World Journal of Innovation and Modern Technology, Vol. 7, Issue 2, (Apr) 2024
ISSN 2682-5910
DOI: 10.53469/wjimt.2024.07(02).13

Enabling Intelligent Decision Making and Optimization in Enterprises through Data Pipelines

Chunhe Ni[1*], Chenwei Zhang[2], Wenran Lu[3], Hongbo Wang[4], Jiang Wu[5]

[1]Computer Science, University of Texas at Dallas, Richardson, TX, USA
[2]Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL, USA
[3]Electrical Engineering, University of Texas at Austin, Austin, TX, USA
[4]Computer Science, University of Southern California, Los Angeles, CA, USA
[5]Computer Science, University of Southern California, Los Angeles, CA, USA
[*]*Corresponding author Email: nichunhe@outlook.com*

Abstract: ***This article explores the pivotal role of data pipelines in driving intelligent decision-making and optimization within manufacturing enterprises. It delves into the transformative potential of technologies such as the Internet of Things (IoT), big data analytics, and artificial intelligence (AI) in revolutionizing operations. By seamlessly integrating diverse data sources through intricate pipeline networks, enterprises can gain invaluable insights into market dynamics, consumer preferences, and operational efficiencies. Real-time data processing capabilities empower enterprises to dynamically adjust production processes, anticipate demand fluctuations, and optimize resource allocation, leading to significant improvements in efficiency, quality, and competitiveness. The article emphasizes the importance of establishing a sound data infrastructure, cultivating data analysis talents, developing data governance policies, promoting a data-driven decision-making culture, and continuously optimizing data application scenarios. Through methodologies like data cleaning, improving accuracy, ensuring consistency, and enhancing efficiency, enterprises can harness the full potential of data to achieve sustainable growth and competitive advantage in today's digital age.***

Keywords: Data pipelines; Intelligent decision-making; Data cleaning; Data analysis.

1. INTRODUCTION

Competitive business landscape, the collection, analysis, and utilization of data have emerged as pivotal means for enterprises to gain a competitive edge. With the advent of technologies such as the Internet of Things (IoT), big data analytics, and artificial intelligence (AI), the possibilities for data-driven intelligent operations have become endless. This paradigm shift allows manufacturing enterprises to not only understand market demands better but also to optimize product designs, enhance production efficiency and quality, thus propelling them to stand out amidst the competition.

The integration of data pipelines has become instrumental in this journey towards data-driven intelligence. By seamlessly collecting and presenting data, manufacturing enterprises can gain invaluable insights into market dynamics, consumer preferences, and operational efficiencies. This foundational understanding paves the way for informed decision-making at every level of the organization.

Furthermore, the convergence of [1] IoT, big data, and AI technologies enables real-time data collection and processing on an unprecedented scale. This real-time capability empowers enterprises to dynamically adjust production [2] processes, anticipate demand fluctuations, and optimize resource allocation. Consequently, manufacturing processes can be automated and streamlined, leading to significant improvements in production efficiency, product quality, cost reduction, and market responsiveness.

In this article, we will delve deeper into how data pipelines serve as the backbone for intelligent decision-making and optimization in manufacturing enterprises. We will explore the transformative potential of these technologies, their practical applications, and the benefits they offer in terms of efficiency, quality, and competitiveness. Through real-world examples and case studies, we aim to illustrate how leveraging data pipelines can propel enterprises towards greater agility, innovation, and sustainable growth in today's rapidly evolving market landscape.

2. RELATED WORK

Contemporary data infrastructure has evolved into a vast and intricate ecosystem that plays a pivotal role in modern operations. The exponential expansion of data supply and demand has engendered a proliferation of diverse data sources, ranging from social media platforms and [3] IT logs to IoT sensors and real-time streaming data. To accommodate this deluge of information, novel storage solutions such as data lakes have emerged, offering scalability, ease of deployment, and cost-effectiveness.

2.1 Data Pipeline

Facilitating the seamless integration of this multitude of data sources with data lakes is the establishment of intricate data pipeline networks. These networks serve as conduits, linking data lakes to various data sources and applications. This connectivity necessitates the adoption of an array of technologies including Extract, Load, Transform (ELT) [4] processes, Change Data Capture [5] (CDC), application programming interfaces [6] (APIs), and event streams. Such technologies are deployed to meet escalating demands for real-time operations and AI-driven predictive analytics.

ETL pipelines extract data from their source systems/databases, transform the data, and then load it into a data warehouse or database from which to derive business insights.The data pipeline includes the ETL pipeline because for the data pipeline, the destination of the data does not have to be a database or data warehouse, but can also be other applications, and supports the orchestration, management, and use of the data throughout the organization.The steps in a data pipeline typically include extraction, transformation, composition, validation, visualization, and other such data analysis processes [7-9]. Without a data pipeline, these processes require a lot of time-consuming and tedious manual steps and leave room for human error. The best analogy for a data pipeline is a conveyor belt, which efficiently and accurately delivers data to every step of the process. For example, data pipelines help data flow efficiently from SaaS applications to data warehouses and so on.

Figure 1: Data pipeline architecture diagram

A data pipeline is an operation of a set of processing steps that integrates raw data from multiple sources into one destination for storage, business intelligence (BI), data analytics, and visualization. If the data is not currently loaded into the data platform, the data ingestion takes place at the beginning of the pipeline. Each step then generates an output that serves as the input to the next step. This process continues until the pipeline is completed. A data pipeline consists of three key elements: the data source, one or more processing steps, and the destination location. In some data pipelines, the destination location may be referred to as the receiver. Data pipelines allow data to flow from applications to data warehouses, from data lakes to analytics databases, or into payment processing systems [10]. A data pipeline may also have the same data source and destination location, making the pipeline purely about modifying the data set. Whenever data is processed between points A and B (or B, C, and D), there is a data pipeline between them.

2.2 Data-Driven Decision Making

The core of data-driven decision-making lies in mining the patterns and trends hidden behind the massive amount of information. In nursing work, this may include basic patient information, medical history, indicators during nursing operations, and even diverse data sources such as patient satisfaction feedback [11]. Through the systematic collection, integration and analysis of these data, it can reveal the potential problems, bottlenecks and optimization space in the nursing process. For example, in ward management, through real-time monitoring and analysis of key data such as bed turnover rate and nurse work efficiency, managers can quickly locate the links that affect work efficiency, and adjust human resource allocation accordingly to reduce waiting time and improve patient flow speed. At the same time, it can also be found through data analysis that the pressure of nursing staff is too large in a specific period or a specific operation, and then adjust the scheduling system or provide special training to ensure the stable and efficient quality of nursing service.

Data-driven Decision Making (DDD) [12] is a methodology that relies primarily on Data analysis and interpretation in the decision-making process, rather than relying on intuition or personal experience. This is in contrast to traditional experience-driven, intuition-driven, or bias-driven decision-making. By collecting and analyzing users' viewing habits, ratings, searches, and other behavioral data, Netflix has developed highly personalized recommendation algorithms to enhance the user experience and increase user viewing time. This data-driven approach is also being applied to deciding which movies and [13] TV shows should be purchased or produced. For example, its original series "House of Cards" is based on the results of extensive user data analysis.

2.3 Enterprise intelligent decision

Data-driven analytics can not only help companies more accurately understand market trends, competitive environments and customer needs, so as to make more targeted strategic planning. For example, by analyzing data such as market share, growth rate, etc., companies can identify new business opportunities and potential threats. In addition, enterprises can also adjust product strategy and pricing strategy to meet market demand through the analysis of customer demand and behavior data. It can also support the optimization of enterprise operations, which are the core links of enterprises [14]. Optimizing the operation process can reduce costs and improve efficiency, thus enhancing the competitiveness of enterprises. By analyzing internal operational data, companies can identify bottlenecks, waste, and room for improvement in their processes. For example, through the analysis of production data, enterprises can optimize the production schedule and improve the utilization of equipment; Through the analysis of logistics data, enterprises can reduce inventory costs and improve distribution efficiency.

Secondly, while improving customer satisfaction, it can also promote innovation [15]. Customers are fundamental to the survival and development of enterprises, and improving customer satisfaction is the primary task of enterprises. Data analytics can help companies gain a deeper understanding of customer behavior and needs to provide more personalized and tailored products and services [16]. For example, through the analysis of customer purchase data, enterprises can discover customer preferences and consumption habits, and then adjust product lines and marketing strategies. In addition, data can also be used to optimize the customer experience, for example, through the analysis of website access data, companies can improve website usability, customer satisfaction and loyalty. Data can provide inspiration and support for business innovation. By analyzing data such as customer feedback and market trends, companies can discover new business opportunities and areas of innovation. In addition, data can help companies evaluate the feasibility and potential value of innovative projects. For example, through the analysis of market demand and competitive situation, enterprises can determine the development direction and market positioning of new products.

In order to fully leverage the central role of data in business decisions, companies need to implement the following data-driven decision strategies:

(1) Establish a sound data infrastructure

Enterprises need to establish a sound data infrastructure, including data collection, storage, management and analysis. This includes adopting appropriate data storage technologies to ensure data security and integrity; Improve the speed and accuracy of data analysis with efficient data processing and analysis tools [17-19].

(2) Cultivate data analysis talents

Enterprises need to train talents with data analysis skills, including data analysts, data scientists, etc. These talents can skillfully use data analysis tools to mine valuable information from massive data and provide support for corporate decision-making.

(3) Develop a data governance policy

Enterprises need to develop a data governance strategy to ensure the quality and reliability of data. This includes data cleaning, data standardization, data security and so on. Only high-quality data can provide effective support for enterprise decision-making.

(4) Promote a data-driven decision-making culture

Companies need to promote a data-driven decision-making culture and encourage employees to use data in their daily work. This means breaking away from the traditional experience-driven, intuition-driven decision-making model and relying more on data and analytical results to make decisions.

(5) Continuously optimize data application scenarios

Enterprises need to continuously pay attention to the application scenarios of data in different business areas to discover the value of data. As the business evolves and changes, enterprises need to constantly adjust their data application strategies to adapt to new market environments and customer needs.

3. METHODOLOGY

In the process of intelligent decision making, data plays a crucial role [20]. Enterprises must fully recognize the central role of data and actively invest in data infrastructure construction, talent training and data governance. Only by implementing data-driven decision-making strategies can companies better respond to market changes and achieve sustainable growth and competitive advantage. In the digital age of the future, data will become a key factor in business success.

First, we will establish an efficient data collection and storage system to ensure the accuracy, completeness and timeliness of data. This involves the construction of data infrastructure, including hardware and software selection, deployment and maintenance. Secondly, we will focus on talent training and establish a professional data team, including data analysts, data engineers and data scientists, who will be responsible for data cleaning, analysis and mining to provide reliable data support for decision-making [21]. At the same time, we will also strengthen data governance to ensure data security and compliance. This includes the establishment of data management systems, the development of data use policies, and the use of technology to enhance data protection and monitoring.

Finally, we will implement data-driven decision-making strategies, use data analysis and machine learning technologies to dig deep into massive data, find hidden rules and values in the data, and provide scientific basis and accurate guidance for corporate decision-making. Through the implementation of the above methodology, we will help enterprises establish a sound data pipeline, realize the intelligent use of data, improve the accuracy and efficiency of decision-making, and thus succeed in the fierce market competition.

3.1 Data Cleaning

Data Cleaning is often seen as a key preparation step for data-driven decisions, with the goal of finding and correcting errors and inconsistencies in data to improve data quality. As datasets grow, ensuring data cleanliness and integrity becomes more challenging. Understanding the importance of data cleansing and how to do it becomes critical.

Data cleaning is a crucial part of the data processing PipeLine [22]. Although it is often thought of as a relatively simple process that only involves deleting bad data, the absence of a comprehensive data cleansing process can produce suboptimal [23] or outright wrong results, even with the most advanced analytical techniques or algorithms.

Having clean and accurate data provides a solid foundation for any data-driven work. Good data cleaning effect can play a role in all stages of data application. Here are seven reasons why data cleansing is especially important: Accuracy: Incorrect data may lead to incorrect analysis and conclusions. By cleaning, you can ensure that the data accurately reflects the real scenario.

Consistency: Data sets are typically sourced from a variety of sources and may have different formats or standards. Cleaning ensures consistency in data sets and makes analysis and model training easier.

Efficiency: Clean data simplifies the analysis process and speeds up model training [24]. When trained on a clean data set, the algorithm converges faster and requires fewer computational resources.

Reliability: Data cleansing can reduce the likelihood of false results or misleading patterns, ensuring that models and analyses are more credible.

Correlation: Cleaning helps filter out irrelevant or redundant features or records, ensuring that the analysis or model is focused on relevant data.

Interpretability: Models trained on clean data are often more interpretable, making it easier to diagnose problems, interpret results, and gain actionable insights.

Optimization: High-quality data often means that models can be more straightforward and practical, reducing the risk of overfitting and enhancing generalization.

Data cleansing plays a key role in ensuring the accuracy and credibility of data-driven decisions. By adopting appropriate data cleaning methods, the quality of data can be improved to the greatest extent and provide a reliable basis for subsequent analysis and model construction.

3.2 Improve the accuracy of intelligent decision making

Using data-driven analytics to identify and deal with outliers is one of the key steps in ensuring the accuracy of business decisions. Outliers can lead to a decline in data quality, which can affect the outcome of subsequent analysis and decision making. In a data-driven environment, outliers are identified and processed to ensure the accuracy and reliability of the data in order to provide a trusted basis for decision making. By employing techniques such as interquartile spacing (IQR) [25], outliers can be detected and corrected in a timely manner, thus ensuring the quality of the data set. Such data quality control measures provide enterprises with a more reliable data basis, which in turn improves the accuracy and credibility of decisions. Therefore, having outlier detection and processing as part of a data-driven analytics process is critical to the success of an enterprise's decision making process.

Data error: For example, the age value in the data set contains a negative number, which may be caused by a manual entry error or a system failure. Code converts data with a negative age to its absolute value. This correction ensures that the data in the "age" column conforms to the valid age range in the real world, avoiding potential problems in subsequent analysis or model training.

```
df.loc[df['age'] < 0, 'age'] = df['age'].abs
```

Text data errors: For example, there may be typographical errors in text data, especially when dealing with open-ended text responses or project names. Use text processing techniques, such as spell checking or similarity matching, to detect and correct typographical errors. Ensuring the accuracy of textual data is critical to maintaining data integrity, especially where open text is involved.

```
typo_mapping = {"Applle": "Apple", "Bannana": "Banana"}

df['fruit'] = df['fruit'].replace(typo_mapping)
```

The above processing methods can effectively deal with data inaccuracies and ensure the quality and integrity of data sets, thus providing a reliable basis for subsequent analysis and modeling.

3.3 Consistency

Consistency plays a key role in data processing. As datasets grow in size, data tends to come from multiple sources, which inevitably introduces inconsistencies. Such inconsistencies are not just external appearances; deeper effects can include introducing noise, reducing analytical capabilities, and even leading to erroneous conclusions. In such an environment, ensuring data consistency becomes critical.

Consistency is more than just surface tweaking of data, it involves ensuring that each entry is treated according to uniform standards. Such uniform processing eliminates the risk that certain data points may unduly influence the results due to their format or standard. Consistent data sets provide a reliable basis for seamless data integration and simplified analysis, keeping data consistent across different links, ensuring reliability and accuracy for subsequent operations. As a result, consistency cleaning plays a vital role in improving overall data quality while paving the way for data-driven tasks.

A common consistency problem is the date, converting the date column to a consistent date format (or the same time zone, time zone consistency is also particularly important); By standardizing dates, we can more easily perform time series analysis and other date-specific operations.

Text data processing usually faces big challenges, such as case inconsistencies, abbreviations, ambiguity processing, etc. Unified text data representation is very important, especially for text mining, clustering, classification and other tasks have a great impact. For example, "Apple, "" APPLE," and "Apple" would be treated as the same entity, eliminating potential duplications and misclassifications.

Text data processing usually faces big challenges, such as case inconsistencies, abbreviations, ambiguity processing, etc. Unified text data representation is very important, especially for text mining, clustering, classification and other tasks have a great impact. For example, "Apple, "" APPLE," and "Apple" would be treated as the same entity, eliminating potential duplications and misclassifications.

Data pipeline analysis is a method of collecting, cleaning, transforming, and analyzing data through automated processes. It can improve the consistency of enterprise decision-making, mainly in the following aspects:

Data consistency: Data pipelines ensure data consistency because they are processed through predefined processes and rules. This means that no matter where the data comes from, it will have the same format and quality standards after being processed through the data pipeline. This consistency ensures that the data used in decision analysis is accurate, reliable, and comparable.

Process consistency: The automated flow of the data pipeline ensures consistency in data processing and analysis. Whenever an analysis task is performed, the data pipeline processes the data according to the same steps and rules. This eliminates human error and inconsistencies and ensures that every decision is made based on the same data processing processes and criteria.

Real-time consistency: Data pipelines can help ensure real-time consistency of data. By collecting and updating data from data sources on a regular or real-time basis and transferring it to an analytics system for processing, data pipelines ensure that the data on which decisions are based is always up to date and most accurate. This helps avoid the risk of making decisions based on outdated or inconsistent data.

Standardized consistency: Data pipelines promote standardized consistency of data. By applying consistent specifications and standards to data processing, data pipelines ensure that data between different data sources is processed and stored in a uniform format and structure. This enables more consistent and comparable data analysis and decision making across different departments or business units.

In summary, data pipeline analysis can improve the consistency of intelligent decision making in enterprises by ensuring consistency of data, process, real-time and standardization. This helps ensure that every decision is based on the same, reliable data and processing processes, which improves the quality and credibility of decisions.

3.4 Efficiency

When dealing with large and complex data, efficient data processing is crucial. Redundant records, incorrect data types, or inconsistent formats can impose a significant burden on computing resources. Not only does this increase the computation time, but it may also slow down the entire analysis process, requiring additional processing and handling of special cases. By cleaning the data properly, performance can be optimized so that algorithms and processing operations run smoothly and quickly.

By adjusting the data type of the column, you can optimize the data storage. For example, converting integer values stored in the default 64-bit integer format to 32-bit format effectively reduces memory usage while ensuring that the data integrity of columns with smaller integer ranges is not affected.

In many datasets, text or classification columns are often represented as strings. However, when these columns contain fewer unique values, representing them as strings is inefficient both in terms of memory and computational speed. Consider a "gender" column with entries such as "male," "female," and "non-binary." Instead of repeatedly storing these strings in millions of lines, use a more space-efficient representation. The following code snippet takes advantage of the data types in pandas, which internally map unique strings to integers and store only integer

values, greatly reducing memory usage. Not only does this help save memory, but it also speeds up operations such as sorting and grouping the column, since integer operations are generally more efficient than string operations.

Sparse data structures come into play when a dataset has a large percentage of missing or zero values. In a standard data structure, each of these missing or zero values will still occupy memory. In contrast, sparse data structures store only non-zero or non-missing values and their locations, saving a lot of memory.

Therefore, in the intelligent decision-making of enterprises, data analysis has the core advantage of improving efficiency. By digging deep into and analyzing massive amounts of data, data analytics can quickly and accurately provide business insights and trend predictions, enabling decision-makers to make informed decisions quickly, saving time and resources [27]. Optimizing resource allocation and reducing risk is an important function of data analysis, which helps enterprises use resources more effectively, reduce unnecessary costs and risks, and thus improve the efficiency of enterprise operations. By providing reliable support and guidance to businesses, data analytics promotes growth and competitiveness, enabling them to stay ahead of the curve in a competitive market.

The real value of data lies in its quality. Ensuring accuracy, consistency, efficiency, reliability, relevance, interpretability, and optimization is critical. Data quality serves as the cornerstone of robust analytics. Prioritizing the cleaning, optimization, and generation of high-quality data enhances the robustness and trustworthiness of data analytics and models, thereby facilitating informed decisions and impactful insights.

Moreover, high-quality data analysis can enable companies to make intelligent decisions. By harnessing the combination of big data and artificial intelligence technology, enterprises can unearth hidden patterns and rules from vast datasets, providing accurate predictions and recommendations for decision-makers. This form of intelligent decision-making not only enhances operational efficiency but also mitigates decision-making risks, thereby providing solid support for the sustainable development of enterprises.

4. CONCLUSION

In today's competitive business environment, enterprise data analytics is a powerful engine driving intelligent decision making. Through careful data cleaning and optimization, companies can gain a deeper understanding of market trends, customer behavior, and competitive dynamics [28-30]. This in-depth insight not only helps to capture business opportunities in a timely manner, but also enables to identify potential risks in advance and develop appropriate response strategies.

To sum up, the advantages and functions of enterprise data analysis for intelligent decision-making cannot be ignored. By ensuring data quality and optimizing data analytics and models, companies can better grasp market dynamics, improve decision-making efficiency, and achieve sustainable development. Therefore, enterprises should pay attention to data quality management and continuously improve data analysis capabilities to cope with increasingly complex business challenges.

In conclusion, the integration of data pipelines and data-driven decision-making in manufacturing enterprises holds the promise of enhancing competitiveness through real-time insights, operational efficiency, and innovation. By leveraging these technologies, companies can navigate market dynamics with agility, driving sustainable growth and staying ahead in the digital era.

ACKNOWLEDGEMENT

Thanks to the research team of Xu, J., Wu, B., Huang, J., et al., who conducted in-depth research on the practical application of advanced cloud services and generative AI systems in the field of medical image analysis in literature [1]. Their work has provided us with valuable insights and expanded our understanding of the potential of big data analytics and AI applications in medicine. Thanks to their hard work and pioneering spirit, they have made important contributions to promoting the development of the field of medical image analysis.

Thanks to Wu, Binbin, et al. "and his team for the enterprise digital intelligent remote control system based on Industrial Internet of Things proposed in literature [2]. Their research shows us the potential of iot technology in the field of enterprise intelligence and provides new ideas and methods for achieving digital transformation. Thanks to their in-depth exploration of the industrial Internet of Things and digital intelligent systems, they have made important contributions to promoting the development of enterprise intelligence.

Through their in-depth research and innovative thinking, I was able to gain a deeper understanding of the application and potential of advanced cloud services, generative artificial intelligence systems, and industrial Internet of Things technologies in different fields. Their results have provided valuable references and references for my writing, enabling me to better explore and express related topics in the fields of big data analysis and artificial intelligence. Thank you again for their contributions and efforts, and I will continue to work hard to make progress and contribute more in these areas

REFERENCES

[1] Xu, J., Wu, B., Huang, J., Gong, Y., Zhang, Y., & Liu, B. (2024). Practical Applications of Advanced Cloud Services and Generative AI Systems in Medical Image Analysis. arXiv preprint arXiv:2403.17549.
[2] Wu, Binbin, et al. "Enterprise Digital Intelligent Remote Control System Based on Industrial Internet of Things." (2024).
[3] Zhang, Y., Liu, B., Gong, Y., Huang, J., Xu, J., & Wan, W. (2024). Application of Machine Learning Optimization in Cloud Computing Resource Scheduling and Management. arXiv preprint arXiv:2402.17216.
[4] Xu, Z., Gong, Y., Zhou, Y., Bao, Q., & Qian, W. (2024). Enhancing Kubernetes Automated Scheduling with Deep Learning and Reinforcement Techniques for Large-Scale Cloud Computing Optimization. arXiv preprint arXiv:2403.07905.
[5] Zhou, Y., Tan, K., Shen, X., & He, Z. (2024). A Protein Structure Prediction Approach Leveraging Transformer and CNN Integration. arXiv preprint arXiv:2402.19095.
[6] Sun, Guolin, et al. "Revised reinforcement learning based on anchor graph hashing for autonomous cell activation in cloud-RANs." Future Generation Computer Systems 104 (2020): 60-73.
[7] Li, H., Wang, X., Feng, Y., Qi, Y., & Tian, J. (2024). Integration Methods and Advantages of Machine Learning with Cloud Data Warehouses. International Journal of Computer Science and Information Technology, 2(1), 348-358.
[8] Lei, H., Chen, Z., Yang, P., Shui, Z., & Wang, B. (2024). Real-time Anomaly Target Detection and Recognition in Intelligent Surveillance Systems based on SLAM.
[9] Gong, Y., Huang, J., Liu, B., Xu, J., Wu, B., & Zhang, Y. (2024). Dynamic Resource Allocation for Virtual Machine Migration Optimization using Machine Learning. arXiv preprint arXiv:2403.13619.
[10] Liu, B. (2023). Based on intelligent advertising recommendation and abnormal advertising monitoring system in the field of machine learning. International Journal of Computer Science and Information Technology, 1(1), 17-23.
[11] Pan, Y., Wu, B., Zheng, H., Zong, Y., & Wang, C. (2024, March). THE APPLICATION OF SOCIAL MEDIA SENTIMENT ANALYSIS BASED ON NATURAL LANGUAGE PROCESSING TO CHARITY. In The 11th International scientific and practical conference “Advanced technologies for the implementation of educational initiatives”(March 19–22, 2024) Boston, USA. International Science Group. 2024. 254 p. (p. 216).
[12] K. Xu, X. Wang, Z. Hu and Z. Zhang, "3D Face Recognition Based on Twin Neural Network Combining Deep Map and Texture," 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China, 2019, pp. 1665-1668, doi: 10.1109/ICCT46805.2019.8947113.
[13] Shi, Peng, Yulin Cui, Kangming Xu, Mingmei Zhang, and Lianhong Ding. 2019. "Data Consistency Theory and Case Study for Scientific Big Data" Information 10, no. 4: 137. https://doi.org/10.3390/info10040137.
[14] Zhenghua Hu, Xianmei Wang, Kangming Xu, and Pu Dong. 2020. Real-time Target Tracking Based on PCANet-CSK Algorithm. In Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence (CSAI '19). Association for Computing Machinery, New York, NY, USA, 343–346. https://doi.org/10.1145/3374587.3374607.
[15] Medication Recommendation System Based on Natural Language Processing for Patient Emotion Analysis. (2024). Academic Journal of Science and Technology, 10(1), 62-68. https://doi.org/10.54097/v160aa61
[16] Li, X., Zheng, H., Chen, J., Zong, Y., & Yu, L. (2024). User Interaction Interface Design and Innovation Based on Artificial Intelligence Technology. Journal of Theory and Practice of Engineering Science, 4(03), 1-8.
[17] Xu, K., Zhou, H., Zheng, H., Zhu, M., & Xin, Q. (2024). Intelligent Classification and Personalized Recommendation of E-commerce Products Based on Machine Learning. arXiv preprint arXiv:2403.19345.
[18] Zhou, H., Xu, K., Bao, Q., Lou, Y., & Qian, W. (2024). Application of Conversational Intelligent Reporting System Based on Artificial Intelligence and Large Language Models. Journal of Theory and Practice of Engineering Science, 4(03), 176-182.
[19] Li, L., Xu, K., Zhou, H., & Wang, Y. (2024). Independent Grouped Information Expert Model: A Personalized Recommendation Algorithm Based on Deep Learning.

[20] Che, C., Lin, Q., Zhao, X., Huang, J., & Yu, L. (2023, September). Enhancing Multimodal Understanding with CLIP-Based Image-to-Text Transformation. In Proceedings of the 2023 6th International Conference on Big Data Technologies (pp. 414-418).
[21] Huang, Zengyi, et al. "Research on Generative Artificial Intelligence for Virtual Financial Robo-Advisor." Academic Journal of Science and Technology 10.1 (2024): 74-80.
[22] Chen, J. (2022). The Reform of School Education and Teaching Under the “Double Reduction” Policy. Scientific and Social Research, 4(2), 42-45. (Feb 2022)
[23] Huang, Zengyi, et al. "Application of Machine Learning-Based K-Means Clustering for Financial Fraud Detection." Academic Journal of Science and Technology 10.1 (2024): 33-39.
[24] Yu, H., Huo, S., Zhu, M., Gong, Y., & Xiang, Y. (2024). Machine Learning-Based Vehicle Intention Trajectory Recognition and Prediction for Autonomous Driving. arXiv preprint arXiv:2402.16036.
[25] He, Z., Shen, X., Zhou, Y., & Wang, Y. Application of K-means clustering based on artificial intelligence in gene statistics of biological information engineering.
[26] Wu, Y., Jin, Z., Shi, C., Liang, P., & Zhan, T. (2024). Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis. arXiv preprint arXiv:2403.08217.
[27] Shi, C., Liang, P., Wu, Y., Zhan, T., & Jin, Z. (2024). Maximizing User Experience with LLMOps-Driven Personalized Recommendation Systems. arXiv preprint arXiv:2404.00903.
[28] Wu, C., Cui, J., Xu, X., & Song, D. (2023). The influence of virtual environment on thermal perception: physical reaction and subjective thermal perception on outdoor scenarios in virtual reality. International Journal of Biometeorology, 67(8), 1291-1301.
[29] Huang, Zengyi, et al. "Application of Machine Learning-Based K-Means Clustering for Financial Fraud Detection." Academic Journal of Science and Technology 10.1 (2024): 33-39.
[30] Du, S., Qian, W., Zhang, Y., Shen, Z., & Zhu, M. (2024, February). IMPROVING SCIENCE QUESTION RANKING WITH MODEL AND RETRIEVAL-AUGMENTED GENERATION. In The 6th International scientific and practical conference “Old and new technologies of learning development in modern conditions”(February 13-16, 2024) Berlin, Germany. International Science Group. 2024. 345 p. (p. 252).