

Reading List

Week 1 — From LLMs to Coding Agents

1. **The Llama 3 Herd of Models** — Grattafiori et al., 2024 [arXiv](#)
2. **SWE-agent: Agent-Computer Interfaces Enable Automated Software Engineering** — Yang et al., 2024 [arXiv](#)

Question: What does a foundation model need to become a useful software-engineering agent?

Week 2 — Agent Architecture

3. **ReAct: Synergizing Reasoning and Acting in Language Models** — Yao et al., 2023 [arXiv](#)
4. **Agentless: Demystifying LLM-based Software Engineering Agents** — Xia et al., 2024 [arXiv](#)

Question: Does sophisticated agent architecture actually improve software engineering?

Week 3 — Repository Understanding

5. **RepoCoder: Repository-Level Code Completion Through Iterative Retrieval and Generation** — Zhang et al., EMNLP 2023 [ACL Anthology](#)
6. **AutoCodeRover: Autonomous Program Improvement** — Zhang et al., 2024 [arXiv](#)

Question: How does an agent determine which parts of a repository it needs to understand?

Week 4 — Context Engineering

7. **RepoGraph: Enhancing AI Software Engineering with Repository-level Code Graphs** — Ouyang et al., ICLR 2025 [ICLR paper](#) [arXiv](#)
8. **Context as a Tool: Context Management for Long-Horizon SWE-Agents** — Liu et al., Findings of ACL 2026 [ACL Anthology](#)

Question: Is context management becoming a first-class capability of coding agents?

Week 5 — Tools & Action Spaces

9. **CodeAct: Executable Code Actions Elicit Better LLM Agents** — Wang et al., 2024 [arXiv](#)
10. **CODESTRUCT: Code Agents over Structured Action Spaces** — Kim et al., ACL 2026 [ACL Anthology](#)

Question: Should an agent interact with a codebase through text, executable actions, or structured program representations?

Week 6 — Planning, Search & Long-Horizon Work

11. **CodePlan: Repository-level Coding using LLMs and Planning** — Bairi et al., FSE 2024 [Microsoft Research](#)
12. **SWE-Dev: Building Software Engineering Agents with Training and Inference Scaling** — Findings of ACL 2025 [ACL Anthology](#)

Question: How should agents plan and execute work that spans many steps and files?

Week 7 — Verification & Feedback

13. **Agentic Rubrics as Contextual Verifiers for SWE Agents** — ACL 2026 [ACL Anthology](#)
14. **RLEF: Grounding Code LLMs in Execution Feedback with Reinforcement Learning** — Gehring et al., ICML 2025 [arXiv](#)

Question: How does an agent determine whether its solution is correct, and how should feedback change its behavior?

Week 8 — What Should We Measure?

15. **SWE-bench: Can Language Models Resolve Real-World GitHub Issues?** — Jimenez et al., ICLR 2024 [arXiv](#)

Plus SWE-bench Verified, the human-validated subset: [OpenAI — SWE-bench Verified](#)

16. **OmniCode: A Benchmark for Evaluating Software Development Agents** — Sonwane et al., Findings of ACL 2026 [ACL Anthology](#)

Question: What do coding-agent benchmarks actually measure, and what do they systematically miss?

Week 9 — Harness Engineering

17. **Harness Engineering: Anatomy, Architecture, and Evolution of Coding Agents — A Source-Code Study of Eleven Systems** — Barbaste, Darrigol, Vu & Wiltberger, 2026 [arXiv](#)

18. **Agentic Harness Engineering: Observability-Driven Automatic Evolution of Coding-Agent Harnesses** — Lin et al., 2026 [arXiv](#)

Question: If the model is only one component of an agent, what should the surrounding runtime—the *harness*—actually do?

Week 10 — Where Coding Agents Are Going

19. **From Laboratory to Real-World Applications: Benchmarking Agentic Code Reasoning at the Repository Level** — Li, Su & Lyu, ACL 2026 [ACL Anthology](#)
20. **Contract-Coding: Towards Repo-Level Generation via Structured Symbolic Paradigm** — 2026 [ACL Anthology](#)

Question: What abstractions will be required for agents to progress from patch generation to genuine repository-scale software engineering?