

Coding with AI help

Peter Williamson

Overview

1. Hook up copilot (or open source codeium) to your editor (VSCode, Xcode, emacs, vim, etc)

This will provide completions as you type

2. Hook up Chat GPT/Gemini/Claude/Llama (etc) to your editor

This will give you a way to make requests

Stories

Copilot, square, and squareroot

Square

I start typing - the part in bold, and I get

.h file:

```
int square(int x);
```

.cpp file:

```
int square(int x) {  
    return x * x;  
}
```

Square test

I start typing in a gtest file, and I get:

```
TEST_F(CMakeHelloTestTest, Square) {  
    EXPECT_EQ(4, square(2));  
    EXPECT_EQ(9, square(3));  
    EXPECT_EQ(16, square(4));  
}
```

It is correct! The test passes!

Square root

Again I start typing

.h file:

```
int square_root(int x);    // Oops, should probably have used a double here
```

.cpp file:

```
int square_root(int x) {  
    return x * x;  
}
```

Notice there is something a bit off about the body of the function...

Square root test

Then in the test file:

```
TEST_F(CMakeHelloTestTest, SquareRoot) {  
    EXPECT_EQ(2, square_root(4));  
    EXPECT_EQ(3, square_root(9));  
    EXPECT_EQ(4, square_root(16));  
}
```

This is reasonable, and it fails as it should.

Stories - copilot

- Sometimes it reads my TODOs and fixes them!
- Many times it just offers to finish my comment - not that useful
- Sometimes it gets things exactly right (more often than not)
- Sometimes it hallucinates badly as we saw
- I started writing a custom stepper class, and it gave me just what I was thinking about (maybe it used a TODO comment in another file?)

Stories - GPT

Able to suggest how to modify the file I am currently editing, but it loses my comments in its suggestion, so I can't just cut and paste

Helpful for research, but it can hallucinate here too.

How do I do X? ChatGPT may have a good answer

Find the bug in my code (and it did)

Local LLMs - they actually work

I tried using llama.cpp and llamafire technology to run the llama 2 LLMs on my windows machine with a GPU - the 70B model did actually run and produce valid C++ code, but it was pretty slow - slower than I could have typed it myself, whereas ChatGPT 4o is pretty fast.

However, Llama 3 is much faster and seems to produce pretty good code (maybe on a par with GPT 3.5?). Maybe someday local LLMs can be built into our editors.

Conclusions

- Using AI to help you code really is the wave of the future. Productivity will improve, but maybe not by as much as 2x as some are suggesting. At least not yet.
- AIs still hallucinate - don't trust what it gives you. Think of it as you need to code review everything it writes. Can use deprecated APIs
- Still not great for proprietary code bases, but local LLMs can help
- Can save time typing (co-pilot), and with research when working in a new area
- Copilot is magical when it happens, but it often doesn't happen in time to save me typing
- Prompt engineering still matters to get good results even for coding requests.
- LMK if you'd like a longer talk with more how-to in it.