SHUXIN (ERICA) ZHOU
ericashuxin@gmail.com | LinkedIn

EDUCATION

New Jersey Institute of Technology
PhD of Computer Science — Jan. 2020 — Dec. 2024
New York Institute of Technology
Master of Computer Science with ***Distinction*** — Jan. 2017 — Dec. 2019
Arkansas State University
Master of Public Administration — Aug. 2012 — Aug. 2015
Xiamen University
Bachelor of Management (Financial Management) — Oct. 2007 — Jul. 2011

EXPERIENCE

St. Francis College — Brooklyn

Assistant Professor of Computer Science and Cybersecurity Department — Sep. 2025 —

- Teach for undergraduate courses: IT-2105 Programming II and IT-2510 Database Management System.
- Teach for graduate courses: CS-6001 Foundations of Programming, CS-6002 Computer Organization, CS-6003 Analysis of Algorithms, CYB-6000 Critical Infrastructure Security Foundation, CYB6100 Transportation Systems Cybersecurity
- Designed and developed several core courses for Computer Science Master Program and Cybersecurity Master Program.
- Collaborate with designing & developing of Ph.D. in IT program
- Serve for Curriculum Committee for reviewing the developed courses

Senior Adjunct Lecturer — Jan. 2025 — Aug. 2025

- Taught for graduate courses: CS-6001 Foundations of Programming, CS-6002 Computer Organization, CS-6003 Analysis of Algorithms, CS-6004 Programming Languages and CS-6006 Software Engineering.
- Designed and led a Data Structure and Algorithm course focused on algorithmic logic, data structure manipulation, and analytical reasoning, preparing students for advanced roles in data analysis and machine learning.
- Guided project-based learning to develop skills in algorithm design, program efficiency, and modular coding, equipping students with problem-solving techniques for data-driven applications.
- Combined real-world programming challenges to strengthen students' understanding of IT and software engineering concepts, essential for effective data system management.

New York Institute of Technology — New York, Jan. 2025 — May. 2025

Adjunct Assistant Professor

- Designed and taught “Introduction of Data Science Programming” course that integrated ethical considerations in data science, covering basic knowledge of python programming, object-oriented programming, and responsible usage in alignment with industry standards and societal expectations.

- Combined theoretical instruction with hands-on exercises to strengthen critical thinking, statistical analysis, and data interpretation skills, preparing students to solve complex, real-world data challenges.
- Supervised group projects focused on data collection and pre-processing, rigorous statistical analysis, and compelling data visualization, highlighting the role of data science and analytics in addressing societal issues and driving meaningful impact.

New Jersey Institute of Technology Newark, Jan. 2020 — Dec. 2024

Researcher in SABOC Lab

- Design and develop an Optimized Cluster-focused Combination annotator for annotating the EHRs text by referencing the user-defined interface terminology which achieve significant improvement on the text coverage and time efficiency by reducing the annotation execution time to second level compared to traditional annotation tools.
- Developed Clinical and Medical Ontology Hierarchical Relation Predicting Pipeline with Large Language Models (LLM) which achieved a notable 96% macro accuracy with the preliminary research on varied LLM such as: Llama2-7b/2-13b, Openchat_3.5, Vicuna-13b-v1.5, Zephyr-7b-beta.
- Extracted the medical relevant phrases from the open-sourced Cardiology patient notes and Covid-19 Radiology notes by using the self-designed concatenation and anchoring methodology.
- Constructed the cardiology and Covid-19 interface ontology using protégé for annotation of the electronic health records. Experimented with the Mimic public source, EHR database.
- Developed a hybrid (rule-base & machine learning--using spaCy/NLTK) Concept Extraction tool for Covid19 terminology.

Lab Instructor

- Instructed for CS103 Business Analysis with Python (Undergraduate Level)
- Instructed for CS115 Introduction of Computer Science: C++ programming. (Undergraduate Level)

Tutor

- Tutored for CS331 Database course (Undergraduate Level)
- Tutored for CS631 Database course (Graduate Level)

Teaching Assistant

- Assisted with teaching for CS331 Database Systems (Undergraduate Level)
- Assisted with teaching for CS634 Data Mining (Graduate Level)
- Assisted with teaching for CS413 Advanced Database (Undergraduate Level)
- Assisted with teaching for CS631 Database Management (Graduate Level)

Regeneron Pharmaceuticals Tarrytown, Jan. 2023 — Aug. 2023

Biomedical Informatics Scientist Co-op

- Designed & Developed Name Entity Recognizer (NER) Plus Ontology Resolution Pipeline for identifying disease, cell and tissue related phrases from the large gene expression text dataset (Gene Expression Omnibus) and resolve such free-form phrases to existing Disease, Cell and Tissue ontologies

concepts by using Spark NLP Pre-trained Language Model on Databrick which achieved 95% F1 score outperformed among all the other state-of-arts NLP models.
- Designed & Developed Disease Abbreviation Identification Pipeline to detect and disambiguate the meaning of disease abbreviation in the GEO dataset.
- Researched and trained on LDA Topic Modeling on classifying disease term in terms of Disease Ontology.
- Researched and developed prompts for BioGPT Pre-trained Model for medical abbreviation detecting and ambiguating.

New York Institute of Technology New York NY, Jan. 2019 — Dec. 2019

Graduate Assistant

- Researched on optimizing data retrieval performance of SNOMED CT via deep learning.
- Evaluated the textual search performance based on the collected data.
- Scrapped (dynamic loading) the data from SNOMED CT browsers using python.
- Tutored for Data Structure, Introduction of Data Mining, and Database Systems courses.

Montefiore IT Yonkers NY, Jul. 2019 — Sep. 2019

Information Security Engineer Intern

- Integrated Montefiore HR staging system with Microsoft Identity Manager (MIM) and optimizing the user authentication modules by using Data Parallelism and Dataflow on .Net platform;
- Set the websites access policy, and designed the content filter, internet access policies using Blue Coat Proxy and assisted with upgrading proxy version and pulling over policies to multiple servers; Customized the policies based on different using groups and different devices;
- Used IBM QRadar and XForce (SIEM) monitoring cybersecurity events for updating the blocked list.
- Built Microsoft SharePoint portal for Information Security Department with two versions: public (merged to company entire site for only viewing) and internal (able to modify by authorized internal users, such as department manager), which has features including: Wiki Knowledge Base, Exception Tracker Flow, Project Tracker, Server Status Tracker, etc.

PUBLICATIONS

S. Zhou, M. Dehkordi, W. Yao, H. Liu, P. Sen, F. Deek, and Y. Perl, "Enhancing Electronic Health Records Annotation with a Cluster-Focused Combination Algorithm and Interface Terminologies," in *HEALTHINF 2025*, Springer, Under Processing.

M. K. H. Dehkordi, **S. Zhou**, Y. Perl, F. Deek, H. Zhe, Vipina K Keloth, H. Liu, G. Elhanan, L. Luke, and A.J. Einstein. Curation of a Cardiology Interface Terminology for Highlighting Electronic Health Records Using Machine Learning. BMC Medical Informatics and Decision Making. 2025. (Pending Revision)

S. Zhou, P. Sen and H. Liu, Y. Perl, M. K. H. Dehkordi. "CFC Annotator: A Cluster-Focused Combination Algorithm for Annotating Electronic Health Records by Referencing Interface Terminology". In Proceedings of the 18th International Joint Conference on Biomedical Engineering

Systems and Technologies - Volume 2: HEALTHINF, ISBN 978-989-758-731-3, ISSN 2184-4305, pages 195-206. DOI: 10.5220/0013244500003911

M. K. H. Dehkordi, **S. Zhou**, Y. Perl, F.P. Deek, A.J. Einstein, G. Elhanan, Z. He, H. Liu, "Enhancing patient Comprehension: An effective sequential prompting approach to simplifying EHRs using LLMs." *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2024): 6370-6377.

N. M. Kollapally, M. K. H. Dehkordi, Y. Perl, J. Geller, F.P. Deek, H. Liu, V. K. Keloth, **S. Zhou**, "Using clinical entity recognition for curating an interface terminology to aid fast skimming of EHRs." *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2024): 6427-6434.

H. Liu, **S. Zhou*,** Z. Chen, Y. Perl and J. Wang, "Using Generative Large Language Models for Hierarchical Relationship Prediction in Medical Ontologies," *2024 IEEE 12th International Conference on Healthcare Informatics (ICHI)*, Orlando, FL, USA, 2024, pp. 248-256, doi: 10.1109/ICHI61247.2024.00040.

S. Zhou*, M. K. H. Dehkordi*, Y. Perl, J. Geller, A. J. Einstein, G, Elhanan, V. K. keloth, et al., "Using annotation for computerized support for fast skimming of cardiology electronic health record notes," *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Istanbul, Turkiye, 2023, pp. 4043-4050, doi: 10.1109/BIBM58861.2023.10385289.

V.K. Keloth*, **S. Zhou***, L. Lindemann, L. Zheng, G. Elhanan, A.J. Einstein, J. Geller, Y. Perl. Mining of EHR for interface terminology concepts for annotating EHRs of COVID patients. BMC Med Inform Decis Mak. 2023 Feb 24;23(Suppl 1):40. doi: 10.1186/s12911-023-02136-0. PMID: 36829139; PMCID: PMC9951157.

V.K. Keloth*, **S. Zhou***, L. Lindemann, L. Zheng, G. Elhanan, A.J. Einstein, J. Geller, Y. Perl. "Mining Concepts for a COVID Interface Terminology for Annotation of EHRs," *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA, 2020, pp. 3753-3760, doi: 10.1109/BigData50022.2020.9377981.

V.K. Keloth, **S. Zhou**, G. Elhanan, Y. Chen, J. Geller, Y. Perl. "Generating Training Data for Concept-Mining for an 'Interface Terminology' Annotating Cardiology EHRs," *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Seoul, Korea (South), 2020, pp. 1728-1735, doi: 10.1109/BIBM49941.2020.9313435.

S. Zhou, C. Chen and H. H. Gu, "Testing and Evaluating SNOMED CT Web Browsers' Textual Search Feature" *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, CA, USA, 2019, pp. 1954-1961, doi: 10.1109/BIBM47256.2019.8983355.

RESEARCH PROJECT EXPERIENCE

- **Disease, Cell & Tissue Phrases Identification & Ontology Resolution (Regeneron Project).** Identifying the disease, cell and tissue concepts from the free text dataset, and resolved such phrases to

controlled ontology concepts by using NER Model (feature extractor using Bert Embedding & sBert Embedding & Bert Chunk Embedding) and Ontology Resolution Model using Jaccard/Cosine/Euclidean distance function.

- **Cardiology Comprehensive Terminology Constructing via Mining the Mimic-III Cardiology Event Notes.** Collecting and processing/cleaning Mimic-III event notes, extracting concepts by using designed pipeline (mentioned in COVID-19 project) for constructing a comprehensive Cardiology terminology graph.
- **COVID-19 High Granular Chunk Extracting & Annotating System.** Mined the covid-19 concepts from the radiology notes, generated higher granular phrases (terminology) via an innovated hybrid approach – designed "concatenation" and "anchoring" algorithms with NLTK stopping word system; Designed an automatic terminology reviewing system (using TextRank and spaCy) to collect valid terminology and annotated them via calling BioPortal/Mgrep annotator function resulting in html format.

TECHNICAL SKILLS

- **Programming Languages:** Java, Python, SQL, C++
- **Machine Learning (Natural Language Processing):**
  - ML Platform & Packages: NLTK, SpaCy, TensorFlow, PyTorch, Scikit-Learn, SparkNLP, Spark
  - NLP Models: Name Entity Recognition Model, Ontology Resolution, BERT, Large Language Models
- **Web Technologies:** JavaScript, XHTML+CSS, AngularJS
- **Database:** MySQL, Oracle pl/sql(SQLdeveloper), MongoDB(Studio3T)
- **Data Analysis Packages (Python):** numpy, pandas, scipy, matplotlib, umap, scikit-learn
- **Others**: Ontology Construction, Protégé

COURSE QUALIFED TO TEACH

- **Introduction to Data Mining**
  - A course introduces basic techniques in data mining exploring the data preprocessing, data analysis, classification, clustering et al.
- **Machine Learning and Natural Language Processing**
  - Advanced Level course covering different tasks and topics in natural language processing, include sequence-base ML models, attention-base models and convention ML models
- **Database Management**
  - A course on database design, management, and optimization, with a focus on modern database technologies.
- **Programming Language (Java or Python)**
  - A course on programming languages with a focus on Python or SQL, covering syntax, data structures, and practical applications.

OTHERS

- **Certificate**
  - Natural Language Processing Specialization / OpenAI. License number: U329EE4GBKVF
  - Machine Learning/Stanford University. License number: CQT2YZ6MR6EF
  - Biomedical Research Investigators/MIT. License number: 35131660
- **Review Services**
  - AMIA-Annual Symposium 2021/2022/2023/2025/2026: Invited Peer Reviewer;
  - AMIA-CIC 2024: Invited Peer Reviewer
  - AMIA-Amplify 2026: Invited Peer Reviewer
  - IEEE-ICHI 2023/2024/2026: Invited Reviewer
  - IEEE-CHASE 2024/2025/2026: Invited Reviewer
  - IEEE-ICC 2026: Invited Reviewer
- **Other Services**
  - IEEE-BIBM 2024/2025/2026: Invited Program Committee
  - IEEE-ICHI 2025: Invited Program Committee
  - BIOSTEC 2025 – BIODEVICES Conference Session Chair
  - BIOSTEC 2025 – HEALTHINF Conference Session Chair
- **Patent**
  - Curating an interface terminology for annotation of Electronic Health Records (EHRs) of a medical discipline. US No. 63/232,872
- **Presentation**
  - **HEALTHINFO 2025, "**CFC Annotator: A Cluster-Focused Combination Algorithm for Annotating Electronic Health Records by Referencing Interface Terminology"
  - **BIBM 2019, "**Testing and Evaluating SNOMED CT Web Browsers' Textual Search Feature"
- **Award**
  - New York Institute of Technology 2019 Computer Science Graduate Faculty Award
  - St. Francis College 2026 Faculty Research Award