Martin Weyssow

Research Scientist (Postdoc), Singapore Management University
email: mweyssow@smu.edu.sg ◇ web: martin-wey.github.io ◇ github: https://github.com/martin-wey

RESEARCH INTERESTS

Deep learning, code generation, vulnerability detection, reinforcement learning from human/ai feedback, large language models, and continual learning.

EDUCATION

Université de Montréal, Canada *09/2020 - 09/2024*
Ph.D., Computer Science
Advisor: Prof. Houari Sahraoui
Thesis topic: Efficient, continual, and preference tuning of LLMs for code generation
GPA: 4.3/4.3, Awarded with highest distinction by the jury

University of Namur, Belgium *09/2018 - 06/2020*
MSc., Computer Science and Data Science
Thesis topic: Deep learning and static program analysis for code completion
Awarded with highest distinction by the jury

University of Namur, Belgium *09/2015 - 06/2018*
BSc., Computer Science, Minors in Mathematics

EXPERIENCE

SCIS, Singapore Management University *10/2024 - current*
Research Scientist (Postdoc)

- Member of the Centre for Research on Intelligent Software Engineering (RISE) at SMU.
- Conducting research on LLMs for software vulnerability detection, while supervising and mentoring students on related research projects.
- Exploring post-training techniques, including knowledge distillation, RLAIF, and continual learning, to enhance model effectiveness and alignment with real-world requirements.

SCIS, Singapore Management University *10/2022 - 01/2023*
Ph.D. Visiting Researcher

- Led a new international collaboration between the Université de Montréal and SMU.
- Secured an NSERC Alliance International Catalyst Grant to support the collaborative research initiative, demonstrating grant-writing and project development skills.
- Conducted collaborative research at the SOftware Analytics Research (SOAR) lab under the mentorship of Prof. David Lo, focusing on parameter-efficient fine-tuning and continual learning for code generation with language models.

DIRO, Université de Montréal *09/2019 - 09/2020*
MSc. Research Intern

- Led a new international collaboration between the University of Namur and the Université de Montréal.
- Successfully completed four advanced graduate courses at DIRO, specializing in machine learning and natural language processing.
- Conducted master's thesis research on the application of deep learning integrated with static program analysis to enhance the performance of code recommendation systems.
- Defended the thesis at the University of Namur, earning a mention of excellence from the jury.

SCHOLARSHIPS & GRANTS

Throughout my studies I received prestigious scholarships, including the Google Scholarship for Excellence and the FRQ-PBEEE scholarship.

Scholarship for Excellence, DIRO	*2023–2024*
Canada NSERC Alliance International Catalyst Grant	*10/2022 – 01/2023*
Scholarship for Excellence, DIRO	*2022–2023*
FRQ Merit Scholarship for Foreign Students (PBEEE)	*05/2022 - 09/2024*
Google Scholarship for Excellence, DIRO	*05/2021 - 05/2022*
Scholarship for Excellence, DIRO	*2021–2022*
Scholarship for Excellence, DIRO	*2020–2021*
Research Scholarship, Mitacs Globalink Canada	*01-05/2020*

COMMUNITY SERVICES

EASE, GenSys Workshop – *Program Committee*	*2025*
EASE, EQUISA Workshop – *Program Committee*	*2025*
ICLR, Deep Learning for Code Workshop – *Program Committee*	*2025*
ACM Transactions on Software Engineering and Methodology (TOSEM, journal) – *Reviewer*	*2024 - current*
Empirical Software Engineering (EMSE, journal) – *Reviewer*	*2024 - current*
FORGE conference – *Program Committee*	*2024*
Automated Software Engineering (AUSE, journal) – *Reviewer*	*2023*
IEEE/ACM Automated Software Engineering (ASE, conference) – *Reviewer*	*2022*
ICLR, Deep Learning for Code Workshop – *Program Committee*	*2022*
Software and System Modeling (SOSYM, journal) – *Reviewer*	*2021*
IEEE SANER Conference – *Reviewer*	*2021*
ACM/IEEE MODELS Conference – *Student Volunteer*	*2020*
IEEE Transactions on Software Engineering (TSE) – *Reviewer*	*2020*

STUDENTS

Aton Kamanda, MSc. at DIRO, Université de Montréal — *09/2021 - 06/2023*
Neural interfaces for deep neural networks

Lucas Maes, MSc. at DIRO, Université de Montréal — *09/2021 - 06/2023*
Self-supervised learning for code representation learning

Bastien Nicolas, MSc. at University of Namur — *01/2021 - 06/2021*
Learning from Code Flow Dependencies using Graph Neural Networks for Code

PUBLICATIONS

C: Conference, J: Journal, P: Preprint

[P4] **LessLeak-Bench: A First Investigation of Data Leakage in LLMs Across 83 Software Engineering Benchmarks**. X. Zhou, M. Weyssow, R. Widyasari, T. Zhang, J. He, Y. Lyu, J. Chang, B. Zhang, D. Huang, D. Lo. *arXiv preprint arXiv:2502.06215 (under revision)*

[C7] **Harnessing Large Language Models for Curated Code Reviews**. OB. Sghaier, M. Weyssow, H. Sahraoui. *22nd International Conference on Mining Software Repositories (MSR 2025)*

[C6] **A Functional Software Reference Architecture for LLM-Integrated Systems**. A. Bucaioni, M. Weyssow, J. He, Y. Lyu, D. Lo. *22nd IEEE International Conference on Software Architecture (ICSA-NEMI 2025)*

[J3] **Exploring Parameter-Efficient Fine-Tuning Techniques for Code Generation with Large Language Models**. M. Weyssow, X. Zhou, K. Kim, D. Lo, H. Sahraoui. *ACM Transactions on Software Engineering and Methodology (TOSEM 2025).*

[P3] **CleanVul: Automatic Function-Level Vulnerability Detection in Code Commits Using LLM Heuristics**. Y. Li, T. Zhang, R. Widyasari, YN. Tun, HH. Nguyen, T. Bui, IC. Irsan, Y. Cheng, X. Lan, HW. Ang, F. Liauw, M. Weyssow, HJ. Kang, EL. Ouh, LK. Shar, D. Lo (2024). *arXiv preprint arXiv:2411.17274 (under revision)*

Aligning Language Models to Code: Exploring Efficient, Temporal, and Preference Alignment for Code Generation. M. Weyssow (2024). *PhD Thesis*

[P2] **CodeUltraFeedback: An LLM-as-a-Judge Dataset for Aligning Large Language Models to Coding Preferences**. M. Weyssow, A. Kamanda, X. Zhou, H. Sahraoui. *arXiv preprint arXiv:2403.09032* (under revision)

[C5] **CodeLL: A Lifelong Learning Dataset to Support the Co-Evolution of Data and Language Models of Code**. M. Weyssow, C. Di Sipio, D. Di Ruscio, H. Sahraoui. *21st International Conference on Mining Software Repositories (MSR 2024, Data and Tool Showcase Track)*

[C4] **On the Usage of Continual Learning for Out-of-Distribution Generalization in Pre-trained Languages Models of Code**. M. Weyssow, X. Zhou, K. Kim, D. Lo, H. Sahraoui. *The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2023)*

[C3] **AST-Probe: Recovering abstract syntax trees from hidden representations of pre-trained language models**. J. A. H. López *, M. Weyssow *, J. S. Cuadrado, H. Sahraoui. *37th IEEE/ACM International Conference on Automated Software Engineering (ASE 2022)* *equal contributions

[C2] **Better Modeling the Programming World with Code Concept Graphs-augmented Multi-modal Learning**. M. Weyssow, H. Sahraoui, B. Liu. *44th IEEE International Conference on Software Engineering, New Ideas and Emerging Results (ICSE-NIER 2022)*

[J2] **Recommending Metamodel Concepts during Modeling Activities with Pre-Trained Language Models**. M. Weyssow, H. Sahraoui, E. Syriani. *Software and Systems Modeling (SoSym 2022), theme issue on AI-enhanced Model-Driven Engineering*

[J1] **Opportunities in intelligent modeling assistance**. G. Mussbacher, B. Combemale, J. Kienzle, S. Abrahão, H. Ali, N. Bencomo, M. Búr, L. Burgueño, G. Engels, P. Jeanjean, JM. Jézéquel, T. Kühn, S. Mosser, H. Sahraoui, E. Syriani, D. Varró, M. Weyssow. *Software and Systems Modeling (SoSym 2020)*

[P1] **Combining Code Embedding with Static Analysis for Function-Call Completion**. M. Weyssow, H. Sahraoui, B. Vanderose, B. Frénay *arXiv preprint arXiv:2008.03731* (2020)

[C1] **Towards an Assessment Grid for Intelligent Modeling Assistance**. G. Mussbacher, B. Combemale, S. Abrahão, N. Bencomo, L. Burgueño, G. Engels, J. Kienzle, T. Kühn, S. Mosser, H. Sahraoui, M. Weyssow. *MODELS 2020, 2nd Workshop on Artificial Intelligence and Model-driven Engineering*