

SHIXUAN LIN

shixuan.lin@nyu.edu | shixuanlin.com | Qingdao, China

PROFILE

Recent graduate conducting independent applied research in AI safety evaluation, content safety, and agentic AI security. Independently built a multilingual system for toxic and xenophobic language detection and an AI Safety Testing Framework covering standards mapping, threat modeling, adversarial testing, multidimensional metrics, and mitigation validation. Current work examines failure modes in tool-using LLM agents and translates technical findings into operational controls and digital workflows.

EDUCATION

New York University Sep. 2023 – Jan. 2025
Master of Science in Management and Systems, Concentration in Database Technologies New York, NY, U.S.A.

GPA: 3.90/4.00 **Selected Coursework:** Database Design and Management; Information Technology and Data; Information Security Management; Quantitative Models for Decision-Making.

Macau University of Science and Technology Sep. 2019 – Jun. 2023
Bachelor of Business Administration in Finance Macau SAR, China

GPA: 3.42/4.00 **WES GPA:** 3.70/4.00 **Selected Coursework:** Corporate Finance; Asset Pricing; Financial Derivatives; Financial Econometrics; Banking Management; Operations Management.

Fudan University Sep. 2022 – Dec. 2022
Hong Kong and Macau Ten Thousand Student Interflow Programme Shanghai, China

Selected Coursework: Python Programming and Econometrics.

RESEARCH AND PROJECTS

Safety Evaluation of Tool-Using LLM Agents Jul. 2026 – Present
Independent Researcher Remote

- Ongoing independent research on safety evaluation for tool-using language-model agents.
- The public record describes the scope only; methods, traces, findings, code, data, and manuscripts are not disclosed.

AI Safety Testing Framework for LLM and Agentic Systems Sep. 2025 – Dec. 2025
Independent Researcher and Project Lead Remote

- Designed a four-layer evaluation framework mapping the EU AI Act, NIST AI RMF, ISO/IEC 42001 and 23894, and UN principles to operational testing requirements for LLM and agentic systems.
- Defined critical capability levels, risk tiers, seven trustworthiness dimensions, adversarial test categories, safety and deployment levels, and mitigation-verification procedures.
- Produced a testing playbook, compliance scorecard, safety-case and risk-analysis templates, and a web prototype; piloted the workflow on three case-study systems.

AI-Driven Media Analysis for Toxic and Xenophobic Language Detection Aug. 2024 – Dec. 2024
Independent Researcher, UNICC-Sponsored NYU Applied Project New York, NY, U.S.A.

- Built a multilingual NLP system using GPT-4 and a fine-tuned BERT model to detect explicit and implicit xenophobic language, toxicity, misinformation, and biased framing in media content.
- Developed preprocessing and evaluation pipelines for PDF, Word, and plain-text inputs, including toxicity categorization, sentiment analysis, 1–10 severity scoring, fact and terminology validation, and topic classification.
- Integrated three modules into a Streamlit web application with API-based workflows, achieving 87% xenophobic-language detection accuracy, 90% fact and language validation accuracy, 85% thematic categorization accuracy, and sub-two-second response time.

Forecasting the Term Structure of Interest Rates with Dynamic SV Models Dec. 2022 – Jul. 2023
First Author Macau SAR, China

- Collected and cleaned ten years of Chinese bond-market data and constructed six time-series models to analyze and forecast interest-rate behavior.
- Used Python, R, SPSS, Excel, and LaTeX for data preparation, modelling, statistical analysis, visualization, and academic writing.

The Determinants of Carbon Emissions in Macau

Co-Author and Empirical Researcher

Jul. 2022 – Feb. 2023

Macau SAR, China

- Conducted data collection, factor decomposition, model testing, and comparative analysis using STIRPAT, EKC, and LMDI models in Stata and EViews.
- Contributed to the methodology, empirical analysis, interpretation, and conclusions; the research was published by Springer Nature.

PROFESSIONAL EXPERIENCE

Honghuayuan Fabric

Product and Operations Manager

Jan. 2025 – Present

Qingdao, China

- Lead end-to-end development of a B2B digital platform (WeChat Mini Program, back-office system, company website): backend (Fastify, Prisma, PostgreSQL, Redis), deployment (PM2, Nginx), and a 34-page uni-app/Vue 3/TypeScript Mini Program.
- Design the workflow and data architecture across inventory, procurement, production, sales, and finance modules; own the full product lifecycle from requirements through deployment and adoption.

Greenworks

Market Strategy Analyst, NYU Real World Client Project

Sep. 2024 – Dec. 2024

New York, NY, U.S.A.

- Analyzed industry dynamics, competitors, consumer preferences, and market-entry opportunities in North American and European outdoor power-equipment markets.
- Developed recommendations for customer segmentation, product positioning, pricing, promotion, and sales-channel strategy and presented findings to company and academic stakeholders.

HMD Global U.S.A.

Market Data Analyst Intern

May 2023 – Aug. 2023

San Diego, CA, U.S.A.

- Conducted market and product analysis for Nokia mobile devices, identifying customer preferences, competitive trends, and product opportunities.
- Used Python and SQL for web data collection, cleaning, transformation, querying, and analysis of sales, market, and customer-feedback datasets.
- Managed internal databases and analyzed website traffic metrics to produce visual reports for management decision-making.

Shanghai Longyi Enterprise Management Group

Consulting Intern

May 2022 – Jun. 2022

Shanghai, China

- Analyzed China's sugar-free food and beverage industry and estimated market size using corporate disclosures, industry reports, and public secondary data.
- Investigated consumer preferences, raw-material R&D trends, policies, and supply chains; contributed to a 9,930-word consulting report.

Huaxia Bank, Qingdao Branch

Key Institutions Business Department Intern

Jul. 2021 – Aug. 2021

Qingdao, China

- Analyzed branch-issued bonds and collected financial data for three key institutional clients to support internal financial review.
- Organized corporate records and assisted with debt-maturity analysis, meeting minutes, and reporting documentation.

PUBLICATION

Liang, W., Pan, R., Cao, Y., Wang, W., **Lin, S.**, and Zhao, Y. (2023). "The Determinants of Carbon Emissions in Macau: Based on the Analysis of STIRPAT, EKC, and LMDI Models." *International Conference on Informatics Engineering and Information Science*, Springer Nature Singapore, pp. 152–161. DOI: 10.1007/978-981-99-3618-2_15.

TECHNICAL SKILLS

Programming and Data: Python, SQL, MySQL, PostgreSQL, R, JavaScript, HTML/CSS, Pandas, NumPy, Matplotlib, Streamlit, PyMuPDF, python-docx, SPSS, Stata, EViews, MATLAB, Excel.

AI Safety and NLP: LLM evaluation, BERT, GPT-4, text classification, toxicity and xenophobia detection, sentiment analysis, topic modeling, prompt injection testing, jailbreak testing, RAG security, agent safety, adversarial test design, risk taxonomies, and experimental design.

Product and Research Tools: Product requirements, business-process modelling, workflow design, acceptance testing, data governance, Git, JupyterLab, Bloomberg, Wind, Visio, and Microsoft Office.

ACTIVITIES AND LANGUAGES

Memberships: China Society of Image and Graphics; Systems Engineering Society of China.

Languages: Mandarin Chinese (Native); English (Professional working proficiency).