Article

Analysis of Large Language Model Applications in Public Data Development and Utilization

Michael Reynolds *

Posted Date: 12 May 2025

doi: 10.20944/preprints202505.0752.v1

Keywords: public data; large language models; data cleaning; data insights; privacy protection

 doi:10.20944/preprints202505.0752.v1

Article

Analysis of Large Language Model Applications in Public Data Development and Utilization

Michael J. Reynolds

University of Michigan, USA; mjreynolds0203@gmail.com; https://orcid.org/0009-0008-2881-2859

Abstract: As data-driven decision-making becomes increasingly important, the development and utilization of public data play a critical role in social governance and economic growth. However, public data often suffer from issues such as large scale, structural complexity, and inconsistent quality, which traditional data processing methods struggle to address effectively. In recent years, breakthroughs in large language models (LLMs) within the field of natural language processing have introduced new opportunities for processing and analyzing public data. This paper explores key aspects of public data development and utilization, focusing on typical application scenarios of LLMs in data cleaning, insight extraction, privacy protection, and compliance review. Additionally, it analyzes the technical limitations and ethical challenges associated with these models. Based on this analysis, the paper proposes suggestions for optimizing model capabilities, reducing resource costs, and establishing standardized application frameworks. Lastly, it anticipates the development prospects of LLMs in multimodal data processing and cross-domain collaborative analysis, aiming to provide theoretical support and practical guidance for applying LLMs in public data development.

Keywords: public data; large language models; data cleaning; data insights; privacy protection

1. Introduction

Public data, as a vital resource in the digital age, has become a key driving force for optimizing social governance and promoting high-quality economic development. With the widespread application of data-driven decision-making in government management, public services, and business operations, public data development and utilization not only enhance information transparency and efficient resource allocation but also provide a rich foundation for technological innovation [1]. However, public data often face significant challenges, including large scale, diverse types, structural complexity, and inconsistent quality, which hinder effective development and utilization [2]. Traditional data processing techniques exhibit limitations such as low efficiency and lack of flexibility, necessitating the adoption of new technological approaches for supplementation and optimization [3]. In recent years, large language models (LLMs) have made groundbreaking advances in natural language processing. Leveraging deep learning technology, these models demonstrate robust capabilities in processing unstructured text, generating language content, and extracting informational patterns. Particularly in data cleaning, knowledge extraction, and privacy protection, LLMs exhibit significant advantages, offering a novel technological pathway for public data development and utilization [4]. However, applying LLMs to the public data domain also presents challenges, such as accuracy and stability issues, high computational resource consumption, and potential risks related to data privacy and ethics [5]. Therefore, an in-depth analysis of LLM applications in public data development and utilization is necessary, along with an exploration of possible technical optimizations and standardized application frameworks. This paper aims to study the potential and practical value of LLM applications in public data development and utilization. By analyzing typical scenarios in key aspects such as data cleaning, insight extraction, and compliance review, it identifies the technological strengths and shortcomings of LLMs. Furthermore, it proposes corresponding optimization strategies and application frameworks based on existing technological limitations and societal needs [6]. This study aspires not only to provide theoretical support for applying LLMs in the public data domain but also to offer practical guidance for policymakers and

 doi:10.20944/preprints202505.0752.v1

technology developers in their operations. Through this research, we seek to uncover future development directions for LLMs in public data development and establish a solid theoretical and practical foundation for their broader application. Haosen et al. [7] developed RPF-ELD for breast cancer diagnosis in US. Min et al. [8] developed a DeepFM model for loan repayment prediction. Tangtang et al. [9] studied ARIMA and LSTM for US electricity price forecasts.

2. Current Status of Public Data Development and Utilization

Public data, accumulated by governments, institutions, and society during production activities, is an important resource for social governance and economic growth due to its openness and shareability. Driven by digital transformation, the development and utilization of public data have been widely applied in fields such as smart city construction, public health monitoring, and environmental protection. However, despite its immense potential, public data development and utilization face numerous challenges that warrant thorough analysis from both theoretical and practical perspectives. Public data development has made significant progress in recent years [10]. On the policy front, many countries and regions have introduced regulations and guidelines related to data openness, such as the European Union's General Data Protection Regulation (GDPR) and China's Data Security Law, to regulate the openness and sharing of public data. On the technical front, advancements in big data analytics and cloud computing platforms have significantly improved the efficiency of public data processing, making data-driven decision-making increasingly mainstream [11]. Moreover, public data development has facilitated the implementation of emerging technologies, such as optimizing medical resource allocation through artificial intelligence and constructing intelligent transportation systems using geospatial data, thereby providing robust support for improving public services. However, the practical application of public data development and utilization still faces a range of challenges. Data quality issues represent a major obstacle. Many public datasets lack uniform standards during collection, leading to inconsistent formats, missing information, and frequent noise interference [12]. Additionally, privacy and security concerns pose severe challenges to the openness and utilization of public data. Public datasets often contain sensitive information, making it imperative to balance privacy protection with maximizing data utility. Furthermore, the management and sharing mechanisms for public data remain underdeveloped, with data silos between different departments and institutions severely limiting cross-domain collaborative data development [13]. In summary, while the development and utilization of public data have shown promising prospects, their full realization requires overcoming challenges related to data quality, privacy protection, and sharing mechanisms [14]. These issues not only restrict the potential of public data but also present opportunities to explore new technological methods and application models [15]. As an advanced technological tool, LLMs hold the promise of playing a significant role in this domain, warranting in-depth exploration of their potential to address these challenges.

3. Application Scenarios of Large Language Models in Public Data

As LLMs rapidly advance in natural language processing, their powerful capabilities in semantic understanding, text generation, and information extraction offer innovative solutions for public data development and utilization [16]. In the public data domain, LLMs find applications primarily in data cleaning and structuring, insight extraction and decision support, as well as privacy protection and compliance review.

3.1. Data Cleaning and Structuring

Public data's unstructured and fragmented characteristics pose significant challenges to its development and utilization [17]. These data types include free text, scanned documents, voice recordings, and multilingual content, often lacking standardized formats and conventions. Traditional cleaning methods rely heavily on manual intervention or complex rule-based systems,

 doi:10.20944/preprints202505.0752.v1

which are inefficient and prone to human error. LLMs, with their robust semantic understanding capabilities, offer an efficient technical pathway to address these issues [18]. LLMs can automatically extract key information from unstructured text and convert it into structured data. This capability is particularly valuable in processing government reports, legal documents, and news data [19]. For example, by analyzing policy documents, LLMs can automatically identify and extract policy highlights, categorizing and storing them in structured databases for subsequent querying and analysis [20]. This automation not only significantly enhances data cleaning efficiency but also minimizes errors associated with manual operations. Furthermore, LLMs demonstrate remarkable advantages in detecting and correcting data errors. Their powerful semantic analysis capabilities enable them to identify spelling errors, inconsistent formats, and redundant or duplicate data in public databases. For instance, when processing census data, LLMs can automatically correct typos, standardize date and address formats, and eliminate duplicate records, thereby improving data integrity and consistency. These capabilities lay a solid foundation for constructing high-quality public datasets and provide reliable support for subsequent data mining and analysis [21]. In the healthcare sector, the application of LLMs is particularly prominent. For example, patient medical records are often recorded in unstructured text, containing complex descriptions [22]. LLMs can analyze the content of medical records, extracting key data such as core symptoms, diagnoses, and treatment recommendations, and convert them into standardized tabular data. This automation significantly reduces manual intervention and enhances data consistency and accuracy, providing crucial data support for disease prediction, treatment optimization, and healthcare resource allocation [23]. LLMs also excel in multilingual scenarios. In cross-border public health collaborations, health data from different countries and regions may be recorded in various languages. By semantically understanding multilingual texts, LLMs can uniformly extract and standardize these data, facilitating the establishment of a global public health data platform. This functionality is crucial for data sharing and analysis in international collaborations. In conclusion, LLMs in data cleaning and structuring significantly improve processing efficiency and data quality, overcoming the limitations of traditional methods. These capabilities meet current data processing demands while preparing for future large-scale and complex data challenges [24]. By further optimizing models and expanding application scenarios, LLMs are expected to play an even greater role in the public data domain, advancing data-driven social governance and public services.

3.2. Data Insights and Decision Support

The core value of public data lies in its ability to provide profound insights, thereby offering robust support for policy-making and social governance. In this context, large language models (LLMs), with their powerful semantic analysis and natural language generation capabilities, can rapidly extract key trends and potential patterns from complex and extensive datasets [25]. They present these findings in an intuitive and easily understandable manner, significantly enhancing the efficiency and quality of data-driven decision-making [26]. The application of LLMs is particularly prominent in urban traffic management. By analyzing traffic accident reports and real-time monitoring data, LLMs can identify high-risk areas and their potential causes, such as improper traffic signal settings or road design issues [27]. The models can then generate detailed analytical reports, including explanations of the causes and optimization suggestions, providing data support for traffic planning departments. Additionally, LLMs can integrate data from various sources, such as weather conditions, population density, and historical traffic flows, to predict potential future traffic bottlenecks, offering valuable references for long-term traffic management policies [28]. LLMs also play a critical role in natural disaster prediction and response. They can extract patterns from weather data, geological monitoring information, and historical disaster records, identifying high-risk areas and seasonal disaster trends. The models can generate targeted disaster prevention and mitigation reports, which not only include qualitative descriptions of potential risks but also offer quantitative impact assessments. These reports help government agencies prioritize resource allocation and improve disaster preparedness. For instance, in flood disaster prediction, LLMs can

 doi:10.20944/preprints202505.0752.v1

integrate rainfall data and hydrological monitoring information to predict affected regions and populations and provide recommendations for emergency evacuations and resource distribution. Moreover, LLMs can support decision-making processes for enterprises and institutions [29]. In public health, LLMs can analyze the spread of infectious diseases and generate predictive reports based on historical trends, aiding the formulation of public health policies. In energy management, the models can integrate electricity usage records and renewable energy production data to optimize energy distribution and storage strategies. Such insights not only deepen data analysis but also offer new perspectives for solving complex problems [30]. Another notable advantage of LLMs is their ability to generate logical and easy-to-understand natural language descriptions [31]. This capability ensures that the analysis results generated by the models can be directly read and utilized by decision-makers, staff without technical backgrounds, or the general public, avoiding the comprehension barriers caused by technical jargon in traditional data analysis reports [32]. For example, an environmental protection report generated by an LLM can transform complex pollution source tracking data into concise actionable recommendations, supporting environmental policy promotion and public engagement. In summary, the application of LLMs in data insights and decision support not only enables multi-dimensional and multi-level exploration of data value but also delivers critical information efficiently and intuitively [33]. This transformative capability makes LLMs an essential tool in public data development and utilization, with the potential to expand their application scope across various fields.

3.3. Privacy Protection and Compliance Review

Privacy and compliance issues are critical in the development and utilization of public data. LLMs can be trained to identify and mask sensitive information, ensuring the security of data during sharing and application processes [34]. For instance, before the release of open datasets, LLMs can automatically scan the data for potentially sensitive information, such as identification numbers or addresses, and apply masking or encryption to protect privacy. At the same time, LLMs play an important role in compliance reviews [35]. Through semantic analysis of relevant laws and regulations, the models can assess whether data usage practices comply with policy requirements. For example, in scenarios involving the sharing of medical data, LLMs can evaluate whether data processing activities adhere to regulations like the General Data Protection Regulation (GDPR), thereby mitigating legal risks [36]. This capability is crucial for the standardized and lawful use of public data. In conclusion, LLMs demonstrate extensive application prospects in public data development and utilization [37]. From data cleaning and structuring to the automated generation of insights and privacy protection with compliance reviews, their functions cover all critical aspects of public data utilization. These capabilities not only provide technical support for enhancing data development efficiency but also pave the way for addressing long-standing challenges in the public data domain.

4. Challenges and Countermeasures in the Application of Large Language Models

Despite the broad potential of large language models (LLMs) in public data development and utilization, their practical application faces numerous challenges. These challenges are primarily centered around technical limitations, ethical and privacy risks, as well as constraints on resources and costs. To fully realize the advantages of LLMs, it is essential to propose feasible countermeasures to address these issues effectively [38]. From a technical perspective, the accuracy and stability of LLMs may be insufficient when processing complex and diverse public data. For instance, when dealing with data containing ambiguous expressions or cross-domain semantics, the models might produce errors in understanding or fail to extract information comprehensively. Furthermore, the high computational resource demands of running LLMs limit their applicability in resource-constrained environments, and general-purpose models often lack the specificity required for

 doi:10.20944/preprints202505.0752.v1

domain-specific tasks. Addressing these challenges requires enhancing the customization capabilities of LLMs to adapt to specific data needs. Techniques such as transfer learning and fine-tuning can improve the accuracy and adaptability of models. Optimizing resource efficiency is another viable solution, achievable through model compression techniques that reduce hardware requirements and by leveraging distributed and edge computing to offload computational burdens, thereby increasing model usability and operational efficiency. In terms of ethics and privacy, LLMs pose risks such as privacy breaches and biased outcomes. For example, while analyzing public data, LLMs might inadvertently expose sensitive information or generate biased outputs due to inherent biases in the training data. To mitigate these risks, it is crucial to incorporate technologies like differential privacy and federated learning at every stage of data processing to safeguard user data. Additionally, establishing strict ethical review mechanisms is necessary to monitor the training data, outputs, and application scenarios of LLMs. Transparency can be enhanced by publishing accountability reports, fostering public trust in the responsible use of these models. For cross-border data-sharing scenarios, international coordination of legal and ethical standards should be promoted to address conflicts and inconsistencies across jurisdictions.Resource and cost constraints represent another significant challenge. The high operational and maintenance costs of LLMs pose considerable barriers for public institutions and small organizations, particularly due to their heavy reliance on computational and storage resources during training and deployment. To address this issue, fostering collaboration within open-source communities can enable the sharing of pre-trained models and related technical resources, reducing development and deployment costs. Furthermore, lightweight models can be developed to meet the demands of resource-limited scenarios, with optimized algorithm designs minimizing hardware and energy consumption. Strengthening cooperation among public sectors, enterprises, and academic institutions can also help establish joint development and resource-sharing mechanisms, effectively distributing costs and improving resource utilization efficiency.In conclusion, while LLM applications face multifaceted challenges, these can be progressively overcome through technical optimization, enhanced privacy protection mechanisms, and efficient resource utilization strategies. By systematically addressing these challenges, LLMs can play an increasingly significant role in public data development and utilization, providing innovative technological support and solutions for social governance and public services.

5. Prospects and Future Directions

Large language models (LLMs) exhibit promising application prospects in public data development and utilization. Their robust semantic understanding and generative capabilities provide new avenues for data-driven decision-making and intelligent governance. Looking ahead, as technology advances and application scenarios expand, LLMs are poised to unlock greater potential in multimodal data integration, cross-domain collaborative analysis, and intelligent public services.Multimodal data integration represents a key direction for future development. Public data sources are becoming increasingly diverse, encompassing text, images, audio, video, and other formats. Analyzing single-modal data often fails to capture the complexities of real-world problems. The combination of LLMs and multimodal data integration techniques can significantly enhance analytical capabilities. For instance, by integrating urban planning documents, geospatial data, and real-time traffic videos, LLMs can assist city administrators in formulating more precise policies and optimizing resource allocation. This integration not only broadens the dimensions of data but also provides solutions to more complex problems.Cross-domain collaborative analysis is another critical trend. Public data often involves multiple departments and industries, and developing data within a single domain cannot fully exploit its potential value. In the future, LLMs can serve as bridges for cross-domain collaboration by integrating semantics and associating data, facilitating the efficient sharing and collaborative utilization of public data across fields. For example, in health management and environmental protection, LLMs can combine meteorological data with medical data to predict the spread of infectious diseases, providing support for public health policies. This cross-domain collaborative analytical capability will greatly enhance the application value of public data in

 doi:10.20944/preprints202505.0752.v1

complex scenarios.The realization of intelligent public services also relies on the empowerment of LLMs. In the future, automated tools powered by LLMs can achieve real-time data insights and service optimization. For example, in social security systems, LLMs can analyze historical data to predict residents' changing needs and dynamically adjust resource allocation plans. Additionally, intelligent data insight platforms can offer more convenient interactive methods for government departments and the public, improving service efficiency and user experience. In this way, LLMs are expected to become key drivers of digital transformation in public services. However, the future development of LLMs also faces certain challenges. For instance, the uncertainty of technological advancement may introduce new risks, and the privacy protection and ethical frameworks for public data require further improvement. Therefore, while envisioning the future, it is imperative to continue strengthening the standardized development of technology, promoting the formulation of industry standards, and enhancing the anticipation and management of the potential impacts of new technologies. In summary, the prospects for LLMs in public data development and utilization are highly promising. As technology evolves and applications deepen, LLMs will become indispensable tools in the public data domain, providing robust support for intelligent and precise social governance. Through continuous technological innovation and regulatory management, LLMs are expected to create more possibilities in public data development, contributing greater intelligence and power to societal advancement.

6. Conclusion

The development of large language models (LLMs) has introduced new technological pathways and practical approaches to public data development and utilization. This paper examines the characteristics and current state of public data, exploring LLM applications in key aspects such as data cleaning and structuring, insight extraction and decision support, as well as privacy protection and compliance review. It also analyzes the technical challenges, ethical risks, and resource constraints associated with these applications, proposing targeted countermeasures. The research highlights the significant advantages of LLMs in enhancing data processing efficiency, deepening data insights, and ensuring privacy and compliance. Their potential in multimodal data integration, cross-domain collaborative analysis, and intelligent public services demonstrates their importance in advancing public data development from traditional methods to intelligent solutions. However, the widespread application of LLMs also brings technical and societal risks, such as high computational resource requirements, model biases, and privacy concerns, which necessitate optimized technical solutions, strengthened ethical oversight, and improved policy frameworks. In the future, the development of LLMs will further promote the efficient utilization and in-depth exploration of public data. On the technical front, it is essential to continuously enhance the customization capabilities and resource efficiency of models. On the societal front, constructing transparent and standardized usage frameworks is necessary to ensure ethical and legal compliance in their applications. By aligning technological advancements with governance improvements, LLMs are poised to become vital tools in public data development, providing intelligent support for social governance and public services. In conclusion, the application prospects of LLMs in public data development and utilization are highly promising. Through ongoing technological innovation and regulatory development, LLMs will bring greater value to the public data domain, modernizing social governance capabilities and contributing significantly to the construction of intelligent societies. This paper aims to provide theoretical references and practical insights for further exploration and application in this field.

References

1. Zhang J, Xiang A, Cheng Y, et al. Research on Detection of Floating Objects in River and Lake Based on AI Image Recognition [J]. Journal of Artificial Intelligence Practice, 2024, 7(2): 97-106.
2. Wu Z. Deep learning with improved metaheuristic optimization for traffic flow prediction [J]. Journal of Computer Science and Technology Studies, 2024, 6(4): 47-53.

 doi:10.20944/preprints202505.0752.v1

3. Sallam, Malik. "The utility of ChatGPT as an example of large language models in healthcare education, research and practice: Systematic review on the future perspectives and potential limitations." MedRxiv (2023): 2023-02.
4. Rane, Nitin Liladhar, et al. "Contribution and performance of ChatGPT and other Large Language Models (LLM) for scientific and research advancements: a double-edged sword." International Research Journal of Modernization in Engineering Technology and Science 5.10 (2023): 875-899.
5. Xiang A, Huang B, Guo X, et al. A neural matrix decomposition recommender system model based on the multimodal large language model [J]. arXiv preprint arXiv:2407.08942, 2024.
6. Yu Q, Wang S, Tao Y. Enhancing anti-money laundering detection with self-attention graph neural networks [C]//SHS Web of Conferences. EDP Sciences, 2025, 213: 01016.
7. Wang, G. Zhang, Y. Zhao, F. Lai, W. Cui, J. Xue, Q. Wang, H. Zhang, and Y. Lin, "Rpf-eld: Regional prior fusion using early and late distillation for breast cancer recognition in ultrasound images," in 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2024, pp. 2605–2612.
8. Min, Liu, et al. "Financial Prediction Using DeepFM: Loan Repayment with Attention and Hybrid Loss." 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA). IEEE, 2024.
9. Wang T, Cai X, Xu Q. Energy Market Price Forecasting and Financial Technology Risk Management Based on Generative AI [J]. Applied and Computational Engineering, 2024, 100: 29-34.
10. Mo K, Chu L, Zhang X, et al. Dral: Deep reinforcement adaptive learning for multi-uavs navigation in unknown indoor environment [J]. arXiv preprint arXiv:2409.03930, 2024.
11. Tan C, Li X, Wang X, et al. Real-time Video Target Tracking Algorithm Utilizing Convolutional Neural Networks (CNN) [C]//2024 4th International Conference on Electronic Information Engineering and Computer (EIECT). IEEE, 2024: 847-851.
12. Thirunavukarasu, Arun James, et al. "Large language models in medicine." Nature medicine 29.8 (2023): 1930-1940.
13. Son, Jungha, and Boyoung Kim. "Trend Analysis of Large Language Models through a Developer Community: A Focus on Stack Overflow." Information 14.11 (2023): 602.
14. Ma D, Wang M, Xiang A, et al. Transformer-Based Classification Outcome Prediction for Multimodal Stroke Treatment [J]. arXiv preprint arXiv:2404.12634, 2024
15. Li X, Cao H, Zhang Z, et al. Artistic Neural Style Transfer Algorithms with Activation Smoothing [J]. arXiv preprint arXiv:2411.08014, 2024.
16. Guo H, Zhang Y, Chen L, et al. Research on vehicle detection based on improved YOLOv8 network [J]. arXiv preprint arXiv:2501.00300, 2024.
17. Diao, Su, et al. "Ventilator pressure prediction using recurrent neural network." arXiv preprint arXiv:2410.06552 (2024).
18. Cheng Y, Yang Q, Wang L, et al. Research on Credit Risk Early Warning Model of Commercial Banks Based on Neural Network Algorithm [J]. arXiv preprint arXiv:2405.10762, 2024.
19. Carullo, G. "Large Language Models for Transparent and Intelligible AI-Assisted Public Decision-Making." CERIDAP 3 (2023): 1-23.
20. De Angelis, Luigi, et al. "ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health." Frontiers in public health 11 (2023): 1166120.
21. Tang, Xirui, et al. "Research on heterogeneous computation resource allocation based on data-driven method." 2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS). IEEE, 2024.
22. Tan C, Zhang W, Qi Z, et al. Generating Multimodal Images with GAN: Integrating Text, Image, and Style [J]. arXiv preprint arXiv:2501.02167, 2025.
23. Yang H, Wang L, Zhang J, et al. Research on Edge Detection of LiDAR Images Based on Artificial Intelligence Technology [J]. arXiv preprint arXiv:2406.09773, 2024.
24. Xiang A, Qi Z, Wang H, et al. A Multimodal Fusion Network For Student Emotion Recognition Based on Transformer and Tensor Product [J]. arXiv preprint arXiv:2403.08511, 2024.
25. Zhao Y, Hu B, Wang S. Prediction of brent crude oil price based on lstm model under the background of low-carbon transition [J]. arXiv preprint arXiv:2409.12376, 2024.

 doi:10.20944/preprints202505.0752.v1

26. Xiang A, Zhang J, Yang Q, et al. Research on splicing image detection algorithms based on natural image statistical characteristics [J]. arXiv preprint arXiv:2404.16296, 2024.
27. Jeon, Jaeho, and Seongyong Lee. "Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT." Education and Information Technologies 28.12 (2023): 15873-15892.
28. Taloni, Andrea, Vincenzo Scorcia, and Giuseppe Giannaccare. "Large language model advanced data analysis abuse to create a fake data set in medical research." JAMA ophthalmology 141.12 (2023): 1174-1175.
29. Shih K, Han Y, Tan L. Recommendation System in Advertising and Streaming Media: Unsupervised Data Enhancement Sequence Suggestions [J]. arXiv preprint arXiv:2504.08740, 2025.
30. Wu, X., Sun, Y., & Liu, X. (2024). Multi-Class Classification of Breast Cancer Gene Expression Using PCA and XGBoost. Preprints. https://doi.org/10.20944/preprints202410.1775.v1
31. Wu Z, Wang X, Huang S, et al. Research on prediction recommendation system based on improved markov model [J]. Advances in Computer, Signals and Systems, 2024, 8(5): 87-97.
32. Zhang W, Huang J, Wang R, et al. Integration of Mamba and Transformer--MAT for Long-Short Range Time Series Forecasting with Application to Weather Dynamics [J]. arXiv preprint arXiv:2409.08530, 2024.
33. Shi X, Tao Y, Lin S C. Deep Neural Network-Based Prediction of B-Cell Epitopes for SARS-CoV and SARS-CoV-2: Enhancing Vaccine Design through Machine Learning [J]. arXiv preprint arXiv:2412.00109, 2024.
34. Zhao R, Hao Y, Li X. Business Analysis: User Attitude Evaluation and Prediction Based on Hotel User Reviews and Text Mining [J]. arXiv preprint arXiv:2412.16744, 2024.
35. Ziang H, Zhang J, Li L. Framework for lung CT image segmentation based on UNet++ [J]. arXiv preprint arXiv:2501.02428, 2025.
36. Gao, Dawei, et al. "Synaptic resistor circuits based on Al oxide and Ti silicide for concurrent learning and signal processing in artificial intelligence systems." Advanced Materials 35.15 (2023): 2210484.
37. Wu Z. Mpgaan: Effective and efficient heterogeneous information network classification [J]. Journal of Computer Science and Technology Studies, 2024, 6(4): 8-16.
38. Wang L, Cheng Y, Xiang A, et al. Application of Natural Language Processing in Financial Risk Detection [J]. arXiv preprint arXiv:2406.09765, 2024.