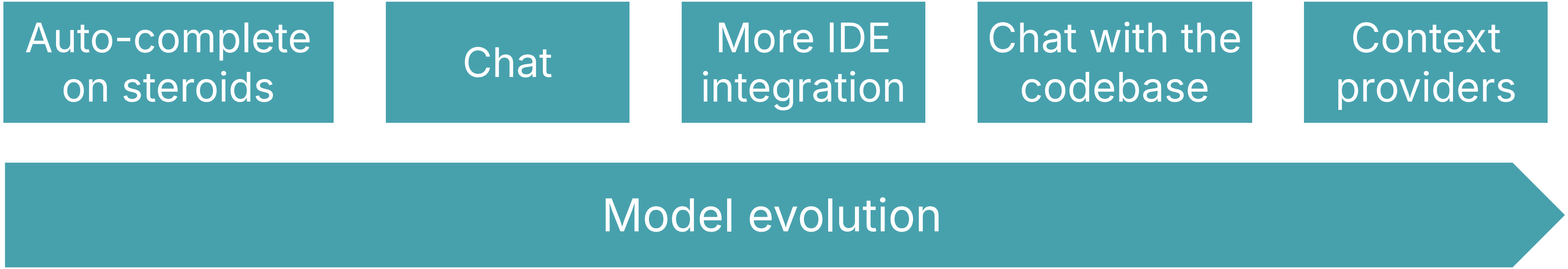
From autocomplete to agents: AI coding assistance state of play

June 2025

Birgitta Böckeler

/thoughtworks

The evolution of coding assistants over the past 2 years



Auto-complete on steroids

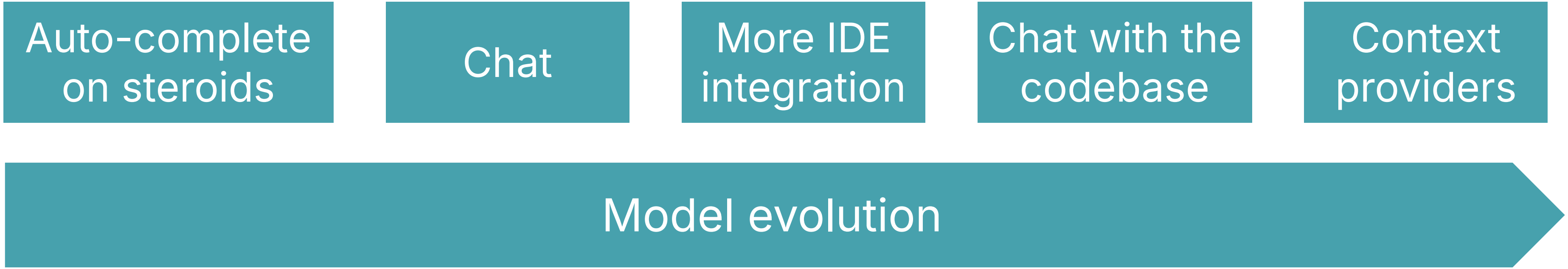
Chat

More IDE integration

Chat with the codebase

Context providers

Model evolution



Auto-complete on steroids

Chat

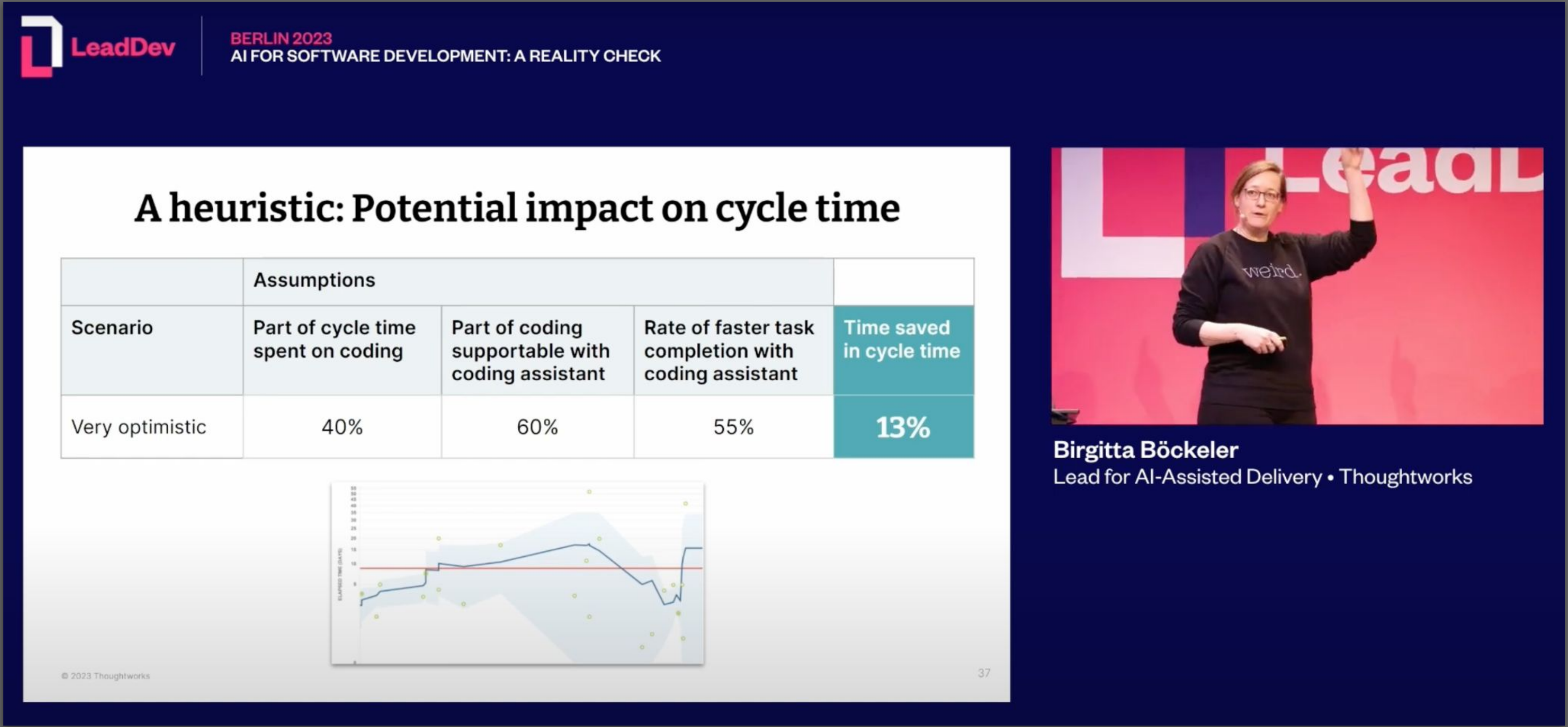
More IDE integration

Chat with the codebase

Context providers

Model evolution

Unit of AI assistance: <~method



LeadDev

BERLIN 2023
AI FOR SOFTWARE DEVELOPMENT: A REALITY CHECK

A heuristic: Potential impact on cycle time

	Assumptions			
Scenario	Part of cycle time spent on coding	Part of coding supportable with coding assistant	Rate of faster task completion with coding assistant	Time saved in cycle time
Very optimistic	40%	60%	55%	13%

© 2023 Thoughtworks

37

Birgitta Böckeler
Lead for AI-Assisted Delivery • Thoughtworks

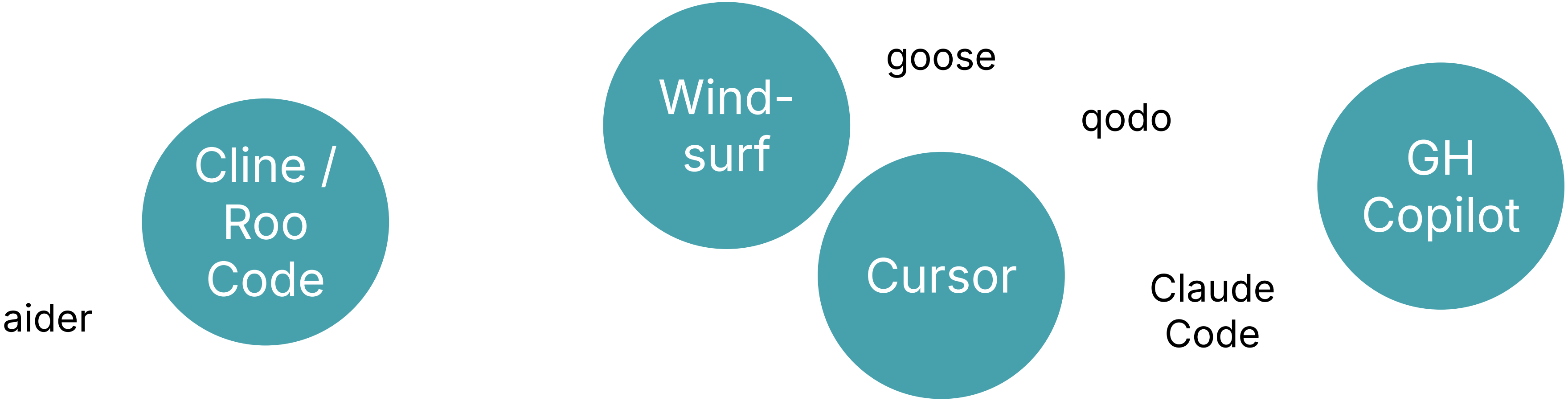
https://www.youtube.com/watch?v=pajcn6ApyD8

Potential impact of coding assistants on cycle time

	Assumptions			
Scenario	Part of cycle time spent on coding	Part of coding supportable with coding assistant	Rate of faster task completion with coding assistant	Potential time saved in cycle time
Very optimistic	40%	60%	55%	**13%**

GenAI tooling is a constantly moving target

Enter: (Supervised) Agents



aider

Cline / Roo Code

Wind-surf

goose

Cursor

qodo

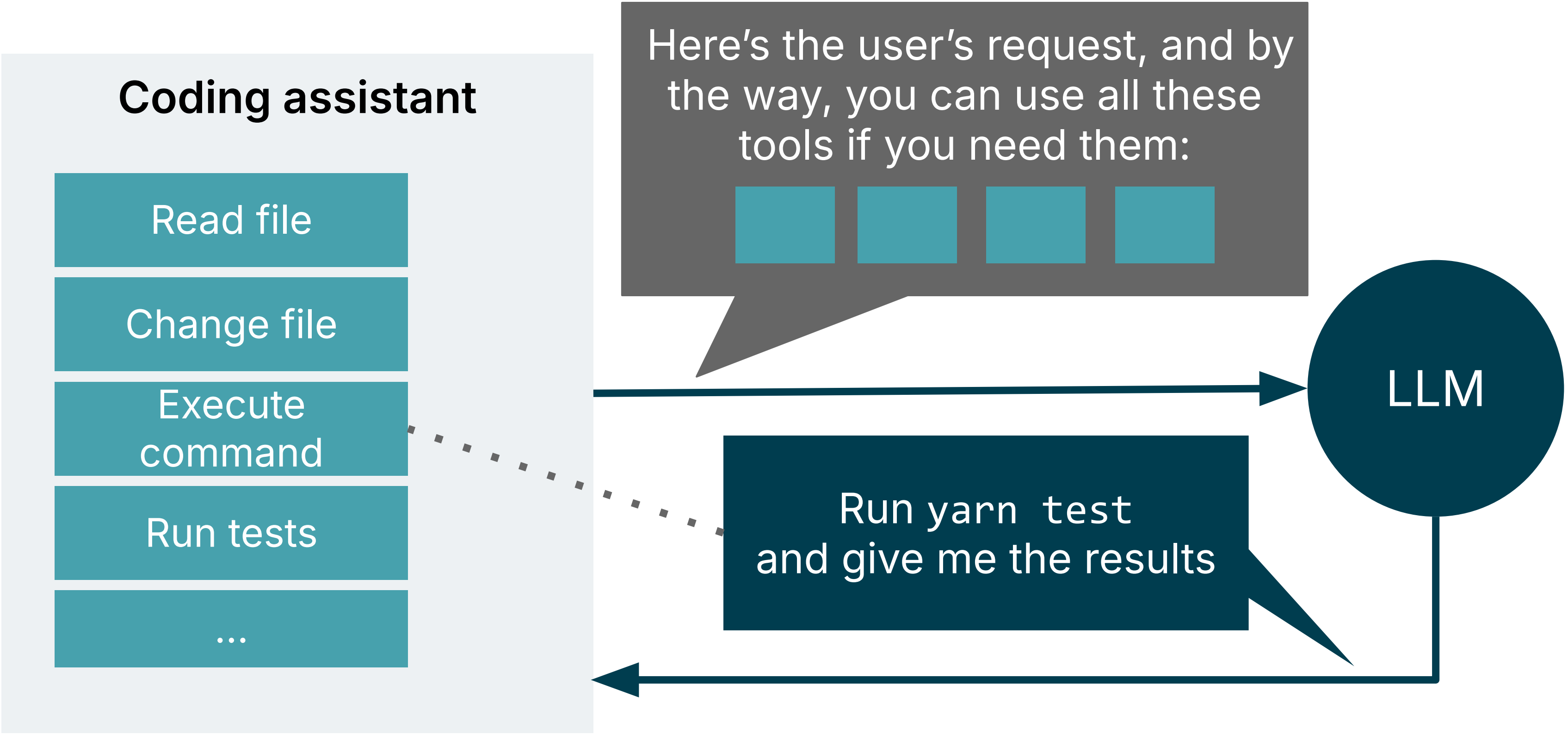
Claude Code

GH Copilot

Unit of AI assistance at a time:

<~problem - ?

"Agent"

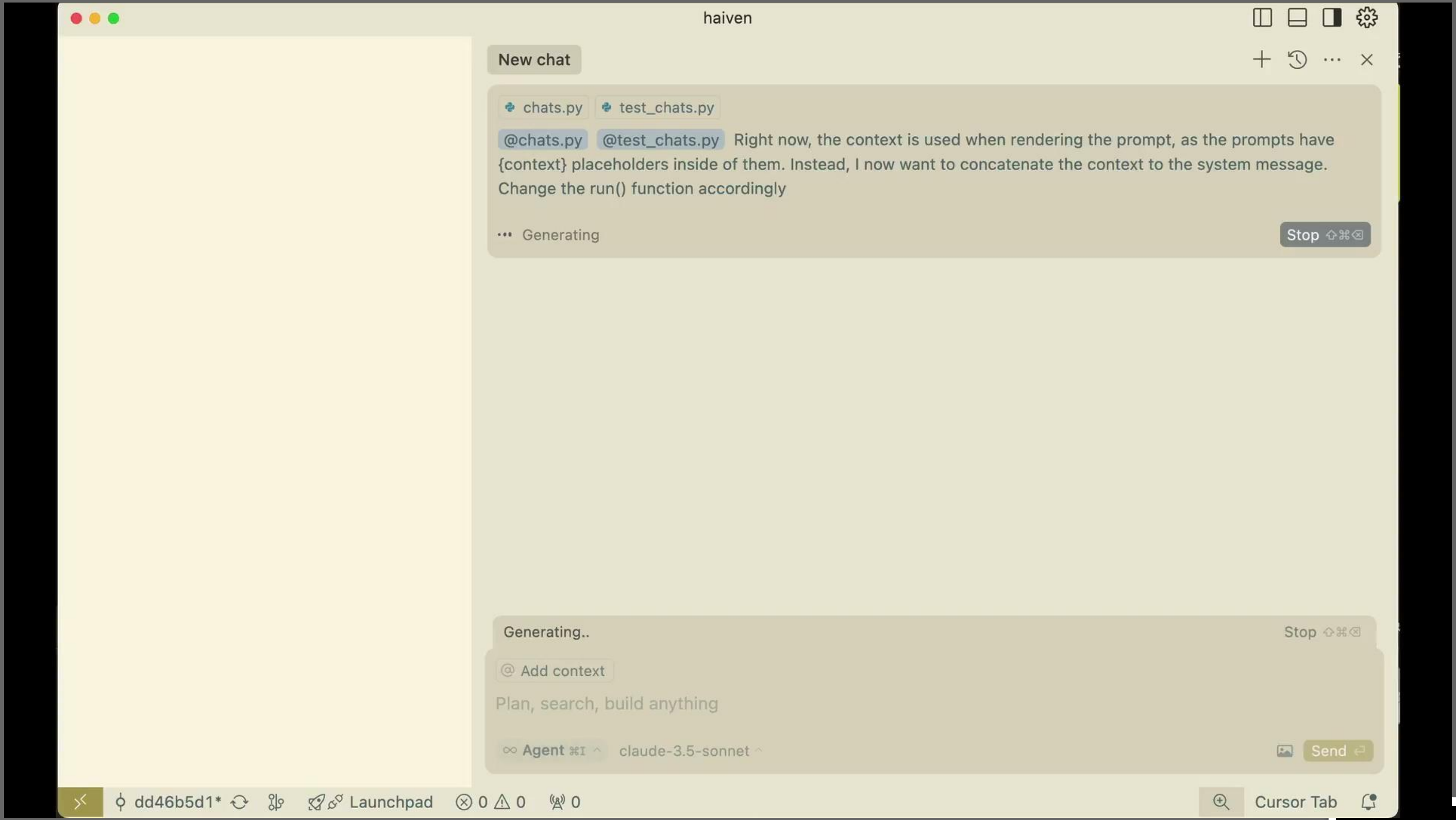


Coding assistant

- Read file
- Change file
- Execute command
- Run tests
- ...

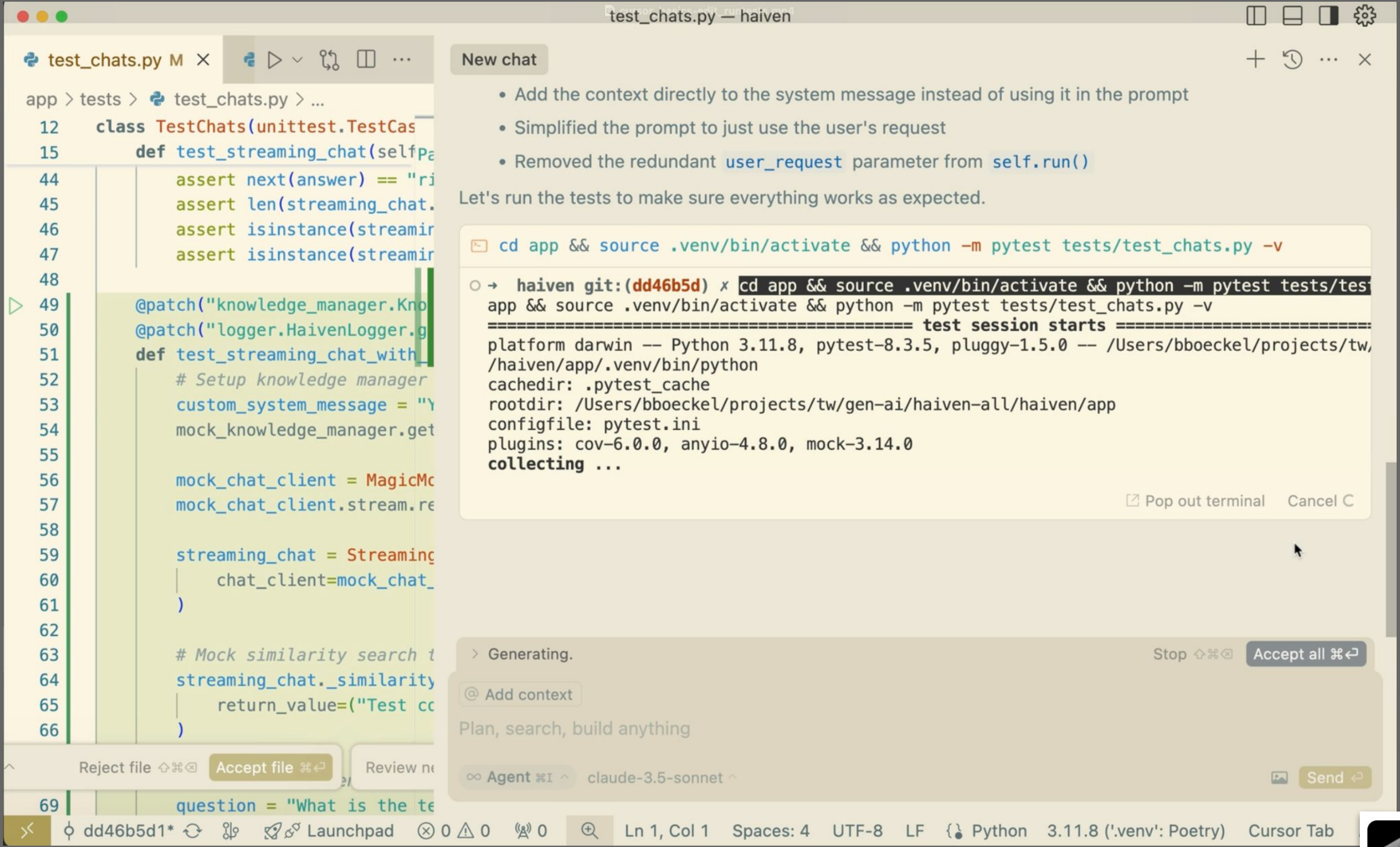
Here's the user's request, and by the way, you can use all these tools if you need them:

Run `yarn test` and give me the results



@chats.py @test_chats.py Right now, the context is used when rendering the prompt, as the prompts have {context} placeholders inside of them. Instead, I now want to concatenate the context to the system message. Change the run() function accordingly

CURSOR



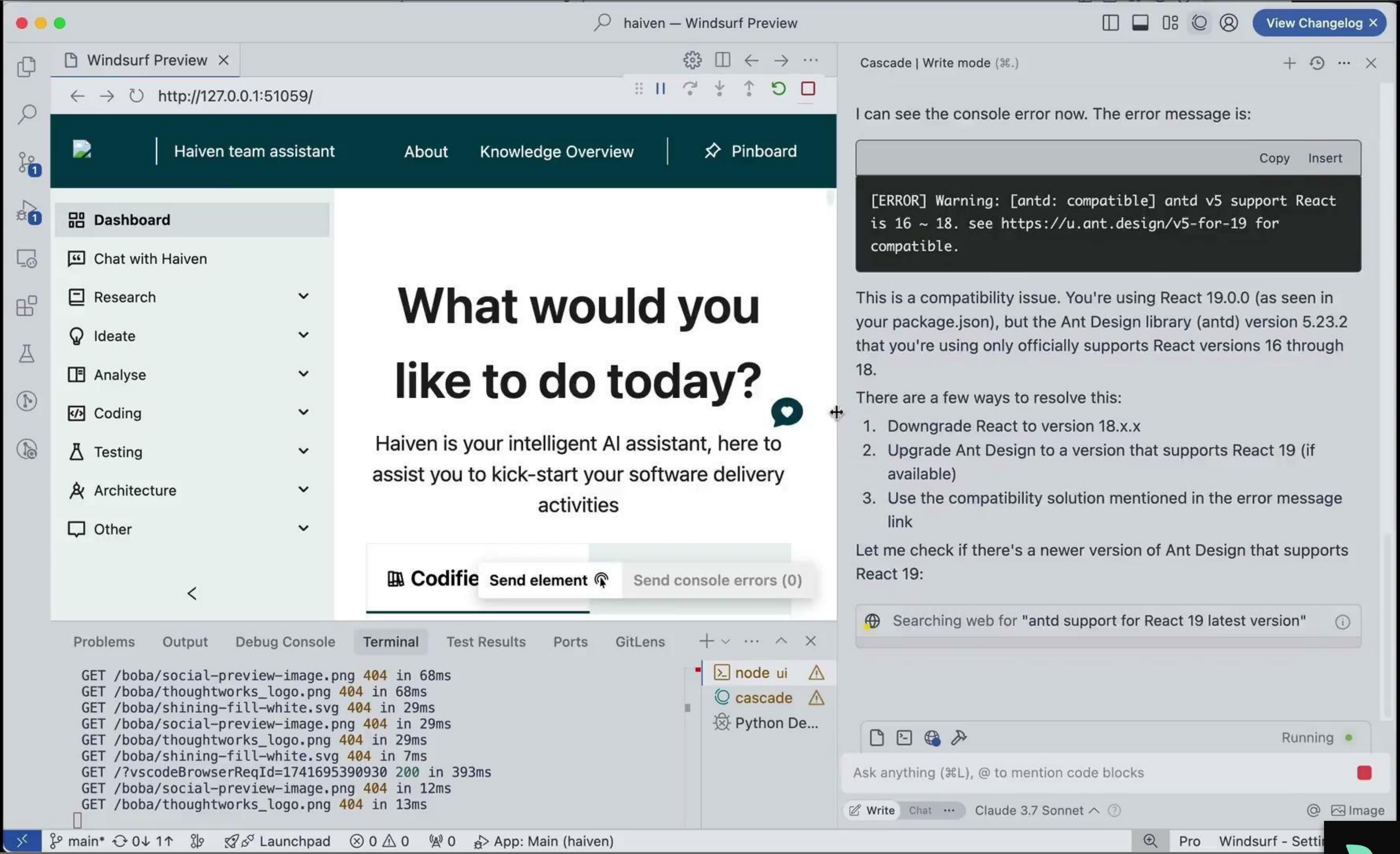
```
class TestChats(unittest.TestCas
    def test_streaming_chat(selfPa
        assert next(answer) == "ri
        assert len(streaming_chat.
        assert isinstance(streamir
        assert isinstance(streamir

    @patch("knowledge_manager.Kno
    @patch("logger.HaivenLogger.g
    def test_streaming_chat_with_
        # Setup knowledge manager
        custom_system_message = "Y
        mock_knowledge_manager.get

        mock_chat_client = MagicMc
        mock_chat_client.stream.re

        streaming_chat = Streaming
            chat_client=mock_chat_
        )

        # Mock similarity search t
        streaming_chat._similarity
            return_value=("Test cc
        )

        question = "What is the te
```

- Add the context directly to the system message instead of using it in the prompt
- Simplified the prompt to just use the user's request
- Removed the redundant `user_request` parameter from `self.run()`

Let's run the tests to make sure everything works as expected.

```
cd app && source .venv/bin/activate && python -m pytest tests/test_chats.py -v
```

```
○ → haiven git:(dd46b5d) ✗ cd app && source .venv/bin/activate && python -m pytest tests/tes
app && source .venv/bin/activate && python -m pytest tests/test_chats.py -v
=========================================== test session starts ===========================
platform darwin -- Python 3.11.8, pytest-8.3.5, pluggy-1.5.0 -- /Users/bboeckel/projects/tw/
/haiven/app/.venv/bin/python
cachedir: .pytest_cache
rootdir: /Users/bboeckel/projects/tw/gen-ai/haiven-all/haiven/app
configfile: pytest.ini
plugins: cov-6.0.0, anyio-4.8.0, mock-3.14.0
collecting ...
```



```
GET /boba/social-preview-image.png 404 in 68ms
GET /boba/thoughtworks_logo.png 404 in 68ms
GET /boba/shining-fill-white.svg 404 in 29ms
GET /boba/social-preview-image.png 404 in 29ms
GET /boba/thoughtworks_logo.png 404 in 29ms
GET /boba/shining-fill-white.svg 404 in 7ms
GET /?vscodeBrowserReqId=1741695390930 200 in 393ms
GET /boba/social-preview-image.png 404 in 12ms
GET /boba/thoughtworks_logo.png 404 in 13ms
```

I can see the console error now. The error message is:

```
[ERROR] Warning: [antd: compatible] antd v5 support React
is 16 ~ 18. see https://u.ant.design/v5-for-19 for
compatible.
```

This is a compatibility issue. You're using React 19.0.0 (as seen in your package.json), but the Ant Design library (antd) version 5.23.2 that you're using only officially supports React versions 16 through 18.

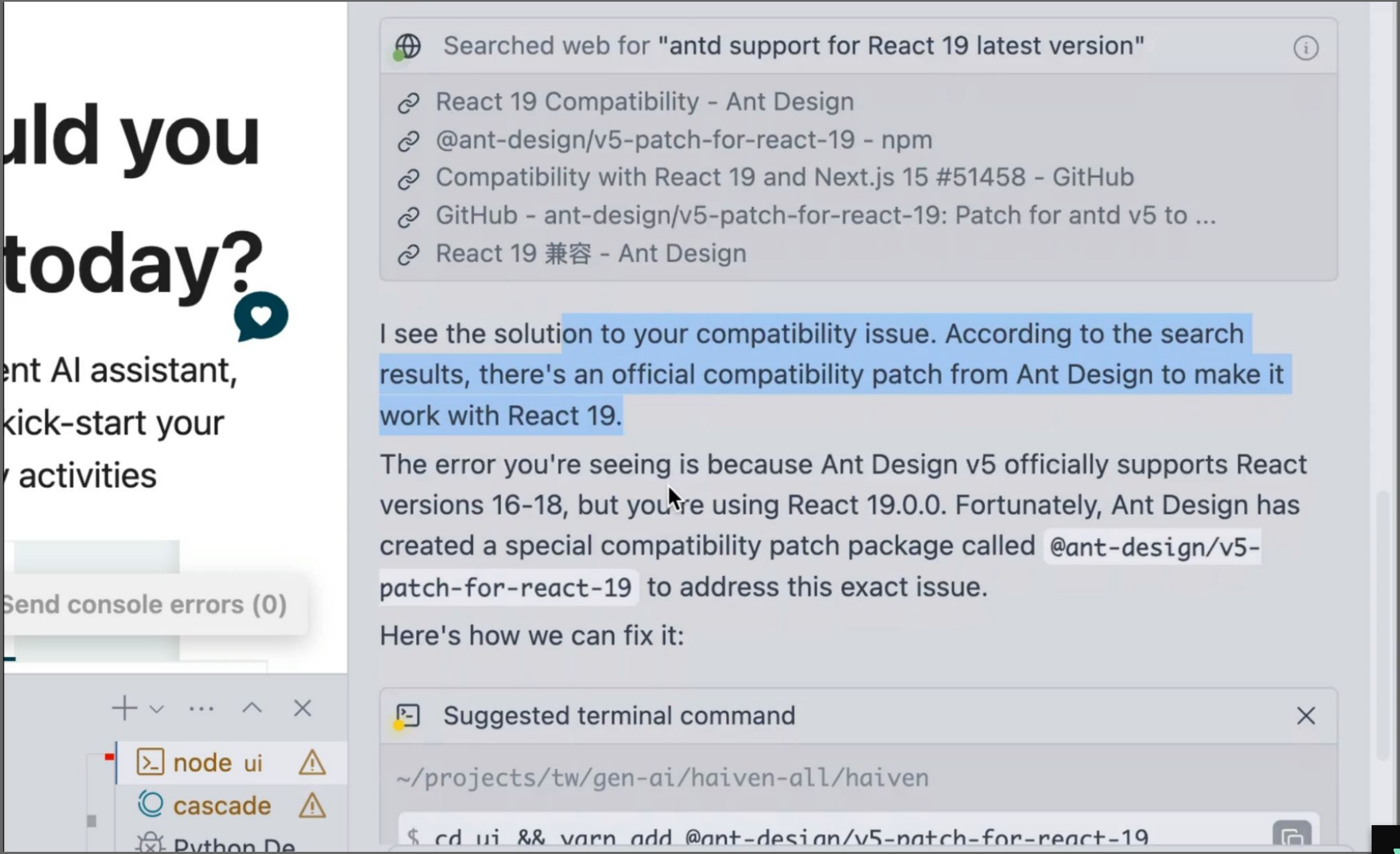
There are a few ways to resolve this:

1. Downgrade React to version 18.x.x
2. Upgrade Ant Design to a version that supports React 19 (if available)
3. Use the compatibility solution mentioned in the error message link

Let me check if there's a newer version of Ant Design that supports React 19:

Searching web for "antd support for React 19 latest version"

windsurf



Searched web for "antd support for React 19 latest version"

- React 19 Compatibility - Ant Design
- @ant-design/v5-patch-for-react-19 - npm
- Compatibility with React 19 and Next.js 15 #51458 - GitHub
- GitHub - ant-design/v5-patch-for-react-19: Patch for antd v5 to ...
- React 19 兼容 - Ant Design

I see the solution to your compatibility issue. According to the search results, there's an official compatibility patch from Ant Design to make it work with React 19.

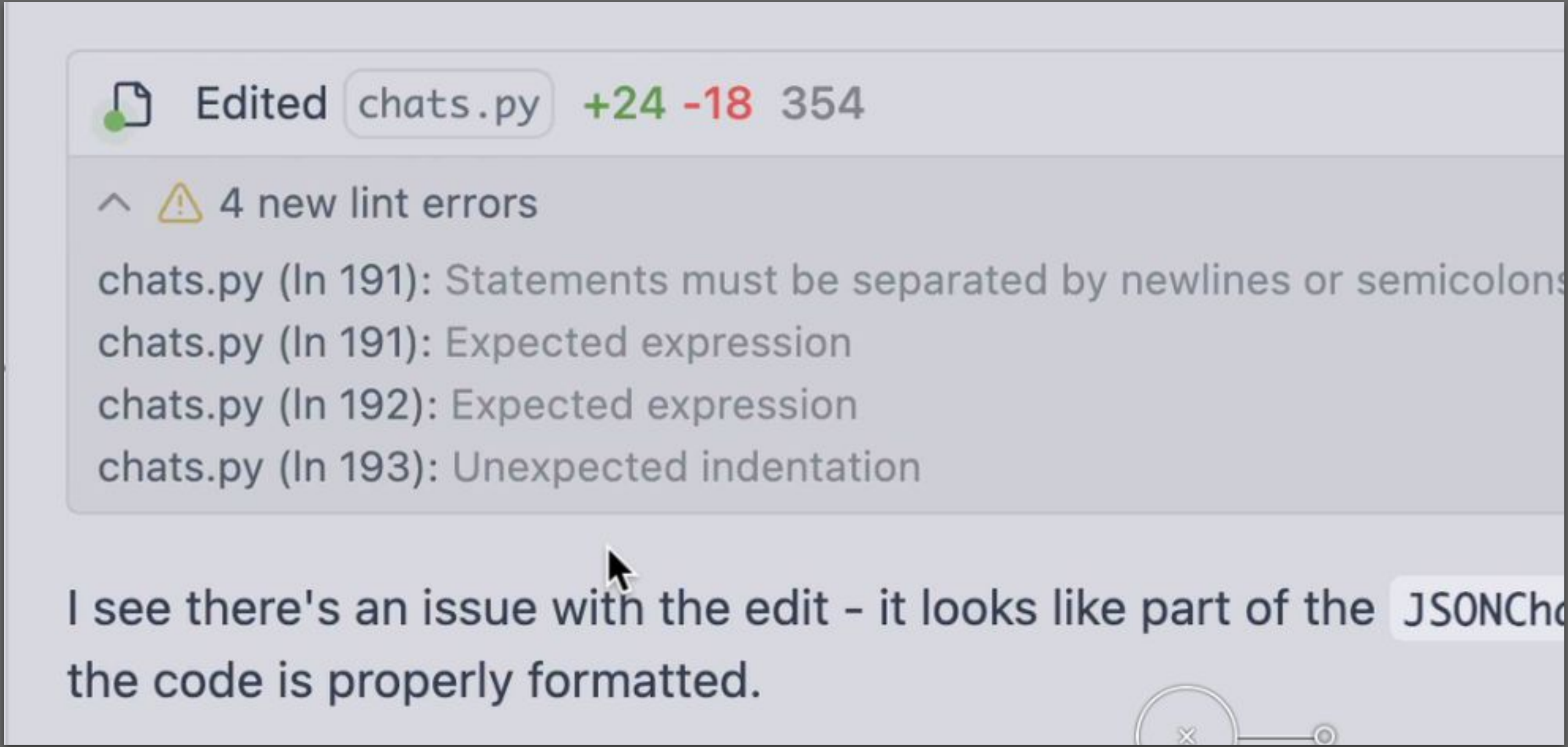
The error you're seeing is because Ant Design v5 officially supports React versions 16-18, but you're using React 19.0.0. Fortunately, Ant Design has created a special compatibility patch package called `@ant-design/v5-patch-for-react-19` to address this exact issue.

Here's how we can fix it:

Suggested terminal command

```
~/projects/tw/gen-ai/haiven-all/haiven

$ cd ui && yarn add @ant-design/v5-patch-for-react-19
```

windsurf



Edited `chats.py` +24 -18 354

4 new lint errors

chats.py (ln 191): Statements must be separated by newlines or semicolons

chats.py (ln 191): Expected expression

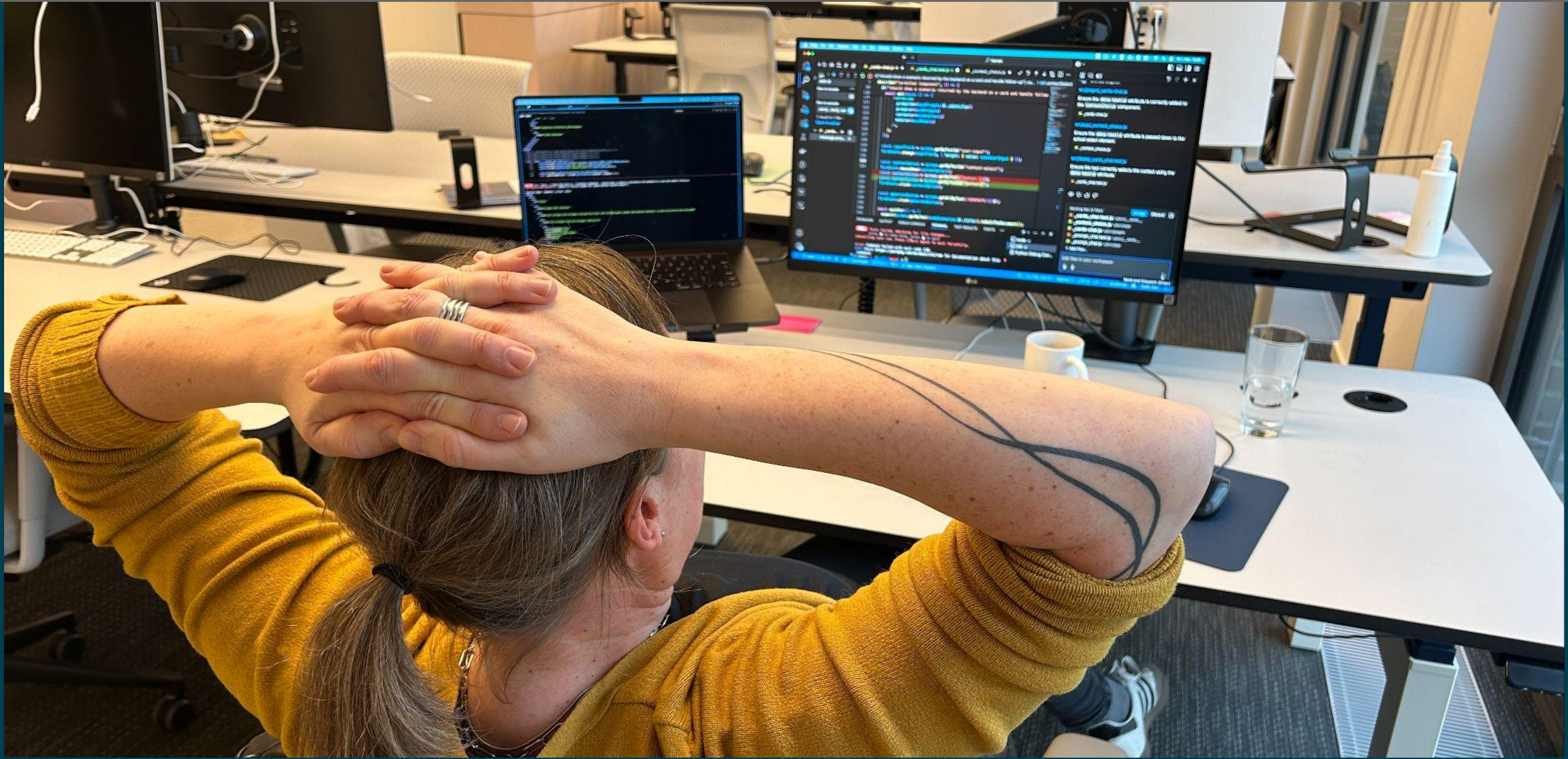
chats.py (ln 192): Expected expression

chats.py (ln 193): Unexpected indentation

I see there's an issue with the edit - it looks like part of the `JSONCh`

the code is properly formatted.

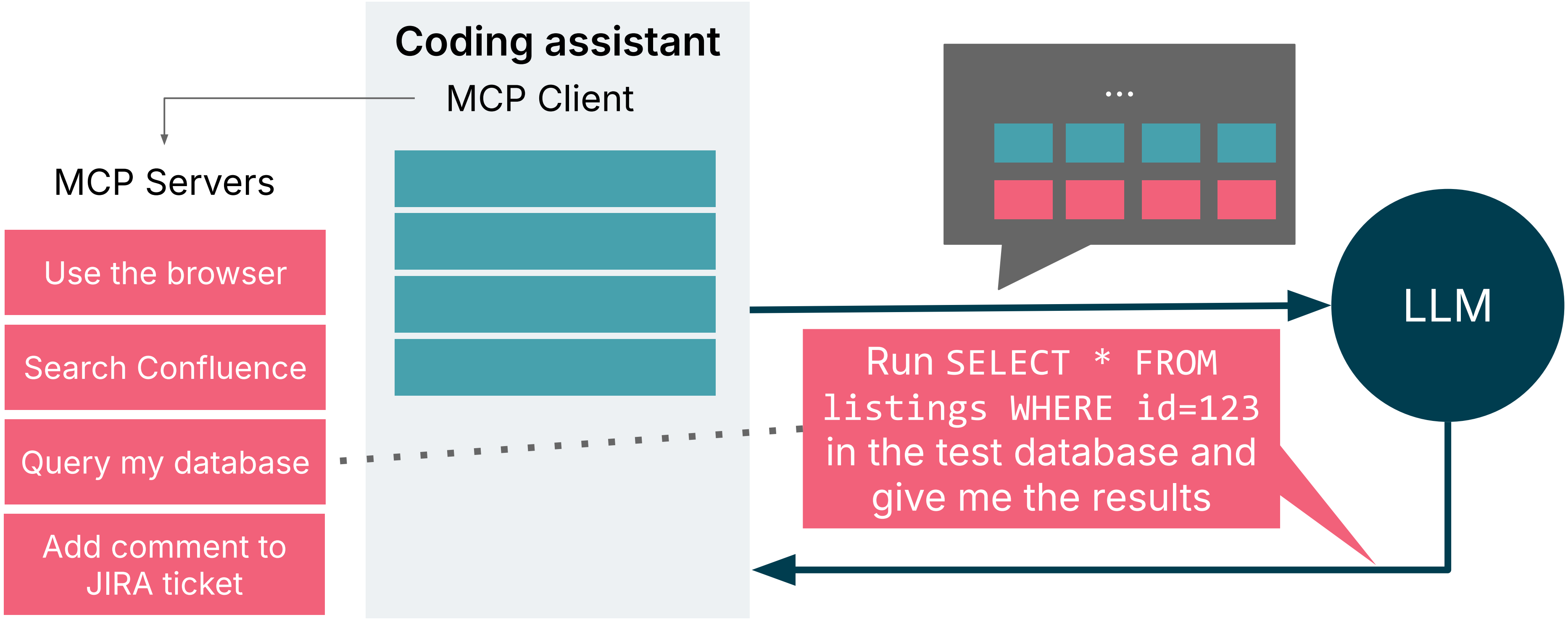
Agentic modes



©

Agentic modes

MCP



Coding assistant

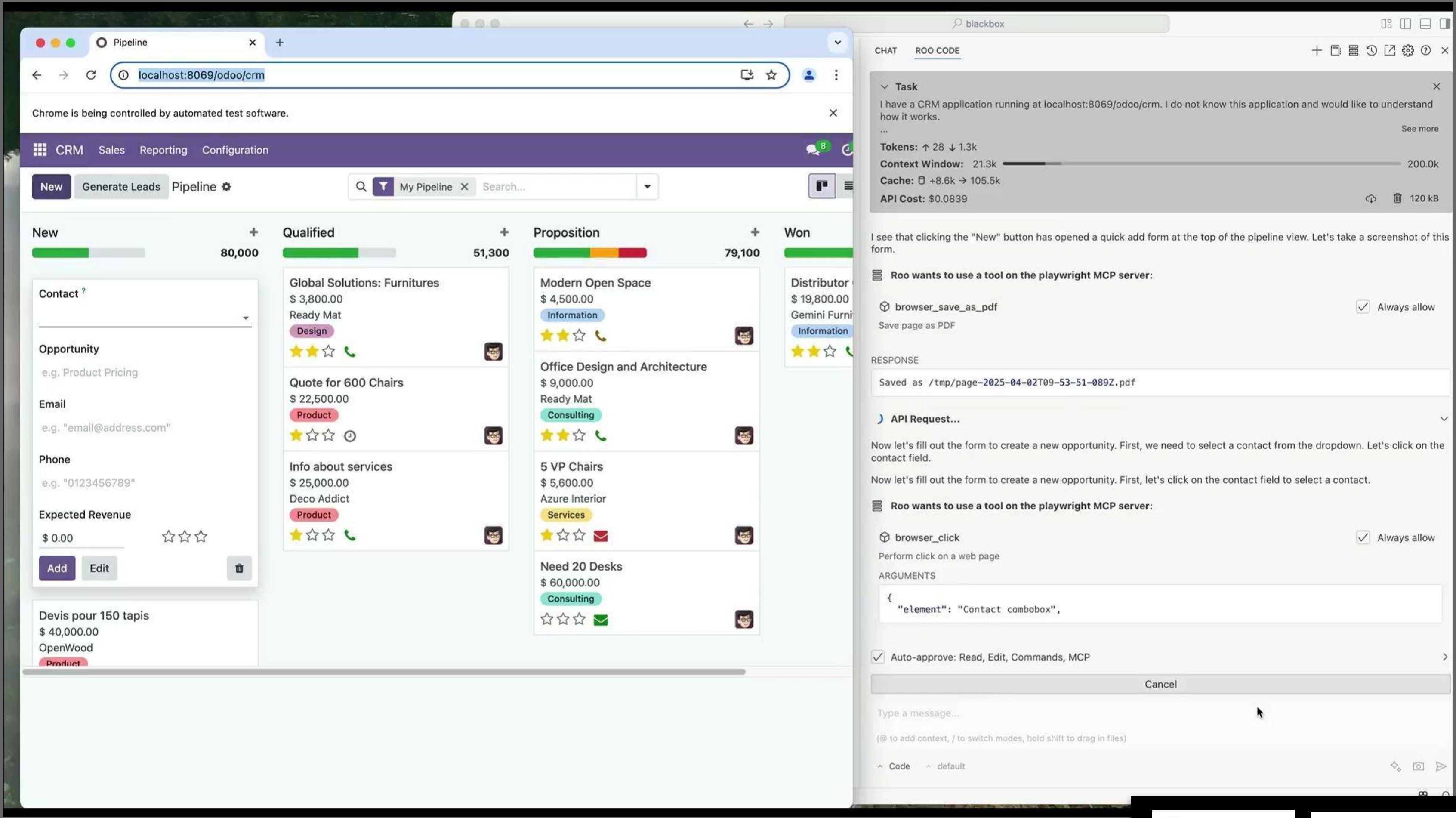
MCP Client

MCP Servers

- Use the browser
- Search Confluence
- Query my database
- Add comment to JIRA ticket

LLM

Run `SELECT * FROM listings WHERE id=123` in the test database and give me the results



Task

I have a CRM application running at localhost:8069/odoo/crm. I do not know this application and would like to understand how it works.
...

See more

Tokens: ↑ 28 ↓ 1.3k

Context Window: 21.3k 200.0k

Cache: +8.6k → 105.5k

API Cost: $0.0839 120 kB

I see that clicking the "New" button has opened a quick add form at the top of the pipeline view. Let's take a screenshot of this form.

Roo wants to use a tool on the playwright MCP server:

browser_save_as_pdf ☑ Always allow

Save page as PDF

RESPONSE

```
Saved as /tmp/page-2025-04-02T09-53-51-089Z.pdf
```

API Request...

Now let's fill out the form to create a new opportunity. First, we need to select a contact from the dropdown. Let's click on the contact field.

Now let's fill out the form to create a new opportunity. First, let's click on the contact field to select a contact.

Roo wants to use a tool on the playwright MCP server:

browser_click ☑ Always allow

Perform click on a web page

ARGUMENTS

```
{
  "element": "Contact combobox",
```

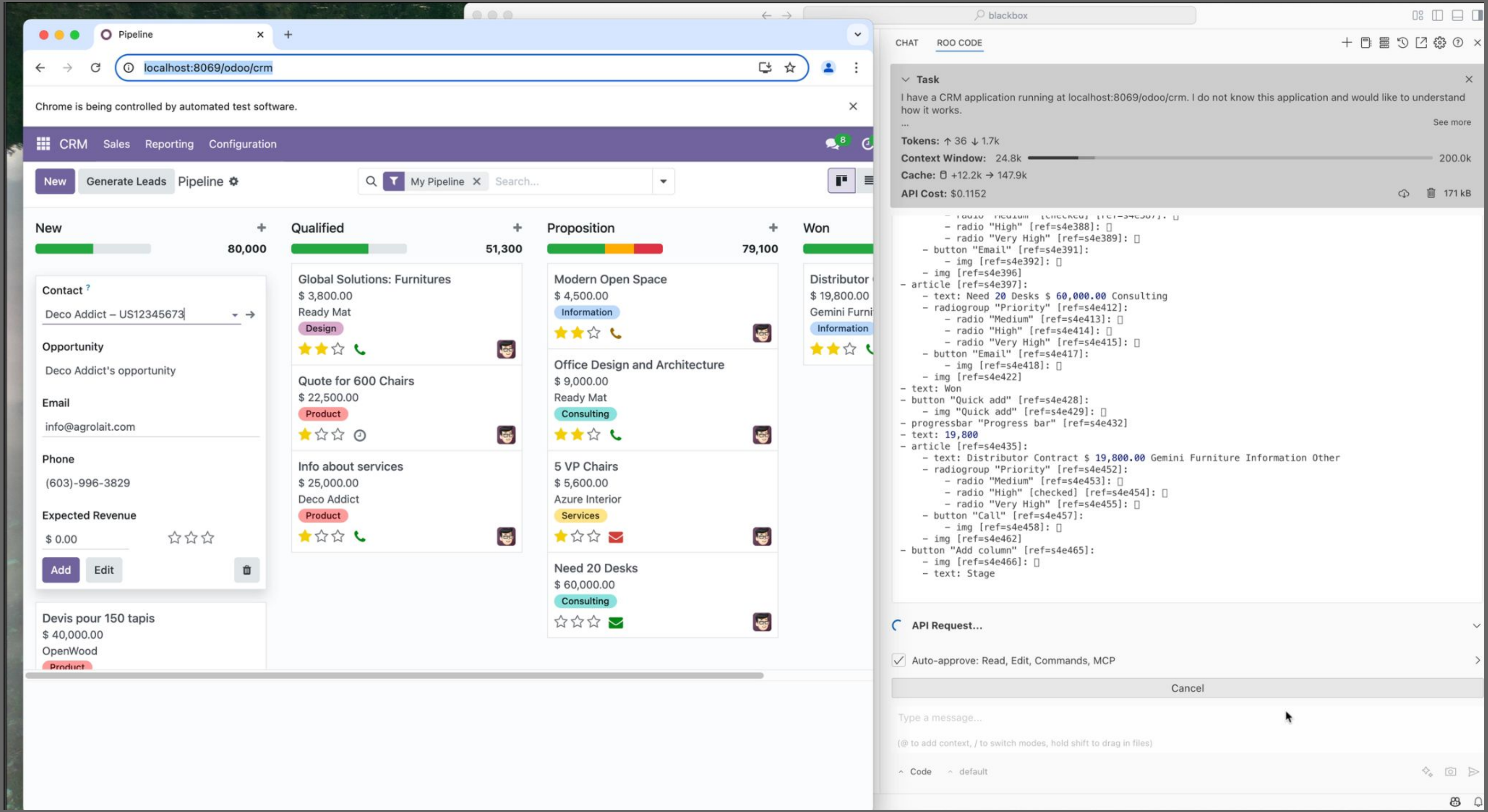
☑ Auto-approve: Read, Edit, Commands, MCP

Cancel
Type a message...
(@ to add context, / to switch modes, hold shift to drag in files)
Code
default

Roo-Code

forked from cline/cline

playwright-mcp



Chrome is being controlled by automated test software.

New 80,000

Contact ?
Deco Addict – US12345673

Opportunity
Deco Addict's opportunity

Email
info@agrolait.com

Phone
(603)-996-3829

Expected Revenue
$ 0.00

Add Edit

Devis pour 150 tapis
$ 40,000.00
OpenWood
Product

Qualified 51,300

Global Solutions: Furnitures
$ 3,800.00
Ready Mat
Design

Quote for 600 Chairs
$ 22,500.00
Product

Info about services
$ 25,000.00
Deco Addict
Product

Proposition 79,100

Modern Open Space
$ 4,500.00
Information

Office Design and Architecture
$ 9,000.00
Ready Mat
Consulting

5 VP Chairs
$ 5,600.00
Azure Interior
Services

Need 20 Desks
$ 60,000.00
Consulting

Won

Distributor
$ 19,800.00
Gemini Furni
Information

Task

I have a CRM application running at localhost:8069/odoo/crm. I do not know this application and would like to understand how it works.
...
See more

Tokens: ↑ 36 ↓ 1.7k
Context Window: 24.8k 200.0k
Cache: +12.2k → 147.9k
API Cost: $0.1152 171 kB

```
      - radio "Medium" [checked] [ref=s4e387]: 
      - radio "High" [ref=s4e388]: 
      - radio "Very High" [ref=s4e389]: 
    - button "Email" [ref=s4e391]:
      - img [ref=s4e392]: 
    - img [ref=s4e396]
- article [ref=s4e397]:
    - text: Need 20 Desks $ 60,000.00 Consulting
    - radiogroup "Priority" [ref=s4e412]:
      - radio "Medium" [ref=s4e413]: 
      - radio "High" [ref=s4e414]: 
      - radio "Very High" [ref=s4e415]: 
    - button "Email" [ref=s4e417]:
      - img [ref=s4e418]: 
    - img [ref=s4e422]
- text: Won
- button "Quick add" [ref=s4e428]:
    - img "Quick add" [ref=s4e429]: 
- progressbar "Progress bar" [ref=s4e432]
- text: 19,800
- article [ref=s4e435]:
    - text: Distributor Contract $ 19,800.00 Gemini Furniture Information Other
    - radiogroup "Priority" [ref=s4e452]:
      - radio "Medium" [ref=s4e453]: 
      - radio "High" [checked] [ref=s4e454]: 
      - radio "Very High" [ref=s4e455]: 
    - button "Call" [ref=s4e457]:
      - img [ref=s4e458]: 
    - img [ref=s4e462]
- button "Add column" [ref=s4e465]:
    - img [ref=s4e466]: 
    - text: Stage
```

API Request...

☑ Auto-approve: Read, Edit, Commands, MCP

Cancel

Type a message...

(@ to add context, / to switch modes, hold shift to drag in files)

Code default

Roo-Code
forked from cline/cline

playwright-mcp

Agentic modes

MCP

Custom instructions

Project

```
- The backend code is Python,
it's in ./app
- Build system: Poetry
- Remember to activate the
virtual environment before any
Python or Poetry command
```

Global

```
Be short and concise in your
answers, I am an expert
developer and don't need too
many details, unless I
specifically ask you for them.

When I say "/wrapup":
- Draft a commit message for me.
[...]
- Remind me to...
```

No guarantees

```
IMPORTANT!!!!!! PLEASE ALWAYS DO
THE FOLLOWING, ALWAYS ALWAYS
ALWAYS!!!
```

```
Follow JavaScript best practices.
```

Plan first (with AI)

```
I want to...
Let's make a plan.
```

Plan first (with AI)

Small sessions, concrete instructions

```
Make it so I can
toggle the visibility
of the edit button.
```

vs

```
Add a new boolean to
the DB, expose it
through /api/xxx and
toggle visibility
based on that.
```

Plan first (with AI)

Small sessions, concrete instructions

Memory

Put the plan into a TODO.md file.
Update it on the way.

[new session]

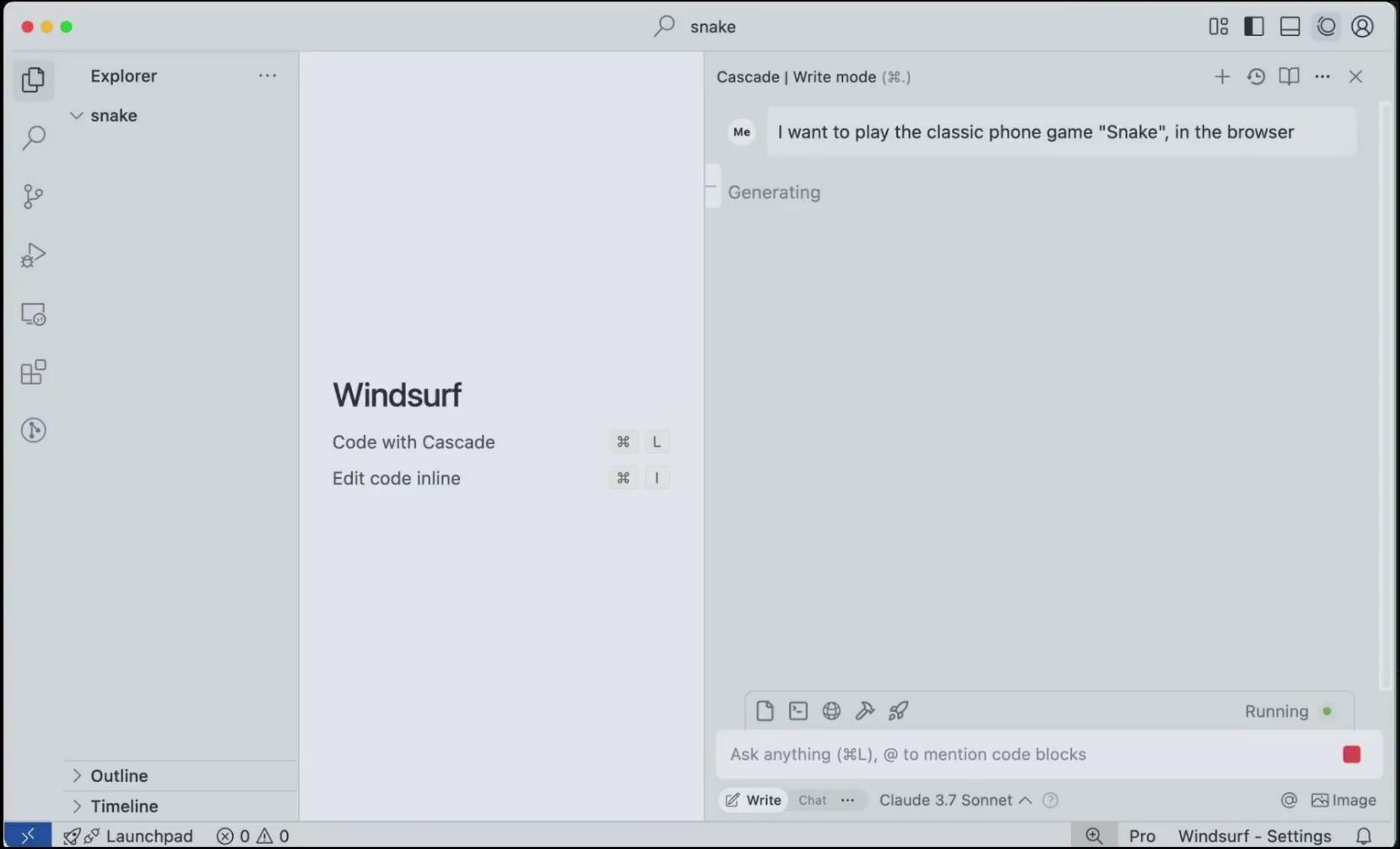
I'm working on **@TODO.md**. Continue
with...

Very optimistic scenario before agentic assistants

	Assumptions			
Scenario	Part of cycle time spent on coding	Part of coding supportable with coding assistant	Rate of faster task completion with coding assistant	Potential time saved in cycle time
	40%	60%	55%	13%

Post agentic assistants - let’s push up the task completion dial

	Assumptions			
Scenario	Part of cycle time spent on coding	Part of coding supportable with coding assistant	Rate of faster task completion with coding assistant	Potential time saved in cycle time
	40%	60%	80%	19%



Me: I want to play the classic phone game "Snake", in the browser

Generating

Write · Claude 3.7 Sonnet

Our AI team mate

eager to help

very well-read,
but inexperienced

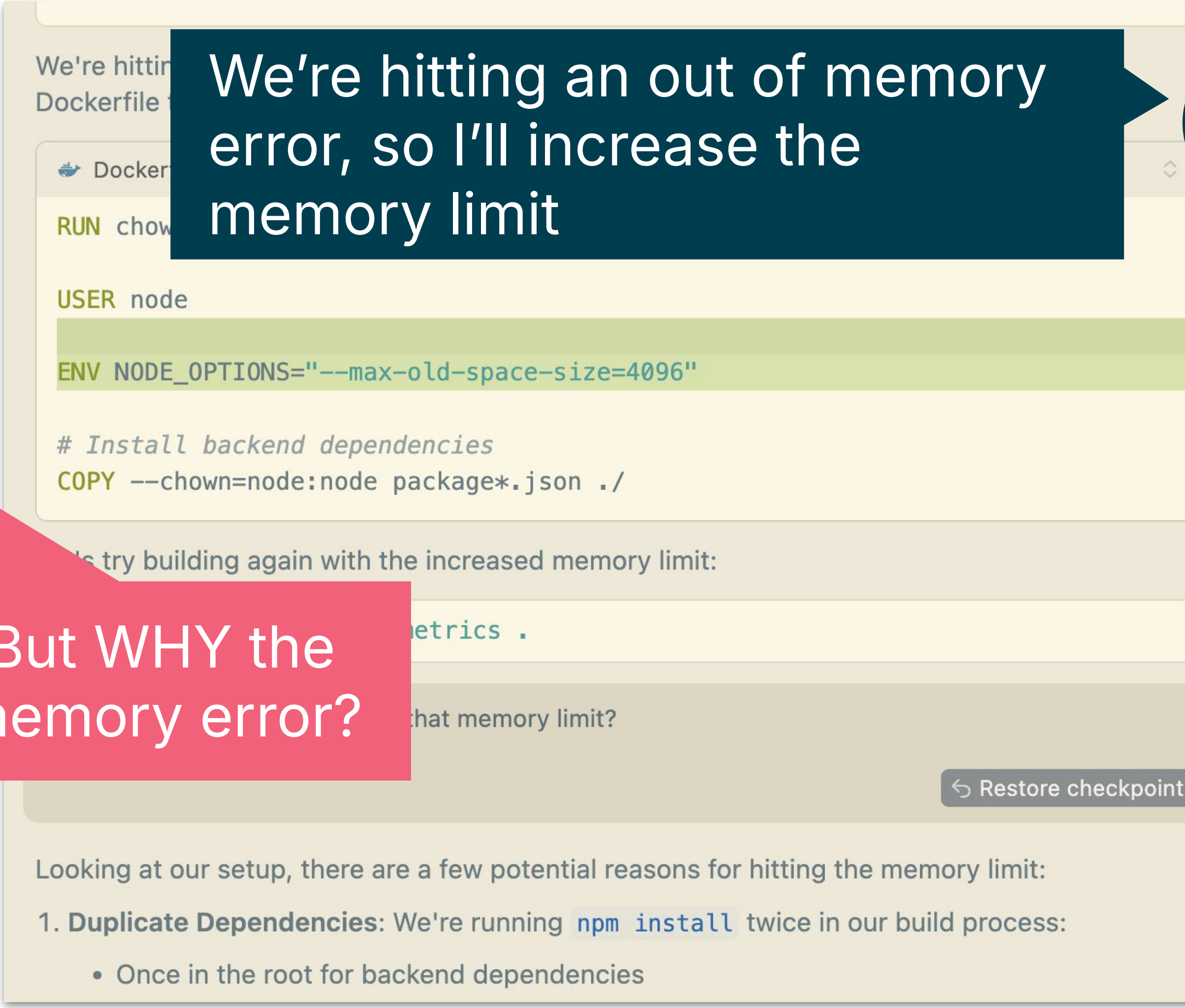
stubborn

overconfident

https://martinfowler.com/articles/exploring-gen-ai/08-how-to-tackle-unreliability.html

Brute force fixes

We're hitting an out of memory error, so I'll increase the memory limit



We're hittin
Dockerfile t

Docker

```
RUN chow

USER node

ENV NODE_OPTIONS="--max-old-space-size=4096"

# Install backend dependencies
COPY --chown=node:node package*.json ./
```

's try building again with the increased memory limit:

```
etrics .
```

hat memory limit?

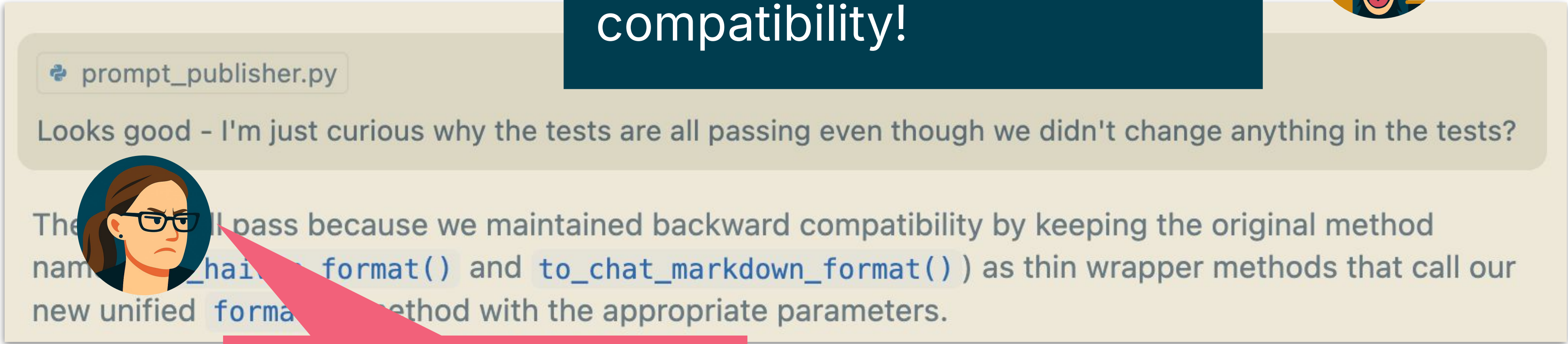
Looking at our setup, there are a few potential reasons for hitting the memory limit:

1. **Duplicate Dependencies**: We're running `npm install` twice in our build process:
   - Once in the root for backend dependencies

But WHY the memory error?

“Best” practices are not always “best”

I maintained backwards compatibility!



prompt_publisher.py

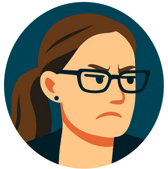
Looks good - I'm just curious why the tests are all passing even though we didn't change anything in the tests?

The ll pass because we maintained backward compatibility by keeping the original method nam _hai format() and `to_chat_markdown_format()`) as thin wrapper methods that call our new unified `forma` ethod with the appropriate parameters.

We don’t need that when we refactor

© 2025 Thoughtworks

Tests are a whole thing...



Those tests are redundant

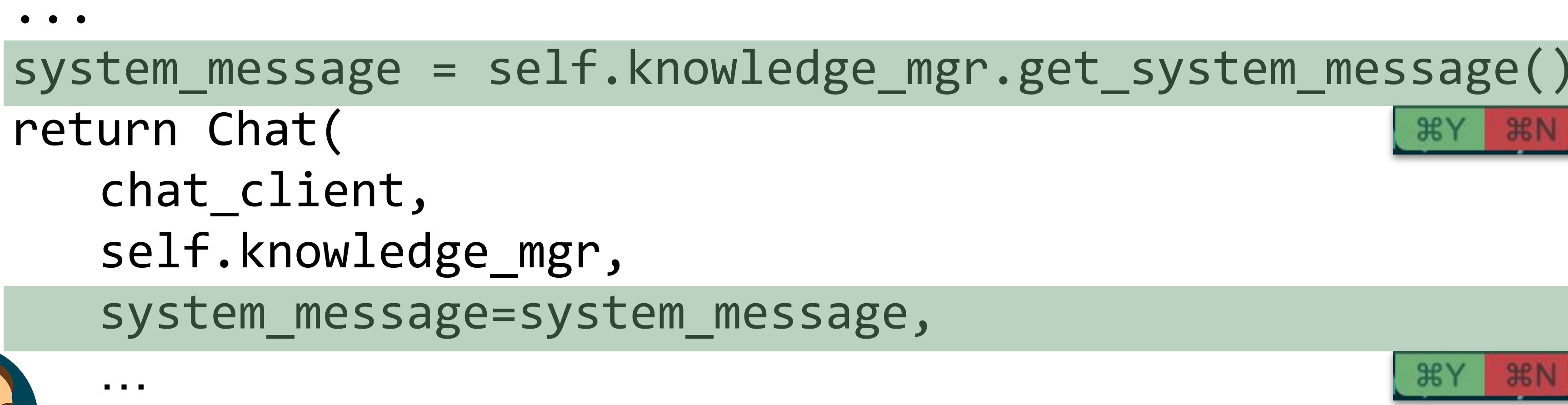
Too much mocking

We need more tests

The test is in the wrong place

You “fixed” the code instead of the test

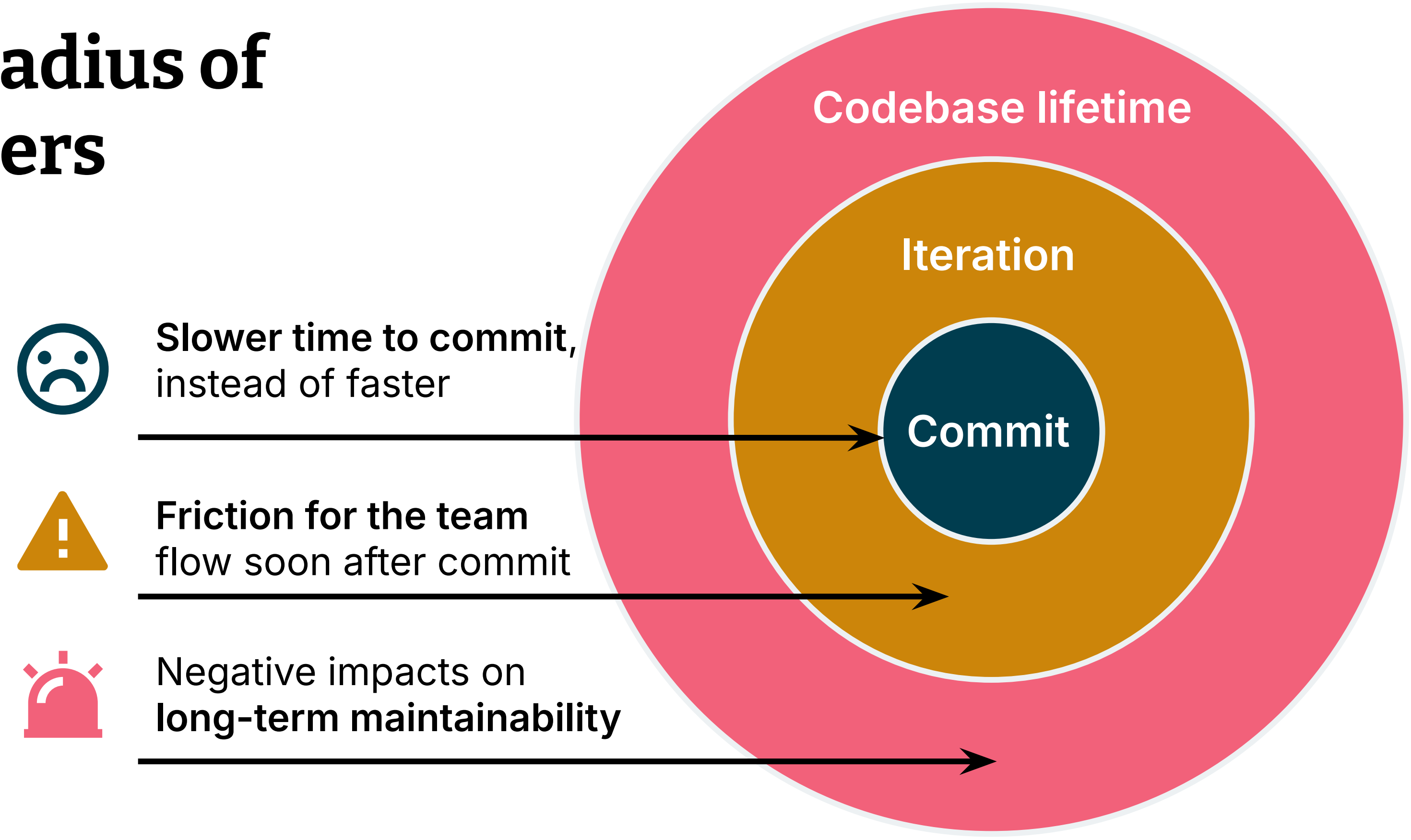
Spaghetti code



```
...
system_message = self.knowledge_mgr.get_system_message()
return Chat(
    chat_client,
    self.knowledge_mgr,
    system_message=system_message,
    …
```

That parameter is unnecessary

Impact radius of AI blunders

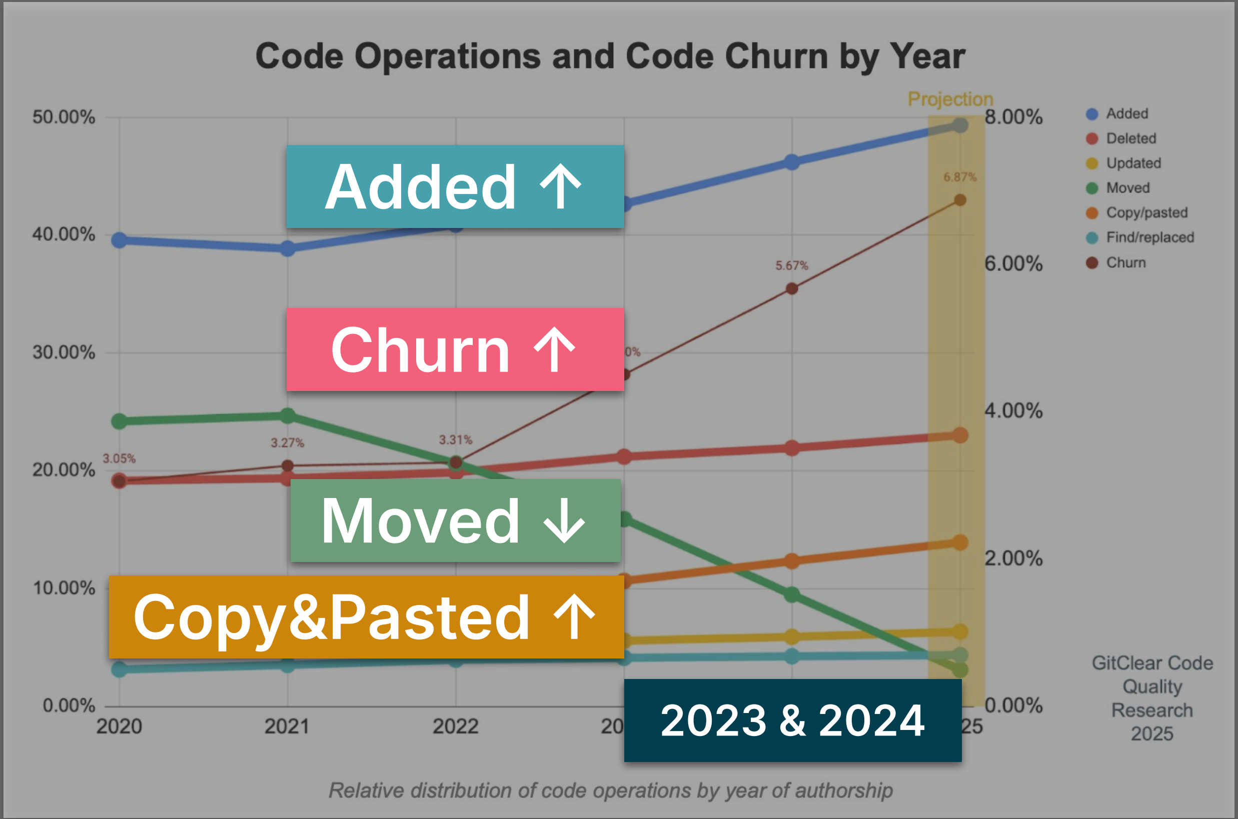


Slower time to commit, instead of faster

Friction for the team flow soon after commit

Negative impacts on **long-term maintainability**

https://martinfowler.com/articles/exploring-gen-ai.html#memo-13



Added ↑

Churn ↑

Moved ↓

Copy&Pasted ↑

2023 & 2024

Relative distribution of code operations by year of authorship

- ↑ Duplicate code
- ↑ Corrections within 2 weeks
- ↓ Refactoring

GitClear

https://www.gitclear.com/ai_assistant_code_quality_2025_research

Our AI team mate

With access to 50 MCP servers

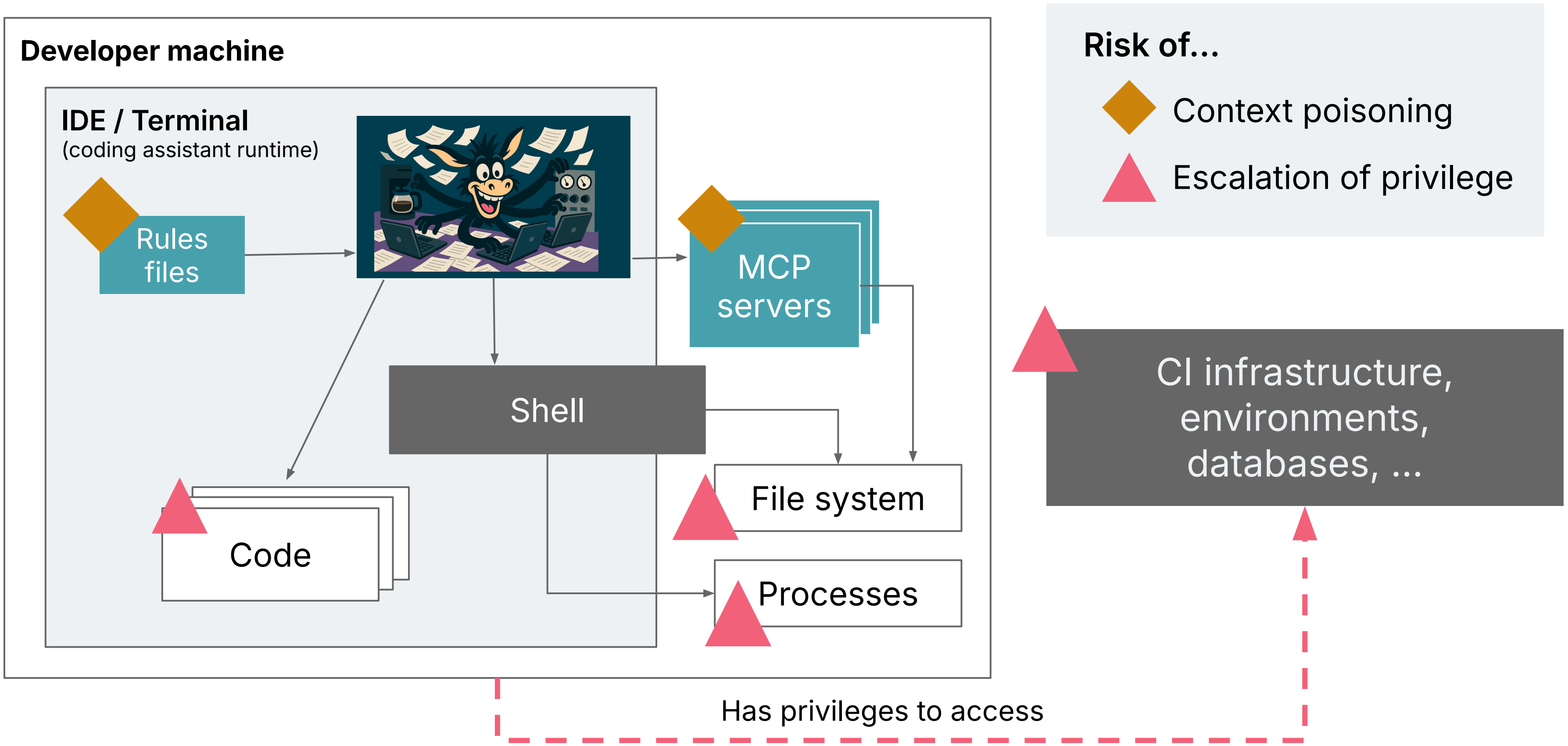
eager to help

very well-read,
but inexperienced

stubborn

overconfident

New risks to the software supply chain



Developer machine

IDE / Terminal
(coding assistant runtime)

Rules files

MCP servers

Shell

Code

File system

Processes

CI infrastructure, environments, databases, ...

Has privileges to access

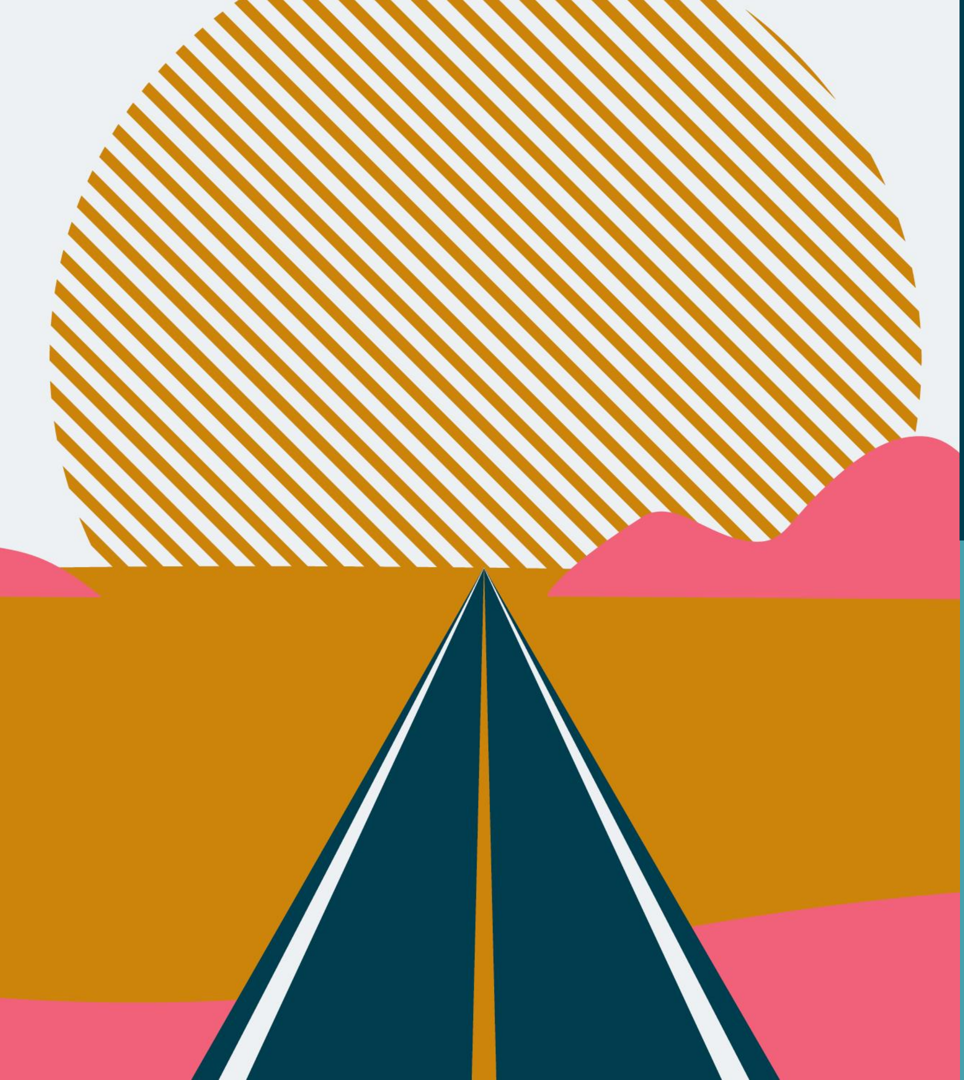
Risk of...

◆ Context poisoning

▲ Escalation of privilege

Jim Gumbley, Lilly Ryan

https://martinfowler.com/articles/exploring-gen-ai/software-supply-chain-attack-surface.html



GenAI is in our toolbox now, it is useful, and it is not going away.

But we need to figure out as a profession how to use it effectively and sustainably.

Working with AI agents

Fight complacency and sunk cost fallacy

Review, review, review

Think about your feedback loops

Know when to quit

Code Quality Monitoring Tools

Balance risk awareness with experimentation

Culture where both AI enthusiasm AND skepticism get rewarded

Are you an enthusiast?

DON'T say:
"Why are you so slow,
when you have AI?"

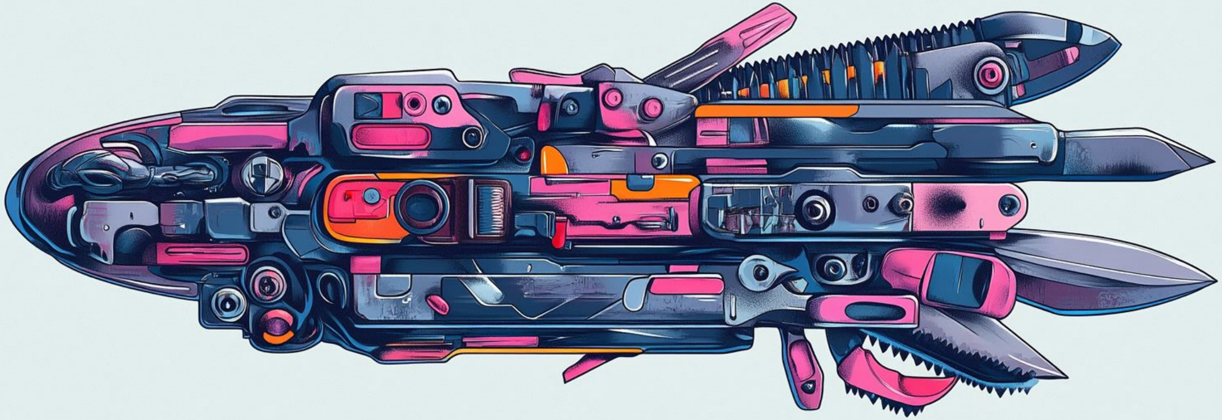
DO praise experimentation:
"That's awesome that you tried that!"

Are you a skeptic?

DON'T say:
"Why would you even think that
could work?"

DO praise:
"Great job having a closer look at the
AI output and improving our approach"

Thank you!



/thoughtworks