Martin Weyssow

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

RESEARCH INTERESTS

Large language models, code generation, vulnerability detection, reinforcement learning from human/ai feedback (RLHF/RLAIF), LLM alignment, and continual learning.

EDUCATION

Université de Montréal, Canada *09/2020 - 09/2024*
Ph.D., Computer Science
Thesis topic: Efficient, continual, and preference fine-tuning of LLMs for code generation
GPA: 4.3/4.3, highest distinction

University of Namur, Belgium *09/2018 - 06/2020*
MSc., Computer Science and Data Science
graduated with *magna cum laude*

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

EXPERIENCE

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

- Contributing as a 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 curriculum learning, to enhance model effectiveness and alignment with real-world requirements.

SCIS, Singapore Management University *10/2022 - 01/2023*
PhD Visiting Researcher

- Established a new international collaboration between the Université de Montréal and SMU, fostering long-term research partnerships.
- 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*
Research intern

- Established 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 with distinction 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

ACM Transactions on Software Engineering and Methodology (TOSEM, journal) – *Reviewer*	*2024*
Empirical Software Engineering (EMSE, journal) – *Reviewer*	*2024*
FORGE conference – *Program Committee*	*2024*
Automated Software Engineering (ASE, journal) – *Reviewer*	*2023*
IEEE/ACM Automated Software Engineering (ASE, conference) – *Reviewer*	*2022*
International Conference on Learning Representations (ICLR), DL4C 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 Université de Montréal *09/2021 - 06/2023*
Neural interfaces for deep neural networks

Lucas Maes, MSc. at 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

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)
CleanVul: Automatic Function-Level Vulnerability Detection in Code Commits Using LLM Heuristics
arXiv preprint arXiv:2411.17274
under review

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

M. Weyssow, A. Kamanda, & H. Sahraoui (2024)
CodeUltraFeedback: An LLM-as-a-Judge Dataset for Aligning Large Language Models to Coding Preferences
arXiv preprint arXiv:2403.09032
under review

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

M. Weyssow, X. Zhou, K. Kim, D. Lo, & H. Sahraoui (2023)
Exploring parameter-efficient fine-tuning techniques for code generation with large language models
arXiv preprint arXiv:2308.10462
under review

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

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

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

M. Weyssow, H. Sahraoui & E. Syriani (2021)
Recommending Metamodel Concepts during Modeling Activities with Pre-Trained Language Models
Software and Systems Modeling, theme issue on AI-enhanced Model-Driven Engineering

Mussbacher, G., Combemale, B., Kienzle, J. et al. (2020)
Opportunities in intelligent modeling assistance
Software and Systems Modeling

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

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