Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ28955

Application and Challenges of Artificial Intelligence in Data Analysis: A Case Study of Uber's Ride-sharing Algorithm

Xuru Yang

Guangzhou SCA School, Guangzhou, China
yangxuru123@163.com

Abstract. With the in-depth development of the big data era, artificial intelligence (AI) technology, relying on its powerful data processing and pattern recognition capabilities, has become a core driver for advancing business analysis towards intelligence and automation. This paper aims to systematically explore the specific application models, significant advantages, and potential limitations of AI technology in the field of data analysis. First, from the perspective of technological development, the paper outlines the key technical components of current cutting-edge business analysis and explains the inevitability of AI becoming an indispensable element therein. Second, taking Uber's ride-sharing algorithm as a typical case, it deeply analyzes its core principle—the "Trip Chain"—and working mechanism, revealing how AI achieves accurate resource matching and route planning in a highly dynamic environment to improve overall system efficiency and user experience. Finally, the paper dialectically analyzes the advantages and disadvantages of AI applications and concludes that AI technology significantly enhances the scientific nature and intelligence level of business decisions by improving the efficiency, scale, and depth of data analysis. However, its widespread application still needs to address multiple constraints in terms of technology itself and social acceptance. Technology is a lever, while humans remain the fulcrum. The future of AI-driven data analysis belongs to practitioners who skillfully wield these tools without succumbing to technological determinism.

Keywords: Artificial Intelligence, Data Analysis, Uber's Ride-Sharing Algorithm, Trip Chain

1. Introduction

In the context of the deepening digital economy, data has gradually become one of the core factors of production for enterprises. However, facing data resources with diverse sources, different structures, and huge scales, how to efficiently and accurately extract valuable information and transform it into effective support for business decisions has become a key challenge for enterprises. The limitations of traditional data analysis methods in processing high-dimensional, real-time, and unstructured data have led to the rapid rise of artificial intelligence (AI) technology [1], which has increasingly become a cutting-edge driving force for promoting the evolution of the business analysis field toward automation and intelligence.

In this context, this study focuses on the specific application mechanisms and practical effects of artificial intelligence in data analysis, particularly taking Uber's ride-sharing as an example, to

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.TJ28955

deeply explore how it utilizes core models such as "Trip Chain" to achieve multi-objective real-time optimization and dynamic resource allocation. By adopting a method combining case analysis and literature research, this paper aims to systematically address the following key questions: What are the advantages of the mechanism of AI technology in enhancing data analysis efficiency and decision-making intelligence? What technical bottlenecks and ethical constraints does it face during the process of complex system optimization? This study not only helps deepen the theoretical understanding of the AI-driven analytical paradigm, but also provides practical references for enterprises to build efficient and reliable data intelligence systems in real scenarios and has certain theoretical significance and application value.

2. Cutting-edge technologies in business analysis

2.1. Cutting-edge technologies

In recent years, data intelligence technology has continued to drive profound changes in enterprise decision-making methods. Enhanced Data Intelligence automatically completes data cleaning, feature engineering, and model selection through technologies such as AutoML, and is widely used in fields such as sales forecasting and supply chain risk early warning. Graph Neural Networks focus on mining complex patterns in relational data and perform prominently in anti-fraud (e.g., identifying credit card gang fraud) and social recommendation systems (e.g., relationship chain-based product recommendations). Generative AI relies on large language models (e.g., GPT-4) to generate analysis reports and simulate decision scenarios, enabling automated business insights and strategy simulation. At the same time, causal inference technology reveals the causal mechanisms between variables through methods such as difference-in-differences and instrumental variables, providing reliable explanations for issues such as advertising effect attribution and policy evaluation. These technologies together constitute the core architecture of a new generation of data-driven decision-making, significantly improving the automation, depth, and interpretability of analysis.

2.2. Reasons for being cutting-edge technologies

The reason why AI technology has become a cutting-edge development direction in the business analysis field mainly stems from the joint promotion of three forces: First, at the academic research level, cutting-edge papers published in top international conferences (e.g., KDD, NeurIPS) continuously verify the effectiveness and innovation of various AI methods [2], providing a solid theoretical foundation for technological development. Second, at the industry practice level, leading enterprises (e.g., Alibaba, Amazon) have proven the significant effects of AI technology in improving operational efficiency and optimizing resource allocation through large-scale implementation cases [3], and provided clear evidence of Return on Investment (ROI). Finally, at the technology popularization level, the continuous emergence of open-source tools and frameworks (e.g., FedML, DoWhy) has significantly lowered the application threshold of AI technology, promoted technological democratization, and enabled more enterprises and researchers to participate in and benefit from the innovation and application of intelligent analysis technology. These three aspects complement each other and jointly constitute an important support for the continuous development and wide application of AI technology in the business analysis field.

DOI: 10.54254/2755-2721/2026.TJ28955

3. Application of AI technology in data analysis: uber's ride-sharing algorithm

3.1. Concept of uber's ride-sharing algorithm

Dynamic ride-sharing, as a classic large-scale real-time optimization problem, had its core challenges defined by the academic community long before Uber and other enterprises commercialized it [4]. The operation of Uber's ride-sharing algorithm is a multi-stage collaborative process aimed at efficiently matching passengers with drivers and optimizing travel routes and costs. Its core process includes: Firstly, the system collects information such as the starting point, destination, and departure time of passengers; Secondly, route planning is carried out based on map APIs and real-time traffic data, and path search and optimization algorithms (such as merging orders and avoiding congestion) are used to plan efficient routes; Then, dynamic cost assessment is conducted based on route length, time cost, and traffic conditions to determine reasonable fees; Finally, driver matching is completed based on factors such as driver reputation, real-time location, and availability to improve service safety and completion rate. The entire process effectively enhances ride-sharing efficiency and user experience through data-driven and algorithmic decision-making.

3.2. Core concept of uber's ride-sharing

The essence of the ride-sharing algorithm is to plan a shared route for a vehicle, which, like a chain, passes through all passengers' pick-up and drop-off points in sequence.

For example, the original route of the vehicle is A -> B.

A new request is added: A passenger gets on at C and gets off at D.

The algorithm will calculate multiple possible routes:

• A -> C -> B -> D (may involve excessive detours)

• A -> C -> D -> B (more reasonable)

• C -> A -> B -> D (another sequence)

The algorithm will select the optimal one from these options.

3.3. Detailed working steps of uber's ride-sharing algorithm

The Uber carpool matching algorithm, as a typical large-scale real-time ridesharing optimization system, can be formally described as a multi-stage sequential decision-making process. When the system receives a new passenger's travel request (including the origin O and destination D), it first triggers the data collection module to obtain the real-time information of the available vehicle fleet (including their current geographical location, passenger-carrying status, remaining seats, etc.) and the passenger requests waiting for matching.

Subsequently, the algorithm enters the feasibility screening and matching generation stage. For each candidate vehicle, the system enumerates all possible ways to insert the new passenger's (O, D) pair into its existing itinerary chain, generating a candidate insertion scheme set. This process needs to satisfy several hard constraints: passenger capacity constraint (the number of passengers after insertion does not exceed the maximum seating capacity), time window constraint (the total travel

DOI: 10.54254/2755-2721/2026.TJ28955

time of each passenger does not exceed their maximum acceptable detour threshold), and route logical consistency constraint (to avoid invalid paths such as "getting off first and then getting on").

In the cost evaluation and optimal allocation stage, the algorithm calculates the multi-objective cost function for each feasible matching scheme. This function is usually a weighted combination of the following factors: the incremental total travel distance or time of the system, the additional travel delay for all passengers due to the matching, the degree of alignment between the new passenger's route and the vehicle's original route (used to measure the degree of detour), and the utility of this matching for the overall operational efficiency of the platform. Finally, the system solves this combined optimization problem based on the principle of cost minimization or utility maximization to determine the global optimal matching scheme.

The final stage is real-time decision-making and multi-end notification: The system pushes the updated route information to the driver's navigation interface and sends a matching confirmation message to the new passenger, including the estimated price and estimated arrival time (ETA). At the same time, to avoid experience degradation, the system also sends a lightweight notification to the existing passengers in the vehicle, indicating that their route has been slightly adjusted due to the carpool matching, thereby improving system efficiency while maintaining user satisfaction.

4. Advantages and limitations of AI applications in data analysis

4.1. Significant advantages

Artificial intelligence has triggered a significant efficiency revolution in the field of data analysis by automating data processing procedures (such as data cleaning, missing value filling, and anomaly detection), reducing processing time by over 70%. This has greatly enhanced the efficiency of analysis, for instance, in the automatic cleaning of e-commerce orders [5]. In terms of deep insight, AI demonstrates outstanding capabilities in complex pattern recognition and prediction: Graph Neural Networks (GNN) can parse multi-entity relationship networks and accurately identify money laundering gangs, with an accuracy rate exceeding 96% [6]; spatial-temporal models (such as STGCN) reduce retail sales prediction errors by significantly integrating multiple sources of information, like weather and promotions, to a level of 5.3%, far superior to the error level of traditional models, which is over 12%. Moreover, AI technology is also continuously lowering the threshold for analysis; for example, tools like Tableau GPT support data querying and analysis through natural language [7], enabling non-technical background personnel to directly participate in data exploration, such as quickly obtaining insights by asking, "Explain the reasons for the decline in sales in the third quarter". In conclusion, AI is comprehensively promoting the leap-type development of data analysis capabilities from three dimensions: efficiency, accuracy, and ease of use.

4.2. The challenges confronted

Although artificial intelligence has demonstrated significant advantages in data analysis, it still encounters various technical obstacles and risks in practical applications. At the technical level, deep learning models generally have the problem of "black box" decision-making, lacking interpretability, making it difficult to verify the rationality of their conclusions in high-risk scenarios such as medical diagnosis [8]; at the same time, the performance of the models highly depends on the quality of the training data, and biases in the data can easily lead to algorithmic discrimination; for example, credit models have a higher rejection rate for minority ethnic groups [9]. Furthermore,

DOI: 10.54254/2755-2721/2026.TJ28955

the generalization ability of AI systems is still insufficient. In extreme scenarios not covered by the training data (such as in agricultural remote sensing models when facing rare disaster weather), there may be a significant increase in errors.

In terms of security and ethics, the risk of privacy leakage cannot be ignored: even if technologies such as federated learning are adopted to prevent the direct transmission of raw data, the parameter interaction process may still indirectly expose users' behavioral patterns [10]; and in fields like autonomous driving, the ambiguity in defining the responsibility for AI decision-making errors also poses difficulties for legal accountability after accidents. On the other hand, the human-machine collaboration mechanism is not yet fully developed. Excessive reliance on AI may cause grassroots personnel to abandon necessary review and judgment. For instance, statisticians directly adopt the automatic cleaning results without conducting manual verification, resulting in logically inconsistent data being included in the report, which affects the reliability of decision-making. These issues indicate that the further development of AI technology not only requires breaking through the limitations at the algorithmic level, but also needs to achieve systematic innovation and construction in terms of ethical norms, security guarantees, and human-machine collaboration mechanisms.

5. Conclusion

This study has confirmed that artificial intelligence technology has completely transformed the basic paradigm of data analysis, shifting it from the traditional static, retrospective reporting model to a dynamic, real-time, and highly autonomous intelligent decision-making system. Typical applications such as Uber's carpooling algorithm clearly demonstrate how AI handles highly complex, multi-objective, and real-time changing business issues. This algorithm successfully resolves the problem of efficient matching between passengers and drivers in large-scale two-way markets through the "shared trip chain" model and the real-time optimization mechanism under multiple constraints. It not only significantly improves the utilization rate of road resources and the overall efficiency of the system, but also achieves a win-win situation for users, drivers, and the platform. This case fully demonstrates that AI technology not only has superiority in the theoretical aspect, but also shows strong engineering application potential in real business scenarios.

However, we must clearly recognize that technology is not omnipotent. The effectiveness of AI systems highly depends on high-quality training data, model design that fully considers business scenarios, and value orientation with ethical consciousness. The biases, omissions or noise in the data can significantly undermine the performance of the model, and even lead to discriminatory decisions; while if the model design neglects interpretability and robustness, it may cause systemic risks during actual deployment. For this reason, we emphasize that technology is fundamentally an enabling tool, a lever to expand human capabilities, and humans should always be the ones in control of this lever. Especially in critical decision-making stages, the final judgment power, value weighing, and ethical responsibility must be in the hands of humans. Technology should only provide support rather than replace the decision-making position of humans.

In the future, data analysis driven by AI will focus on developing in three directions: Firstly, developing more reliable AI technologies, such as Explainable Artificial Intelligence (XAI), to make the decision-making process of complex models transparent, traceable and auditable; Secondly, developing more efficient and energy-saving algorithms to meet the demands of ultra-large-scale and ultra-high real-time data processing; Thirdly, building a closer and more natural human-machine collaboration mechanism, enabling AI systems to form complementary advantages with human experts and jointly address complex decision-making environments. Under this trend, future data analysts will no longer merely be users of tools, but rather practitioners who are capable of critical

DOI: 10.54254/2755-2721/2026.TJ28955

thinking, able to correctly pose questions and understand the boundaries of technology. They need to master advanced AI technology tools as well as maintain an insight into the original business intentions, and be able to make more responsible and intelligent strategic decisions in human-machine collaboration. Only by combining technological empowerment with humanistic spirit can we truly promote the development of the data analysis field towards a more mature, reliable and sustainable direction.

References

[1] Fieyue Wang. (2018). Data Analysis and Decision Support System Based on Artificial Intelligence. Acta Automatica Sinica, 44(5), 789-801.

[2] Mei, S., & Zhu, X. (2015). Using machine learning to predict the success of crowdfunding campaigns. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1859-1868.

[3] Amazon Science. (2022). How Amazon is using causality to understand the impact of its business decisions. https: //www.amazon.science/latest-news/how-amazon-is-using-causality-to-understand-the-impact-of-its-business-decisions

[4] Agatz, N., Erera, A. L., Savelsbergh, M. W., & Wang, X. (2012). Optimization for dynamic ride-sharing: A review. European Journal of Operational Research, 223(2), 295-303.

[5] FB13713612741. (2025) JD API Data Governance Guide: In-depth Practical Analysis of Multi-Source Heterogeneous E-commerce Data Cleaning - JD Processes PB-level Transaction Data Daily [EB/OL]. CSDN blog. https: //blog.csdn.net/FB13713612741/article/details/146566426

[6] Liu, Z., Chen, C., Yang, X., Zhou, J., Li, X., & Song, L. (2018). Heterogeneous graph neural networks for malicious account detection. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 2077-2085.

[7] Pilgun, M. (2025). Explicit and Implicit Knowledge in Large-Scale Linguistic Data and Digital Footprints from Social Networks. Big Data and Cognitive Computing, 9(4), 6-7.

[8] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.

[9] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), 1-35.

[10] Shalan, M., Hasan, M.R., Bai, Y., & Li, J. (2025). Enhancing Smart Home Security: Blockchain-Enabled Federated Learning with Knowledge Distillation for Intrusion Detection. Smart Cities, 8(1), 5-6.