Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ31736

The Technological Evolution, Applications and Future Prospects of Representative Artificial Intelligence Models

Moqian Shi

Roys, Hangzhou, China
shimoqian123@outlook.com

Abstract. This article aims to systematically sort out the development trajectory of artificial intelligence technology from theory to application, and clarify the evolution of its core model, the current challenges and the future development trend, in order to provide a clear reference for in-depth research in this field. For this reason, this article first traces the development process and milestone events of artificial intelligence since the mid-20th century. Then, the core concepts and architectures of four representative model types (Multi-Layer Perceptrons (MLPs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Transformer) and the successful application cases in key fields such as health care and finance are analyzed. In addition, through structured analysis, this article deeply explores the three core challenges faced by the current artificial intelligence system in the process of moving towards reliable and reliable deployment: model interpretability ("black box" problem), insufficient generalization ability and security vulnerabilities. In response to these problems, this article explores how key technical paths such as artificial intelligence, migration learning and federal learning can provide a blueprint for building the next generation of reliable artificial intelligence systems. Analysis shows that the development of artificial intelligence is shifting from the simple pursuit of performance breakthroughs to the construction of robust, safe and understandable intelligent systems. This transformation requires the coordinated development of algorithm innovation and ethical, legal and social frameworks.

Keywords: AI Survey, Deep learning, Model Development, Trustworthy AI

1. Introduction

From its inception in the mid-twentieth century, Artificial Intelligence (AI)-this emerging scientific field aimed at simulating, extending, and augmenting human intelligence has evolved from initial theoretical concepts into today's extensive practical applications. The goal to train machines in doing hard and complicated things that need human intelligence was very ambitious. Newest developments in AI have been huge mainly because of massive data availability, high computational power, and brilliant algorithmic improvements.

AI technology is still very much a hodgepodge of several root disciplines like image recognition, speech processing, and natural language processing to mention a few [1]. It has already made an extraordinary and brilliant landmark. The use of AI technology has been recognized as the major

DOI: 10.54254/2755-2721/2026.TJ31736

driving force behind the drastic changes in social production and even in the daily lives of people through its applications in e-commerce, medical diagnosis, and risk management in finance, etc. [2]. Nonetheless, as this technological system gets more and more complex and advanced, a thorough and transparent review is needed to systematically follow its developmental path, key technology changes, and state-of-the-art applications. As a result, the current review aims to deliver an organized presentation and integration of the artificial intelligence field, particularly dealing with its technology, existing problems, and upcoming possibilities, thereby providing a supportive reference for the studies in question.

The course of artificial intelligence development has been characterized by great ups and downs. Initially, the focus of research was on 'expert systems' based on symbolic logic that tried to mimic the skills of experts within a limited domain (like medical diagnosis) by 'programming' human specialists' knowledge into computers [3]. At the same time, the multi-layer perceptron model advocated by the connectionist movement established groundwork for the future deep learning revolution. Nevertheless, the actual use of these initial technologies was restricted by the limitations of the available computational power and data.

Undoubtedly, the face-off between world champion Go player Lee Sedol and DeepMind's AlphaGo in 2016 was the watershed event that signaled the rebirth of AI [4]. It was one of those historical times that was not only to make all AI in the public eye, but also to launch the AI investments in industrial and academic setup. AI applications have rapidly increased in different fields since that time: in healthcare, AI is used to examine medical images for helping in cancer screening; in chemistry and materials science it quickens the process of creating new molecules and discovering new drugs; in physics it is dealing with huge amounts of data coming from particle colliders; and self-driving cars projects such as Baidu's Apollo [5] and Tesla's [6] Autopilot are illustrating the highest level of AI usage in difficult real-life scenarios.

The great achievements in these areas are mainly attributed to the constant enhancement of the fundamental machine learning algorithms. The multilayer perceptron, being the first feedforward neural network, was able to sort out the linear separability problem. After that, it was a proposal of convolutional networks tailored specifically for grid-based data such as images was made. Via local connections and weight sharing, among other methods, they greatly improved the performance and efficiency of image recognition tasks [7]. Transformer architecture, which comes from natural language processing, has been one of the most significant enablers of handling long sequence data through its self-attention mechanism in the last few years. In addition to serving as the foundation for sizable language models like the GPT series, its ideas have been applied to domains like computer vision [8].

The past decade saw ample amount of studies in the AI area, yet, systematic reviews that could portray the path of ideas from their initial stages until the rise of high-tech innovations like Transformers are still rare. In this regard, the present paper intends to provide and that is why it brings in an extensive retrospective and summary of the technological development history and the current breakthroughs in artificial intelligence.

The left paper is made up of three additional parts. This paper will mainly guide the reader through the second section where the key representative methods along with their application areas will be discussed. The third section deals with debates and analyses of the current drawbacks of AI and future prospects. Finally, the fourth section presents the entire study as a conclusion in a nutshell.

DOI: 10.54254/2755-2721/2026.TJ31736

2. Representative machine learning models

2.1. Multi-Layer Perceptron

The MLP or Multi-layer Perceptron is the very first neural network architecture and the most important one among all basic deep learning models, thus paving the way for the creation of future intricate ones. The working of MLP is that it applies multiple fully connected layers along with non-linear activation functions to learn complex input-output mapping relations.

Mostly, a Multi-Layer Perceptron (MLP) is built with a minimum of the three layers: an input layer, a hidden layer, and an output layer.The complete interconnection of all the neurons in the neighboring layers is one of the most noticeable characteristics of MLPs. Essentially, the operation of an MLP is twofold, the first phase being the forward propagation during which the input data travels backwards through the layers starting from the input layer.The neurons of the present layer get all the outputs of the neurons of the previous layer, and each neuron performs the operations of computing weighted sum and applying nonlinear transformation through an activation function like ReLU or Sigmoid one after another. Second, in the process of training, the optimizers, e.g., the backpropagation algorithm [9] and gradient descent, make a series of adjustments to the entire weight and bias configuration of the network in accordance with the error (loss function) which is the difference between the predictions of the model and the real values.

MLP stands for multi-layer perceptron, and these models are very strong in the case of tabular structured data processing where features are not easily related spatially or temporally. In the medical field, MLPs can be trained to provide a prognosis of patient susceptibility to diabetes or heart disease based on age, blood pressure, and cholesterol levels among other factors [10].Indicatively, the main area of MLP utilization is in the light of its diverse applications in the financial sector, and one of them is credit scoring to mention the least. Here, the models determine the applicants' credit worthiness by taking into consideration various parameters like income, credit history, current debts along with some other personal attributes thus turning the whole loan issuance process into a more machine-like and uniform one [11].

2.2. Convolutional Neural Networks

For computers, images are just a cold matrix of millions of numbers representing colors. In order to explain this information, Convolutional Neural Networks (CNNs) have been developed [7].

CNN is a neural network design specifically meant for handling grid-like data. The main contribution of CNN over the past decades is the integration of two principles: local connection and parameter sharing. This will allow greatly diminished, model parameters and letting the network recognize such patterns on a global scale [7]. Traditionally, the framework of CNN has three main parts in sequence:

(1) At the beginning of the process come the convolutional layers, carrying on local feature extraction for example, edges and textures through the convolution operations while the learnable convolution kernel simultaneously walks through the input image.

(2) In contrast, the pooling layer is a way to shrink the size of the data and simultaneously retain the features' invariance; maximum pooling or average pooling methods are employed for gradually reducing the feature map's dimensions, which is how it works. This not only retains the most significant information, but also reduces the calculation volume and gives the model a certain degree of translation invariance.

(3) Full connection layers are usually located at the end of the network. They combine the advanced features from the previous convolutional layer and the pooled layer to complete the final classification or regression task [7].

CNN has completely changed the field of computer vision. In medical image analysis, CNN constitutes the core of an automated diagnosis system. It can analyze X-rays, MRI and CT scan images to detect pathological characteristics such as tumors and fractures with high accuracy, and provide strong support for radiologists [12]. In the field of financial technology, CNN is used to analyze visual financial data - for example, generating trading signals by identifying specific patterns in stock market heat maps or K-line charts, adding a powerful tool to quantitative investment strategies [13].

2.3. Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are a kind of neural network used to process serial data. It is characterized by the introduction of circular connection, which gives the network short-term memory ability, so that it can integrate historical information into the current calculation process. The RNN architecture works in such a manner that at each instant the input is presented to the network and simultaneously it receives the hidden state of the previous time step to determine the present output and further refresh the hidden state.This method of cycle can successfully simulate the time correlated to the sequence. Nevertheless, the conventional RNN suffers from the issue of gradients vanishing or exploding, which in turn limits its performance in learning the long-term dependencies.To tackle this issue, new and better structures were introduced, for instance, Long Short-Term Memory (LSTM) [14]. The incorporation of a gate control system that allowed information to be remembered or forgotten led to these networks being capable of processing very long sequences to a much greater extent.

RNN and its variants are good for various applications in sequence data processing. In medical monitoring where an RNN is used, continuous physiological parameter time series (like ECG) are analyzed to give a prior signal for critical situations such as arrhythmia [15].RNN in the area of financial technology has made a big step forward by creating a time series predictive model that successfully predicts the near future changes of financial items like stocks and forex through the examination of historical price data, thus supporting the practice of quantitative trading [16].

2.4. Transformer

Transformer architecture represents a revolutionary breakthrough in the field of neural network design [8]. It completely abandons the circular structure and relies on a mechanism called self-attention to process sequence data. This allows the computer to consider and weigh the importance of all other elements in the sequence at the same time when processing any single element in the sequence, thus capturing global dependencies in one step - which is completely different from RNN's sequential processing method. The self-attention mechanism lies at the center of the Transformer architecture, which is a multi-head one and thus allows the model to focus on the different representation subspaces simultaneously with their respective meanings (e.g., in the case of a word processing its syntax and semantic context can be simultaneously attended to).

In fact, the typical Transformer architecture is developed over the encoder-decoder paradigm [8]. The encoder, which consists of a number of identical layers stacked one over the other, is responsible for transforming the incoming sequence into a set of representations that are very rich in context. In order to ensure training stability and accelerated convergence, each encoder layer usually

DOI: 10.54254/2755-2721/2026.TJ31736

contains a multi-head self-attention sublayer and a feedforward neural network sublayer, and residual connection and layer normalization are applied after each sublayer. Decoders are also composed of similar stacking layers.The resulting target sequence is produced in an autoregressive way by the encoder's output and the mask mechanism is used to limit the circulation of the current word from the future information. Nevertheless, the model cannot detect the word order, hence, the position code has to be included in the input word embedding to enable the model to know the precise position of each word in the sequence.

So far, Transformer has become a practical standard for modern natural language processing and even multimodal fields. It is the basic building block of large-scale language models such as GPT and BERT [17].Transformer-based models significantly assist medical research by allowing the rapid analysis of extensive scientific literature and protein sequences, thus ultimately expediting the whole pharmaceutical research process [18].The financial analysts are mostly very busy in reading huge financial papers, the talking sessions of the related firms, and news to do the market mood and corporate risk assessments very well which results in a totally transparent investment decision process.

3. Discussion

The application of AI in large-scale and high-risk environments is still confronted with some serious problems, mainly because of the extremely rapid evolution of AI models like multilayer perceptrons, convolutional neural networks, recurrent neural networks, and transformers, which have already been improved a lot. The involvement of AI in high-risk situations and large-scale operations still confronts a lot of issues, mainly because of the rapid growth of the AI models such as multilayer perceptrons, convolutional neural networks, recurrent neural networks, and transformers that have undergone a considerable amount of improvement already [19].

3.1. Core challenge analysis

At present, the state-of-the-art AI systems, especially the deep neural networks, generally have the "black box" problem as one of their main drawbacks-the decision-making process and feature representations of the systems are still not revealed to the human eye [20]. The models' capability of performing highly complex nonlinear transformations is the main reason behind their very limited interpretability. To demonstrate, convolutional neural networks employed in medical imaging diagnostics could provide tumor localization with excellent precision, but doctors won't be able to pinpoint the exact radiologic characteristics (e.g., certain textures or edge blurriness) that the model is basing its decisions on. The "black-box effect" has major consequences: in the fields of healthcare, finance, and law, where the risk is high, unclarified decisions do not gain credibility and therefore, AI's ability is limited. If a model does not work properly, it gets hard for the developers to find out the exact reasons for the problem and fix it; besides, this also causes a dilemma in accountability-when the AI causes bad things to happen, it is difficult to tell who should be accountable.

Simultaneously, the model exhibits considerable constraints regarding its ability to generalize which in turn restricts its widespread utilization to a certain extent. For the model to be excellent in performance, one of the major prerequisites is that the training and test datasets should be subjected to the Independence and Identical Distribution (i.i.d.) Hypothesis. Though, real-world data distributions often differ greatly. Like this can be illustrated a self-driving car; if its training data contained only one area then it would not be able to transfer to another area probably due to

DOI: 10.54254/2755-2721/2026.TJ31736

different traffic rules or road conditions. It would suggest that for every alternative case, data accumulation and model retraining would be done again thus, making the whole process very costly and consequently, the acceptance and spreading of the technology would take a longer time.

In the end, the model's protection is the one that determines whether the system is viable or not. The two most telling points for this are as follows:

On the one hand, the model is vulnerable to adversarial attacks, and, on the other hand, the input data that undergoes such subtle and nearly invisible alterations can result in very serious misclassification that would thereby exposing the ambiguity of the decision-making limits.

On the other hand, there is a chance of data privacy breach. The model might learn confidential details and during the communication with the user might derive the user's private information which in turn could lead to severe compliance and ethical issues.

3.2. Future technology pathways

To overcome these difficulties, the new research frontiers give rise to the following important directions: The very first, the creation of explainable AI technology forms the foundation for establishing a dependable model. Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), which are after-the-fact interpretation methods, have turned into functional instruments that elucidate individual predictions by means of the feature contributions' quantification. Exploring the neural symbol system is a wider trend that is the combination of the deep learning's perceptibility with the symbolic reasoning's interpretability and logic, thus, creating a reasoning chain which is humanly and verifiably understandable.

Enhancing the transferability and domain adaptability of the learning process is the next step in the development of generalization ability. Migration learning relies on the vast knowledge of the source domain to support the learning of the target domain, which in turn, significantly cuts down the need for labelled data in the new situations. Domain adaptation technology further focuses on aligning the distribution of cross-domain characteristics, so that the model can automatically adapt to distribution differences - this is the core technology to realize the vision of "one learning, applying everywhere" [21].

In addition, the federal learning paradigm has achieved a balance between data privacy protection and value sharing. Federal learning enables multiple participants to work together to train global models by exchanging encrypted model parameters without transferring data outside the local boundary. The approach not only enables the unification of different data values located in various places but also ensures that data security and privacy are respected. It is pointed out that this technique opens up new and imaginative paths for the collaboration of the compliance in health care and finance where the sensitivity of the matters is great [22].

Taking everything into account, the AI development has shifted from the so-called "unrestrained growth" performance-centered phase to the new era of 'meticulous cultivation' that emphasizes the qualities of reliability, universality, security, and trustworthiness. Winning over these barriers not only needs the creation of new algorithms but also the combined progress in technology, ethics, law, and social awareness.

4. Conclusion

The study presented in this paper takes a stepwise and comprehensive look back at the past of artificial intelligence covering its evolution from the very beginning when it was merely a theoretical idea to being utilized in practical ways, and it also features the major transitions of the standard models across such dimensions as MLP, CNN, RNN, and Transformer, and their influence on and usefulness in several fields including healthcare and finance among others. Despite the present-day AI that has already gone through a lot of tests and analyses, the AI explainability, generalization capabilities, and safety still stick to the core challenges of AI. Technologies such as explainable AI and federated learning generalization will eventually be the ones that will build the trust in and the reliance upon AI, if they are accompanied by ethical and safe considerations.

References

[1] McCarthy, J., Minsky, M.L., Rochester, N. and Shannon, C.E. (2006) A proposal for the Dartmouth summer research project on artificial intelligence. AI Magazine, 27, 12–14.

[2] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M. and Thrun, S. (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542, 115–118.

[3] Russell, S. and Norvig, P. (2020) Artificial Intelligence: A Modern Approach. 4th Edition, Pearson.

[4] Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016) Mastering the game of Go with deep neural networks and tree search. Nature, 529, 484–489.

[5] Apollo (2025) Apollo—An open autonomous driving platform. Available from: https: //apollo.auto/index.html

[6] Tesla (2025) Autopilot. Available from: https: //www.tesla.com/autopilot

[7] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86, 2278–2324.

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017) Attention is all you need. Advances in Neural Information Processing Systems, 5998–6008.

[9] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning representations by back-propagating errors. Nature, 323, 533–536.

[10] Chushig-Muzo, D., Córdova-Suárez, M., Hidalgo-Andrade, P., et al. (2024) Interpretable data-driven approach based on feature selection methods and GAN-based models for cardiovascular risk prediction in diabetic patients. IEEE Access, 12, 87073–87087.

[11] Fukumi, M., et al. (2015) Three-MLP Ensemble Re-RX algorithm and recent classifiers for credit-risk evaluation. Proceedings of the International Joint Conference on Neural Networks, 1–8.

[12] Goh, Y.T., Chua, H.N., Jasser, M.B., Wong, R.T.K. and Issa, B. (2025) Multi-class classification of lung cancer using streamlined convolutional neural networks. Proceedings of the IEEE International Colloquium on Signal Processing and Its Applications, 301–306.

[13] Gadhiya, U., Faldu, P., Darji, K., Obaidat, M.S., Gupta, R. and Tanwar, S. (2024) CNN-based application recognition to enhance network governance for financial networks. Proceedings of the International Conference on Computer, Information and Telecommunication Systems, 1–5.

[14] Hochreiter, S. and Schmidhuber, J. (1997) Long short-term memory. Neural Computation, 9, 1735–1780.

[15] Jovanovic, L., et al. (2024) Metaheuristic optimized electrocardiography time-series anomaly classification with recurrent and long short-term neural networks. International Journal of Hybrid Intelligent Systems, 1–26.

[16] Srivastava, S., et al. (2024) Leveraging RNNs and LSTMs for synchronization analysis in the Indian stock market: A threshold-based classification approach. arXiv preprint, arXiv: 2409.06728.

[17] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the North American Chapter of the Association for Computational Linguistics, 4171–4186.

[18] Jumper, J., Evans, R., Pritzel, A., et al. (2021) Highly accurate protein structure prediction with AlphaFold2. Nature, 596, 583–589.

[19] Kozlov, D., Sahoo, S.S., Sánchez, V.P., et al. (2024) Establishing and evaluating trustworthy AI: Overview and research challenges. Frontiers in Big Data, 7, Article 1467222.

[20] Chander, B., John, C., Warrier, L. and Gopalakrishnan, K. (2024) Toward trustworthy artificial intelligence (TAI) in the context of explainability and robustness. ACM Computing Surveys, 56, Article 228.

[21] Zhu, D., Bu, Q., Zhu, Z., Zhang, Y. and Wang, Z. (2024) Advancing autonomy through lifelong learning: A survey of autonomous intelligent systems. Frontiers in Neurorobotics, 18, Article 1385778.

[22] Langford, M.A., Zilberman, S. and Cheng, B.H.C. (2024) Anunnaki: A modular framework for developing trusted artificial intelligence. ACM Transactions on Autonomous and Adaptive Systems, 19, Article 11.