

Dayi Lin

PRINCIPAL RESEARCHER · SOFTWARE ENGINEERING FOR AI SYSTEMS

☎ (+1) 613-770-8659 | ✉ heydayi@gmail.com | 🏠 lindayi.me

🎓 Google Scholar: [Dayi Lin](#) | 🔗 LinkedIn: [dayilin](#)

*Good traditional software is never shipped with “it probably works and I don’t know why”.
Good agentic foundation model systems shouldn’t either.*

Highlights

- **Lead a 20-person industrial research team on Software Engineering for Compound AI Systems**, innovating advanced technologies that drastically improve the productivity and quality when building foundation model applications (including multi-agent systems), with an annual research budget of \$2M+ USD.
- **Experienced in interdisciplinary research across software engineering and AI**, with a track record of patents and publications in top venues including ICSE, FSE, TSE, TOSEM and EMSE; serve as AIware’24 PC Co-Chair and Steering Committee member, Co-organizer of GAS’22-’24, and top venue PC member/reviewer (e.g., TOSEM, JSS, EMSE, FSE, ICSE-SEIP, CAIN).
- My talk at FM+SE Summit 2024 on rethinking engineering for foundation model systems, detailing 10 research directions of our work: https://youtu.be/3gnD0sAo_pk.

Experience

Principal Researcher — Huawei Canada Research Institute

Toronto, Canada

SOFTWARE ENGINEERING FOR CONVENTIONAL AI AND FOUNDATION MODEL BASED SYSTEMS

Oct. 2020 - Present

- **Lead a research team of 20 researchers with \$2M USD annual investment** to innovate on methods, techniques and tools for productively engineering high-quality and trustworthy AI/foundation model systems, with a system and interdisciplinary perspective rather than only focusing on improving models
- **Lead the research prototype development of full-stack foundation model application engineering solutions**, including multi-agent development studios, multi-agent native framework, and application graph IR compiler. Full-stack prototype adopted and productionized by Huawei Cloud as a commercial product [FMArts Studio](#)
- **Deliver research breakthroughs across the full lifecycle of foundation model based systems:**
 - **[Design & Development] Increase prompt optimization productivity by 10x** through proposed multi-granularity prompt explanation and attribution-based targeted prompt mutation techniques
 - **[Design & Development] Improve task success rate through enhanced user intent understanding** with proposed Theory-of-Mind agent cognitive architecture, addressing ambiguity in natural language instructions
 - **[Quality Assurance] Achieve debuggability of implicit agent reasoning processes** by proposed semantic agent observability techniques, proven effective by fixing 2 failed cases of AutoCodeRover on SWE-Bench
 - **[Quality Assurance] Improve LLM-as-judge accuracy by 6.2% and significantly reduce labelling effort** through proposed constitution-based judging methods
 - **[Deploy & Serve] Reduce strong model reliance (e.g., GPT-4 fallback calls) by 34%** in layered model architectures through our dynamic routing and in-context continual learning approach, while maintaining performance
 - **[Deploy & Serve] Reduce toxicity by 29.7% and IP infringement by 56.2% in LLM output** through our real-time decoding safeguarding approach

Data Scientist — Prodigy Game

Toronto, Canada

PLAYER BEHAVIOUR MODELLING | DATA-DRIVEN GAME DESIGN

Feb. 2019 - Oct. 2020

- Conducted research on player behaviour modelling; modelled complex in-game user behaviour at both user and session levels, with data from 70+ million users who generate 300+ million events per day
- Conducted research on player segmentation; profiled players based on their in-game behaviour for personalized gaming experiences; identified an opportunity to lift membership conversion by 2-3x on a major segment of users

Data Researcher — Ford Motor & Blackberry

DATA-DRIVEN INFOTAINMENT SYSTEM DESIGN

Waterloo, Canada

Jan. 2017 - Apr. 2017

- Designed Markov Chain-based algorithms, in combination with binary classifiers, to identify orphaned or wrongly recognized voice commands and mistakes in user behaviours (e.g., misclicks)
- Analyzed user behavioural data from infotainment systems in Ford vehicles to provide feedback to design and development teams, improving the user-perceived quality of the system

Selected US Patents

| | |
|--|---------|
| Methods, Apparatuses, and Computer-Readable Media for Interpreting a Prompt of a Foundation Model | Filed |
| Hierarchical Dynamic Planning of Foundation Model Agents | Filed |
| Systems, Apparatuses, Methods, and Non-Transitory Computer-Readable Storage Media Employing Similarity-Based Filtering for Foundation Models | Filed |
| Machine Learning Asset Management | Filed |
| Augmentation of Synthetic Labels with Human-generated Data to Avoid Model Collapse | Filed |
| Theory-of-Mind-based Goal Alignment Agent | Filed |
| Systems, methods, and non-transitory computer-readable storage devices for detecting and analyzing data clones in tabular datasets | Granted |
| Systems, methods, and non-transitory computer-readable storage devices for training deep learning and neural network models using overfitting detection and prevention | Granted |

Selected Publications (SEE GOOGLE SCHOLAR FOR FULL LIST)

| | |
|--|------------------------------|
| Towards AI-Native Software Engineering (SE 3.0): A Vision and a Challenge Roadmap AE HASSAN, GA OLIVA, D LIN , B CHEN, ZM JIANG (ARXIV:2410.06107) | ICSE-SEIP, 2025 Submitted |
| A Framework for Real-time Safeguarding the Text Generation of Large Language Model X DONG, D LIN , S WANG, AE HASSAN (ARXIV:2404.19048) | ICSE-SEIP, 2025 Submitted |
| Rethinking Software Engineering in the Era of Foundation Models: A Curated Catalogue of Challenges in the Development of Trustworthy FMware AE HASSAN, D LIN , GK RAJBAHADUR, K GALLABA, ... WM ABDULLAH | FSE Industry, 2024 |
| SimClone: Detecting Tabular Data Clones using Value Similarity X YANG, GK RAJBAHADUR, D LIN , S WANG, ZM JIANG | TOSEM, 2024 |
| Towards Training Reproducible Deep Learning Models B CHEN, M WEN, Y SHI, D LIN , GK RAJBAHADUR, ZM JIANG | ICSE, 2022 |
| Towards a consistent interpretation of AIOps models Y LYU, GK RAJBAHADUR, D LIN , GK RAJBAHADUR, B CHEN, ZM JIANG | TOSEM, 2022 |
| The Impact of Data Merging on the Interpretation of Cross-Project Just-In-Time Defect Models D LIN , C TANTITHAMTHAVORN, AE HASSAN | TSE, 2021 |

Education

Ph.D. in Computer Science — Queen's University

SUPERVISOR: AHMED E. HASSAN, ULTRA-LARGE SCALE SOFTWARE SYSTEMS SPECIALIZATION

Canada

2015 - 2018