

Xuan Chen

Graduate Research Assistant @ Duke University | Databases Focus

 Homepage: cx0222.cn  GitHub: [cx0222](https://github.com/cx0222)  E-mail: xc241@duke.edu

RESEARCH INTERESTS

- Query Optimization
- Vector Databases (for RAG)
- Data Center Networks

PUBLICATION

ONE-PASS ROBUST QUERY OPTIMIZATION VIA STRUCTURE-AWARE SUBPLAN REUSE (IN PREPARATION)

Authors: **Xuan Chen**, Haibo Xiu, Jun Yang

Target Conference: 2025 ACM SIGMOD International Conference on Management of Data

DELTA-AUG: CROSS-MODAL HARD FEATURE MINING FOR FEW-SHOT LEARNING

Authors: **Xuan Chen**

Conference: 2024 6th Asia Conference on Machine Learning and Computing

RESEARCH PROJECTS

ROBUST QUERY OPTIMIZATION IN DATABASE SYSTEMS

Advisor: Prof. Jun Yang, Dept. of Computer Science, Duke University | May 2025 - Present | Durham, NC

- Pioneering new paradigms for cost-based query optimization under uncertainty, rethinking how optimizers make robust decisions in dynamic data environments.
- Exploring next-generation frameworks for robust and adaptive query planning, with emphasis on lightweight, efficient, and structure-aware approaches.
- Expanding the boundaries of PostgreSQL's optimizer to enable richer plan spaces and more robust query execution strategies.

MINI-AMAZON: A DATABASE-CENTRIC E-COMMERCE SYSTEM

Advisor: Prof. Jun Yang, Dept. of Computer Science, Duke University | Jan 2025 - May 2025 | Durham, NC

- Built a full-stack data system integrating PostgreSQL with pgvector to enable Retrieval-Augmented Generation (RAG) for large language model applications.
- Enhanced the system with real-time analytics and advanced search, using Spark Streaming + Kafka for data streaming and Lucene for multi-field fuzzy retrieval.

CONTENT MODERATION IN WEB AND NETWORKED SYSTEMS

Advisor: Prof. Xiaowei Yang, Dept. of Computer Science, Duke University | Nov 2024 - Present | Durham, NC

- Developing a system based on the novel LLM(VLM)-as-a-judge concept to evaluate online content across text, image, and video modalities.
- Building a full ML system pipeline encompassing data acquisition, cleaning, VLM tuning, and prompt/inference evaluation focusing on performance improvement.

CROSS-MODAL HARD FEATURE MINING FOR FEW-SHOT LEARNING

Central University of Finance and Economics | Jun 2023 - Jan 2024 | Beijing, China

- Identified few-shot learning challenges in VLMs, highlighting hard classes' impact.
- Proposed DeltaAug – a cross-modal augmentation strategy leveraging CLIP features to improve performance, achieving state-of-the-art results across benchmarks.

EDUCATION

DUKE UNIVERSITY

MASTER OF SCIENCE IN COMPUTER SCIENCE

Aug 2024 - Present | Durham, NC

Overall GPA: 3.81/4.0

CENTRAL UNIVERSITY OF FINANCE AND ECONOMICS

BACHELOR OF DATA SCIENCE
(GUOZHI HSU TALENT PROGRAM IN DATA SCIENCE AND BUSINESS ANALYTICS)

Sep 2020 - Jun 2024 | Beijing, China

Overall GPA: 3.73/4.0 (WES)

HONORS

PROGRAMMING

2023 – The 6th “Hello World” Algorithm and Programming Competition, Special Prize (Team Award, 1st Place)

MATHEMATICS

2022 – Mathematics Competition for Chinese College Students, Second Prize

ACADEMIC RESEARCH

2021 – Academic Research Scholarship, First Prize

SKILLS

ACADEMIC

- C, Linux and DB Kernels
- Data Modelling and Visualization
- LLMs and VLMs
- Python and PyTorch

DEVELOPER

- PostgreSQL & MySQL
- Java & JDBC
- React Native & Vue 3