

REVOLUTIONISING TRANSLATION TECHNOLOGY: A COMPARATIVE STUDY OF VARIANT TRANSFORMER MODELS - BERT, GPT AND T5

Zaki, Muhammad Zayyanu

French Department, Faculty of Arts, Usmanu Danfodiyo University, Sokoto, Nigeria

ABSTRACT

Recently, transformer-based models have reshaped the landscape of Natural Language Processing (NLP), particularly in the domain of Machine Translation (MT). This study explores three revolutionary transformer models: Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformer (GPT), and Text-to-Text Transfer Transformer (T5). The study delves into their architecture, capabilities, and applications in the context of translation technology. The study begins by discussing the evolution of machine translation from rule-based to statistical machine translation and finally to transformer models. The models have distinct architectures and purposes which pushed the limits of MT and have been instrumental in revolutionising the field. The study found significant contributions of the models in the advancement of NLP tasks including translation technology. Using comparative approach, the study further elaborates on each model's design and utility. BERT is strong in excelling in tasks requiring a deep understanding of the context. GPT is excellent for tasks such as text generation, translation and creative writing. While the strengths of T5 is text-to-text framework by simplifying the task-specific architectures, making it easy to perform different NLP tasks. Recognising these models' unique features allows translators to select the best one for particular translation tasks and adjust them for better accuracy, fluency, and cultural relevance in translations. The study concludes that the models bridge language barriers, improve cross-cultural communication and pave way for more accurate and natural translations in the future. The study also points out that language processing models are continually evolving but understanding BERT, GPT, and T5's specific features is key for ongoing development in translation technology.

KEYWORDS

Transformer model, BERT, GPT, T5, Translation technology

1. INTRODUCTION

The translation landscape has undergone a dramatic transformation due to the emergence of powerful transformer-based models like Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformer (GPT), and Text-to-Text Transfer Transformer (T5). These models have revolutionized Natural Language Processing (NLP) tasks, particularly machine translation, by leveraging the “attention” mechanism. Unlike traditional sequential models, transformers excel at understanding long-range dependencies within sentences, enabling them to capture complex grammatical structures and nuances crucial for accurate translation, (Vaswani et al., 2017). Pre-training on massive datasets as these models are pre-trained on vast amounts of text data, allowing them to learn general language representations that can be fine-tuned for specific translation tasks (Devlin et al., 2019; Radford et al., 2019; Raffel et al., 2020). And, achieving state-of-the-art performance. Transformer models have consistently outperformed previous approaches in benchmark translation tasks, demonstrating

significant improvements in fluency, grammatical correctness, and overall quality, (Ott et al., 2018; Sockeye Team, 2019).

In today's interconnected world, language barriers often pose significant challenges in communication, trade, and understanding across diverse cultures and languages. With the swift progression of Artificial Intelligence (AI) and NLP techniques, there has been a revolutionary transformation in translation technology. This transformation has been marked by the advent of sophisticated transformer-based models: BERT, GPT, and T5. These models have significantly enhanced the accuracy and efficiency of Machine Translation (MT), leading to a paradigm shift in the way languages are translated and understood. Zaki, (24) further explains that MT is a branch of Computational Linguistics (CL) or Natural Language Processing (NLP) that studies the use of software to convert text or speech across natural languages. It is a web-based software that converts text into a variety of target languages throughout the world.

The objective of this comparative study is to delve into the intricacies of these cutting-edge transformer models and analyse their respective strengths and limitations in the context of translation tasks. BERT, GPT, and T5 represent the pinnacle of NLP, each offering unique approaches to language representation and understanding. By comparing these models comprehensively, the study aims at providing valuable insights into their performance, enabling a deeper understanding of their applications in real-world scenarios.

The study begins by exploring the historical evolution of MT and the challenges faced by traditional methods. It provides a backdrop to the emergence of transformer-based models, elucidating the underlying principles that differentiate them from earlier approaches. Understanding the context is essential to appreciate the significance of these advancements in translation technology.

Transformer Models provides a detailed explanation of the BERT, GPT, and T5 transformer architectures. It delves into their core components, including attention mechanisms, encoder-decoder structures, and pre-training techniques. A comparative analysis of these components sets the stage for evaluating their impact on translation tasks. BERT in Translation is a bidirectional model that excels in capturing contextual information from both left and right context words. This study explores how BERT has been utilised in translation tasks, highlighting its strengths and limitations. Some researches demonstrate its effectiveness in handling specific language pairs and nuanced translations.

GPT in Translation is a generative model that focuses on generating coherent and contextually relevant translations. The study examines into the applications of GPT in MT, emphasising its ability to produce fluent and contextually appropriate translations. Real-world use cases to showcase the power of GPT in handling complex sentence structures and idiomatic expressions. T5 in Translation is a text-to-text transfer model that represents a versatile approach to translation, treating all tasks as text generation problems. It explores how T5 has been leveraged for translation tasks, emphasising its flexibility in handling diverse languages and translation domains. Comparative studies between T5 and traditional translation models highlight its superiority in various scenarios.

The comprehensive comparative study aims at guiding researchers, translators, and enthusiasts with a nuanced understanding of how BERT, GPT, and T5 revolutionise translation technology. Through critical analysis and real-world examples, this study illuminates the transformative potential of these models, paving the way for a more connected and linguistically inclusive global community.

Machine Translation (MT) has come a long way since its inception, with significant advancements driven by various techniques and models. One of the pivotal milestones in the evolution of MT is the development of transformer models, which have greatly enhanced translation quality and efficiency. A preliminary assessment is presented on how MT has evolved, leading up to the transformative role of transformer models. MT research began in the mid-20th century with rule-based approaches. Early systems depend on linguistic rules and dictionaries to translate text from one language to another. However, these systems were limited by the complexity of language and often produced translations of poor quality.

In the 1990s and 2000s, Statistical Machine Translation (SMT) emerged as a dominant paradigm. SMT systems used statistical models to learn patterns from large bilingual corpora. These models, such as phrase-based models, improved translation quality significantly by capturing statistical relationships between phrases in different languages. Around 2014, Neural Networks (NNs) revolutionised MT with the introduction of Neural Machine Translation (NMT) models. Unlike rule-based and statistical methods, NMT used deep learning techniques to directly learn the mapping from one language to another. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks were initially employed for this purpose, providing better translation quality compared to earlier methods.

The breakthrough came in 2017 with the introduction of the Transformer model, as described in the paper “Attention is All You Need” by Vaswani et al., (2017). Unlike previous architectures, transformers relied on self-attention mechanisms, allowing the model to weigh the importance of different words in the input sentence when generating the translation, (D’Souza, 54). This attention mechanism enabled transformers to capture long-range dependencies and improved the quality of translations significantly. Some benefits of Transformer Models in Translation are:

- Transformers can process input sequences in parallel, making them much faster than sequential models like RNNs. This parallelisation greatly enhanced the efficiency of translation systems.
- They excel at capturing long-range dependencies in language, allowing them to generate more contextually accurate translations, especially for complex sentences.
- They can be scaled up to handle large amounts of data, leading to the development of massive pre-trained models (GPT and BERT), which have further improved translation quality through transfer learning.
- They have been extended to handle multimodal translation tasks, where both text and images are translated simultaneously. This capability is crucial for applications like image captioning and multilingual visual recognition.

Since the introduction of transformers, research in MT has continued to advance. Techniques like self-supervised learning, reinforcement learning, and iterative back-translation have been employed to further enhance translation quality and address challenges related to low-resource languages and domain adaptation. The evolution of MT from rule-based systems to statistical methods and, finally, to transformer models has significantly improved translation quality and efficiency. According to D’Souza, (52) transformers, with their ability to capture long-range dependencies and process input data in parallel, have played a pivotal role in shaping the modern landscape of MT. Ongoing research and advancements continue to refine translation systems, making them more accurate, versatile, and applicable in various real-world scenarios.

Moreover, understanding the nuances of BERT, GPT, and T5 models is crucial in the context of translation technology, as these models represent significant advancements in NLP and have

distinct characteristics that make them suitable for various translation tasks. Let us break down the importance of understanding these models in the context of translation technology:

BERT:

- BERT is considered to comprehend the milieu of words in a sentence. It reads text bidirectionally (putting into consideration both left and right context in all layers) and captures the relationships between words.
- Understanding BERT's contextual embeddings is essential for fine-tuning translation models. Translators can benefit from these embeddings to handle complex sentence structures and ambiguous phrases in different languages.
- BERT's ability to grasp the semantic meaning of words and phrases aids in more accurate translations, especially for languages with intricate nuances.

GPT:

- GPT models are generative and can produce coherent and contextually relevant text. This characteristic is useful for generating translations fluently and naturally.
- It generates text autoregressively, meaning it predicts the next word based on the preceding context. Understanding this sequential nature is vital for translators to create fluent translations that maintain coherence and meaning.
- It's creative text generation abilities can be harnessed to explore diverse ways of expressing ideas and concepts in different languages, making translations more engaging and culturally appropriate.

T5:

- T5 treats all NLP tasks, including translation, as text-to-text tasks. This unified framework simplifies the translation process, as both source and target languages are treated as text inputs, allowing for consistent handling of diverse language pairs.
- It's ability to learn task-agnostic representations of text allows for efficient transfer learning. Translators can leverage pre-trained T5 models to adapt to specific translation tasks, benefiting from the model's general language understanding capabilities.
- Translators can fine-tune T5 models for specific translation domains or styles, tailoring the translation output to meet specific requirements, such as technical, literary, or conversational translations.

Understanding the unique features and capabilities of BERT, GPT, and T5 models empowers translators to choose the right model for specific translation tasks. This knowledge also enables the fine-tuning and customisation of these models to improve the accuracy, fluency, and cultural appropriateness of translations in diverse linguistic contexts. Keeping pace with advancements in these NLP models is essential for the continuous improvement of translation technology, ensuring high-quality translations that resonate with the target audience. For Zaki, (27) NLP "is the ability of computers to understand human language. Natural language is human language and computers can analyse, understand, alter and generate it". This is further used for translation purposes in MT as corpus, text or data.

Moreover, the performance of these models on translation tasks can vary based on the specific dataset, training techniques, and evaluation metrics. Generally, T5, being specifically designed for text generation tasks like translation, often outperforms BERT and GPT on translation-related benchmarks. However, it's essential to note that the field of NLP is rapidly evolving, and newer models and techniques might have been developed. It is against this background that the researcher finds it essential to explore the power of these transformer models in translation technology.

The rationale behind the study is to evaluate and compare the transformer models in translation technology.

The objectives of the study are to:

- i. Identify the transformer models in terms of quality and effectiveness in translation tasks,
- ii. Evaluate their efficacy in translation technology,
- iii. Compare the transformer models and their contribution in revolutionising translation technology.

The justification of the study is their architecture, capabilities, and applications in the context of translation technology. The study considers the returns of transformer models such as instantaneous input sequence processing, handling language's long-range dependencies, and managing numerous translation tasks. It is through recognition of these transformer models' exceptional structures that permits translators to select the best transformer model for particular translation tasks and adjust them for better accuracy, fluency, and cultural relevance in translation processes.

2. LITERATURE REVIEW: TRANSFORMER MODELS - A DEEP DIVE

The review offers an explanation of transformer architecture, focusing on the key components such as attention mechanisms, encoder-decoder structure, and self-attention mechanisms. It provides an in-depth analysis of BERT, GPT, and T5 models, exploring their unique features, training methodologies, and underlying principles. BERT, GPT, and T5 models are Large Language Models (LLMs) used for various NLP tasks. In the context of email spam detection, LLMs have shown superior performance compared to traditional machine learning techniques such as Naïve Bayes and Light GBM, especially in scenarios with limited training samples. Spam-T5, a fine-tuned Flan-T5 model, outperforms other LLMs and baseline models in detecting email spam, particularly when training samples are scarce, (Varginia, et al., 2021). In the field of relation extraction for drug-protein interactions, BERT-based models and T5 models have been explored. Larger BERT-based models have generally performed better, while the T5 text-to-text approach shows promising results and has room for further research, (Jianmo et al., 2021). T5 models have also been investigated for generating sentence embeddings, with encoder-only models outperforming BERT-based sentence embeddings on transfer tasks and Semantic Textual Similarity (STS). Scaling up T5 to billions of parameters consistently improves downstream task performance, (Xin, et al, 2021). In predicting drug-protein interactions, an ensemble model combining fine-tuned BERT, sentence BERT, and T5 models achieved high performance, with the best model achieving an F-1 score of 0.753, (Peng et al, 2019). With the above literatures, these models are proved to be efficient and provide promising results compared to the previous models.

Furthermore, GPT models have proven extraordinary competences for Natural Language Generation (NLG) and MT together with GPT-3. They attain competitive translation quality for high-resource languages but have restricted capabilities for low-resource languages. Hybrid approaches that join GPT models with other translation systems can enhance translation quality further, Maysa, (2023). GPT-3 has been evaluated for translating specialised Arabic text to English and has shown generally comprehensible translations but struggles with capturing nuances and cultural context, (Hendy., A, et al., 2023). Chat GPT, another GPT model, has been evaluated for its understanding ability and performs well on inference tasks but inadequate in tackling rewording and resemblance tasks. Overall, GPT models have promising potential for

translation tasks, but further research is needed to expand their abilities and address their limitations, (Hasin, R., et al., 2023) and (Mamatha, A., et al., 2023). It is based on these reasons that translators need to understand these transformer models potentials and application for better results.

Also, BERT, GPT, and T5 transformer architectures are all related to NLP. BERT combines Transformers with a neighbor-attention mechanism to improve relation extraction tasks in biomedical literature, (Po-Ting, 2021). GPT is a Transformer-based language model that has achieved groundbreaking results in tasks like poetry generation and summarisation, (Topal, 2021). Transformers, in general, have revolutionised NLP by addressing issues like vanishing gradient problems and enabling parallelisation in sentence processing, (Grail, 2021). These architectures have been applied to various tasks, including downstream tasks in NLP, such as text generation and summarisation, (Zheng, 2021). BERT, in particular, is an implementation of the Transformer architecture developed by Google, (Turton, 2021).

2.1. Comparative Study of the Variant Models

A comparative analysis of the variants BERT, GPT, and T5 models is made below: BERT is a transformer-based model designed for Natural Language Understanding (NLU) tasks. For Zaki, (27) NLU implies “the understanding of language by linguists and translators”. It is fully linguistics, as it deals with each system of phonology, morphology, syntax and pragmatics.

It pre-trains a language model on a large corpus of text in a bidirectional manner, enabling it to capture context from both the left and right sides of a word. The Strengths in BERT excel in tasks requiring a deep understanding of the context, such as question answering and text completion. Its bidirectional approach allows it to capture intricate relationships between words in a sentence. The limitation of BERT requires large amounts of data and computational resources for training. It processes text sequentially, making it computationally intensive and slower for long texts.

GPT, developed by Open AI, is another transformer-based model introduced in 2018. Unlike BERT, GPT is designed for NLG tasks. For Zaki, (27), NLG is a “computer process that generates natural text and speech from pre-defined data”. It is pre-trained to predict the next word in a sentence, enabling it to generate coherent and contextually appropriate text. The strengths of GPT are excellent for tasks like text generation, translation, and creative writing. It generates text auto regressively, meaning it predicts one word at a time, which can be advantageous for certain applications. The limitations of GPT’s unidirectional nature might limit its understanding of context, as it only considers the preceding words in a sentence. It might face challenges in tasks requiring precise comprehension and extraction of information.

T5, introduced by Google Research in 2019, is a versatile transformer-based model. Unlike BERT and GPT, T5 frames all NLP tasks as text-to-text tasks, unifying different tasks under a common text-based format. It is pre-trained to convert one form of text into another, allowing it to handle a wide array of tasks. The strengths of T5’s text-to-text framework simplify the task-specific architectures, making it highly flexible and easy to apply to various NLP tasks. It achieves state-of-the-art performance across multiple benchmarks due to its unified architecture. The limitations of T5’s performance might be prejudiced by the superiority and variety of the training data for diverse tasks. Training and fine-tuning T5 models can be resource-intensive, especially for large-scale applications.

The variant models BERT, GPT, and T5 have been compared in various studies. A study found that BERT achieved higher accuracy compared to other models on the Stanford Question Answering Dataset (SQuAD), (Melek, 2023). Another study evaluated the performance of GPT

and BERT models in detecting protein-protein interactions (PPIs) and found that BERT-based models achieved the best overall performance, (Devshree, 2020). GPT-4, despite not being explicitly trained for biomedical texts, showed similar performance to the best BERT models in detecting PPIs, (Hasin, 2023). Additionally, a comparative analysis of DL models for sentiment prediction in customer reviews found that fine-tuned BERT outperformed other DL models in terms of accuracy and performance measures, (Anandan, 2022). Overall, these studies highlight the effectiveness of BERT and GPT models in various NLP tasks, including question answering, PPI identification, and sentiment prediction.

Furthermore, BERT excels in tasks requiring deep contextual understanding, making it suitable for applications like question answering and sentiment analysis. GPT is ideal for text generation tasks, such as creative writing and story generation, where coherent and contextually relevant text is essential. T5 offers a unified solution for various NLP tasks with its text-to-text framework, enabling easy adaptation and fine-tuning for specific applications.

3. APPLICATIONS OF BERT, GPT, AND T5 TRANSFORMERS MODELS IN TRANSLATION TECHNOLOGY

The variant models BERT, GPT, and T5 have practical applications in the field of translation technology. GPT has been used for question-answering systems and can be applied to further NLP tasks such as text classification, Named Entity Recognition (NER), and language translation, (Dai, 2023). According to Zaki, (26) NER is “the procedure that a machine follows in finding the name entities”. The subtask of information extraction in Artificial Intelligence (AI) known as NER looks for and verifies named entities mentioned in the unstructured text that fall into pre - defined categories like names of people, organisations, places, medical codes, time expressions, quantities, monetary values, and percentages. BERT has been explored for Neural Machine Translation (NMT) and has shown promising results when used as contextual embedding in the encoder and decoder of the NMT model, (Zhu, 2020),(Sabharwal et al., 2021). It has been used for supervised NMT tasks, achieving state-of-the-art results on benchmark datasets, (Clinchant, 2019) and (Garg, 2020). T5 (text-to-text transfer transformer) can be used for translation tasks, as it has been shown to achieve high translation quality when fine-tuned on translation datasets.

BERT, GPT, and T5 are advanced NLP models with practical applications in translation technology. Some practical applications of these models in the field of translation are: BERT-based models have been used for improving translation quality by generating contextual embeddings of words and phrases in departed and arrived languages. GPT-based models can be fine-tuned for translation tasks, where the model generates fluent and contextually relevant translations given a source text. T5 models can be applied to translation tasks by framing translation as a text-to-text problem, where the input is a text prompt in the departed language, and the output is the translated text in the arrived language. The field of AI and NLP is rapidly evolving, and new applications and advancements. A transformer model is presented in the figure below:

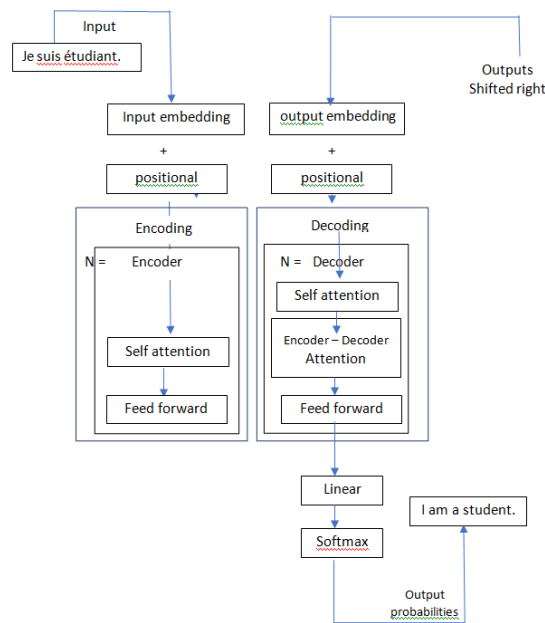


Figure 1: Transformer Model
Source: Reddy, 2023

The translation process in the transformer model starts from a sentence or text in the source language as input and ends in the target language as output. The transfer follows certain procedures from encoder, input, and output embeddings such as self-attention, feed-forward, and to decoder following linear and SoftMax calculating output probabilities for the translation result.

4. METHODOLOGY

The population of the study focuses on the comparison between BERT, GPT, and T5 transformer models in revolutionising translation. The study tries to establish the application of these models in translation technology. It is based on the facts and results of the models in translations and from translation experts. The theory of meaning and the study applies a comparative, scientific and technical approach to compare and analyse the facts about transformer models.

5. RESEARCH FINDINGS

The transformer-based models BERT, GPT, and T5 are powerful and have significantly contributed to the advancement of NLP tasks, including translation technology. While they have distinct architectures and purposes, they have collectively pushed the boundaries of MT and have been instrumental in revolutionising the field. They are logically presented based on their introduction as follows: BERT was introduced by Google in 2018 and revolutionised the way researchers approached NLU tasks. Unlike previous models, BERT is bidirectional and can understand the context of a word based on its surrounding words in a sentence. It has been used in various ways to enhance translation technology, particularly in the area of contextual word embeddings. Translation models utilising BERT embeddings can generate more accurate and contextually relevant translations, by understanding the context of words in both source and target languages.

T5 was introduced by Google in 2019 and takes a unified approach to various NLP tasks, including translation. Instead of treating translation as a sequence-to-sequence task, T5 frames all NLP tasks as text-to-text tasks. This means that both the input and output are treated as text strings. For translation, the source language text is treated as the input text, and the target language text is treated as the output text. This approach allows T5 to handle translation consistently with other NLP tasks. T5 models have achieved state-of-the-art results in MT tasks, by pre-training on a large corpus of text and fine-tuning on translation-specific data.

GPT was developed by Open AI and focuses on generating coherent and contextually relevant text based on a given prompt. While GPT is not specifically designed for translation tasks, its ability to generate human-like text has been harnessed in certain translation applications. GPT-based systems can provide reasonably good translations, especially for shorter texts. By conditioning the model on a source language prompt and allowing it to generate text in the target language, However, its unidirectional nature (it generates text from left to right) can limit its effectiveness for some translation tasks where understanding the entire sentence context is crucial.

These variant models brought about the revolution from their ability to capture complex linguistic patterns, contextual nuances and semantic meanings in both source and target languages. Researchers, practitioners and developers continue to develop and build upon these innovations leading to more advancements in MT systems.

6. RESEARCH IMPLICATIONS

BERT, GPT, and T5 are all powerful transformer-based models that have significantly impacted various NLP tasks, including translation technology. Some of the implications of these models in the field of translation are:

- BERT, GPT, and T5 models have demonstrated superior performance in understanding context and generating fluent and contextually relevant translations. These models can capture complex linguistic patterns and nuances, leading to improved translation quality, especially for ambiguous or context-dependent phrases.
- BERT, being a bidirectional model, captures contextual information effectively. It understands the meaning of words in the context of surrounding words, enabling it to produce contextually accurate translations. This is particularly useful for languages with ambiguous word meanings.
- GPT is a generative model, that can produce coherent and contextually appropriate translations. Its ability to generate text sequentially allows it to create fluent translations that follow the natural flow of the target language. GPT-based models can generate longer translations with consistent style and tone.
- T5, based on a text-to-text approach, treats all NLP tasks, including translation, as converting one kind of text to another. This framework allows T5 to handle translation in a unified manner, making it versatile and adaptable to various language pairs and domains. T5's ability to frame translation as a text-generation task contributes to its effectiveness in this area.
- GPT and T5 models have shown promising results in few-shot and zero-shot translation scenarios. Few-shot translation involves providing the model with a few examples of the translation task, allowing it to generalise and translate similar phrases accurately. Zero-

shot translation involves translating language pairs the model has never seen during training. Both capabilities open the door for more flexible and adaptable translation systems.

- These transformer models can be fine-tuned for multiple languages, enabling the development of multilingual translation systems. This is especially valuable for languages with limited labeled data, as these models can leverage the knowledge learned from high-resource languages to advance translation quality for low-resource languages.
- BERT, GPT, and T5 models can be fine-tuned on specific domains or topics, allowing developers to create domain-specific translation systems. The customisation enhances the accuracy and relevance of translations in specialised fields such as legal, medical, or technical translations.

While these models offer remarkable capabilities, challenges such as biases in training data, ethical concerns related to content manipulation, and the potential for reinforcing existing stereotypes in translations need to be addressed. Researchers and practitioners must be mindful of these issues while deploying these models in real-world applications. In summary, BERT, GPT, and T5 transformer models have significantly advanced the field of translation technology by providing state-of-the-art solutions for various translation challenges. Their ability to understand context, generate fluent translations, handle multilingual tasks, and adapt to specific domains makes them pivotal in the development of advanced and versatile translation systems. However, it is essential to handle ethical concerns and biases to guarantee accountable and reasonable use of these technologies in translation applications.

7. CONCLUSION

The conclusion summarises the key findings of the study and emphasises the significance of BERT, GPT, and T5 models in revolutionising translation technology. It highlights the potential of these models to bridge language barriers, improve cross-cultural communication, and pave the way for more accurate and natural translations in the future. This study has helped to provide readers and translators with a thorough understanding of the transformative impact of BERT, GPT, and T5 models on translation technology, offering valuable insights for researchers, practitioners, and enthusiasts in the field of NLP and MT.

8. RECOMMENDATIONS

The rapid advancement of AI-driven translation tools necessitates translators to modify their approaches to optimise the advantages and minimise the drawbacks of these technologies. When AI is used effectively and with a thorough awareness of both its strengths and weaknesses, it can greatly improve human-AI cooperation and collaboration in translation. The suggestions in this study are meant to assist educators of translation in preparing language specialists to operate efficiently using state-of-the-art technologies, they should do the following:

- Pay attention to creative translation and specialised translation fields,
- Offer a thorough investigation of AI-based translation systems,
- Develop computational and programming abilities,
- Pay attention to proofreading and modifying translations,
- Create more challenging evaluation assignments.

9. RESEARCH CHALLENGES AND FUTURE DIRECTION

The study discusses the challenges faced by transformer models in translation tasks, such as handling rare languages, idiomatic expressions, and context-aware translations. It also explores potential solutions and future directions, including model fine-tuning, transfer learning, and hybrid approaches, to address these challenges and further enhance translation technology.

REFERENCES

- [1] Anandan C, et al., (2022). Comparative Analysis of BERT-base Transformers and Deep Learning Sentiment Prediction Models. doi: 10.1109/smart55829.2022.10047651 Po-Ting, L, et al., (2021). BERT-GT: Cross-sentence n-ary relation extraction with BERT and Graph Transformer. arXiv: Computation and Language.
- [2] Devlin, J., et al., (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
- [3] Clinchant, S, et al., (2019). On the use of BERT for Neural Machine Translation. arXiv: Computation and Language,
- [4] D'Souza, J. A Review of Transformer Models. Artificial Intelligence, (2023).
- [5] Dai. Y, et al., (2023). Syntactic Knowledge via Graph Attention with BERT in Machine Translation. arXiv.org, doi: 10.48550/arXiv.2305.13413
- [6] Devlin, J., et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL), (2019) pages 4171-4186.
- [7] Devshree, P., et al., (2020). Comparative Study of Machine Learning Models and BERT on SQuAD. arXiv: Computation and Language,
- [8] Garg, A. et al., (2020). NEWS Article Summarization with Pretrained Transformer. doi: 10.1007/978-981-16-0401-0_15
- [9] Grail., Q, (2021). Globalizing BERT-based Transformer Architectures for Long Document Summarization. doi: 10.18653/V1/2021.EACL-MAIN.154.
- [10] Hendy., A, et al., (2023). How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation. arXiv.org, doi: 10.48550/arXiv.2302.09210
- [11] Hasin, R., et al., (2023). Evaluation of GPT and BERT-based models on identifying protein- protein interactions in biomedical text. arXiv.org, doi: 10.48550/arXiv.2303.17728
- [12] Jianmo, N., et al., (2021). Sentence-T5: Scalable Sentence Encoders from Pre-trained Text-to- Text Models. arXiv: Computation and Language,
- [13] Koehn, P., Neural Machine Translation. 1st Edition. Cambridge: Cambridge University Press, 2020.
- [14] Maysa, B, (2023). Exploring the Effectiveness of GPT-3 in Translating Specialized Religious Text from Arabic to English: A Comparative Study with Human Translation. Journal of Translation and Language Studies, doi: 10.48185/jtls.v4i2.762
- [15] Mamatha, A., et al., (2023). A Comparative Study on Transformer-based News Summarization. doi: 10.1109/DeSE58274.2023.10099798
- [16] Melek, K, (2023). AI in Medical Education: A Comparative Analysis of GPT-4 and GPT-3.5 on Turkish Medical Specialization Exam Performance. medRxiv, doi: 10.1101/2023.07.12.23292564
- [17] Mamatha, A., et al., (2023). A Comparative Study on Transformer-based News Summarization. doi: 10.1109/DeSE58274.2023.10099798
- [18] Nwanjoku, A.C. et al., A Reflection on the Practice of Auto-Translation and Self-Translation in the Twenty-First Century. Case Studies Journal. (2021) Vol. 10 (8) pages 24-42
- [19] Ott, M. et al., Fairseq: A Fast, Extensible Toolkit for Sequence Modeling. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), (2018) pages 48-53.
- [20] Peng, S, et al., (2019). Simple BERT Models for Relation Extraction and Semantic Role Labeling.. arXiv: Computation and Language, Radford, A., et al., (2018). Improving Language Understanding by Generative Pre-training. OpenAI Blog.
- [21] Raffel, C., et al., Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Artificial Intelligence Research, (2020) 67:1-67.

- [22] Radford, P., et al., Language Models are Few-Shot Learners. arXiv preprint arXiv:1905.13677, (2019)
- [23] Reddy, S., Transformer Models and BERT Model: Overview. Advanced Solutions Lab, Google Cloud. (Video). 2023
- [24] Sabharwal, N, et al., (2021). BERT Model Applications: Other Tasks. doi: 10.1007/978-1-4842-6664-9_6
- [25] Siu, S.C., (2023) “Revolutionizing Translation with AI: Unravelling Neural Machine Translation and Generative Pre-trained Large Language Models”.
- [26] Sockeye Team, A Toolkit for Neural Machine Translation. arXiv preprint arXiv:1704.00459 (2019) Sockeye
- [27] Topal., M. O. et al., (2021). Exploring Transformers in Natural Language Generation: GPT, BERT, and XLNet.. arXiv: Computation and Language.
- [28] Turton., J. (2021). Deriving Contextualised Semantic Features from BERT (and Other Transformer Model) Embeddings. doi: 10.18653/V1/2021.REPL4NLP-1.26
- [29] Vaswani, A., et al., Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), (2017) pages 5998-6008.
- [30] Virginia, A., et al., (2021). Text Mining Drug/Chemical-Protein Interactions using an Ensemble of BERT and T5-Based Models. arXiv: Computation and Language.
- [31] Xin, S, et al., (2021). Text Mining Drug-Protein Interactions using an Ensemble of BERT, Sentence BERT and T5 models. bioRxiv, doi: 10.1101/2021.10.26.465944
- [32] Zheng., X, et al., (2021). Adapting GPT, GPT-2 and BERT Language Models for Speech Recognition. arXiv: Computation and Language.
- [33] Zhu, J., et al., (2020). Incorporating BERT into Neural Machine Translation. arXiv: Computation and Language,
- [34] Zaki, M. Z. A Concise Handbook of Modern Translation Technology Terms. Maldov: Lambert Academic Publishing, 2023.
- [35] A Pragmatic Approach to the Translation of the Qur’an in Relation to Modern Technology. GAS Journal of Religious Studies (GASJRS), Vol. 1 (1) (2024) pages 1-12.
- [36] Explaining Some Fundamentals of Translation Technology. GAS Journal of Arts Humanities and Social Sciences (GASJAHSS) Vol. 2 (3) (2024) pages 177-185
- [37] Zaki. M. Z. et al., Multimodal and Multimedia: An Evaluation of Revoicing in Agent Raghav TV Series of Hausa in Arewa24. Journal of Translation and Language Studies 5 (1) (2024) pages 23-31.
- [38] “Understanding Terminologies of CAT Tools and Machine Translation Applications”. Case Studies Journal (2021) Volume 10, Issue 12, pages 30-39.
- [39] “Appreciating Online Software-based Machine Translation: Google Translator”. International Journal of Multidisciplinary Academic Research. (2021) Vol. 2 (2) pages 1-7.
- [40] “Recourse to Modern Technology – The EduERP Usage: An Appraisal of UDUS Reports Portal”. NUFJOL : Northern Inter-University French Journal, Revue Française Inter- Universitaire du Nord . (2019) Vol. 6 No 1, pages. 169-188.
- [41] “Translation and Modern Technologies: An Appraisal of Some Machine Translation”. Degel: Journal of Faculty of Arts and Islamic Studies. (2017) Vol. 15, Issues 1.

ABBREVIATIONS

| | | |
|-------|---|--|
| AI | - | Artificial Intelligence |
| BERT | - | Bidirectional Encoder Representations from Transformer |
| CL | - | Computational Linguistics |
| GPT | - | Generative Pre-trained Transformer |
| LLMs | - | Large Language Models |
| LSTM | - | Long Short-Term Memory |
| MT | - | Machine Translation |
| NLG | - | Natural Language Generation |
| NLP | - | Natural Language Processing |
| NLU | - | Natural Language Understanding |
| NMT | - | Neural Machine Translation |
| NNs | - | Neural Networks |
| RNNs | - | Recurrent Neural Networks |
| SMT | - | Statistical Machine Translation |
| SQuAD | - | Stanford Question Answering Dataset |
| STS | - | Semantic Textual Similarity |
| T5 | - | Text-to-Text Transfer Transformer |