A Comprehensive Review of Deep Learning Architectures for Task specific Analysis

Sahil Bhardwaj
School of Engineering and Technology, KR Mangalam University, Gurugram, India
Rishav Kumar
School of Engineering and Technology, KR Mangalam University, Gurugram, India,
Ayush Verma
School of Engineering and Technology, KR Mangalam University, Gurugram, India
Dr. Solanki Gupta
School of Engineering and Technology, KR Mangalam University, Gurugram, India

Abstract

Deep learning has truly changed the game across numerous fields, reshaping how we tackle complex challenges by providing highly precise and efficient solutions tailored to particular needs. Just picture a system that can create text, summarize information, translate languages, classify data, answer questions, and even reason—deep learning makes all of this a reality. In this review, we took a closer look at different deep learning architectures and see how they drive these various applications. We analysed the past studies and reveal the datasets that power these models, as well as the design principles that influence their performance. Throughout this we emphasized the strengths that set these architectures apart, along with the limitations that pose challenges to their effectiveness. This review acts as a guide for researchers, practitioners, and industry professionals, helping them choose and adapt the right deep learning models for specific tasks.

Keywords: *Deep Learning, Deep Learning Architectures, Task Specific Review, Systematic Review*

1. Introduction

Deep learning is a subfield of machine learning and artificial intelligence that attempts to model the way humans learns from information to extract patterns for decision-making. Neural networks that have multiple layers are used to deal with large data. Through these layers, deep learning models can learn complicated Patterns and representations, making them efficient for many applications, including image recognition, natural language processing (NLP), and speech recognition. The very idea that deep learning embodies is that it allows machines to learn directly from data in their raw form, such as an image with its associated text or audio, without human intervention in feature extraction. This process favours neural networks made up of nodes or interconnected neurons, which adjust their weight and biases during training to minimize errors and increase accuracy. What holds the transition from classical machine learning to deep learning is that the classical machine is doing manual feature extraction, while deep networks are learning directly from raw data. Classical machine learning works great with small datasets but often struggles to infer on complicated patterns, while deep learning works exceptionally well with large datasets, achieving reasonable accuracy for image recognition andNLP. On the other hand, deep learning does require the use ofGraphics Processing Units (GPUs) orTensor Processing Units (TPUs) for processing power, whereas traditional machine learning could run just fine on standardCentral Processing Units(CPUs). Traditionally machine learning methods are applied to structured data, while deep

DOI: https://doi.org/10.5281/zenodo.15110559

learning suits unstructured data applications. This evolution has come into place under the influence of computing power becoming inexpensive, the existence of big data, and user-friendly frameworks such as TensorFlow and PyTorch. Task-specific analysis in deep learning is important since tasks have individual architectural needs for achieving the best performance. Text generation, for instance, necessitates contextual coherence and flow, so Transformers [99] such as Generative Pre-trained Transformer (GPT)[100] are best suited because they process sequential data. Summarization is about extracting the gist of a text, where sequence-to-sequence models with attentionsuch as Text-to-Text Transfer Transformer (T5)[101] orBidirectional and Auto-Regressive Transformers(BART)[102] are used to emphasize significant input segments. Translation needs precise language mapping, which is strength of encoder-decoder architectures. Classification is aided by less complex architectures such as Convolutional Neural Networks(CNNs) or fine-tuned Bidirectional Encoder Representations from Transformers (BERT)[103] models for effective feature extraction. Question answering (QA) requires contextual awareness to provide accurate answers, utilizing models such as BERT with attention. Reasoning requires logical conclusions and multi-step processes, necessitating sophisticated models such as GPT-4 [104] with memory layers. Adapting architectures to task requirements provides improved performance and more accurate outcomes. Deep learning applications in any specific task encounter challenges such as limited availability of data and lack of quality annotation, which leads to problems in model training and generalization. While complex architectures prevent overfitting at times, highly skilled regularization may be demanded. These models are often black boxes, making them hard to interpret; interpretability is critical for sensitive tasks like healthcare. Generalization is difficult, requiring intense fine-tuning and transfer learning. Computational demands are high, thus increasing the cost and energy. Other ethical concerns include the biases embedded in them, which may lead to failure in achieving fair outcomes. Real-time tasks face problems caused by latency, making their deployment in the interactive environment harder. Besides, all these challenges require model design, data preparation, and constant monitoring to be addressed. A review paper that systematically takes into account these questions, datasets employed, the rationale behind their design, and the pros and cons of various models would provide valuable insights into problems related to task-specific deep learning applications.Therefore, the major objectives of this study are to explore:

- The deep learning architectures that are most commonly used across different tasks.
- The datasets utilized in these studies and the principles behind their design.
- The strengths and weaknesses of various models.

To start the review, we identified key real-world applications of deep learning, including text generation, text classification, reading comprehension, summarization, reasoning, translation, and question answering, as fundamental tasks for analysis.By looking at the functionality and performance of model architectures with respect to these tasks, this review would help us understand model complexity and overfitting issues and reach suggestions for reasonable regularization strategies. An evaluation of strong and weak points of different models would help in proposing interpretable and generalizable architectures by reducing the behavior of a deep learning model as a black box. It would also offer guidance for data quality and availability by addressing the most efficient datasets and consequent

DOI: https://doi.org/10.5281/zenodo.15110559

drawbacks, thereby leading to better strategies for dataset selection and augmentation. Further, this review study would serve as the perfect guide forunderstandingthe efficient model for different task-specific deep learning applications.

2. Systematic Review& Analysis

To provide a detailedoverview of deep learning architectures tailored for specific tasks,we decided to use a systematic review methodology. We selected the seven popular tasks namely Reading Comprehension, Translation, Summarization, Question Answering, Reasoning, Generation, and Classification. For data collection, we systematically searched "*Google Scholar*" Database using a combination of keywords like "*deep learning architectures", "task-specific applications, "text generation", "text classification", "reading comprehension", "summarization", "reasoning", "translation"* and *"Question Answering"*. This method offers a structured way to analyze existing literature, emphasizing the identification, selection, and synthesis of studies that enhance our understanding of deep learning applications (Figure 1).A thorough analysis of individual task is elaborated in the following subsections,

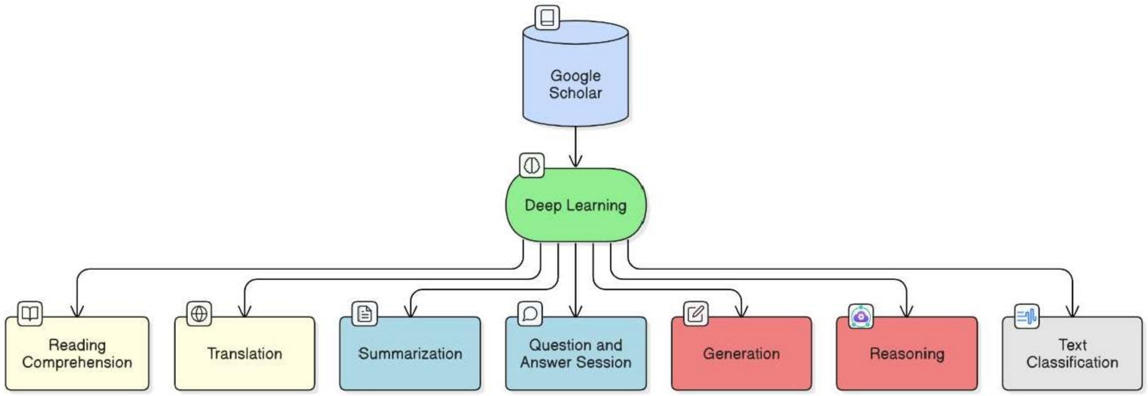

Figure 1: Research Design and methodology

2.1 Task 1-Reading Comprehension

Reading comprehension (RC) is an assignment brought forth to measure the extent to which a machine is capable of interpreting natural languages by having the machine respond to questions about a presented context, and it has the power to change the manner in which humans and machines communicate with one another. The application of deep learning expands over various industries that copy human reading and understanding abilities. Healthcare, Education, Legal and compliance, Finance, News Media and Government are some of the popular examples of this [1]. It also increases productivity and accuracy in reviewing legal documents and ensuring compliance, saving time and reducing human error [2].

Before the commencement of deep learning, traditional reading comprehension systems were used which relied on rule-based approaches and shallow machine learning techniques. Rule-based approaches were based on predefined rules and heuristics and this might rely on keyword or simple algorithms to find out and extract essential information from text.Shallow machine learning approaches used basic machine learning techniques such as Support Vector Machines (SVMs), Decision Trees, and Naive Bayes classifiers. Features like word frequencies, n-grams, and syntactic structures were often used to represent text. There were limitations of these pre-deep learning approaches. Traditional methods could not fully understand the context or subtle nuances of language, which is crucial for reading comprehension. Rule-based systems were inflexible, while shallow machine learning

DOI: https://doi.org/10.5281/zenodo.15110559

ISSN No- 2584-2706

methods relied heavily on handcrafted features that didn't generalize well across tasks. As the amount of text data grew, these approaches struggled to scale and didn't perform well in comparison to human-level comprehension. Deep learning revolutionized reading comprehension by offering models that could automatically learn from huge amounts of data and adapt to various language complexities. The major turning point came with the introduction of Recurrent Neural Network (RNNs), Long Short-Term Memory (LSTM)[3] networks, and later, Transformer-based models. It has its limitations too. It requires large datasets and high computational power for training. It can inherit biases from the training data, leading to unfair or discriminatory outcomes. A detailed overview of various models used in RC, its limitations and key takeaways are provided in Table 1.

Table 1: Summary of Popular Studies on Reading Comprehension Task

Models	Datasets	Key takeaways	Limitations	References
BERT, RoBERTa DistilBERT, ALBERT	ReClor	1. ALBERT is the best model in this paper, so far. 2. Polytuplet loss improves accuracy by 5.6%-11.7% over baseline models like ALBERT, BERT, and DistilBERT.	1. The comparison is limited to baseline models, without evaluating techniques. 2. Scalability to large datasets or real-world tasks remains unaddressed.	**[5]**
T5 base model BART base model GPT-2 model	Fairytale QA Corpus Textbook Question Answering (TQA) dataset	1. The paper compares different neural architectures for automatic question generation based on reading comprehension passages 2. Highlights the strengths and weaknesses of various Question Generation models.	1. The paper focuses on a narrow set of models, lacking comparison with a broader range of question generation techniques. 2. The evaluation metrics used might not fully capture the complexity of question quality.	**[6]**
Stanford AR GA Reader	Who-Did-What (WDW) Children's Book Test (CBT)	1. The paper compares word embedding techniques like GloVe, Word2Vec and fastText for reading comprehension tasks. 2.Embedding effectiveness varies based on task and dataset.	1. Focus is less on word embedding models. 2. The paper doesn't provide detailed insights into why some embeddings perform better than others.	**[7]**

DOI: https://doi.org/10.5281/zenodo.15110559

RCNs BERT GPT RNN CNN	CNN &Daily Mail CBT LAMBADA CLOTH RACE SQuAD	1. The paper proposes a neural network-based model that reads and understands a passage to answer questions without requiring task-specific feature engineering. 2. It utilizes attention mechanisms to focus on relevant parts of the text, improving the model's comprehension ability.	1. The model's achievement heavily relies on the quality and size of the training data. 2. The model requires significant computational resources, especially for large-scale datasets and complex attention mechanisms.	**[8]**
BERT RoBERTa Cross-Document Reasoning Models Textual Entailment Models	TriviaQA Web DuReader	1. The paper introduces a model that performs reading comprehension across multiple documents, capturing information from diverse sources to answer questions. 2. Attention mechanisms have been used to attention the model's reasoning process at the most relevant parts of the files.	1. The method includes complicated architectures that can be computationally expensive, requiring sizable resources for schooling and inference, particularly with huge file sets. 2. The model doesn't provide fine-grained control over which documents or pieces of information are prioritized in the reasoning process.	**[9]**
Co-match BERT	MCTest CNN/Daily Mail RACE	1. BERT performed higher accuracy compared to the Co-match model on the Vietnamese corpus. 2. It targets multiple-choice reading comprehension questions, where the model selects the most appropriate answer based on the given passage.	1. The study is tailored to Vietnamese, limiting its applicability to other languages with different linguistic structures. 2. The paper primarily compares a few deep learning models (RNNs, LSTMs, and BERT), without considering a broader range of models or alternative architectures.	**[10]**
T5 BERT	DROP	1. The proposed method demonstrates significant improvements over	1. Performance is reliant on the availability of annotated sub-	**[11]**

DOI: https://doi.org/10.5281/zenodo.15110559

		baseline models, achieving higher F1 scores on the hard subset of the DROP dataset. 2. A single model is used for both question decomposition and reading comprehension, simplifying the architecture.	questions, and weak supervision can only partially alleviate the data limitation. 2. The success of the model relies on the accuracy of the question decomposition process, which remains a challenging task.	
Bi-GRU Encoder	CNN/Daily Mail	1. Improves query-document interaction for improved answer selection. 2. Surpasses state-of-the-art models on CNN/Daily Mail and CBTest datasets.	1. The system can misunderstand ambiguous requests. 2. Attention mechanisms incurcomputation cost.	**[12]**
LSTMs	SQuAD	1. Splits MRC into Cloze-fashion, multi-preference, span-prediction, and free-form question answering. 2. Pre-educated models (BERT, GPT,) outperform baseline strategies in contextual comprehension.	1. Difficulty in dealing with lengthy text passages, resulting in loss of contextual pertinence. 2. Needs huge-scale labelled datasets to prevent overfitting.	**[13]**
GPT-FT	COSMOS QA	1. The paper introduces a model that integrates contextual commonsense knowledge to improve machine reading comprehension, enabling better understanding beyond explicit information in the text. 2. The model tailoring knowledge application to the specific reading passage and question.	1. The model might struggle to generalize to very diverse or uncommon knowledge that is not well-represented in the commonsense knowledge base. 2. Integrating contextual commonsense reasoning adds computational complexity, which can slow down training and inference times.	**[14]**

2.2 Task 2-Translation

Translation deals with transforming information from one language to another. The primary objective is to automatically

DOI: https://doi.org/10.5281/zenodo.15110559

translate text from one language into another via deep learning models. The

translation is an example of sequence-to-sequence learning, wherein both the input and output are word sequences. The translation has a wide range of applications in diverse fields. Computer aids such as Microsoft Translator and Google Translate have made cross-lingual communication simpler. These processes utilize neural machine translation based on deep learning systems to translate text, audio, and images instantly[15]. Companies offer customer services in a variety of languages instantly by using translational model[16]. Translational model used in hospitals and clinics assist health professionals to interpret with patients communicating in other languages thus eliminating miscommunication [17]. Prior to the onset of deep learning, machine translation and language processing involved rule-based methods and statistical approach. This method was dependent upon linguistic rules as well as a dictionary for interpretation of text into languages. Rule-Based system made use of the rules of grammar to transfer the words and expressions of a word from one word to another word. Statistical Machine Translation arrived later in the 1980s and was based on probability models to figure out the most appropriate translation of the given sentence against texts that have been translated among languages. It has some limitations to it. The translation was not up to the point and was always grammatically improper, particularly to long sentences. Rule-basedmethod took hard work and didn't scale very efficiently to new words. Statistical techniques had competitors regarding handling new terminology and uncommon languages. With the onset of deep learning, issues were mainly vanquished via Neural Machine Translation, where an enormous neural net is employed in modelling translation of complete sentences. NMT, particularly with the arrival of the Transformer model, really improved translation to a large degree by understanding the context and dependency over long distances. A detailed overview of models utilized in Translation, its limits and important lessons are summarized in Table 2.

Table 2: Summary of Popular Studies on Translation Task

Models	Datasets	Key Takeaways	Limitations	References
Deep Transition RNNs Stacked RNNs Bi-Deep RNN Architecture	WMT'15 English-German (En-De) Byte Pair Encoding (BPE) WMT'14 English-French (En-Fr)	1. The research proves that extra profound architectures, specifically deep transition RNNs and stacked RNNs, decorate neural gadget translation (NMT) accuracy. 2. Increasing version intensity (up to eight layers) assists in taking snap shots greater state-of-the-art linguistic styles, enhancing translation accuracy.	1. Training deep NMT models is time-consuming and computationally steeply-priced, making them much less scalable for large-scale programs. 2. Although the consequences are encouraging, the generalizability of the method to different language pairs or domain names is but to be hooked up.	**[18]**
Vanilla Seq2Seq Model	WMT'15 English-German Task	1. Using deeper fashions, mainly with interest mechanisms,	1. The Transformer and bidirectional models required longer	**[19]**

DOI: https://doi.org/10.5281/zenodo.15110559

RNNs GRUs LSTMs		leads to seriously better translation excellent. 2.The models showed study overall performance across both English-German and English-French, confirming their versatility.	education times. 2.Taken a look at frequently specializes duties in English-German and English-French translation duties, which can be every pretty excessive-resource language pairs.	
Transformer Models DLCL	WMT'16 English-German (En-De) Task NIST'12 Chinese-English (Zh-En-Small) Task WMT'18 Chinese-English (Zh-En-Large) Task	1.Deepening the Transformer model results in improved translation quality, particularly when utilizing dynamic layer combinations (DLCL). 2.Employing a dynamic combination of layers in the encoder and decoder enhances translation quality over conventional fixed-layer models.	1.Deep models such as Transformer-Deep are highly computationally demanding and may prove difficult to train on regular hardware because of processing and memory constraints. 2.Because of the model's depth, it is hard to process a complete batch on one GPU.	**[20]**
DTMT Vanilla Encoder-Decoder Transformer Model	WMT'14 English-German (En-De) IWSLT'15 Multi30k	1.Bloating the Transformer model results improves translation, particularly when utilizing dynamic layer combinations DLCL. 2. Using a dynamic combination of layers within the encoder and decoder pairs well with translation spectacularity compared to traditional constant-layer models.	1.Deep models such as Transformer-Deep can be challenging to train on regular hardware because of processing and memory limitations. 2.Due to the version's severity, it's miles hard to prepare a complete set on a solitary GPU.	**[21]**
Vanilla Encoder-Decoder Models BERT GPT CNNs LSTM GRU	WMT IWSLT Flickr30k and COCO	1.Pretrained models can be transferred to NMT tasks and thus are very effective for low-parallel-data languages.2.Pretraining language models like BERT and GPT has been shown to enhance performance on the	1.NMT models need large data sets to effectively train. 2.NMT models can inherit and pass on biases in the training data, resulting in biased or unfair translations. This is a major issue in	**[22]**

DOI: https://doi.org/10.5281/zenodo.15110559

		task of translation.	applications where fairness and neutrality are important.	
Seq2Seq Transformer Models	American Sign Language (ASL)	1. These pairs scored better than different styles, with better BLEU rankings on the GSL dataset. 2. Higher models proved robust ability on much less controlled ASL and CSL datasets, showing versatility.	1. Performance can fluctuate with less managed facts units because of variability in signing patterns and recording environments. 2. Advanced models, such as transformers, necessitate substantial computational resources throughout both the training and inference phases.	**[23]**
Neural Machine Translation	WMT (Workshop on Machine Translation)	1. Translation Adequacy: In blind tests, CUBBITT performed better than professional human translators in maintaining the original meaning of the text. 2. Fluency Comparison: Human translations were graded as more fluent.	1. Fluency Gap: There is still a narrow fluency gap between CUBBITT's outputs and those of human professionals. 2. Domain Specificity: The performance of the system has been mostly tested on news articles, and its performance on other domains or language pairs might need to be evaluated.	**[24]**
Global Memory Module	IWSLT	**1.** The model presented here greatly enhances translation quality by efficiently capturing and making use of both local and global context information. **2.** Integrating grammatical dependencies with the attention mechanism enhances context representation, resulting in more precise translations**.**	1. The "end-to-end" design of deep learning models may result in poor interpretability of learning outcomes, making it hard to know the decision-making process. 2. Although the model is good on the IWSLT dataset, its generalization to other datasets or real-world use needs to be verified.	**[25]**

DOI: https://doi.org/10.5281/zenodo.15110559

ISSN No- 2584-2706

RBMT and SMT	United Nations Parallel Corpus	1. The move from rule-based and statistical models to neural models has tremendously improved the quality of translations. 2. Combination of various MT paradigms can effectively cope with particular issues, like low-resource languages	1. NMT systems need huge quantities of good quality parallel data, and such parallel data may not exist for all language pairs. 2. Even with progress, getting high-quality translations for low-resource languages is still a major issue.	**[26]**
Transformer-Based NMT	Translation Corpus (TC)	1. The multi-challenge mastering method enhances MAP by using 16% in comparison to the baseline transformer. 2. Evades overfitting to TC terminology, producing translations relevant to each corpora.	1. The model works well but keeps quite low MAP rankings as a result of having few education epochs and dataset. 2. Speaks to gaining access to a retrieval corpus (RC) index, which hinders schooling index.	**[27]**

2.3 Task 3-Summarization

Summarization is the process of creating a brief and coherent summary of a longer text without losing its core meaning. It is the process of extracting important information and removing unnecessary details to create a shorter version that still maintains the key points of the original text. Summarization methods, especially in deep learning, have numerous applications in real-world situations. For example, news aggregation sites, where short summaries of long news stories are given to readers. This provides faster reading of news while ensuring the vital content[28]. Summarization is employed for assisting researchers, students, and professionals in maintaining pace with scientific literature in massive quantities. Summarization platforms can summarize research papers into the main findings, abstracts, or even a summary of a paper, with the aim of saving time and making research easier to access[29]. Blogging websites and social media websites utilize summarization to create short summaries of posts to enable users to quickly scan through content without the need to read entire posts[30]. Pre-deep learning summarization strategies were predominantly based on rule-based and statistical processes, including keyword extraction-based extractive summarization, sentence rank algorithms (e.g., TF-IDF), and heuristic methods that marked up salient sentences by their occurrence or location within a document. These methods had the limitation that they could not interpret the contextual or semantic nature of the content. They had difficulty in generating coherent abstracts, typically producingincomplete results, since they did not look at the more profound relationships between words or phrases. Moreover, these approaches tended to be computationally costly and failed to generalize across various languages. They also did not cope-up with complex sentence structures, synonyms, or paraphrases, which have been effectively handled by deep learning models. Deep

DOI: https://doi.org/10.5281/zenodo.15110559

learning methods for text summarization utilize strong neural network architectures, specifically sequence-to-sequence models and transformers, to produce more coherent and contextually correct summaries. Seq2Seq models, based on encoders and decoders (usually with LSTMs or GRUs), can map input text into a fixed-size representation and output a summary by predicting each word in sequence. The transformer architecture where models such as BERT, GPT, and T5 come into play, has transformed summarization by employing self-attention to tackle whole documents in parallel and capture long-range dependencies.A detailed overview of various models used in Summarization, its limitations, and key takeaways are provided in Table 3.

Table 3: Summary of popular Studies on Summarization Task

Models	Data Sets	Key takeaways	Limitations	References
STFIDF TBS	BillSum, IN-ABS and IN-EXT	1. Legal structures vary by areas, so there is a need for models evolved on jurisdiction-particular statistics to stay accurate and relevant. 2. With prison documents written in diverse languages, institutions including the European Union, there may be growing call for for fashions able to doing multilingual and pass-lingual summarization obligations.	1. There is a great scarcity of huge-scale, outstanding datasets across most jurisdictions and languages, making it hard to build sturdy legal summarization fashions. 2. Traditional assessment metrics may fail to effectively capture actual correctness and legal soundness of summaries that are vital in prisoneventualities**.**	**[31]**
RNN Extractor and Seq2Seq Extractor Cheng & Lapata Model	Reddit and AMI	1. Position Bias: Sentence function is the main responsibility that summarization models must bear. 2. Word averaging is just as good as CNNs/RNNs.	1. Performance isn't always consistent throughout domain names. 2. Models are prone to overfitting dataset-specific characteristics.	**[32]**
GoogleNet and AlexNet LSTM	YouTube	1. Deep learning outperforms conventional	1. Prevention of duplicate or Unwanted files.	**[33]**

DOI: https://doi.org/10.5281/zenodo.15110559

		approaches (CNNs, RNNs, Transformers enhance summarization). 2. Supervised models are precise but require large labelled datasets (SumMe, TVSum).	2.Preservation of meaningful and Contextually appropriate segments.	
Seq2Seq	DUC (Document Understanding Conferences)	1. Getting to know that strategies have dramatically progressed the overall performance of MDS systems. 2. The authors advocate a brand-new taxonomy classifying neural community design methods for MDS.	1. Super datasets required for powerful training of deep mastering models. 2. Deep studying algorithms for MDS tend to call for loads of computational assets, consequently less suitable for researchers with confined facilities.	**[34]**
RNNs and BERT SUM, T5, PEGASUS	Gigaword	1. It discusses the evolution of models from RNNs and LSTMs to more advanced transformer models, showing improvements in generating coherent and concise summaries. 2.ROUGE-1, ROUGE-2, and ROUGE-L are the maximum broadly used metrics to evaluate summarization pleasant	1. Traditional evaluation metrics like ROUGE may not fully capture the quality of abstractive summaries, especially when it comes to factual accuracy and coherence. 2.The models often face challenges when dealing with out-of-vocabulary (OOV) words, which can negatively impact summary quality, especially in specialized domains	**[35]**
Attention Mechanisms	Pre-Training and Fine-Tuning	1. The method below attention utilizes deep fashions that are trained extensively on big	1. The fulfilment of the approach is largely dependent on the presence of first rate, large-scale datasets for pre-training and	**[36]**

DOI: https://doi.org/10.5281/zenodo.15110559

		datasets through pre-training and excellent-tuned with domain-particular net pages. 2. The technique indicates area adaptability, effectively moving to extraordinary net domain names by way of exceptional-tuning pre-trained fashions with little domain-unique facts.	satisfactory-tuning. 2.Although the technique contains extraordinary fields, a few specialized domains with precise terminologies or frameworks may want in addition adjustment.	
GPT-2 BERT	Udacity Lecture Transcripts	1. BERT plays higher than traditional tacticsin summarizing lectures. 2.K-Means clustering lets in for dynamic adjustment of summary duration in line with consumer desire.	1. Difficulty with prolonged lectures (a hundred sentences may lose context). 2. Computationally highly priced (BERT could be very useful resource-in depth).	**[37]**
Coverage Models	Gigaword	1. Integration of attention mechanism and pointer-generator network has enormously enhanced the generated summary's quality. 2. Having access to large and high-quality datasets is imperative for training good summarization models.	1. Abstractive models have the possibility of creating information that does not exist in the source material, creating possible inaccuracy. 2. Advanced deep learning models take a lot of computational resources, and this may not be readily available to all researchers**.**	**[38]**
Graph-Based	TAC (Text	1. Dependent on	1. It may not be	**[39]**

DOI: https://doi.org/10.5281/zenodo.15110559

Methods Template-Based Methods	Analysis Conference)	the selection of pre-existing sentences; While using is less complex to use, it can be repetitive and incompatible. 2.Creates new sentences that forms the content of the text; Greater is flexible but more difficult because it asks for herbal language era's abilities.	readable to see that they can be based on literal sentences, which can also bring about excesses and lack of glide. 2.Sophisticated natural language production strategies and a large amount of education fabrics require; They are also interrupted with the help of problems in preserving the data up-to-date	
BERT BiGRU	IMDb Reviews	1. Model design selection must conform to the inherent nature of the text classification task and consider the size of the sequence and the significance of the reference.	1. It takes a large amountof computational resources, especially to train and inferior transformer-based architecture. 2. Model can overfit training data, especially when working with small datasets.	**[40]**

2.4 Task 4-Question and Answer Session

A question answer (QA) session is the step where a machine model is asked to comprehend a provided text (or set of texts) and answer particular questions accordingly based on that. QA systems have numerous applications in everyday situations.Virtual assistance such as Amazon Alexa, Google Assistant, and Apple Siri employ QA systems to respond to user queries and carry out actions based on input in natural language. Chatbots employed in customer support systems also depend on QA models to aid users[41]. Inthe medical sector, QA systems are applied for medical question answering, assisting doctors, medical students, and even patients to receive correct information from medical literature, clinical guidelines[42]. QA systems can assist media organizations in automating the process of summarization and extracting salient information from news stories. They can be utilized to

respond to questions regarding events, individuals and issues reported in the news[43]. Prior to deep learning QA systems, there existed a range of pre-deep learning QA methods. These methods mostly depended on rule-based techniques, conventional machine learning, and statistical models. These pre-deep learning techniques formed the foundation in the development of contemporary QA systems and formed the basis of subsequent more complex methodologies like deep learning, which subsequently displaced or substituted many of these methods using more advanced models that better handle context and semantics. Pre-deep learning methods forQA systems were very limited. These systems tended to be inflexible,

DOI: https://doi.org/10.5281/zenodo.15110559

involving manual rule definition or pre-defined knowledge bases and did not do well with complex or vague queries. They did not have deep contextual understanding, did not handle word ambiguity well and were unable to handle synonyms, paraphrases, or complicated sentence structures well. These problems eventually resulted in the emergence of deep learning models, which were capable of coping with natural language variability more effectively and offering more accurate, adaptive QA solutions.The initial models such as RNN and LSTM networks were employed for sequential text processing, while attention models enabled models to concentrate on significant parts of the text [44]. Transformer models, specifically BERT [45], brought about bidirectional context comprehension, greatly transforming performance over QA tasks. Models such as T5 [46] have also improved QA by solving tasks as text-to-text or producing answers outright. These models have performed well in open-domain QA using large pre-trained models and fine-tuning them for specific tasks. Their capacity to capture local and global dependencies in the text has placed them at the state-of-the-art for QA systems.A detailed overview of various models used in QA, their limitations, and key takeaways are listed in Table 4.

Table 4: Summary of popular Studies on Question-Answering Task

Models	Datasets	Key takeaways	Limitations	References
BERT Hybrid Models	MEDIQA	1.Deep mastering has immensely improved the performance of scientific QA systems to recognize and bring natural language more efficiently. 2.Blending similar approaches, such as retrieval-based solely and understanding-based solely models, is likely to yield improved outcomes than a single approach.	1.Medical vocabulary and the complexity of medical language are challenging for herbal language processing fashions. 2.It is difficult to quantify overall performance of medical QA systems because medical recommendation is subjective and there may be variations in correct solutions.	**[47]**
Dataset-Specific Optimized Models Inter-Sentences Architecture	TREC QA	1.Architectures that model question-answer interactions at earlier stages (word or subsquence level) work better. 2.The paper brings to the attention that different architectures provide different performance, and there is a focus on the correct choice based on the application to be addressed.	1.The performance is only measured for the TREC QA dataset and therefore may restrict the generality of the results to other datasets or domains. 2.Four provided architectures comprise the scope of the research, although potential models different from them are not taken into account.	**[48]**

DOI: https://doi.org/10.5281/zenodo.15110559

FCNs and LSTM	PASCAL VOC	1. Deep learning algorithms have greatly enhanced the accuracy of semantic segmentation operations compared to conventional techniques. 2. The encoder-decoder architecture, i.e., the network architecture, is crucial in maintaining the equilibrium between localisation accuracy and context capture.	1. A model can be trained on a particular dataset but can be poor in another scenario or domain. 2. Model will require domain adaptation methods or other training sets.	**[49]**
Autoencoders DBNs	Electronic Health Records	1. Deep learning algorithms have much enhanced the accuracy of disease diagnosis and prognosis. 2. Deep learning allows heterogeneous data sources to be integrated, offering an integrated view of patient health.	1. Utilization of sensitive patient information poses concerns related to privacy and security**.** 2. The accuracy of these models highly relies on the quality as well as the availability of data, which in healthcare applications can prove to be a limiting factor**.**	**[50]**
Information Retrieval and Deep Neural Network	WikiQA	1. The discipline has moved from the classical IR-based approach to integrating deep learning methods, resulting in huge leaps in comprehending and creating correct answers. 2. Merging IR and DNN techniques has the potential to capitalize on both approaches.	1. Deep learning model training and deployment require immense computational resources, which might be out of reach for some organizations.	**[51]**
GRU Dynamic Memory Networks	MCTest Dataset (Microsoft)	1. Attention-based models assist in extracting information pertinent to answering questions. 2. Sequence-to-sequence models work well to produce multi-	1. Training certain models, such as Dynamic Memory Networks, is computationally costly. 2.Performance is constrained by fixed memory sizes on long-	**[52]**

DOI: https://doi.org/10.5281/zenodo.15110559

		word responses.	context tasks.	
GANs	Genomic Databases	1. Deep learning algorithms can enhance the accuracy of disease detection and diagnosis. 2. Models allow for customized treatment protocols based on specific patient information.	1. The management of sensitive patient data requires strict privacy practices. 2. Deep learning models tend to be "black boxes" and difficult to interpret their decision-making processes.	**[53]**
MRC	SQuAD (Stanford Question Answering Dataset)	1. The incorporation of deep learning methods, particularly neural networks, has greatly enhanced the performance of open-domain QA systems. 2. A range of models, such as MRC, knowledge-based, and hybrid models, serve various aspects of QA tasks, and the choice of suitable models depends on the application.	1. Certain models struggle to scale to large datasets or process the enormous amount of information present in open-domain environments. 2. Models can still be challenged by grasping subtle contexts or unclear questions and provide the wrong answers.	**[54]**
TPRN	SQuAD	1. TPRN encodes grammar-like structures without explicit annotation. 2. Symbol-role binding enhances readability by linking words with grammatical functions.	1. Lower accuracy than BiDAF (~2% loss of F1 score). 2. Takes large computational power for training and tuning.	**[55]**
SGD Elastic Averaging SGD	TREC QA	1. Distributed deep learning speeds up training procedures. 2. Optimization algorithm performance is inconsistent; whereas certain ones such as EASGD perform well under distributed environments	1. Although improved, the speedup from increased workers is sublinear, which means returns diminish as more workers are added. 2. Distributed training brings communication overhead, which can negate the advantages of parallelism, particularly in high-	**[56]**

DOI: https://doi.org/10.5281/zenodo.15110559

ISSN No- 2584-2706

			latency environments.	

2.5 Task 5-Generation

Generation in deep learning architecture focuses on creating new content, whether it's text or even synthetic data, and usesadvanced neural networks like RNN, LSTM, Transformer, and GRUs to generate human-like text.These models help to understand grammar, context and pattern, enabling them to produce coherent and contextually relevant text. The application of deep learning expands over various industries and used into daily life making tasks easy,faster, smarter and more efficient. We can easily write an emailand can chat with virtual assistant [57].We can easily get AI generated content[58]. Chatbots likeChatgpt andGemini used in daily life providing recommendations for various purposes by generating ideas through texts. There are some writing tools exists in real world like jasper and writesonic which helps in generating ideas and news article[59]. There are AI generated voiceovers like Amazon polly, Google Text to speech which is used inconverting text to speech and speech to text. Before the existence of deep learning, traditional methods were used. Examples of these models are Rule-based system model, Template-based approach, NN-gram language model,Hidden Markov Models (HMM) ruled-based system. There are some limitations of it.These modelslacked flexibility asthey cannot create and generate dynamic text and it is difficult to update neededmanual rule.N-gram language model is used in predictive typing models in early mobile keyboards. Limitations of this model was the explosive memory requirement because of large amount of dataset used in it. So, this model cannot understand the long-termcontext. HMM used probability-based models for part-of-speech tagging and basic sentence formation and also helps in predict word. But there are certain limitations like this needed labelled dataset. After the existence of deep learning, all the problems faced by these modelshave been solved.Deep learning revolutionized text generation by introducing neural networks that learn patterns, context, and semantics automatically. Unlike traditional rule-based or statistical models, deep learning can understand context**,** generate coherent text, and adapt dynamically. N-gram model solve the problem of long-range dependent paragraph using RNN and LSTM or transformer like GPT and BERT.The problem of updating manual-based rule is also solved by neural network learn pattern. N-gram models failed with rare words or new phrases and could not generate creative or out-of-the-box text.To resolve this problem, they use word embeddings (like Word2Vec, GloVe, and Transformer embeddings) to understand word relationships and they can generate completely new, creative sentences.A detailed overview of various models used in Generation, its limitations and key takeaways is summarized in Table 5.

Table 5: Summary of popular Studies on Generation Task

Models	Datasets	Key takeaways	Limitations	References
GANs	Speech Data	1.The usage of GANs is to generate new and unique architectural shapes, going past traditional design methods.	1.Mapping 3D architectural designs into graph representations can be challenging. 2.Deep neural network training, particularly GANs, on graph-based facts calls for excessive	**[60]**

DOI: https://doi.org/10.5281/zenodo.15110559

		2.The article suggests a graph-primarily based gadget learning approach for 3D architectural layout spaces.	levels of computation.	
VAEs	Synthetic Data Output	1. The proposed structure, SenseGen, generates synthetic sensor statistics, permitting information argumentation and device mastering version education. 2.Employing deep studying models, SenseGen analyses state-of-the-art patterns in sensor readings.	1. The validity of synthesized records generated closely depends on the domain the original information comes from. 2. It continues to be challenging to assess the pleasant and usability of artificial statistics.	**[61]**
RNNs Autoencoders	Audio-Based Datasets	1. Deep learning can facilitate diverse purposes like melody composition, polyphony, accompaniment, and counterpoint. 2.Generation strategies such as single-step feedforward processes, iterative feedforward strategies, sampling strategies, and input manipulation are employed to control the music generation process.	1. The diversity-coherence trade-off in generated music is still an issue. 2. The majority of current models cannot integrate real-time user feedback to any extent.	**[62]**
Point Cloud-Based Models and Voxel-Based Models	ShapeNet and ModelNet	1. Deep learning has greatly improved the ability to create complex and sundry 3-D shapes	1. Computational models based totally on deep learning for the technology of 3-D form require substantial computational sources,	**[63]**

DOI: https://doi.org/10.5281/zenodo.15110559

		out of reach of conventional modelling. 2.The survey categorizes cutting-edge models into several classes, giving a scientific evaluation of the methodologies in the field.	which may not be less costly for everybody who's either a practitioner or a researcher. 2.The models won't generalize to new instructions of shapes, proscribing their application	
LSTM and Hybrid Models	Stock Market Data and Electricity Consumption Dataset	1. Deep models including RNNs and LSTM networks are proven to outperform conventional statistical models in the modelling of sophisticated temporal dependencies. 2. Blending deep learning models with classical forecasting techniques or other machine learningtechniques can result in better forecasting accuracy and stability.	1. Deep learning models need vast amounts of high-quality training data, which in time series applications may not always be available**.** 2. Deep learning models, if not properly regularized and validated, can overfit with small datasets**.**	**[64]**
RNNs and CNNs SaShiMi	Music Generation and Unconditional Speech Generation	1. It integrates S4 layers with a multiscale structure to enable efficient modelling of long-range dependencies in audio data. 2. Resolves S4 autoregressive generation stability by modifying parameterization, keeping it stable	*Autoregressive Instability***:** The standard S4 models are unstable during autoregressive generation and need to be parameter-tuned**.**	**[65]**

DOI: https://doi.org/10.5281/zenodo.15110559

		while generating audio.		
VAE MedGAN	MIMIC-III & Sutter EHR	1. VAEs and GANs are state-of-the-art methods in artificial data generation. 2. Synthetic privacy-preserving data ensures secure data sharing.	1. Models learn biases from actual datasets. 2. Excessive resource utilization for training generative models.	**[66]**
LSTM	ImageNet	1. Chainer introduces "Define-by-Run" execution, making deep learning models more flexible and easier to use. 2. Optimized GPU computation using CuPy for speeding up deep learning training.	1. Models will be prone to inherit real-world dataset biases. 2. Vast resource demand for training generative models.	**[67]**
BERT RoBERTa	PY150, GitHub	1. CodeXGLUE has 14 datasets for 10 programming tasks such as code search, code translation, and bug detection. **2.** Integration of pretrained models: uses CodeBERT, CodeGPT, and Encoder-Decoder as baselines.	*Not having Real-World Edge Cases: Certain* datasets are generated synthetically and lack actual real-world variations in coding.	**[68]**
RNN	WikiText-2	*Inclusion of Recurrent Neural Networks (RNNs):*The bigger architecture includes RNN layers within the Transformer model, in the hope of better capturing sequential relationships.	*Risk of Overfitting:* The bigger model, with its greater number of parameters, can be more overfitting, particularly when dealing with smaller datasets**.**	**[69]**

DOI: https://doi.org/10.5281/zenodo.15110559

2.6 Task 6-Reasoning

Reasoning refers to the application of deep neural network to enhance the ability of machines to perform logical inference,problem solving and decision making.It integrates deep learning techniques with reasoning process by allowing AI models to understand reasoning pattern and give answer or conclusion and also simulate thought process like human. It is also used in graph reasoning, common sense reasoning, and knowledge reasoning. The application of deep learning expands over various industries that copy understand reasoning and find accurate answer ability. It includes autonomous system which help in self-driving cars, medicaldiagnosis systemthat help doctors to assist in diagnosing disease by analysing medical record, financial analysis system which enhance the fraud detection and stock market intelligent decision system.In today's world, Tesla's AI model uses deep learning reasoning to analysis road condition and make drive decision[70]. IBM also uses deep learning reasoning in healthcare helping doctors within treatment[71]. Bank uses deep learning reasoning model to detect fraudulent transaction[72].Before deep learning, reasoning tasks were primarily handled by using following system:Rule-Based Systems-AI systems were built using hand-crafted rules and logical reasoning (e.g.expert-systems, knowledge-based systems),Traditional ML–Algorithms such as decision trees, SVMs and Bayesian networks were used to model relationships in data. Logic-Based Reasoning such as first-order logic (FOL) and probabilistic graphical models, were used to infer conclusions from structured data. But there are limitations which this system cannot handle it. In Rule-based systems, rule should be manually defined which is not possible for complex tasks. Traditional ML models need extensive manual feature selection that was time-consuming and domain-specific also. These methods were inefficient in handling unstructured data like images, videos, and natural language.TheLogic-based reasoning systems mostly failed when encountering new scenarios or missing data. When deep learning came into existence, it overcome all kind of such problems like Automatic Feature Extraction. Deep learning models learn representations directly from raw data and eliminating the need for manual feature engineering.Neural networks, particularly architectures like CNNs and RNNs/Transformers[73] enable reasoning over complex data types. Deep learning models, especially transformer-based models like GPT, BERT[74]can process vast amounts of data efficiently.A detailed overview of various models used in Reasoning, its limitations, and key takeaways are listed in Table 6.

Table 6: Summary of Popular Studies on Reasoning Task

Models	Dataset	Key takeaways	limitation	References
GPT-3,PaLM, Codex, RoBERTa, T5, Transformer, Seq2Seq and Minerva, GPT-3, MWP-BERT, Bhaskara, NaturalProver, UniGeo,	MathQA, CoqGym, GEOSTheoremQA ScienceQA	1.GPT-3,PaLM (Minerva), Codex have shown advanced reasoning talents. 2.GPT-3, Minerva, MWP-BERT, Codex carry out properly but	1.Pre-educated Language Models Are Not Optimized for Math Reasoning. 2.Lack of Consistency and Robustness in Mathematical Reasoning	**[75]**

DOI: https://doi.org/10.5281/zenodo.15110559

FinQANet		aren't optimized for math reasoning.		
BERT RoBERTa T5 GPT-3 T5-11B RoBERTa GPT-3	DeepMind Mathematics Dataset SVAMP HOList,ParaRules	1. Transformers Achieve High Performance however, Lack True Reasoning. 2. Models like BERT, GPT-three, RoBERTa, and T5 carry out properly on many NLP tasks however battle with deep reasoning.	1. Math Word Problems (MWPs) are challenging, and performance drops when questions are barely changed (e.g. SVAMP dataset). 2. Fail on lengthy sequences requiring reminiscence.	**[76]**
BERT, T5, RoBERTa Graph Neural Networks (GNNs), Relational Networks Memory-Augmented Neural Networks (MANN)	CLEVR ATOMIC DeepMind Mathematics Dataset	1. Neural networks like Transformers (BERT, T5, RoBERTa) are good at learning statistical patterns in data. 2. They lack deliberate logical reasoning and fail to systematically generalize beyond their training data.	1. Deep learning is still at the surface level (lacks true reasoning). 2. Neural networks rely on statistical correlations rather than true logical deductions.	**[77]**
RTNs & RNTNs	DBpedia	1. Classical logic-based reasoning is correct but slow and does not handle incomplete information well. 2. Relational Tensor Networks – RTNs can substitute rule-based reasoning	1. RTNs do not guarantee strict logical correctness like traditional reasoning systems. 2. RTNs struggle with nested logical rules, negation, and deep inference chains.	**[78]**

DOI: https://doi.org/10.5281/zenodo.15110559

		for faster and more scalable ontology inference.		
CNN, LSTM, Transformer and EDNNs, FDNNs, RDNNs	MNIST DatasetsMedicalAI DatasetsCybersecurity Datasets (e.g., CICIDS2017, NSL-KDD)	1. Traditional notion/proof theories were used for reasoning below uncertainty. 2. Belief Theories Can Enhance Deep Learning Models and Three varieties of uncertainty-aware deep studying fashions mentioned.	1. Handling noisy or opposed statistics stays a mission, specifically in hostile assaults on AI models. 2. Uncertainty estimation methods can extend biases if not cautiously designed.	**[79]**
CNN, MLP, LSTM and EDNNs, FDNNs, RDNNs	Sandia Matrices**,**RAVEN-FAIRPGM	1. Understanding uncertainty is fundamental to effective selection-making in AI and deep mastering. 2. Fuzzy Deep Neural Networks (FDNNs) – Uses Fuzzy Logic for indistinct data and Rough Deep Neural Networks (RDNNs) – Uses Rough Set Theory to version imprecise or incomplete records	1. Deep studying models overfit unique RPM systems rather than gaining knowledge of real abstract reasoning. 2. PGM dataset exposes generalization disasters, as many models fail on feature distributions.	**[80]**

DOI: https://doi.org/10.5281/zenodo.15110559

Transformer	Multiple VQA datasets,	1. Transformer-based approach that enhances visual reasoning through self-attention and co-attention mechanisms model iteratively refines its understanding of images and text. 2. Using custom tokens improves how the model integrates visual and textual features for better comprehension.	1. The model struggles slightly with counting and numerical comparison tasks, achieving lower accuracy in "Compare numbers" type question on the CLEVR dataset. 2. While Transformers excel in capturing relationships, they are computationally expensive and require high-end hardware for training and inference.	**[81]**
Hybrid Neural-Logical Model	Sudoku Datasets Protein MPNN dataset	1. Deep studying with logical reasoning permits solving NP-tough issues more effectively. 2.E-NPLL loss overcomes barriers of conventional pseudo-loglikelihood features, enabling better logical constraints.	1. While the approach is scalable, fixing very big NP-hard problems with many variables nevertheless poses computational stressful conditions, especially in the course of inference. 2. The E-NPLL loss, however deciding on the proper good enough charge (wide sort of omitted variables) is crucial, affecting convergence tempo.	**[82]**
RNNs, LSTM	Kinsources, IMDb,CLUTRR (Commonsense Reasoning Benchmark	1. The paper offers a hybrid version that combines Neural Networks with	1. It struggles while dealing with large-scale understanding bases with	**[83]**

DOI: https://doi.org/10.5281/zenodo.15110559

		First-Order Predicate Logic for analogical reasoning. 2. Analogical Reasoning Outperforms Traditional Deductive and Inductive Methods.	hundreds of thousands of facts. 2.The model is quite specialized for logical inference and dependent symbolic responsibilities, making it less effective for unstructured records.	
CNN, RNN	Clevr	1. CBN-based models beat humans on CLEVR (97.6% accuracy). 2. Acquire multi-step reasoning without being taught explicit compositional structure.	1. Struggling with longer, more involved problems. 2. There is no direct hierarchical modelling, as compared to domain-specific architectures.	**[84]**

2.7 Task 7-Text Classification

Text classification is a type of NLP task through which the deep learning algorithm predicts pre-tagged categories or labels for text data. It is widely used in task analysis such as sentiment analysis [89,94], spam filtering [85], topic classification [97], intent detection [87], and document classification [86]. Text type in deep learning has confirmed enormous effects throughout multiple domains, presenting greater automation, security, and efficiency in numerous applications. In sentiment evaluation, groups leverage class models to research patron evaluations from product opinions and social media, facilitating advanced service techniques and user level [105]. Similarly, junk mail detection employs text categories for filtering phishing emails and fraudulent SMS, thereby strengthening cybersecurity measures [85]. In huge-scale document management, subject matter categorization aids in organizing news articles, criminal documents, and studies papers, allowing efficient content material retrieval and recommendation systems [86]. Intent reputation in AI-driven conversational structures, which include chatbots and virtual assistants, enhances computerized question resolution by way of classifying user intents, optimizing response accuracy [87]. Moreover, medical textual content type assists in categorizing clinical information, sickness diagnosis reviews and drug discovery research, thereby supporting healthcare specialists in selection-making [88]. The developing challenge of misinformation is addressed through faux information detection fashions, which assess content credibility and prevent the unfolding of fake data on virtual structures [87]. These packages spotlight the transformative function of textual content classification, powered with the aid of deep gaining knowledge of architectures which include CNNs, RNNs, LSTMs, and Transformers, in advancing statistics processing and selection-making across industries. Traditional text

DOI: https://doi.org/10.5281/zenodo.15110559

ISSN No- 2584-2706

classification methods were based on machine learning models such as Naïve Bayes, SVMs, Decision Trees, and Latent Dirichlet Allocation (LDA), usually in conjunction with feature extraction methods such as Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF). These methods suffered from various shortcomings such as too much reliance on manual feature engineering, inability to retain word order and contextual sense, poor accuracy for complex and large datasets, and poor generalizability for wide-ranging applications. To counter these shortcomings, deep learning methods such as RNNs, LSTMs and Gated Recurrent Units (GRU) were introduced to maintain sequential dependencies in text, to counter the issue of the absence of awareness about context in conventional models.A detailed overview of various models used in Classification, its limitations and key takeaways are provided in Table 7.

Table 7: Summary of Popular Studies on Classification Task

Models	Datasets	Key takeaways	Limitations	References
LSTM, GRU Hybrid Models	IMDB, Yelp, Amazon Reviews	1.Covers sentiment evaluation, information categorization, QA, and natural language inference (NLI). 2.Organized into categories like RNNs, CNNs, Transformers, Capsule Networks, and Siamese Networks. Benchmarks provide insights into fine-performing models for precise NLP duties.	1.Large-scale deep studying fashions require enormous assets for education and inference. 2.Models rely upon massive classified datasets, making them difficult to use in low-aid settings**.**	**[89]**
BERT Seq2Seq	TREC QA, Bing	1.In this article we are learning how to integrate deep learning techniques especially about transformer models such as BERT and GPT which has enhanced the MRC capabilities. 2.There are large data sets in this paper which generously helps in development and in training of MRC models efficiently.	1.Sometimes struggles in understanding of contexts, especially in longerpassages. 2.It requires large and high-quality datasets which may not be easily available.	**[90]**
DBNs and LSTM	Stanford Natural Language Inference (SNLI)	1.This paper extracts the hierarchical features of models very effectively. 2.DBNs is good for text extraction and classification.	Sometimes models struggle in sequencing data efficiently. 2.Requires time for training as data sets have large amount of data.	**[91]**

DOI: https://doi.org/10.5281/zenodo.15110559

DCNN CNN HAN (Sentiment)	Sohu News	1. This paper shows how deep learning improves text classification by eliminating the text manually and enhancing accuracy. 2. CNN and RNN captures sequential dependencies and enhance interpretability.	1. Models requires high power for performance. 2. Models are lacking in transparency in decision making topics.	**[92]**
Caps-Net-based	Amazon Reviews (User product reviews) YouTube Music Ratings	1. Caps-Net improve the CNN classification as it maintains relationship and avoid pooling operation. 2. It uses a special feature called gated sharing unit which filters out irrelevant features and improve efficiency.	1. It is more resource intensive and complex than RNN and CNN method. 2. It depends heavily on high quality datasets.	**[93]**
MLP and CNN	Sougo Lab's Sohu News	1. Text classification is important in spam filtering, sentiment analysis, and information retrieval. 2. The application of CNN & RNN architectures enhances text classification accuracy.	1. Limited to Chinese content only; other languages subject to varying penalties. 2. Pretrained embeddings required for best accuracy.	**[94]**
CRNN HAN (Sentiment) VDCNN	Yahoo Answers	1. VDCNN is a improved version of CNN for text classification and improve performance over CNNs. 2. Directly works on Character instead of words for better working on different languages.	1. As it uses advance version of CNN, hence require more power and time. 2. It struggles with tasks that requires long range dependencies.	**[95]**
BERT MTL (Sentiment)	IMDB, MR, Amazon	1. AMTL model helps in improving feature separation between shared and task specific spaces. 2. As data is shared, so it can be reused for more tasks.	1. If data are not set in proper order than it can face overfitting issues.	**[96]**

DOI: https://doi.org/10.5281/zenodo.15110559

RNNs	EHRs	1. The study show us how we can deal with imbalanced class distribution with help of text classification. 2. Various models are used to classify texts.	1. As it uses specific datasets, so finding may not apply on other domains. 2. It uses medical notes which compromises with personal information of others and raises privacy concerns.	**[97]**
Bi-LSTM (Sentiment) DCLSTM	8,292 news articles	1. These models outperforms older models which helps in achieving higher accuracy. 2. It uses common features of CNkdmN, LSTM and MLP by combining them which helps in capturing better relationship in text data.	1. Although 8,292 news articles were used for data set, but still, this is small number for training and may affect the results. 2. Multiple deep learning models used which obviously increases the load on machine and affect performance	**[98]**

3. Discussion & Conclusion

This study systematically reviews deep learning tasks, including reading comprehension, translation, text generation, question answering, reasoning, summarization, and classification. It highlights the significant advancements made in these areas, alongside the persistent challenges that remain. Over time, numerous generalized and specialized models and diverse datasets have been developed and utilized to address these specific tasks effectively. However, the variability in design and methodology across these models and datasets demonstrates the complexity of developing solutions that can be universally applied. A snapshot of the observed models and datasets is provided in Figure S1-S8.The models under review, ranging from transformers to RNNs, all have something to offer in terms of strength, with some being more scalable, and yet others having greater context-sensitive task accuracy. Core datasets, which have been instrumental in model training, are central to determining the outcome of deep learning models. They form the basis for measuring model performance, directing researchers toward solving specific challenges inherent in various domains. Yet, dataset design itself brings with it limitations—bias, domain specificity, and generalization issues—that must be given careful thought when choosing or designing datasets for task-specific use. Assessment of models and their respective datasets point to the general trend of incremental performance improvement over multiple tasks, with transformer-based models, especially BERT, GPT, and T5, dominating tasks such as reading comprehension, question answering, and summarization. These models tend to utilize large-scale, high-variance datasets like SQuAD, GLUE, and CNN/Daily Mail, offering an abundance of training data but also revealing some

DOI: https://doi.org/10.5281/zenodo.15110559

limitations in terms of domain transferability and bias. For generation and translation tasks, GPT-based models excel but continue to struggle with nuances of languages and produce contextually consistent results in lengthy text formats. Reasoning tasks, though making advancements through models such as T5 and GPT-3, continue to need improvements in terms of logical reasoning strength and common-sense inference. Summarization and classification work perform well on extractive and abstractive tasks. Despite the strengths, concerns exist for factual accuracy and coherence retention in automatic summaries, particularly in domain-specialized cases. Conversely, many state-of-the-art models lack proper handling of aspects like data bias, model bias, heavy computationally demanding work, and lesser explainability, which negatively contribute to their utilization in practical use cases where there is a demand for transparency as well as economic resource optimization.

Although task-specific models have demonstrated significant accuracy and efficiency improvements, the review calls out a number of key areas for further investigation. These include improving generalization across languages and domains, dealing with ethical issues such as bias and fairness, and enhancing interpretability to provide transparency in decision-making. Moreover, merging multimodal datasets and models, applying transfer learning, and making models adaptable to low-resource languages are areas that show promise for further developing deep learning models in these tasks.

References

1. Rane, N. L., Mallick, S. K., Kaya, O., & Rane, J. (2024). Applications of deep learning in healthcare, finance, agriculture, retail, energy, manufacturing, and transportation.
2. Truyens, M., & Van Eecke, P. (2014). Legal aspects of text mining. *Computer law & security review*, *30*(2), 153-170
3. Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, *27*.
4. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training (2018)*.
5. Lu, J., & Rodriguez, I. (2024, March). Deep Manifold Learning for Reading Comprehension and Logical Reasoning Tasks with Polytuplet Loss. In *Future of Information and Communication Conference* (pp. 337-345). Cham: Springer Nature Switzerland
6. Perkoff, E. M., Bhattacharyya, A., Cai, J., & Cao, J. (2023, July). Comparing Neural Question Generation Architectures for Reading Comprehension. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)* (pp. 556-566).
7. Dhingra, B., Liu, H., Salakhutdinov, R., & Cohen, W. W. (2017). A comparative study of word embeddings for reading comprehension. *arXiv preprint arXiv:1703.00993*.
8. Liu, S., Zhang, X., Zhang, S., Wang, H., & Zhang, W. (2019). Neural machine reading comprehension: Methods and trends. *Applied Sciences*, *9*(18), 3698.
9. Ren, F., Liu, Y., Li, B., Wang, Z., Guo, Y., Liu, S., ... & Wang, B. (2022). Deep understanding based multi-document machine reading comprehension. *Transactions on Asian and Low-Resource Language Information Processing*, *21*(5), 1-21.
10. Luu, S. T., Van Nguyen, K., Nguyen, A. G. T., & Nguyen, N. L. T. (2021, January). An experimental study of deep neural network models for vietnamese multiple-choice reading comprehension. In *2020 IEEE Eighth International Conference on*

DOI: https://doi.org/10.5281/zenodo.15110559

Communications and Electronics (ICCE) (pp. 282-287). IEEE.
11. Guo, X. Y., Li, Y. F., & Haffari, G. (2022). Complex reading comprehension through question decomposition. *arXiv preprint arXiv:2211.03277*.
12. Cui, Y., Chen, Z., Wei, S., Wang, S., Liu, T., & Hu, G. (2016). Attention-over-attention neural networks for reading comprehension. *arXiv preprint arXiv:1607.04423*.
13. Qiu, B., Chen, X., Xu, J., & Sun, Y. (2019). A survey on neural machine reading comprehension. *arXiv preprint arXiv:1906.03824*.
14. Huang, L., Bras, R. L., Bhagavatula, C., & Choi, Y. (2019). Cosmos QA: Machine reading comprehension with contextual commonsense reasoning. *arXiv preprint arXiv:1909.00277*.
15. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
16. Csaky, R. (2019). Deep learning based chatbot models. *arXiv preprint arXiv:1908.08835*.
17. Li, L. (2024, August). In-Depth Exploration of the Advantages of Neural Networks in English Machine Translation. In *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 1418-1423). IEEE.
18. Barone, A. V. M., Helcl, J., Sennrich, R., Haddow, B., & Birch, A. (2017). Deep architectures for neural machine translation. *arXiv preprint arXiv:1707.07631*.
19. Britz, D., Goldie, A., Luong, M. T., & Le, Q. (2017). Massive exploration of neural machine translation architectures. *arXiv preprint arXiv:1703.03906*.
20. Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D. F., & Chao, L. S. (2019). Learning deep transformer models for machine translation. *arXiv preprint arXiv:1906.01787*
21. Meng, F., & Zhang, J. (2019, July). DTMT: A novel deep transition architecture for neural machine translation. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 224-231).
22. Yang, S., Wang, Y., & Chu, X. (2020). A survey of deep learning techniques for neural machine translation. *arXiv preprint arXiv:2002.07526*.
23. Ananthanarayana, T., Srivastava, P., Chintha, A., Santha, A., Landy, B., Panaro, J., ... & Nwogu, I. (2021). Deep learning methods for sign language translation. *ACM Transactions on Accessible Computing (TACCESS)*, *14*(4), 1-30.
24. Popel, M., Tomkova, M., Tomek, J., Kaiser, Ł., Uszkoreit, J., Bojar, O., & Žabokrtský, Z. (2020). Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature communications*, *11*(1), 1-15.
25. Zhang, T. (2022). Deep learning classification model for English translation styles introducing attention mechanism. *Mathematical Problems in Engineering*, *2022*(1), 6798505.
26. Sharma, S., Diwakar, M., Singh, P., Singh, V., Kadry, S., & Kim, J. (2023). Machine translation systems based on classical-statistical-deep-learning approaches. *Electronics*, *12*(7), 1716
27. Sarwar, S. M., Bonab, H., & Allan, J. (2019). A multi-task architecture on relevance-based neural query translation. *arXiv preprint arXiv:1906.06849*
28. Abdel-Salam, S., & Rafea, A. (2022). Performance study on extractive text summarization using BERT models. Information, 13(2), 67
29. Nallapati, R., Zhou, B., Gulcehre, C., & Xiang, B. (2016). Abstractive text

DOI: https://doi.org/10.5281/zenodo.15110559

summarization using sequence-to-sequence rnns and beyond. arXiv preprint arXiv:1602.06023
30. Luca, M. (2015). User-generated content and social media. In Handbook of media Economics (Vol. 1, pp. 563-592). North-Holland.
31. Anand, D., & Wagh, R. (2022). Effective deep learning approaches for summarization of legal texts. *Journal of King Saud University-Computer and Information Sciences*, *34*(5), 2141-2150
32. Kedzie, C., McKeown, K., & Daume III, H. (2018). Content selection in deep learning models of summarization. *arXiv preprint arXiv:1810.12343*.
33. Apostolidis, E., Adamantidou, E., Metsai, A. I., Mezaris, V., & Patras, I. (2021). Video summarization using deep neural networks: A survey. *Proceedings of the IEEE*, *109*(11), 1838-1863
34. Ma, C., Zhang, W. E., Guo, M., Wang, H., & Sheng, Q. Z. (2022). Multi-document summarization via deep learning techniques: A survey. *ACM Computing Surveys*, *55*(5), 1-37
35. Suleiman, D., & Awajan, A. (2020). Deep learning based abstractive text summarization: approaches, datasets, evaluation measures, and challenges. *Mathematical problems in engineering*, *2020*(1), 9365340
36. Liu, M., Ma, Z., Li, J., Wu, Y. C., & Wang, X. (2024). Deep-Learning-Based Pre-training and Refined Tuning for Web Summarization Software. *IEEE Access*
37. Miller, D. (2019). Leveraging BERT for extractive text summarization on lectures. *arXiv preprint arXiv:1906.04165*.
38. Zhang, M., Zhou, G., Yu, W., Huang, N., & Liu, W. (2022). A comprehensive survey of abstractive text summarization based on deep learning. *Computational intelligence and neuroscience*, *2022*(1), 7132226
39. Zhong, S. H., Liu, Y., Li, B., & Long, J. (2015). Query-oriented unsupervised multi-document summarization via deep learning model. *Expert systems with applications*, *42*(21), 8146-8155
40. Bano, S., Khalid, S., Tairan, N. M., Shah, H., & Khattak, H. A. (2023). Summarization of scholarly articles using BERT and BiGRU: Deep learning-based extractive approach. *Journal of King Saud University-Computer and Information Sciences*, *35*(9), 101739
41. Sukhbaatar, S., Weston, J., & Fergus, R. (2015). End-to-end memory networks.). Improvements in neural information processing systems, 28
42. Huang, K., Altosaar, J., & Ranganath, R. (2019). Clinicalbert: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:1904.05342
43. Yin, D., Dong, L., Cheng, H., Liu, X., Chang, K. W., Wei, F., & Gao, J. (2022). A survey of knowledge-intensive nlp with pre-trained language models. arXiv preprint arXiv:2202.08772
44. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
45. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186)
46. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, *21*(140), 1-67.

DOI: https://doi.org/10.5281/zenodo.15110559

47. Mutabazi, E., Ni, J., Tang, G., & Cao, W. (2021). A review on medical textual question answering systems based on deep learning approaches. *Applied Sciences*, *11*(12), 5456
48. Hakimov, S., Jebbara, S., & Cimiano, P. (2019, January). Evaluating architectural choices for deep learning approaches for question answering over knowledge bases. In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)* (pp. 110-113). IEEE
49. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., & Garcia-Rodriguez, J. (2018). A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing*, *70*, 41-65.
50. Zhang, C., Lai, Y., Feng, Y., & Zhao, D. (2021). A review of deep learning in question answering over knowledge bases. *AI Open*, *2*, 205-215.
51. Abbasiantaeb, Z., & Momtazi, S. (2021). Text-based question answering from information retrieval and deep neural network perspectives: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(6), e1412
52. Stroh, E., & Mathur, P. (2016). Question answering using deep learning. *Unpublished*
53. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, *19*(6), 1236-1246
54. - Huang, Z., Xu, S., Hu, M., Wang, X., Qiu, J., Fu, Y., ... & Wang, C. (2020). Recent trends in deep learning based open-domain textual question answering systems. *IEEE Access*, *8*, 94341-94356
55. - Palangi, H., Smolensky, P., He, X., Deng, L., & Redmond, W. A. (2017). Deep learning of grammatically-interpretable representations through question-answering. *arXiv preprint arXiv:1705.08432*.
56. Feng, M., Xiang, B., & Zhou, B. (2016, October). Distributed deep learning for question answering. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 2413-2416).
57. Jeyaraj, S., & Raghuveera, T. (2022). A deep learning based end-to-end system (F-Gen) for automated email FAQ generation. *Expert Systems with Applications*, *187*, 115896.
58. Liu, J., Snodgrass, S., Khalifa, A., Risi, S., Yannakakis, G. N., & Togelius, J. (2021). Deep learning for procedural content generation. *Neural Computing and Applications*, *33*(1), 19-37.
59. Muhammed, A., Hassan, M., Soliman, W., Ibrahim, A., & Abdelaal, S. H. (2025). The potential use of deep learning in performing autocorrection of setup errors in patients receiving radiotherapy. *Radiography*, *31*(2), 102881
60. As, I., Pal, S., & Basu, P. (2018). Artificial intelligence in architecture: Generating conceptual design via deep learning. *International Journal of Architectural Computing*, *16*(4), 306-327.
61. Alzantot, M., Chakraborty, S., & Srivastava, M. (2017, March). Sensegen: A deep learning architecture for synthetic sensor data generation. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)* (pp. 188-193). IEEE.
62. Briot, J. P., Hadjeres, G., & Pachet, F. D. (2020). *Deep learning techniques for music generation* (Vol. 1). Heidelberg: Springer
63. Wang, A., Dong, J., Lee, L. H., Shen, J., & Hui, P. (2024). A Survey on Deep Learning for Design and Generation of Virtual Architecture. *ACM Computing Surveys*, *57*(2), 1-41.

DOI: https://doi.org/10.5281/zenodo.15110559

64. Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: a survey. *Big Data*, *9*(1), 3-21.
65. Goel, K., Gu, A., Donahue, C., & Ré, C. (2022, June). It's raw! audio generation with state-space models. In *International Conference on Machine Learning* (pp. 7616-7633). PMLR
66. Lu, Y., Shen, M., Wang, H., Wang, X., van Rechem, C., Fu, T., & Wei, W. (2023). Machine learning for synthetic data generation: a review. *arXiv preprint arXiv:2302.04062*.
67. Tokui, S., Oono, K., Hido, S., & Clayton, J. (2015, December). Chainer: a next-generation open source framework for deep learning. In *Proceedings of workshop on machine learning systems (LearningSys) in the twenty-ninth annual conference on neural information processing systems (NIPS)* (Vol. 5, pp. 1-6).
68. Lu, S., Guo, D., Ren, S., Huang, J., Svyatkovskiy, A., Blanco, A., ... & Liu, S. (2021). Codexglue: A machine learning benchmark dataset for code understanding and generation. *arXiv preprint arXiv:2102.04664*
69. Li, H., Adele, Y. C., Liu, Y., Tang, D., Lei, Z., & Li, W. (2019, November). An augmented transformer architecture for natural language generation tasks. In *2019 International Conference on Data Mining Workshops (ICDMW)* (pp. 1-7). IEEE.
70. Hind, S. (2019). Digital navigation and the driving-machine: supervision, calculation, optimization, and recognition. *Mobilities*, *14*(4), 401-417
71. Jayatilake, S. M. D. A. C., & Ganegoda, G. U. (2021). Involvement of machine learning tools in healthcare decision making. *Journal of healthcare engineering*, *2021*(1), 6679512.
72. Parkar, E., Gite, S., Mishra, S., Pradhan, B., & Alamri, A. (2024). Comparative study of deep learning explainability and causal ai for fraud detection. *International Journal on Smart Sensing and Intelligent Systems*, (1).
73. Guo, Z., Wan, Z., Zhang, Q., Zhao, X., Zhang, Q., Kaplan, L. M., ... & Cho, J. H. (2024). A survey on uncertainty reasoning and quantification in belief theory and its application to deep learning. *Information Fusion*, *101*, 101987.
74. Helwe, C., Clavel, C., & Suchanek, F. (2021). Reasoning with transformer-based models: Deep learning, but shallow reasoning. In *International Conference on Automated Knowledge Base Construction (AKBC)*.)
75. Lu, P., Qiu, L., Yu, W., Welleck, S., & Chang, K. W. (2022). A survey of deep learning for mathematical reasoning. *arXiv preprint arXiv:2212.10535*.
76. Helwe, C., Clavel, C., & Suchanek, F. (2021). Reasoning with transformer-based models: Deep learning, but shallow reasoning. In *International Conference on Automated Knowledge Base Construction (AKBC)*.
77. Tran, T., Le, V., Le, H., & Le, T. M. (2021, August). From deep learning to deep reasoning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (pp. 4076-4077).
78. Hohenecker, P., & Lukasiewicz, T. (2017). Deep learning for ontology reasoning. *arXiv preprint arXiv:1705.10342*.
79. Guo, Z., Wan, Z., Zhang, Q., Zhao, X., Zhang, Q., Kaplan, L. M., ... & Cho, J. H. (2024). A survey on uncertainty reasoning and quantification in belief theory and its application to deep learning. *Information Fusion*, *101*, 101987.
80. *A Survey on Raven's Progressive Matrices*- Małkiński, M., & Mańdziuk, J. (2022). Deep learning methods for abstract visual reasoning: A survey on

DOI: https://doi.org/10.5281/zenodo.15110559

raven's progressive matrices. *ACM Computing Surveys*.
81. Zakari, R. Y., Owusu, J. W., Qin, K., He, T., & Luo, G. (2025). Seeing and Reasoning: A Simple Deep Learning Approach to Visual Question Answering. *Big Data Mining and Analytics*, *8*(2), 458-478.
82. Defresne, M., Barbe, S., & Schiex, T. (2023). Scalable coupling of deep learning with logical reasoning. *arXiv preprint arXiv:2305.07617*.
83. *Based Symbolic Processing* - Honda, H., & Hagiwara, M. (2021). Analogical reasoning with deep learning-based symbolic processing. *IEEE Access*, *9*, 121859-121870.
84. Perez, E., De Vries, H., Strub, F., Dumoulin, V., & Courville, A. (2017). Learning visual reasoning without strong priors. *arXiv preprint arXiv:1707.03017*.
85. Almeida, T. A., Hidalgo, J. M. G., & Yamakami, A. (2011, September). Contributions to the study of SMS spam filtering: new collection and results. In Proceedings of the 11th ACM symposium on Document engineering (pp. 259-262)
86. Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. In European conference on machine learning (pp. 137-142). Berlin, Heidelberg: Springer Berlin Heidelberg.
87. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.
88. Uzuner, Ö., Luo, Y., & Szolovits, P. (2007). Evaluating the state-of-the-art in automatic de-identification. Journal of the American Medical Informatics Association, 14(5), 550-563.
89. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning--based text classification: a comprehensive review. *ACM computing surveys (CSUR)*, *54*(3), 1-40.
90. Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., ... & He, L. (2022). A survey on text classification: From traditional to deep learning. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *13*(2), 1-41.
91. Zulqarnain, M., Ghazali, R., Hassim, Y. M. M., & Rehan, M. (2020). A comparative review on deep learning models for text classification. *Indones. J. Electr. Eng. Comput. Sci*, *19*(1), 325-335
92. Wu, H., Liu, Y., & Wang, J. (2020). Review of Text Classification Methods on Deep Learning. *Computers, Materials & Continua*, *63*(3).
93. Jacob, I. J. (2020). Performance evaluation of caps-net based multitask learning architecture for text classification. *Journal of Artificial Intelligence*, *2*(01), 1-10.
94. Li, S., & Gong, B. (2021). Word embedding and text classification based on deep learning methods. In *MATEC web of conferences* (Vol. 336, p. 06022). EDP Sciences. Conneau, A., Schwenk, H., Barrault, L., & Lecun, Y. (2016). Very deep convolutional networks for text classification. *arXiv preprint arXiv:1606.01781*.
95. Liu, P., Qiu, X., & Huang, X. (2017). Adversarial multi-task learning for text classification. *arXiv preprint arXiv:1704.05742*.
96. Lu, H., Ehwerhemuepha, L., & Rakovski, C. (2022). A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. *BMC medical research methodology*, *22*(1), 181.
97. Zhang, M. (2021). Applications of deep learning in news text classification. *Scientific Programming*, *2021*(1), 6095354.
98. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017).

DOI: https://doi.org/10.5281/zenodo.15110559

Attention is all you need. *Advances in neural information processing systems*, *30*.

99. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
100. Colin, R. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, *21*.
101. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.
102. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).
103. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774.*
104. Bing, L. (2012). Sentiment analysis and opinion mining (synthesis lectures on human language technologies). University of Illinois: Chicago, IL, USA.

Supplementary Data

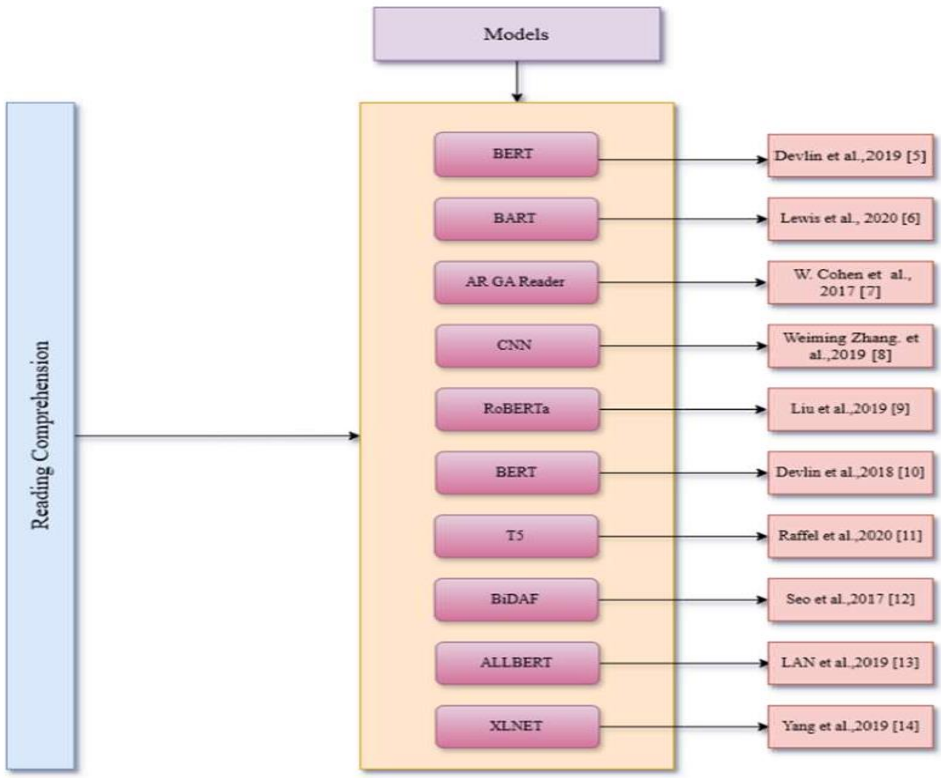

Figure S1: Popular Deep Learning Models for the Reading Comprehension Task

DOI: https://doi.org/10.5281/zenodo.15110559

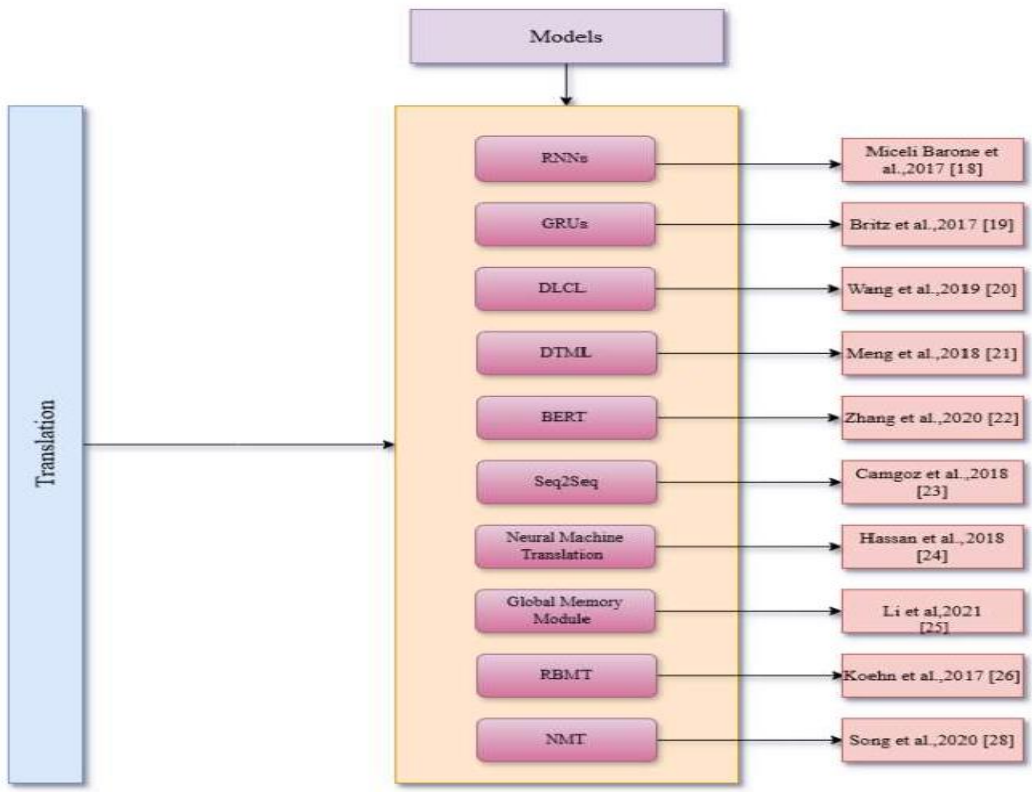

Figure S2: Popular Deep Learning Models for the Translation Task

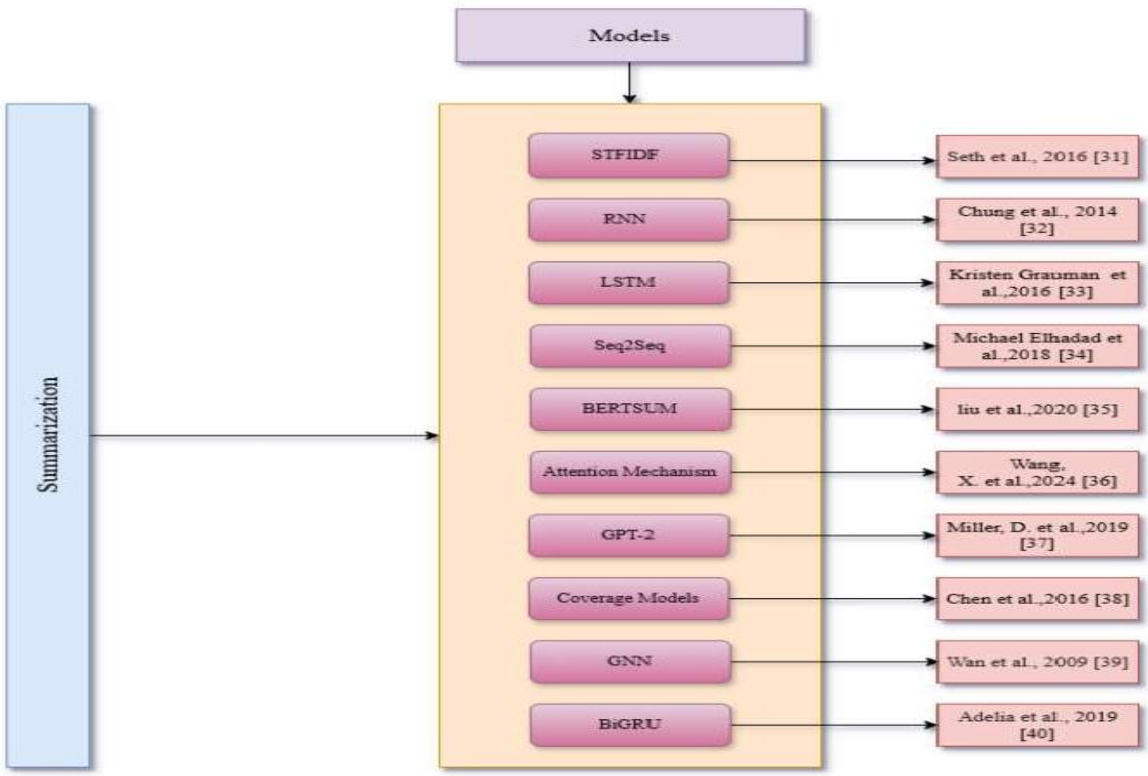

Figure S3: Popular Deep Learning Models for the Summarization Task

DOI: https://doi.org/10.5281/zenodo.15110559

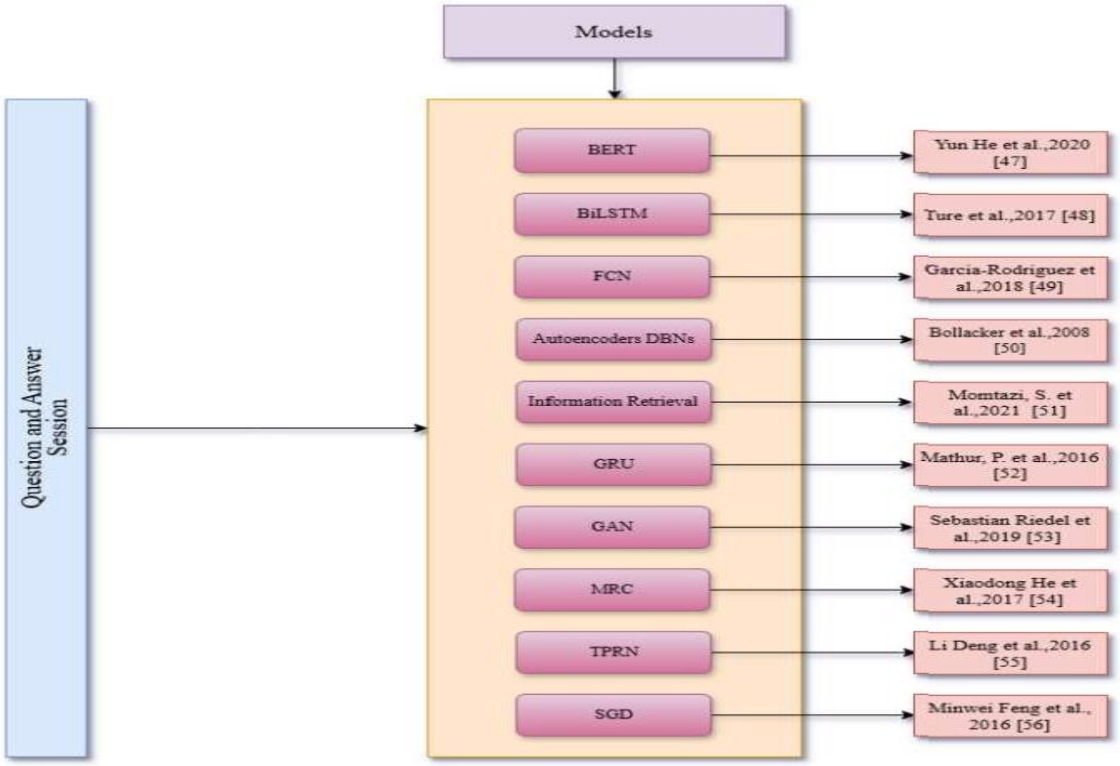

Figure S4: Popular Deep Learning Models for the Question & Answering Task

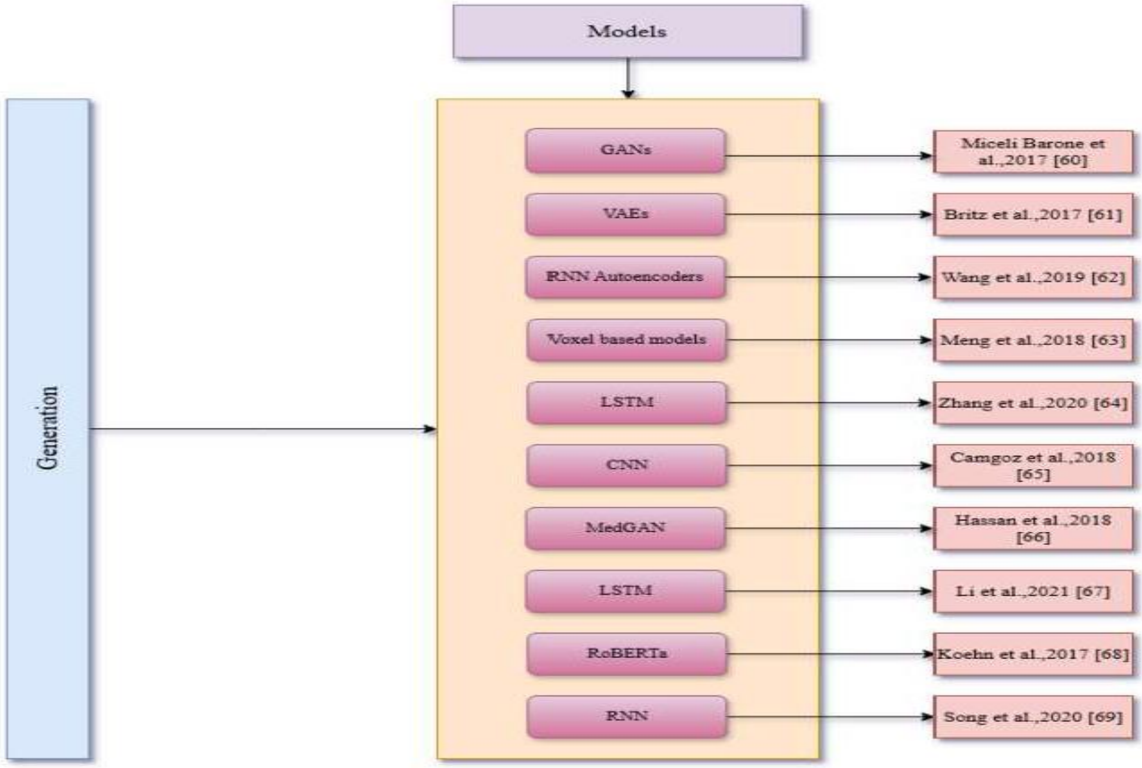

Figure S5: Popular Deep Learning Models for the Generation Task

DOI: https://doi.org/10.5281/zenodo.15110559

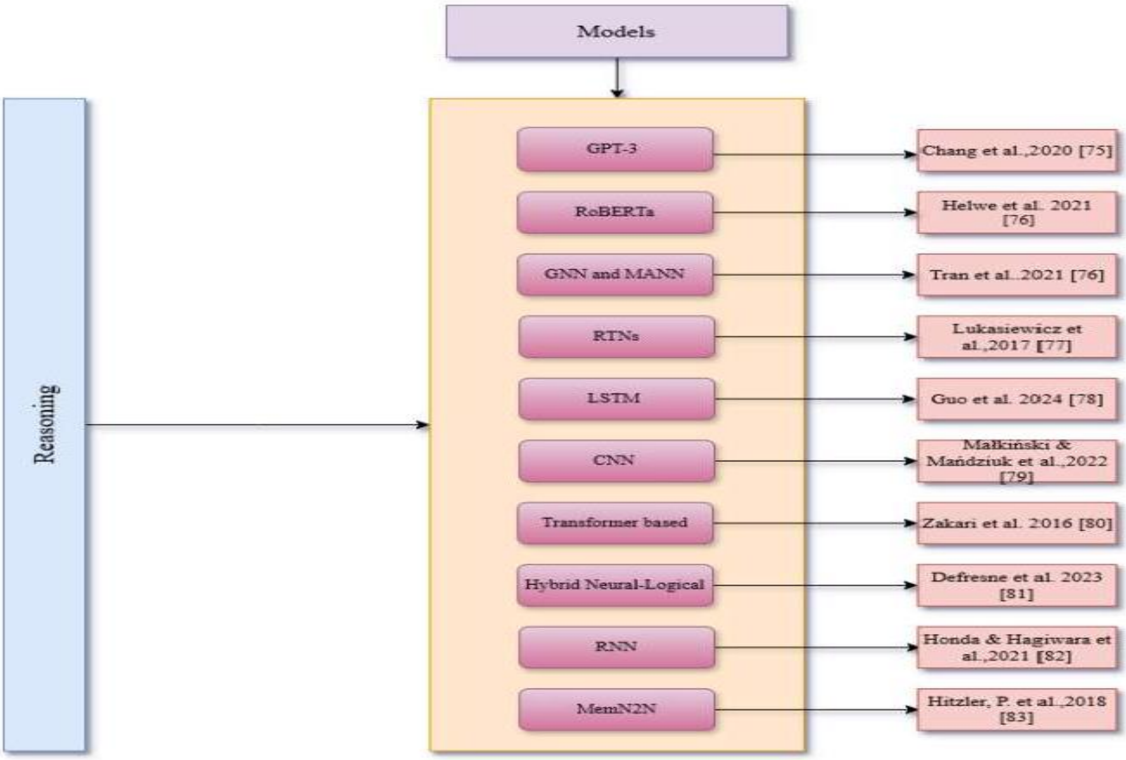

Figure S6: Popular Deep Learning Models for the Reasoning Task

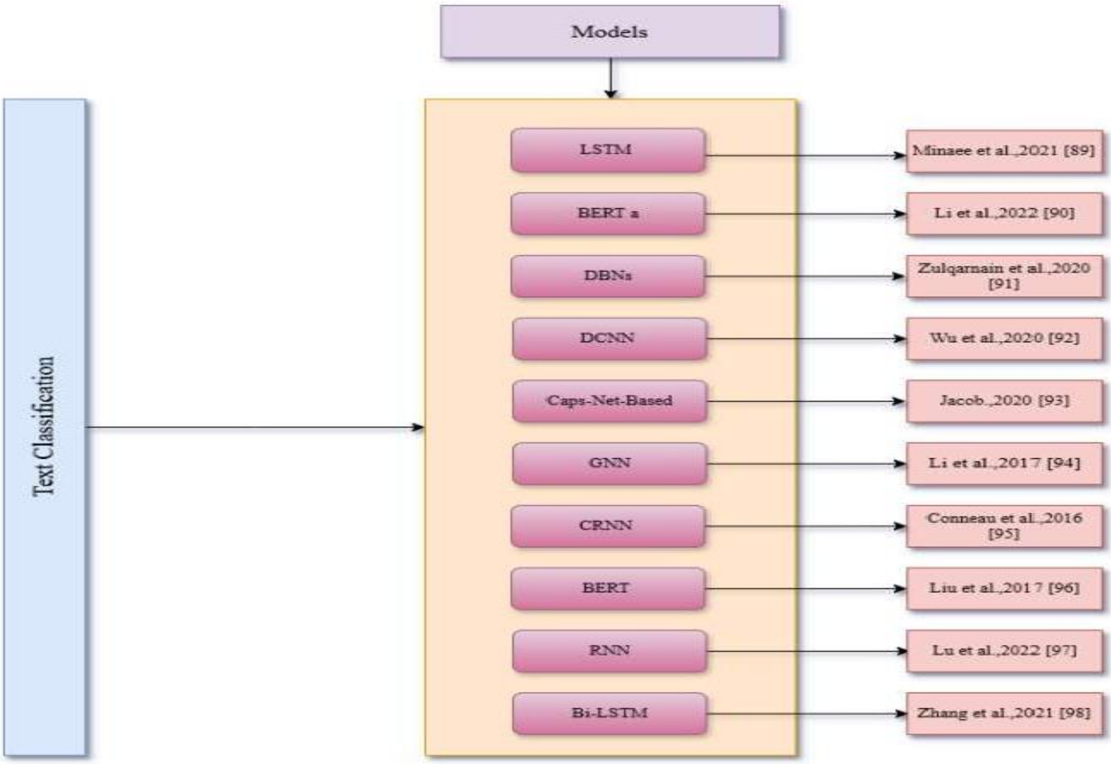

Figure S7: Popular Deep Learning Models for the Text Classification Task

DOI: https://doi.org/10.5281/zenodo.15110559

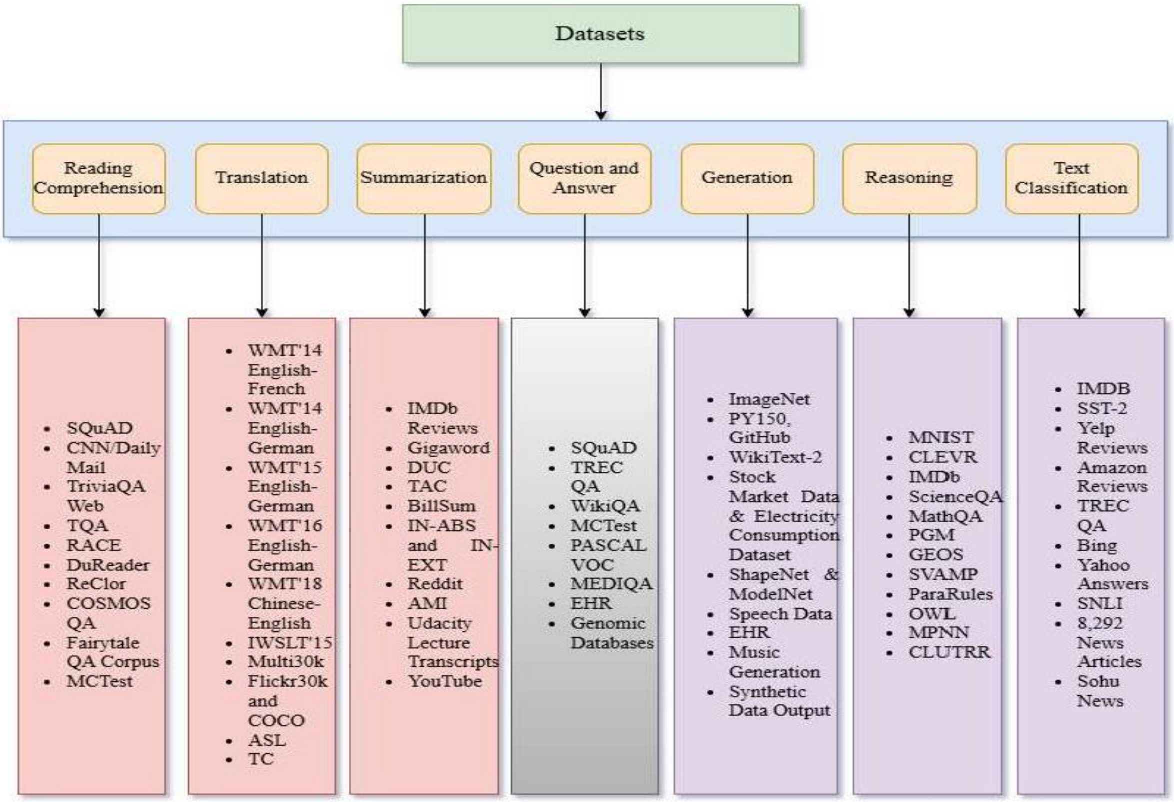

Figure S8: Popular Datasets used for different Deep Learning Tasks

DOI: https://doi.org/10.5281/zenodo.15110559