

Dayi Lin

SENIOR SOFTWARE ENGINEERING RESEARCHER · DATA SCIENTIST
☎ (+1) 613-770-8659 | ✉ dayi.lin@queensu.ca | 🏠 <https://lindayi.me>
🎓 Dayi Lin | 🌐 dayilin | 🐦 @Dr_DayiLin | 📷 lindayi

Addressing challenges in the **modern software development lifecycle** of **PC games** and **machine learning systems**.

Highlights

- Proven effectiveness in both **industry** (experienced data scientist and industry researcher across multiple business domains) and **academia** (top-tier journal publications)
- Attracted **prestigious media coverage** including Kotaku, PC Gamers, Gamasutra, and national newspapers; over 32K reads on ResearchGate
- **Research interests:** empirical software engineering, game engineering, machine learning for software engineering, software engineering for AI systems

Education

Ph.D. — Queen's University

ON, Canada

COMPUTER SCIENCE | ULTRA-LARGE SOFTWARE SYSTEM SPECIALIZATION

Sept. 2015 - Jan. 2019

- Supervisor: Prof. Ahmed E. Hassan (Software Analysis and Intelligence Lab)
- Thesis Topic: *How Can Game Developers Leverage Data from Online Distribution Platforms? A Case Study of the Steam Platform*

B.Eng. — Xi'an University of Posts and Telecommunications (XUPT)

Shaanxi, China

COMPUTER SCIENCE AND TECHNOLOGY

Sept. 2011 - Jun. 2015

- Alumnus of *National Excellent Engineer Education and Training Program*
- 98% Percentile, holder of *Outstanding Student Leader* and *Outstanding Graduates* awards

Publications

JOURNAL PUBLICATIONS (PEER-REVIEWED)

An Empirical Study of Trends of Popular Virtual Reality Games and Their Complaints

T-G, 2021

R EPP, **D LIN**, CP BEZEMER

- IEEE Transactions on Games (T-G), 2021

An Empirical Study of the Characteristics of Popular Minecraft Mods

EMSE, 2020

D LEE, GK RAJBAHADUR, **D LIN**, M SAYAGH, CP BEZEMER, AE HASSAN

- Empirical Software Engineering (EMSE), 2020

Building the Perfect Game – An Empirical Study of Game Modifications

EMSE, 2020

D LEE, **D LIN**, CP BEZEMER, AE HASSAN

- Empirical Software Engineering (EMSE), 2020

- Identifying Gameplay Videos that Exhibit Bugs in Computer Games** *EMSE, 2019*
D LIN, CP BEZEMER, AE HASSAN
 • Empirical Software Engineering (EMSE), 2019
- An Empirical Study of Game Reviews on the Steam Platform** *EMSE, 2018*
D LIN, CP BEZEMER, Y ZOU, AE HASSAN
 • Empirical Software Engineering (EMSE), 2018
- An Empirical Study of Early Access Games on the Steam Platform** *EMSE, 2018*
D LIN, CP BEZEMER, AE HASSAN
 • Empirical Software Engineering (EMSE), 2018
- Studying the Urgent Updates of Popular Games on the Steam Platform** *EMSE, 2017*
D LIN, CP BEZEMER, AE HASSAN
 • Empirical Software Engineering (EMSE), 2017

CONFERENCE PUBLICATIONS (PEER-REVIEWED)

- Chinese Word Segmentation based on Conditional Random Fields with Character Clustering** *IALP, 2016*
 L DU, X LI, C LIU, R LIU, X FAN, J YANG, **D LIN**, M WEI
 • The 20th International Conference on Asian Language Processing (IALP), 2016
- Chinese Term Extraction from Web Pages based on Expected Point-wise Mutual Information** *ICNC-FSKD, 2016*
 L DU, X LI, **D LIN**
 • The 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016

THESIS

- How Can Game Developers Leverage Data from Online Distribution Platforms? A Case Study of the Steam Platform** *2019*
D LIN
 • Ph.D. Thesis, 2019

Experience

AS RESEARCHER

- Senior Researcher — Toronto Centre for Software Excellence, Huawei** *ON, Canada*
 SOFTWARE ENGINEERING FOR AI SYSTEMS | AIOps *Oct. 2020 - Present*
- Conduct research on software engineering theories and practices for developing and maintaining artificial intelligence systems
 - Maintain collaboration with internal product teams and external research partners

Research Assistant — SAIL Lab, Queen's University

ON, Canada

MINING SOFTWARE REPOSITORIES | GAME ENGINEERING | DEFECT PREDICTION

Sept. 2015 - Jan. 2019

- Conducted research on the modern software development lifecycle of a unique software genre – PC games
- Combined both qualitative methods (e.g., Grounded Theory) and quantitative methods (e.g., Natural Language Processing and Machine Learning) to uncover patterns and insights in game store data. Proposed practical suggestions that help game developers produce higher quality games and improve user satisfaction
- Built an intelligent tool to automatically identify game bugs from gameplay videos, with a median average precision at 10 of 0.91
- Improved the interpretation and goodness-of-fit by 100% for just-in-time cross-project defect models using context-aware mixed-effect modelling

Research Assistant — Text Mining Group, XUPT

Shaanxi, China

NATURAL LANGUAGE PROCESSING | COMPLEX HETEROGENEOUS NETWORK ANALYSIS

Sept. 2012 - Jun. 2015

- Designed and developed a visualization and query engine for knowledge graphs
- Proposed a novel approach to rank entities in an academic network by constructing and mining a large-scale heterogeneous information network

AS DATA SCIENTIST

Data Scientist — Prodigy Game

ON, Canada

USER BEHAVIOUR MODELLING | MACHINE LEARNING | EXPERIMENT DESIGN

Feb. 2019 - Oct. 2020

- Conducted research on player behavior modelling; modelled complex in-game user behavior 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 behavior to empower personalized gaming experiences; identified an opportunity to lift membership conversion on a major segment of users by 2 to 3 times
- Sponsor investigator of NSERC Alliance Grant project on data-driven automated game quality assurance. Facilitate the project and support students on the project as the industry partner
- Modelled users' usage pattern of key features and their compound impact on core game metrics (e.g., user retention); provided in-depth data-driven insights and predictions that power product iterations
- Collaborated with product and user research teams to design, conduct and analyze in-game experiments for new releases

Data Researcher — Ford Motor & Blackberry

ON, Canada

CO-OP STUDENT | ALGORITHM DESIGN | STATISTICAL MODELLING

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 behaviors (e.g., mis-clicks)
- Analyzed user behavioral data from infotainment systems in Ford vehicles to provide feedback to design and development teams, improving the user-perceived quality of the system
- Instrumented the source code of a legacy large-scale distributed system and used log mining techniques to collect and stream necessary data for online analysis

AS DEVELOPER

Research & Development Engineer — Alibaba Group

Zhejiang, China

INTERN | BACK-END WEB DEVELOPMENT

Jun. 2014 - Sept. 2014

- Developed the backend and internal tools for Member Entertainment System and the membership module of Tmall (one of the largest e-commerce websites in China), which supports millions of page views per day

Teaching

Academic Assistant — Smith School of Business, Queen's University

ON, Canada

MMA-865: BIG DATA ANALYSIS | MMA-869: MACHINE LEARNING & ARTIFICIAL INTELLIGENCE

Aug. 2019 - May. 2020

- Prepared course materials for the two courses
- Gave lectures and tutorials about using Spark and Spark MLlib to process and model data at scale
- Designed and migrated the teaching environment of MMA program to Microsoft Azure cloud

Head Teaching Assistant — School of Computing, Queen's University

ON, Canada

CISC-326: SOFTWARE / GAME ARCHITECTURE

Sept. 2016 - Dec. 2018

- Prepared course materials for the Software / Game Architecture course
- Gave lectures about software requirement engineering and modern code analysis tools (*Understand*)
- Led four other teaching assistants in organizing the 200-student class

Mentoring

Selina Wan

ON, Canada

INTERN AT TORONTO CENTRE FOR SOFTWARE EXCELLENCE, HUAWEI

Jan. 2021 - Present

- Research Topic: *Software Engineering for AI Systems*

Daniel Lee

ON, Canada

RESEARCH-BASED M.SC. STUDENT AT SAIL LAB, QUEEN'S UNIVERSITY

Jan. 2018 - Nov. 2019

- Thesis Topic: *Studying the Distribution Platforms for Game Mods to Create Active Modding Communities*

Presentations

Building the perfect game—an empirical study of game modifications

ICSE

JOURNAL FIRST PRESENTATION

May 2021

- The 43rd International Conference on Software Engineering (ICSE), 2021

An Empirical Study of Game Reviews on the Steam Platform

Université de Montréal

INVITED TALK

Nov. 2018

- Workshop on Video Game Criticism, Université de Montréal

An Empirical Study of Early Access Games on the Steam Platform

ICSE

JOURNAL FIRST PRESENTATION

May 2018

- The 40th International Conference on Software Engineering (ICSE), 2018

Studying the Urgent Updates of Popular Games on the Steam Platform

ICSME

JOURNAL FIRST PRESENTATION

Sept. 2017

- The 33rd International Conference on Software Maintenance and Evolution (ICSME), 2017

Research on Games: Studying the Steam Platform

Queen's University

GUEST LECTURE

Oct. 2016

- CISC-880: Mining Software Repositories, School of Computing, Queen's University

Research on Games: Studying the Steam Platform

Queen's University

GUEST LECTURE

Oct. 2016

- CISC-326: Game Architecture, School of Computing, Queen's University

Media Coverages

SELECTED COVERAGE (FULL LIST AVAILABLE ON MY WEBSITE)

Big study of 10 million Steam reviews is absolutely fascinating

Kotaku

LUKE PLUNKETT, KOTAKU

<http://bit.ly/2u1n9La>

Research Paper on Steam Early Access Reveals 5 Lessons For Developers

Kotaku

LOGAN BOOKER, KOTAKU

<http://bit.ly/20aA6t7>

Steam study finds users leave negative reviews more quickly than positive ones

PC Gamer

AUSTIN WOOD, PC GAMER

<http://bit.ly/30xfwbG>

Study suggests Steam reviewers are bothered more by bad game design than bugs

Rock Paper Shotgun

MATT COX, ROCK PAPER SHOTGUN

<http://bit.ly/389Pde1>

Studie zu Steam-Reviews: User sudern schneller, ärgern sich kaum über bugs

Der Standard

DER STANDARD

<http://bit.ly/35SYHZz>

Skills

Domains	Software Engineering, Mining Software Repositories, Data Science, Machine Learning
Research Methods	Quantitative & Qualitative Analysis, Grounded Theory, Experimental Research Design
Programming	Python, R, Java, C/C++, PHP, SQL, LaTeX, HTML/CSS
Frameworks	Spark, MLib, GraphX, Kafka, Hadoop, Numpy/Scipy, Scikit-Learn
Languages	English, Mandarin

Honors & Awards

2019	Invited Participant , Shonan Meeting No. 156: Software Engineering for ML Systems, NII Japan
2018	Invited Participant , Workshop on Video Game Criticism, Université de Montréal
2015-2019	Recipient , Queen's Graduate Award, Queen's University
2017-2018	Recipient , Conference Travel Award, Queen's University
2018	Winner and Avande Sponsored Prize Recipient , QHacks
2016	Winner , Limestone City Hacks
2015	Outstanding Graduates , Xi'an University of Posts and Telecommunications

Services

2020	Reviewer , IEEE Transactions on Games (T-G) Journal
2020	Co-reviewer , Empirical Software Engineering (EMSE) Journal

References

Dr. Ahmed E. Hassan

Professor

QUEEN'S UNIVERSITY

Email: ahmed@cs.queensu.ca

Website: <http://sail.cs.queensu.ca>

Dr. Cor-Paul Bezemer

Assistant Professor

UNIVERSITY OF ALBERTA

Email: bezemer@ualberta.ca

Website: <https://www.ece.ualberta.ca/~bezemer/>