

2025 IEEE/ACM International Workshop on Large Language Models for Code (LLM4Code 2025)

**Ottawa, Ontario, Canada
3 May 2025**



**IEEE Catalog Number: CFP254VN-POD
ISBN: 979-8-3315-2616-0**

**Copyright © 2025 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP254VN-POD
ISBN (Print-On-Demand):	979-8-3315-2616-0
ISBN (Online):	979-8-3315-2615-3

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2025 IEEE/ACM International Workshop on Large Language Models for Code (LLM4Code) **LLM4Code 2025**

Table of Contents

Research Papers

Are Large Language Models Memorizing Bug Benchmarks?	1
<i>Daniel Ramos (Carnegie Mellon University), Claudia Mamede (Carnegie Mellon University), Kush Jain (Carnegie Mellon University), Paulo Canelas (Carnegie Mellon University), Catarina Gamboa (Carnegie Mellon University), and Claire Le Goues (Carnegie Mellon University)</i>	
RepairBench: Leaderboard of Frontier Models for Program Repair	9
<i>André Silva (KTH Royal Institute of Technology) and Martin Monperrus (KTH Royal Institute of Technology)</i>	
COSMosFL: Ensemble of Small Language Models for Fault Localisation	17
<i>Hyunjoon Cho (KAIST, Republic of Korea), Sungmin Kang (KAIST, Republic of Korea), Gabin An (KAIST, Republic of Korea), and Shin Yoo (KAIST, Republic of Korea)</i>	
With a Little Help from My (LLM) Friends: Enhancing Static Analysis with LLMs to Detect Software Vulnerabilities	25
<i>Amy Munson (University of California, San Diego), Juanita Gomez (University of California, Santa Cruz), and Alvaro A. Cardenas (University of California, Santa Cruz)</i>	
CWEVAL: Outcome-driven Evaluation on Functionality and Security of LLM Code Generation	33
<i>Jinjun Peng (Columbia University), Leyi Cui (Columbia University), Kele Huang (Columbia University), Junfeng Yang (Columbia University), and Baishakhi Ray (Columbia University)</i>	
Automating the Detection of Code Vulnerabilities by Analyzing GitHub Issues	41
<i>Daniele Cipollone (Delft University of Technology, Netherlands), Changjie Wang (KTH Royal Institute of Technology, Sweden), Mariano Scazzariello (RISE Research Institutes of Sweden, Sweden), Simone Ferlin (Red Hat, Sweden), Maliheh Izadi (Delft University of Technology, Netherlands), Dejan Kostic (KTH Royal Institute of Technology, Sweden), and Marco Chiesa (KTH Royal Institute of Technology, Sweden)</i>	
From Zero to Sixty at the Speed of RAG: Improving YAML Recipe Generation via Retrieval	49
<i>Farima Farmahinifarahani (J.P. Morgan AI Research, USA), Petr Babkin (J.P. Morgan AI Research, USA), Salwa Alamir (J.P. Morgan AI Research, UK), and Xiaomo Liu (J.P. Morgan AI Research, USA)</i>	

SC-Bench: A Large-Scale Dataset for Smart Contract Auditing	57
<i>Shihao Xia (The Pennsylvania State University), Mengting He (The Pennsylvania State University), Linhai Song (The Pennsylvania State University), and Yiying Zhang (University of California, San Diego)</i>	
Training LLMs for Generating IEC 61131-3 Structured Text with Online Feedback	65
<i>Aaron Haag (Siemens AG), Bertram Fuchs (Siemens AG), Altay Kaçan (Siemens AG), and Oliver Lohse (Siemens AG)</i>	
Proving the Coding Interview: A Benchmark for Formally Verified Code Generation	72
<i>Quinn Dougherty (Beneficial AI Foundation, USA) and Ronak Mehta (Independent, USA)</i>	
LLM-ProS: Analyzing Large Language Models' Performance in Competitive Problem Solving	80
<i>Md Sifat Hossain (University of Dhaka, Bangladesh), Anika Tabassum (University of Dhaka, Bangladesh), Md Fahim Arefin (University of Dhaka, Bangladesh), and Tarannum Shaila Zaman (University of Maryland, USA)</i>	
YABLoCo: Yet Another Benchmark for Long Context Code Generation	88
<i>Aidar Valeev (Innopolis University, Russia), Vladimir Ivanov (Innopolis University, Russia), Roman Garaev (Innopolis University, Russia), Vadim Lomshakov (Steklov Institute of Mathematics, Russia), Irina Piontkovskaya (Huawei Noah's Ark Lab, Russia), and Israel Adewuyi (Innopolis University, Russia)</i>	
Evaluating Language Models for Computer Graphics Code Completion	96
<i>Jan Kels (Heinrich Heine University Düsseldorf), Abdelhalim Hafedh Dahou (GESIS – Leibniz-Institute for the Social Sciences), and Brigitte Mathiak (GESIS – Leibniz-Institute for the Social Sciences)</i>	
Cracks in The Stack: Hidden Vulnerabilities and Licensing Risks in LLM Pre-Training Datasets	104
<i>Mahmoud Jahanshahi (University of Tennessee Knoxville, USA) and Audris Mockus (University of Tennessee Knoxville, USA)</i>	
Analysis of Student-LLM Interaction in a Software Engineering Project	112
<i>Agrawal Naman (National University of Singapore), Ridwan Shariffdeen (National University of Singapore), Guanlin Wang (National University of Singapore), Sanka Rasnayaka (National University of Singapore), and Ganesh Neelakanta Iyer (National University of Singapore)</i>	
METAMON: Finding Inconsistencies between Program Documentation and Behavior using Metamorphic LLM Queries	120
<i>Hyeonseok Lee (KAIST), Gabin An (KAIST), and Shin Yoo (KAIST)</i>	
CoCoNUT: Structural Code Understanding does not fall out of a tree	128
<i>Claas Beger (Cornell University, USA) and Saikat Dutta (Cornell University, USA)</i>	
From Theory to Practice: Code Generation Using LLMs for CAPEC and CWE Frameworks	137
<i>Murtuza Shahzad (Northern Illinois University), Joseph Wilson (Northern Illinois University), Ibrahim Al Azher (Northern Illinois University), Hamed Alhoori (Northern Illinois University), and Mona Rahimi (Northern Illinois University)</i>	

Hierarchical Repository-Level Code Summarization for Business Applications Using Local LLMs	145
<i>Nilesh Dhulshette (TCS Research,India), Sapan Shah (TCS Research,India), and Vinay Kulkarni (TCS Research,India)</i>	
Code Summarization Beyond Function Level	153
<i>Vladimir Makharev (Innopolis University, AIRI, Russia) and Vladimir Ivanov (Innopolis University, Russia)</i>	
Is More or Less Automation Better? An Investigation into the LLM4TDD Process	161
<i>Sanyogita Piya (The University of Texas at Arlington), Anahita Samadi (The University of Texas at Arlington), and Allison Sullivan (The University of Texas at Arlington)</i>	
Knowledge Graph Based Repository-Level Code Generation	169
<i>Mihir Athale (Northeastern University, USA) and Vishal Vaddina (Quantiphi Inc., Canada)</i>	
Leveraging LLMs for Legacy Code Modernization: Evaluation of LLM-Generated Documentation	177
<i>Colin Diggs (The MITRE Corporation), Michael Doyle (The MITRE Corporation), Amit Madan (The MITRE Corporation), Eric Scott (The MITRE Corporation), Emily Escamilla (The MITRE Corporation), Jacob Zimmer (The MITRE Corporation), Naveed Nekoo (The MITRE Corporation), Paul Ursino (The MITRE Corporation), Michael Bartholf (The MITRE Corporation), Zachary Robin (The MITRE Corporation), Anand Patel (The MITRE Corporation), Chris Glasz (The MITRE Corporation), William Macke (The MITRE Corporation), Paul Kirk (The MITRE Corporation), Jasper Phillips (The MITRE Corporation), Arun Sridharan (The MITRE Corporation), Doug Wendt (The MITRE Corporation), Scott Rosen (The MITRE Corporation), Nitin Naik (The MITRE Corporation), Justin Brunelle (The MITRE Corporation), and Samruddhi Thaker (The MITRE Corporation)</i>	
Deriving Coding-Specific Sub-Models from LLMs using Resource-Efficient Pruning	185
<i>Laura Puccioni (Spotify), Alireza Farshin (NVIDIA), Mariano Scazzariello (RISE Research Institutes of Sweden), Changjie Wang (KTH Royal Institute of Technology), Marco Chiesa (KTH Royal Institute of Technology), and Dejan Kostic (KTH Royal Institute of Technology)</i>	
The Shortcomings of Code LLMs in Modeling Code Properties	193
<i>Srivishnu Pyda (University of Maryland), Daniel Nichols (University of Maryland), and Abhinav Bhatele (University of Maryland)</i>	
Mix-of-Language-Experts Architecture for Multilingual Programming	200
<i>Yifan Zong (University of Waterloo, Canada), Yuntian Deng (University of Waterloo, Canada), and Pengyu Nie (University of Waterloo, Canada)</i>	

Position Papers

Do Code LLMs Understand Design Patterns?	209
<i>Zhenyu Pan (Northwestern University), Xuefeng Song (Northwestern University), Yunkun Wang (Zhejiang University), Rongyu Cao (Alibaba Group), Binhua Li (Alibaba Group), Yongbin Li (Alibaba Group), and Han Liu (Northwestern University)</i>	

From Scientific Texts to Verifiable Code: Automating the Process with Transformers	213
<i>Changjie Wang (KTH Royal Institute of Technology), Mariano</i>	
<i>Scazzariello (RISE Research Institutes of Sweden), and Marco Chiesa</i>	
<i>(KTH Royal Institute of Technology)</i>	
Author Index	217