

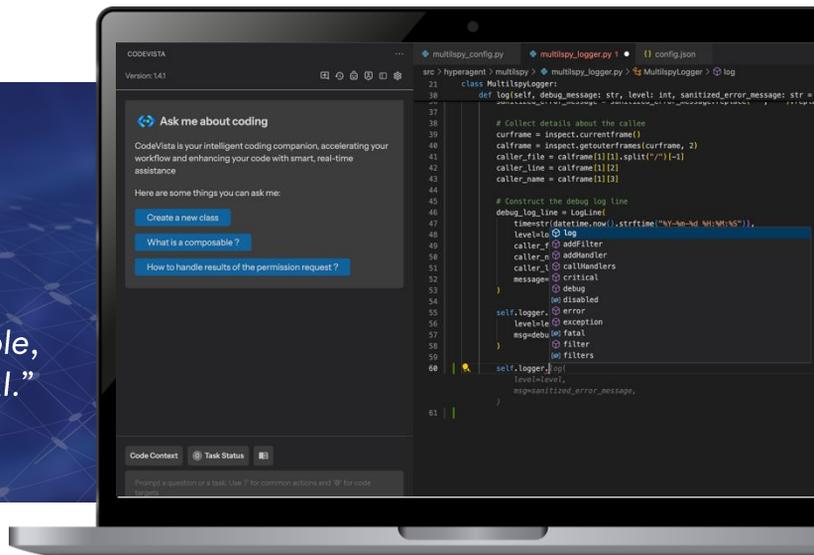


# AGENTIC CODING ASSISTANT

Empower Every Developer with AI Agents



It's not just about coding faster – it's about building smarter, scalable, and maintainable software with AI."



**CodeVista** is an AI-powered coding assistant designed to revolutionize the Software Development Life Cycle (SDLC) with Agentic AI. Seamlessly integrated into your choice of IDE, CodeVista is your intelligent pair-programming navigator, enhancing productivity through context-aware automation and real-time code assistance. With AI agents embedded across every stage of the SDLC, CodeVista automates repetitive tasks, streamlines navigation, and optimizes development efficiency.



60%

Productivity boost



80%

Faster feature development



30%

Reduction in bugs/issues



50%

Less time stuck on problems



10x

more code reuse

## Here's to Effortless Software Development

### Provide Context-Aware Assistance

Transform the entire codebase into a Code Property Graph (CPG), enabling it to understand and map the relationships within repository.

### Turn Chat into Actionable Insights

Enable a smarter and more efficient coding process by analyzing and learning from past tasks.

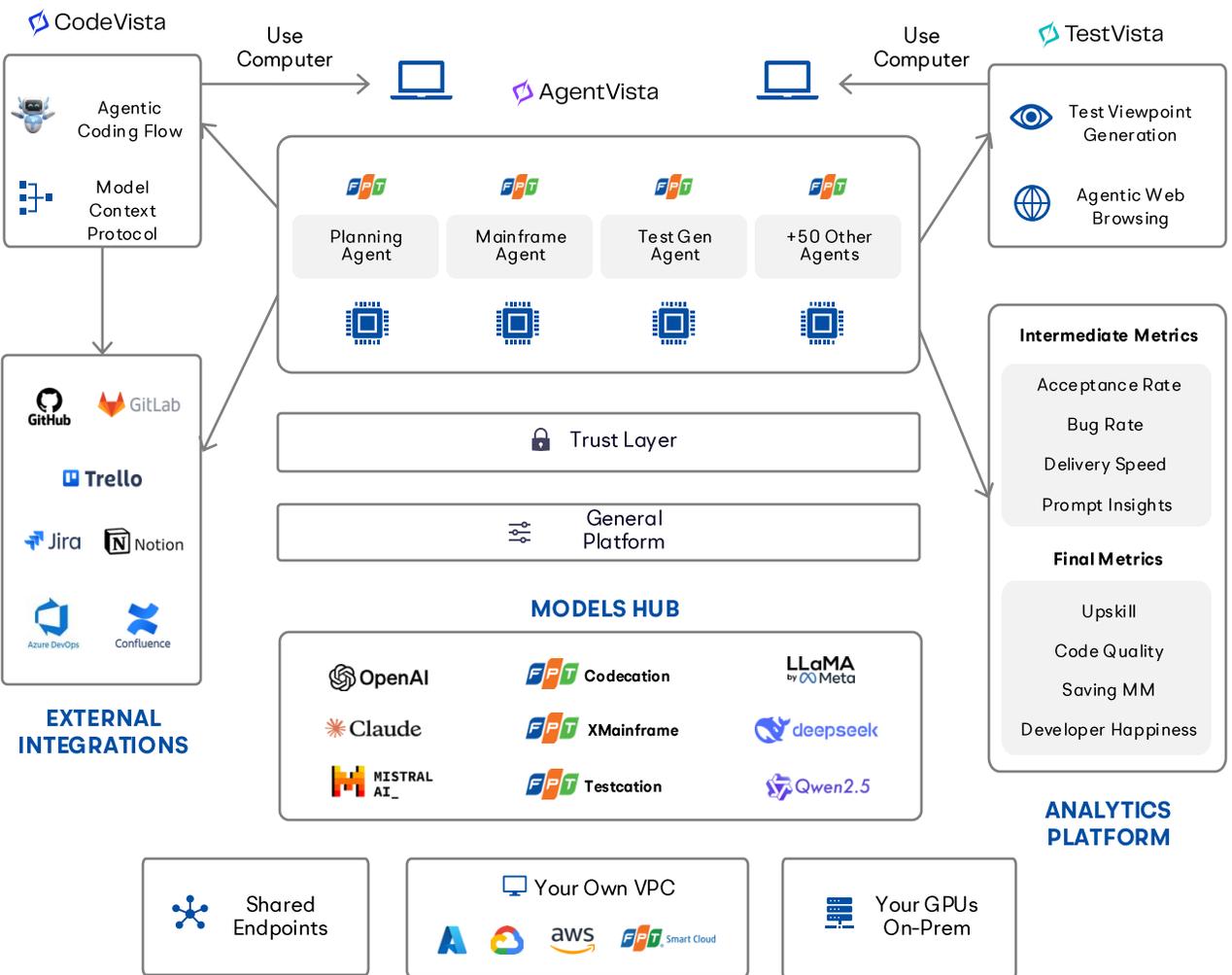
### Automate Tasks Within Workspace

Integrates seamlessly into developer's workspace and the conversation workflow, enabling task automation directly in IDE.

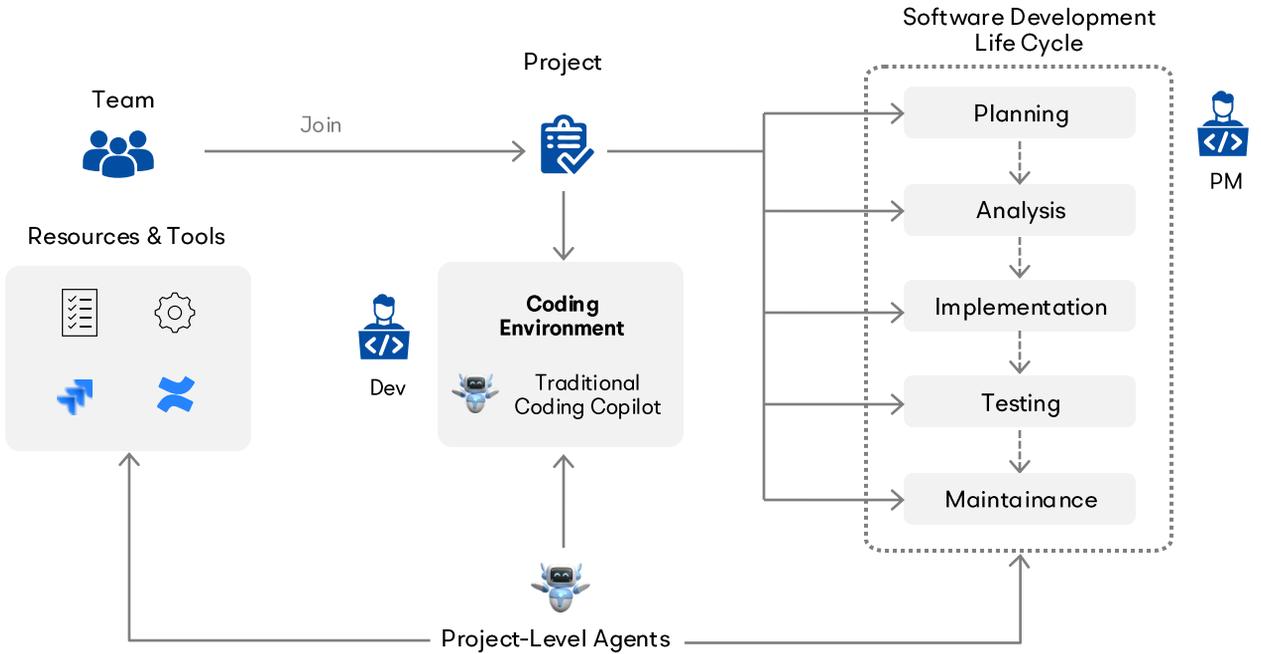
### Allow Customizable Agents

Empowers developers to create, customize, and deploy AI agents tailored to their workflows.

## System Architecture



## Towards a Vision to Build AI as Product/Project Development Team



## Available on Various Integrated Development Environments (IDEs)

- |  |   |   |   |   |   |  |  |
|--|---|---|---|---|---|--|--|
| <br>Visual Studio | <br>Visual Studio Code | <br>Android Studio | <br>PhpStorm | <br>GoLand   | <br>DataGrip     | <br>Aqua  |  |
| <br>CLion          | <br>RustRover          | <br>ReSharper      | <br>WebStorm | <br>RubyMine | <br>IntelliJ IDEA | <br>Rider | <br>PyCharm |



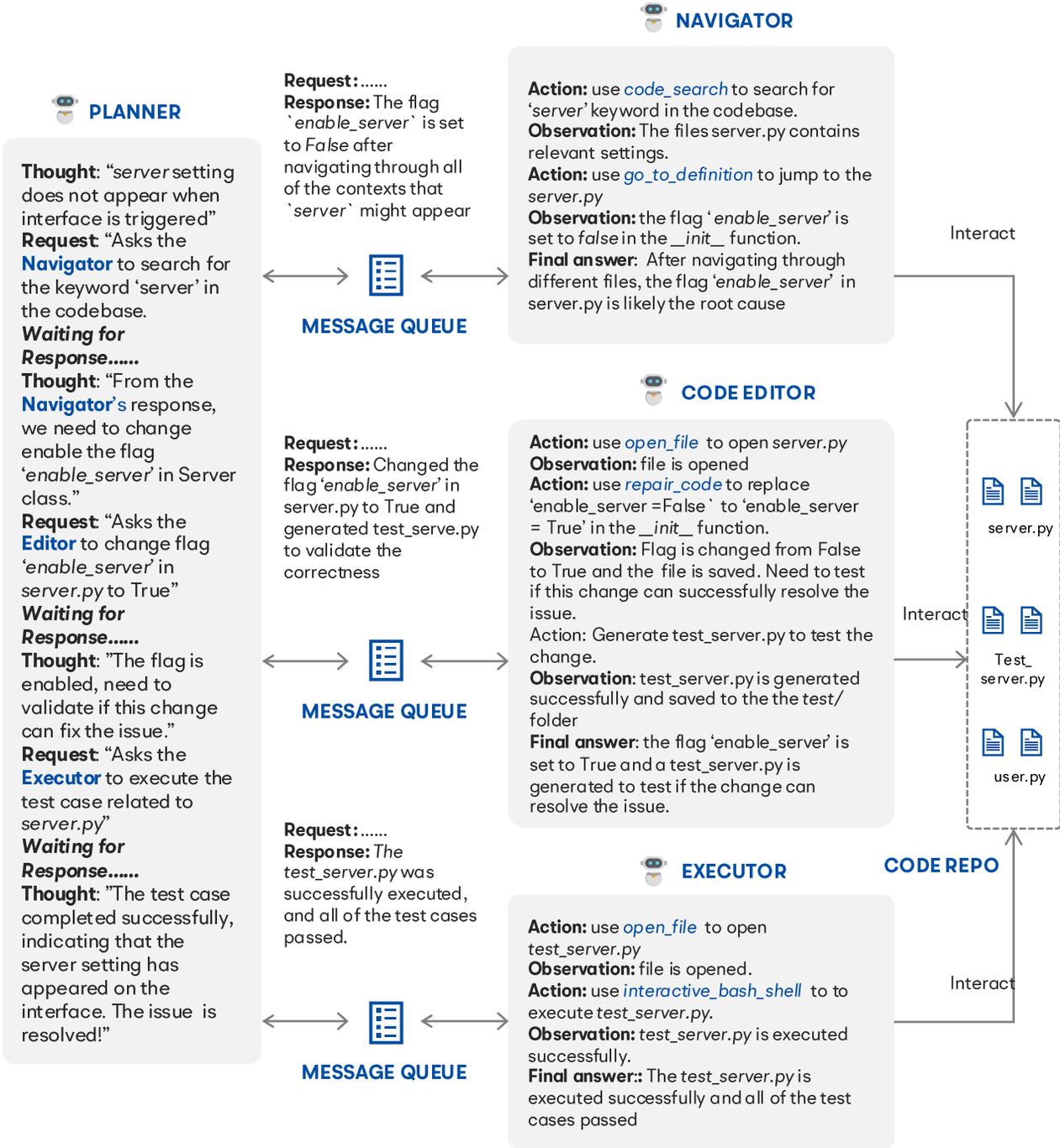
## Our Portfolio of Features

Feature	SaaS (Cloud-Based)	Private Cloud	On-Premises	Hybrid	Embedded SDK/API	Local Standalone (Air-Gapped)
<b>Best For</b>	Individuals/Small/Medium Teams	Enterprises needing isolation	Strict compliance industries	Cloud & on-prem balance	Integrating into platforms	Secure, offline environments
<b>Infrastructure</b>	Managed by FPT on AWS/Azure/GCP	Dedicated instance on customer's cloud	Customer-managed servers	Core cloud, sensitive data on-prem	IDEs, CI/CD pipelines	Individual developer machines
<b>Setup Time</b>	Instant	2-3 days	1-2 weeks	1-2 weeks	Instant	Instant
<b>Architecture</b>	Multi-tenant	Single-tenant	Single-tenant	Hybrid	Modular	Standalone desktop application
<b>Data Isolation</b>	Shared	Dedicated VPC	Full control	Local for sensitive data	Handled by customer	Fully local, no external access
<b>Security Compliance</b>	SOC 2, ISO 27001, GDPR	Customizable policies	Customer-defined	End-to-end encryption	Encrypted API calls	Complete offline environment
<b>Authentication</b>	OAuth2, SSO, IAM	Enterprise IAM (Okta, Azure AD)	LDAP, SSO	Hybrid IAM support	API keys, OAuth	Local machine authentication
<b>Backup &amp; Recovery</b>	Managed, daily backups	Customer-defined	Local backups	Split based on environment	Customer-managed	Local backups only
<b>Monitoring &amp; Logging</b>	Standard dashboards	Customizable	On-prem tools	Cloud + local insights	User's existing platforms	Local logs, no external monitoring
<b>Performance Scaling</b>	Automatic	Adjustable by user	Limited by on-prem setup	Hybrid load balancing	Based on platform setup	Limited by machine resources
<b>Maintenance</b>	Managed by CodeVista	Customer-managed	Customer-managed	Split responsibility	Not applicable	User-managed software updates
<b>Cost Model</b>	Per-user subscription	Annual subscription + infra	One-time license + maintenance	Hybrid cost structure	Pay-per-use/API calls	Per-device license
<b>Version Control</b>	Always latest	User-controlled	User-controlled	Cloud latest, local user	User-controlled	Manual software updates
<b>LLM Support</b>	Cloud LLM (Azure OpenAI, Claude or our self-hosted LLM on cloud)	Users Controlled	Users Controlled	Users Controlled	Users Controlled	Local LLM on personal machine

# HyperAgent

## Generalist Software Agents to Solve Software Engineering Tasks

- HyperAgent is an innovative multi-agent framework designed to solve software engineering tasks at scale.
- HyperAgent replicates how human developers approach and solve coding tasks in their daily routines.
- HyperAgent achieves state-of-the-art results on well-known benchmarks such as SWE-Bench and Defects4J.



Method	Verified (%)	Lite (%)	AvgTime	AvgCost (\$)
AutoCodeRover + GPT-40	28.80	22.7	720	0.68
SWE-Agent + Claude 3.5 Sonnet	33.60	23.00	-	1.79
SWE-Agent + GPT-40	23.20	18.33	-	2.55
Agentless+GPT-40	33.20	24.30	-	0.34
<b>HYPERAGENT-Lite-1</b>	<b>30.20</b>	<b>25.33</b>	<b>106</b>	<b>0.45</b>
HYPERAGENT-Lite-2	16.00	11.00	108	0.76
HYPERAGENT-Full-1	<b>33.00</b>	<b>26.00</b>	320	1.82
HYPERAGENT-Full-2	31.40	25.00	210	2.01
HYPERAGENT-Full-3	18.33	12.00	245	0.89

Dataset	Tool	Tot Bugs	Correct Fixes	Correct %
Defects4J v1.2	HYPERAGENT	395	<b>82</b>	<b>20.8%</b>
	RepairAgent		74	18.7%
	ITER		57	14.4%
	SelfAPR		64	16.2%
Defects4J v2	HYPERAGENT	440	<b>110</b>	<b>25.0%</b>
	RepairAgent		90	20.5%
	SelfAPR		46	10.5%

HyperAgent - Best Model for Github Issue Resolving on SWE-Bench

HyperAgent - Best Model for Bug Fixing on Java Defects4J x

Method	Acc@1	Cost (\$)
Ochiai (Zou et al., 2019)	20.25%	-
DeepFL (Li et al., 2019)	33.90%	-
Dstar (Wong et al., 2012)	33.90%	-
Grace (Zou et al., 2019)	49.36%	-
AutoFL (Kang et al., 2024)	51.00%	-
<b>HYPERAGENT-Lite-1</b>	<b>59.70%</b>	<b>0.18</b>

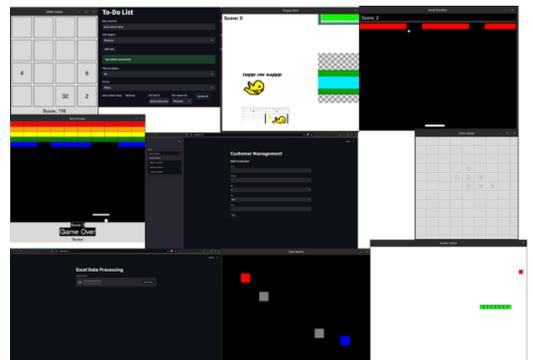
HyperAgent - Best Model for Bug Localization

## AgileCoder

### Multi-Agents Collaboration for App Creation

- AgileCoder is a multi-agent framework designed for automated app creation.
- AgileCoder mimics how software development teams operate under the Agile methodology, incorporating roles like PMs, Developers, Testers, and Scrum Masters.
- AgileCoder can generate realistic, fully executable applications from a single prompt.

Category		Model	Dataset Performance	
			HumanEval	MBPP
LLMs (prompting)		CodeGeeX-13B	18.9	26.9
		PaLM Coder-540B	43.9	32.3
		DeepSeeker-33B-Inst	79.3	70.0
		GPT-3.5 Turbo	60.3	52.2
		Claude 3 Haiku	75.9	80.4
		GPT 4	80.1	80.1
LLM-based Agents	With GPT-3.5 Turbo	ChatDev	61.79	74.80
		Meta GPT	62.80	74.73
		<b>AgileCoder</b>	<b>70.53</b>	<b>80.92</b>
	With Claude 3 Haiku	ChatDev	76.83	70.96
		<b>AgileCoder</b>	<b>79.27</b>	<b>84.31</b>
	With GPT 4	MetaGPT	85.9	87.7
		<b>AgileCoder</b>	<b>90.85</b>	-



Can generate a wide range of software with high accuracy

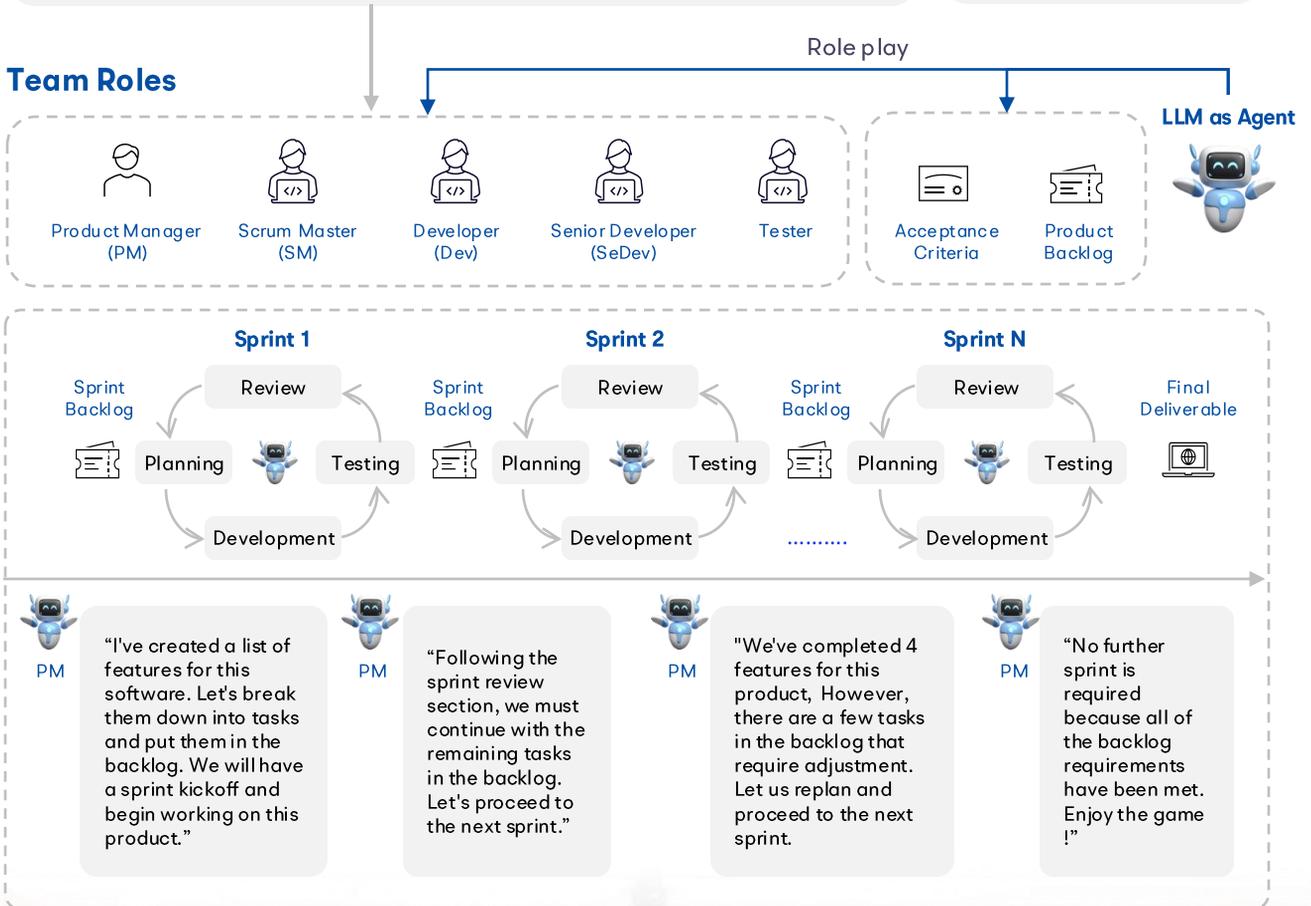
Best Performance on Code Generation Benchmarks

## Software Specs

Create a Snake game where a player navigates a snake across a grid to consume food, growing in size while avoiding self-collision and wall hits. Implement basic elements like a moving snake, randomly placed food, and game-over logic. Ensure smooth controls for direction changes and visually distinguish between the snake, food, and game area. Optionally, introduce features to enhance gameplay, such as difficulty levels and score tracking.

You are acting as the Product Manager, You are familiar with product design and software development life cycle ....

### Team Roles



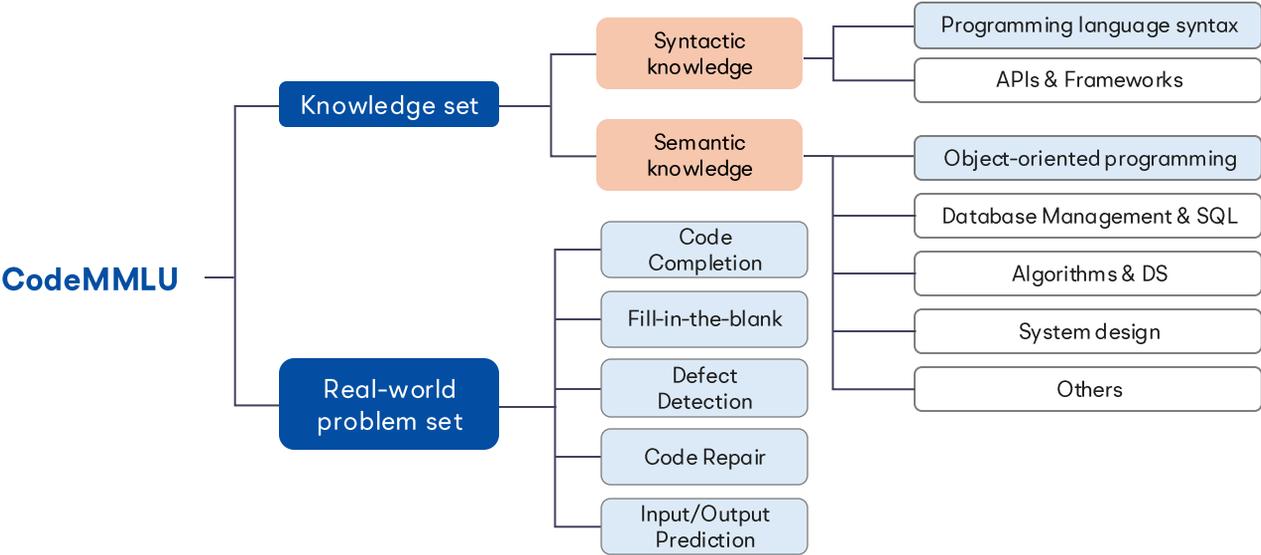
AgileCoder: Multi-Agents Collaboration Framework for App Creation based on Agile Methodology



# CODEMMLU

## Benchmark for Assessing Code Understanding & Reasoning Capabilities of CodeLLMs

- CodeMMLU is a novel benchmark designed to evaluate the code reasoning and understanding capabilities of LLMs.
- CodeMMLU presents a significant challenge, even for advanced reasoning models like DeepSeek R1 and Claude Sonnet 3.7, making it a rigorous test of code intelligence.
- Key Insight: CodeMMLU reveals a unique correlation—models with strong Computer Science knowledge perform better on Software Engineering tasks, mirroring human expertise. This insight is not captured by existing benchmarks.
- CodeMMLU is set to become the standard benchmark for evaluating LLMs in the software domain, complementing MMLU for general knowledge assessment.

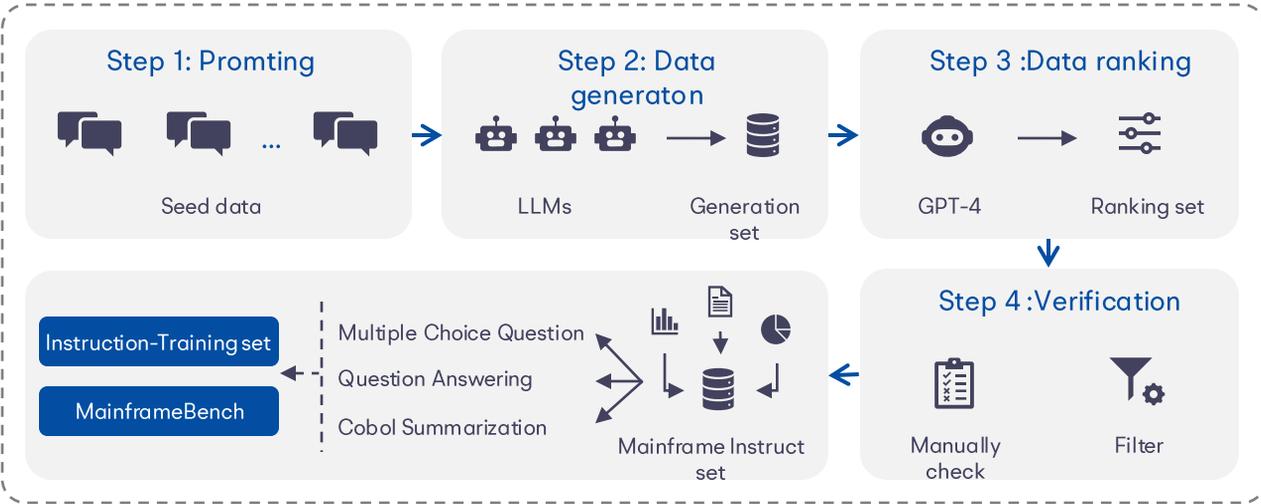


Family	Model name	Size (B)	Knowledge test			CodeMMLU
			Syntactic	Semantic	Fundamental test	
<b>Closed-source models</b>						
Claude	Claude3.7 Sonnet	-	52.78	<b>76.26</b>	<u>60.92</u>	<b>61.65</b>
	Claude3.5 Sonnet	-	52.23	73.45	58.56	59.81
	Claude3.5 Haiku	-	49.24	68.20	57.83	57.25
	Claude3 Sonnet	-	<b>67.22</b>	66.08	38.26	53.97
GPT	GPT 03-mini	-	53.08	75.50	<b>62.77</b>	<b>62.36</b>
	GPT 4o	-	50.63	69.61	53.89	56.40
	GPT 4o-mini	-	48.66	55.90	20.33	38.43
	GPT-3.5-turbo	-	<u>61.68</u>	53.65	45.26	51.70
<b>Open-source models</b>						
Llama	Llama3.3 70B Inst	70	44.31	52.76	30.96	40.66
	Llama3.1 405B Inst	405	50.82	<b>71.41</b>	<b>57.10</b>	<b>58.23</b>
	Llama3 70B Inst	70	46.94	62.64	53.15	53.19
	CodeLlama34B Inst	34	56.81	46.93	23.55	38.73
DeepSeek	DeepSeek R1	671	42.39	56.77	38.08	43.91
	DeepSeek V3	685	48.30	57.53	45.06	49.08
	DeepSeekCoder 33B Inst	33	53.65	45.43	21.46	36.60
	DeepSeekMoE 16B Chat	16.4	31.74	35.42	27.32	31.01
Mistral	Mistral7B Inst (v0.3)	7	54.42	51.25	31.85	43.33
	Mixtral 8x7B Inst	46.7	<u>61.17</u>	54.89	24.09	42.96
	Codestral 22B	22	60.34	52.10	37.85	47.60
Phi	Phi4	14	45.34	57.46	47.82	49.19
	Phi4 Mini Inst	12	41.94	51.59	19.75	34.85
Qwen	Qwen2.5 14B Inst	14	46.38	58.70	51.49	51.38
	QwQ 32B Preview	57	<b>61.34</b>	57.48	30.48	46.34
	QwenCoder2.5 32B Inst	32	50.63	<u>62.61</u>	<b>53.82</b>	<b>56.40</b>

# XMainframe

## The First-in-the-World Foundation Models for Mainframe Modernization

- XMainframe delivers state-of-the-art performance on tasks like COBOL code summarization and mainframe-related question answering.
- XMainframe outperforms larger models, including 33B models and even GPT-4, in mainframe knowledge tasks -all with an efficient 7B model size.
- We introduce MainframeBench, a standardized benchmark designed to assess an LLM’s ability to understand mainframe systems.



Model	Accuracy (%)
GPT-4 73.90	73.90
GPT-3.5	74.56
Mixtral-Instruct 8x7B	68.12
Mistral-Instruct 7B	69.29
Neural-Chat	66.35
DeepSeek-Coder-Instruct 6.7B	47.49
DeepSeek-Coder-Instruct 33B	53.29
XMainframe-Instruct 7B	68.57
<b>XMainframe-Instruct 10.5B</b>	<b>77.89</b>

Model	BERTScore	RougeL	Meteor	BLEU-4
GPT-4 73.90	0.85	0.22	0.34	7.42
GPT-3.5	0.88	0.28	0.34	11.37
Mistral-Instruct 7B	0.85	0.12	0.15	3.61
Neural-Chat	0.88	0.27	0.34	11.07
DeepSeek-Coder-Instruct 6.7B	0.85	0.22	0.32	7.72
DeepSeek-Coder-Instruct 33B	0.85	0.21	0.31	7.55
XMainframe-Instruct 7B	0.89	0.41	0.56	22.23
<b>XMainframe-Instruct 10.5B</b>	<b>0.96</b>	<b>0.74</b>	<b>0.74</b>	<b>62.58</b>

Model	MAP	F1-Score	BERTScore	RougeL	Meteor	BLEU-4
GPT-4 73.90	0.12	0.19	0.88	0.18	0.34	5.71
GPT-3.5	0.14	0.22	0.89	0.21	0.38	7.36
Mixtral-Instruct 8x7B	0.27	0.31	0.9	0.29	0.38	11.36
Mistral-Instruct 7B	0.12	0.19	0.87	0.18	0.34	5.74
Neural-Chat	0.13	0.21	0.88	0.2	0.36	6.45
DeepSeek-Coder-Instruct 6.7B	0.09	0.15	0.86	0.14	0.30	4.09
DeepSeek-Coder-Instruct 33B	0.09	0.15	0.86	0.15	0.31	4.41
XMainframe-Instruct 7B	0.45	0.42	0.92	0.4	0.42	20.43
<b>XMainframe-Instruct 10.5B</b>	<b>0.43</b>	<b>0.42</b>	<b>0.92</b>	<b>0.4</b>	<b>0.42</b>	<b>20.93</b>

## Case Study: Bringing the Power of Generative AI to FPT Developers

### BUSINESS DESCRIPTION

AI is revolutionizing software development, with 82% of developers globally using AI tools for coding. AI-driven coding tools attracted \$1 billion in 2024, and the global AI market, valued at \$391 billion in 2025, is projected to grow to \$1.81 trillion by 2030. These statistics highlight AI's dominance in reshaping the software engineering landscape. As Vietnam's leading IT company, FPT has strategically incorporated AI into its software development processes, achieving notable advancements.

### BUSINESS CONTEXT

- **Objective:** Assist developers with coding tasks and enhance their productivity and work experience.
- **Challenge:** Develop an in-house generative AI tool that meets FPT's security and functionality requirements, while ensuring effective oversight and control of team usage, and demonstrating a strong return on investment.

### TECHNOLOGY HIGHLIGHTS

#### Automate Tasks Within Workspace

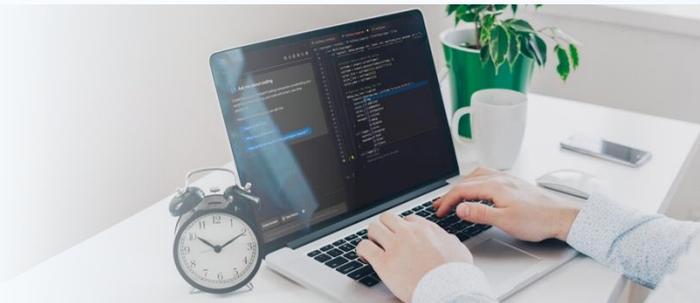
CodeVista introduces the Repo Navigator, which transforms codebases into semantic graphs. This technology enables precise understanding of code relationships, dependencies, and structures, forming the foundation for intelligent navigation, repository-level Q&A, and mainframe modernization.

#### Agent-Based Orchestration

CodeVista incorporates agentic workflows, allowing users to leverage a marketplace of specialized AI agents. These agents perform tasks such as code search, editing, execution, and optimization, offering a scalable and autonomous development experience.

#### Scalable AI Model Integration

CodeVista supports the integration of large language models (LLMs) optimized for code-related tasks. This includes task-specific tuning for repository-level completion, semantic search, and fault localization, achieving superior accuracy and developer assistance.



#### Contextual Awareness Across Modalities

With its External Context Integration feature, CodeVista enhances productivity by incorporating relevant information from diverse sources, including external files, documents, and online resources. This ensures comprehensive contextual understanding in solving software engineering challenges.

#### Advanced Analytics and Insights

The platform includes a robust analytics engine that tracks usage patterns, identifies bottlenecks in workflows, and generates actionable insights. These insights help optimize the software development lifecycle and improve decision-making for development teams.

### OUTCOMES



The official internal tool at FPT Software



Gained pilot commitment from **1,500+** projects



**48%** efficiency increase in code generation tasks



Saved **830** man-months of effort



Processed **1.3** million prompts



Over **3.5** million lines of code was accepted.

### Contact us

 FPT Tower, 10 Pham Van Bach Str., Cau Giay, Ha Noi, Viet Nam

 Support.aic@fpt.com

 <http://www.fpt-aicenter.com/>

 <http://codevista.ai>