

AI Agent Coding Capacities

Agenda

- Demo
- Known Issues & Limitations
- Future Work
- Q & A

Demo

- Basic

- [create issue, clone and scaffolding project](#)
- small tasks
 - [clean branches](#)
 - [remove a file](#)

- Medium

- [create TODO fasterapi](#)
 - [TODO 2](#), pull request, tests, fix issues
 - TODO 3, [test workflow 1](#) (6:07), [test workflow 2](#)
 - [create pr and squash](#)
 - [merge commit](#), [update a commit message](#)

Known Issues & Limitations

- Important info easily loss
 - carry minimal important info
- Predefined Workflow Required
 - dev environment and workflows vary widely
- Error Handling / Feedback loop
 - Easily fall into the failure loop, hard to coverage
 - Relevant information missing
- Dependency / Environment Management
- High Token Usages
 - extensive context for understanding, e.g. requirements or current environment

Known Issues & Limitations (Details)

- Summarization is verbose and contains duplicate info
- Memory is not that useful
 - Hard to decide what level of information to extract
- Not reflect often
 - a little bit over confidence
 - system prompt, autonomy
 - not recognize there's been failures for many times
- Not optimized for coding
 - only “short-term” context
 - security

Learning takeaway

- Use minimal prompt
 - some content might unexpectedly impact the agent capacities
- Use minimal tools
 - help to reduce too many choice, easily to allow agent to decide
 - reduce token usage
-

Future Work

- General

- Summarization: make it concise and relevant
 - reducing verbose and duplicate info in summarization
- Memory: make it more structured, information is scattered,
 - proper project and task management
 - explore available RAG tools
- Reflection
 - Improve the dynamics necessary reflection rate
 - Purpose:
 - problem solving
 - continuity
- Evaluation: establish evaluation process, for better experimentation and iteration
- Strategy to stay coherent, e.g. different level task management

- Coding Capacities:

- Add software engineering predefined workflow
- Increase codebase awareness

Q & A

- ?

Appendix - Token Composition

- Vertical
 - system prompt
 - tool use definition
 - project or task info
 - latest message list and immediate goal
- Horizontal
 - planning
 - reflection
 - memory update / summarization
 - retrieval
- Operation
 - development: trivial and error, not converge fast
 - input file or log can be large unexpectedly