaia.2017.8504
4. Alloatti F, Di Caro L, Sportelli G (2019) Real Life Application of a Question Answering System Using BERT Language Model. In: Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 250–253
5. Alomari A, Idris N, Sabri AQM, Alsmadi I (2022) Deep reinforcement and transfer learning for abstractive text summarization: A review. Comput Speech Lang 71:101276. https://doi.org/10.1016/j.csl.2021.101276
6. Alsaleh A, Althabiti S, Alshammari I, et al (2022) LK2022 at Qur'an QA 2022: Simple Transformers Model for Finding Answers to Questions from Qur'an. In: Proceedings ofthe OSACT 2022 Workshop @LREC2022. Eur Lang Res Assoc (ELRA), Marseille, pp 120–125
7. Ammanabrolu P, Tien E, Cheung W, et al (2019) Guided Neural Language Generation for Automated Storytelling. 46–55.https://doi.org/10.18653/v1/w19-3405
8. Anderson P, Fernando B, Johnson M, Gould S (2016) SPICE: Semantic propositional image caption evaluation. Lect Notes ComputSci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 9909 LNCS:382–398. https://doi.org/10.1007/978-3-319-46454-1_24
9. Anderson P, He X, Buehler C, et al (2018) Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 6077–6086.https://doi.org/10.1109/CVPR.2018.00636
10. Asghar N, Poupart P, Hoey J, et al (2018) Affective neural response generation. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 10772 LNCS:154–166. https://doi.org/10.1007/978-3-319-76941-7_12
11. Bahdanau D, Cho K, Bengio Y (2015) Neural Machine Translation by Jointly Learning to Align and Translate. 3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc 1–15

12. Bapna A, Chen MX, Firat O, et al (2020) Training deeper neural machine translation models with transparent attention. Proc 2018 Conf Empir Methods Nat Lang Process EMNLP 2018 3028–3033. https://doi.org/10.18653/v1/d18-1338
13. Barrull R, Kalita J (2020) Abstractive and mixed summarization for long-single documents. http://arxiv.org/abs/200701918 1–9
14. Basu S, Ramachandran GS, Keskar NS, Varshney LR (2021) Mirostat: A Neural Text Decoding Algorithm that Directly Controls Perplexity. ArXiv 200714966:1–25
15. Baumel T, Eyal M, Elhadad M (2018) Query Focused Abstractive Summarization: Incorporating Query Relevance, Multi-Document Coverage, and Summary Length Constraints into seq2seq Models. arXiv:180107704. arXiv:1801.07704
16. Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Networks 5:157–166. https://doi.org/10.1109/72.279181
17. Bott S, Saggion H, Figueroa D (2012) A hybrid system for spanish text simplification. 3rd Work Speech Lang Process Assist Technol SLPAT 2012 2012 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol NAACL-HLT 2012 - Proc 75–84
18. Bowman SR, Vilnis L, Vinyals O, et al (2016) Generating sentences from a continuous space. CoNLL 2016 - 20th SIGNLL Conf Comput Nat Lang Learn Proc 10–21. https://doi.org/10.18653/v1/k16-1002
19. Bradbury J, Merity S, Xiong C, Socher R (2016) Quasi-Recurrent Neural Networks. 5th Int Conf Learn Represent 1–11
20. Brown TB, Mann B, Ryder N, et al (2020) Language Models are Few-Shot Learners. Adv Neural Inf Process Syst
21. Buck C, Bulian J, Ciaramita M, et al (2018) Ask the Right Questions: Active Question Reformulation with Reinforcement Learning. In: 6th International Conference on Learning Representations, ICLR 2018. Conference Track Proceedings (2018), pp 1–15
22. Cao Z, Luo C, Li W, Li S (2017) Joint Copying and Restricted Generation for Paraphrase. In: 31st AAAI Conference on Artificial Intelligence, AAAI 2017. AAAI, pp 3152–3158
23. Cao S, Wang L (2021) Controllable Open-ended Question Generation with A New Question Type Ontology. arXiv 6424–6439. http://arxiv.org/abs/2107.00152
24. Celikyilmaz A, Clark E, Gao J (2020) Evaluation of Text Generation: A Survey. 1–75.http://arxiv.org/abs/2006.14799
25. Chen S, Beeferman D, Rosenfeld R (1998) Evaluation metrics for language models. Proc DARPA Broadcast News Transcr Underst Work 275– 280
26. Chen J, Xiao G, Han X, Chen H (2021) Controllable and Editable Neural Story Plot Generation via Control-and-Edit Transformer. IEEE Access 9:96692–96699. https://doi.org/10.1109/ACCESS.2021.3094263
27. Chen Y, Xu L, Liu K, et al (2015) Event extraction via dynamic multi-pooling convolutional neural networks. ACL-IJCNLP 2015 - 53rd Annu Meet Assoc Comput Linguist 7th Int Jt Conf Nat Lang Process Asian Fed Nat Lang Process Proc Conf 1:167–176. https://doi.org/10.3115/v1/p15-1017
28. Cheng J, Lapata M (2016) Neural Summarization by Extracting Sentences and Words. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 484–494
29. Cho K, van Merriënboer B, Gulcehre C, et al (2014) Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. Empir Methods Nat Lang Process (EMNLP), Assoc Comput Linguist 1724–1734. https://doi.org/10.1128/jcm.28.9.2159-.1990
30. Cho WS, Zhang Y, Rao S, et al (2021) Contrastive Multi-document Question Generation. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 12–30
31. Chung J, Gulcehre C, Cho K, Bengio Y (2015) Gated Feedback Recurrent Neural Networks. In: 32nd International Conference on Machine Learning, ICML 2015. ICML
32. Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. ArXiv 14123555:1–9
33. Clark E, Celikyilmaz A, Smith NA (2020) Sentence mover's similarity: Automatic evaluation for multi-sentence texts. ACL 2019 - 57th Annu Meet Assoc Comput Linguist Proc Conf 2748–2760. https://doi.org/10.18653/v1/p19-1264
34. Clark E, Ji Y, Smith NA (2018) Neural Text Generation in Stories Using Entity Representations as Context. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 2250–2260

35. Clinchant S, Jung KW, Nikoulina V (2019) On the use of BERT for Neural Machine Translation. In: Proceedings of the 3rd Workshop on Neural Generation and Translation. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 108–117
36. Cui Q, Wu S, Liu Q et al (2020) MV-RNN: A Multi-View Recurrent Neural Network for Sequential Recommendation. IEEE Trans Knowl Data Eng 32:317–331. https://doi.org/10.1109/TKDE.2018.2881260
37. Dai B, Fidler S, Urtasun R, Lin D (2017) Towards Diverse and Natural Image Descriptions via a Conditional GAN. Proc IEEE Int Conf Comput Vis 2017-Octob:2989–2998. https://doi.org/10.1109/ICCV.2017.323
38. Dauphin YN, Fan A, Auli M, Grangier D (2017) Language modeling with gated convolutional networks. 34th Int Conf Mach Learn ICML 2017 2:1551–1559
39. Denil M, Demiraj A, Kalchbrenner N et al (2014) Modelling, Visualising and Summarising Documents with a Single Convolutional Neural Network. ArXiv 14063830:1–10
40. Devlin J, Chang M, Lee K, Toutanova K (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:181004805v2 4171–4186. https://doi.org/10.18653/v1/N19-1423
41. Dinan E, Roller S, Shuster K, et al (2019) Wizard of Wikipedia: Knowledge-Powered Conversational agents. In: ICLR. pp 1–18
42. Donahue J, Hendricks LA, Rohrbach M et al (2017) Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. IEEE Trans Pattern Anal Mach Intell 39:677–691. https://doi.org/10.1109/TPAMI.2016.2599174
43. Dong L, Mallinson J, Reddy S, Lapata M (2017) Learning to paraphrase for question answering. arXiv 875–886
44. Dong L, Wei F, Zhou M, Xu K (2015) Question Answering over Freebase with Multi-Column Convolutional Neural Networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 260–269
45. Dozat T (2016) INCORPORATING NESTEROV MOMENTUM INTO ADAM. In: ICLR Workshop. ICLR, pp 2013–2016
46. Du X, Cardie C (2017) Identifying Where to Focus in Reading Comprehension for Neural Question Generation. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 2067–2073
47. Du X, Shao J, Cardie C (2017) Learning to Ask: Neural Question Generation for Reading Comprehension. arXiv:170500106v1
48. Duan N, Tang D, Chen P, Zhou M (2017) Question Generation for Question Answering. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 866–874
49. Duchi JC, Bartlett PL, Wainwright MJ (2012) Randomized smoothing for (parallel) stochastic optimization. In: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC). IEEE, pp 5442–5444
50. Dwivedi SK, Singh V (2013) Research and Reviews in Question Answering System. Procedia Technol 10:417–424. https://doi.org/10.1016/j.protcy.2013.12.378
51. Evans R, Grefenstette E (2018) Learning Explanatory Rules from Noisy Data. J Artif Intell Res 61:1–64. https://doi.org/10.1613/jair.5714
52. Faizan A, Lohmann S (2018) Automatic generation of multiple choice questions from slide content using linked data. ACM Int Conf Proceeding Ser doi 10(1145/3227609):3227656
53. Fan A, Lewis M, Dauphin Y (2018) Hierarchical neural story generation. ACL 2018 - 56th Annu Meet AssocComput Linguist Proc Conf (Long Pap 1:889–898. https://doi.org/10.18653/v1/p18-1082
54. Feng B, Liu D, Sun Y (2021) Evolving transformer architecture for neural machine translation. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. ACM, New York, NY, USA, pp 273–274
55. Frome A, Corrado GS, Shelens J et al (2018) DeViSE: A Deep Visual-Semantic Embedding Model Andrea. Phys C Supercond its Appl. https://doi.org/10.1016/0921-4534(95)00110-7
56. Fung P, Bertero D, Xu P, et al (2014) Empathetic Dialog Systems. In: The International Conference on Language Resources and Evaluation. European Language Resources Association. European Language Resources Association
57. Gambhir M, Gupta V (2017) Recent automatic text summarization techniques. Artif Intell Rev 47:1–66. https://doi.org/10.1007/s10462-016-9475-9

58. Gao P, Li H, Li S, et al (2018) Question-Guided Hybrid Convolution for Visual Question Answering. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 11205 LNCS:485–501. https://doi.org/10.1007/978-3-030-01246-5_29
59. Garbacea C, Mei Q (2020) Neural Language Generation: Formulation, Methods, and Evaluation. http://arxiv.org/abs/200715780
60. Gardent C, Kow E (2007) A symbolic approach to near-deterministic surface realisation using tree adjoining grammar. In: ACL 2007 - Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics. pp 328–335
61. Gehring J, Auli M, Grangier D, et al (2017) Convolutional sequence to sequence learning. 34th Int Conf Mach Learn ICML 2017 3:2029–2042
62. Goldberg Y (2016) A Primer on Neural Network Models for Natural Language Processing. J Artif Intell Res 57:345–420. https://doi.org/10.1613/jair.4992
63. Goyal T, Li JJ, Durrett G (2022) News Summarization and Evaluation in the Era of GPT-3. arXiv. http://arxiv.org/abs/2209.12356
64. Grechishnikova D (2021) Transformer neural network for protein-specific de novo drug generation as a machine translation problem. Sci Rep 11:321. https://doi.org/10.1038/s41598-020-79682-4
65. Gu J, Bradbury J, Xiong C, et al (2017) Non-Autoregressive Neural Machine Translation. Proc 2018 Conf Empir Methods Nat Lang Process 479–488
66. Guan J, Wang Y, Huang M (2019) Story Ending Generation with Incremental Encoding and Commonsense Knowledge. Proc AAAI Conf Artif Intell 33:6473–6480. https://doi.org/10.1609/aaai.v33i01.33016473
67. Gupta A, Agarwal A, Singh P, Rai P (2018) A deep generative framework for paraphrase generation. 32nd AAAI Conf Artif Intell AAAI 2018 5149–5156
68. Harrison B, Purdy C, Riedl MO (2021) Toward Automated Story Generation with Markov Chain Monte Carlo Methods and Deep Neural Networks. In: AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. AAAI, pp 191–197
69. Harrison V, Walker M (2018) Neural generation of diverse questions using answer focus, contextual and linguistic features. In: Proceedings ofThe 11th International Natural Language Generation Conference. Association for Computational Linguistics, Tilburg, The Netherlands, pp 296–306
70. Hashimoto TB, Zhang H, Liang P (2019) Unifying human and statistical evaluation for natural language generation. NAACL HLT 2019 - 2019 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol - Proc Conf 1:1689–1701. https://doi.org/10.18653/v1/n19-1169
71. He X, Deng L (2017) Deep Learning for VisuaLunDerstanDing Deep Learning for Image-to-Text Generation. IEEE Signal Process Mag 109–116. https://doi.org/10.1109/MSP.2017.2741510
72. Helcl J, Haddow B, Birch A (2022) Non-Autoregressive Machine Translation: It’s Not as Fast as it Seems. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 1780–1790
73. Hidasi B, Quadrana M, Karatzoglou A, Tikk D (2016) Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations. In: Proceedings of the 10th ACM Conference on Recommender Systems. ACM, New York, NY, USA, pp 241–248
74. Hochreiter S, Schmidhuber J (21997) Long Short-Term Memory. Neural Comput. https://doi.org/10.17582/journal.pjz/2018.50.6.2199.2207
75. Holtzman A, Buys J, Du L, et al (2019) The Curious Case of Neural Text Degeneration. CEUR Workshop Proc 2540:
76. Huang C, Zaïane OR, Trabelsi A, Dziri N (2018) Automatic dialogue generation with expressed emotions. NAACL HLT 2018 - 2018 Conf North Am Chapter AssocComput Linguist Hum Lang Technol - Proc Conf 2:49–54. https://doi.org/10.18653/v1/n18-2008
77. Iyyer M, Wieting J, Gimpel K, Zettlemoyer L (2018) Adversarial example generation with syntactically controlled paraphrase networks. NAACL HLT 2018 - 2018 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol - Proc Conf 1:1875–1885. https://doi.org/10.18653/v1/n18-1170
78. Jain P, Agrawal P, Mishra A, et al (2017) Story Generation from Sequence of Independent Short Descriptions. ArXiv.https://doi.org/10.48550/arXiv.1707.05501
79. Jha S, Sudhakar A, Singh AK (2018) Learning cross-lingual phonological and orthagraphic adaptations: A case study in improving neural machine translation between low-resource languages. arXiv 1–48. https://doi.org/10.15398/jlm.v7i2.214
80. Jin J, Fu K, Cui R, et al (2015) Aligning where to see and what to tell: image caption with region-based attention and scene factorization. 1–20
81. Jozefowicz R, Vinyals O, Schuster M, et al (2016) Exploring the Limits of Language Modeling. arXiv:160202410

82. Kalchbrenner N, Blunsom P (2013) Recurrent continuous translation models. EMNLP 2013 - 2013 Conf Empir Methods Nat Lang Process Proc Conf 1700–1709
83. Kamal Eddine M, Shang G, Tixier A, Vazirgiannis M (2022) FrugalScore: Learning Cheaper, Lighter and Faster Evaluation Metrics for Automatic Text Generation. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 1305–1318
84. Kannan A, Kurach K, Ravi S, et al (2016) Smart Reply. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp 955–964
85. Karpathy A, Joulin A, Fei-Fei L (2014) Deep Fragment Embeddings for Bidirectional Image Sentence Mapping. 27th International Conference on Neural Information Processing Systems (NIPS'14). MIT Press, Cambridge, MA, USA, pp 1889–1897
86. Keneshloo Y, Shi T, Ramakrishnan N et al (2020) Deep Reinforcement Learning for Sequence-to-Sequence Models 31:2469–2489
87. Khamparia A, Pandey B, Tiwari S et al (2020) An Integrated Hybrid CNN–RNN Model for Visual Description and Generation of Captions. Circuits, Syst Signal Process 39:776–788. https://doi.org/10.1007/s00034-019-01306-8
88. Kim Y, Lee H, Shin J, Jung K (2019) Improving Neural Question Generation Using Answer Separation. Thirty-Third AAAI Conf Artif Intell Improv
89. Kingma DP, Ba J (2015) Adam: A Method for Stochastic Optimization. In: 3rd International Conference on Learning Representations. ICLR 2015, pp 1–15
90. Kiros R, Salakhutdinov R, Zemel R (2014) Multimodal neural language models. 31st Int Conf Mach Learn ICML 2014 3:2012–2025
91. Kiros R, Salakhutdinov R, Zemel RS (2014) Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models. ArXiv 14112539:1–13
92. Kitchenham B, Pearl Brereton O, Budgen D et al (2009) Systematic literature reviews in software engineering - A systematic literature review. Inf Softw Technol 51:7–15. https://doi.org/10.1016/j.infsof.2008.09.009
93. Knight K, Marcu D (2000) Statistics-Based Summarization - Step One: Sentence Compression. In: Knight2000StatisticsBasedS. American Association for Artificial Intelligence (www.aaai.org), pp 703–710
94. Kumar A, Irsoy O, Ondruska P, et al (2016) Ask me anything: Dynamic memory networks for natural language processing. 33rd Int Conf Mach Learn ICML 2016 3:2068–2078
95. Kumar V, Ramakrishnan G, Li YF (2019) Putting the horse before the cart: A generator-evaluator framework for question generation from text. CoNLL 2019 - 23rd ConfComput Nat Lang Learn Proc Conf 812–821. https://doi.org/10.18653/v1/k19-1076
96. Lavie A, Agarwal A (2005) METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. In: Proceedings of the Second Workshop on Statistical Machine Translation
97. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539
98. Lee J, Liang B, Fong H (2021) Restatement and Question Generation for Counsellor Chatbot. In: Proceedings of the 1st Workshop on NLP for Positive Impact. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 1–7
99. Lee J, Yoon W, Kim S, et al (2019) BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 1–7. https://doi.org/10.1093/bioinformatics/btz682
100. Lelkes AD, Tran VQ, Yu C (2021) Quiz-Style Question Generation for News Stories. In: Proceedings of the Web Conference 2021. ACM, New York, NY, USA, pp 2501–2511
101. Lemberger P (2020) Deep Learning Models for Automatic Summarization. http://arxiv.org/abs/200511988 1–13
102. Lewis M, Liu Y, Goyal N, et al (2020) BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. 7871–7880. https://doi.org/10.18653/v1/2020.acl-main.703
103. Lewis M, Yarats D, Dauphin YN, et al (2017) Deal or no deal? End-to-end learning for negotiation dialogues.EMNLP 2017 - Conf Empir Methods Nat Lang Process Proc 2443–2453. https://doi.org/10.18653/v1/d17-1259
104. Li J, Galley M, Brockett C, et al (2016) A diversity-promoting objective function for neural conversation models. 2016 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol NAACL HLT 2016 - Proc Conf 110–119. https://doi.org/10.18653/v1/n16-1014

105. Li Z, Jiang X, Shang L, Li H (2018) Paraphrase Generation with Deep Reinforcement Learning. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 3865–3878
106. Li B, Lee-urban S, Johnston G, Riedl MO (2013) Story Generation with Crowdsourced Plot Graphs. AAAI, pp 598–604
107. Li Y, Li K, Ning H, et al (2021) Towards an Online Empathetic Chatbot with Emotion Causes. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York, NY, USA, pp 2041–2045
108. Li J, Luong MT, Jurafsky D (2015) A hierarchical neural Autoencoder for paragraphs and documents. ACL-IJCNLP 2015 - 53rd Annu Meet Assoc Comput Linguist 7th Int Jt Conf Nat Lang Process Asian Fed Nat Lang Process Proc Conf 1:1106–1115. https://doi.org/10.3115/v1/p15-1107
109. Li J, Monroe W, Jurafsky D (2016) A Simple, Fast Diverse Decoding Algorithm for Neural Generation. ArXiv: 161108562
110. Li J, Monroe W, Ritter A, et al (2016) Deep reinforcement learning for dialogue generation. EMNLP 2016 - Conf Empir Methods Nat Lang Process Proc 1192–1202. https://doi.org/10.18653/v1/d16-1127
111. Li J, Monroe W, Shi T, et al (2017) Adversarial learning for neural dialogue generation. EMNLP 2017 - Conf Empir Methods Nat Lang Process Proc 2157–2169.https://doi.org/10.18653/v1/d17-1230
112. Li S, Tao Z, Li K, Fu Y (2019) Visual to Text: Survey of Image and Video Captioning. IEEE Trans Emerg Top Comput Intell 3:297–312. https://doi.org/10.1109/TETCI.2019.2892755
113. Liao K, Lebanoff L, Liu F (2018) Abstract Meaning Representation for Multi-Document Summarization. In: International Conference on Computational Linguistics. Santa Fe, New Mexico, USA, pp 1178–1190
114. Lin C-Y (2004) ROUGE: A Package for Automatic Evaluation of Summaries. In: Text Summarization Branches Out. Association for Computational Linguistics, Barcelona, Spain, pp 74–81
115. Liu P, Huang C, Mou L (2022) Learning Non-Autoregressive Models from Search for Unsupervised Sentence Summarization. arXiv 7916–7929. https://doi.org/10.18653/v1/2022.acl-long.545
116. Liu Y, Lapata M (2019) Text Summarization with Pretrained Encoders. arXiv. http://arxiv.org/abs/1908.08345
117. Liu X, Lei W, Lv J, Zhou J (2022) Abstract Rule Learning for Paraphrase Generation. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, California, pp 4273–4279
118. Liu PJ, Saleh M, Pot E, et al (2018) Generating Wikipedia by Summarizing Long Sequences. In: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings. pp 1–18
119. Liu J, Shen D, Zhang Y, et al (2021) What Makes Good In-Context Examples for GPT-3? DeeLIO 2022 - Deep Learn Insid Out 3rd Work Knowl Extr Integr Deep Learn Archit Proc Work 3:100–114. http://arxiv.org/abs/2101.06804
120. Liu W, Wang Z, Liu X et al (2017) A survey of deep neural network architectures and their applications. Neurocomputing 234:11–26. https://doi.org/10.1016/j.neucom.2016.12.038
121. Lopyrev K (2015) Generating News Headlines with Recurrent Neural Networks. ArXiv 151201712:1–9
122. Lu J, Yang J, Batra D, Parikh D (2016) Hierarchical Question-Image Co-Attention for Visual Question Answering. Adv Neural Inf Process Syst 289–297
123. Lu S, Zhu Y, Zhang W, et al (2018) Neural Text Generation: Past, Present and Beyond. http://arxiv.org/abs/180307133
124. Luong M-T, Pham H, D. Manning C (2015) Effective Approaches to Attention-based Neural Machine Translation. In: Proceedings ofthe 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, pp 1412–1421
125. Ma S, Sun X, Li W, et al (2018) Query and output: Generating words by querying distributed word representations for paraphrase generation. NAACL HLT 2018 - 2018 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol - Proc Conf 1:196–206. https://doi.org/10.18653/v1/n18-1018
126. Makav B, Kilic V (2019) A New Image Captioning Approach for Visually Impaired People. ELECO 2019 - 11th Int Conf Electr Electron Eng 945–949. https://doi.org/10.23919/ELECO47770.2019.8990630
127. Mao J, Xu W, Yang Y, et al (2015) Deep captioning with multimodal recurrent neural networks (m-RNN). 3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc 1090:1–17
128. Martin LJ, Ammanabrolu P, Wang X, et al (2018) Event representations for automated story generation with deep neural nets. 32nd AAAI Conf Artif Intell AAAI 2018 868–875

129. Mehta P, Arora G, Majumder P (2018) Attention based Sentence Extraction from Scientific Articles using Pseudo-Labeled data. Assoc Comput Mach 2–5. https://doi.org/10.48550/arXiv.1802.04675
130. Michalopoulos G, Chen H, Wong A (2020) Where's the Question? A Multi-channel Deep Convolutional Neural Network for Question Identification in Textual Data.215–226. https://doi.org/10.18653/v1/2020.clinicalnlp-1.24
131. Mou L, Song Y, Yan R, et al (2016) Sequence to Backward and Forward Sequences: A Content-Introducing Approach to Generative Short-Text Conversation. In: COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers. COLING, pp 3349–3358
132. Mridha MF, Lima AA, Nur K et al (2021) A Survey of Automatic Text Summarization: Progress, Process and Challenges. IEEE Access 9:156043–156070. https://doi.org/10.1109/ACCESS.2021.3129786
133. Nag D, Das B, Dash PS, et al (2015) From word embeddings to document distances. In: 32nd International Conference on International Conference on Machine Learning. ICML'15, Lille, France, pp 957–966
134. Nallapati R, Zhai F, Zhou B (2017) SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents. In: 31st AAAI Conference on Artificial Intelligence, AAAI 2017. pp 3075–3081
135. Nallapati R, Zhou B, dos Santos C, et al (2016) Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. In: Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 280–290
136. Narayan S, Cohen SB, Lapata M (2018) Ranking Sentences for Extractive Summarization with Reinforcement Learning. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 1747–1759
137. Narayan S, Gardent C (2012) Structure-Driven Lexicalist Generation. 24th International Confer- ence in Computational Linguistics (COLING). Mumbai, India, pp 100–113
138. Narayan S, Gardent C (2020) Deep Learning Approaches to Text Production. Synth Lect Hum Lang Technol 13:1–199. https://doi.org/10.2200/S00979ED1V01Y201912HLT044
139. Narayan S, Gardent C, Narayan S, et al (2015) Hybrid Simplification using Deep Semantics and Machine Translation To cite this version : HAL Id : hal-01109581
140. Nguyen A (2021) Language Model Evaluation in Open-ended Text Generation. arXiv. http://arxiv.org/abs/2108.03578
141. Niu T, Bansal M (2018) Polite dialogue generation without parallel data. arXiv. https://doi.org/10.1162/tacl_a_00027
142. PadmaPriya G, Duraiswamy K (2014) AN APPROACH FOR TEXT SUMMARIZATION USING DEEP LEARNING ALGORITHM. J Comput Sci 10:1–9. https://doi.org/10.3844/jcssp.2014.1.9
143. Papineni K, Roukos S, Ward T, Zhu W-J (2002) BLEU: a Method for Automatic Evaluation of Machine Translation. In: 40th Annual Meeting of the Association for Computational Linguistics (ACL). ACL, pp 311–318
144. Park HJ, Lee JS, Ko JG (2020) Achieving Real-Time Sign Language Translation Using a Smartphone's True Depth Images. In: 12th International Conference on Communication Systems & Networks (COMSNETS). IEEE, pp 622–625
145. Parveen D, Mesgar M, Strube M (2016) Generating coherent summaries of scientific articles using coherence patterns. EMNLP 2016 - Conf Empir Methods Nat Lang Process Proc 772–783. https://doi.org/10.18653/v1/d16-1074
146. Pascanu R, Mikolov T, Bengio Y (2018) On the difficulty of training recurrent neural networks. Phylogenetic Divers Appl Challenges Biodivers Sci 41–71. https://doi.org/10.1007/978-3-319-93145-6_3
147. Paulus R, Xiong C, Socher R (2017) A Deep Reinforced Model for Abstractive Summarization. 6th Int Conf Learn Represent ICLR 2018 - Conf Track Proc 1–12
148. Pauws S, Gatt A, Krahmer E, Reiter E (2019) Making effective use of healthcare data using data-to-text technology. Data Sci Healthc Methodol Appl 119–145. https://doi.org/10.1007/978-3-030-05249-2_4
149. Pawade D, Sakhapara A, Jain M et al (2018) Story Scrambler - Automatic Text Generation Using Word Level RNN-LSTM. Int J Inf Technol Comput Sci 10:44–53. https://doi.org/10.5815/ijitcs.2018.06.05
150. Pedersoli M, Lucas T, Schmid C, Verbeek J (2017) Areas of Attention for Image Captioning. Proc IEEE Int Conf Comput Vis 2017-Octob:1251–1259. https://doi.org/10.1109/ICCV.2017.140

151. Peng N, Ghazvininejad M, May J, Knight K (2018) Towards Controllable Story Generation. In: Proceedings of the First Workshop on Storytelling. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 43–49
152. Peng B, Yao K (2015) Recurrent Neural Networks with External Memory for Language Understanding. arXiv:150600195v1
153. Peng D, Zhou M, Liu C, Ai J (2020) Human–machine dialogue modelling with the fusion of word- and sentence-level emotions. Knowledge-Based Syst 192:105319. https://doi.org/10.1016/j.knosys.2019.105319
154. Portet F, Reiter E, Gatt A et al (2009) Automatic generation of textual summaries from neonatal intensive care data. Artif Intell 173:789–816. https://doi.org/10.1016/j.artint.2008.12.002
155. Prakash A, Hasan SA, Lee K, et al (2016) Neural paraphrase generation with stacked residual LSTM Networks. COLING 2016 - 26th Int Conf Comput Linguist Proc COLING 2016 Tech Pap 2923–2934
156. Przybocki M, Peterson K, Bronsart S, Sanders G (2009) The NIST 2008 metrics for machine translation challenge-overview, methodology, metrics, and results. Mach Transl 23:71–103. https://doi.org/10.1007/s10590-009-9065-6
157. Qi W, Gong Y, Jiao J, et al (2021) BANG: Bridging Autoregressive and Non-autoregressive Generation with Large Scale Pretraining
158. Qian N (1999) On the momentum term in gradient descent learning algorithms. Neural Netw 12:145–151. https://doi.org/10.1016/S0893-6080(98)00116-6
159. Qian Q, Huang M, Zhao H, et al (2018) Assigning personality/identity to a chatting machine for coherent conversation generation. Proc Twenty-Seventh Int Jt Conf Artif Intell 4279–4285
160. Qian L, Qiu L, Zhang W, et al (2019) Exploring Diverse Expressions for Paraphrase Generation. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 3171–3180
161. Qian L, Zhou H, Bao Y, et al (2021) Glancing Transformer for Non-Autoregressive Neural Machine Translation. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 1993–2003
162. Radford A, Narasimhan K (2018) Improving Language Understanding by Generative Pre-Training
163. Raffel C, Shazeer N, Roberts A, et al (2020) Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. arXiv:191010683v2 1–67
164. Rajasekar AA, Garera N (2021) Answer Generation for Questions With Multiple Information Sources in E-Commerce. Proc Flip DS Conf 1:
165. Rajpurkar P, Zhang J, Lopyrev K, Liang P (2016) SQuAD: 100,000+ Questions for Machine Comprehension of Text. Proc 2016 Conf Empir Methods Nat Lang Process 2383–2392.https://doi.org/10.18653/v1/D16-1264
166. Ranzato M, Chopra S, Auli M, Zaremba W (2016) Sequence Level Training with Recurrent Neural Networks. In: 4th International Conference on Learning Representations, ICLR. ICLR, pp 1–16
167. Rashkin H, Smith EM, Li M, Boureau YL (2020) Towards empathetic open-domain conversation models: A new benchmark and dataset. ACL 2019 - 57th Annu Meet Assoc Comput Linguist Proc Conf 5370–5381. https://doi.org/10.18653/v1/p19-1534
168. Reiter E, Dale R (1997) Building applied natural language generation systems. Nat Lang Eng 3:57–87. https://doi.org/10.1017/S1351324997001502
169. Ren Z, Wang X, Zhang N, et al (2017) Deep reinforcement learning-based image captioning with embedding reward. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017 2017-Janua:1151–1159. https://doi.org/10.1109/CVPR.2017.128
170. Roemmele M (2016) Writing Stories with Help from Recurrent Neural Networks. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16) Writing. AAAI, pp 4311–4312
171. Roemmele M, Gordon AS (2015) Interactive Storytelling. Springer International Publishing, Cham
172. Rush AM, Chopra S, Weston J (2015) A Neural Attention Model for Abstractive Sentence Summarization. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 379–389
173. Santhanam S, Shaikh S (2019) A Survey of Natural Language Generation Techniques with a Focus on Dialogue Systems - Past. Present and Future Directions A Survey of Natural Language Generation Techniques with a Focus on Dialogue Systems - Past, Present and Future Directions. https://doi.org/10.5087/dad.DOINUMBER

174. Saxena SS, Saranya G, Aggarwal D (2020) A Convolutional Recurrent Neural Network (CRNN) Based Approach for Text Recognition and Conversion of Text To Speech in Various Indian Languages. Int J Adv Sci Technol 29:2770–2776
175. Schuster M, Paliwal KK (1997) Bidirectional recurrent neural networks. IEEE Trans Signal Process 45:2673–2681. https://doi.org/10.1109/78.650093
176. Scialom T, Hill F (2021) BEAMetrics: A Benchmark for Language Generation Evaluation Evaluation. arXiv 1–20
177. See A, Liu PJ, Manning CD (2017) Get To The Point: Summarization with Pointer-Generator Networks. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 1073–1083
178. Serban I V., Sordoni A, Bengio Y, et al (2016) Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models. In: 30th AAAI Conference on Artificial Intelligence, AAAI 2016. AAAI Press, pp 3776–3783
179. Shetty R, Rohrbach M, Hendricks LA, et al (2017) Speaking the Same Language: Matching Machine to Human Captions by Adversarial Training. Proc IEEE Int Conf Comput Vis 2017-Octob:4155–4164. https://doi.org/10.1109/ICCV.2017.445
180. Song L, Wang Z, Hamza W, et al (2018) Leveraging Context Information for Natural Question Generation. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 569–574
181. Sordoni A, Galley M, Auli M, et al (2015) A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 196–205
182. Sriram A, Jun H, Satheesh S, Coates A (2018) Cold fusion: Training Seq2seq models together with language models. Proc AnnuConf Int Speech Commun Assoc INTERSPEECH 2018-Septe:387–391. https://doi.org/10.21437/Interspeech.2018-1392
183. Stasaski K, Rathod M, Tu T, et al (2021) Automatically Generating Cause-and-Effect Questions from Passages. Proc 16th Work Innov Use NLP Build Educ Appl BEA 2021 - held conjunction with 16th Conf Eur Chapter Assoc Comput Linguist EACL 2021 158–170
184. Su Y, Wang Y, Cai D et al (2021) PROTOTYPE-TO-STYLE: Dialogue Generation with Style-Aware Editing on Retrieval Memory. IEEE/ACM Trans Audio Speech Lang Process 29:2152–2161. https://doi.org/10.1109/TASLP.2021.3087948
185. Subramanian S, Wang T, Yuan X, et al (2018) Neural Models for Key Phrase Extraction and Question Generation. In: Proceedings of the Workshop on Machine Reading for Question Answering. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 78–88
186. Sun X, Liu J, Lyu Y, et al (2018) Answer-focused and Position-aware Neural Question Generation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 3930–3939
187. Sundermeyer M, Alkhouli T, Wuebker J, Ney H (2014) Translation Modeling with Bidirectional Recurrent Neural Networks Human Language Technology and Pattern Recognition Group. In: Emnlp2014. ACL, pp 14–25
188. Sutskever I, Vinyals O, Le QV (2014) Sequence to Sequence Learning with Neural Networks. Adv Neural Inf Process Syst 4:3104–3112
189. Tambwekar P, Dhuliawala M, Martin LJ, et al (2019) Controllable Neural Story Plot Generation via Reward Shaping. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, California, pp 5982–5988
190. Tian C, Wang Y, Cheng H, et al (2020) Train Once, and Decode As You Like. In: Proceedings of the 28th International Conference on Computational Linguistics. International Committee on Computational Linguistics, Stroudsburg, PA, USA, pp 280–293
191. Tian Z, Yan R, Mou L, et al (2017) How to make context more useful? An empirical study on context-Aware neural conversational models. ACL 2017 - 55th Annu Meet Assoc Comput Linguist Proc Conf (Long Pap 2:231–236. https://doi.org/10.18653/v1/P17-2036
192. Tu Z, Lu Z, Yang L, et al (2016) Modeling coverage for neural machine translation. 54th Annu Meet Assoc Comput Linguist ACL 2016 - Long Pap 1:76–85. https://doi.org/10.18653/v1/p16-1008
193. Upadhya BA, Udupa S, Kamath SS (2019) Deep Neural Network Models for Question Classification in Community Question-Answering Forums. 2019 10th Int Conf Comput Commun Netw Technol ICCCNT 2019 6–11. https://doi.org/10.1109/ICCCNT45670.2019.8944861

194. Vasisht S, Tirthani V, Eppa A, et al (2022) Automatic FAQ Generation Using Text-to-Text Transformer Model. 2022 3rd Int Conf Emerg Technol INCET 2022 1–7. https://doi.org/10.1109/INCET54531.2022.9823967
195. Vaswani A, Shazeer N, Parmar N, et al (2017) Attention Is All You Need. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, arXiv:1706.03762v5. CA, USEA
196. Vedantam R, Zitnick CL, Parikh D (2015) CIDEr: Consensus-based image description evaluation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 4566–4575
197. Vijayakumar AK, Cogswell M, Selvaraju RR, et al (2018) Diverse Beam Search: Decoding Diverse Solutions from Neural Sequence Models. In: 32nd AAAI Conference on Artificial Intelligence, AAAI 2018. pp 1–16
198. Vijayakumar AK, Cogswell M, Selvaraju RR, et al (2016) Diverse Beam Search: Decoding Diverse Solutions from Neural Sequence Models. In: 32nd AAAI Conference on Artificial Intelligence, AAAI 2018. pp 7371–7379
199. Vinyals O, Toshev A, Bengio S, Erhan D (2015) Show and tell: A neural image caption generator. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 3156–3164
200. Wang Q, Li B, Xiao T, et al (2019) Learning Deep Transformer Models for Machine Translation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 1810–1822
201. Wang C, Yang H, Bartz C, Meinel C (2016) Image captioning with deep bidirectional LSTMs. MM 2016 - Proc 2016 ACM Multimed Conf 988–997. https://doi.org/10.1145/2964284.2964299
202. Wang P, Yang A, Men R, et al (2022) OFA: Unifying Architectures, Tasks, and Modalities Through a Simple Sequence-to-Sequence Learning Framework. arXiv. http://arxiv.org/abs/2202.03052
203. Wang W, Yang N, Wei F, et al (2017) Gated Self-Matching Networks for Reading Comprehension and Question Answering. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 189–198
204. Wang T, Yuan X, Trischler A (2017) A Joint Model for Question Answering and Question Generation. arXiv:170601450v1
205. Welleck S, Kulikov I, Kim J, et al (2020) Consistency of a recurrent language model with respect to incomplete decoding. EMNLP 2020 - 2020 Conf Empir Methods Nat Lang Process Proc Conf 5553–5568.https://doi.org/10.18653/v1/2020.emnlp-main.448
206. Weston J, Chopra S, Bordes A (2015) Memory Networks. 3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc 1–15
207. Wilt C, Thayer J, Ruml W (2010) A comparison of greedy search algorithms. In: Proceedings of the 3rd Annual Symposium on Combinatorial Search, SoCS 2010. SoCS 2010, pp 129–136
208. Wiseman S, Shieber SM, Rush AM (2018) Learning Neural Templates for Text Generation. Proc 2018 Conf Empir Methods Nat Lang Process EMNLP 2018 3174–3187. https://doi.org/10.18653/v1/d18-1356
209. Wolf T, Sanh V, Chaumond J, Delangue C (2019) TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents. arXiv
210. Wołk K, Koržinek D (2017) Comparison and adaptation of automatic evaluation metrics for quality assessment of re-speaking. Comput Sci 18:129–144. https://doi.org/10.7494/csci.2017.18.2.129
211. Woodsend K, Lapata M (2010) Automatic generation of story highlights. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pp 565–574
212. Wu Y, Hu B (2018) Learning to Extract Coherent Summary via Deep Reinforcement Learning. In: The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18) Learning. Association for the Advancement of Artificial Intelligence, pp 5602–5609
213. Wu J, Ouyang L, Ziegler DM, et al (2021) Recursively Summarizing Books with Human Feedback
214. Wu Y, Schuster M, Chen Z, et al (2016) Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arXiv:160908144 1–23
215. Xiao Y, Wu L, Guo J, et al (2022) A Survey on Non-Autoregressive Generation for Neural Machine Translation and Beyond. arXiv 00:1–25. http://arxiv.org/abs/2204.09269
216. Xie Z (2017) Neural Text Generation: A Practical Guide. http://arxiv.org/abs/180307133 1–21
217. Xie Y, Le L, Zhou Y, Raghavan VV (2018) Deep Learning for Natural Language Processing. Handb Stat 38:317–328. https://doi.org/10.1016/bs.host.2018.05.001
218. Xing C, Wu W, Wu Y, et al (2017) Topic aware neural response generation. 31st AAAI Conf Artif Intell AAAI 2017 3351–3357

219. Xiong C, Merity S, Socher R (2016) Dynamic memory networks for visual and textual question answering. 33rd Int Conf Mach Learn ICML 2016 5:3574–3583
220. Xu K, Ba JL, Kiros R, et al (2015) Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In: International Conference on Machine Learning. JMLR: W&CP
221. Xu W, Li C, Lee M, Zhang C (2020) Multi-task learning for abstractive text summarization with key information guide network. EURASIP J Adv Signal Process 2020:16. https://doi.org/10.1186/s13634-020-00674-7
222. Yamada K, Knight K (2001) A syntax-based statistical translation model. 523–530.https://doi.org/10.3115/1073012.1073079
223. Yan Z, Duan N, Bao J, et al (2016) DocChat: An information retrieval approach for chatbot engines using unstructured documents. 54th Annu Meet Assoc Comput Linguist ACL 2016 - Long Pap 1:516–525. https://doi.org/10.18653/v1/p16-1049
224. Yang Q, Huo Z, Shen D, et al (2019) An End-to-End Generative Architecture for Paraphrase Generation. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 3130–3140
225. Yang W, Xie Y, Lin A, et al (2019) End-to-End Open-Domain Question Answering with BERTserini. https://doi.org/10.18653/v1/N19-4013
226. Yao T, Pan Y, Li Y, Mei T (2017) Incorporating copying mechanism in image captioning for learning novel objects. Proc - 30th IEEE ConfComput Vis Pattern Recognition, CVPR 2017 2017-Janua:5263–5271. https://doi.org/10.1109/CVPR.2017.559
227. Yao L, Peng N, Weischedel R, et al (2019) Plan-and-Write: Towards Better Automatic Storytelling. In: Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, pp 7378–7385
228. Yao K, Zweig G, Peng B (2015) Attention with Intention for a Neural Network Conversation Model. ArXiv, abs/151008565 1–7
229. Yin C, Qian B, Wei J, et al (2019) Automatic Generation of Medical Imaging Diagnostic Report with Hierarchical Recurrent Neural Network. In: 2019 IEEE International Conference on Data Mining (ICDM). IEEE, pp 728–737
230. You Q, Jin H, Wang Z, et al (2016) Image captioning with semantic attention. Proc IEEE Comput-Soc Conf Comput Vis Pattern Recognit 2016-Decem:4651–4659. https://doi.org/10.1109/CVPR.2016.503
231. Yu L, Zhang W, Wang J, Yu Y (2017) SeqGAN : Sequence Generative Adversarial Nets with Policy Gradient. In: 31st AAAI Conference on Artificial Intelligence. AAAI, pp 2852–2858
232. Yu W, Zhu C, Li Z et al (2022) A Survey of Knowledge-Enhanced Text Generation. ACM Comput Surv 1:1–44. https://doi.org/10.1145/3512467
233. Yuan X, Wang T, Gulcehre C, et al (2017) Machine comprehension by text-to-text neural question generation. arXiv 15–25. https://doi.org/10.18653/v1/w17-2603
234. Zeiler MD (2012) ADADELTA: An Adaptive Learning Rate Method. ArXiv: 12125701
235. Zhang S, Dinan E, Urbanek J, et al (2018) Personalizing dialogue agents: I have a dog, do you have pets too? ACL 2018 - 56th Annu Meet Assoc Comput Linguist Proc Conf (Long Pap 1:2204–2213. https://doi.org/10.18653/v1/p18-1205
236. Zhang T, Kishore V, Wu F, et al (2020) BERTScore: Evaluating Text Generation with BERT. arXiv:190409675 1–41
237. Zhang X, Lapata M (2017) Sentence Simplification with Deep Reinforcement Learning. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 584–594
238. Zhang L, Sung F, Liu F, et al (2017) Actor-Critic Sequence Training for Image Captioning. ArXiv: 170609601
239. Zhang J, Tan J, Wan X (2018) Towards a Neural Network Approach to Abstractive Multi-Document Summarization. arXiv:180107704
240. Zhang J, Zhao Y, Saleh M, Liu PJ (2019) PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. 37th IntConf Mach Learn ICML 2020 PartF16814:11265–11276
241. Zhao Y, Ni X, Ding Y, Ke Q (2018) Paragraph-level neural question generation with maxout pointer and gated self-attention networks. Proc 2018 Conf Empir Methods Nat Lang Process EMNLP 2018 3901–3910. https://doi.org/10.18653/v1/d18-1424
242. Zhou H, Huang M, Zhang T, et al (2018) Emotional chatting machine: Emotional conversation generation with internal and external memory. 32nd AAAI Conf Artif Intell AAAI 2018 730–738
243. Zhou H, Huang M, Zhang T, et al (2018) Emotional chatting machine: Emotional conversation generation with internal and external memory. Thirty-Second AAAI Conf Artif Intell 730–738

244. Zhou X, Wang WY (2018) MOJITALK: Generating Emotional Responses at Scale. In: Proceedings of the 56th Annual Meeting ofthe Association for Computational Linguistics (Long Papers). Assoc Comput Ling. Melbourne, Australia, pp 1128–1137
245. Zhou Q, Yang N, Wei F, et al (2017) Neural Question Generation from Text: A Preliminary Study. arXiv:170401792v3 [csCL]