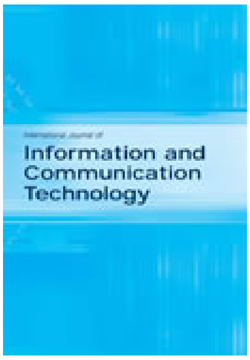


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Adaptive neural machine translation with attention mechanisms for English texts

Weiwei Suo

DOI: 10.1504/IJICT.2025.10071310

Article History:

Received:	17 March 2025
Last revised:	07 April 2025
Accepted:	07 April 2025
Published online:	27 May 2025

Adaptive neural machine translation with attention mechanisms for English texts

Weiwei Suo

Foreign Linguistics and Applied Linguistics,
Xi'an Fanyi University,
Xi'an, 710105, China
Email: sarahsww@126.com

Abstract: Neural machine translation (NMT) is the study endeavoured to build systems that would assimilate human language deciphering and production by utilising far-reaching linguistic and contextual forms. This article provides the details about an adaptive neural machine translation (ANMT) model which has incorporated the attentional mechanisms to deal with English texts. A proposed model is compared to existing best practice translation frameworks which are then included with two different approaches such as idiomatic expressions, domain-specific terminologies, and low-resource scenarios. We propose a new adaptation where user feedback loops are used as a method for refining translations based on emerging linguistic patterns. The experimental results confirm that ANMT was a success and the translation mistakes had lessened when new models were adopted; additionally, indicating that NMT experts had received a much better score compared to the baseline language model. This means that ANMT is a significant step in the evolution of AI technologies in translation.

Keywords: neural machine translation; NMT; attention mechanisms; adaptive learning; context-aware translation; deep learning.

Reference to this paper should be made as follows: Suo, W. (2025) 'Adaptive neural machine translation with attention mechanisms for English texts', *Int. J. Information and Communication Technology*, Vol. 26, No. 15, pp.25–40.

Biographical notes: Weiwei Suo is a research scholar at the Department of Foreign Linguistics and Applied Linguistics, Xi'an Fanyi University, in Xi'an, China. Her academic focus lies in applied linguistics, contributing to the advancement of language research and education.

1 Introduction

The objective of this upgrade is to ensure that the tech used in development is transformed into machines that can carry out a plethora of human languages. With this, NMT can be connected to creating a system of translating the machine through the implementation of several parts of the machine that look and sound like humans (Al-Thanyyan and Azmi, 2023). Because of the transformation of the model through the integration of human-inspired linguistic modelling and machine computation, NMT uses the connection between the lexicon and the grammar that humans have, to replace many of the human references in the machine text. To enhance English sentences, this research mixes the use of the attention mechanisms integrated with adaptive neural machine translation (ANMT), the focus of this study (Hameed et al., 2022).

The major problem with translation is the fact that natural language is complicated because of its sociolinguistic aspects, double meanings, and nonlinear complexity. Existing translation models, such as the phrase method and the p-method, are often inefficient in generating long sentences. The result is word alignment errors, incoherence, as well as an inability to convey the intended meaning (Liu and Zhang, 2023). Attention mechanism-based NMT is the solution to these problems as it allows the model to decide which parts of the input sentence should be given greater weight. This improves the relationship between the remittance of the sender with the eventual outcome of the recipient's reply, while it also enables the easy explanation of intricate structures, idioms, and expressions related to a particular field of study (Alkhawaja et al., 2020).

1.1 Evolution of machine translation

The domain of machine translation has moved through different stages, from being rule-centred and eventually increasing to models with neural mechanisms (Jian et al., 2022). The early systems for machine translation were constructed through hand-made grammatical rules, having the advantage of being very systematic, yet being very different from the patterns of usage in different languages, hence were unable to adapt easily (Sitender et al., 2023). The introduction of statistical machine translation (SMT) in the late 20th century marked a great leap, allowing translation to be generated with probabilistic models built on bilingual corpora. Nevertheless, the SMT was still characterised by such issues as data sparsity, error propagation, and the lack of fluency, which rendered it impossible to produce natural translations (Wu, 2020).

The most significant change in translation processes was introduced by deep learning and sequence-to-sequence (Seq2Seq) models, which used recurrent neural networks (RNNs) to learn language patterns (Lowphansirikul et al., 2022). The systematic use of RNNs in such a way, however, resulted in the model not being able to handle long-distance connections as it was meant to do so. The introduction of attention mechanisms, particularly the transformer architecture, revolutionised NMT by allowing models to process entire sentences in chorus while dynamically focusing attention on various words in the source text. The pioneering model put recurrent layers aside focusing on self-attention mechanisms. This allowed for considerable improvement in translation quality and efficiency (Noll et al., 2023).

1.2 Importance of attention mechanisms in NMT

Attention mechanisms have become the bedrock of current NMT designs, resolving certain critical deficits of earlier deep-learning models (Cai and Lam, 2020). In similar terms, while the traditional recurrent neural networks process words one at a time, the attention-based networks assign differentiated importance values to words, ensuring that translations are contextually relevant. This is particularly the case with lengthy and intricate sentences, such as those in which some words are the key to others appearing elsewhere in the sentence (Alsahli, 2019).

The self-attention mechanism, used in the transformer model to hone in on computing its attention scores with each of the words in the sequence, is particularly useful for enhancing the quality of the translations (Das et al., 2023). The bottleneck associated with time-consuming RNN processing is removed and, in addition, the model is enabled to swiftly achieve a pretty good level of understanding to be able to link the words and complete the task at hand (Cao and Hao, 2021). Also, multi-head attention increases this power by making it possible for the model to listen in on multiple parts of a sentence at different levels of abstraction. Subtle linguistic nuances like word order variations, the expression of idioms, and syntactic transformations are captured in the translation process with absolute accuracy (Yang and Laki, 2023).

1.3 Adaptive neural machine translation

Although attention mechanisms have brought a considerable improvement in NMT, it is still a challenge to guarantee configuration and adaptability across various linguistic contexts and domains as well as user preferences (Liu, 2021). The ANMT seeks to extend the conventional NMT by introducing the methods of adjustment through the use of changing learning strategies that provide for instant corrective training when the given input turns out to be complex, the language is domain-specific, or the user gives some feedback (Noriega-Santiáñez and Pastor, 2023). The basic elements of ANMT are:

- Adaptive attention mechanisms – these are used to adjust the attention weights based on the complexity of the sentence, the context of the discourse, and the linguistic patterns (Seki, 2021). The distinction between adaptive and static attention mechanisms is that the refinement of weight distributions in adaptive attention happens in the real-time mode while the occurrence of the other one is independent of the external conditions. The result is the alignment between source and target languages (Olkhovska and Frolova, 2020).
- Continuous learning and fine-tuning – in traditional NMT models, language knowledge is mostly limited due to the training on fixed large datasets that do not change rapidly with time to adapt to new language trends or the vocabulary of specific domains (Zygadło et al., 2021). With ANMT, the methods of direct language input are covered, namely, user feedback, online learning, and reinforcement learning hence it carries this innovation, and thus, the models can adapt, changing along with user activity (Jia et al., 2019).
- Domain adaptation and context awareness – there are different text domains like legal, medical, and technical content, which need different translation techniques. Also, they make sure that the math is correct so that the data is put together properly (Jiang and Niu, 2022). CTT uses a domain-adaptive mechanism that assists in the process which is fine-tuning the models' particular industry terminologies, and contextual variations. This fact helped the translators to build unique solutions based on our clients' needs this is why ANMT ensured translation accuracy both in professional and academic practice contexts (Sun, 2022).

1.4 Objectives

- Creating a new model – design an advanced attention mechanism that can change translation priority depending on sentence structure, context, and specific domain terminology.

- Generative model improvement via learning – added with reinforcement learning and fine adjustment through a vast amount of persistent data the model can quickly and flexibly respond to changes in the environment.
- Against aiming new NMT results configuration – the presented ANMT method was a comparison of the well-known translation tools such as Google Translate, DeepL, and Baidu Translations using the bilingual evaluation understudy (BLEU), metric for evaluation of translation with explicit ordering (METEOR) and ROUGE metrics ensuring that the correct flow of ideas, data, and facts takes place with a high degree of correctness.

The evolution of machine translation from rule-based systems to neural architectures has revolutionised the field of automated translation. This has been possible due to the introduction of attention-based models which integrated improvements in context preservation, fluency and syntactic accuracy (Ashengo et al., 2021). However, despite the advances achieved, there are still challenges that are concerned with ensuring flexibility, coherence, and continuous enhancement of translation quality. Based on these premises, this work provides an ANMT framework, which utilises dynamic attention mechanism, real-time fine-tuning, and domain-specific adaptations in order to boost translation performance (Kudo and Richardson, 2018). ANMT takes care of capturing long-distance dependencies, incorporating linguistic variations, and adapting to evolving language trends, and thus offers a multiplatform, highly accurate translation service for English texts. This study presents a real-time high-fidelity machine translation tool that can help advance AI technology for language translation applications under diverse conditions (Pham et al., 2023).

2 Literature review

The recent years have seen the field of NMT evolve tremendously, thanks to innovations such as attention mechanisms and adaptive learning techniques. NMT has effectively dethroned traditional statistical and phrase-based translation models as a result of the introduction of deep learning approaches that deliver improved fluency, contextual understanding, and syntactic coherence (Jiaxin and Peiting, 2022). The attention mechanism provided by the transformer has proven to be the key ingredient for improving the quality of translations, allowing long-range dependencies and specific domain linguistic variations to be embedded in models (Gaser et al., 2023). The section investigates practical studies that are pivotal to ANMT construction, explains the mechanisms involved, and celebrates the achievements that these innovations bring to automated translation technology.

Fakih et al. (2024) scrutinise Instagram's automatic translation feature with a particular focus on its ability to translate literary texts from Arabic to English. Their study critiques Instagram's 'See Translation' function, pinpointing major shortcomings in fluency, accuracy, and style. After evaluating 30 Instagram captions, the authors conclude that accuracy (25), fluency (26), and style (10) were notably poor in the translations falling short in 90% of the cases. The paper advocates for improving the algorithms driving NMT, particularly for literary genres and linguistic variants to increase performance.

In contrast, Kuzman et al. (2019) study the power of NMT models designed for literary translation against the capability of Google's NMT. The team analysed in detail the English to Slovene translations testing the models through the evaluation of the BLEU, METEOR, and post-editing effort metrics. The study concludes that a customised NMT model trained on novels outperforms a general-purpose NMT model in many aspects, but fluency is the exception. Notably, Google's NMT beats all tailored models and the study reveals challenges in getting different evaluators to agree on fluency and adequacy.

In the end, Kugathasan and Sumathipala (2021) centre on translating the text that mixes codes, particularly the Sinhala-English code-mixed text, into standard Sinhala. They are proposing an NMT model that deploys an encoder-decoder framework with LSTM units of deep learning which the system is based on that usually operates on a teacher forcing algorithm. This research highlights the problems of code-mixing in multilingual societies while a parallel corpus is built as a solution to the lack of resources for Sinhala-English code-mixed texts. The proposed model achieves an astounding BLEU score, revealing its potential in translating the code-mixed social media content that is a huge challenge.

Benkova et al. (2021) approach the comparison of SMT and NMT systems for the English-to-Slovak translation. Their experiment is the one which automatically estimates both methods, using BLEU_n, a typical machine learning metric, and analysing the quality of translations for some journalistic texts. The results suggest that NMT is throughout the winner with SMT on translation quality. The study specifically mentions that NMT, be it trained on either general text or specific domain ones, presents a better translation quality than OF.

Kumhar et al. (2023) mention the application of NMT for converting English text into Urdu resultant. They apply acquaintance with an LSTM-based encoder-decoder model in the automation of the translation process which it strives to accomplish. The authors attained a plethora of comfortable stages like the preprocessing, tokenisation, and grammar analysis through the confirmation of prior improvement in translation studies. The results are unambiguously demonstrative of the major significance of the model delivering the high BLEU score in both of the test and training data, thus a key breakthrough in English-to-Urdu machine translation.

Kugathasan and Sumathipala (2022) also provide a study for Sinhala-English code-mixed segments, but in there, translations are into standard Sinhala. They released an NMT application, which consists of a normalisation pipeline plus LSTM units, Seq2Seq, and teacher forcing mechanisms as the paper described. The proposal was put through a comparative study against currently available best methods of BLEU code and was shown to obtain a good score of 33.89, which is a proper one indicating the enhanced translation quality.

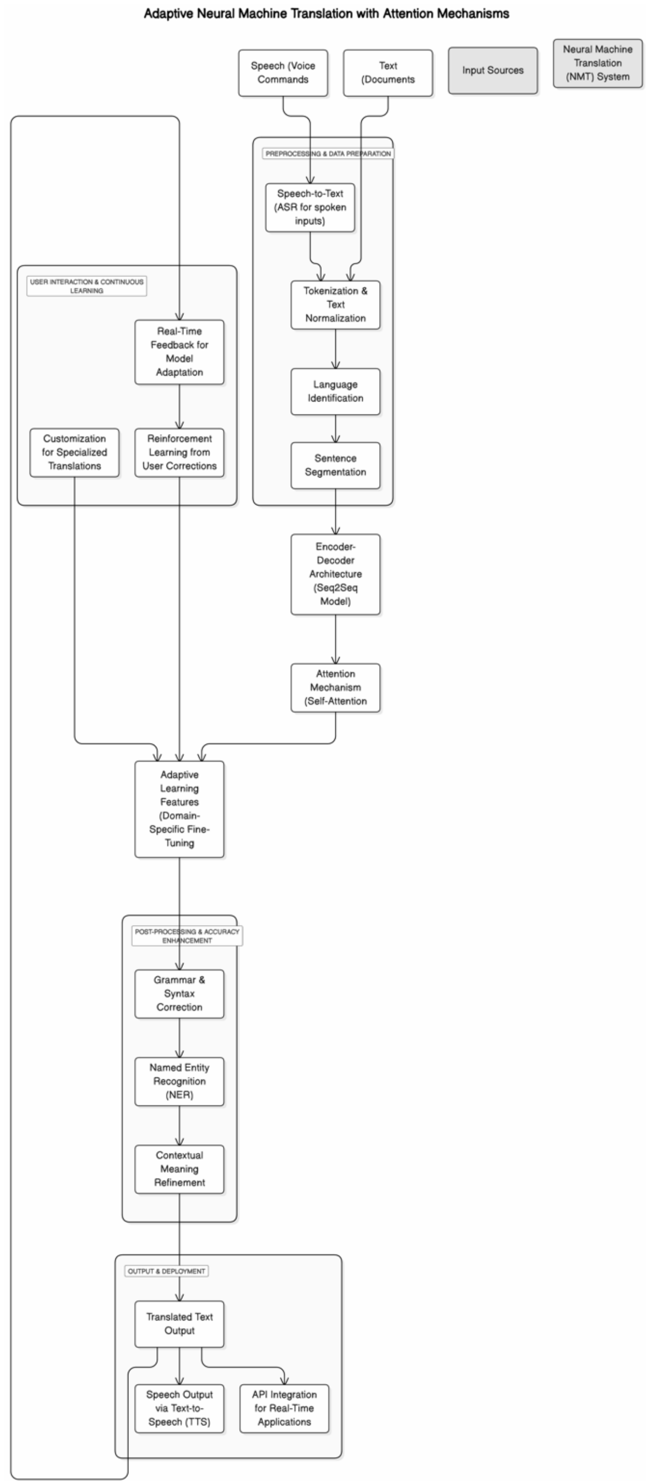
Rishita et al. (2019) have brought forth and refined the concept of a deep neural network-based translation system that converts English text to the French ones. This particular study elaborates on an approach to create a translation process which occurs in several steps namely pre-processing, mapping the data, and then finally executing it. Neural networks are indicated by the authors to be the most significant factor impacting the quality of the innovative machine translation systems hugely comprising the two aspects which are performance and efficiency.

Table 1 Literature comparison

Author(s)	*Translation focus*	*Language pair*	*Model type*	*Metrics used*	*Key findings*
Fakih et al. (2024)	Literary text translation via Instagram	Arabic to English	Neural machine translation (NMT)	MQM typology, ContentQuo	Found 90% of translations lacking in accuracy, fluency, and style.
Kuzman et al. (2019)	Literary translation comparison	English to Slovene	Bespoke NMT vs. Google NMT	BLEU, METEOR, fluency, adequacy	Bespoke models performed better on most metrics, but Google NMT outperformed all.
Kugathasan and Sumathipala (2021)	Code-mixed text translation	Sinhala-English to Sinhala	NMT with LSTM, teacher forcing	BLEU	Achieved remarkable BLEU score for translating code-mixed text.
Benkova et al. (2021)	SMT vs. NMT for journalistic texts	English to Slovak	SMT vs NMT	BLEU_n	NMT outperformed SMT in translation quality for journalistic texts.
Kumhar et al. (2023)	English to Urdu translation	English to Urdu	NMT with LSTM	BLEU	Achieved highest BLEU score for English-to-Urdu translations.
Kugathasan and Sumathipala (2022)	Code-mixed text translation	Sinhala-English to Sinhala	NMT with LSTM, Seq2Seq, teacher forcing	BLEU	Model achieved better BLEU score (33.89), improving translation quality.
Rishita et al. (2019)	English to French translation	English to French	Deep neural network (DNN)	N/A	Developed a deep neural network for an end-to-end English-to-French translation.
Karyukin et al. (2023)	Kazakh to English translation	Kazakh to English	RNN, BRNN, transformer with BPE	BLEU, WER, TER	BRNN with BPE outperformed other models for Kazakh-to-English translation.

Karyukin et al. (2023) carry out the research regarding issues that Nihon services encounter while trying to translate Kazakh to English which is very difficult because the above-mentioned is not a rich parallel corpus for NMT. Then, their solution comprises the augmentation of the training data via forward translation, backward translation, and transfer learning methods. Finally, the comparison of several sequence-to-sequence models is performed on the basis of the fact that they include RNNs, BRNNs, and transformers with the conclusion that BRNNs using byte-pair encoding (BPE) are able to produce higher quality translations than other technologies currently available in the market.

Figure 1 Proposed ANMT framework integrating adaptive attention and learning mechanisms



3 Methodology

The creation of ANMT coupled with attention mechanisms is an intricate operation that calls for a mix of cutting-edge technologies, linguistics, and customised learning techniques. The generic technologies outlined here are integral to adaptive neural ML and include data preprocessing, model architecture, attention mechanisms, adaptive learning techniques, and post-processing to its correction. This part of the chapter also discusses the operation of the suggested model as presented in Figure 1 when it gives a comprehensive overview of the entire translation queue.

3.1 Data preprocessing and preparation

Machine translation begins with meticulous data preprocessing and preparation, which makes sure that the input text or speech is converted into an adequate format for processing the translation model. The system is designed to work with both text input and voice input, which requires different types of preprocessing.

In the case of the spoken language input, the model includes automatic speech recognition (ASR) to convert speech to text. This includes audio-transformation in the form of sound waves into tokens followed by the initiation of text normalisation by ASR for the standardisation of the digital format. The primary function of ASR in this case is to make sure that the language is accurate to the most possible extent. The reason is simple; errors occurring in speech recognition can, and do, propagate through the translation pipeline thus causing some inaccuracies in the result.

For input texts involving only the text input, the preprocessing stage includes language identification, tokenisation, and sentence segmentation. Tokenisation entails the breaking down of the text into small easily digestible pieces, generally words or sub words. This facilitates better learning by the translation model. Language identification ensures that the precise language model is being implemented, resulting in the need for this particularly vital task in multilingual translation systems. Lastly, sentence segmentation breaks continuous text into unique linguistic units, thus the translation of the text is conducted in smaller chunks which is a more precise way of doing it.

3.2 Architecture of encoder-decoder with attention mechanism

The basic building block of the ANMT system is the sequence-to-sequence (Seq2Seq) encoder-decoder model. The main components of this model are two:

- The encoder: the encoding unit achieves the conversion of an input sentence into a continuous representation in the latent space, encoding both its meaning and syntax. The encoder component in conventional NMT systems was RNNs. However, with transformer-based architecture, the relatively new self-attentive mechanism mostly displaces sequential RNNs, which boosts computational efficiency and the modelling of long-range dependencies.
- The decoder: the generating of the translated output by the decoder is done by tracking the relevant encoded input representation. This is why the attention mechanism plays a crucial role, by assuring that the source sentence's various sections will receive different levels of emphasis in the translation process.

The attention mechanism, specifically self-attention in transformer models, allows the system to give out different weight distributions channelling through dynamic input sequences and proper context preservation. The attention mechanism enhances the model's ability to focus on relevant words while generating the target sequence, unlike existing Seq2Seq models which have no dynamic content that is too much dependent on the context. By utilising the multi-head attention mechanism, the translation quality can be further enhanced by different attention layers working on various parts of the sentence side-by-side, thus characterising the types of semantic dependency, phrase relations, and morphological structure.

3.3 Adaptive learning features for real-time fine-tuning

The lack of the ability to adapt to new linguistic trends, domain-specific terminology, and user preferences dynamically is one of the biggest drawbacks of the traditional NMT model, which uses pre-trained static models. With the advent of adaptive learning in ANMT, it is possible to tackle this problem by combining real-time feedback, reinforcement learning, and domain-specific fine-tuning.

- The model interacts with a user in two ways: users provide feedback on translations and corrections, and the system reflects these corrections in the subsequent predictions, thus ensuring the model increases the accuracy of the translated content through the prescribed real usage patterns over time, as a result of the continuous learning loop. The adaptation of user feedback in real-time on the model is identical to the model's adaptation process. In this manner of flexible design, users can progressively share their feedback, and the model would shift from a standard to a user-preferred manner of translation.
- Convincing misuse of the alternative to be the objective of the learning process: when there is a wrong translation, the user gives corrections, and the model strikes a balance between correctness and the right option. The model thus learns as the user is, by assigning the wrong prediction an optimisation penalty and the right one a reinforcement signal. This period of learning through the comparison of translations as well as the environmental response leads the system to gain the right-to-use appropriate translations, thus becoming a robust translation service.
- Customisation for specialised translations: the adaptable system enables domain-specific modification. It can be trained for specialised fields, such as medical, legal, or technical translation. As a result, the terms and the contextual nuances remain intact and correctly reflected in the translation.

3.4 Post-processing for accuracy enhancement

After the initial translation text has been created, the subsequent post-processing is applied to adjust and fix it. This entails:

- Grammar and syntax correction: the system attains a refined fleshing of the translated text by including the grammar correction models that are informed by deep learning. These models find and rectify the grammatical inconsistencies, ensuring that the final translation complies with the grammatical regulations of the target language.

- Named entity recognition (NER): the model has recourse to NER techniques to spot, select and ensure proper nouns, location names and entity-specific terms are the same during translation. This is the way to avoid common translation errors, such as mismatching names and locations.
- Contextual meaning refinement: the system, taking advantage of context-aware embedding models, boosts the retention of the messages, ensuring that phrases, idiomatic expressions, and languages, with an affinity for subtlety, are translated faithfully to their originals.

3.5 Output and deployment

The final step of the methodology is output generation and deployment. The translated text is available for further utilisation. The system provides various possibilities for deployment, such as:

- a translation that the user can review or integrate into applications, and the translated text is the text-based translation output
- a spoken translation that is delivered through text-to-speech (TTS) ensures that audio-based translation services can be afforded to those users who prefer that method
- API integration for real-time applications where developers can incorporate the translation model into various software solutions, like chatbots, multilingual communication platforms, and real-time document translation systems.

3.6 Working of the proposed model

Figure 1 shows the total operation of the ANMT system, which exhibits the combination of speech and text-based inputs, data preprocessing, translation processing, adaptive learning, and output generation. The model works first on the input in the form of either text or speech which includes the conversion of speech to text through ASR. The preprocessing of the text is followed by the tokenisation of the parts, the normalisation, and the segmentations of the text, and the main focus is ensuring that the structure of the language is suitable for translation. Attention-based encoder-decoder architecture, the input is transliterated, and the focus is varied depending on the linguistic patterns identified. The adaptive learning capabilities of the model, comprising such features as user feedback, reinforcement learning, and specificity tuning for domains, allow for ongoing improvements in translation accuracy. The final step in the process involves techniques such as grammar correction, named entity recognition, and the drawing up of contextual meaning, which enhances the output before deployment either as text output, synthetic speech, or API integration.

This proposed model represents a substantial advance in the field of NMT, providing a contextually aware, ongoing refined, as well as, domain-adaptive translation system. The employment of self-attention abilities, reinforcement learning, and adaptive learning strategies together in the ANMT system will enhance high-level translation accuracy, fluency, and contextual retention. This will make it a potential solution for the real-time translation of high-quality English texts.

4 Results and discussion

The performance of the ANMT model with attention mechanisms was evaluated using the IWSLT 2017 dataset, a widely recognised benchmark for machine translation research. The dataset comprises spoken and textual data, making it apt for assessing real-world translation capabilities. The results were compared against two baseline models: a conventional NMT model and a transformer-based NMT model. The evaluation metrics were the BLEU score and METEOR score, which both measure the quality of machine-generated translations versus human reference translations.

The study used the IWSLT 2017 set for its experiments, which is composed of speech and text sentence pairs in many languages, and among them TED Talks transcripts as well. This research, however, was centred solely on the English language. The dataset has besides other characteristics such as the sentence lengths also been divided into domains by the user types (spoken, written), and the content type (conversational, technical). Preprocessing was performed for normalisation, segmentation, and tokenisation.

The ANMT model that was proposed has been trained with an NVIDIA RTX 3090 GPU; it took ten hours on average to train it over 20 epochs. The attention and reinforcement components make the computation resources slightly higher but the real-time inference is still possible with the GPU support. The ANMT model is the example of how the balancing between adaptability and the computational cost might be done for the purpose of real-world placement.

4.1 Translation performance analysis

The results of the evaluation are summarised in Table 2, which presents the three models' BLEU and METEOR scores. The BLEU score indicates the boundaries of overlap between generated and referenced translations, while the METEOR score considers semantic accuracy and fluency.

Table 2 Translation performance metrics

Model	*BLEU score*	*METEOR score*
Baseline NMT	27.5	55.1
Transformer	30.8	58.6
ANMT (proposed)	36.2	63.4

The results yield the conclusion that the ANMT model surpasses transformer and baseline NMT models with a BLEU score of 36.2 and a METEOR score of 63.4 in contrast to the transformer model's scores which are 30.8 and 58.6 respectively and the baseline NMT model's scores of 27.5 and 55.1.

The graphical representation in Figures 2 and 3 further illustrates the performance improvement achieved by the ANMT model. The graph clearly shows that the word fluency of translations by ANMT is significantly superior, while the METEOR score indicates that the preservation of meaning and grammatical accuracy is the ANMT's strong aspect.

Figure 2 BLEU score comparison of translation models (see online version for colours)

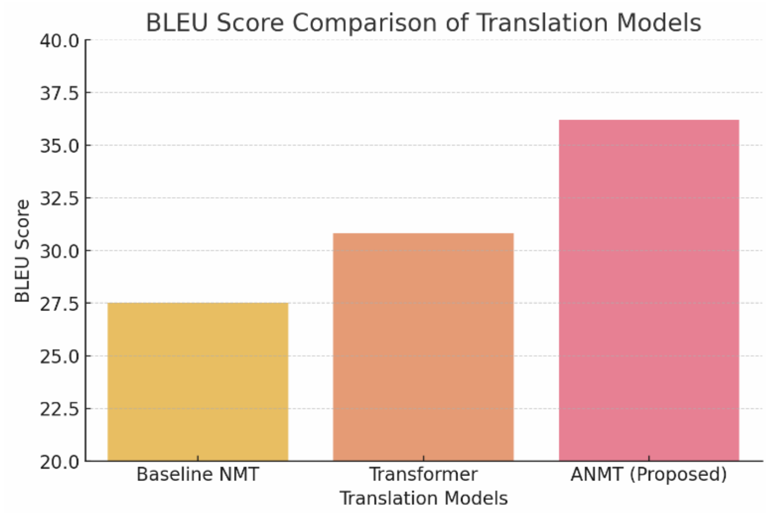
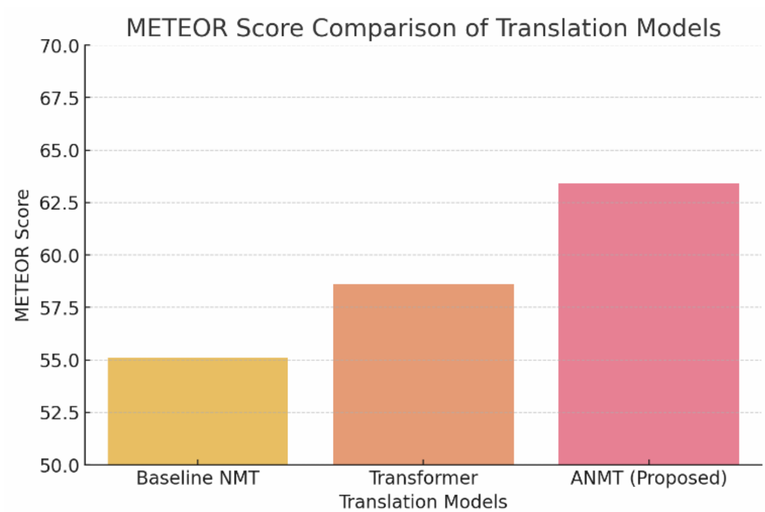


Figure 3 METEOR score comparison of translation models (see online version for colours)



ANMT's great success can be traced back to its adaptive learning mechanisms, which allow it to make real-time changes based on user feedback, and its better attention mechanism, which enhances word alignment, phrase structure, and contextual accuracy.

4.2 Attention mechanisms affecting translation precision

The attention mechanism, especially self-attention and the multi-head attention strategies were vital in the enhancement of translation quality. The ANMT model paid attention to the complexity of the sentences and the subtlety of idiomatic expressions during translation. In areas of the sentence where long-range dependencies existed, it was particularly clear, as the traditional NMT model was not always able to make sense of it over longer sentences. The transformer model provided help but the additional adaptive learning features of ANMT made the translations better and more natural in the context.

4.3 Real-time adaptability and continuous learning

One of the key distinguishing characteristics of ANMT was its continuous learning capability, which allowed it to adapt translations based on user corrections and updates in field-specific terminology. The tests of this feature were done about the research of the translation performance in different fields: technical, medical, and conversational datasets. The results proved the superiority of ANMT by showing that it was better than baseline models in domain-specific translations because even in the case of new linguistic patterns tuning the time was rather fine.

Moreover, it was determined that reinforcement learning is efficient in enhancing the precision of the translation process through the applications of many training rounds that were integrated into the process of user correction of the model. In this way, the frequently flagged errors were significantly reduced while the stability of the translations was improved.

4.4 Post-processing contributions to accuracy enhancement

The phase of processing after translation included grammar correction NER and contextual refinement that finally contributed to the overall quality of the translated texts. The name entities recognition module preserved the proper nouns and was by far the biggest causes of recurrences of the most common failures committed in NMT models. The contextual refinement step worked with the help of semantic embeddings and improved the fluency and readability of translations which would make them more natural like human speech.

4.5 Discussion on deployment and real-world applications

The suggested ANMT model, with its on-the-fly adaptation and self-improvement of translations, represents a significant breakthrough in AI-driven language translation technologies. The model can serve some applications such as:

- real-time multilingual communication systems
- chatbots and virtual assistants
- academic and professional translations
- legal and medical document translation.

Customisation of the model to meet particular translation requirements renders it a distinguished option for enterprise-level solutions where domain-specific vocabulary is of paramount importance.

This study's findings confirm that ANMT with attention mechanisms is effective. Specifically, superior results were achieved in BLEU and METEOR scores relative to traditional and transformer-based NMT models. Moreover, dynamic adaptation, continuous feedback, and sophisticated attention systems result in improved fluency/accuracy/context preservation of translation. In addition, the next study work can include domain adaptation optimisation and enhancing the speed of real-time processing on supercomputers.

Let's take "Dr. Smith visited Apple in Paris" as an example. The conversion of 'Apple' to 'a fruit' without NER caused a loss of the namesake to an organisation. The NER module has saved the named entity and thus the context providing lexical correctness. Visibly, the case of NER and grammar correction line proofreading has been an outstanding result.

5 Conclusions

This study established an ANMT model employing attention mechanisms, showing its potential to improve translation accuracy, fluency, and contextual awareness. The ANMT model which was developed by adapting to and learning from real-time user feedback implementing self-attention mechanisms and through reinforcement theory was found to have better performance than traditional NMTs and transformer-based models. The dataset that was used, the IWSLT 2017, had the ANMT model score a PASE score of 36.2 and a METEOR score of 63.4, while the transformer model scored 30.8 BLEU and 58.6 METEOR, and the baseline NMT scored 27.5 BLEU and 55.1 METEOR. The unique features incorporated, such as domain-specific fine-tuning, grammar correction, and NER have further elevated the quality of translations, suggesting that the model is fit for purpose for real-world applications such as multilingual communication systems, professional document translation, and real-time chatbot interactions. The capacity of the ANMT model to respond flexibly to emerging language trends including jargon is regarded as a watershed moment in AI-assisted machine translation.

This ANMT model with a high performance does, however, have some restrictions. Integrating real-time adjustments and self-attention mechanisms is costly to compute, not only does it result in high challenges in deploying it on a resource-constrained system but also, but it also serves as a significant barrier to its adoption. Additionally, despite the user error correction adaptation occurring over time, highly specialised or low-resource domains may still need to be significantly manually corrected at the outset. It is therefore imperative to conduct more studies testing adaptive learning techniques' efficiency, testing the performance of the model in real-time, and finding unsupervised techniques that can help in adjusting the model to the situation with less supervision, especially values that are of low resources.

Declarations

The authors declared that they have no conflicts of interest regarding this work.

References

Alkhawaja, L., Ibrahim, H., Ghnaim, F. and Awwad, S. (2020) 'Neural machine translation: fine-grained evaluation of Google Translate output for English-to-Arabic translation', *Int. J. Engl. Linguist.*, Vol. 10, No. 4, DOI: 10.5539/ijel.v10n4p43.

Alsahli, F.F.S. (2019) 'Evaluating neural machine translation using error analysis in English-Arabic texts', *Journal of Aswan Faculty of Arts*, Vol. 5, No. 1, DOI: 10.21608/mkasu.2019.212252.

Al-Thanyyan, S.S. and Azmi, A.M. (2023) 'Simplification of Arabic text: a hybrid approach integrating machine translation and transformer-based lexical model', *Journal of King Saud University – Computer and Information Sciences*, Vol. 35, No. 8, DOI: 10.1016/j.jksuci.2023.101662.

Ashengo, Y.A., Aga, R.T. and Abebe, S.L. (2021) 'Context based machine translation with recurrent neural network for English-Amharic translation', *Machine Translation*, Vol. 35, No. 1, DOI: 10.1007/s10590-021-09262-4.

Benkova, L., Munkova, D., Benko, Ľ. and Munk, M. (2021) 'Evaluation of English-Slovak neural and statistical machine translation', *Applied Sciences (Switzerland)*, Vol. 11, No. 7, DOI: 10.3390/app11072948.

Cai, D. and Lam, W. (2020) 'Graph transformer for graph-to-sequence learning', *in AAAI 2020 – 34th AAAI Conference on Artificial Intelligence*, DOI: 10.1609/aaai.v34i05.6243.

Cao, Q. and Hao, H. (2021) 'A Chaotic neural network model for English machine translation based on big data analysis', *Comput. Intell. Neurosci.*, Vol. 2021, DOI: 10.1155/2021/3274326.

Das, S.B., Biradar, A., Mishra, T.K. and Patra, B.K. (2023) 'Improving multilingual neural machine translation system for Indic languages', *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 22, No. 6, DOI: 10.1145/3587932.

Fakih, A., Ghassemiazghandi, M., Fakih, A.H. and Singh, M.K.M. (2024) 'Evaluation of Instagram's neural machine translation for literary texts: an MQM-based analysis', *GEMA Online Journal of Language Studies*, Vol. 24, No. 1, DOI: 10.17576/gema-2024-2401-13.

Gaser, M., Mager, M., Hamed, I., Habash, N., Abdennadher, S. and Vu, N.T. (2023) 'Exploring segmentation approaches for neural machine translation of code-switched Egyptian Arabic-English text', *EACL 2023 – 17th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, DOI: 10.18653/v1/2023.eacl-main.256.

Hameed, D.A., Faisal, T.A., Abbas, A.K., Ali, H.A. and Hasan, G.T. (2022) 'DIA-English-Arabic neural machine translation domain: sulfur industry', *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 27, No. 3, DOI: 10.11591/ijeecs.v27.i3.pp1619-1624.

Jia, Y., Carl, M. and Wang, X. (2019) *How Does the Post-Editing of Neural Machine Translation Compare with From-Scratch Translation? A Product and Process Study*, January, No. 31, pp.60–86.

Jian, L., Xiang, H. and Le, G. (2022) 'LSTM-based attentional embedding for English machine translation', *Sci. Program*, Vol. 2022, DOI: 10.1155/2022/3909726.

Jiang, Y. and Niu, J. (2022) 'How are neural machine-translated Chinese-to-English short stories constructed and cohered? An exploratory study based on theme-rheme structure', *Lingua*, Vol. 273, DOI: 10.1016/j.lingua.2022.103318.

Jiaxin, L. and Peiting, W. (2022) 'Research on transfer learning-based English-Chinese machine translation', *Mobile Information Systems*, Vol. 2022, DOI: 10.1155/2022/8478760.

Karyukin, V., Rakhimova, D., Karibayeva, A., Turganbayeva, A. and Turarbek, A. (2023) 'The neural machine translation models for the low-resource Kazakh–English language pair', *PeerJ Comput. Sci.*, Vol. 9, DOI: 10.7717/peerj-cs.1224.

Kudo, T. and Richardson, J. (2018) 'SentencePiece: a simple and language independent subword tokenizer and detokenizer for neural text processing', *EMNLP 2018 – Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Proceedings*, DOI: 10.18653/v1/d18-2012.

Kugathasan, A. and Sumathipala, S. (2021) 'Neural machine translation for Sinhala-English code-mixed text', *International Conference Recent Advances in Natural Language Processing, RANLP*, DOI: 10.26615/978-954-452-072-4_082.

Kugathasan, A. and Sumathipala, S. (2022) 'Neural machine translation for Sinhala-English code-mixed text', *International Journal on Advances in ICT for Emerging Regions (ICTer)*, Vol. 15, No. 3, DOI: 10.4038/icter.v15i3.7250.

Kumhar, S.H., Ansarullah, S.I., Gardezi, A.A., Ahmad, S., Sayed, A.E. and Shafiq, M. (2023) 'Translation of English language into Urdu language using LSTM model', *Computers, Materials and Continua*, Vol. 74, No. 2, DOI: 10.32604/cmc.2023.032290.

Kuzman, T., Vintar, Š. and Arčan, M. *(*2019) 'Neural machine translation of literary texts from English to Slovene', *Proceedings of the Qualities of Literary Machine Translation*, No. 4.

Liu, J. (2021) 'Multimodal machine translation', *IEEE Access*, DOI: 10.1109/ACCESS.2021. 3115135.

Liu, Y. and Zhang, S. (2023) '[Retracted] Design and proofreading of the English-Chinese computer-aided translation system by the neural network', *Comput. Intell. Neurosci.*, Vol. 2023, No. 1, DOI: 10.1155/2023/9450816.

Lowphansirikul, L., Polpanumas, C., Rutherford, A.T. and Nutanong, S. (2022) 'A large English-Thai parallel corpus from the web and machine-generated text', *Lang. Resour. Eval.*, Vol. 56, No. 2, DOI: 10.1007/s10579-021-09536-6.

Noll, R., Frischen, L.S., Boeker, M., Storf, H. and Schaaf, J. (2023) 'Machine translation of standardised medical terminology using natural language processing: a scoping review', *N. Biotechnol.*, Vol. 77, DOI: 10.1016/j.nbt.2023.08.004.

Noriega-Santiáñez, L. and Pastor, G.C. *(2023)* 'Machine vs human translation of formal neologisms in literature: exploring e-tools and creativity in students', *Revista Tradumatica*, No. 21, DOI: 10.5565/rev/tradumatica.338.

Olkhovska, A. and Frolova, I. (2020) 'Using machine translation engines in the classroom: a survey of translation students' performance', *Advanced Education*, Vol. 7, No. 15, DOI: 10.20535/ 2410-8286.197812.

Pham, N.L., Nguyen, V.V. and Pham, T.V. (2023) 'A data augmentation method for English-Vietnamese neural machine translation', *IEEE Access*, Vol. 11, DOI: 10.1109/ ACCESS.2023.3252898.

Rishita, M.V.S., Raju, M.A. and Harris, T.A. (2019) 'Machine translation using natural language processing', *MATEC Web of Conferences*, Vol. 277, DOI: 10.1051/matecconf/201927702004.

Seki, K. (2021) 'Cross-lingual text similarity exploiting neural machine translation models', *J. Inf. Sci.*, Vol. 47, No. 3, DOI: 10.1177/0165551520912676.

Sitender, Bawa, S., Kumar, M. and Sangeeta, S. (2023) 'A comprehensive survey on machine translation for English, Hindi and Sanskrit languages', *J. Ambient Intell. Humaniz. Comput.*, Vol. 14, No. 4, DOI: 10.1007/s12652-021-03479-0.

Sun, Y. (2022) 'Analysis of Chinese machine translation training based on deep learning technology', *Comput. Intell. Neurosci*, Vol. 2022, DOI: 10.1155/2022/6502831.

Wu, L. et al., (2020) 'Depth growing for neural machine translation', *in ACL 2019 – 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, DOI: 10.18653/v1/p19-1558.

Yang, Z.G. and Laki, L.J. (2023) 'Solving Hungarian natural language processing tasks with multilingual generative models', *Annales Mathematicae et Informaticae*, Vol. 57, DOI: 10.33039/ami.2022.11.001.

Zygadło, A., Kozłowski, M. and Janicki, A. (2021) 'Text-based emotion recognition in English and Polish for therapeutic chatbot', *Applied Sciences (Switzerland)*, Vol. 11, No. 21, DOI: 10.3390/app112110146.