International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN (Online): 2347-5552, Volume-12, Issue-2, March 2024
https://doi.org/10.55524/ijircst.2024.12.6.12
Article ID IJIRD-1363, Pages 89-94
http://www.ijircst.org

Optimizing Supply Chain Demand Forecasting and Inventory Management Using Large Language Models

Tianyu Lu[1], Emily Garcia[2], and Jackson Lee[3]

[1] Computer Science, Northeastern University, MA, USA
[2] Business Administration California State University, Los Angeles, CSULA, USA
[3] Computer Technology, Loyola Marymount University, LMU, USA

Correspondence should be addressed to Tianyu Lu; rexcarry036@gmail.com

Received 26 October 2024; Revised 10 November 2024; Accepted 25 November 2024

ABSTRACT- This paper explores the potential for optimizing the Internet of Things by integrating machine learning and computer vision technologies and its implications for U.S. national security and economic competitiveness. First, the application of machine learning in IoT device optimization is introduced, emphasizing its ability to improve system intelligence and efficiency. Secondly, the critical role of computer vision technology in monitoring and reacting to changes in the physical environment is discussed, especially in security applications, such as the protection of national infrastructure and border security. Finally, the strategic significance of integrating these technologies in national security strategy and economic development is analyzed, and the direction and challenge of future research are put forward.

.

KEYWORDS- Machine learning; Computer Vision; Internet of Things Optimization; American National Security

I. INTRODUCTION

Traditionally, it has been difficult for enterprises to detect anomalies in their supply chain management promptly, such as demand fluctuations, supply disruptions, transport anomalies, or stock outages, which can severely impact the business. However, this situation is changing fundamentally with the widespread application of artificial intelligence (AI) algorithms, especially with the intervention of large models. [1]AI algorithms can quickly identify abnormal patterns in the data and issue immediate warnings, enabling enterprises to react and adjust strategies promptly to ensure the continuous and stable operation of the supply chain. This real-time analysis and early warning mechanism enhances the supply chain's resilience and brings more business opportunities and growth potential for enterprises. Some enterprises rely on the excellent forecasting ability of AI algorithms to predict future market demand changes and supply fluctuations accurately. [2-3]This foresight allows companies to develop adaptive strategies in advance to effectively adjust inventory levels, significantly reducing inventory holding costs and avoiding economic losses due to excess or shortage. In addition, large AI models also show outstanding capabilities in intelligent scheduling and route planning. Using deep learning and complex algorithms, they can analyze multiple transportation factors such as traffic flow, road conditions, and weather in real-time to optimize a vehicle's routing [4]. This intelligent route planning significantly reduces transportation time and costs and dramatically improves the accuracy and efficiency of distribution. As a result, companies can deliver products to customers more quickly and cost-effectively, enhancing market competitiveness. Therefore, this paper is based on the increasing complexity of modern supply chains, covering multiple levels of suppliers, customers, and service providers. The widespread application of large language models has enabled partial automation in decision-making and has led to significant efficiency gains and cost reductions in many industries. However, some automated processes still require the involvement of business operators to understand and explain decisions, provide "what if" analysis, and fully interact with program managers, data scientists, and system optimization engineers.

II. RELATED WORK

A. Supply Chain Demand Forecasting

Supply chain demand forecasting is critical in enterprise operations related to production plans, inventory levels, and customer satisfaction. [5]This article will give you an in-depth analysis of the primary supply chain demand forecasting methods to help you improve forecasting accuracy and optimize business operations. Demand forecasting plays a crucial role in enterprise operation, affecting production planning and raw material purchase arrangement and directly relating to inventory control, production efficiency, and customer satisfaction. However, the accuracy of forecasts is often challenged because market changes and complexity usually make the forecast results unsatisfactory. [6]In general, predicting the demand for product categories is more accurate than for individual products, and the accuracy of short-term forecasts is also significantly higher than long-term forecasts. When faced with the uncertainty of a forecast, companies often take various approaches to improve accuracy. The forecasting methods are mainly divided into two categories: qualitative forecasting and quantitative forecasting. Qualitative forecasting relies on experience and intuition and is suitable

for situations that lack historical data, such as new products or long-term strategic plans. Such methods include the historical analogy method, expert opinion method, and Delphi method[7]. In contrast, quantitative forecasting is based on mathematical models and statistical analysis, using historical data to predict future demand, such as time series analysis and regression analysis. In practical applications, historical analogy makes predictions by comparing the development trend of similar events, such as predicting the market reaction to the sales of similar products before launching a new product. The Law of expert opinion is evaluated and corrected by assembling a team, using the views and analysis of in-house professionals, combined with the results of past predictions. The Delphi method collects expert opinions anonymously and finally reaches a consensus prediction result through multiple feedbacks and corrections, improving the prediction's reliability and accuracy. Companies can take several key steps to improve the accuracy of forecasts [8]. The first is to collect comprehensive and accurate historical data to provide reliable support for forecasting models. The second is to choose the appropriate forecasting method according to the actual situation of the enterprise to avoid unthinkingly following the trend and misjudgement. At the same time, strengthening communication and cooperation with upstream and downstream enterprises in the supply chain to jointly cope with market changes is also an effective way to improve forecasting accuracy [9-10]. Finally, regular evaluation and adjustment of forecasting results and models and continuous optimization of forecasting strategies can help enterprises better respond to the dynamic market environment and changes in customer demand.

B. Supply Chain Inventory Management

Since the middle and late 20th century, human society began to enter the era of post-industrial society and move forward to the society of knowledge economy [11]. Enterprises are facing a new competitive environment: technological progress is getting faster and faster, the market and lab or competition is global, the requirements of users are becoming more and more demanding, and the research and development of products are becoming more and more difficult. Achieving low-cost operations based on quickly and effectively meeting customer needs is a complex problem many enterprises face. Under this background, inventory control has become a hot research object in management [12]. Under this background, inventory management has become a research hotspot in management circles. Inventory is the link between each member in the supply chain, and inventory control and management is an integral part of the entire supply chain management. Node enterprises on the supply chain, starting from the supply of raw materials through the processing, assembly, distribution, and other processes of different chain enterprises, deliver products to the hands of final customers[13]. To meet customer demand promptly, avoid stock shortages, or deal with supply chain uncertainties, enterprises must have a certain amount of inventory. The purpose of inventory management is to control the inventory level of enterprises under the premise of maintaining high customer service, reducing the inventory level as much as possible, reducing enterprises' cost burden, and improving enterprises' market competitiveness. Therefore, an optimal inventory strategy is sought to reduce the inventory level of the entire supply chain, reduce the inventory cost of each node, and achieve the purpose of obtaining individual benefits from the overall benefits. Inventory management is one of the most critical operations in manufacturing enterprises. The most significant difference between the manufacturing industry and the circulation enterprise is that the manufacturing industry takes "production" or "manufacturing" as the main body[14]. Manufacturing production is the conversion of tangible inputs into tangible outputs through physical or chemical processes, thereby increasing the added value of products. Inventory management is an essential part of the resource mobilization of manufacturing enterprises because resources and materials occupy the most significant proportion of funds, and other machinery and equipment, plant, and workforce impact profits are not as good as those of materials. The organizational mission of manufacturing enterprises is to produce products that can satisfy customers according to market demand; if the inventory link loses the proper function, it will be unable to promptly supply the "quality" and "quantity" materials required for production. The imbalance of the two links of sales, production, and marketing goals cannot be achieved[15].Inventory costs account for the most significant proportion of manufacturing costs and impact profits the most. The production of manufacturing enterprises is the first input cost and then transformed into finished products to make profits, so the role of material cost is no small matter. [16]In general, the assembly and processing industries, whose material costs often account for more than half of the total manufacturing costs of a small number of capital-intensive or technology-intensive sectors. Therefore, regarding the principle of focus management, material cost should be the focus of management, and the impact on costs and profits is also the key to the success or failure of the most significant business. Inventory control cannot be timely, appropriate, and suitable for material supply, which will seriously reduce the productivity of enterprises. Productivity is the central idea of manufacturing. Manufacturing enterprises must use management methods and skills, talents, equipment, and materials that can be adequately and effectively used to improve performance and create profits.[17-19]Inventory is a significant financial burden for enterprises. For general manufacturing enterprises, inventory usually accounts for a high ratio of the total assets of enterprises, which is not a small burden on the capital of enterprises. Even if there is no interest burden on the capital accumulated by the manufacturing enterprise for inventory, it will lose its due opportunity benefit from the economic point of view. If the inventory can be reduced or accelerated, inventory turnover can significantly reduce the manufacturing cost of enterprises.

C. Application of Large Language Model In Supply Chain

Large Language model (LLM) is a large deep learning model pre-trained on massive data[20]. The underlying architecture, Transformer, consists of a collection of neural networks containing encoders and decoders with self-attention mechanisms. Encoders and decoders extract meaning from text sequences and understand the relationships between words and phrases within the text. (See figure 1)

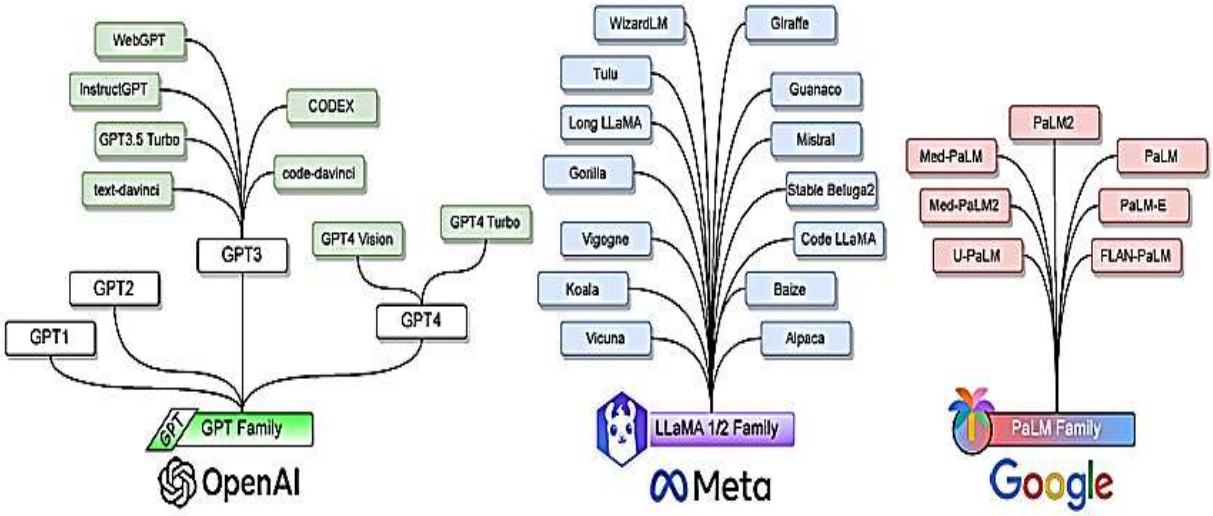

Figure 1: Significant language Model Network Architecture

D. Application Scenarios Of Extensive Language Model Optimization In The Supply Chain

Demand forecasting: Predicting future product demand is the key to supply chain management. Large language models can analyze historical sales data, market trends, seasonal changes, and other factors to help predict future demand. Accurate demand forecasting can reduce inventory costs, excess or shortage, and improve supply chain efficiency.[21-22]Supplier evaluation: The large language model can analyze the historical performance, reputation, delivery time, price, and other supplier factors to help the enterprise choose the right supplier. In addition, it can monitor the real-time performance of suppliers, identify potential problems, and make adjustments in time. Warehousing and logistics optimization: [23]The large language model can analyze the layout of logistics networks and storage facilities and make optimization suggestions. It can also adjust inventory distribution, shipping routes, and delivery times based on real-time demand to reduce shipping costs and improve customer satisfaction.

III. PRACTICAL LARGE-SCALE LANGUAGE MODEL APPLICATION CASES IN SUPPLY CHAIN

The paddler may want to know why a need is not being met. They often bring the issue to the project manager, who must involve domain experts, such as data scientists or engineers developing optimized systems. The domain expert may then need to write additional code and often re-run the optimization to extract the relevant insights. The entire process can be very time-consuming for all parties involved and cause significant delays in the decision-making process.In some applications, teams maintain custom tools, allowing decision-makers to make certain decisions. For example, application dashboards can provide visibility or even allow enforcement of certain actions (for example, fixing a specific vendor for a requirement). [24]However, due to the engineering overhead of maintaining these tools, they are usually limited to the most common use cases.

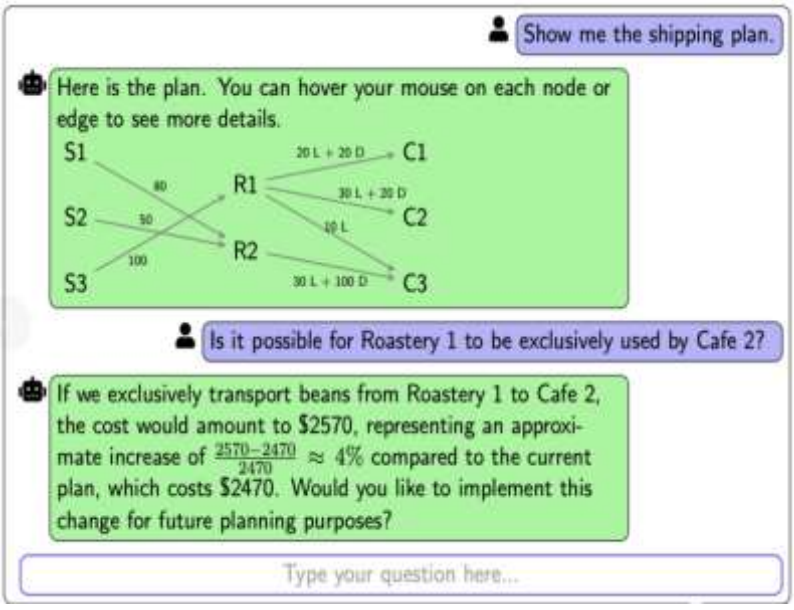

Figure 2: OptiGuide Interfac

A) Supply Chain Optimization

Now, describe a simple supply chain example that is helpful to illustrate our approach. Supply chain. Consider a coffee roasting company that roasts two types of coffee (light and dark). The company sources green coffee beans from three suppliers, roasts them at one of its two roasting facilities, and then ships them to its three retail locations for customer sale. The goal is to meet the needs of each retail location while minimizing total costs. The total price includes purchasing the coffee from the supplier, roasting at each facility, and shipping the final product to the retail location. Figure 3 shows an illustration.The model statement. We can model this problem as a mixed integer programming problem. Let x represent the number of units purchased from suppliers for roaster r, and y represent the quantities of light and dark roasters shipped from roaster r to retail location l, respectively[25-26]. Each supplier has a capacity C, and each retail location l has demand D for light and dark roast, respectively. The cost per unit shipped from suppliers to roasting facility r is c, the freight per unit shipped from roasting facility r to retail location l is g, and the roasting cost per unit of light roast coffee and dark roast coffee in facility r is h, respectively. The optimization problem is as follows:

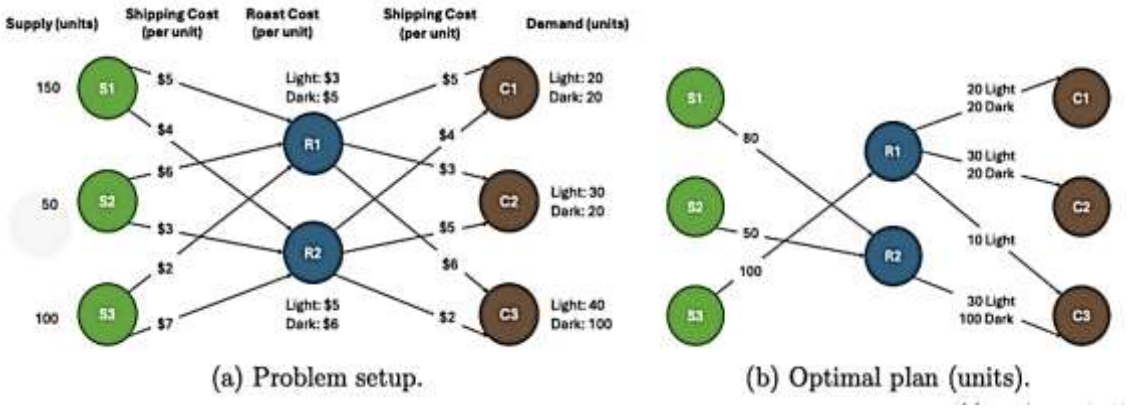

Figure 3: A Simple Supply Chain Example: Coffee Roasting Company

Interpretability. Now, let's look at the example in Figure 3. The optimal solution is shown in Figure 3b. In the optimal plan, both roasters produce light and dark roasters. The raw materials for the first baking facility come from supplier 3, while the second comes from suppliers 1 and 2. The first two retail locations then get all their coffee from the first roaster, while two roasters supply the third retail location.Due to the rapid growth of the cloud computing industry, cloud providers constantly deploy additional capabilities to meet demand. This is achieved by acquiring clusters of new servers and deploying them in the data center. Azure's supply chain covers various processes, including demand forecasting, strategic foresight, hardware semantic search, fulfilment planning, and document management. Due to complexity and scale, Azure supply chain optimization is assigned to different subsystems. We focus on an Intelligent Fulfillment System (IFS) subsystem responsible for distributing and shipping servers from the warehouse to the data center.

B. Supply Chain Optimization Decision

The main decision. For each cloud capacity requirement, significant decisions include (i) the hardware vendor to meet the demand,(ii) the cluster's delivery schedule - precisely the cluster's docking date (which determines the date of shipment from the warehouse), and (iii) where the cluster is deployed in the data center (choosing the row of servers to place the cluster). [27-28]The goal is to minimize the total cost of multiple components, such as being early or late compared to the cluster's ideal docking date and shipping cost, while complying with many constraints. Constraints include vendor and data center capacity constraints, location preference constraints for requirements, and compatibility. The underlying optimization problem is formulated as a mixed integer programming problem with a total input data size of about 500 MB. The optimal solution is obtained every hour using Gurobi. See Appendix A for more details on optimization issues. Stakeholders. The primary users of IFS are planners. They have business background knowledge, so when they receive an optimized result, they can confirm that it meets the business needs (or override decisions) and ensure that decisions are executed as planned. However, due to the increasing complexity of the underlying optimization problem and the global scale of decision-making (hundreds of data centers), there needs to be more clarity of reasoning for each decision. As a result, planners often contact engineers (including data scientists) who develop optimized systems to gain additional insights. Planners and engineers frequently engage in multiple rounds of interaction to understand the problem or explore the scenario.

IV. CONCLUSION

Despite the growing use of artificial intelligence in art and design, it only incompletely replaces mortal generators but plays a reciprocal and enhanced part. In the future, mortal-machine collaboration will come the mainstream mode of cultural design invention. In this cooperative relationship, the AI system gives full play to its advantages of literacy and processing massive data to give material accumulation, rule mining, and creative generation support for creation. At the same time, contrivers calculate on their professional experience and aesthetic judgment to screen, optimize, and ameliorate the workshop generated by artificial intelligence and fit particular independent style and emotional

experience. The man-machine combination is realized through innovative generalities. In visual design, contrivers can use artificial intelligence systems to snappily and efficiently complete original creation and also concentrate on high- position creative design and heightening to achieve complementarity and community between the two sides. In terms of interactive design, artificial intelligence can simulate and analyze user behavior data, propose interactive logic and interface layout suggestions consistent with user habits, and suggest designers combine them with the overall product design concept to optimize transformation and innovation. In conclusion, the innovative operation of AI in interactive product design has significantly bettered design effectiveness and stoner experience. Traditional design is limited to a single and one- sided, and the preface of artificial intelligence technology makes the design process more diversified and intertwined. By integrating product, consumption, and distribution, platforms like Netflix show how AI can transfigure design from functional to applicable and from decentralized to unified to maximize stoner requirements. In the future, with the farther development of digital media technology, the country and assiduity's emphasis on interactive product design and stoner experience will drive further invention. Artificial intelligence will continue to change design generalities and practices and profoundly affect the stoner's commerce and perceived moxie in the digital media terrain, driving the advancement and operation of digital media technologies.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] L. Li, Y. Zhang, J. Wang, and X. Ke, "Deep learning-based network traffic anomaly detection: A study in IoT environments," 2024. Available from: https://doi.org/10.53469/wjimt.2024.07(06).03

[2] G. Cao, Y. Zhang, Q. Lou, and G. Wang, "Optimization of high-frequency trading strategies using deep reinforcement learning," *Journal of Artificial Intelligence General Science* (JAIGS), vol. 6, no. 1, pp. 230–257, 2024, ISSN: 3006-4023. Available from: https://doi.org/10.60087/jaigs.v6i1.247

[3] G. Wang, X. Ni, Q. Shen, and M. Yang, "Leveraging large language models for context-aware product discovery in e-commerce search systems," *Journal of Knowledge Learning and Science Technology*, vol. 3, no. 4, 2024, ISSN: 2959-6386 (online). Available from: https://doi.org/10.60087/jklst.v3.n4.p300

[4] H. Zhang, et al., "Enhancing facial micro-expression recognition in low-light conditions using attention-guided deep learning," *Journal of Economic Theory and Business Management*, vol. 1, no. 5, pp. 12–22, 2024. Available from: https://doi.org/10.5281/zenodo.13933725

[5] J. Wang, T. Lu, L. Li, and D. Huang, "Enhancing personalized search with AI: A hybrid approach integrating deep learning and cloud computing," *International Journal of Innovative Research in Computer Science & Technology*, vol. 12, no. 5, pp. 127–138, 2024. Available from: https://doi.org/10.5281/zenodo.13998900

[6] S. Zhou, W. Zheng, Y. Xu, and Y. Liu, "Enhancing user experience in VR environments through AI-driven adaptive UI design," *Journal of Artificial Intelligence General Science* (JAIGS), vol. 6, no. 1, pp. 59–82, 2024. Available from: https://doi.org/10.60087/jaigs.v6i1.230

[7] M. Yang, D. Huang, H. Zhang, and W. Zheng, "AI-enabled precision medicine: Optimizing treatment strategies through genomic data analysis," *Journal of Computer Technology and Applied Mathematics*, vol. 1, no. 3, pp. 73–84, 2024. Available from: https://doi.org/10.5281/zenodo.13380619

[8] X. Wen, Q. Shen, W. Zheng, and H. Zhang, "AI-driven solar energy generation and smart grid integration: A holistic approach to enhancing renewable energy efficiency," *International Journal of Innovative Research in Engineering and Management*, vol. 11, no. 4, pp. 55–66, 2024. Available from: https://doi.org/10.55524/ijirem.2024.11.4.8

[9] S. Zhou, B. Yuan, K. Xu, M. Zhang, and W. Zheng, "The impact of pricing schemes on cloud computing and distributed systems," *Journal of Knowledge Learning and Science Technology*, vol. 3, no. 3, pp. 193–205, 2024. Available from: https://doi.org/10.60087/jklst.v3.n3.p206-224

[10] Y. Zhang, W. Bi, and R. Song, "Research on deep learning-based authentication methods for e-signature verification in financial documents," *Academic Journal of Sociology and Management*, vol. 2, no. 6, pp. 35–43, 2024. Available from: https://doi.org/10.5281/zenodo.14161744

[11] Z. Zhou, S. Xia, M. Shu, and H. Zhou, "Fine-grained abnormality detection and natural language description of medical CT images using large language models," *International Journal of Innovative Research in Computer Science & Technology*, vol. 12, no. 6, pp. 52–62, 2024. Available from: https://doi.org/10.55524/ijircst.2024.12.6.8

[12] Y. Zhang, Y. Liu, and S. Zheng, "A graph neural network-based approach for detecting fraudulent small-value high-frequency accounting transactions," Academic Journal of Sociology and Management, *vol. 2, no. 6, pp. 25–34, 2024. Available from:* https://doi.org/10.5281/zenodo.14161459

[13] S. Huang, Y. Liang, F. Shen, and F. Gao, "Research on federated learning's contribution to trustworthy and responsible artificial intelligence," in *Proceedings of the 2024 3rd International Symposium on Robotics, Artificial Intelligence and Information Engineering*, July 2024, pp. 125–129. Available from: https://doi.org/10.1145/3689299.3689322

[14] K. Yu, Q. Shen, Q. Lou, Y. Zhang, and X. Ni, "A deep reinforcement learning approach to enhancing liquidity in the US municipal bond market: An intelligent agent-based trading system," *International Journal of Engineering and Management Research*, vol. 14, no. 5, pp. 113–126, 2024. Available from: https://doi.org/10.5281/zenodo.14184756

[15] Y. Wang, Y. Zhou, H. Ji, Z. He, and X. Shen, "Construction and application of artificial intelligence crowdsourcing map based on multi-track GPS data," in *2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE)*, IEEE, March 2024, pp. 1425–1429. Available from: https://doi.org/10.1109/ICAACE61206.2024.10548953

[16] A. Akbar, N. Peoples, H. Xie, P. Sergot, H. Hussein, W. F. Peacock IV, and Z. Rafique, "Thrombolytic administration for acute ischemic stroke: What processes can be optimized?" *McGill Journal of Medicine*, vol. 20, no. 2, 2022. Available from: https://doi.org/10.26443/mjm.v20i2.881

[17] Z. Feng, M. Ge, and Q. Meng, "Enhancing energy efficiency in green buildings through artificial intelligence," *Applied Science and Engineering Journal for Advanced Research*, vol. 3, no. 5, pp. 10–17, 2024. Available from: https://www.preprints.org/manuscript/202408.1489

[18] J. Chen, J. Xiao, and W. Xu, "A hybrid stacking method for short-term price forecasting in electricity trading market," in *2024 8th International Conference on*

Information Technology, Information Systems and Electrical Engineering (ICITISEE), IEEE, 2024, pp. 1–5. Available from: https://doi.org/10.1109/ICITISEE63424.2024.10730623

[19] Y. Zhang, H. Xie, S. Zhuang, and X. Zhan, "Image processing and optimization using deep learning-based generative adversarial networks (GANs)," *Journal of Artificial Intelligence General Science* (JAIGS), vol. 5, no. 1, pp. 50–62, 2024. Available from: https://doi.org/10.60087/jaigs.v5i1.163

[20] T. Lu, M. Jin, M. Yang, and D. Huang, "Deep learning-based prediction of critical parameters in CHO cell culture process and its application in monoclonal antibody production," *International Journal of Advance in Applied Science Research*, vol. 3, pp. 108–123, 2024. Available from: https://h-tsp.com/index.php/ijaasr/article/view/69

[21] S. Xia, Y. Zhu, S. Zheng, T. Lu, and X. Ke, "A deep learning-based model for P2P microloan default risk prediction," *International Journal of Innovative Research in Engineering and Management*, vol. 11, no. 5, pp. 110–120, 2024. Available from: https://doi.org/10.55524/ijirem.2024.11.5.16

[22] J. Xiao, T. Deng, and S. Bi, "Comparative analysis of LSTM, GRU, and transformer models for stock price prediction," *arXiv preprint*, arXiv:2411.05790, 2024. Available from: https://doi.org/10.48550/arXiv.2411.05790

[23] W. Zheng, M. Yang, D. Huang, and M. Jin, "A deep learning approach for optimizing monoclonal antibody production process parameters," *International Journal of Innovative Research in Computer Science & Technology*, vol. 12, no. 6, pp. 18–29, 2024. Available from: https://doi.org/10.55524/ijircst.2024.12.6.4

[24] X. Ma, J. Wang, X. Ni, and J. Shi, "Machine learning approaches for enhancing customer retention and sales forecasting in the biopharmaceutical industry: A case study," *International Journal of Engineering and Management Research*, vol. 14, no. 5, pp. 58–75, 2024. Available from: https://doi.org/10.5281/zenodo.14053620

[25] S. Huang, S. Diao, H. Zhao, and L. Xu, f federated learning to AI development," in *The 24th International Scientific and Practical Conference 'Technologies of Scientists and Implementation of Modern Methods'*, June 2024, pp. 358. Available from: http://dx.doi.org/10.20944/preprints202407.0551.v1

[26] H. Zheng, J. Wu, R. Song, L. Guo, and Z. Xu, "Predicting financial enterprise stocks and economic data trends using machine learning time series analysis," *Applied and Computational Engineering*, vol. 87, pp. 26–32, 2024. Available from: https://www.preprints.org/manuscript/202407.0895

[27] H. Zheng, K. Xu, M. Zhang, H. Tan, and H. Li, "Efficient resource allocation in cloud computing environments using AI-driven predictive analytics," *Applied and Computational Engineering*, vol. 82, pp. 6–12, 2024. Available from: https://doi.org/10.54254/2755-2721/82/2024GLG0055

[28] B. Wang, H. Zheng, K. Qian, X. Zhan, and J. Wang, "Edge computing and AI-driven intelligent traffic monitoring and optimization," *Applied and Computational Engineering*, vol. 77, pp. 225–230, 2024. Available from: https://doi.org/10.54254/2755-2721/77/2024MA0062