Thesis Proposal for MPhil Degree

Student Name:	Weilin Ruan
ID Number:	50018083
Group Project:	Connected Transportation Information System (CTIS)
Project Manager:	Wenxun Hu
Project Supervisor:	Wenxun Hu
Individual Project:	Retrieval-Augmented Universal Models for Spatio-temporal Data
Thesis Supervisor(s):	Yuxuan Liang
Student's Thrust & Hub:	Data Science and Analysis, Information Hub

Dec. 2024

The Hong Kong University of Science and Technology (Guangzhou)

Content

Part I Introduction to the Group Project

1.1 Background and Objective

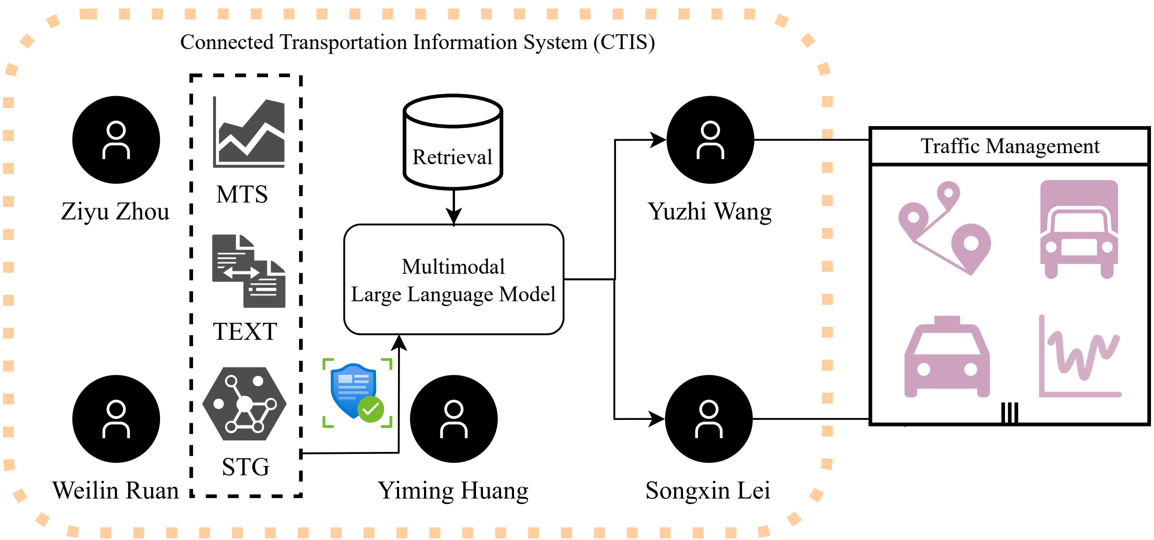

Figure 1: The outline of the composition and structure of our group project.

The rapid urbanization of modern cities has created unprecedented challenges for transportation systems [1]. With increasing population density and vehicle usage, cities are facing issues such as traffic congestion, delays, safety hazards, and environmental degradation [2]. These challenges not only affect the efficiency of transportation systems but also compromise public safety and hinder sustainable urban development. Addressing these issues requires innovative solutions that leverage advancements in intelligent technologies [3].

The field of Intelligent Transportation Systems (ITS) has emerged as a transformative approach to modern transportation management [4]. ITS integrates advanced sensing, communication, and data analysis technologies to create smarter and more adaptive transportation frameworks [5]. By utilizing real-time traffic monitoring, predictive analytics, and connected vehicle technologies, ITS aims to optimize traffic flow, enhance safety, and promote sustainability. ITS encompasses various key branches, including traffic management, autonomous vehicle support, infrastructure monitoring, and multimodal data fusion for urban planning [6].

Building on the vision of ITS, our group project, Connected Transportation Information System (CTIS), aims to develop an integrated platform that combines real-time traffic data, Vehicle-to-Everything (V2X) communication, and predictive analytics [3]. This platform seeks to address critical challenges in urban transportation systems and provide innovative solutions for optimizing traffic flow, enhancing road safety, and reducing environmental impact.

Modern transportation systems face several pressing challenges [7]. Increasing traffic demand in urban areas has resulted in severe congestion and delays, which not only affect the efficiency of daily commutes but also lead to significant economic losses [8]. Safety concerns remain a critical issue, as the lack of real-time hazard detection and preventive measures contributes to frequent accidents. Furthermore, inefficient route planning and traffic management result in excessive energy consumption and carbon emissions, exacerbating environmental degradation.

To address these challenges, the CTIS project adopts a comprehensive and interdisciplinary approach. The project focuses on integrating real-time communication and data-sharing systems to establish seamless interactions between vehicles, infrastructure, and management centers. Predictive analytics and machine learning technologies will be employed to analyze dynamic traffic patterns, detect potential hazards, and provide actionable insights to optimize traffic flow [9]. Additionally, the project incorporates V2X communication technologies to enhance connectivity and support autonomous vehicle operations, while multimodal data fusion will enable urban planners to make informed decisions for long-term mobility solutions.

The CTIS project is structured to achieve the following key objectives:

1. **Real-time Traffic Optimization:** Develop advanced algorithms for intelligent traffic management to reduce congestion and delays [3].

2. **Safety Enhancement:** Integrate V2X communication technologies and real-time monitoring to improve hazard detection and accident prevention capabilities [7].

3. **Sustainability Promotion:** Facilitate efficient route planning and traffic management to reduce energy consumption and carbon emissions [5].

4. **Urban Mobility Support:** Utilize multimodal data fusion to generate actionable insights for urban planning and decision-making [6].

1.2 Significance

The CTIS project represents a transformative initiative addressing critical challenges in urban transportation systems while advancing the evolution of Intelligent Transportation Systems (ITS) [1]. Its significance lies in integrating advanced sensing, communication, and data analytics to create smarter, safer, and more sustainable transportation networks [3]. By synthesizing these technologies, CTIS provides solutions for both immediate urban mobility challenges and future demands of adaptive, technology-driven cities [10].

The project's significance can be categorized into several key dimensions:

- **Reducing Congestion and Improving Efficiency:** CTIS leverages real-time traffic data and predictive algorithms to optimize traffic flow, minimize delays, and enhance the overall efficiency of urban transportation systems [10]. Recent studies indicate that such intelligent systems can reduce average travel times by up to 20% in dense urban areas[6].
- **Improving Road Safety:** Through the integration of connected vehicle technologies with real-time hazard detection, the project significantly reduces accident risks and enhances safety for all road users, including vulnerable road users such as pedestrians and cyclists.
- **Promoting Environmental Sustainability:** Efficient traffic management and route optimization contribute to reduced fuel consumption and carbon emissions. This aligns with global environmental initiatives and supports sustainable urban development goals.
- **Advancing Urban Planning:** CTIS generates multimodal data-driven insights that enable urban planners to design more resilient cities and optimize infrastructure development. These insights facilitate evidence-based decision-making in urban development.
- **Driving Technological Innovation:** The development of CTIS catalyzes innovation in transportation research, creating scalable and adaptable solutions applicable across diverse urban environments [1].

Beyond addressing immediate transportation challenges, CTIS establishes a scalable framework for intelligent transportation management that can evolve with changing urban needs [3]. Its holistic approach ensures that the project serves not only as a response to current transportation problems but also as a foundation for the continuous advancement of ITS [11]. By promoting intelligent traffic systems and sustainable urban mobility, CTIS contributes significantly to realizing a more adaptive, environmentally responsible, and technology-driven future in urban transportation.

1.3 Project Composition

The initiative encompasses multiple synergistic research streams, each contributing distinct methodological innovations and technological breakthroughs to the overall framework.

1. Weilin Ruan: **Retrieval-Augmented Universal Models for Spatio-Temporal Data**
   This research stream advances a novel framework for processing and analyzing extensive spatio-temporal datasets. The methodology addresses the complexities of integrating heterogeneous urban data streams, including remote sensing imagery, sensor networks, and transit monitoring systems. The proposed architecture employs retrieval-augmented methodologies to enhance model performance while maintaining interpretability in spatio-temporal analysis.

2. Ziyu Zhou: **Frequency-Enhanced Framework for Time Series Forecasting**
   This stream introduces WaveTS, a wavelet-based forecasting architecture optimized for traffic prediction. The framework decomposes complex traffic patterns across multiple frequency domains, enabling edge-deployable solutions for real-time congestion prediction and traffic safety enhancement.

3. Yiming Huang: **Safety Protocols and Robustness Assessment**
   This research establishes comprehensive validation protocols through mechanistic interpretability, focusing on hallucination detection and mitigation strategies. The framework introduces novel foundations for system performance assessment and addresses information bias through multi-source integration techniques.

4. Yuzhi Wang: **LLM-Enhanced Urban Agent Simulation**
   This component explores the integration of large language models in urban agent-based simulations. Through advanced prompt engineering and retrieval mechanisms, the research models complex human decision-making patterns in urban environments, providing insights into mobility patterns and infrastructure optimization.

5. Songxin Lei: **Collaborative Public Resource Allocation**
   This research stream develops optimization strategies for mobile public resource deployment. The methodology synthesizes potential game theory with spatio-temporal neural architectures to ensure optimal resource distribution, generating actionable strategies for urban resource management.

These complementary research streams converge to form a comprehensive approach toward sustainable urban mobility, addressing key challenges in transportation while promoting system efficiency and safety protocols.

1.4 Project Connections

The individual research streams within the Connected Transportation Information System are strategically integrated to form a unified and efficient framework:

- **Weilin's Predictive Analysis:** The retrieval-augmented framework serves as a foundational component, enabling efficient processing of multi-scale spatio-temporal data streams. This architecture enhances downstream applications through contextual data integration and cross-domain knowledge transfer. The framework augments Ziyu's temporal modeling by providing relevant historical patterns, strengthens Songxin's infrastructure planning through enhanced spatial representations, and supports Yuzhi's trajectory analysis via transferable pre-trained models. The resulting cross-city generalization capabilities improve prediction accuracy for vehicle-pedestrian interactions and autonomous navigation systems.

- **Ziyu's Strategic Deployment:** The multivariate time series modeling framework processes diverse data streams within the ecosystem, incorporating textual insights and sensor metrics for enhanced network-wide forecasting and optimization.

- **Yiming's Safety Protocols:** The framework implements comprehensive safety protocols addressing model hallucination, robustness verification, and debiasing mechanisms to ensure reliable system performance.

- **Yuzhi's Urban Simulation:** The generative agent-based modeling creates digital twin environments through LLM integration, enabling dynamic urban scenario evaluation and system optimization.

- **Songxin's Resource Management:** This component implements environment-aware decision-making protocols, optimizing mobile resource deployment through synthesized predictions and operational insights.

1.5 Project Milestones

The CTIS development follows a strategic four-phase timeline, integrating individual research components into a unified transportation intelligence system.

Phase 1	**Foundation: Data Acquisition and Model Design (Months 0-3)**
Core Focus	Development of foundational models and data collection framework
Key Deliverables	• Retrieval-augmented spatio-temporal data integration system • Advanced time series prediction frameworks • Safety verification and robustness assessment protocols • Urban agent simulation and behavior modeling • Resource optimization and allocation algorithms

Phase 2	**Integration: Cross-Component Synthesis (Months 3-6)**
Core Focus	Component integration and data pipeline establishment
Key Deliverables	• Integration of prediction models with agent simulations • Development of unified data processing pipelines • Implementation of cross-module verification systems

Phase 3	**Validation: System Testing (Months 6-9)**
Core Focus	System validation and performance assessment
Key Deliverables	• Integrated system testing and validation • Multi-scenario simulation evaluation • Comprehensive safety and robustness verification

Phase 4	**Implementation: Deployment Strategy (Months 9-12)**
Core Focus	System deployment and optimization
Key Deliverables	• Urban area pilot implementation • System performance optimization • Future development roadmap

Part II Proposal of the Individual Project

2.1 Significance and Relevance of the Individual Project to the Group Project

2.1.1 Complementary Role

The individual project, **Retrieval-Augmented Universal Models for Spatio-temporal Data (RAST)**, plays a pivotal role in enhancing the overall objectives of the group project, Connected Transportation Information System (CTIS). While the group project focuses on creating an integrated platform to improve urban mobility through intelligent algorithms and predictive analytics, this individual project addresses the critical challenge of generalization and adaptability across diverse urban environments.

The proposed RAST framework enhances the group project by integrating pre-trained universal models with dynamic retrieval mechanisms. This allows for efficient handling of complex spatio-temporal tasks, such as traffic flow prediction, hazard detection, and urban mobility optimization. By dynamically retrieving relevant historical data or external knowledge during inference, the RAST framework reduces computational redundancy, improves prediction accuracy, and supports data-driven decision-making processes across various CTIS modules [12, 13].

2.1.2 Value Addition

This project introduces innovative advancements in spatio-temporal modeling through the application of retrieval-augmented generation (RAG) techniques, which are traditionally applied in natural language processing but adapted here for urban spatio-temporal tasks. The dynamic retrieval mechanism enhances the adaptability and scalability of pre-trained universal models, making them more effective for complex urban transportation scenarios.

The key contributions of this project to the group project can be summarized as follows:

- **Improved Generalization across Urban Environments**: The RAST framework enables the CTIS platform to generalize across diverse cities and urban regions, addressing variations in traffic patterns, road structures, and sensor coverage.

- **Optimized Data Utilization**: By leveraging retrieval mechanisms, the RAST model reduces the dependency on large-scale labeled datasets, ensuring robust performance even with limited data availability.

- **Enhanced Predictive Accuracy and Efficiency**: Through retrieval-augmented techniques, the framework achieves higher accuracy in tasks such as traffic flow prediction and hazard detection, providing reliable inputs for real-time traffic management and long-term urban planning [5].

2.1.3 Interdependence

The individual project is tightly interwoven with other components of the group project, ensuring a cohesive and efficient system. The interdependencies are outlined as follows:

- **Data Sharing**: The retrieval mechanism in the RAST framework relies on high-quality spatio-temporal datasets, which are also utilized in other CTIS components for real-time traffic monitoring, congestion analysis, and safety enhancement. This shared data infrastructure ensures consistency and mutual reinforcement between individual and group projects.

- **Algorithmic Synergy**: The RAST framework's outputs, such as spatio-temporal predictions and retrieved contextual insights, directly support other group project modules. For example, the temporal forecasts generated by the RAST model can improve the accuracy of traffic optimization algorithms and hazard detection systems.

- **Integrated Platform Contributions**: The RAST model's outputs are seamlessly integrated into the CTIS platform, providing actionable insights for various sub-tasks, including traffic flow optimization, multimodal urban planning, and emergency response management.

This interdependence ensures a holistic approach to addressing urban transportation challenges. By combining the retrieval-augmented capabilities of the RAST model with the broader objectives of the CTIS framework, the collaborative effort contributes to real-time traffic optimization, enhanced road safety, and sustainable urban development.

2.2 Statement of the Individual Project in Details

2.2.1 Literature/Market Review and Problem Definition

Spatio-temporal prediction serves as a fundamental cornerstone in urban mobility optimization, enabling critical tasks such as traffic forecasting, hazard detection, and infrastructure planning. Traditional statistical approaches, including ARIMA and conventional machine learning methods, frequently exhibit limitations in cross-regional generalization due to the inherent complexity and dynamic nature of urban environments. While recent advances in deep learning, particularly Spatio-Temporal Graph Neural Networks (STGNNs), have substantially enhanced modeling capabilities, they continue to face significant challenges in achieving an optimal balance between **generalization performance** and **computational efficiency**.

The advent of **large-scale pre-trained models** in natural language processing (NLP) and computer vision has revolutionized the landscape of spatio-temporal data modeling. However, despite their remarkable success, these models encounter several critical limitations when applied to urban mobility tasks:

- **Data Heterogeneity**: Urban environments exhibit substantial variability across different regions, presenting significant challenges for universal models to adapt effectively without extensive localized fine-tuning.
- **Computational Complexity**: The inherently high-dimensional nature of spatio-temporal data introduces substantial computational overhead during both training and inference phases, potentially compromising system scalability.
- **Dynamic Environments**: Contemporary urban landscapes undergo rapid evolution, necessitating models capable of dynamic adaptation to emerging patterns and shifting data distributions.

To address these fundamental challenges, this project introduces innovative **retrieval-augmented techniques** to enhance both the adaptability and efficiency of universal spatio-temporal models. Through dynamic retrieval of relevant historical data and external knowledge during inference, our proposed Retrieval-Augmented Spatio-temporal (RAST) framework achieves superior prediction accuracy while significantly reducing computational redundancy. This novel approach demonstrates particular effectiveness in challenging scenarios such as cross-region applications and data-scarce environments, where conventional methodologies often exhibit suboptimal performance [12, 13].

2.2.2 Objective and Scope of the Project

This research initiative aims to develop an innovative **Retrieval-Augmented Spatio-temporal (RAST) Model** that seamlessly integrates pre-trained universal models with dynamic retrieval

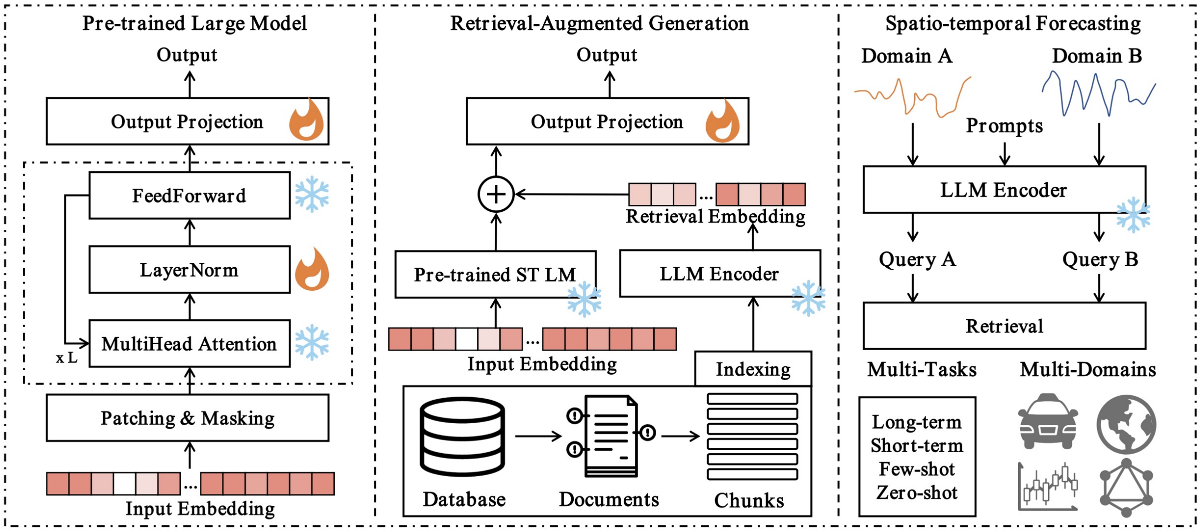

Figure 2: The framework of Retrieval-Augmented Universal Models.

mechanisms. The proposed model specifically focuses on enabling robust generalization and adaptive capabilities across diverse urban environments, thereby supporting the comprehensive vision of the *Connected Transportation Information System (CTIS)* platform.

To validate our approach, we will leverage the **LargeST benchmark dataset** [14], which provides comprehensive transportation data across multiple regions in California, including Greater Los Angeles (GLA), Greater Bay Area (GBA), and San Diego (SD). This extensive dataset, comprising over 301M samples from 8,600 sensors across different urban areas, offers an ideal testbed for our research objectives.

1. **Framework Development**

    - Design and implement a scalable architecture supporting efficient cross-region generalization, capable of handling diverse urban contexts (e.g., GLA: 3,834 points, GBA: 2,352 points)
    - Develop comprehensive pre-training strategies optimized for large-scale urban datasets spanning extended timeframes (35,040 time slices over one year)
    - Create dynamic retrieval mechanisms for accessing historical data and external knowledge across multiple metropolitan regions

2. **Technical Implementation**

    - Capture long-term spatio-temporal dependencies through advanced pre-training techniques, leveraging the extensive temporal coverage (01/01/2019-12/31/2019)

- Implement adaptive retrieval mechanisms optimized for real-time inference across varying sensor densities
- Integrate the RAST framework within the CTIS platform architecture, ensuring scalability from local (716 points in SD) to regional scales (8,600 points in CA)

3. **Validation and Evaluation**

- Validate cross-region generalization capability using the diverse geographic distribution of sensors across California.
- Evaluate the effectiveness of retrieval mechanisms using the rich temporal and spatial patterns present in the LargeST dataset.
- Assess model scalability through comprehensive testing across different regional scales, from individual cities to state-wide applications
- Benchmark performance against existing approaches using standardized metrics on the LargeST dataset

2.2.3 Research Method and Justification

The research methodology encompasses three interconnected technical components that form the foundation of our approach to spatio-temporal modeling. The architecture is designed to address the fundamental challenges in urban mobility prediction while leveraging the latest advances in deep learning and retrieval-augmented generation.

The first component focuses on the development of a **pre-trained large model architecture.** At its core, the design implements a transformer-based structure incorporating multiple sophisticated elements for robust spatio-temporal learning. The architecture employs Multi-Head Attention layers to capture complex spatial dependencies across urban regions, complemented by LayerNorm mechanisms that ensure training stability. FeedForward networks handle feature transformation processes, while specialized Patching and Masking mechanisms effectively manage missing data scenarios common in urban sensing networks. This component culminates in efficient output projection layers optimized for various spatio-temporal tasks, establishing a robust foundation for urban mobility prediction.

The second component advances the **retrieval-augmented generation framework**, introducing innovative mechanisms for dynamic information access and integration. This system begins with a comprehensive database architecture designed to store and index historical spatio-temporal patterns efficiently. Through carefully designed document chunking strategies and optimized indexing mechanisms, the system enables rapid retrieval of relevant historical data during inference. A sophisticated LLM encoder processes both input embeddings and retrieval results, while custom fusion mechanisms seamlessly integrate pre-trained model outputs with retrieved information. This integration significantly enhances the model's ability to adapt to diverse urban contexts and changing mobility patterns.

The **validation and evaluation strategy** forms the third critical component, leveraging the comprehensive LargeST benchmark dataset to ensure rigorous assessment of the framework's capabilities. The evaluation spans multiple geographic scales, from local urban areas to regional networks, utilizing data from diverse California regions including Greater Los Angeles (3,834 points), Greater Bay Area (2,352 points), and San Diego (716 points). The temporal validation encompasses 35,040 time slices over a full year (01/01/2019-12/31/2019), providing extensive coverage of seasonal and temporal patterns. Performance assessment includes comprehensive metrics examining prediction accuracy, computational efficiency, and cross-region generalization capabilities. The evaluation framework includes systematic comparisons against traditional statistical methods such as ARIMA and VAR, as well as modern deep learning approaches including STGNNs and standard Transformers.

This methodological approach ensures the RAST framework achieves robust performance across diverse urban environments while maintaining efficient handling of large-scale spatio-temporal data. The integration of pre-trained models with retrieval-augmented generation techniques, validated through the extensive LargeST dataset, provides a comprehensive solution to the challenges of urban mobility prediction. The methodology specifically addresses the critical balance between model generalization and computational efficiency, while ensuring practical applicability in real-world transportation systems. Through this structured approach, the framework demonstrates significant advantages in both theoretical advancement and practical implementation within urban computing environments.

2.2.4 Execution Plan

The project will be executed over a 12-month period, divided into the following phases:

- **Phase 1 (Months 1–3)**: Data collection, preprocessing, and integration. This phase establishes the foundation for building the RAST framework.

- **Phase 2 (Months 4–6)**: Development of the RAST framework, including the implementation of the retrieval mechanism and the integration of pre-trained models.

- **Phase 3 (Months 7–9)**: Pre-training and fine-tuning the framework on urban datasets. This phase includes optimizing the retrieval mechanism for spatio-temporal tasks.

- **Phase 4 (Months 10–12)**: Evaluation and validation of the framework, followed by its integration into the CTIS platform. This phase also involves generating reports to summarize findings and outcomes.

2.2.5 Intended Outcomes

The expected outcomes of this project are as follows:

- A retrieval-augmented spatio-temporal framework (RAST) that significantly enhances generalization and adaptability across diverse urban environments.

- Improved prediction accuracy and computational efficiency for urban mobility tasks, particularly in cross-region and data-scarce scenarios.

- Seamless integration of the RAST framework into the CTIS platform, providing actionable insights to support data-driven urban mobility optimization.

2.3 Project Milestones

The milestones of my individual project, *"Retrieval-Augmented Universal Models for Spatio-temporal Data"*, are outlined below: **Milestone 1: Data Collection and Preprocessing (Months 1–2)**

- Collect relevant spatio-temporal datasets, including traffic flow, weather data, and urban mobility patterns from multiple urban regions.

- Design and implement a preprocessing pipeline to clean, normalize, and integrate datasets, ensuring compatibility with the RAST framework.

- Analyze dataset quality to identify potential gaps and ensure readiness for model development.

Milestone 2: Framework Design and Development (Months 3–5)

- Design the core architecture of the RAST model, integrating dynamic retrieval mechanisms for enhanced spatio-temporal learning.
- Implement pre-training pipelines to extract long-term spatio-temporal dependencies and patterns from diverse datasets.
- Conduct initial testing to evaluate the framework's modularity, scalability, and computational efficiency.

Milestone 3: Model Training and Fine-tuning (Months 6–8)

- Train the pre-trained RAST model on large-scale urban datasets to ensure generalization across diverse regions.
- Fine-tune the model for specific urban mobility tasks, such as traffic flow prediction and congestion analysis.
- Validate the model's performance using metrics such as prediction accuracy, computational efficiency, and adaptability to dynamic environments.

Milestone 4: Deployment and Evaluation (Months 9–11)

- Integrate the RAST framework with the group project's *Connected Transportation Information System (CTIS)* platform.
- Conduct field testing across diverse urban settings to validate the framework's adaptability and effectiveness in real-world scenarios.
- Collaborate with team members to ensure seamless integration of outputs and alignment with group project objectives.

Milestone 5: Visualization and Reporting (Month 12)

- Develop a user-friendly visualization system to facilitate analysis and interpretation of spatio-temporal predictions.
- Prepare a comprehensive project report summarizing results, insights, and recommendations for real-world implementation.
- Present findings and demonstrate the impact of the RAST framework to stakeholders and team members.

Milestone	Individual Project: RAST Framework Development	Group Project: Urban Mobility Ecosystem
Milestone 1: Data Collection and Preprocessing	Collect and preprocess spatio-temporal datasets for model training and evaluation.	Data collection across members' domains (e.g., traffic, weather, urban planning). Initial dataset alignment among team members.
Milestone 2: Framework Design and Development	Design and implement the RAST framework, including retrieval mechanisms and pre-training pipelines.	Establish baseline models and frameworks in individual domains. Initial integration of shared components into the group platform.
Milestone 3: Model Training and Fine-tuning	Train and fine-tune the RAST model on multi-region datasets. Validate its performance on tasks such as traffic flow prediction.	Integration of fine-tuned models from members into the group project platform. Collaborative testing and optimization of shared components.
Milestone 4: Deployment and Evaluation	Deploy the RAST framework in real-world scenarios and validate its adaptability across urban regions.	Deploy the integrated mobility platform. Conduct pilot testing to evaluate system performance, scalability, and effectiveness.
Milestone 5: Visualization and Reporting	Develop a visualization system for analyzing spatio-temporal predictions and prepare the final report.	Complete platform integration with visualization features. Summarize project outcomes and provide recommendations for future improvements.

2.4 Budget Plan

2.4.1 Estimated Budget

The budget for the individual project, *"Retrieval-Augmented Universal Models for Spatio-temporal Data"*, is estimated to be **15,000 RMB**. This budget is allocated to cover key aspects of the project, including data acquisition, computational resources, software development, and miscellaneous expenses. The detailed breakdown is shown in Table 3.

Table 3: Estimated Budget for the Individual Project

Category	Budget (RMB)
Data Acquisition and Preprocessing	3,000
- Data acquisition from public and private sources	2,000
- Data preprocessing and integration services	1,000
Computational Resources	7,000
- Cloud computing (GPU rentals, storage)	5,000
- Optimization of computational infrastructure	2,000
Software Development and Tools	4,000
- Framework development and tool integration	3,000
- Open-source software licenses	1,000
Miscellaneous Expenses	1,000
- Contingency fund for unexpected costs	700
- Communication and internet expenses	300
Total	**15,000**

2.4.2 Budget Breakdown

The budget is divided into four major categories to ensure efficient resource allocation:

1. Data Acquisition and Preprocessing (3,000 RMB):

- Data Acquisition (2,000 RMB): Includes costs for obtaining spatio-temporal datasets from public and private sources. Efforts will focus on leveraging publicly available datasets to minimize expenses.

- Data Preprocessing (1,000 RMB): Covers the cost of cleaning, normalizing, and integrating datasets for compatibility with the RAST framework.

2. Computational Resources (7,000 RMB):

- Cloud Computing (5,000 RMB): Includes GPU rentals and storage for training and evaluating the model.

- Infrastructure Optimization (2,000 RMB): Covers efforts to optimize computational pipelines and reduce resource overhead.

3. Software Development and Tools (4,000 RMB):

- Framework Development (3,000 RMB): Includes costs for implementing the RAST model and integrating retrieval mechanisms.
- Open-source Tools (1,000 RMB): Covers licensing costs for necessary software tools.

4. Miscellaneous Expenses (1,000 RMB):

- Contingency Fund (700 RMB): Reserved for unforeseen expenses, ensuring smooth project execution.
- Communication (300 RMB): Includes internet and communication costs for coordinating with team members and accessing cloud resources.

2.4.3 Cost-Effectiveness

The proposed budget prioritizes cost-effectiveness while ensuring high-quality project outcomes. Key strategies include:

- **Utilizing Publicly Available Datasets:** Reduces data acquisition costs by leveraging open-access datasets and public resources.
- **Efficient Training Techniques:** Adopts optimized training and fine-tuning processes to minimize computational overhead and cloud computing expenses.
- **Relying on Open-source Tools:** Limits software development costs by leveraging well-established open-source frameworks and tools.

These strategies ensure the project remains within the allocated budget while achieving its objectives of developing an adaptable and efficient retrieval-augmented spatio-temporal framework.

2.5 Risk Analysis and Mitigation

2.5.1 Potential Risks or Challenges

The project may face the following risks and challenges during its development and implementation:

- **Data Scarcity or Quality Issues:** Limited access to high-quality, labeled spatio-temporal datasets may hinder effective model training and evaluation. Additionally, existing datasets may contain incomplete, inconsistent, or noisy data.
- **High Computational Costs:** Training large-scale models with retrieval mechanisms is resource-intensive, requiring significant GPU and storage resources, which could lead to delays or budget overruns.
- **Model Overfitting or Retrieval Errors:** The retrieval mechanism may retrieve irrelevant or noisy data, causing the model to overfit to specific patterns and reducing its generalization capability in cross-region scenarios.
- **Integration Challenges:** Difficulties may arise when integrating the RAST framework with the group project's *Connected Transportation Information System (CTIS)* platform, due to compatibility issues or differing technical requirements.
- **Dynamic Urban Environments:** Rapid changes in urban mobility patterns (e.g., new traffic regulations or infrastructure changes) may reduce the relevance of pre-trained data and affect model adaptability.

2.5.2 Impact of the Risks

The identified risks carry significant implications for the project's success and implementation timeline. Data quality issues could fundamentally impact the model's performance, particularly in challenging scenarios such as few-shot or zero-shot learning applications. This limitation becomes especially critical when attempting to generate accurate predictions for new regions or unprecedented urban situations.

The computational resource demands pose both immediate and long-term challenges. Beyond potential project delays and budget overruns, these constraints could significantly limit our ability to experiment with different model configurations and retrieval strategies, potentially affecting the optimization of the framework's performance. Poor retrieval performance could substantially impact the model's real-world applicability, compromising its adaptability and accuracy in dynamic urban environments.

Integration challenges with the broader CTIS platform could create cascading effects throughout the group project. Delays in framework integration might impede overall project progress and postpone critical field testing phases, affecting the timeline for system deployment and validation. Moreover, the rapidly evolving nature of urban environments poses ongoing challenges to model maintenance and effectiveness. As city dynamics continue to evolve, the pre-trained models may require more frequent updates and fine-tuning than initially anticipated, potentially increasing long-term operational costs and complexity.

These impacts underscore the importance of implementing robust mitigation strategies and maintaining flexibility in our development approach. Through careful planning and proactive risk management, we aim to address these challenges while ensuring the RAST framework achieves its intended objectives in advancing urban mobility prediction capabilities.

References

[1] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, pp. 1–55, 2014.

[2] M. Zhao, H. Cai, Z. Qiao, and X. Xu, "Influence of urban expansion on the urban heat island effect in shanghai," *International Journal of Geographical Information Science*, vol. 30, no. 12, pp. 2421–2441, 2016.

[3] S. Wang, J. Cao, and S. Y. Philip, "Deep learning for spatio-temporal data mining: A survey," *IEEE transactions on knowledge and data engineering*, vol. 34, no. 8, pp. 3681–3700, 2020.

[4] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383–398, 2018.

[5] Q. Zhang, J. Chang, G. Meng, S. Xiang, and C. Pan, "Spatio-temporal graph structure learning for traffic forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 1177–1185.

[6] Y. Liang, K. Ouyang, L. Jing, S. Ruan, Y. Liu, J. Zhang, D. S. Rosenblum, and Y. Zheng, "Urbanfm: Inferring fine-grained urban flows," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 3132–3142.

[7] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, "Connecting the dots: Multivariate time series forecasting with graph neural networks," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 753–763.

[8] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.

[9] Y. Yang and T. Hospedales, "Deep multi-task representation learning: A tensor factorisation approach," 2016.

[10] L. Cai, K. Janowicz, G. Mai, B. Yan, and R. Zhu, "Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting," *Transactions in GIS*, vol. 24, no. 3, pp. 736–755, 2020.

[11] X. Zheng, W. Chen, P. Wang, D. Shen, S. Chen, X. Wang, Q. Zhang, and L. Yang, "Big data for social transportation," *IEEE transactions on intelligent transportation systems*, vol. 17, no. 3, pp. 620–630, 2015.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[13] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.

[14] X. Liu, Y. Xia, Y. Liang, J. Hu, Y. Wang, L. Bai, C. Huang, Z. Liu, B. Hooi, and R. Zimmermann, "Largest: A benchmark dataset for large-scale traffic forecasting," in *Advances in Neural Information Processing Systems*, 2023.