

YUNSHUN ZHONG

35 St George St, Toronto, ON M5S 1A4 | +1 (437)-776-7220 | yunshun.zhong@mail.utoronto.ca

EDUCATION

DOCTOR OF PHILOSOPHY: Civil and Mineral Engineering <i>University of Toronto</i>	Expected Completion: 05/2025 Toronto, ON
MASTER OF SCIENCE IN ENGINEERING: Civil and Environmental Engineering <i>University of California, Berkeley (UCB)</i> <i>Graduated with Graduate Certificate in Applied Data Science</i>	05/2020 Berkeley, USA
HONOURS BACHELOR OF ENGINEERING: Civil Engineering <i>Harbin Institute of Technology (HIT)</i>	07/2018 Harbin, China
EXCHANGE STUDY <i>University of California, Berkeley (UCB)</i>	08/2016-08/2017 Berkeley, USA

TECHNICAL SKILLS

Programming Languages: *Advanced:* Python, Java, R, SQL | *Proficient:* MATLAB, C++

Data Science & Machine Learning (ML): Python (eg. Scikit-learn, Statsmodels, Numpy, SciPy, Pandas, Seaborn, OpenCV, Scikit-image, Optuna), A/B testing, Data science pipeline (cleansing, wrangling, visualization, modelling, interpretation), Time series, Hypothesis testing, OOP, Git, Cloud Platform (GCP), Spark

Natural Language Processing: Pretrain and fine-tune large language models (LLMs), LLM in-context learning, retrieval augmented generation (RAG) (Huggingface, PyTorch, Tensorflow, NLTK, OpenAI, Langchain)

Learning Management Systems (LMS): Canvas, Blackboard, Markus, Moodle

SOFT SKILLS

- University teaching & curriculum development
- Graduate & Post Graduate mentoring
- Higher education & student-centered Learning
- Cultural awareness, equality, diversity, and inclusion
- Academic research analysis and scholarly writing
- Critical thinking and conflict resolution

PUBLICATIONS

Published and Submitted Journal Paper:

Zhong, Y., & El-Diraby, T. (2025). Leveraging Large Language Models with Retrieval-Augmented Generation for Trend Analysis in Construction Management Research. *Automation in Construction* (under review).

Zhong, Y., & El-Diraby, T. (2025). An Agency-specific Project Authoring Advisor: An LLM-based RAG System. *Advanced Engineering Informatics* (under review).

Zhong, Y., & Goodfellow, S. D. (2024). Domain-specific language models pre-trained on construction management systems corpora. *Automation in Construction*, 160, 105316.

Wang, H., Zhong, P., Xiu, D., Zhong, Y., Peng, D., & Xu, Q. (2022). Monitoring tilting angle of the slope surface to predict loess fall landslide: an on-site evidence from Heifangtai loess fall landslide in Gansu Province, China. *Landslides*, 1-11.

Wang, H., Zhuo, T., Zhong, P., Wei, C., Zou, D., & Zhong, Y. (2021). A novel wireless underground transceiver for landslide internal parameter monitoring based on magnetic induction. *International Journal of Circuit Theory and Applications*, 49(6), 1549-1558.

Mukai, K., Zhong, Y., Hubbard, P., & Soga, K. (2021). A preliminary study of environmental monitoring using embedded sensors in the soil. *Japanese Geotechnical Society Special Publication*, 9(5), 164-168.

Wang, H., Nie, D., Tuo, X., & Zhong, Y. (2020). Research on crack monitoring at the trailing edge of landslides based on image processing. *Landslides*, 17(4), 985-1007.

Conference Paper and Patent:

Zhong, Y., & El-Diraby, T. (2024). Generative project question answering system: triangulating three approaches for project authoring. *18th buildingSMART International Summit*.

Zhong, Y., & El-Diraby, T. (2022). Shoreline Recognition Using Machine Learning Techniques. *IOP Conference Series*:

RESEARCH AND PROFESSIONAL EXPERIENCE

NLP (Retrieval Augmented Generation) Research Assistant **07/2023-present**

Toronto Region Conservation Agency (TRCA)

Toronto, ON

Developed a Generative Question Answering (QA) System for TRCA's Technical Documents leveraging RAG and LLMs.

- Engineered a QA System: Designed and implemented a QA system using retrieval-augmented generation, combining large language models with an advanced information retrieval system.
- Team Management: Managed a team of graduate and undergraduate assistants, overseeing their daily routine research processes and ensured the operation and completeness of projects.
- Optimized Data Processing: Conducted data cleaning and preprocessing for TRCA's technical documents, ensuring a clean, accessible corpus for the QA system.
- Innovated in Prompt Engineering: Developed and optimized passage retrieval mechanisms and domain-specific prompt engineering to guide LLMs in generating accurate, context-aware responses.
- Developing User-Friendly GUI: Spearheaded the creation of an interactive GUI for the QA system to enhance user engagement, accessibility, and personalized interaction through a chatbot interface.

Data Scientist Intern – Global Risk Analytics Team

05/2024-08/2024

Royal Bank of Canada

Toronto, ON

Developed geospatial data pipelines, ML models, and wildfire simulation framework to predict future fire risk and assess economic impact under future climate scenarios.

- Data Collection and Dataset Curation: Sourced, cleaned, and analyzed global geospatial data, including fire intensity, FWI, elevation, and land cover from satellite and reanalysis sources; merged and aligned data, and grouped it into fire events using DBSCAN and time-based clustering for ML model training.
- ML Model Development: Trained kernelized logistic regression and XGBoost models to predict future FWI under various climate scenarios, optimized with Optuna and geospatial cross-validation.
- Wildfire Simulation: Built a wildfire season simulation framework by modifying source code of CLIMADA library and incorporating ML, probabilistic models and cellular automaton to assess future fire risk, supported enhanced economic impact assessments for wildfire management.

Data Scientist Intern – Fix Income Team

01/2024-05/2024

Guardian Capital

Toronto, ON

Developed ML and deep learning pipelines for predicting Non-Farm Payroll (NFP) and US Treasury Yield, focusing on enhancing predictive accuracy and efficiency.

- Constructed ML Pipeline: Designed and implemented an end-to-end ML pipeline, incorporating various models such as Kernelized Linear Regression, Histogram-based Gradient Boosting Regressor, and Random Forest Regressor, meeting an urgent and competing deadline while maintaining high accuracy and quality.
- Implemented Advanced Data Processing: Integrated sophisticated data processing methods, utilizing roll-forward partitioning and lagged features to clean and prepare data for optimal model performance.
- Optimized Model Performance: Conducted hyperparameter optimization using various search methods such as cross validation with halving random search to fine-tune the models and improve predictive accuracy, achieved significant predictive accuracy improvement.
- Expanding with Deep Learning: utilized pre-trained transformer models (TimeGPT1) to predict NFP time series data, aiming to leverage the latest advancements in artificial intelligence for financial time series forecasting.

NLP (LLM pre-training and fine-tuning) Research Assistant

05/2022-06/2023

University of Toronto

Toronto, ON

Developed the first corpus in construction management domain and pre-trained domain-specific large language models

- Corpus Development: Successfully collected, cleaned, and pre-processed data from 732 journal papers to develop a comprehensive domain-specific corpus for construction management.
- Model training: Engineered an end-to-end pipeline for the pre-training and fine-tuning of domain-specific pre-trained models (PTMs), optimizing with minimal preprocessing and hyperparameter adjustments.

- ♦ Performance Improvement in NLP Tasks: Achieved significant improvements in Text Classification (TC) and Named Entity Recognition (NER) tasks within the construction management systems (CMS) domain, increasing the F1 score by 5.9% and 8.5% to 75.3% and 95.4%, respectively.
- ♦ Domain-Specific Model Advancements: Contributed to the advancement of NLP applications in the CMS domain by obtaining domain-specific PTMs that demonstrate enhanced performance on specialized tasks.

Machine Learning Research Assistant

05/2021-03/2022

University of Toronto

Toronto, ON

Develop automatic shoreline recognition system from aerial images using advanced ML algorithms

- ♦ Shoreline images are labeled and pre-processed to be used in ML algorithms.
- ♦ Random Forest, XGBoost, and LGBM algorithms are implemented in shoreline detection. The averaged semantic segmentation accuracy for the above algorithms is 95.6%, 96.0%, and 94.8%, respectively.
- ♦ Enhanced final shoreline accuracy by post-processing and applying Gaussian Edge Detection algorithm on results from ML algorithms.

Crack Monitoring at the Trailing Edge of Landslide Based on Image Processing

08/2018-12/2019

Wireless Sensor Network Research Group (Supervised by Prof. Kenichi Soga and Prof. Honghui Wang)

- ♦ Processed the trailing edge of a landslide (TEL) images including image preprocessing, OTSU algorithm, and Canny edge detection to identify the outline of the TEL.
- ♦ Analyzed crack motion through azimuth and displacement measurements.
- ♦ Calculated the changes before and after the motion of cracks using interval median comparison algorithm.
- ♦ Applied our method to a 3D simulation model, a gravel model, a soil model, and a collapsed body.

Green Infrastructure Monitoring with Embedded and Fiber Optic Sensors

08/2018-05/2019

Wireless Sensor Network (WSN) and Distributed Fiber Optic Research Group (Supervised by Prof. Kenichi Soga)

- ♦ Tested and analyzed WSN signal propagation in soil under varying saturation levels.
- ♦ Conducted laboratory experiments to measure and analyze daily strain changes in the soil using WSN units and fiber optic sensors.
- ♦ Designed and 3D-printed anchors to mobilize fiber optic sensors integration in soil monitoring systems.

Editor:

- ♦ *Automation in Construction* (2024-present)
- ♦ *42nd Conference on IT in Construction* (2025-present)

TEACHING EXPERIENCE (Ordered by Descending Course Level)

Department of Computer Science:

Natural Language Computing (CSC 2511) – TA, Course Developer, Substitute Lecturer

01/2024-05/2024, 08/2024-12/2024

- ♦ Designed and wrote tests for neural machine translation assignment using Transformer and RNN.
- ♦ Gave lectures on introduction to Pytorch, RNN and Transformer.
- ♦ Tutored around 100 students including teaching tutorials and assisting with questions.
- ♦ Invigilated, and graded exams.

Capstone Design Project (CS 490) – TA

08/2023-12/2023

- ♦ Mentored senior students on machine learning/deep learning computer vision capstone projects

Introduction to Artificial Intelligence (CSC 384) – TA

01/2022-04/2022

Introduction to Machine Learning (CS 311) – TA

09/2022-12/2022

Introduction to Data Structures and Algorithms (CS 263) – TA

01/2023-04/2023

Software Design (CS 207) – TA

08/2023-12/2023

Introduction to Computer Science (CS 148) – TA

05/2023-08/2023

Principles and Techniques of Data Science (CS 100 at UC Berkeley) – TA

08/2021-12/2021

- ♦ Conducted lab/tutorial sessions, office hours for 30+ students
- ♦ Graded assignments/exams; integrated practical examples

Introduction to Data Science (GGR 274) – TA

01/2022-05/2022

- ♦ Developed automated grading systems; maintained MarkUs testing environment

Introduction to Computer Programming (CS 108) – TA and Substitute Lecturer 09/2022-12/2022, 08/2023-12/2023

- ♦ Lectured on Python testing/debugging; assisted in interactive content development
- ♦ Graded assignments/exams, held office hours

Writing Development Initiative in Computer Science (WDI) – TA 01/2023-04/2023

- ♦ Graded and provided feedback for writing assignments on structure, writing mechanics, and audience expectations

Department of Statistical Sciences:

TA for Regression Analysis (STA 302) 08/2023-12/2023, 08/2024-12/2024

TA for Introduction to Statistical Reasoning and Data Science (STA 130) 01/2025-04/2025

- ♦ Designed and delivered tutorials for ~70 students (STA 302) and ~30 students (STA 130)
- ♦ Held office hours, graded assignments, invigilated, and graded exams

Department of Civil & Mineral Engineering:

Survey Camp (CME 358) – TA and Substitute Lecturer (08/2022) 08/2022

- ♦ Gave lectures to and help students on highway curve surveying methods and equipment usage
- ♦ Graded assignments, invigilated, and graded exams

Engineering Mathematics II (CME 262) – TA 01/2023-04/2023

Mechanics (CIV 100) – TA 08/2021-12/2021, 08/2023-12/2023, 08/2024-12/2024

Elementary Fluid Mechanics (CIV 100 at UC Berkeley) – TA 08/2018-12/2018

- ♦ Delivered tutorials, graded assignments/exams, held office hours
- ♦ Maintained a high average student satisfaction score of 4.85/5 in evaluations for CIV 100 at UC Berkeley

SELECTED HONORS & AWARDS

School of Cities Fellowship (\$2,000 | *University of Toronto*) 05/2022

Connaught International Scholarship (\$10,000 top up per year for 4 years | *University of Toronto*) 2021-2025

Jane Lewis Fellowship (66,000 USD | UC Berkeley) 2019-2020

2nd Place, 2nd Place Final Product, 3rd Place Design Paper, 1st Place Oral Presentation, 3rd Place Men's Sprint Race in ASCE National Concrete Canoe Competition (Mid-Pacific) 04/2017

4th place and Best Design Report in the 9th Inter-University International Structural Design Competition 07/2016

First Place in the Mathematics Contest Modeling 05/2016