Journal of Electrical Systems
and Information Technology

REVIEW

Open Access

Overview of deep learning and large language models in machine translation: a special perspective on the Arabic language

Sanaa Abou Elhamayed[1*] and Mohamed Nour[1]

*Correspondence:
sanaa-hamayed@hotmail.com

[1] Electronics Research Institute, Cairo, Egypt

Abstract

This work aims to present an overview of using some artificial intelligence (AI) models in machine translation (MT). This work aims to integrate machine learning (ML), deep learning (DL), large language models (LLMs) to enhance machine translation between natural languages. The focus is directed to present the neural-based machine translation (NMT), and some DL models are presented. The bidirectional-encoder-representation from transformer (BERT) and LLMs are presented to utilize the big amount of textual data to learn translation patterns. The main measurable criteria that are used to evaluate the performance of MT and Arabic machine translation (AMT) are also presented. Some linguistic and technical challenges of MT and AMT are discussed. Some key points of future works in NMT are mentioned. A comparative study among some recent published related works is presented. A critical survey is presented to show the important role of DL and LLMs in MT. Some open-source toolkits, datasets and some commercial MT systems are collected and briefly presented. This work is expected to be useful for those people interested to know the up-to-date knowledge of MT and the potential role of DL and LLM in automatic translation.

Keywords: Neural machine translation, Deep learning, Large language models, Translation to and from Arabic language, Open-source toolkits

Introduction and background

Natural language processing (NLP) is a vital area of research which focuses on how to use the computer and information technology to process, manipulate, and understand natural languages. A lot of researchers consider that the study of NLP is an important area of artificial intelligence. NLP involves studying the main linguistic aspects of human–human and human–machine communication. In the early years, the research efforts of NLP were directed toward the automatic analysis on natural language structures. In recent years, NLP had addressed great research works with machine and deep learning techniques. Examples of the NLP applications include, but not limited to, text mining, text classification, sentiment analysis, language recognizers, chatbots, text analysis, text generation, question-answering systems, text summarization, machine translation, and others [1–3].

Textual machine translation is a vital sub-field of natural language processing which is concerned with translating a given text from one language to another such that both the given text and the translated one have the same meaning. MT was conducted and handled by several researchers since more than three decades ago. A lot of MT systems were developed for several languages and millions of people gained access. In MT, both the source and target languages have a wide array of linguistic dissimilarities. The MT process can analyze the given text in source language and can generate the corresponding text in the target language. The MT process deals with very important themes such as morphology, syntax, semantics, additional varieties of grammatical complexities of natural languages, and others [4, 5].

Arabic is one of the common natural languages in the world. Arabic is the native language of hundreds of millions of speakers. Arabic is one of the official six languages of the United Nations. In recent years, the amount of research work concerning the translation to and from the Arabic language has dramatically increased and become important for several applications. The amount of research work of MT to and from the Arabic language is still less compared to some other natural languages like English. The Arabic morphology, syntax, semantics, other linguistic aspects and others made the translation to and from Arabic language is a big challenge [6].

There are several approaches of machine translation. One of the famous machine translation approaches is that one based on rule-based machine translation (RBMT). Such type is based on the linguistic knowledge which is encoded by experts [5]. Another approach of machine translation is that one called statistical machine translation (SMT). SMT can learn latent structures such as word alignment or phrases directly from parallel corpora. In SMT, the word from the target language is a translation of the source language word set with several possibilities. In SMT, there are two important themes namely decoding complexity and target language reordering [7]. Neural machine translation (NMT) is a good approach because it achieved promising results in many translation tasks. NMT is considered a fully automated neural network-based translation technology. NMT depends on encoders and decoders. NMT can employ continuous representations instead of discrete symbolic representations in SMT. A lot of researchers consider that NMT is the main pivot behind several commercial MT systems. Some researchers mentioned that NMT is superior to the conventional approaches [7–9].

This survey is valuable for those persons interested in MT research work as it highlights on the followings:

- Challenges of MT
- MT-based approaches and datasets
- Performance evaluation of MT approaches using different metrics
- Listing some key points as future work directions in MT in general and AMT in particular.

The organization of this research work will be as follows: Section "Challenges of machine translation" mentions some key points of the challenges of machine translation. Sections "The main models of machine translation" and "Deep learning in textual language translation" present respectively the main models of MT and some deep learning approaches for MT. Section “Large Language Models and ChatGPT in Arabic machine translation” presents the role of large language models and ChatGPT in machine translation. Some resources and toolkits are presented in Sect. “Some resources and NMT toolkits for Arabic machine translation”. The main common evaluation metrics are briefly mentioned in Sect. “Metrics for evaluating machine translation systems”. A brief comparison of some machine and deep learning approaches of translation to and from Arabic language is presented in Sect. “Comparison of some machine and deep learning approaches to translate to and from Arabic text”. The observations from the adopted related works and some commercial Arabic machine translation systems are highlighted respectively in Sects. “Observations from the related work and tables" and "Some commercial Arabic machine translation systems”. Finally, Sects." Discussion" and “Conclusion” present respectively the discussion and conclusion.

The main contribution of this work can be briefly mentioned as follows:

- In their development, MT approaches are those approaches based on rules, statistics, and neural networks. Neural machine translation (NMT) has shown better results for many translation language pairs than the others. This includes the translation to and from the Arabic language.
- Artificial intelligence (AI) and generative artificial intelligence (GAI) are promising in the new age of machine learning and intelligent approaches. They can generate novel content such as text.
- Deep learning (DL) models such as recurrent neural networks (RNNs) can efficiently model the sequence of data, so RNNs have achieved the state-of-the-art performance in developing neural machine translation.
- Large language models (LLMs) are artificial neural networks that exploit the transformer architecture. LLMs can be used to generate text which is necessary in any automatic translation.
- The integration of LLMs in MT is a promising paradigm. LLMs such as ChatGPT can present translation with high quality and promising accuracy.
- The emphasis on Arabic machine translation (AMT) is valuable. Dealing with Arabic language is still limited compared to other languages like English for example. Morphology, syntax, and semantics of Arabic language have special features and characterization.
- BLEU is an important score to identify the reliability and quality of MT systems.

Challenges of machine translation

As mentioned before, text translation from a natural language to another is important for several applications. By analyzing the output from a translation process, it was found that the quality or accuracy of MT is affected by key parameters. Those parameters may present improvement or degradation in the translation accuracy depending on how to handle such parameters. Examples of those parameters are word morphology, complexity of sentences’ structures, different categories of ambiguities, meaning understanding, dialects, domain dependence, sentence length, and others. Some challenges of MT are briefly mentioned as follows:

Accuracy of Machine Translation (MT)

The majority of MT models produce some errors in translation. The errors can take place due to omissions, meaning changes, existence of ambiguities, complexity in sentences' structures, misunderstanding of meaning of phrases and sentences, syntax errors, pronoun resolution errors, segmentation errors, idiom translation errors and others [1, 10, 11].

Translation of Long Sentences

If a source sentence is long, the quality or accuracy of translation may decrease. This can take place in long sentences and also in those long sentences containing pronouns, i.e., long sentences have significant effects on translation accuracy [8].

Features of Source and Target Languages

Features and characterization of the source and target languages are considered a big challenge in MT. A natural language like Arabic has specific features. Examples of such features are rich morphology, disjoint and joint pronouns, word sense disambiguation, word order, unvocalized and vocalized words, Arabic dialects, multiple meanings of an Arabic word, category ambiguity, structure ambiguity, semantic ambiguity, semantic understanding, of phrases and sentences, and others [4, 8]. Moreover, bad interpretation of Arabic words, with joint pronouns' has a bad effect on the translation accuracy [6. 7]. Identification of the correct sense of an Arabic word with different meaning is also a challenge. Till now, there are some difficulties in sense disambiguation of Arabic words. Variations in word order within a single Arabic sentence can yield diverse meanings [6, 9].

Limitations of Neural Machine Translation (NMT)

Although NMT presented better performance than those approaches based on rule-based and statistical-based, the results of NMT are still far from satisfactory. Converting discrete and continuous representation into each other is still a problem in NMT [9].

Domain Dependence

Translation quality is also affected by the domain of text. Out-of-domain text leads to degradation in translation accuracy because MT models are typically trained on specific domains.

Size of Training Dataset

The size and quality of training data are also important. Adopting large and high quality datasets can improve translation accuracy while small datasets lead to poor performance.

Translation of Complex Sentences

Translation of complex sentences is considered a big challenge. Complex sentences are more difficult to present reasonable accurate translation like those occurred in simple and compound sentences. Complex sentences, in a natural language, most probably have complex grammatical structures [6].

The main models of machine translation

As mentioned before, textual machine translation is a process which aims at translating sentences of a source language (S) into text of a target language (T). The models and/or paradigms of machine translation have been continuously improved to make the translation process more efficient. Due to the development of machine translation approaches,

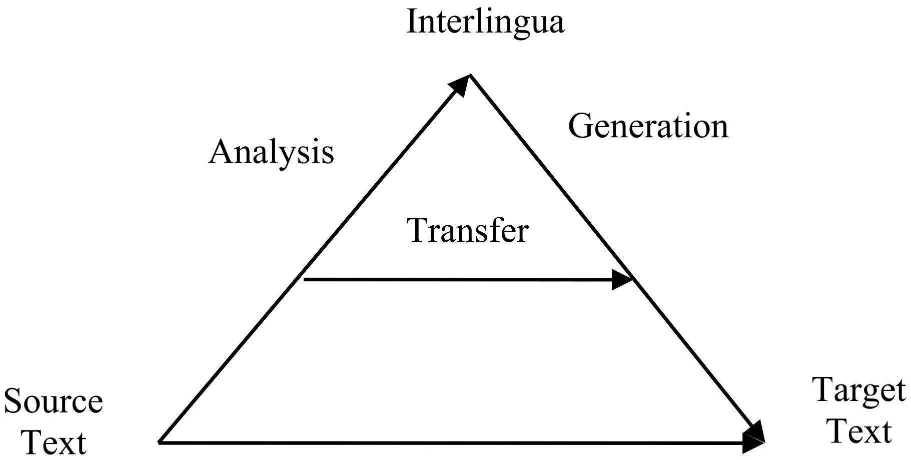

Fig. 1 Paths of rule-based machine translation [3, 12]

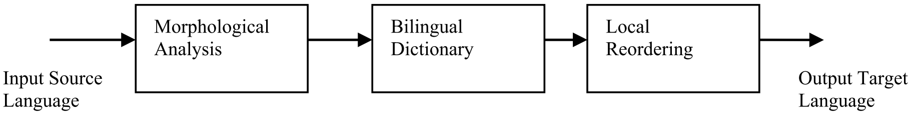

Fig. 2 Key elements of direct machine translation [12]

there are three main models, namely rule-based machine translation (RBMT), statistical-based machine translation (SMT), and neural-based machine translation (NMT). The following sub-sections are briefly highlighting and discussing such models.

Rule-based machine translation

This category is the first model and/or class of machine translation. RBMS is based on the hypothesis that a word in a language has its corresponding word in another language with the same meaning. The translation from a source language to a target language is handled as word replacement. Every word in the source text/sentence should take its corresponding location in the target sentence. RBMT is based on the syntax rules of both the source and target languages [1, 2].

RBMT has three classes and/or paths, namely direct translation, transfer translation, and interlingua translation. Figure 1 shows the main paths of RBMT.

Direct translation

Direct translation is the simplest method to translate a source language to a target language. This type of machine translation is known as word-to-word translation. This is because a sentence in the source language can be translated into a sentence in the target language without adopting any analysis beyond the morphological level. The source sentence/text can be morphologically analyzed, then the translation is done word by word using a bilingual dictionary. Such dictionary is considered as a mapping between each word in the source text and its corresponding one in the target text. Each word in the dictionary may be associated with some simple rules for local reordering. Figure 2 briefly presents the main building blocks and/or elements for direct machine translation.

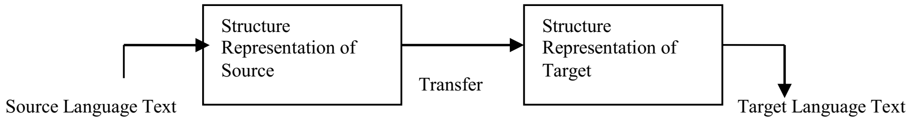
Structure Representation of Source
Transfer
Structure Representation of Target
Source Language Text
Target Language Text

Fig. 3 Elements of transfer machine translation [3, 12]

Transfer translation

This type of translation involves some sorts of lexical, syntactic and semantic information of the source text. The translation between the source and target languages is done on three steps, namely analysis, transfer and generation. In the analysis step, morphological analysis and syntactic analysis are done for the source sentence to create an internal representation. In the transfer step, a set of rules (or transfer rules) are adopted to transform the structural representation of the input sentence into the target one. In the generation step, the output translation is generated from the target representation using a bilingual dictionary. Figure 3 briefly shows the three steps of the transfer machine translation [3, 12].

Interlingua translation

This type of machine translation is using a representation called interlingua which is independent of any natural language. This type of translation is focused on the meaning representation instead of building a specific language-related representation of the source text and then transforming it to the target text. Interlingua translation by this concept has only two steps namely analysis and generation steps. The transfer step in this translation type is becoming no longer needed. The analysis step is adopting the morphological, syntactic, and semantic analysis to transform the source text into a meaning representation. Such representation is called interlingua or abstractive representation. In the generation step, the target output can be generated from the interlingua representation. Figure 4 briefly shows the steps of machine translation interlingua-based [3, 12].

Statistical-based machine translation

Statistical-based machine translation (SMT) is concerned with the translation process from a statistical perspective. SMT aims at statistically finding the words and/or phrases which have the same meaning via a bilingual corpus. SMT showed significant success in different domains like industry, for example, SMT can divide a given sentence into a set of sub-sentences and each can be replaced by a target word or phrase. SMT can build statistical models from a collection of datasets of sentences-aligned parallel corpus [1, 6].

Phrase-based SMT (PBSMT) is a significant type of SMT which showed better performance than that based on word-to-word translation. PBSMT uses an important component namely a phrase-based lexicon. That lexicon is built from the training dataset which is a bilingual corpus, i.e., the phrase-based lexicon pairs phrases in the source language which phrases in the target language. PBSMT has three phases and/or models: translation model, language model, and decoder model. The translation model is trained on a bilingual corpus to estimate the probability of the source sentence being a translated version of the target sentence. The language model is trained on monolingual corpora to improve the output translation. The decoder model computes the maximum probability of product of both the language and translation models [6, 12]. SMT is based on the probability distribution of strings in the target language corresponding to those ones in the source language. The string with maximum probability match can be taken to translate the sentence.

Moreover and using Bayes theorem, the translation process can be represented as follows: given a source sentence s with its translation t, the highest probability sentence is chosen as the best translation using Eq. (1) (Bayes Theorem) [2, 3, 12].

$$p(t/s) = \frac{P(s|t)P(t)}{P(t)} \tag{1}$$

where $P(t)$ is the probability of the language model that ensures the fluency of the generated target output and P(s|t) is the probability of the translation model that ensures the accuracy.

Some researchers mentioned that SMT is involving three translation levels: word-based, phrase-based, and syntax-based. SMT word-based can translate a source sentence word by word to construct a target sentence. Phrase-based SMT can translate the source language phrase to a phrase in the target language. The syntax-based SMT is based on translating syntactic units rather than single words or sequence of words [6]. It is easy to say that the language model in SMT is dedicated to compute P(t), while the translation model is focused to compute P(s|t) to ensure the translation accuracy between the source and target sentences. The decoder model, on the other hand, can find the most probable translation output P(t|s) for the input sentence s from the space of all possible translations of s [12].

Reference [6] mentioned that SMT can build statistical models from a collection of datasets constituting of sentence-aligned parallel corpus called multilingual resource. A brief description about each translation type of SMT is mentioned as follows:

SMT word-based

SMT word-based can translate the source language word by word or more in the target language. [13] mentioned that statistical techniques can be used to translate text from one language to another. That translation is based on a complex glossary of correspondences of fixed locations, i.e., the glossary maps words and phrases to corresponding translations. The SMT word-based concluded that a probabilistic method is required to identify the corresponding words in the target and source sentence. [13] mentioned also that SMT word-based has failed to deal with cases, gender and other themes such as homonymy. Every single word was translated in a single true way.

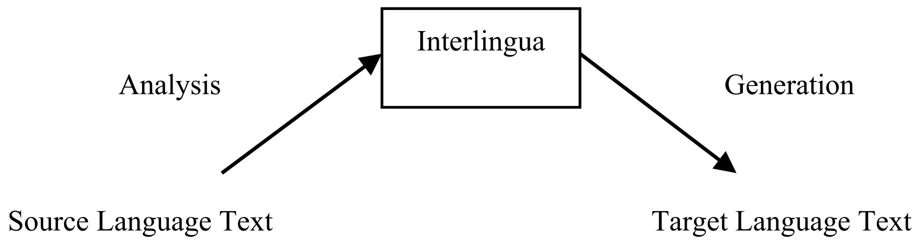

Fig. 4 Elements of machine translation interlingua-based [3, 12]

SMT phrase-based

In SMT phrase-based, there is no restriction of translating a source sentence into a target sentence word by word. A lexical unit is a sequence of words of any length as opposed to a single word in SMT word-based. A score or a weight is given to each pair of units: one unit from the source language and one unit from the target language. Phrase-based translation has a phrase translation probability table to map phrases in the source language to phrases in the target language. Such table is learnt from word alignment models using bilingual corpus [13].

SMT syntax-based

SMT syntax-based can also be used in translating the source language to a target language. This model of translation is briefly based on the idea of translating syntactic units, rather than single words or sequence of words. [6] mentioned that a hierarchical phrase-based model can be constructed to combine the benefits of both SMT phrase-based and SMT syntax-based.

Neural-based machine translation

Neural machine translation (NMT) has achieved recently significant results. NMT is recommended to be used as it can solve the problems of word order errors, morphological errors found in SMT, and syntactic errors. NMT is appreciated to be used for facing the problem of time consuming training and decoding process, i.e., using NMT, multiple features can be trained without prior knowledge and sentence structure can be optimized to obtain better results [2].

The concept of neural networks is based on that idea of how a human brain works. A human brain has billions of neurons connected together. An artificial neural network consists of some neurons connected together via a network. The connection between neurons is represented by weights. Information can pass through the connections to or from neurons which allow a neural network to train, learn and predict. The activation of the artificial neuron S4 can be represented by Eq. (2), where W_i is the weight of each input neuron S_i (S_1, S_2, S_3), while S_4 is the output neuron.

$$S_4 = \sum_{i=}^{3} (S_i \times w_i) \tag{2}$$

The NMT models can be trained on large datasets. Such datasets contain sentences in the source language and their translation in the target language. By training the NMT models, words can be predicted that may appear before or after a given word. Each word can be represented in a vector (sometimes called word embeddings) of real numbers that captures the semantic relationships between a word and the others. The vectors are used as inputs to the neural network model to understand the text meaning [14].

The NMT has two main modules, namely the encoder and decoder modules. The encoder module represents words of a sentence (of the source language) into a sequence of vectors. The sequence of vectors is representing the meaning of words. The encoded sentence can be passed to the decoder network to generate the translation sentence (of the target language) word by word, i.e., The next word is predicted by the previous one. Such prediction can be passed through the neural network to enhance the predictions. This process is repeated till the translation sentence is generated [15].

In other words, it is easy to say that NMT basically aims at building a model for transforming the input sequence into an output sequence via an encoder and decoder architecture. The encoder can transform the input sequence of words to a new form which can be stored and understood by the machine. That form can be taken as text features and then can be converted back by the decoder into a sequence of words in the output [3], i.e., the encoder inputs and encodes a source language sentence into a fixed-length vector in each hidden state. The decoder makes the reverse work by transforming the hidden state vector to the target sentence word by word [1, 6, 12].

Moreover, the language model is considered an important theme in machine translation. Developing an efficient language model can improve the results of the translation process. The language model is based on the Markov assumption as shown in Eq. (3) [1].

$$P(x_1, x_2, \ldots x_T) = \Pi_{l=1}^{L} P\left(x_l | x_1, \ldots\ldots\ldots x_{l-1}\right)$$

$$\Pi_{l=1}^{L} P\left(x_l | x_{l-n}, \ldots\ldots x_{l-1}\right) \tag{3}$$

where x_1, x_2, ...x_l are a sequence of words in a sentence, L is the sentence length, and n is the total number of words which is chosen to simplify the model. It is clear from Eq. (3) that the probability of a sentence equals to the multiplication of probability of each word [1].

Based on the Markov assumption, the translation probability of a source language sentence is modeled into the target language sentence. Assume that a source language sentence $S = \{s_1, s_2, \ldots.s_n\}$ and a target language sentence $T = \{t_1, t_2, \ldots..t_n\}$, the encoder can encode all the words from S into a set of hidden states (h_1, h_2, ...h_n) and passes the fixed size vector N to the decoder. The translation probability with a single neural network can be represented by Eq. (4).

$$P\,(T|S) = \prod_{i=1}^{n} P\,(t_i|\, t_{<i}, S) \tag{4}$$

where $t_{<i}$ is the sequence preceding the ith target word and n is the words in the sentence.

Moreover, NMT can be formed using different architectures.

Also, there are several types of deep learning models. This includes but not limited to convolutional neural networks (CNNs), artificial neural network (ANN), recurrent neural networks (RNNs), and others. This work is focused on the RNN models which are adopted and handled for the translation process.

Recurrent neural network (RNN)

The recurrent neural network (RNN) is mainly composed of encoder and decoder with similar working of sequence-to-sequence learning. RNN can be used for machine translation because it is relevant to sequence modeling tasks. RNN is a recurrent encoder–decoder where the encoder produces the embedding words in the input sentence. The decoder uses attention to compute the target embedding words by using the information of the input words and then generates the target words [14, 16]. RNN is considered a chain of repeating neural network modules. Each module has a hidden state and an output which feeds back to the input of the next module. The hidden state is updated one by one. Each word in the input sentence is converted to vectors and entered to the RNN one by one. The current input vector and the previous hidden state are taken by RNN as inputs that a new hidden state and an output vector are produced [14]. RNN can predict the future words of the sequence based on the past words due to the hidden state which captures the information that the network has learned about the sequence up to that point. Moreover, there are different types of RNN model. Examples of such models are long short-term memory (LSTM), bidirectional LSTM (BiLSTM), and gated recurrent unit (GRU). RNN-based NMT can be applied as feature extractor to compress the source sentence into a feature vector.

Long short-term memory (LSTM) RNN has several types: LSTM is one of them. LSTM, as shown in Fig. 5, has a forget gate, an input gate, and an output gate. Such gates are used to solve vanishing gradient problem [16, 17]. The forget gate can control the forgetting data from previous cell state. The input gate is used to control the new information that added to the cell state while the output gate can control the output from the cell state to the next layer of the network.

The LSTM model can be briefly represented by the following set of mathematical equations:

$$f_t = \sigma\left(W_f h_{t-1} + U_f x_t + b_f\right) \tag{5}$$

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \tag{6}$$

$$k_t = \tanh(W_k h_{t-1} + U_k x_t + b_k) \tag{7}$$

$$O_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \tag{8}$$

$$C_t = C_{t-1} \odot f_t + i_t \odot k_t \tag{9}$$

where the above symbols are x_t: word representation at time t, C_t: memory state, i_t: input gate, f_t: forget gate, O_t: output gate, h_t: hidden state, tanh: tangent hyperbolic function, σ: sigmoid activation function, ⊙: element wise Hadamard product, W U: weight matrices, b: bias.

Equation (5) is used to evaluate the forget gate, and Eqs. (6,7) are used to evaluate the input gate, while Eq. (8) evaluates the output gate. Equation (9) is used to evaluate the vectors f, i, k which represent the LSTM. For more details, the reader can refer to [16, 17].

Bidirectional long short-term memory (BiLSTM) The bidirectional long short-term memory, abbreviated as BiLSTM, is another type of RNN. The BiLSTM has two LSTM cells. An LSTM has a memory cell to store information about the past and three gates to control information flow in and out the cell. In BiLSTM, as shown in Fig. 6, two LSTM cells are used to process the input sequence. The training is done like that happens in LSTM but in BiLSTM, the training is done in both forward and backward directions. The output of the two LSTM cells is combined to produce a single output which represents the entire input sequence. The BiLSTM model can be mathematically represented as follows:

$$h \rightarrow= \tanh(W_{xh} \rightarrow x_t + U_h \rightarrow_h \rightarrow h_{t-1} + b_h \rightarrow) \tag{10}$$

$$h \leftarrow= \tanh(W_{xh} \leftarrow x_t + U_h \leftarrow_h \leftarrow h_{t-1} + b_h \leftarrow) \tag{11}$$

where $h_t = [h^{\rightarrow}, h^{\leftarrow}]$ and $h^{\rightarrow}$, $h^{\leftarrow}$ are the forward hidden vector sequence and the backward hidden vector sequence, respectively. For more details, the reader can refer to [16, 18].

Gated recurrent unit (GRU) The gated recurrent unit, abbreviated as GRU, is another type of RNN. GRU is approximately like the LSTM except it has only two gates: a reset gate and an update gate as shown in Fig. 7. The reset gate is used to control how much forget from the previous hidden state. The update gate is used to control how much add from the new input and how much keep from the previous hidden state [16, 18]. The GRU model can be briefly represented by the following equations:

$${}_{t}^{u} = \sigma(U_u h_{t-1} + W_u x_t) \tag{12}$$

$$r_t = \sigma(U_r h_{t-1} + W_r x_t) \tag{13}$$

u_t: update gate, r_t: reset gate.

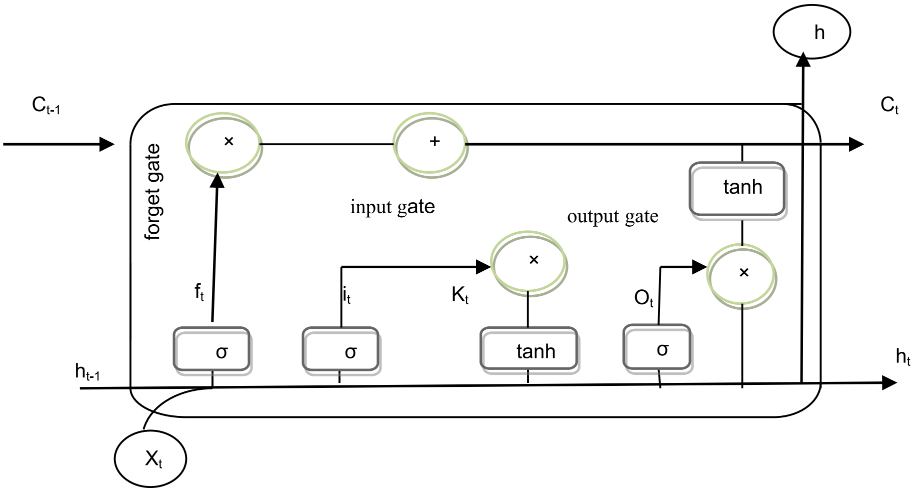

Fig. 5 Architecture of LSTM model [16, 17]

u_t in Eq. (12) is used to evaluate the update gate, while r_t in Eq. 13 is used to evaluate the reset gate. The reset gate, the update gate, and the candidate hidden state are computed from the current input and the previous hidden state. Then, the candidate hidden state is modulated by the reset gate. The previous hidden state and the candidate hidden state combine and compute the new hidden state and the update gate weights this new hidden state. The reset gate and the update gate can learn during training and allow the GRU cell to control its memory. For more details, the reader can refer to [16, 18].

Convolutional neural network (CNN)

The convolutional neural network (CNN)-based NMT is considered a significant concrete architecture as it achieved fruitful results for the word-based machine translation but along with RNN. Compared with RNN, CNN-based NMT models have an advantage in training speed. CNN structure allows parallel computations for its different filters when handling the input data. CNN also can resolve the gradient vanishing problem.

Self-attention-based transformer

This model consists of a stack of layers. Each layer can utilize the self-attention to extract information from the whole sentence. Self-attention can be extended to the processing of input sequences and output sequences. Using this model words, dependency between source sequences with target sequence can be calculated. Self-attention can calculate the words dependency inside the sequence and get an attention-based sequence representation. The transformer-based language models such as bidirectional encoder representation from transformer (BERT) can extend the function of attention to encompass the main task [1, 6, 12].

Deep learning in textual language translation

As mentioned before, a neural network consists of some connected layers. Each layer has a set of nodes, sometimes called neurons, which are concerned with the data processing. A deep learning architecture has several hidden layers inside the neural network design.

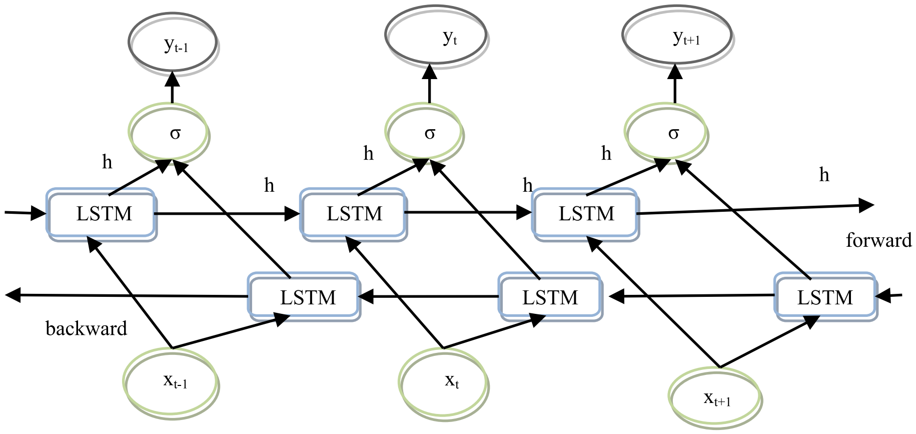

Fig. 6 Structure of BiLSTM [18]

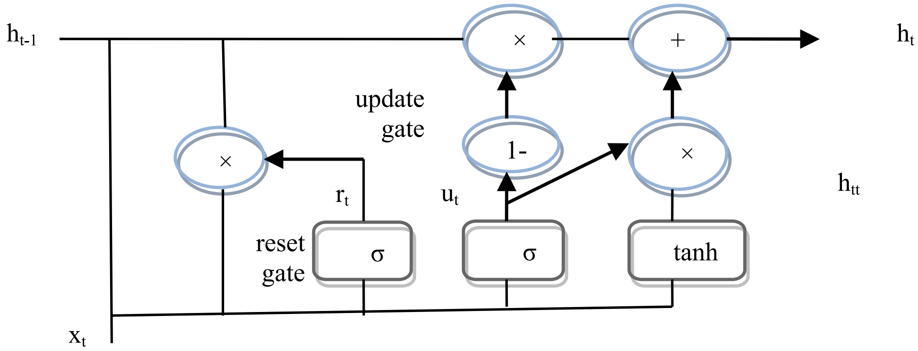

Fig. 7 Structure of GRU [16, 18]

Machine and deep learning approaches are very useful and important in language translation. The use of deep learning models such as transformers plays a vital role also in language translation. This section presents a survey of some deep learning approaches in textual language translation (in general) and Arabic machine translation (in particular). Some of the research efforts presented by other scientists are briefly mentioned as follows:

The work in [4] used a multitask learning model to translate Arabic dialects and modern standard Arabic (MSA) to MSA and English language, respectively. The authors used a BiLSTM encoder for each source language and one BiLSTM decoder shared for all language pairs. The decoder has several common LSTM hidden layers which activated by the ReLU activation function. They used data by concatenating the Levantine dialects together and the Maghrebi dialects together from PADIC corpus and MPCA corpus, respectively. They compared the quality of the single neural machine translation with the multitask learning model by BLEU score. The authors found that the performance of the multitask learning model was higher than the single NMT with more than 15%.

The work in [6] mentioned that NMT depends on the transformer model and multiple attention mechanism. For Arabic language, some authors developed two transformer-based language models namely AraBERT and GigaBERT. The two models are directed to solve a masked language modeling task to predict a masked word from its context. Moreover, the models can be used to resolve a next sentence prediction task.

The work in [19] mentioned that training NMT without large parallel corpora is considered a challenge. Researchers have explored various methodologies to enhance robustness utilize monolingual data and improve efficiency of the translation model.

The work in [20] presented a NMT model incorporating gate mechanisms to optimize monolingual data utilization during training. The model separated the monolingual and parallel data using a gate to distinguish between the types of input sequences. The experimental results on English-German language pairs showed improvements compared to robust baseline models.

The work in [21] presented an approach to develop resilience in both the encoder and decoder elements of NMT models, ensuring consistency in behavior between the original and distributed input. The evaluation of the translation process for English-German and English-French showed significant gains in translation quality.

The work in [22] designed a system called DIA for translating text from English to Arabic. The authors used two dimensional models to enter the meaning of past and future words to improve the encoder output. The encoder was recurrent neural network while the decoder was attention network based on LSTM. The authors compared the performance of DIA translator with Google translator. The authors found that DIA translator improved the accuracy of translation by 40% more than Google translator.

The work in [23] used TURJUMAN, a neural toolkit for translating twenty languages into modern standard Arabic. TURJUMAN is an open-source package based on Python supported by commands for Arabic neural machine translation. The authors used AraOPUS-20 which was extracted from the open parallel corpus (OPUS). OPUS contains 90 languages one of them is Arabic. They used AraOPUS-20 dataset to develop TURJUMAN.

The work in [24] used Google translate for the translation of Egyptian Arabic in a corpus of 280 song lyrics into English. She used the parallel corpora of standard Arabic and English for training. She adopted Google translate which achieved a BLEU score of only 14.69, although a huge amounts of parallel corpora of standard Arabic and English were used in training. She mentioned that the researchers should create more dialect-specific corpora and add them to the standard Arabic-English parallel corpora to be used in training.

The work in [25] used an encoder–decoder model for Arabizi detection and transliteration of Egyptian SMS/Chat. The encoder has two layers namely BiGRU and GRU. The decoder has two GRU. The author used a transliteration corpus called (ARABIZICORPUS) and BOLT Egyptian Arabic SMS/Chat. The hasher metrics were used to measure the accuracy of word level and BLEU score was used to measure the quality of the transliteration at the end. The author noticed that the faster systems had lower quality and vice versa.

The work in [18] used 16 models of different combinations of LSTM, BiLSTM, GRU, and BiGRU for Arabic machine translation. They compared the translation of Arabic to English using encoder–decoder models based on the architectures of the four models with an attention mechanism. They used OPUS dataset, namely UN and metric BLEU to score the results. From the experiments, the authors found that the encoder of BiGRU architecture, decoder of BiLSTM architecture, and the attention mechanism showed the best results in terms of BLEU score and computational speed.

The work in [16] compared Vanilla RNN, LSTM, BiLSTM, and GRU architectures for Arabic word sense disambiguation. The authors used a lexical resource of Arabic WordNet (AWN). For their comparison, they used batch size, word embedding size, and number of epochs during training. They mentioned that the performance of the GRU model outperforms the other ones.

The work in [17] compared between LSTM, BiLSTM, and GRU models in generating the Arabic poem. They mentioned that GRU was trained faster and better performance than that of LSTM. This investigation is not help only in Arabic generation, but also for many Arabic language tasks including Arabic machine translation.

The work in [26] built models using feed forward neural network (FFNN) and recurrent neural network with modification of embeddings and encoding. They used Teshkeela corpus which consists of 2.3 M words. The authors compared the performance of the adopted models with diacritization and without diacritization by BLEU score. The experiments were made on neural machine translation with diacritization and without. The results showed that the FFNN models were promising with the embeddings. From the results, they proved that the translation was better with Arabic diacritization.

The work in [27] proposed an attention-based encoder–decoder model for machine transliteration between Arabic and English. The authors compared the performance of their model with a model based on statistical machine translation. The encoder–decoder model was GRU which did not have an activation function but had two gates to control the flow of information throughout the network. The encoder was bidirectional of two GRU cells. They used four English-Arabic parallel corpora from lingfil website. The authors used python NLTK toolkit for both Arabic and English normalization and word tokenization. The results showed that the authors' model outperformed the other models.

The work in [28] compared the quality of translation for Arabic Levantine to English using RNN, LSTM, and GRU. The dataset used was 1200 tweets in Syrian dialect. The authors used the source code from GitHub and made some changes for the models. They compared the performance of the quality of the translation for the adopted models using BLEU score. The results showed that GRU had achieved the best performance.

The work in [29] adopted some pre-trained NMT models using different transformer architectures and ChatGPT. The authors developed a parallel corpus for Arabic-English in the financial domain and fine-tuned several models and made them available for researchers. The quality of the adopted models was evaluated by automatic metrics and human evaluation. The results of their experiments were promising and the models performed better with back-translated data than that on synthetic data.

The work in [30] used generative pre-trained transformers (GPT) models to translate 50 natural languages into English text. The authors evaluated 16 open-source GPT models, which are available to users as cloud bases resources. They considered a reasonably large benchmark of the translations of these models and compared them against the Google translate API. The quality of the GPT models is measured by using METEOR, GLEU, and BLEU metrics. The authors used TED Talk transcript dataset for 50 of the foreign languages to be translated into English. The promising performance of the 16 GPT models was: ReMM-V2-L2-13B and L1ama2-chat-AYT-13B for any tuple of language and translation quality metric.

The work in [31] proposed an approach based on transformer architecture pre-trained language model for English and Arabic translation in the public health domain. The authors used OPUS after multi-filtering to get high quality dataset and used it for training. They randomly selected 50,000 segments from the original dataset for testing. The authors considered the BLEU metric as automatic evaluation and human evaluation by an expert in the domain whose native is Arabic. The authors used the baseline MT to start some parallel data from nothing. The authors employed text generation with LLM in the target language. The pre-trained language model was used to have enough data in the target language. The target sentences were translated back to obtain the parallel source sentences. They combined the two cases and evaluated the results. The evaluation involved 5–6 BLEU on Arabic to English translation and 2–3 BLEU on English to Arabic translation. The human translation corroborated the results. The authors' approach improved the quality of the translation of the in-domain test set.

The work in [32] evaluated the performance of ChatGPT, GPT-3.5, and text-davinci-002 for machine translation across 18 language pairs with low and high resource languages. The performance was evaluated automatically and by humans to indicate and provide the weaknesses and strengths of GPT models. The authors used WMT-22 test sets for 16 languages, and for the rest of the languages, they used WMT21 test sets. They combined both GPT models and NMT models to improve the performance. They used Microsoft Translator and then GPT as fallback system to improve the quality of the translation. The authors noticed from the results that the adopted GPT models produced fluent translation for the high resource language but faced some challenges for low-resource language. The hybrid approach improved the quality of the translation and GPT models perform better for some languages than others.

The work in [33] used GPT-4 to automatically post-edit the output of NMT (MS-Translator) across several language pairs. The authors adopted WMT-22 datasets for their experiments. The results were evaluated by both human experts in bilingual languages and automatic evaluation metrics. From the human evaluation, GPT-4 improved the performance of the translation but, there were better performance of languages pairs than others. Using the BLEU metric evaluation the authors found that the quality of the initial translation was lower than zero-shot (without needing any instructions) translation of large language model (LLM). The post-editing translation leaded to lower quality of final translation.

The work in [34] proposed an approach to enhance GPT-4 through in-context learning to increase the accuracy of machine translation without additional fine-tuning. The author aimed to provide translations not only have the same meaning but also contextual rich and linguistically sophisticated. The author used OPUS-100, 10,000 instances for training and the first 100 sentences for testing for each language pair. The BLEU metric was used to evaluate the overall translation quality, while COMET was adopted to know the fluency and the semantic accuracy of the output. Three experiments were applied to assess the GPT-4 translation across Chinese, Japanese, and Vietnamese to English. In the first experiment GPT-4 was processed with no prior examples are provided. In the second experiment, GPT-4 was processed with random examples meaning where the translation was left to chance. The third experiment applied the adopted authors' approach. The term-frequency-inverse document frequency (TF-IDF) matrix was used to calculate the cosine similarity scores between the user prompt and the different sentences in the dataset and fed the top four most relevant examples to GPT-4 as context for translation. From the experimental results, the author's approach was the best in BLEU and COMET scores compared to the other two experiments.

The work in [35] used PADIC corpus which contains six Arabic dialects and has 273 K words and 6400 parallel sentences in their experiments. The authors made experiments to get the relationships between Arabic dialects. The authors made the training with GRU encoder–decoder model and selected 32 sentences randomly. They compared between many-to-one multitask learning, one-to-many multitask learning, SMT, and single-task learning. One-to-many means one shared encoder to all dialects and a decoder for each dialect. Many-to-one means encoder for each dialects and one decoder for all target languages. They used the BLEU score for the comparison and found that the best one was for many-to-one then one-to-many.

The work in [36] translated Amharic language to Arabic language using LSTM model and GRU model. They constructed small parallel corpora of Amharic-Arabic language due to their lack. They developed LSTM and GRU models by using attention-based encoder–decoder architecture and comparing them with Google translation system by BLEU score. From the results it was shown that the LSTM outperforms both the GRU and Google translation system.

Large language models and ChatGPT in Arabic machine translation

Large language models (LLMs) are efficient models for several linguistic tasks. Several LLMs were presented and introduced. Examples of such LLMs include, but not limited to, GPT-3, ChatGPT-4, ChatGPT The work in [37], and others. LLMs are considered expensive to pre-train and deploy. They can adopt research and non-commercial deployment. ChatGPT can surpass Google translate on many translation pairs. ChatGPT can match the performance of fully supervised models for document-level translation.

The work in [38] presented an evaluation study of Bard and ChatGPT on machine translation of ten Arabic varieties. Their study involved diverse Arabic varieties such as classic Arabic, modern standard Arabic, and some country-level dialectal variants such as Algerian and Egyptian Arabic. The authors' work presented an evaluation of LLMs on machine translation from major Arabic varieties into English. The study presented some sort of challenges with dialects for which minimal public datasets exist. The study, on the other hand, showed better performance and/or better translation of dialects than some of those commercial systems. The authors' study assessed performance on Bard on natural language processing tasks in general and on Arabic machine translation in particular. The authors presented a multi-Arabic dataset for machine translation that non-explored to existing LLM.

The work in [39] mentioned that translation software models based on artificial intelligence are available. Examples of such models are (Google Translate, Bing, Microsoft Translator, DeepL, Amazon Translate, and others). Also, artificial intelligence can be applied to develop a lot of applications such as GPT-3, Playground, ChatGPT-4, Chatsonic, and ChatGPT. The authors presented a study to identify the differences between human and artificial intelligence translation in the legal field. The authors' work also aimed to evaluate the quality of artificial intelligence translations for legal documents. The authors chose a collection of legal texts from various contracts. Such texts were allocated to legal translators and subjected to artificial intelligence translation systems. The authors examined the differences between artificial intelligence and human translation. They discussed also the strengths and weaknesses of the two approaches.

The work in [40] mentioned that artificial intelligence has an important role in the development of neural machine translation models. The development of ChatGPT is one of such models. The author presented his work to determine the role of ChatGPT in Arabic-Turkish translation. The source text was written in Arabic language where the generated translated text was written in Turkish language. The author mentioned that the error margin is lower in human translations than that of the artificial intelligence translations. The translations based on artificial intelligence, on the other hand, are faster and less costly compared with that of human translations. The study concluded that artificial intelligence translations are difficult to replace the human translations. Also, neural machine translation cannot present acceptable translation in all application domains. The development of ChatGPT has a positive effect to make the translation work easier. The author concluded that ChatGPT translation outputs varied by the text type. ChatGPT was promising in the translation of plain text of Arabic-Turkish language pairs. ChatGPT presented some sort of errors when adopting technical text translations and literary text translation.

The work in [41] presented an evaluation of large language models on discourse modeling. The authors mentioned that the machine translation tasks presented better performance than before by adopting the pre-trained models such as BERT and GPT. The authors' effort adopted the document-level machine translation with LLMs. The study was focused toward the effects of context-aware prompts. The authors presented a critical comparison between the translation performances of ChatGPT with some commercial machine translation systems. The authors concluded that LLMs showed a promising performance and also presented a significant potential to be a fruitful paradigm for document-level translation. They concluded also that GPT-4 presented a better performance than those adopted commercial machine translation systems in terms of human evaluation. Moreover, the capture and utilization of discourse knowledge using ChatGPT are considered a big challenge.

The work in [42] presented a comparative study of LLMs and Google Translate. The study was focused on evaluating the effectiveness and precision of translation represented by Google Translate and that translation presented by ChatGPT-3.5 and ChatGPT-4. The study was directed to translate the academic abstracts between English and Arabic in bidirectional translation. The abstracts were collected from indexed journals namely: the journal of Arabic literature and Al-Istihlal journal. The author's work also was focused on some important themes such as semantic integrity, syntactic coherence, and technical adequacy. The author concluded that ChatGPT-4 presented better performance than that of ChatGPT-3.5 for Arabic-English translation. The performance of both ChatGPT-3.5 and ChatGPT-4 was approximately the same for English-Arabic translation. ChatGPT, on the other hand, lacks adequacy in providing a good text output for low-sources languages, i.e., ChatGPT has shown some sort of limitations for low-resource languages.

The work in [43] mentioned that adaptive machine translation is an important type of translation that exploits feedback from users to improve the translation quality. Feedback involves corrections to previous translation terminology and style guides. LLMs have promising features of in-context learning. LLMs can learn to replicate certain input–output generation patterns without further fine-tuning. The authors' work was presented to utilize the in-context learning to improve real time adaptive machine translation. LLMs were adopted to a set of in-domain sentence pairs while translating a new sentence. The authors operated their experimental work across five language pairs: English-Arabic, English-French, English–Spanish, English-Chinese, and

English-Kinyarwanda. The authors concluded that the translation quality can surpass that of strong encoder–decoder machine translation approaches.

The work in [37] presented a technical report discussing the capabilities of GPT-4. GPT-4 is a transformer-based model pre-trained to predict the next token in a document. The model can be fine-tuned using reinforcement learning from human feedback. GPT-4 is a large-scale multimodal model capable of accepting images and text inputs, processing them and producing text outputs. GPT-4 is very important for several linguistic tasks, machine translation is one of them. The report discussed the challenge of developing learning infrastructure and optimization methods that behave predictably across a wide range of scales. GPT-4 outperforms the previous LLMs and also presents a promising performance in several natural languages. GPT-4, on the other hand, has a limitation like the earlier GPT models. GPT-4 is not fully reliable as it can suffer from hallucinations. GPT-4 has a limited context window and does not learn from experience.

The work in [44] mentioned that pre-trained language models (PLMs) are neural network models trained on large amount of textual data in unsupervised manner. The pretraining process enables the models to learn linguistic knowledge from the data. This includes, but not limited to, syntax, semantics, and relationships between words. PLMs play a vital role in several linguistic tasks, machine translation is one of them. PLMs include, but not limited to, BERT, GPT-2, and RoBERTa. The authors' work investigated the error patterns of the state-of-the-art PLMs when translating from English text to Arabic. The error patterns may be in different types such as lexical, morphological, syntactic, semantic, orthographic, reordering and others. The authors compared the performance of Google Translate with some PLMs approaches such as GPT-3.5, GPT-4, Facebook, Marefa, and Helsinki. The authors evaluated the adopted models in a parallel corpus of English-Arabic sentences. Also, they identified the common error patterns and explored their reasons. The authors concluded that the PLMs presented some improvements for English-Arabic translation capabilities. PLMs, on the other hand, have some limitations in handling some Arabic grammar and vocabularies.

Some resources and NMT toolkits for arabic machine translation

Machine translation approaches need datasets and bilingual parallel corpora. These parallel corpora are important for training models. It is observed that most of the Arabic machine translation (AMT) researchers are using the following datasets and corpora which are freely available.

- UN Corpus: This corpus is collected from the official documents of the United Nations (UN). The UN corpus consists of six official languages, one of them is the Arabic language. The UN corpus has around 300 million words for each language [6].
- Arab-Acquis: It is a large dataset used to evaluate machine translation between 22 languages and Arabic. It has over 600,000 words and over 12,000 sentences [6].
- Open Subtitles (OPUS): It is a multilingual parallel corpus of ninety languages. It is a huge data collected from different domains including Arabic-English sentences.
- Tashkeela: Tashkeela is a dataset which has around 75.6 million vocalized Arabic words [6, 26].

Table 1 Comparison of some approaches for Arabic machine translation

References	Languages	Approach Description	Adopted Dataset	Metric	Remark
[Alkhawaja, L., et. al.,2020][51]	English- Arabic	Evaluation of Google Translate using neural machine translation	100 passages	Error analysis and BLEU	Improve translation quality using Google Translate
[Almahasees, Z., 2018][52]	Arabic- English	Google and Microsoft Bing translation of journalistic texts	Examples from Jordan Petra news agency	Error Analysis	Improve translation quality using Google Translate
[Oudah, M., 2019][53]	English- Arabic	Impact of preprocessing on statistical and neural machine translation	LDC2010T12, LDC2010T14 (in domain) and LDC2014T02 (out of domain)/ news	BLEU	Improve translation quality ≈ 56.18% and 37.96% for in domain and out of domain testing
[Ehab, R., et.al.,2019] [54]	English- Arabic	Hybrid machine translation using EBMT& Translation memory	Medical sentences, medical sentences of internal medicine	BLEU	Improve translation quality using EBMT and translation memory
[Ataman, D., et. al.,2019] [55]	English- Arabic	A Latent Morphology Model for neural machine translation	TED talks/ General	BLEU	Improve translation quality by about 51%
[Almansor, E. et. al.,2018] [56]	English < > Arabic	Neural machine translation for translating low-resource languages	90 K sentences from TED, IWSTL 2016/ General	BLEU	Improve translation quality by > 15%
[Farhan, W., et.al.,2019][57]	Arabic- English	Unsupervised dialectal neural machine translation	Multi-parallel corpus of 309 K sentences	BLEU	Improve translation quality by about 32.14%
[Berrichi, S. and Mazroui, A. 2o21] [58]	Arabic- English	Limted vocabulary and long sentences constrained in neural machine translation	Sentence pairs from UN corpus and from Arabic- Acquis/ government	BLEU	Improve translation quality by about 2.81% over the baseline
[Berrichi, S. and Mazroui, A. 2019] [59]	English- Arabic	Word alignment with prior knowledge for machine translation	Sentences from UN corpus and from MulTedCorpus/ law, business	Alignment error rate, BLEU	Improve translation quality by about 5%
[Mahesh, V. and Milam, A., 2020][60]	Arabic- English	Google Translate, System Prompt, World Lingo, and Bing Translate	6 sentences	BLEU	Improve translation quality using Google Translate

BLEU = Bilingual evaluation under study & EBMT = Example-based machine translation

Table 2 Examples of some NMT systems with respective datasets and toolkits

References	Toolkits/language	Model	Language pairs	Datasets
[Laith, H., et. al., 2018] [4]	Tokenizer by Python Training by Python and Keras	Multitask, Single NMT	AD/MSA/E	Concatenated of PADIC and MPCA
[Hadeel, S. and Constantin, O., 2022] [61]	Semi-supervised NMT	Semi-Supervised NMT	DA/MSA/E	OPUS
[Wael, A. and Youns, B., 2018] [62]	OpenNMT	Baseline Transformer	English- Arabic	Arab-Acquis
[Sara, E., et. al., 2017] [63]	GIZA++	PBSMT	E/A	UN
[Donia, G., et. al., 2023] [15]	Keras Transformer Keras Embeddig Layer	Transformer with multi head attention	E/A	Set of Tatoeba Project sentence classified by frequency words
[Diadeen, A., et. al., 2022] [22]	NMT	RNN encoder–attention decoder	E/A	Sulfur industry Domain 1200 sentences
[El Moatez, B., et. al.,2022] [23]	TURJUMAN	Neural AMT	20 language /MSA	Ara OPUS-20 by removing noise from OPUS
[Ali, S., et. al., 2020] [25]		Seq2seq	Transilitration/A	BOLT Egyptian Arabic SMS/Chat and Transilitration (ARABIZICORPUS)
[Nouhaila, B., et. al., 2020] [18]	Farasa as Tokenization	Encoder BiGRU, Decoder BiLSTM and attention mechanism	A/E	OPUS namely UN
[Ali, F., et. al., 2019] [26]	FastText	FFNN and RNN	Arabic- English	Tashkeela corpus
[Mohamed, S. et. al., 2017] [27]	Python NLTK for A character and word tokenization NLTK for E word tokenization OpenNMT	Attention-based encoder–decoder	Transilitration of A E	Python NLTK for A character and word tokenization NLTK for E word tokenization OpenNMT
[Sawsan, A. and Yasmin, A.,2018] [28]	NMT	GRU	DA/E	Build 1200 tweets Syrian dialect
[Youness, M., et. Al., 2021] [35]	PyTorch	Encoger-Decoder GRU M-2-O O-2-M SMT	MSA/DA DA/MSA	PADIC
[Ibrahim G., Shashirekha H., 2019] [36]	OpenNMT	LSTM, GRU	Amharic/A	Constructed from Tanzile corpora
[Weim et. al, 2023] [49]	Python	AraT5 transformer model	DA\MSA	MADARcorpus
[Rania, 2024] [50]	Python	NMT	Egyption Arabic\ English	Selection ofArabic sentences

MT: Phrase-based statistical machine translation, POS: Part of speech tagging, FFNN: Feed-forward neural network, RNN: Recurrent neural network, WER: Word error, CER: Character error, PADIC: Parallel Arabic dialect corpus, M-2-O: many-to-one, O-2-M: One-to-many, S-task: single-task, SMT: Statistical machine translation, MSA: Modern standard Arabic, DA: Dialect Arabic

Table 3 Example of some NMT for Arabic language and their results

References	Model	Evaluation metrics	Batch Size	Epochs
[Laith, H., et. al., 2018] [4]	Multitask without POS, Multitask with POS, and Single NMT	BLEU Score	140, 50, and 140	49, 50, and 120
[Hadeel, S. and Constantin, O., 2022] [61]	Semi-Supervised NMT	BLEU Score	32	100
[Wael, A. and Youns, B., 2018] [62]	Baseline Transformer	BLEU Score	32	75
[Donia, G., et. al., 2023] [15]	Encoder–Decoder Transformer	Accuracy BLEU Score		10
[Diadeen, A., et. al., 2022] [22]	DIA trans. and Google translate	Precision		
[El Moatez, B., et. al.,2022] [23]	TURJUMAN Toolkit	BLEU Score	32	25
[Ali, S., et. al., 2020] [25]	Seq2seqGRU	BLEU Score Accuracy	1024 for LINE2LINE 2048 for WORD2WORD	40
[Nouhaila, B., et. al., 2020] [18]	Encoder BiGRU, Decoder BiLSTM and attention mechanism	BLEU Score	64	20
[Ali, F., et. al., 2019] [26]	FFNN RNN	BLEU Score WER	512 256	300 50
[Mohamed, S. et. al., 2017] [27]	Bi-Att-seq2seq	WER and CER	128	8
[Sawsan, A. and Yasmin, A.,2018] [28]	GRU	BLEU Score	32	
[Youness, M., et. Al., 2021] [35]	M 2 O, O 2 M, Single task, SMT	BLEU Score	32	200
[Ibrahim G., Shashirekha H., 2019] [36]	LSTM, GRU, Google translate	BLEU Score, Accuracy	80	10

- Tatoeba Project: The project has 11.433 number of samples and the maximum length of the target language (Arabic) is 201 while, of the source language (English) is 209 [15].
- PADIC: It is a parallel corpus which contains Arabic dialects. It contains six dialects from Maghreb dialects and Levantines dialects, and MSA. Maghreb dialects such as Moroccan dialect, Tunisia dialect, and Algerian dialects. Levantines dialects such as Syrian, Palestinian, and Lebanese [35]. This dataset is designed to tackle the significant challenge of translating Arabic dialects, which often exhibit substantial linguistic differences from MSA.

Moreover, there are a number of toolkits which are very friendly user licenses. Examples of such toolkits include, but not limited to, the following:

- NMT: It is a user-friendly toolkit which enables the user to add customized models. It has a tutorial for building a competitive NMT model from scratch [9].
- NMT Keras: It focuses on the advancement of sophisticated applications such as interactive NMT and online learning [9].
- FairSequ: It is a sequence modeling toolkit which is based on PyTorch [9].
- FastText: It helps researchers/users to represent and classify the text [6].

- OpenNMT: It is developed by the collaboration of Harvard University and SYSTRAN. It is easy to use and it supports both CPU and GPU [6]. It provides for training and deploying translation models.
- TURJUMAN: It is a neural toolkit and is used to translate 20 languages into Modern Standard Arabic (MSA) [23]. It offers different decoding methods and provides a platform for researchers to build and deploy NMT systems for Arabic.

Metrics for evaluating machine translation systems

To evaluate the quality of the machine translation process, some measurable criteria can be adopted and applied. The most common metrics which are used to evaluate the performance of machine translation systems are: BLEU score, word error rate (WER), and precision.

BLEU Score: The BLEU metric counts the matching words from the output and the reference. It balances between ignoring and requiring word order. It computes the coefficient of brevity penalty with the precision for n-gram of size 1-to-4 and is defined as the following equation [24]:

$$\mathrm{BLEU} = \mathrm{brevity} - \mathrm{penalty} \times \exp \sum_{i=1}^{4} \log \frac{\mathrm{matching}\, i - \mathrm{grams}}{\mathrm{total}\, i - \mathrm{grams\ in}\, MT} \tag{14}$$

$$\mathrm{brevity} - \mathrm{penalty} = \min(1, \frac{\mathrm{output} - \mathrm{length}}{\mathrm{refrence} - \mathrm{length}}) \tag{15}$$

where brevity penalty is based on the ratio between the number of words in the machine translation and reference translation.

Word Error Rate (WER): WER is a metric used to evaluate machine translation. This metric represents the minimum number of editing steps to transform output to reference [45] and it represented by Eq. (16). WER is very low when the word order of the translation is wrong according to the reference [45].

$$\mathrm{WER} = \frac{\mathrm{substitution} + \mathrm{insertion} + \mathrm{deletion}}{\mathrm{reference}_{\mathrm{length}}} \tag{16}$$

where substitution, insertion, and deletion are the replacing word, the adding word, and the dropping word, respectively.

Precision is defined by Eq. (17)

$$\mathrm{Precision} = \frac{TP}{PP} \tag{17}$$

where TP and PP are the true positive and prediction positive, respectively.

Comparison of some machine and deep learning approaches to translate to and from Arabic text

Table 1 outlines the main attributes and features of some deep learning research efforts for Arabic machine translation. Such efforts were published in the literatures. Table 2 presents some NMT systems along with their respective datasets and toolkits. Table 3, on the other hand, specifically focuses on NMT systems developed for the Arabic language and their performance results.

Observations from the Related Work and Tables

- The most important resources for NMT are bilingual parallel corpora such as OPUS. Opus-100 is used for multilingual machine translation.
- The majority of researchers prefer to use automatic evaluation to measure the performance of their models. Automated metrics compare the output of machine translation with the reference translations and they are objective and cheap.
- It is clear that a larger of batch size increases the accuracy and BLEU score of the adopted models.
- Some researchers are interested in domain-specific translation such as, sulfur industry and Arabic sign language. Because of the lack of parallel corpora for their domains, they construct their own dataset.
- Some researchers have shown interest in looking into translating the trend of mixing Arabic with English words (transliteration) on social media.
- Some researchers compare their models with Google translate and prove that their models with RNN architectures outperform it.
- Researchers use different batch sizes and number of epochs according to their models, datasets, and experiments.
- The majority of researchers use OpenNMT toolkit because it is the most user-friendly toolkit.
- We observed that many researchers prefer to use GRU model because it is faster to train and run than LSTM. GRU has fewer parameters than the LSTM.
- In some tasks, the LSTM outperforms GRU.

Some commercial Arabic machine translation systems

Due to the dramatic advances in information technology, the communication between different and cross-cultural societies has been increased. Translation of text from one language to another is a vital task. Machine translation plays an important role in breaking done the language barriers among different people. This is important for business and global cooperation among different societies. The effective collaboration between linguists and computer specialists has succeeded to develop some translation systems. In this regard, several commercial translation systems are currently used. Examples of such translation systems are briefly mentioned as follows:

Sakhr: Sakhr is a machine translation system produced in Kuwait in 1982 by Mohamed AlSharekh and his group in Sakhr Al-Alamia. Sakhr is a bidirectional translation system as it can translate between Arabic and English. Sakhr system is used for online translation, mobile applications and can be used for making professional translation services. The design methodology for Sakhr translator is based on both rule and statistical approaches.

Systran: Systran is a translation system designed by Peter Tone and his team in USA in 1968. Systran was designed to translate among 55 natural languages including the Arabic language. Systran was developed using the rule-based translation approach. Systran can be used to translate text, documents and web pages. Nowadays, Systran is considered a reliable translation system as it achieves high accuracy and reasonable quality of translation.

ChatGPT: ChatGPTcan be used as a translator for multiple language pairings including both Arabic and English Languages. ChatGPT, such as ChatGPT-3.5 and ChatGPT-4, adopt the transformer-based architecture that enables it to learn the syntactic structures of different natural languages. ChatGPT can learn from large corpora including articles, books, and websites and generate output. ChatGPT can handle complex language pairs that are different in their grammar and syntax. ChatGPT has the capability to translate sentences from English to Arabic [44].

Microsoft Translator: Microsoft translator is a multilingual machine translation. Microsoft translator can translate both text and speech through cloud services for business. The service supports text translation between several languages. It is used to translate the Microsoft knowledge base into French, Spanish, Gzerman, and Japanese. Microsoft translator is based on neural machine translation, statistical machine translation and deep neural networks. The outputs of Microsoft translator are evaluated using Bilingual Evaluation Understudy (BLEU) score [https://en.wikipedia.org/wiki/Microsoft_translator].

Amazon Translate: Amazon Translate is considered a text translation service. It is used to translate unstructured text documents and can build applications that work in multiple languages. Amazon translate can translate emails, customer service chat, technical reports, knowledge base articles, posts, and document repositories. Amazon translate can translate text stored in several databases such as Amazon Aurora, Amazon DynamoDB, and Amazon Redshift. [https://docs.aws.amazon.com/translate/latest/dg/what-is-html].

Marefa Translator: Marefa is a model designed to translate English to Arabic. Marefa translator is the first translation model that was issued from Marefa pedia. That translator is using additional Arabic characters to characterize acoustics in English language [https://hugginface.co/marefa-nlp/marefa-mt-en-ar].

Google Translate: Google Translate is a multilingual translator developed by Google to translate text and documents from a natural language to another. Google Translate is considered a web-based free to use translation service. It can translate among several natural languages including the Arabic language. Google Translate was released as a statistical machine translation. Currently Google Translate is using advanced methods such as neural machine translation and deep learning techniques. The functions of Google Translate include several forms of text and media such as written word translation, document translations, mobile applications, and others. Moreover, translation using Google Translate is now available with a cell phone in an offline mode [https://en.wikipedia.org/wiki/Google_Translate, 2024].

Discussion

Machine translation is one of the important areas of computational linguistics. As mentioned in the previous sections there are different approaches of machine translation. Rule-based machine translation (RBMT) systems were the first developed commercial systems. Such systems are feasible for resource-poor languages with little data. RBMT is mainly based on linguistic knowledge which is encoded by experts. The RBMT approach needs the expertise of linguists to write the language rules. The RBMT approach uses rules which dictate the syntactic knowledge. That approach also depends on linguistic resources which deal with information related to morphology, syntax, and semantics of the adopted language. In the RBMT, the language analysis can be done on the syntactic level as well as the semantic level. RBMT systems are relatively simpler to carry out error analysis [5].

Statistical-based machine translation (SMT) is another approach of machine translation. SMT systems have a set of translations to translate the source sentences to their corresponding target sentences, respectively. Decoding complexity and target language reordering are two important themes with SMT. The SMT approach requires a large amount of data which can learn from it and needs less human efforts to create linguistic rules. Approach-based SMT can learn structures such as word alignments or phrases directly from parallel corpora [5, 9].

Another approach of machine translation is that one called neural machine translation (NMT). NMT approach is a promising direction as it achieved the state-of-the-art performance in many translation tasks. NMT is considered as a sequence-to-sequence approach based on encoder–decoder architectures. There are different methods to build efficient encoders and decoders. Examples of such methods are those based on RNN (such as LSTM, BiLSTM, and GRU). In this concern, each output produced by the encoder and decoder can encode information in the input sequence. RNNs are the most efficient family of neural networks. RNNs are a family of sequential models that apply the same state transition function to sequence. [9]. Moreover, NMT based on LSTM model encoder–decoder can translate an Arabic text into English. LSTM can facilitate a given sentence as input rather than just one word as input for prediction. This is relevant and efficient in NMT. [7]. In LSTM, the input gate, output gate, and forget gate can respectively control the model input, model output, and the degree of memory module forgetting at the last moment. The same thing can take place in BiLSTM expect that the BiLSTM can learn in both the forward and reverse directions. The gated recurrent unit (GRU) can also represent the sequential data by considering the previous data. The GRU can use the amount of information seen by the hidden states at the previous steps. Adopting GRU based approaches can facilitate generation of effective sentences using temporal features. This is because it is easy to capture the long-term dependencies between the words of a source sentence [46]. NMT-based approaches are preferred approaches compared with those models based on SMT.

Generative artificial intelligence (GAI) is a promising paradigm that produces novel text content. Large language models (LLMs) can also generate novel text from textual prompts. LLM is an example of GAI. LLMs are deep learning models designed to analyze, process, and generate human languages. Such models can exploit the power of neural networks and massive amount of textual data to learn the complexities of human language like Arabic. LLMs are based on transformer architectures which can handle effectively sequential data. LLMs can be used in a lot of applications, text translation is one of them. LLM architecture has an input layer to encode the input text in a form easy to be understood by the model. The input text is that one written in a source language in case of text translation. LLM architecture has transformer encoder layers which allow the translation model to capture contextual information from the source text. The LLM architecture has an output layer to generate the relevant text in the target language. The generated text is based on the input text and the information learned by the encoder layers [47]. LLMs such as ChatGPT-3.5 and GPT-4 can produce translation exhibit accuracy and fluency to the source input text. Adopting LLMs like GPT-3.5 and GPT-4 textual machine translation can present efficient and good quality translation. LLMs can present better translation quality compared with those traditional translation models [48]. Although ChatGPT has a positive potential to improve text generation process in machine translation, it will not replace completely the human translators.

Moreover and regardless the adopted approach of machine translation, the translation systems based on machine and deep learning can be trained using datasets. The national institute of standard (NIST) offers large collection datasets which are used for training and testing. NIST offers several language pairs datasets for Arabic to English translation, Chinese to English translation, Farsi to English translation, Korean to English translation, Arabic-French newspaper parallel test set, and others [12]. Another source that offers large collections of corpora is the linguistic data consortium (LDC). Examples of the datasets offered by the LDC include, but not limited to, Arabic broadcast new parallel text (LDC2007T24 and LDC2015T07), Arabic newsgroup parallel text (LDC2009T03 and LDC2009T09), Arabic newwise English translation collection (LDC2009T22), Arabic-French parallel text (LDC2018T13, LDC2018T21), and others. Finally, the performance of machine translation systems can be evaluated using some measurable criteria. Examples of such criteria include, but not limited to, accuracy, BLEU score, word error rate, and others.

Conclusion

Textual machine translation can translate a given text in a source language to another text in a target language. Machine translation aims to eliminate the need for people translators using the available models, algorithms, software tools, and linguistic resources. Automatic translation, in general, and Arabic machine translation, in particular, can help people and can partially replace some human translation works. Statistical-based approaches presented better performance than those based on classical-based approaches. Neural machine translation, on the other hand, showed better performance than those based on traditional and statistical-based approaches. RNN-based neural machine translation considered the dominant position in machine translation development and achieved a promising performance. RNN deep learning approaches such as LSTM, BiLSTM, and GRU showed good performance for modeling the sequence data. The encoder–decoder structure of neural machine translation is considered as a mapping theme with the source and target sentences, respectively. Neural machine translation is a promising paradigm for Arabic machine translation. Utilizing the advantages of neural machine translation with LLMs is the promising trend. Merging LLMs with machine translation can improve the translation accuracy due to the improvement of context and languages understanding LLMs presented a promising behavior in developing machine translation and generating language text. LLMs can play an important role to improve the Arabic translation quality and speed. LLMs such as BERT, GPT-3.5 and GPT-4 can exploit using the powerful deep learning approaches and massive data to process and generate human-like language. LLMs can analyze huge amount of textual data and involve deeper context understanding. LLMs are expected to deal with several linguistic complexities which will lead to a better accurate translation. LLMs can handle and quickly translate huge amount of textual data which is suitable for real time applications. Introducing LLMs such as ChatGPT have showed improvement in the translation quality. Using LLMs in translation can adjust sentence structure, rephrase for better flow and generate creative text in the target language. Although neural machine translation showed a better trend compared with those classical approaches, machine translation still needs more efforts. Till now Arabic machine translation is not the same like that translation done by the human level.

Future work

We summarize some trends as future work directions. The directions, include but not limited to, the following:

- Improving semantic representation and contextual understanding of text especially Arabic.
- Evaluating the performance of translation tasks for handling different Arabic dialects.
- Enriching how to apply generative artificial intelligence to generate text (e.g., generation of letters, reports, emails, etc.).
- Enhancing the translation quality using LLMs.
- Developing accurate models and/or systems to translate to and from Arabic text.
- Exploiting advanced alignment methods to present high quality translation of natural languages.
- Investigating how neural machine translation can improve accuracy of translation to and from Arabic.

Abbreviations

RBMT	Rule-based machine translation
SMT	Statistical-based machine translation
NMT	Neural-based machine translation
CNNs	Convolutional neural networks
ANN	Artificial neural network
RNNs	Recurrent neural networks
LSTM	Long short-term-memory
BiLSTM	Bidirectional LSTM
GRU	Gated recurrent unit
CNN	Convolutional neural network
WER	Word error rate
GAI	Generative artificial intelligence

Author contributions

Conceptualization was done by M. N.; methodology was done by S. E. and M. N.; investigation was done by S. E and M. N.; writing—original draft preparation was done by S. E. and M. N.; writing—review and editing was done by S. E and M. N.; supervision was done by M. N. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability

Not applicable.

Declarations

Competing interests

The author of correspondence declares no conflicts of interest.

Received: 1 February 2025 Accepted: 14 May 2025
Published online: 23 June 2025

References

1. Shuoheng Y, Yuxin W, Xiaowen C (2020) A survey of deep learning techniques for neural machine translation., The arXiv: 2002.07526v1:1–21.
2. Guangx-in, W. Application of pre-training and fine-tuning AI models to machine translation: a case study of multilingual text classification in Baidu. A Technical Report Presented to the Universidad De Lisboa-Faculdade De Letras,
3. Ruimeng S (2022) Analysis on the recent trends in machine translation. AMMSAC Highlights Sci Eng Technols 16:40–47
4. Laith H, Seyoung P, Seong P (2018) A neural machine translation model for Arabic dialects that utilizes multitask learning (MTL). Comput Intell Neurosci. https://doi.org/10.1155/2018/7534712
5. Bharathi R, Priya R, Mihael A, John P (2021) A survey of orthographic information in machine translation. SN Computer Sci 2(330):1–19. https://doi.org/10.1007/s42979-021-00723-4
6. Jezia Z, Moutaz S, Somaya AJ (2021) Arabic machine translation: a survey with challenges and future directions. IEEE Access 9:161445–161468
7. Khen D, Agung B, Aji P, Arif N, Anik N, Leonel H (2022) Neural machine translation of Spanish-English food recipes using LSTM. Int J Inform Visual 6(2):290–297
8. Abdullah, A (2018) A recipe for Arabic-English neural machine translation.,The arXiv: 1808.0611gv1:1–5.
9. Zhixing T, Shuo W, Zonghan Y, Gang Ch, Xuancheng H, Maosong S, Yang L (2021) Neural machine translation. A review of methods, resources, and tools, KeAi, AI Open: pp 1–21
10. Evgeny M (2019) The challenges of using neural machine translation for literature. In: Proceeding of Dublin international conference on qualities of literary machine Translation. Dublin, pp 19–23.
11. Minghan W, Jinming Z, Thuy-Trang V, Fatemeh S, Ehsan S, Gholamreza H (2023) Simultaneous machine translation with large language models., The arXiv: 2309.06706v1:1–5.
12. Mohamed S, Farid M, Ahmed G (2020) Arabic machine translation: a survey of the latest trends and challenges. Computer Sci Rev 38(100305):1–22
13. Ankush G, Mayank A (2018) Machine translation: a literature review. arXiv: 1901.01122v1, pp 1–17, Retrieved on December 2018, from https://arxiv.org/pdf/1901.01122,2018.
14. Juan A, Mikel L, Felipe S (2022) Machine translation for everyone: Empowering users in the age of artificial intelligence. Berlin Language Science Press: pp 141–164.
15. Donia G, Marco A, Salud M, Mostafa A (2023) A case study of improving English-Arabic translation using the transformer model. Int J Intell Comput Inform Sci 23(2):105–115
16. Rakia S, Fethi J, Mohamed A (2022) Comparative analysis of recurrent network architectures for Arabic word Sense disambiguation. In Proceedings of the 18th international conference on web information systems and technologies valleta, Malta. pp 272–277.
17. Abdelhamid A, Ikram E (2022) Comparison and generation of a poem in Arabic language using the LSTM, BiLSTM and GRU. J Manag Inform Decision Sci 25:1–8
18. Nouhaila B, Habib A, Abdelhamid I (2020) LSTM vs. GRU for Arabic machine translation. In Proceedings of 12th International conference on soft computing and pattern recognition, Hyderabad, India. pp 156–165.
19. Yasir A, Akbar K, Mohamed B, Abdul Hakim H, Mousab A, Muawia A (2024) The impact of artificial intelligence on language translation: a review. IEEE Access 12:25553–25579
20. Yang Z, Chen W, Wang F, Xu B (2019) Effectively training neural machine translation models with monolingual data. Neurocomputing 333:240–247
21. Cheng Y, Tu Z, Meng F, Zhai J, Liu Y (2018) Towards robust neural machine translation. arXiv:1805.06130.
22. Diadeen A, Tahseen A, Alaa K, Harith A, Ghanim T (2022) DIA-English-Arabic neural machine translation domain: Sulfur industry. Indones J Electr Eng Computer Sci 27(3):1619–1624
23. El Moatez B, Abdellhalim E, Muhammad A (2022) Turjuman: a public toolkit for neural machine translation. Proceed OSACT 20:1–11
24. Abou Mohamed E, Guerrout O (2020) Preprocessing for Arabic neural machine translation. Algeria. (University Ahmed Draia Master dissertation).

25. Ali S, Aiza U, Nizar H (2020) A unified model for arabizidetection and transliteration using sequence to sequence models. In Proceeding of 5th Arabic natural language processing workshop. Barcelona, Spain. Pp 167–177.
26. Ali F, Ibrahim T, Mohamed A (2019) Neural Arabic text diacritization state of the art results and a novel approach for machine translation. In Proceedings of the 6th Workshop of Asian Translation. Hong Kong, China. pp 215–225.
27. Mohamed S, Farid M, Ahmed G (2017) Arabic machine transliteration using an attention-based encoder-decoder model. In Proceedings of 3rd international conference on arabic computational linguistics. Dubai, United Arab Emirates: Elsevir; pp 287–297.
28. Sawsan A, Yasmin A (2018) A proposal sequence to sequence modeling for Arabic dialect machine translation dialect Arabic to English translator. Multi-knowl Electron Compr Educ Sci Publ 8:335–344
29. Emad A, Jezia Z, Fares A (2023) Domain adaptation for Arabic machine translation: the case of financial texts. PP.1–13, December 25. available from https://doi.org/10.48550/arXiv.2309.12863.
30. Elijah P, Vincent U, Lorie M (2024) Automaticted multi-language to English machine translation using generative pre-trained transformers. arXiv:2404.14680V1.
31. Yasmine M, Andy W, Rjwanul H, John D (2022) Domain-specific text generation for machine translation. Proceedings of 5th Biennial Conference of the Association for Machine Translation; 2022; Americas, Orlando USA, pp 14–30.
32. Amr H, Mohamed A, Amr A, Vikas R, Mohamed G, Hitokazu M, Young J, Mohamed A, Hany H (2023) How good are GPT models at machine translation: A comprehensive evaluation. arXiv:2302.0921Ov1.
33. Vikas R, Amr S, Yiren W, Hany H, Arul M (2023) Leveraging GPT-4 for automatic translation Post-Editing. Find Assoc Comput Lingus: EMNLP 2023:12009–12024
34. Yufeng C (2023) Enhancing machine translation through advanced strategy for GPT-4 improvement. arXiv:2311.10765.
35. Youness M, Nada S, Mounir G, Kamel S (2021) Improving machine translation of Arabic dialects through multi-task learning. Adv Artif Intell AIXIA. https://doi.org/10.1007/978-031-08421-8-40.pp.580-590
36. Ibrahim G, Shashirekha H (2019) Amharic-Arabic neural machine translation. arXiv: Computation and Language.
37. Open AI, GPT-4 technical report. A technical report presented by Open AI, arXiv:submit/4812508 [cs.CL], PP. 1–100, 27 March, 2023.
38. Karima K, Samar M, Abdul Waheed, Md Tawkat I, Ahmed O, El-Moatez B, Muhammad A (**2023**) TARJAMAT: evaluation of Bard and ChatGPT on machine translation of ten Arabic varieties., arXiv: 2308.03051v2 [cs.CL], pp.1–24.
39. Ahmed M, Yousef S (2024) Artificial intelligence and human translation: a contrastive study based on legal texts. Heliyon 10:1–14
40. Sezer Y (2023) The efficiency of neural machine translation models in Arabic-Turkish translation types: example of ChatGPT. J Oriental Stud Istanbul Univ Press 43:339–355
41. Longyue W, Chenyang L, Tianbo TI, Zhirui Z, Dian Y, Shuming S, Zhaopeng T (2024) Document level machine translation with large language models. Retrieved in 2024 from https://aclanthology.org/2023.emnlp-main.1036.pdf
42. Mohammed A (2024) Artificial intelligence in academic translation: a comparative study of large language models and google translate. The Psycholinguistics 35(2):134–156
43. Yasmin M, Rejwanul H, John D, Andy W (2024) Adaptive machine translation with large language models. Retrieved in 2024 available from https://arxiv.org/pdf/2301.13294
44. Hend A, Khaloud A, Hesham H (2024) Error analysis of pretrained language models (PLMs) in English-to-Arabic machine translation. Human-Centric Intell Syst. https://doi.org/10.1007/s44230-024-00061-7
45. Aaron L, Derek F, Lidia S (2016) Machine translation evaluation: A survey. arXiv:1605.0451V6[CS.CL], pp 1–16, 19 June, **2016**.
46. Nouhaila B, Ayad H, Adib A, Ibn ElFarouk A (2018) CRAN: A hybrid CNN-RNN attention-based model for Arabic machine translation. LNCS11157, Conceptual Modeling. In Proceeding of Er2018, Xian, China, the 37th International Conference Xian, China. Pp 571–584.
47. Godwin G (2024) Comprehensive overview of large language models (LLMs): grasping their E-ssence, 2023,Retrieved on September 28 from https://osf.10/preprints/osf/xbtzw
48. Wangjian W, Gang H (2023) Exploring prompt engineering with GPT language models for document-level machine translation: insights and findings. In Proceedings of the 8th Conference on Machine Translation (WMT). Association for Computational Linguistics. pp 166–169.
49. Wiem D, Sameh K, Rahma B (2023) A NLP dialects: a comparative study of transformer models. In Proceedings of 1st Arabic natural language processing conference (Arabic NLP2023). Singapore; pp 683–689.
50. Rania A (2024) The negative transfer effect on the neural machine translation of Egyptian Arabic adjuncts into English: the case of Google Translate. Int J Arabic-English Stud (IJAES) 24(1):95–118
51. Alkhawaja L, Ibrahim H, Ghnaim F, Aawwad S (2020) Neural machine translation: few-grained evaluation of Google Translate output for English-to-Arabic translation. Int J English Linguist. https://doi.org/10.5539/ijel.v10n4
52. Almahasees Z (2018) Assessment of Google and microsoft bing translation of journalistic texts. Int J Lang: Literat Linguit 4(3):231–235
53. Oudah M, Almahairi A, Habash N (2019) The impact of preprocessing on Arabic-English statistical and neural machine translation. In Proceedings of Machine Translation Summit XVII (MT Summit) Vol. 1. Dublin, Ireland.,pp. 214–221.
54. Ehab R, Amer E, Gadallah M (2019) English-Arabic hybrid machine translation system using EBMT and translation memory. Int J Adv Comput Sci Appl 10(1):195–203. https://doi.org/10.14569/IJACSA.2019.0100126
55. Ataman D, Aziz W, Birch A (2019) A latent morphology model for open vocabulary neural machine translation, arXiv:1910.13890.
56. Almansor E, Al-Ani A (2018) A hybrid neural machine translation technique for translating low resource languages. Proceedings of International Conference Machine Learning. Data Mining Pattern Recognition Cham. Switzerland, Springer, pp 347–356.
57. Farhan W, Talafha B, Abuammar A, Jaikat R, Al-Ayyoub M, Tarakji A, Toma A (2019) Unsupervised dialectal neural machine translation. Inform Process Management 57(3):102181

58. Berrichi S, Mazroui A (2021) Addressing limited vocabulary and long sentences constraints in English-Arabic neural machine translation. Arabian Journal of Science Engineering. Available from https://o.doi.org.mylibrary.qu.edu.qahttps://doi.org/10.1007/s13369-020-05328-2, Vol.46, no.9, pp.8245–8259.
59. Berrichi S, Mazroui A (2019) Guiding word alignment with prior knowledge to improve English-Arabic machine translation. In Proceedings of 4th International conference for big data internet things, New York, NY, USA: Association for Computing Machinery, pp1-5 https://doi.org/10.1145/3372938.3372957
60. Mahesh V, Milam A (2020) A comparison of free online machine language translators. J Management Sci Bus Intell 5(1):26–31
61. Hadeel S, Constantin O (2022) A semi-supervised approach for a better translation of sentiment in dialectical Arabic UGT. Workshop proceeding of the seventh arabic natural language processing (WANLP) Abu Dhabi, United Arab Emirates, pp 214–224.
62. Wael A, Youns B (2018) Improving English to Arabic machine translation. In Proceedings of Montreal, 32rd conference on neural information processing systems (NIPS), Montreal, Canada, pp 1–17.
63. Sara E, Samhaa R, Doaa H, Mostafa G (2017) Toward building a comprehensive phrase-based English-Arabic statistical machine translation system. Egypt J Lang Eng 4(2):10–26

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.