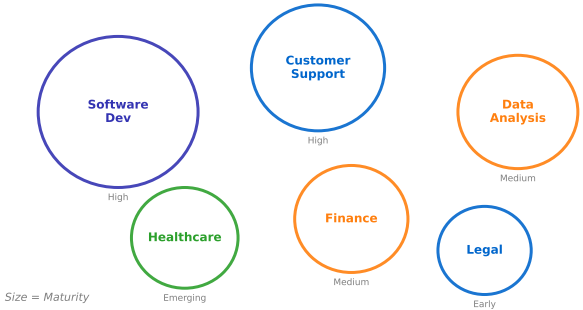
Domain Applications

Week 11: Code, Finance, and Healthcare Agents

PhD Course in Agentic Artificial Intelligence

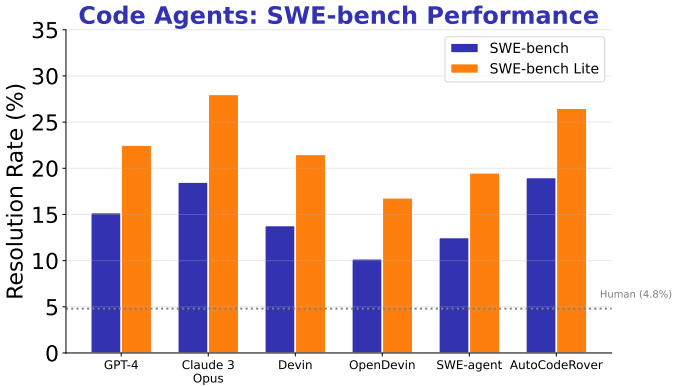
Application Domain Maturity



Softw

development leads in maturity; healthcare is emerging.

Code Agents: SWE-bench Performance

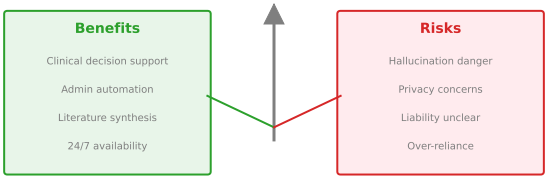


Code agents now outperform average human developers on SWE-bench.

Finan

agents span research, trading, compliance, and operations.

Healthcare Agent Considerations



Human oversight essential | Narrow use cases first

Healt

requires careful balance of benefits vs risks.

Key Papers

This Week

- Jimenez et al. (2024). "SWE-bench: Can Language Models Resolve Real-World GitHub Issues?" arXiv:2310.06770
- Ridnik et al. (2024). "AlphaCodium: Code Generation with Flow Engineering." arXiv:2401.08500
- Li et al. (2024). "FinAgent: A Multimodal Foundation Agent for Financial Trading." arXiv:2402.18485

Focus on SWE-bench paper for understanding code agent evaluation.

Summary

Key Insights

- **Code**: Most mature domain; SWE-bench as gold standard
- **Finance**: High value but requires compliance awareness
- **Healthcare**: High stakes require human oversight

Next Week

- Research Frontiers and Final Projects

expertise + agent capabilities = real-world impact.

(c) Joerg Osterrieder 2025