

AI Coding Assistants for Java Developers



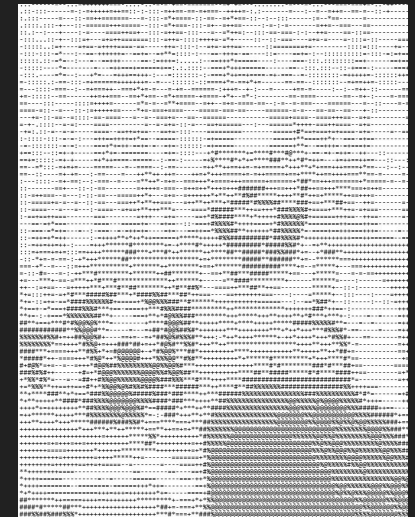
Private AI guy

<https://lu.ma/omahajava>
5:30 to 7pm 16-Jan-2025

<https://github.com/payne/LLM-demos>

<https://bit.ly/16LLM>

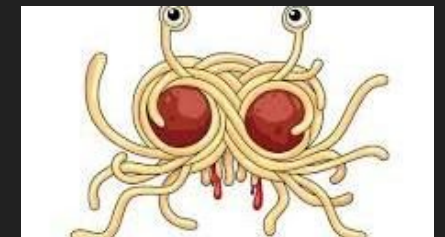
M & A



Java guy

Goals -- Anecdotes about private AI & commercial AI for Java Devs

1. LLMs do not know everything - try asking about HAM radio's world of XLX reflectors and you get squat
2. LLMs for java developers -- we'll report on the ones we tried.
3. IDE integration for LLMs -- Trying Continue.dev and GitHub copilot integrations
4. Web Based UIs are nice. Think of LLMs as an extension of the StackOverflow tradition



But so much more!

<https://bit.ly/16LLM>

M

May the copy
pasta be with
you....

BLUF - Bottom line up front

1. Matt Payne thinks LLMs are just search engines. First there was library books (Remember section QA76 -- so good), then stack overflow, and now Claude.ai. Matt Payne is totally in the bag for Claude.ai -- it's the bee's knees.
2. Aaron Grothe thinks there's more to LLMs than that.... Aaron realized that the first time his co-worker referred to the ChatGPT as "he". "He wrote this code."

Matt Payne's Claude.io history - some highlights

1. **Public:** “[build spring boot java program that uses kafka without zookeeper](#)”
2. **Public:** “[Help me write a nostr client in java](#)”
3. **Private:** “[I'd like to use the northwind database with spring data jdbc. In the famous northwind database schemas what are the Aggregates?](#)”
4. **Private:** “[I want help practicing spring data JPA. What is a classic many to many relationship that I can use for practice?](#)”
5. **Private:** “[If I use spring data rest for the backend to my angular GUI, how can I add server side processing logic when rows are inserted into the database?](#)”
6. **Public:** “Convert the [table in this image](#) to [JSON](#)”

Matt Payne's BARD.google.com history - some highlights

1. Public: [JPA without a nested query.](#)

First prompt: “In spring boot I have a repository that extends JpaRepository. There can be many rows for one foold. I want a method that returns the most recent row according to the requestDate”

Coding assistant

For this we'll be firing up a Private AI coding assistant.

It will run on your local machine

We will turn off telemetry

Components

Ollama - our engine for running Large Language Models

VS Codium - A de-Microsoft'd version of VS Code

Continue - Our AI plugin

<https://bit.ly/16LLM>

A

Hardware Requirements

What do you need to be able to run a private AI coding assistant

Ideally you'll have a discrete GPU, but you can also run ollama on the regular CPU cores

- Mac books with Apple silicon work very well for this. Preferably with 16gb of ram
- My home machine is a HP Z620 with 4 Nvidia k2200 graphics cards - \$20 per card, but not very powerful
- For this demo I'm running on an Acer Aspire 3 with a Ryzen 5700u and built in graphics, right now on cpu, need to get Rocm reenabled

Installing Ollama

Go to <https://ollama.co> and install their LLM engine

```
$ curl -fsSL https://ollama.com/install.sh | sh
```

You trust a shell script off the internet don't you???

If you're running on Windows you can either use the Windows executable or install it in the Windows Subsystem for Linux. Used to be WSL was the way to go, but the Windows stuff is coming along

Private AI Setup on a M1 Mac Mini (16GB RAM)

1. How to run an LLM on your PC, not in the cloud, in less than 10 minutes
 1. `brew install ollama`
 2. `ollama serve`

Installing VSCodium

VS Codium is a modified version of VS Code that has had the telemetry removed. They remove the logos etc. to comply with Microsoft's license

It is to VS Code as Chromium is to Google Chrome

For install instructions please refer to the following

<https://vscodium.com/#install>

Installing Continue

Installing the Continue plugin is most easily done through the Marketplace

<https://marketplace.visualstudio.com/items?itemName=Continue.continue>

Continue homepage

<https://www.continue.dev/>

<https://bit.ly/16LLM>

A

Need to load LLMs for Ollama/Continue

We'll be using three LLMs for this project

General purpose LLM, used for code completion, etc.

```
# ollama pull llama3.2
```

Needed to index codebase on your machine, needed for local chatbot

```
# ollama pull nomic-embed-text
```

<https://bit.ly/16LLM>

A

Need to load LLMs for Ollama/Continue

We'll be using three LLMs for this project

Need a model for tab-autocomplete

```
# ollama pull starcoder2:3b
```

Now we have the 3 models, you can mix/match with other models. E.g. codellama might be an alternative to llama3b

Turn off Telemetry in Continue Plugin

Examples of Telemetry in VS Codium

- Number of Tokens generated
- IDE and OS
- Whether you accept suggestions
- Page Views

You can turn this off if you want

Turn off Telemetry in VS Codium

Turning off telemetry in VS Codium

Telemetry



Continue: Telemetry Enabled



Continue collects anonymous usage data, cleaned of PII, to help us improve the product for our users. Read more at continue.dev >
Telemetry.

Chatting with your Private AI Buddy

We'll start with playing with the example tutorial a bit

Identify the sorting algorithm

Rewrite the code

Autocompletion

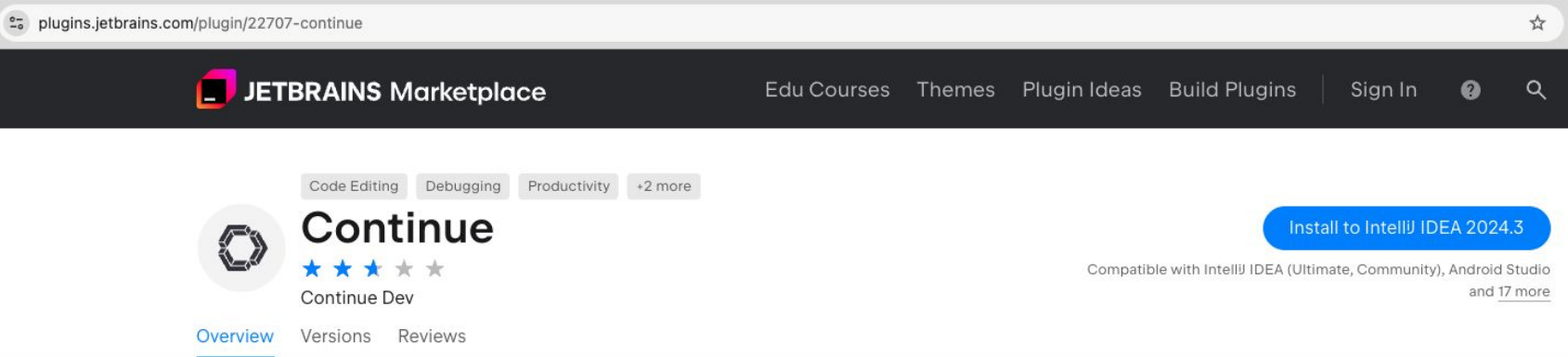
Continue.dev

The screenshot shows the homepage of Continue.dev. At the top, there is a navigation bar with the Continue logo on the left and links for Enterprise, About Us, Docs, and Blog in the center. On the right side of the navigation bar are icons for GitHub, Discord, and a user profile. The main content area features a large heading: "Amplified developers, AI-enhanced development". Below the heading is a subheading: "The leading open-source AI code assistant. You can connect any models and any context to create custom autocomplete and chat experiences inside the IDE". At the bottom of the main content area, there are two prominent buttons: "VS Code" and "JetBrains".

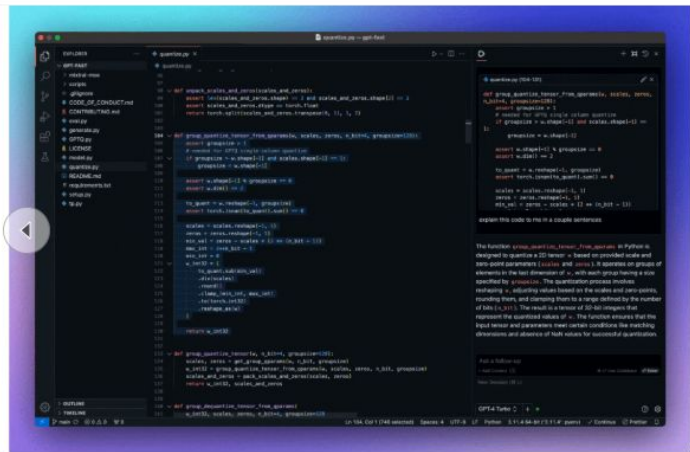
<https://bit.ly/16LLM>

M

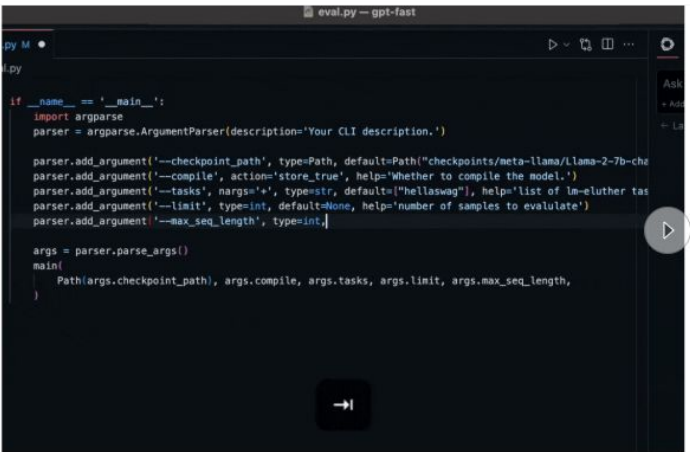
Continue in IntelliJ:



The screenshot shows the JetBrains Marketplace page for the 'Continue' plugin. The page header includes the JetBrains logo and 'Marketplace' text. Navigation links for 'Edu Courses', 'Themes', 'Plugin Ideas', 'Build Plugins', 'Sign In', and search are visible. The plugin card for 'Continue' features a hexagonal logo, the name 'Continue', a star rating of 4.5, and the developer 'Continue Dev'. A blue button says 'Install to IntelliJ IDEA 2024.3'. Below the button, it states 'Compatible with IntelliJ IDEA (Ultimate, Community), Android Studio and 17 more'. Navigation tabs for 'Overview', 'Versions', and 'Reviews' are present.



This screenshot shows the IntelliJ IDEA interface with the Continue AI assistant. The assistant's chat window is open, displaying a conversation about a Python script. The user's prompt is partially visible, and the assistant's response is shown in a light blue box. The background shows a dark-themed code editor with Python code.



This screenshot shows a Python code editor with a script named 'eval.py'. The script uses the 'argparse' module to define command-line arguments for a GPT model. The arguments include '--checkpoint_path', '--compile', '--tasks', '--limit', and '--max_seq_length'. The script also includes a 'main' function that calls 'parser.parse_args()' and a 'Path' object for the checkpoint path.

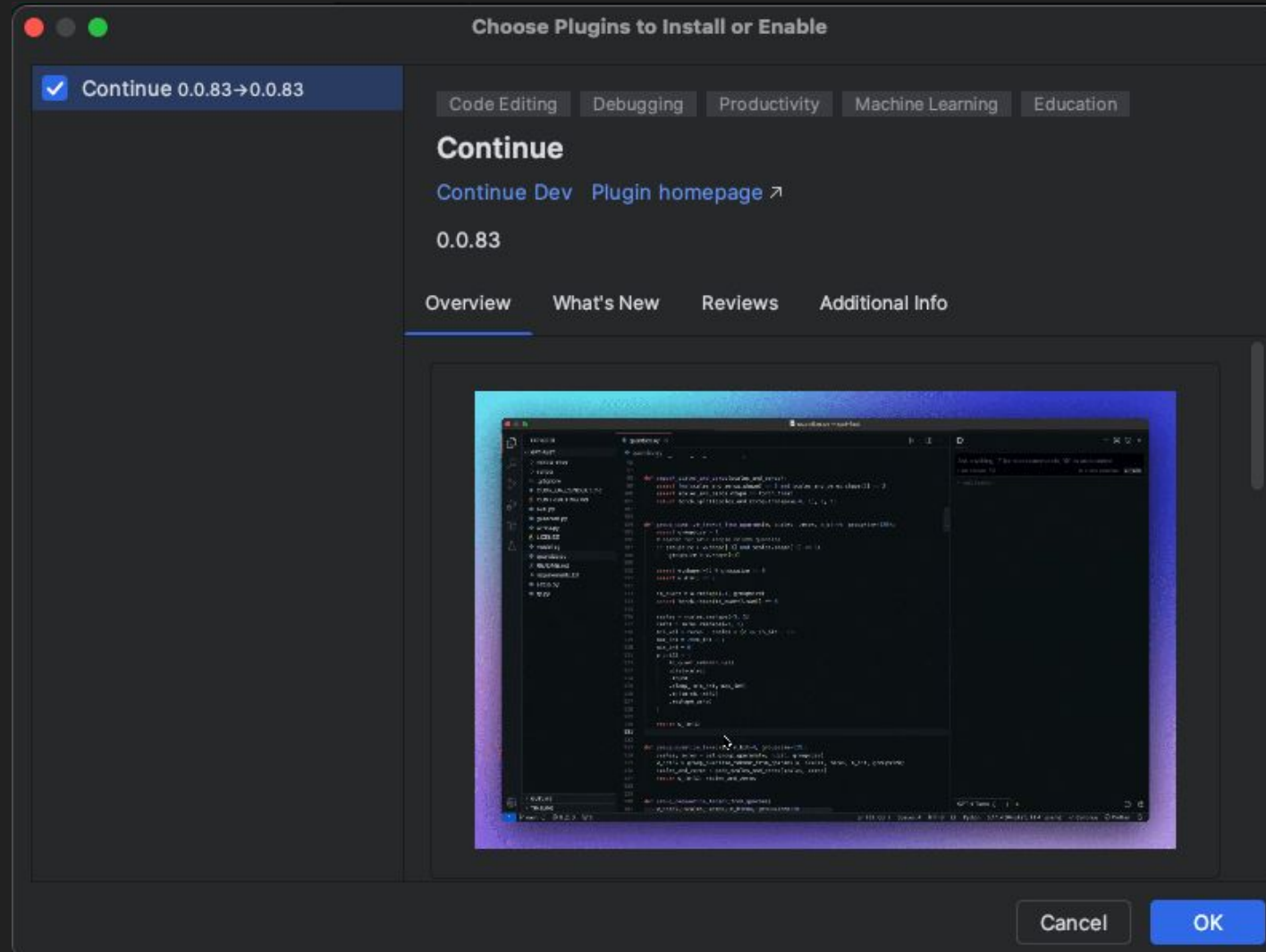
Continue

Continue is the leading open-source AI code assistant. You can connect any models and any context to build custom autocomplete and chat experiences inside **VS Code** and **JetBrains IDEs**.

<https://bit.ly/16LLM>

M

Continue in IntelliJ... Continued.



<https://bit.ly/16LLM>

M

Continue IntelliJ config & offline LLM downloads

The screenshot shows the Continue IDE interface with the 'Local with Ollama' configuration panel open. The panel includes the following elements:

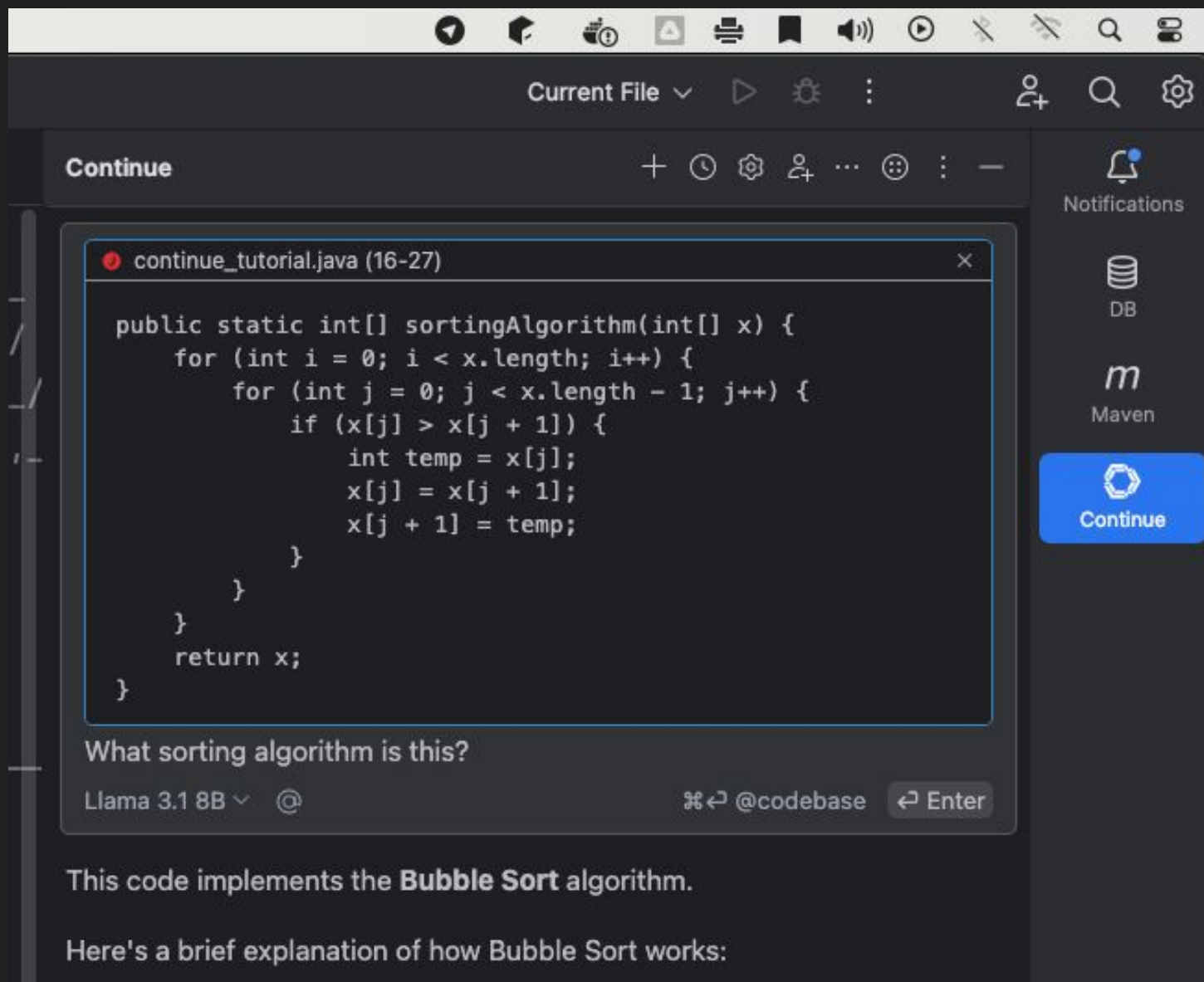
- Install Ollama:** A link to <https://ollama.ai/download> with an external link icon.
- Download Chat model:** The command `ollama pull llama3.1:8b` is shown in a terminal-style box. The size **4.9 GB** is displayed in red text to the right.
- Download Autocomplete model:** The command `ollama pull qwen2.5-coder:1.5b` is shown in a terminal-style box. The size **986 MB** is displayed in red text to the right.
- Connect:** A button at the bottom of the panel with the text 'Connect'.
- Footer:** A note that says 'This will update your [config file](#)'.

The right sidebar of the IDE shows a 'Continue' button in blue, along with icons for Notifications, DB, and Maven.

<https://bit.ly/16LLM>

M

Hi offline! :-)



The screenshot shows a code editor window titled "continue_tutorial.java (16-27)" containing the following Java code:

```
public static int[] sortingAlgorithm(int[] x) {  
    for (int i = 0; i < x.length; i++) {  
        for (int j = 0; j < x.length - 1; j++) {  
            if (x[j] > x[j + 1]) {  
                int temp = x[j];  
                x[j] = x[j + 1];  
                x[j + 1] = temp;  
            }  
        }  
    }  
    return x;  
}
```

Below the code editor, there is an AI assistant interface with the prompt "What sorting algorithm is this?". The assistant's response is:

This code implements the **Bubble Sort** algorithm.

Here's a brief explanation of how Bubble Sort works:

The interface also shows the model "Llama 3.1 8B" and the user "@codebase".

This prompt worked first time OFFLINE. [Code's here on GitHub.](#)

Continue

Add spring boot to this application. I want to serve a RESTful API on port 8080. There should be an h2 database. Entities include student and course. There is a many to many relationship between student and course.

Llama 3.1 8B @codebase Enter

Here's the updated project structure with Spring Boot added:

Project Structure

```
spring-boot-rest-api/  
├── pom.xml (if using Maven)  
├── src/  
│   ├── main/  
│   │   ├── java/  
│   │   │   ├── com/  
│   │   │   │   ├── example/  
│   │   │   │   │   ├── Student.java  
│   │   │   │   │   ├── Course.java  
│   │   │   │   │   ├── StudentRepository.java  
│   │   │   │   │   ├── CourseRepository.java  
│   │   │   │   │   ├── StudentService.java  
│   │   │   │   │   ├── CourseService.java  
│   │   │   │   │   └── RestApiController.java  
│   │   │   └── resources/  
│   │   │       ├── application.properties  
│   │   │       ├── hibernate.cfg.xml  
│   │   │       └── schema.sql  
│   └── test/  
│       ├── java/  
│       │   ├── com/  
│       │   │   └── example/  
│       │   │       ├── StudentTest.java  
│       │   │       ├── CourseTest.java  
│       │   │       └── RestApiControllerTest.java  
└── target/ (compiled classes and other build artifacts)
```

pom.xml (if using Maven)

```
<project xmlns="http://maven.apache.org/POM/4.0.0"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0
```

Ask a follow-up

Llama 3.1 8B @codebase Enter

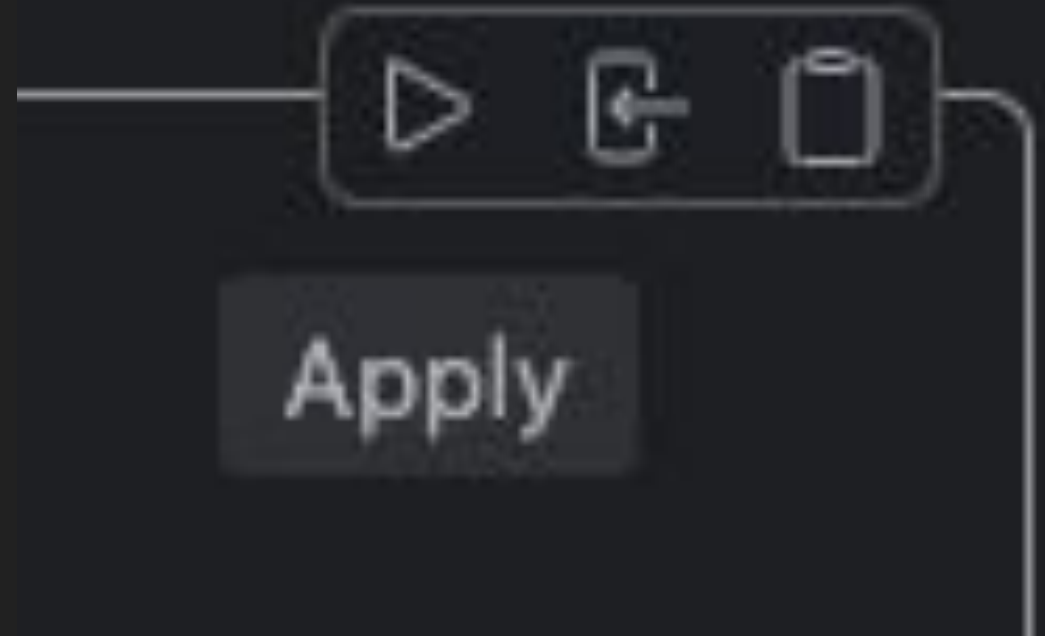
1:1 LF UTF-8 4 spaces NORMAL

<https://bit.ly/16LLM>

M

Nice copy paste widget!


1. Apply (didn't try)
2. Insert at cursor widget
3. Copy the code widget



Prompt: Create a spring boot program that sends reminder emails on a schedule using the FastMail service


1. [Claude.io](#) -- compiles on the first try. [Code is here](#).
2. GitHub CoPilot -- Errors! `mvn install` shows pom.xml errors :-(Code here
3. [ChatGPT -- no pom.xml created. Didn't try to run it.](#)
4. [Free tier of Bard.Google.com did ok.](#)

Copilot now has a free tier --- aka



The video player shows a man with a mustache and blonde hair, wearing a dark hoodie, speaking into a microphone. He is sitting at a desk with a laptop and a mug. The background is dark with several logos: CURSOR (a white cube icon), WINDSURF AI (a green and white wave icon), PearAI (a green pear icon), and JetBrains AI (a white and green logo). The video player interface includes a progress bar at 0:37 / 28:30, a play button, a volume icon, a closed captions icon, a settings icon, a full screen icon, and a share icon.

Copilot is so desperate, they made it free

Theo - t3.gg  373K subscribers

71K views 2 days ago
I thought copilot was going to be useless, until I used it. And now four years later its free?

<https://bit.ly/16LLM>


M

Example w/ Claude.ai Let's Parse... to JSON

← → ↻ 🏠 Not secure coloradodigital.duckdns.org ☆ | 🗄️ 👤

XLX303 Colorado Digital Multiprotocol Reflector Gateway

XLX303 v2.5.3 - Dashboard v2.4.2 [Colorado Digital Main Website](#) / Service uptime: 3 days 09:54:19



democratizing digital by bridging modes

Users / Modules
Repeaters / Nodes (69)
Peers (4)
Modules list
Reflectors list
D-Star live

Callsign		Module					
#	Flag	Callsign	Suffix	DPRS	Via / Peer	Last heard	
1		KA0ZWV	DMR		XLX303 G / BM3102	14.01.2025 20:31	G
2		W0SUN	DMR		XLX303 G / BM3102	14.01.2025 20:12	B
3		W0CDM			W0CDM B	14.01.2025 20:07	K
4		KF0LPT	F70		KF0LPT B	14.01.2025 19:41	D
5		N2BEZ	DMR		XLX303 G / BM3102	14.01.2025 19:06	D
6		KD2QQF	DMR		XLX303 G / BM3102	14.01.2025 18:56	H
7		KB2IDX			KF0LPT B / XLX372	14.01.2025 18:51	I
8		WD0AJG	DMR		XLX303 G / BM3102	14.01.2025 18:25	B
					XLX303	14.01.2025	


CO Severe WX WIRES-X 65045 AllStar 289800 BM 31083 DG-ID 10	Mountain Skynet WIRES-X 00587 BM 31084 DG-ID 11	Olympic Mountain Digital WIRES-X 04429 AllStar 289802 BM 314072 DG-ID 12	Colorado HD WIRES-X 43674 BM 31088 DG-ID 13	NoCO WIRES-X 80851 BM 3171 DG-ID 14	SuperFreq WIRES-X 85958 BM 310869 DG-ID 15	Pi-Star Chat BM 31672 DG-ID 16	Venture Overland WIRES-X 81335 AllStar 54897 BM 31655 DG-ID 17	Interlink TGIF 720 AmComm 31088 PRA XLX720 Interlink DG-ID 18	Special K TGIF 420 AmComm 420 AllStar 58823 DG-ID 20	VHF+ Weak Sign Net D-STAR DG-ID 22
A	B	C	D	E	F	G	H	I	K	M
N0TOA-B	WD0AJG-D	W0CDM-B	KF7CBU-B	W0SUN-B	KOLEE-B	W0SUN-D	W6SPY-B	WD0HDR-B	W0CDM-B	KB0UOA-B
W0CDM-B		WD0HDR-B	W7NEE-B		W4SHB-B		N1RIJ-B	W7FMJ-B	K5NWZ-D	KC0IJH-B
		KD7UXT-C	W6DS-B		KG0REK-B		N7MYW-B			K0FYR-H
			N5FKU-B		N1UKJ-B		KK7DBC-B			
			KOLEE-B		W4SHB-B		KB3HWM-B			
			KD0YD-B							
			KG7FXD-B							
			N9DYO-B							
			KC8I-B							
			N0KMO-B							
			N0AA-B							
			K9DJN-B							
			KF0BDD-B							
			W1VAN-B							

https://bit.ly/16LLM


M

[Claude.io - parse HTML into json.](#) Had off by one & claude fixed it. [5 commits to working code.](#)

yay! Last module and items n-2, n-1 look good with dg-id and node list.

 payne committed 2 weeks ago

Gets the first module correct but has problem with the second.

 payne committed 2 weeks ago

Able to parse xlx303.html but the node column is off by one.

 payne committed 2 weeks ago

copied from <http://coloradodigital.duckdns.org/>

 payne committed 2 weeks ago

start with bootify.io

 payne committed 2 weeks ago

Using AI to go back -- to Java 6!

The screenshot shows an IDE with a project named 'Java17-features-demo'. The main editor displays the file 'ModernJavaDemo.java' with the following code:

```
5 public class ModernJavaDemo {
43     private static void demonstrateLambdaWithCollecti
56
57     // Using lambda with filter and map
58     System.out.println("IT Department employees w
59     employees.stream() Stream<Employee>
60     .filter( Employee e -> e.department()
61     .filter( Employee e -> e.salary() > 80
62     .map(Employee::name) Stream<String>
63     .forEach( String name -> System.out.pr
64
65     // Group employees by department with average
66     Map<String, Double> avgSalaryByDept = employe
67     .collect(Collectors.groupingBy(
68         Employee::department,
69     Collectors.averagingDouble(Em
70     })
```

An AI chat window is open on the right, showing a conversation with 'Llama 3.1 8B'. The chat history includes:

- Context item: 1 context item
- Message: "Here's the modified version of `ModernJavaDemo.java` that uses features and syntax available in Java 6."
- Code block:

```
import java.util.List;
import java.util.Map;

public class ModernJavaDemo {
    public static void main(String[] args) {
        demonstrateSwitchExpressions();
        demonstrateLambdaWithCollections();
    }
}
```
- Message: "Ask a follow-up"

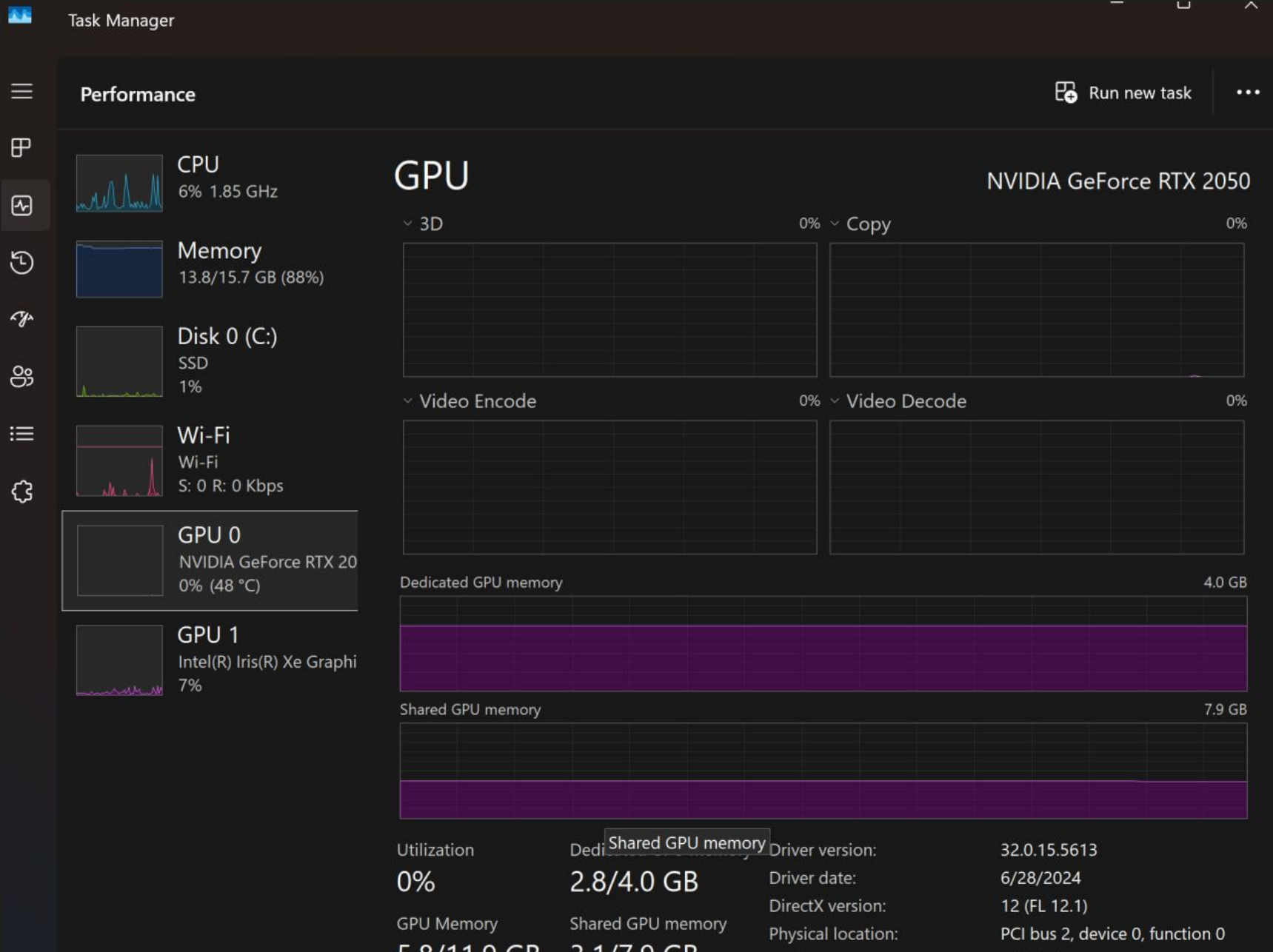
The Run console at the bottom shows the output of the program:

```
Employee count by department:
HR: 2 employees
IT: 3 employees

Process finished with exit code 0
```

The status bar at the bottom indicates the file encoding is UTF-8 and uses 4 spaces for indentation.

Matt Payne's POS Windows 11 laptop performance



Write Better Code

We'll pop out to Ollama for this one just for clarity

Based on work done by Max Woolf - Data Scientist at BuzzFeed

writing java code to find the difference between the smallest and the largest numbers whose digits sum up to 30, given a list of one million random integers between 1 and 100,000.

And we get some code.

Write Better Code

Time to ask for a bit more from the LLM

Can you write more readable code?

Can you write better code?

Can you write faster code?

Can you write an even faster version?

So trusting the first piece of code out of the LLM might not be the best idea.

Why Students Fail to Code with LLMs

There is a paper that discusses this

From the abstract: “(1) students simply lack the technical vocabulary needed to write good prompts, and (2) students do not understand the extent of information that LLMs need to solve code generation tasks.”

Students in this case I think is all of us - need to be able to have a conversation with the LLM, need to know when to question the ai, and how to reiterate. An LLM is not supposed to judge you so don't fail bad asking stupid questions or asking it to explain things.

How to Setup a Private AI with your Code

There are a couple of ways to add your code to an LLM

- Retrieval Augmented Generation - adding your code to an existing data model - adds an additional layer to your ai lookups
- Fine tuning adding your code to an existing model, this needs some power but builds it directly into the LLM
- Create a custom LLM - this is a commitment of time/resources

Does Github Copilot improve Code Quality?

We're still trying to get real data whether using an LLM will improve your code quality.

Github put out a post on this “Does GitHub Copilot improve code quality? Here's what the data says”

JDJarma had a response to it titled “Does GitHub Copilot Improve Code Quality? Here's How We Lie With Statistics”

Does Github Copilot improve Code Quality?

Both posts had good points

Majority of the code was writing API endpoints for a webserver

Not the most challenging of tasks, but one that needs to be done

Github showed a 5% improvement, not sure if that is worth it

Recommend you review them both

There will be a lot more and more in-depth studies in the future of this topic. Still early days

Conclusions & Questions?

LLMs have come a long way in the last few years and are going to be going a long way in the next few years.

Companies see the chance for AI PCs to drive the next hardware refresh cycle.

How many tflops does your CPU have???

Use the tools, but be cautious, hallucinations are real :-)

Links

Continue.dev (Continue plugin) - <https://www.continue.dev/>

VS Codium (Microsoft free VS Code) - <https://vscodium.com/>

VS Code (Microsoft VS Code) - <https://code.visualstudio.com/>

Github Copilot Free (documentation about free github copilot) - <https://docs.github.com/en/copilot/managing-copilot/managing-copilot-as-an-individual-subscriber/about-github-copilot-free>

Register article on Continue - https://www.theregister.com/2024/08/18/self_hosted_github_copilot/

Building a \$300 AI Computer for the GPU-Poor - [https://hackernoon.com/how-to-build-a-\\$300-ai-computer-for-the-gpu-poor](https://hackernoon.com/how-to-build-a-$300-ai-computer-for-the-gpu-poor)

<https://bit.ly/16LLM>

A & M

Links

Write better code - you have to ask -

https://www.theregister.com/2025/01/07/ai_can_write_improved_code_research/

Substance Beats Style: Why Beginning Students Fail to Code with LLMs -

<https://arxiv.org/abs/2410.19792>

Huggingface - the home for AI models - Huggingface.co

Ollama - A great tool to run LLMs - <https://ollama.com/>

Aaron Grothe's - Private AI talk from HDC -

<https://www.grothe.us/presentations/hdc-202408-privateai.pdf>

Links

OpenWebui - nice gui front end to ollama - <https://openwebui.com/>

Github Copilot Code Quality blog post -

<https://github.blog/news-insights/research/does-github-copilot-improve-code-quality-heres-what-the-data-says/>

Github Copilot Code Quality response post -

<https://jadarma.github.io/blog/posts/2024/11/does-github-copilot-improve-code-quality-heres-how-we-lie-with-statistics/>