

Paper ID	Title	Authors	Year	Date Accessed	University/Organization	Country	DOI/Link
Search Query: ("llm agent" AND "issue resolution") OR ("swe-bench" AND "llm agent")							
RQ1: What benchmarks and metrics are commonly used to evaluate the performance of LLM-based agents? RQ2: What types of LLM-based agents are designed to automate the resolution of issue reports, and how do they perform? RQ3: What factors influence the performance of LLM-based agents in automating the resolution of issue reports? RQ4: Is the performance of LLM-based agents consistent across evaluations on multiple datasets?							
P1	MAGIS: LLM-Based Multi-Agent Framework for GitHub Issue Resolution	Wei Tao, Yuchang Zhou, Yanlin Wang, Wenqing Zhang, Hongyu Zhang, Yu Cheng	2024	18.04.2025	Fudan University, University of Macau, Sun Yat-sen University, Chongqing University, The Chinese University of Hong Kong	China, Hong Kong	https://doi.org/10.48550/arXiv.2403.17927
P2	Alibaba LingmaAgent: Improving Automated Issue Resolution via Comprehensive Repository Exploration	Yingwei Ma, Qingting Yang, Rongyu Cao, Beihua Li, Fei Huang, Yongbin Li	2024	18.04.2025	Alibaba Group	China	https://doi.org/10.48550/arXiv.2406.01422
P3	LocAgent: Graph-Guided LLM Agents for Code Localization	Zhaoling Chen, Xiangyu Tang, Gangda Deng, Fang Wu, Jialong Wu, Zhiwei Jiang, Viktor Prasanna, Arman Cohan, Xingyao Wang	2025	19.04.2025	Yale University, University of Southern California, Stanford University, All Hands AI	United States of America	https://doi.org/10.48550/arXiv.2503.09089
P4	CSR-Bench: Benchmarking LLM Agents in Deployment of Computer Science Research Repositories	Yijia Xiao, Runhui Wang, Luyang Kong, Davor Golob, Wei Wang	2025	19.04.2025	University of California, Amazon Web Services	United States of America	https://doi.org/10.48550/arXiv.2502.06111
P5	HyperAgent: Generalist Software Engineering Agents to Solve Coding Tasks at Scale	Huy Nhan Phan, Tien N. Nguyen, Phong X. Nguyen, Nghi D. Q. Bui, Nalin Wadhwani, Akshay Sawarne, Damani Arora, Abhav Mehrotra, Saijea Ugalia, Ramakrishna B Bari, Aditya Kanade, Nagarajan Natarajan	2024	19.04.2025	FPT Software AI Center, The University of Texas at Dallas	United States of America, Vietnam	https://doi.org/10.48550/arXiv.2409.16299
P6	MASA: Modular Architecture for Software-engineering AI Agents	Xiangyan Liu, Bo Lan, Zhiyuan Hu, Yang Liu, Zhicheng Zhang, Fei Wang, Michael Shieh, Wenmeng Zhou	2024	19.04.2025	Microsoft Research India	India	https://doi.org/10.48550/arXiv.2406.11638
P7	CodeXGraph: Bridging Large Language Models and Code Repositories via Code Graph Databases	Reem Aleithan, Haoran Xue, Mohammad Mahdi Mohajer, Elijah Nnorom, Gies Uddin, Song Wang	2024	19.04.2025	National University of Singapore, Xi'an Jiaotong University, Alibaba Group	Singapore, China	https://doi.org/10.48550/arXiv.2408.03910
P8	SWE-Bench-v: Enhanced Coding Benchmark for LLMs	Benhao Huang, Yingzhuo Yu, Jin Huang, Xinglan Zhang, Jieqi Ma	2024	19.04.2025	Lassonde School of Engineering, York University, Shanghai Jiao Tong University, University of Michigan, University of Illinois Urbana-Champaign	Canada, United States of America	https://doi.org/10.48550/arXiv.2410.06992
P9	DCA-Bench: A Benchmark for Dataset Curation Agents	Anmol Gaudam, Kishore Kumar, Adarsh Jha, Mukunda NS, Ishaan Bholia	2024	19.04.2025	SuperAGI Research	India	https://doi.org/10.48550/arXiv.2409.11190
P10	SuperCoder2.0: Technical Report on Exploring the feasibility of LLMs as Autonomous Programmer	Antonios Antonidakis, Albert Orwall, Kexun Zhang, Yuxi Xie, Anirudh Goyal, William Wang	2025	19.04.2025	University of California, Moztless AI, Carnegie Mellon University, National University of Singapore, Meta	United States of America, Singapore, Canada	https://doi.org/10.48550/arXiv.2410.20285
P11	SWE-Search: Enhancing Software Agents with Monte Carlo Tree Search and Iterative Refinement	Siru Ouyang, Wenhao Yu, Kaijin Ma, Zilin Xiao, Zihan Zhang, Mengzhao Jia, Jiawei Han, Hongming Zhang, Dong Yu	2024	19.04.2025	University of Illinois Urbana-Champaign, Tencent AI Seattle Lab, Rice University, University of Notre Dame	United States of America, France	https://doi.org/10.48550/arXiv.2410.14684
P12	Agentless: Demystifying LLM-based Software Engineering Agents	Soren Dunn, Lingming Zhang	2024	19.04.2025	University of Illinois Urbana-Champaign	United States of America	https://doi.org/10.48550/arXiv.2407.01489
P13	OrcaLoc: An LLM Agent Framework for Software Issue Localization	Zhongming Yu, Hejia Zhang, Yujie Zhao, Hanxian Huang, Moxia Yao, Ke Ding, Jishen Zhao	2025	19.04.2025	University of California, Intel Corporation	United States of America	https://doi.org/10.48550/arXiv.2502.00350
P14	CVE-Bench: A Benchmark for AI Agents' Ability to Exploit Real-World Web Application Vulnerabilities	Yuxuan Zhu, Antony Kelleermann, Dylan Bowman, Philip Li, Akul Gupta, Adarsh Dandia, Richard Fang, Connor Jensen, Eric Ihi, Jason Benn, Jet Geronimo, Avi Dvir, Sudhit Rao, Kaicheng Yu, Tim Stone, Daniel Kang	2025	19.04.2025	Stanford School of Computing and Data Science, University of Illinois Urbana-Champaign	United States of America	https://doi.org/10.48550/arXiv.2503.17332
P15	AutoCodeRover: Autonomous Program Improvement	Yuntong Zhang, Haifeng Ruan, Zhiyu Fan, Abhik Roychoudhury	2024	19.04.2025	National University of Singapore	Singapore	https://doi.org/10.48550/arXiv.2404.05427
P16	Diversity Empowers Intelligence: Integrating Expertise of Software Engineering Agents	Keun Zhang, Weiran Yao, Zuxin Liu, Yihao Feng, Jiacheng Xu, Bo Pang, Yingbo Zhou, Shihy Hennecke, Silvio Savarese, Huan Wang, Caiming Xiong	2025	19.04.2025	Salesforce AI Research, Carnegie Mellon University	Unknown, Singapore	https://openreview.net/forum?id=rK2K43Jnbh
P17	Automated Benchmark Generation for Repository-Level Coding Tasks	Konstantinos Vergopoulos, Mark Nikias Müller, Martin Vechev	2025	19.04.2025	LogicStar AI, ETH Zurich	Switzerland	https://doi.org/10.48550/arXiv.2503.07701
P18	SpecRover: Code Intent Extraction via LLMs	Haifeng Ruan, Yuntong Zhang, Abhik Roychoudhury	2024	19.04.2025	National University of Singapore	Singapore	https://doi.org/10.48550/arXiv.2408.02232
P19	MansCode Agent: AI-native Automated Bug Fixing	Yanhou Liu, Fanglei Cao, Xinchun Wang, Jie Liu, Yexuan Shi, Zhao Zhang, Chao Peng	2024	19.04.2025	ByteDance, Harbin Institute of Technology	China	https://doi.org/10.48550/arXiv.2409.00899
P20	PatchPilot: A Stable and Cost-Efficient Agentic Patching Framework	Hongwei Li, Yuefeng Tang, Shiqi Wang, Weibo Guo	2025	19.04.2025	University of California, Meta	United States of America	https://doi.org/10.48550/arXiv.2502.02747
P21	TDD-Bench Verified: Can LLMs Generate Tests for Issues Before They Get Resolved?	Toufique Ahmed, Martin Hirzel, Rangeel Pan, Avraham Shinnar, Saurabh Sinha	2024	20.04.2025	IBM Research	United States of America	https://doi.org/10.48550/arXiv.2412.02863
P22	Thinking Longer, Not Larger: Enhancing Software Engineering Agents via Scaling Test-Time Compute	Yingwei Ma, Yongbin Li, Yihong Dong, Xue Jiang, Rongyu Cao, Jue Chen, Fei Huang, Binhua Li	2025	20.04.2025	Alibaba Group	China	https://doi.org/10.48550/arXiv.2503.23803
P23	Issue2Test: Generating Reproducing Test Cases from Issue Reports	Noor Nasheed, Islem Bouzenia, Michael Pradel, Ali Messab	2025	20.04.2025	University of British Columbia, University of Stuttgart	Canada, Germany	https://doi.org/10.48550/arXiv.2503.16320
P24	RunXiang Cheng, Michele Tufano, Jürgen Cito, José Cambonero, Pat Rondon, Remya Wei, Aaron Sun, Satish Chandra	Runxiang Cheng, Michele Tufano, Jürgen Cito, José Cambonero, Pat Rondon, Remya Wei, Aaron Sun, Satish Chandra	2025	20.04.2025	Google, TU Wien	United States of America, Austria	https://doi.org/10.48550/arXiv.2502.01821
P25	Agentic Bug Reproduction for Effective Automated Program Repair at Google						

