

# Zhencan Peng

✉ [zp128@scarletmail.rutgers.edu](mailto:zp128@scarletmail.rutgers.edu), [pzcdmm@gmail.com](mailto:pzcdmm@gmail.com)

☎ +1 (848) 313-8660

---

## EDUCATION

**Rutgers University**, New Brunswick, NJ, USA

*Ph.D. in Computer Science*

Sep. 2022 – 2026 (expected).

**Shenzhen University**, Shenzhen, China

*B.Eng. in Computer Science, GPA: 4.17/4.5, Rank: 3/177*

Sep. 2018 – Jul. 2022

---

## RESEARCH INTERESTS

- Text alignment and deduplication, especially for large-scale LLM pretraining datasets
- Vector databases and approximate nearest neighbor (ANN) search
- Sparse attention mechanisms for efficient LLM inference

---

## WORK EXPERIENCE

**Huawei 2012 Lab**, Shenzhen, China

*AI Engineer Intern*

Jul. 2025 – Present.

- Improved data deduplication techniques for large-scale pretraining datasets in Noah's Ark Lab

**Alibaba Cloud**, Hangzhou, China

*Research Intern*

Jun. 2024 – Dec. 2024

- Submitted a paper on dynamic range-filtering vector search to VLDB 2025

---

## PUBLICATIONS

**Near-Duplicate Text Alignment under Weighted Jaccard Similarity**

Yuheng Zhang, Miao Qiao, **Zhencan Peng**, Dong Deng

SIGMOD 2026 – ACM International Conference on Management of Data (Submitted)

**LSHAlign: All-Pair Near-Duplicate Text Alignment via Locality-Sensitive Hashing**

Yuheng Zhang\*, **Zhencan Peng\***, Miao Qiao, Dong Deng

SIGMOD 2026 – ACM International Conference on Management of Data (In Preparation)

**Dynamic Range-Filtering Approximate Nearest Neighbor Search**

**Zhencan Peng**, Miao Qiao, Wenchao Zhou, Feifei Li, Dong Deng

VLDB 2025 – International Conference on Very Large Data Bases

**Flux: Unifying Heterogeneous Infrastructure for Alibaba AnalyticDB**

Wei Li, Jiachi Zhang, Ye Yin, Yan Li, Zhanyang Zhu, Yuhao Li, **Zhencan Peng**, Lan Lu, Wenchao Zhou, Liang Lin, Feifei Li

SIGMOD 2025 – ACM International Conference on Management of Data (Industry Paper)

**Near-Duplicate Text Alignment with One Permutation Hash**

**Zhencan Peng**, Yuheng Zhang, Dong Deng

SIGMOD 2025 – ACM International Conference on Management of Data

**Near-Duplicate Sequence Search at Scale for Large Language Model Memorization Evaluation**

**Zhencan Peng**, Zhizhi Wang, Dong Deng

SIGMOD 2023 – ACM International Conference on Management of Data

**ViObject: Harness Passive Vibrations for Daily Object Recognition with Commodity Smartwatches**

Wenqiang Chen, Shupef Lin, **Zhencan Peng**, Farshid Salemi Parizi, Seongkook Heo, Shwetak Patel, Wojciech Matusik, Wei Zhao, John Stankovic

IMWUT 2024 – ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 8, Issue 1

## **Robust Finger Interactions with COTS Smartwatches via Unsupervised Siamese Adaptation**

Wenqiang Chen, Ziqi Wang, Pengrui Quan, **Zhencan Peng**, Shupeí Lin, Mani Srivastava, Wojciech Matusik, John Stankovic  
UIST 2023 – ACM Symposium on User Interface Software and Technology

## **NF-Heart: A Near-field Non-contact Continuous User Authentication System via Ballistocardiogram**

Yandao Huang, Minghui Qiu, Lin Chen, **Zhencan Peng**, Qian Zhang, Kaishun Wu

IMWUT 2023 – ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 7, Issue 1

## **Making Vibration-based On-body Interaction Robust**

Wenqiang Chen, Ziqi Wang, Pengrui Quan, **Zhencan Peng**, Shupeí Lin, Mani Srivastava, John Stankovic

ICCPs 2022 – ACM/IEEE International Conference on Cyber-Physical Systems

## **Toward Device-free and User-independent Fall Detection Using Floor Vibration**

Kaishun Wu, Yandao Huang, Minghui Qiu, **Zhencan Peng**, Lu Wang

TOSN 2023 – ACM Transactions on Sensor Networks, Vol. 19, Issue 1

---

## **AWARDS**

Travel Fellowship — VLDB 2025 Travel Award	2025
Travel Fellowship — SIGMOD 2023 Travel Award	2023
Gold Prize — National Undergraduate Algorithm Design and Programming Challenge	2021
Second Prize — China Collegiate Computing Contest (4C)	2021
Silver Prize — CCF Collegiate Computer Systems & Programming Contest	2020
Bronze Prize — ACM-ICPC Asia Regional Contest	2020
National Scholarship — Awarded by Ministry of Education, P.R. China	2020
Bronze Prize — CCPC Regional Contest (Harbin Site)	2019

---

## **SERVICES**

CS210: Data Management for Data Science — <i>Teaching Assistant</i>	2024
SIGMOD 2023 Student Programming Contest — <i>Organization Member</i>	2023
CS213: Software Methodology — <i>Teaching Assistant</i>	2023
CS539: Database Management Systems — <i>Teaching Assistant</i>	2022