# KAIZHEN TAN

◇ Email: tkz@tongji.edu.cn ◇ Phone: +86 15214390380

## EDUCATION

**Tongji University** Sept. 2021-Present
*Bachelor's Degree, Information Management and Information System, GPA: 90.35/100* Shanghai
**Relevant Courses**: Data Structures, Python/C++/Java Programming, Computer Network and Communication, Operating System, Database Technology and Applications, Advanced Mathematics, Probability and Mathematical Statistics, Applied Statistics, Discrete Mathematics, Data Mining and Data Analysis

## RESEARCH EXPERIENCES

**Spatio-temporal Interaction Mechanism of Human Activities and Traffic Congestion Propagation — Project funded by the National Natural Science Foundation of China** Apr. 2024-Present
*Research Assistant, Advisor: Prof. Wei Huang* *Shanghai*

- To model traffic congestion propagation patterns (TCPP) using geospatial-temporal-semantic knowledge graphs.
- Utilize map matching algorithm to project one month's worth of spatio-temporal taxi trajectory data from Shanghai onto real roads, and calculate traffic state index based on speeds in each time slice to identify congested road segments and establish road network graphs.
- Use community detection algorithm based on spatial and topological similarity to cluster congested segments, and observe significant changes in congested subgraphs over adjacent time slices using ARI and NMI to detect TCPP.
- Identify the correlation and causality between TCPP and urban built environment by calculating weighted values of POIs within each buffer, and quantify the impacts of TCCP through propagation probabilities.

**Analysis of Tourists' Focal Preferences and Recreational Experience in Historic Urban Quarters based on Deep Learning** Jan. 2024-Present
*Student Researcher, Advisor: Prof. Yujia Zhai* *Shanghai*

- To assess tourist focal points and perception evaluation (satisfaction and sentiment tendency) of historic urban quarters in Shanghai using deep learning methods.
- Crawl travel reviews from social media platforms and develop a semantic segmentation model for scenic spot extraction and image classification using pre-trained models such as PSPNet and Seg2Former, integrating image annotation and hyperparameter optimization in the fine-tuning process to evaluate tourists' focal preferences.
- Construct a language model for sentiment analysis and satisfaction assessment by establishing an encoding system of satisfaction factors, annotating textual data, and employing BERT and Mamba with supervised fine-tuning.
- Visualize and present analytical results of images and texts by incorporating techniques such as word frequency analysis, co-occurrence analysis and community detection.

**CFGPT: Chinese Financial Generative Pre-trained Transformer Framework** Jan. 2024-Apr. 2024
*Student Researcher, Advisor: Prof. Dawei Cheng* *Shanghai*

- Established the datasets of CFGPT, an open-sourced Chinese financial large language model (LLM), and achieved model refinement as a prompt engineer in collaboration with Shanghai Artificial Intelligence Laboratory.
- Collected online financial content using proxy-based distributed crawlers, extracted and filtered texts with regular expressions, banned word lists and locality-sensitive hashing algorithm, assembling a pre-training dataset.
- Designed text prompts for supervised fine-tuning training and model optimization, enhancing model generalization ability for 6 specific downstream financial tasks like sentiment analysis, topic decomposition and stock prediction.
- Crafted real-world financial application cases to provide textual support for retrieval-augmented generation (RAG).

**Pavement Disease Recognition Using Object Detection — Global Campus Artificial Intelligence Algorithm Elite Competition** Sept. 2023-Nov. 2023
*Member* *Nanjing*

- Utilized YOLOv8 model for object detection of pavement diseases, added a detection module for small targets of 4x4 pixels and replaced the original loss function with Wise-IoU, enhancing the model's detection sensitivity by 3% in the competition dataset.

- Expanded the dataset through online enhancement of tuning parameters and offline random data augmentation techniques to mitigate the impact of sample imbalance and improve the generalization capabilities and robustness.
- Compared and analyzed different combinations of optimization methods, identified an optimized model encompassing small targets, Wise-IoU, and data augmentation, winning the Second Prize at national level (top 5%).

## ACADEMIC PROJECTS

**Design of a Medical Information Service Platform Website** Jun. 2023-Aug. 2023
*Leader* *Shanghai*

- Led a team of five members to design and build a medical information service platform, providing authoritative medical advice, comprehensive healthcare information, and personalized treatment recommendations.
- Utilized the Bootstrap framework for frontend layout design and the Django framework (Python web) for backend development, and constructed a database containing information on hospitals, users, and diseases with MySQL.
- Developed an algorithm to provide patients with personalized recommendations regarding hospitals and departments based on their symptoms and the geographical locations of hospitals and patients using the Amap API.
- Completed cloud server deployment and successfully launched the project online.

**Agent-based Modeling and Simulation System for Library Seat Selection** Jun. 2023-Aug. 2023
*Leader* *Shanghai*

- Led a team of six members to collect data on the school library environment and readers through onsite investigation and questionnaire surveys, applied K-means clustering to classify reader agent types and fuzzy analytics hierarchy process (FAHP) to analyze demographic characteristics and behavioral decision-making patterns.
- Used Anylogic software to construct a virtual library environment and create agents representing readers and seats, applied Java to simulate the decision-making process for seat selection based on reader attributes and environmental preferences, and simulated library environment changes by generating random events.
- Optimized and validated model inputs and outputs, adjusting reader attributes and seat booking preference parameters for hypothesis tests and model enhancement, increasing grey correlation degree to 0.87.

## INTERNSHIP

**Institute for Infocomm Research (I2R), A*STAR** (Expected) Sep. 2024-Dec. 2024
*Research Officer Intern* *Singapore*

- Expected to collaborate with the research team focused on Intelligent Transportation Solutions.

**Shanghai Qiantan Emerging Industry Research Institute** Jan. 2023-Feb. 2023
*Data Analyst Intern* *Shanghai*

- Used PyQuery and Requests to crawl news headlines from global think tanks, and employed Pandas for extensive data cleaning, transformation, and normalization, addressing missing values and outliers.
- Employed Jieba for word segmentation of news headlines and conducted statistical analysis on high-frequency terms; leveraged machine learning techniques such as regression, random forest, SVM and XGBoost to perform predictive modeling and trend analysis; visualized results using Seaborn and Matplotlib.
- Conducted an industrial study of Liaoyang with economic data including local businesses and recruitment information; applied clustering and association algorithms to obtain demographic profile and requirement diagram.

## AWARDS

Global Sustainability Supply Chain Student Competition, Second Round Jan. 2024
Global Campus Artificial Intelligence Algorithm Elite Competition, Second Prize Dec. 2023

## SKILLS

**Languages:** IELTS 7.5 (L:9, R:8.5, W:6.5, S:6), GRE 331 (V:161, Q:170, AW:3)
**Programming & Tools:** Python, C++, C#, Java, HTML, SQL, Docker, Git, Matlab, Stata
**Coursera Certificates:** IBM Data Science Specialization, Stanford Machine Learning Specialization