A comprehensive overview of the application of artificial intelligence in language learning

Ziru Zhou

School of Cyberspace Security, Hangzhou Dianzi University, Hangzhou, China

21041305@hdu.edu.cn

Abstract. With the rapid advancement of artificial intelligence, AI technology has made significant progress in the field of language. Machine translation has become the dominant method, replacing manual translation due to its convenience and speed. This article will discuss three different aspects: translation, information retrieval, and language artificial intelligence. In the translation section, three distinct translation models will be analyzed, using Google Translate as a foundation. These models have transformed the translation industry and improved accuracy and efficiency. In the information retrieval section, the differences between semantic search involving AI and traditional keyword-based search techniques will be explored. Semantic search, driven by AI, provides more accurate and relevant search results by understanding the context and intent behind user queries. The impact of these advancements on search engine optimization (SEO) practices will also be discussed. Furthermore, the article will delve into the types of speech recognition and classify speech recognition technologies. Finally, the article will summarize the entire content and provide an outlook on future developments.

Keywords: Artificial Intelligence, Natural Language Processing, Machine Learning, Speech recognition.

1. Introduction

Artificial intelligence refers to the use of computer technology to simulate and realize human intelligence. It draws on research from various disciplines including philosophy, psychology, linguistics, neuroscience and computer science. It is built on technologies such as machine learning, deep learning and natural language processing [1,2]. At present, artificial intelligence technology has been widely used in various industries such as intelligent medical care, financial technology, logistics management, and intelligent manufacturing [3]. Artificial intelligence plays the role of an assistant in many aspects of our lives. For example, robotic vacuum cleaners act as cleaning personnel, and voice-based AI serves as a personal assistant.

AI has wide-ranging applications in language learning, such as intelligent translation, personalized language learning, keyword extraction, text data analysis, and recognition. However, despite significant advancements in natural language processing and speech recognition, there are still challenges in the field of language learning [4]. Firstly, language is a complex symbolic system that encompasses multiple levels, including grammar, semantics, and pragmatics. While modern NLP technologies have made progress in grammar analysis and word sense disambiguation, difficulties remain in understanding rich semantics, contextual reasoning, and pragmatic understanding of language. Secondly, language learning

involves differences and diversities among different languages. Each language has its unique grammar structure, vocabulary, and expression methods. Additionally, language learning is also influenced by cultural, social, and historical factors. The use and understanding of language are influenced by cultural background, social context, and historical evolution. Ensuring that AI services in language learning are suitable for users from different cultural backgrounds is also an important issue to be addressed.
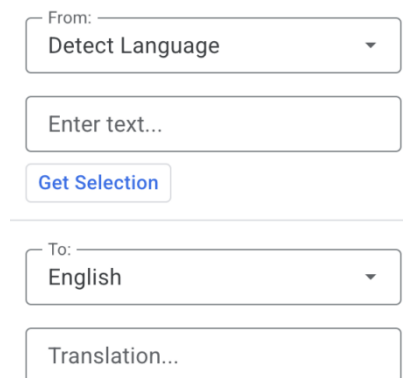
The author noticed that while artificial intelligence has been widely applied in the field of language learning and has been mentioned in relevant literature, there is still a lack of systematic summaries on how artificial intelligence is specifically applied in the field of language learning [5]. This deficiency serves as the main motivation for the formation of this article.

This article provides a comprehensive overview of the application of artificial intelligence in the field of language learning, focusing on three main aspects: translation, information retrieval, and language artificial intelligence. It also summarizes the current challenges and issues faced by artificial intelligence in the field of language learning, as well as prospects for future developments.

2. The Application of Artificial Intelligence in Translation

The development of artificial intelligence technology, along with the background of the information age, has profoundly influenced the research and theory of translation. The objects, subjects, modes, and environments of translation are undergoing transformation [1,4]. As an important application in modern life, this section of the article will discuss the artificial intelligence-related technologies that are applied in widely used translation software in the market, how they are applied, and the potential challenges in using artificial intelligence technology for translation.

Artificial intelligence translation software is a type of software that utilizes machine learning and natural language processing techniques to automatically translate text from one natural language to another. AI translation software achieves high-quality and fast translation by learning and analyzing large corpora of language data. These translation tools typically break down the original text into words or phrases and translate them based on their context and grammar rules [2]. In this section, Google Translate is taken as an example to analyze three technical algorithms (Figure 1).

Figure 1. Google Translate Interface (Original)

Google Translate: Google Translate is a widely used AI-powered translation tool that supports translation between multiple languages. Google Translate can conveniently convert the required text into other languages.

Google Translate uses a deep learning model called Neural Machine Translation. Neural Machine Translation is a machine translation method that utilizes artificial neural networks to predict the likelihood of word sequences and typically models and translates entire sentences within a single integrated model.

Google has developed multiple models based on NMT [6]:

Figure 2. Encoder-Decoder Architecture (Original)

Google Neural Machine Translation (GNMT): Introduced by Google in 2016, GNMT is an NMT model that utilizes an encoder-decoder architecture with attention mechanism (Figure 2).

The encoder-decoder architecture is a fundamental framework used in various natural language processing tasks. It enables sequence translation from one language to another by encoding the source sequence into a fixed-length representation and decoding it to generate the translated sequence. The encoder, utilizing deep neural networks such as RNN or Transformer models, captures the semantic and contextual information of the input sequence. It creates a context vector that contains the encoded information of the input sequence, which serves as the input for the decoder. On the other hand, the decoder takes the context vector generated by the encoder and generates an output sequence. It also employs models such as RNN or Transformer to iteratively decode and generate each token of the output sequence.

GNMT has significantly improved translation quality and has been applied in Google Translate services.
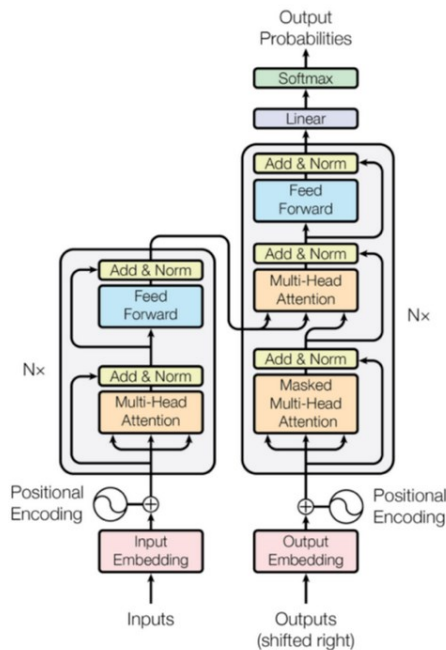
Figure 3. Structure of Transformer Model [4]

Transformer: Transformer is an NMT model based on self-attention mechanism, proposed by Google in 2017. The Transformer model departs from the traditional RNN structure and adopts multi-layer self-attention layers and feed-forward neural network layers. The self-attention mechanism allows the model to flexibly attend to different positions of the input sentence during the encoding and decoding process, capturing long-range dependencies more effectively (Figure 3) [7]. The Transformer model has achieved remarkable performance in translation tasks and has become a significant milestone in the NMT field.

Tensor2Tensor (T2T): T2T is an open-source framework developed by Google for training and deploying NMT models in machine translation and other natural language processing tasks. The T2T framework provides a range of optimized model architectures and training techniques, including Transformer models and other variants. It also offers a set of general-purpose data pre-processing and post-processing functions, making it easier for developers to build and train custom NMT models.

So far, the most widely used NMT model at Google is the Transformer model. It has a sophisticated self-attention mechanism and performs exceptionally well in handling long-range dependencies [8].

Now, Intelligent translation technology faces several challenges, including the differences and equivalence issues caused by language diversity, biases arising from incorrect contextual understanding, and inaccurate expression of specialized vocabulary. These aspects require further exploration and improvement. With the continuous development and maturation of technology, these challenges may be addressed in the future [9].

3. The application of Artificial Intelligence in Information Retrieval.

During information retrieval, artificial intelligence utilizes natural language processing techniques to process and understand human language. Additionally, it uses technologies like Word Embedding to convert words or phrases into vector representations, enabling machines to understand the meaning of textual vocabulary.
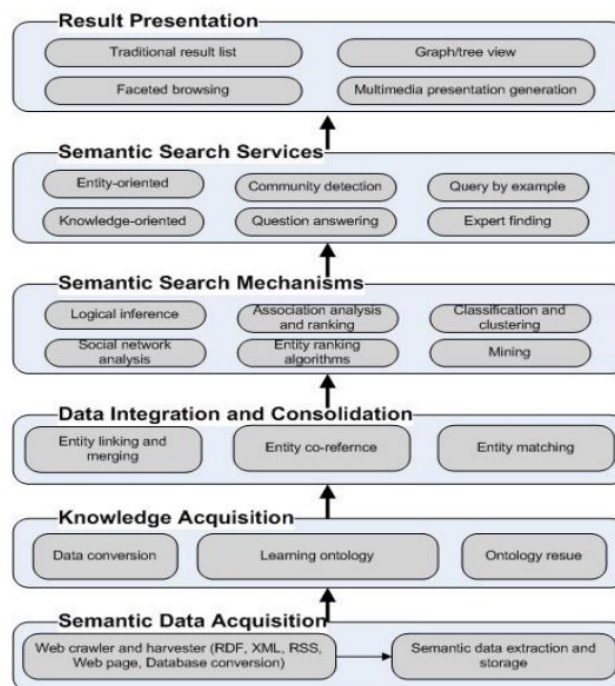
Figure 4. A semantic search framework [6]

Semantic search: Semantic search is an information retrieval technique based on semantic understanding (Figure 4). It utilizes artificial intelligence and natural language processing techniques to comprehend the intent behind user queries and retrieve the most relevant results. Unlike traditional keyword-based searches, semantic search focuses on understanding the context and semantic relationships within queries [10].

Semantic search relies on various natural language processing techniques, such as word sense disambiguation, entity recognition, syntactic analysis, and semantic role labeling, to parse the meaning of queries. By analyzing the words, phrases, and sentence structures in a query, semantic search can understand the user's intent and transform it into relevant concepts and topics [6].

4. The Application of Artificial Intelligence in Speech Recognition

Speech recognition is the process of using algorithms to translate spoken input into a sequence of words or textual signals. The main purpose of this technology is to develop systems for machine input through speech. Currently, speech recognition technology has been widely applied in various tasks that require human-machine interaction.

4.1. Type of Speech Recognition

Speech recognition system can be separated in different classes by describing what type of ullerances they can recognize [11]:

- Isolated Word

Isolated word recognition means that it only accepts a single word or a single piece of speech information at a time, which means there are only two states: "listening" and "not listening."

- Connected Word

Connected word systems have similarities with isolated word systems, but the difference lies in the fact that connected words allow individual utterances to be run together, minimizing pauses.

- Continuous speech

Continuous speech recognition systems allow users to speak continuously and naturally. Such recognizers are the most challenging to create because it means we must use special methods to determine speech boundaries.

- Spontaneous speech

In essence, Spontaneous speech can be regarded as a type of speech that sounds natural and unrehearsed. An ASR (Automatic Speech Recognition) system with spontaneous speech capability should be able to handle various natural speech features, such as word assimilation.

4.2. Speech Recognition Techniques

Speech recognition technology involves converting speech signals into recognizable text or commands, aiming to achieve human-computer interaction and automate speech processing. The core of speech recognition technology lies in speech data, which may contain information such as the speaker's identity and mood. More specifically, it includes specific information such as acoustic features, excitation sources, and behavioral cues. This article will delve into various aspects of speech recognition technology.

- Speaker identity information:

This includes acoustic features, which are primarily related to the speaker's physiological structure, including the vocal cords, throat, mouth, and nasal cavity. Similar to fingerprints, each person's acoustic features are unique and can be used for speaker identification. The excitation source refers to the energy source that generates the speech signal, such as vocal cord vibrations. The characteristics of the excitation source can provide more specific information about the speaker. Behavioral cues can be understood as speech signals that reflect the speaker's personalized behaviors, such as pitch, speaking rate, intonation, and phonetic rhythm. Analyzing these behavioral cues can enable further applications such as emotion recognition and semantic analysis.

- Behavior features in signals:

Apart from displaying speaker identity information, behavior features are also embedded in speech signals. Speech signals are propagated through sound waves, and various characteristics of sound waves can reflect the speaker's behavior. For example, emotional states can be analyzed through variations in pitch, intonation, and other acoustic features in speech signals. Different emotional states, such as happiness, sadness, or nervousness, exhibit varying frequency, intensity, and saturation characteristics in speech signals. Additionally, the speaker's speech rate and rhythm of phonemes can be obtained through analyzing speech signals, revealing the speaker's individual traits and linguistic habits.

- Speech analysis techniques:

Speech analysis technology extracts useful information from speech signals. Segmentation analysis techniques divide the continuous speech signal into fixed-size frames for further processing. The choice

of frame size should capture local features in the speech signal while maintaining sufficient time resolution. Sub-segmentation analysis techniques further divide each frame into smaller sub-frames to obtain finer-grained features, which is particularly useful for analyzing speech at a faster rate or detecting phoneme boundaries. On the other hand, over-segmentation analysis techniques divide the continuous speech signal into longer segments, allowing for capturing more global features such as assessing phonemic coherence, rhythm features, and prosody.

Apart from segmentation analysis, sub-segmentation analysis, and super-segmentation analysis, other speech analysis techniques are applied in speech recognition. For instance, phoneme recognition techniques identify and label each phoneme in speech signals, which is vital for the accuracy and semantic understanding of speech recognition systems. Moreover, feature extraction plays a crucial role in speech signal analysis. Common feature extraction methods include short-term energy, zero-crossing rate, and Mel-frequency cepstral coefficients (MFCC). These features capture essential information in speech signals, encoding and representing the sound.

Addressing noise and variation in speech data is also a significant focus in speech recognition technology research. Noise interferes with speech signals, reducing the accuracy of speech recognition. To mitigate this problem, researchers propose various noise reduction techniques, including model-based and data-driven methods. Furthermore, differences in speech among speakers and varying environmental conditions also lead to variations in speech signals.

With the rapid development of deep learning technology, state-of-the-art speech recognition systems often employ DNNs or RNNs as their models [12]. These models can learn complex representations and patterns in speech, thereby improving the accuracy of speech recognition. Additionally, there are emerging research directions, such as voiceprint recognition, emotion recognition, and cross-language recognition, which are driving advancements in speech recognition technology.

In summary, speech recognition technology is a complex and essential research field that enables human-machine interaction and automated speech processing. Speech data contains various information that reveals the identity of the speaker, including vocal tract characteristics, excitation sources, and behavior-induced speaker-specific features. Behavior features are also embedded in speech signals. Speech analysis plays a vital role in identifying appropriate frame sizes for further analysis and feature extraction. Common techniques include segmentation analysis, sub-segmentation analysis, and super-segmentation analysis. Deep learning technology has significantly improved the accuracy of speech recognition. The future holds promising applications of speech recognition in broader domains like intelligent assistants, smart homes, and medical diagnostics.

5. The issues with artificial intelligence in language learning

- Lack of Contextual Understanding

Language is rich with nuances, idiomatic expressions, cultural references, and context-specific meanings. Understanding these elements is crucial for accurate translation, interpretation, and comprehension. However, AI models primarily rely on statistical patterns and algorithms, which may not fully capture the complex interplay of context in language.

- Insufficient Personalization:

One of the challenges with speech AI is its inability to adapt to the language habits and preferences of different users. While AI systems have made remarkable progress in understanding and processing speech, they often lack the flexibility to accommodate the diverse language habits and variations among individuals.Language habits can include regional accents, dialects, pronunciation variations, and even personal speaking styles. These factors can significantly impact the accuracy and effectiveness of speech AI systems.

- Lack of Emotional Intelligence

While AI systems excel at processing and analyzing data, they struggle to understand and respond to human emotions effectively. Emotional intelligence involves recognizing, understanding, and appropriately responding to emotions, which is a complex and nuanced aspect of human communication.

6. Possible Development of Artificial Intelligence about Language Learning

With the vigorous development and in-depth research of natural language processing, new application directions will continue to emerge. The following are several possible development trends:

- **Personalized Learning:**

AI can further refine its ability to provide personalized learning methods for individual learners. By analyzing learners' performance data, preferences, and goals, AI algorithms can generate customized lesson plans, exercises, and content that cater to the specific needs and learning styles of each learner.

- **Intelligent Conversational Agents:**

AI-powered chatbots and virtual conversation partners can become more sophisticated in their language abilities. These conversational agents can engage in realistic and dynamic conversations with learners, providing valuable practice opportunities, feedback, and language correction.

- **Augmented Reality Language Learning**

AI combined with augmented reality (AR) technology can create immersive language learning experiences. Learners can interact with virtual language environments, objects, and characters, practicing their language skills in a simulated real-world context. AI can provide real-time guidance and feedback within the AR environment.

7. Conclusion

This article comprehensively outlines the applications of artificial intelligence in the field of language learning. Firstly, through the use of machine learning and natural language processing techniques, AI translation software can automatically translate text from one natural language to another. These intelligent software solutions have seamlessly integrated into people's lives. Secondly, artificial intelligence has found widespread applications in information retrieval. Furthermore, AI's application in speech has fundamentally changed the way we interact with technology. Speech AI has enabled accurate and efficient speech recognition systems that can transcribe spoken language into text, enabling hands-free communication and voice-controlled devices.

However, in the field of language learning, artificial intelligence still faces challenges. Language is a complex symbolic system that encompasses multiple levels, including grammar, semantics, and pragmatics. Although modern natural language processing techniques have made progress in grammar analysis and word sense disambiguation, challenges remain in understanding rich semantics, context reasoning, and pragmatic understanding of language. In the future, there is still tremendous potential for the application of artificial intelligence in language learning. Through further research and innovation, translation quality and adaptability can be improved, personalized language learning can be enhanced, and the capabilities of language AI can be strengthened. It is also important to address cultural differences and diversity in language learning, ensuring that AI services can meet the needs of users from different cultural backgrounds.

References

[1] Huashu Wang, & Shijie Liu. (2021). Research on the transformation of translation technology in the era of artificial intelligence. Foreign Language Teaching, 42(5), 87-92.

[2] Wei, Z. (2020, April). The development prospect of English translation software based on artificial intelligence technology. In Journal of Physics: Conference Series (Vol. 1533, No. 3, p. 032081). IOP Publishing.

[3] Long, J., Luo, C., & Chen, R. (2023, July). Multivariate Time Series Anomaly Detection with Improved Encoder-Decoder Based Model. In 2023 IEEE 10th International Conference on Cyber Security and Cloud Computing (CSCloud)/2023 IEEE 9th International Conference on Edge Computing and Scalable Cloud (EdgeCom) (pp. 161-166). IEEE.

[4] Jinchao Zhang. (2018),Why "Transformers" is powerful: A comprehensive analysis of the Google Tensor2Tensor system from model to code,https://cloud.tencent.com/developer/article/1153079

[5] Zhang, M., Li, Z., Zhang, F., & Ma, L. (2022). Adaptive Bidirectional Gray-Scale Center of Gravity Extraction Algorithm of Laser Stripes. Sensors, 22(24), 9567.

[6] Wei, W., Barnaghi, P. M., & Bargiela, A. (2008). Search with meanings: an overview of semantic search systems. International journal of Communications of SIWN, 3, 76-82.

[7] Yamada, M. (2023). Post-editing and a sustainable future for translators. Translation in Transition: Human and machine intelligence, 39.

[8] Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. English for Specific Purposes, 37, 112-121.

[9] Šoštarić, M., Pavlović, N., & Boltužić, F. (2019). Domain adaptation for machine translation involving a low-resource language: Google AutoML vs. from-scratch NMT systems. Translating and the Computer, 41, 113-124.

[10] Bast, Hannah; Buchhold, Björn; Haussmann, Elmar (2016). "Semantic search on text and knowledge bases". Foundations and Trends in Information Retrieval. 10 (2–3): 119–271. doi:10.1561/1500000032. Retrieved 1 December 2018.

[11] Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A review on speech recognition technique. International Journal of Computer Applications, 10(3), 16-24.

[12] Chenhui, C., & Rui, W. (2020). A Survey of Domain Adaptation for Machine Translation. Journal of Information Processing, 61(8).