Tapajit Dey

MTS Researcher · Carnegie Mellon University Software Engineering Institute (SEI-CMU)

📍 4500 Fifth Avenue Room#5300A, Pittsburgh, PA, USA

☎ +1 412-268-6064 | @ tdey@sei.cmu.edu | 🔗 tapjdey.github.io | ⌂ tapjdey | in tapajit-dey | 🎓 Google Scholar - Tapajit Dey

Internationally recognized Software Engineering researcher with 3+ years of post-PhD research experience and 3+ years of Industry work experience across North America, Europe, and Asia.

💼 Work Experience

Carnegie Mellon University Software Engineering Institute
Pittsburgh, PA, USA

Member of Technical Staff (MTS) Researcher — *April 2023 - Present*

Lead and contribute to defining and executing applied research innovations in architecture design, analysis and automation. Accomplishments within 6 months include:

- Spearheaded research endeavors aimed at categorizing open source projects at scale.
- Elevated the existing software refactoring algorithm within an ongoing project on refactoring large and complex software applications by introducing new refactoring capabilities and a novel quality assessment metric. Also led the initiative to revamp the associated GUI tool.
- Led the initiative for validating the refactoring solutions and the GUI tool by conducting comprehensive month-long exercises and feedback sessions with veteran members of the organization.

Lero, SFI Research Center for Software - University of Limerick
Limerick, Ireland

Research Fellow, with *Prof. Brian Fitzgerald* — *January 2022 - March, 2023*

- Spearheaded the Research Project: **Newcomers in Open Source**, a pivotal component of the **€6M** *TREES (Trustworthy Responsible Efficient Engineering of Software) Research Program* executed in collaboration with the Canadian branch of a major multinational corporation.
- Acted as an **inaugural steering committee member** for the Lero Open Source Program Office (OSPO) since 2021, overseeing the establishment and management of Lero's GitHub organization and OSPO website.
- Led Lero's commitment to *Equality, Diversity, and Inclusion (EDI)* as an **inaugural member** of the formal **Diversity Committee**, organizing EDI training sessions and championing a comprehensive 5-point action plan to drive EDI initiatives in 2023.

Postdoctoral Researcher, with *Prof. Brian Fitzgerald* — *Sept. 2020 - Jan. 2022*

- Drove the **InnerSource Adoption** research project, a major part of the **€6M** *TREES Research Program*, facilitating the integration of Open Source culture within a major multinational company.
- Conducted the **Global State of InnerSource survey** with InnerSource Commons members in 2020.
- Organized impactful InnerSource workshops at the company by distinguished professionals from Industry (including Microsoft, Eurospace, Sony Mobile) and Academia.
- Contributed the 'good-first-project' pattern for identifying prospective InnerSource projects to the InnerSource Commons community, which is now officially accepted as an 'initial' pattern.

IBM — India Software Lab
Bangalore, India

Staff Research Engineer — *June 2012 - June 2015*

Held the primary responsibility of overseeing Release & Quality Assurance for IBM's Process Development Kits (PDK). Achievements over 3 years include:

- Led the team in charge of automating the test procedures using **Python**, delivering a remarkable ∼**300x** enhancement in test coverage, a **75%** reduction in turnaround time, a **50%** efficiency boost, and a significant decrease in customer issues.
- Co-developed a novel test methodology, "An Integrated Approach to Process Design Kit (PDK) Verification", which was accepted for **Patent Publication**.
- Mentored colleagues in conducting automated unit-testing of their models and conducted department-wide seminars on basic and advanced Python programming.

🎓 Education _____

University of Tennessee, Knoxville (UTK)   *Knoxville, TN, USA*
PH.D. IN COMPUTER SCIENCE, WITH *Dr. Audris Mockus*   *Aug. 2015 - Aug. 2020*
- *Thesis Title* — Modeling User-Affected Software Properties for Open Source Software Supply Chains.
- Conducted large-scale studies of open source software ecosystems and software supply chains with the goal of better understanding the effects of software dependencies and users on software popularity, quality, issues, and pull requests using data from World of Code, GitHub, and GHTorrent.
- *Research Focus* — Software Repository Mining, Data Analytics, Empirical Software Engineering.

Indian Institute of Technology (IIT) Kharagpur   *Kharagpur, WB, India*
DUAL DEGREE (B.TECH.(HONS.) + M.TECH.) IN ELECTRONICS AND ELECTRICAL
COMMUNICATION ENGINEERING   *July 2007 - May 2012*
- *Specialization* — Microelectronics and VLSI Design

🎓 Honors & Awards _____

2022	**YERUN Open Science Award**, Young European Research Universities Network (YERUN) - Awarded to the LERO Open Science Committee.	*Limerick, Ireland + Virtual*
2022	**Distinguished Reviewer Award**, The 19th Working Conference on Mining Software Repositories (MSR) 2022 - Technical Track	*Pittsburgh, PA, USA + Virtual*
2021	**Distinguished Reviewer Award**, The 18th Working Conference on Mining Software Repositories (MSR) 2021 - Technical Track	*Madrid, Spain - Virtual*
2019	**Best Paper Award**, Proceedings of the 15th International Conference on Predictive Models and Data Analytics in Software Engineering (PROMISE - 2019)	*Recife, Brazil*
2017	**Best Graduate Research Assistant**, University of Tennessee, Knoxville — College of Engineering (Award Amount: $ 2500)	*Knoxville, TN, USA*

✎ Grants/Funding _____

PROJECT: Scaling Code Translation with Generative AI   *Team Member*
FUNDING BODY - CARNEGIE MELLON UNIVERSITY SOFTWARE ENGINEERING   *Year: 2024;*
INSTITUTE (SEI-CMU)   *Amount-$1,500,000.*
- The grant supports a research project (scheduled to start in Fiscal Year 2025) on using Generative AI for translating software from one language to another at scale.
- Worked with James Ivers to draft the funding proposal.

PROJECT: Shift Left with Generative AI   *Team Member*
FUNDING BODY - CARNEGIE MELLON UNIVERSITY SOFTWARE ENGINEERING
INSTITUTE (SEI-CMU)   *Year: 2023; Amount-$500,000.*
- The grant supports a research project (scheduled to start in Fiscal Year 2024) on using Generative AI for automating library replacement at scale.
- Worked with Ipek Ozkaya, James Ivers, and Robert Edman to draft the funding proposal.

PROJECT: Open Research Training Programme: Training for Open Research in an Irish Context (TROPIC)   *Activity Coordinator*
FUNDING BODY - NATIONAL OPEN RESEARCH FORUM (NORF), IRELAND   *Year: 2022; Amount-€190,000.*
- The grant describes a plan for developing and rolling out an Open Science and Research Training curriculum in Ireland across all disciplines.
- Liaised with 10+ partners in Lero, University of Limerick Library, and University of Galway Library to draft the Expression of Interest for the grant.
- Liaised with 30+ partners across 5 institutes in Ireland to draft the final grant proposal.

✎ Teaching Experience

Introduction to Information Technology
CO-INSTRUCTOR

University of Limerick
Fall-2020 — Present

- Co-Instructed the course *(with Prof. Brian Fitzgerald)* designed for **1st and 2nd-year undergraduates** intended to provide an introduction to modern information technology concepts and solutions.
- **Restructured course evaluation** due to COVID restrictions in AY 2020-2021, designed the end-of-semester evaluation projects, detailed grading rubrics, and took care of the grading.
- **Renovated the course** in Fall 2021 to incorporate latest developments in the field and to conform with the Universal Design for Learning (**UDL**) guidelines.

Programming Languages and Systems
TEACHING ASSISTANT

University of Tennessee
Spring-2017

- Graded weekly assignments and the mid-term exam, conducted an introductory lecture on Python.

Fundamentals of Digital Archeology
TEACHING ASSISTANT

University of Tennessee
Fall-2016, Fall-2015

- Helped the professor by grading the assignments, overseeing class activities during his absence.

Operating Systems
TEACHING ASSISTANT

University of Tennessee
Spring-2016

- Helped the professor by grading the weekly assignments and grading the mid-term exam.

✎ Supervising Experience

Doctoral Thesis Co-Supervisor

University of Limerick

- Supervising Robert Healy, a Professional Doctorate Student, together with Prof. Brian Fitzgerald (University of Limerick, Ireland) and Prof. Kieran Conboy (University of Galway, Ireland), working on "Scaling Scrum with stable queuing systems - a mixed methodology evaluation".
- Supervised Swarna Pundir, a Professional Doctorate Student, together with Prof. Conor Ryan (University of Limerick, Ireland) for 1 year, working on "Design of AI Model for Fault Prediction and Debugging during Software Development Life Cycle".

Masters Thesis Supervisor
ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

University of Limerick
Summer-2022

- Supervised a total of **4 Masters students** who worked on the following topics — Udhaya Shankar Subramanian : "Software product-user feedback sentimental analysis on Twitter at cross country level"; Vishnu Priya Nangari : "Text Classification on GITHUB Projects"; Venkata Divya : "A Study of Amazon Products & Product Reviews: Product Popularity, Review Sentiment, and Market Basket Analysis"; Kishore Kumar Lenka : "Movie Revenue Prediction using Machine Learning Algorithms".

Masters Thesis Supervisor
ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

University of Limerick
Summer-2021

- Supervised Swachitha Kommugala, a Masters students who worked on the following topic - "Detecting Toxic Comments on StackOverflow using BERT."

Masters Thesis Committee Member
ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

University of Limerick
Summer-2021 - Present

- Served at Thesis Committee Member of 5 Masters students and 3 Bachelors students.

📖 Publications

Book Chapters

[1] Sadika Amreen, Bogdan Bichescu, Randy Bradley, **Tapajit Dey**, Yuxing Ma, Audris Mockus, Sara Mousavi, and Russell Zaretzki. "A Methodology for Measuring FLOSS Ecosystems." In "Towards Engineering Free/Libre Open Source Software (FLOSS) Ecosystems for Impact and Sustainability", pp. 1-29. Springer, Singapore, 2019. | LINK

Journal Articles

[1] **Tapajit Dey**, Willem Jiang, Brian Fitzgerald "Knights and Gold Stars: A Tale of InnerSource Incentivization", IEEE Software, vol. 39, no. 6, pp. 88-98, Nov.-Dec. 2022. | LINK | *Journal Impact Factor: 3.3 (2022)*

[2] Ahmed Imam, **Tapajit Dey**, Alexander Nolte, Audris Mockus, James D Herbsleb "One-off Events? An Empirical Study of Hackathon Code Creation and Reuse", Empirical Software Engineering 27, no. 7 (2022): 1-49. | LINK | *Journal Impact Factor: 4.1 (2022)*

[3] Yuxing Ma, **Tapajit Dey**, Chris Bogart, Sadika Amreen, Marat Valiev, Adam Tutko,d David Kennard, Russell Zaretzki, Audris Mockus "World of Code: Enabling a Research Workflow for Mining and Analyzing the Universe of Open Source VCS data", Empirical Software Engineering 26, no. 2 (2021): 1-42. | LINK | *Journal Impact Factor: 4.1 (2022)*

[4] Andrey Krutauz, **Tapajit Dey**, Peter Rigby, and Audris Mockus. "Do Code Review Measures Explain the Incidence of Post-Release Defects?" Empirical Software Engineering 25, no.5 (2020): 3323–3356. | LINK | *Journal Impact Factor: 4.1 (2022)*.
Also accepted in Journal-First Track in ESEC/FSE 2020 conference.

[5] **Tapajit Dey** and Audris Mockus. "Deriving a Usage-Independent Software Quality Metric." Empirical Software Engineering (2020): 1-46. | LINK | *Journal Impact Factor: 4.1 (2022)*

Conference Articles

[1] **Tapajit Dey**, Brian Fitzgerald, Sherae Daniel "CROSS: A Contributor-Project Interaction Lifecycle Model for Open Source Software". In Proceedings of the 58th Hawaii International Conference on System Sciences. 2025.| LINK.

[2] **Tapajit Dey**, Jonathan Loungani, James Ivers "Smarter Project Selection for Software Engineering Research". In Proceedings of the 20th International Conference on Predictive Models and Data Analytics in Software Engineering. 2024.| LINK.

[3] Robert Healy, Kieran Conboy, **Tapajit Dey**, Edwin Lewzey, and Brian Fitzgerald "Comparing Stability and Sustainability in Agile Systems." In International Conference on Agile Software Development (pp. 129-143). Springer Nature. 2024. | LINK.

[4] Robert Healy, **Tapajit Dey**, Kieran Conboy, Brian Fitzgerald "A Novel Technique to Assess Agile Systems for Stability." In International Conference on Agile Software Development (XP), pp. 20-33. Springer Nature, 2023. | LINK.
A blog on the same topic was published in InfoQ.

[5] **Tapajit Dey**, Andrey Karnauch, and Audris Mockus "Representation of developer expertise in open source software." In 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), pp. 995-1007. IEEE, 2021. | LINK | Slides | Talk Video.

[6] **Tapajit Dey**, Andrey Karnauch, and Audris Mockus "Replication package for representation of developer expertise in open source software." In 2021 IEEE/ACM 43rd International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), pp. 236-237. IEEE, 2021.

[7] Ahmed Imam, **Tapajit Dey**, Alexander Nolte, Audris Mockus, James D Herbsleb "The Secret Life of Hackathon Code - Where does it come from and where does it go?." In 2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR), pp. 68-79. IEEE, 2021. | LINK.

[8] Ahmed Imam, **Tapajit Dey** "Tracking hackathon code creation and reuse." In 2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR), pp. 615-617. IEEE, 2021. | LINK.

[9] **Tapajit Dey** and Audris Mockus "Effect of technical and social factors on pull request quality for the npm ecosystem." In Proceedings of the 14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 1-11. 2020. | LINK | Slides | Talk Video
Mentioned in the blog by Vadim Markovtsev (Founding Engineer at Poolside, Google Developer Expert in ML-2021)

[10] **Tapajit Dey**, Sara Mousavi, Eduardo Ponce, Tanner Fry, Bogdan Vasilescu, Anna Filippova, and Audris Mockus "Detecting and characterizing bots that commit code." In Proceedings of the 17th International Conference on Mining Software Repositories (MSR), pp. 209-219. ACM, 2020. | LINK.
*Featured as one of the 30 "most interesting and useful" papers in ICSE 2020 conference proceedings in **Dr. Greg Wilson's** (co-founder and the first Executive Director of Software Carpentry) blog: LINK*

[11] Tanner Fry, **Tapajit Dey**, Andrey Karnauch, and Audris Mockus "A dataset and an approach for identity resolution of 38 million author ids extracted from 2b git commits." In Proceedings of the 17th International Conference on Mining Software Repositories (MSR), pp. 518-522. ACM, 2020. | LINK

[12] **Tapajit Dey**, Yuxing Ma, and Audris Mockus. "Patterns of effort contribution and demand and user classification based on participation patterns in npm ecosystem." In Proceedings of the Fifteenth International Conference on Predictive Models and Data Analytics in Software Engineering, pp. 36-45. ACM, 2019. | LINK | Slides

[13] **Tapajit Dey** and Audris Mockus. "Modeling Relationship between Post-Release Faults and Usage in Mobile Software." In Proceedings of the 14th International Conference on Predictive Models and Data Analytics in Software Engineering, pp. 56-65. ACM, 2018. | LINK | Slides

[14] **Tapajit Dey** and Audris Mockus. "Are software dependency supply chain metrics useful in predicting change of popularity of npm packages?." In Proceedings of the 14th International Conference on Predictive Models and Data Analytics in Software Engineering, pp. 66-69. ACM, 2018. | LINK | Slides

[15] **Tapajit Dey**, Jacob Logan Massengill, and Audris Mockus. "Analysis of popularity of game mods: A case study." In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, pp. 133-139. ACM, 2016. | LINK | Presented as a poster.
Mentioned in the Wikipedia article about Video Game Mods.

Workshop Articles

[1] **Tapajit Dey**, Bogdan Vasilescu, and Audris Mockus "An Exploratory Study of Bot Commits." In Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops. pp. 61–65. Association for Computing Machinery, 2020. | LINK | Slides | Talk Video — **Invited Paper**

[2] Yuxing Ma, **Tapajit Dey**, and Audris Mockus. "Modularizing global variable in climate simulation software: position paper." In Proceedings of the International Workshop on Software Engineering for Science, pp. 8-11. ACM, 2016. | LINK

Citations: — As of March,2024, have **452 citations** with **h-index** of 13 and **i-10 index** of 14. See List of Publications from the Google Scholar Page for the latest citation information.

👥 Professional Services and Memberships

- **Program Committee Member** - 47th International Conference on Software Engineering (ICSE) 2025 (Technical Track).

- **Program Committee Member** - 22nd Working Conference on Mining Software Repositories (MSR) 2025 (Technical Track).

- **Program Committee Member** - 46th International Conference on Software Engineering (ICSE) 2024 (Technical Track).

- **Program Committee Member** - 21st Working Conference on Mining Software Repositories (MSR) 2024 (Technical Track).

- **Program Committee Member** - 20th International Conference on Predictive Models and Data Analytics in Software Engineering (PROMISE) 2024.

- **Program Committee Member** - 40th IEEE International Conference on Software Maintenance and Evolution (ICSME) 2024 (Registered Reports Track).

- **Co-Organizer, Co-Chair, Web & Proceedings Chair** - *InnerSoft: 1st International Workshop on InnerSource Software Development* at ICSE 2023.

- **Program Committee Member** - 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE) 2023 (Artifacts-track).

- **Program Committee Member** - 45th International Conference on Software Engineering (ICSE) 2023 (Technical Track, Poster Track, Artifact Evaluation Track).

- **Publicity and Social Media Co-Chair** - 20th Working Conference on Mining Software Repositories (MSR) 2023.

- **Program Committee Mmember** - 39th International Conference on Software Maintenance and Evolution (ICSME) 2023 (Artifact Evaluation Track and ROSE Festival-track).

- **Program Committee Member** - 19th International Conference on Predictive Models and Data Analytics in Software Engineering (PROMISE) 2023.

- **Program Committee Member** - 20th Working Conference on Mining Software Repositories (MSR) 2023 (Mining Challenge Track).

- **Publicity and Social Media Co-Chair** - 19th Working Conference on Mining Software Repositories (MSR) 2022.

- **Program Committee Member** - 19th Working Conference on Mining Software Repositories (MSR) 2022 (Technical Track).

- **Shadow PC Mentor** - 19th Working Conference on Mining Software Repositories (MSR) 2022.

- **Proceedings Chair** - 18th International Conference on Predictive Models and Data Analytics in Software Engineering (PROMISE) 2022.

- **Program Committee Member** - 18th Working Conference on Mining Software Repositories (MSR) 2021 (Technical Track).

- **Program Committee Member** - 17th International Conference on Open Source Systems (OSS) 2021.

- **Panel Member** for the *FM Vision 2030 workshop* at Mexico City, Mexico, November 2023. It was a highly prestigious invite-only workshop attended by world leading industrial and academic experts from many top institutions, like Google, Meta, ServiceNow, Microsoft, University of Waterloo, Queen's University, University of California, etc.

- **Invited Talk** - Discussed the State of the InnerSource in the InnerSource APAC summit in 2020. Talk video link

- **Conducted 27 paper reviews for 8 journals,** including: *Empirical Software Engineering*, ACM Transactions on Software Engineering and Methodology (*TOSEM*), IEEE's Transactions on Software Engineering (*TSE*), *IEEE Software*, etc. *- see* ⓘD *ORCID profile for latest information.*

- **Additional Reviewer** - 16th International Conference on Mining Software Repositories 2019.

- **Additional Reviewer** - 13th International Conference on Mining Software Repositories 2016.

- Member of Association for Computing Machinery (**ACM**), including ACM Special Interest Group in Software Engineering (**SIGSOFT**) and **IEEE Computer Society**.

- Member of **InnerSource Commons**, a global forum of InnerSource practitioners.

- Former (founding) member of **Lero Open Source Programme Office (OSPO)** and former member of **Lero Diversity Committee**.

- Organized the World of Code Hackathon in Limerick, Ireland in 2022 - News Article.