Week 1 09/16

The general questions from abstract

- How helpful are LLMs in large code bases, in terms of code quality, when specific architectures and patterns are followed?

- How can the user manage large context windows with the tools currently available?

- What changes to the flow of software development happen as a result of using LLMs?

1 Goals For Next Week

- Read in-depth AI Agentic Programming: A survey of Techniques, Challenges, and Opportunities

  - includes Agentic architectures and definitions. Vetted to be useful for reading through next week, take note of agentic architectures they mention and methods used

- Overview of Strands agent framework - build a sample agentic system to understand their potential impact on the flow of software development

Notes

- SMART - specific, measurable, achievable, relevant, time-bound

2 Agenda for Weekly Meeting with Advisor

- Discuss findings so far and general approach to research to see if I am on the correct path. Given this is my first attempt at research and thesis work, I would like to verify my methods are correct and will be beneficial for our long term goal.

3 Last weeks progress

- Read several papers from arxiv.com

  - Paradigm shift on Coding Productivity Using GenAI
  - Generative AI in Software Development: An Overview and Evaluation of Modern Coding Tools
  - Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity
  - AI Agentic Programming: A Survey of Techniques, Challenges, and Opportunities
  - To answer "How can the user manage large context windows with tool currently available" I searched for various tools regarding agents - Amazon Agent core and Strands agents
    * these agents are capable of taking specialized context windows for a multi-agent workflow

4 Knowledge and skills accounting

- In "AI Agentic Programming: A Survey..." the authors describe the context issue and note that Agentic systems incorporate external memory mechanisms to store plans, results, tool outputs, and partial progress. This memory allows agents to maintain information from previous tasks helping with the context issue. A table of context window lengths from various IDE and models is shown as well as ones that have "persistent memory"

- In "Paradigm shift on Coding Productivity" the author did a small study at two companies and provided statistics about the growth of AI tool usage increasing by 57% at the start of 2025. Various tools were used - codeium and amazon Q

- In "Measuring the Impact of Early-2025 AI on Experienced..." productivity was shown to decrease when developers were in large code bases. Some potential important notes are that the participants in this study - 93% previously used LLMs but only 44% had prior experience using Cursor IDE. They note the challenge of context window for LLM and the experience / context the dev already has about the codebase

- Strands agentic framework. Notably, the swarm, which is an idea that a group of specialized agents working together can solve problems more effectively than a single agent.

  - each agent has access to full task context
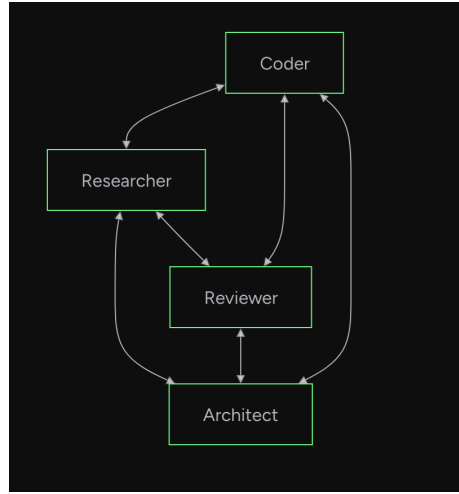 - see history of which agents have worked on task

Figure 1: Strands architecture diagram for Swarm

  - can access shared knowledge contributed by other agents
  - decide when to hand off to another agent

- Asked a few coworkers their respective thoughts to gauge some potential challenges for areas I could dig deeper. Common responses were, "There is no standard thus far in using AI, everyone is using it differently", and "The agents tend to drift off touching unnecessary files or over engineering what should be simple solutions"

5   Approximate Hours Spent

About fifteen minutes to one hour was spent on each research paper depending on depth and quality. Some research papers were not included as I deemed them to be either irrelevant overall, despite having a title that would assume relevance, or poorly written with little citations and examples. About four hours were spent exploring agentic frameworks as well as testing an agent from a course on Hugging Face so I can understand how they work in general without the technical implementation being abstracted too far away. Overall I spent about fifteen to twenty hours to obtain the information presented in this document.

6   How This Week's Effort Is Not a Duplicate Effort

This week's effort is the first in general research. I spent a majority of time gathering information relating to my questions presented in the abstract