



CONTENT

INTRODUCTION

USER

DATASET

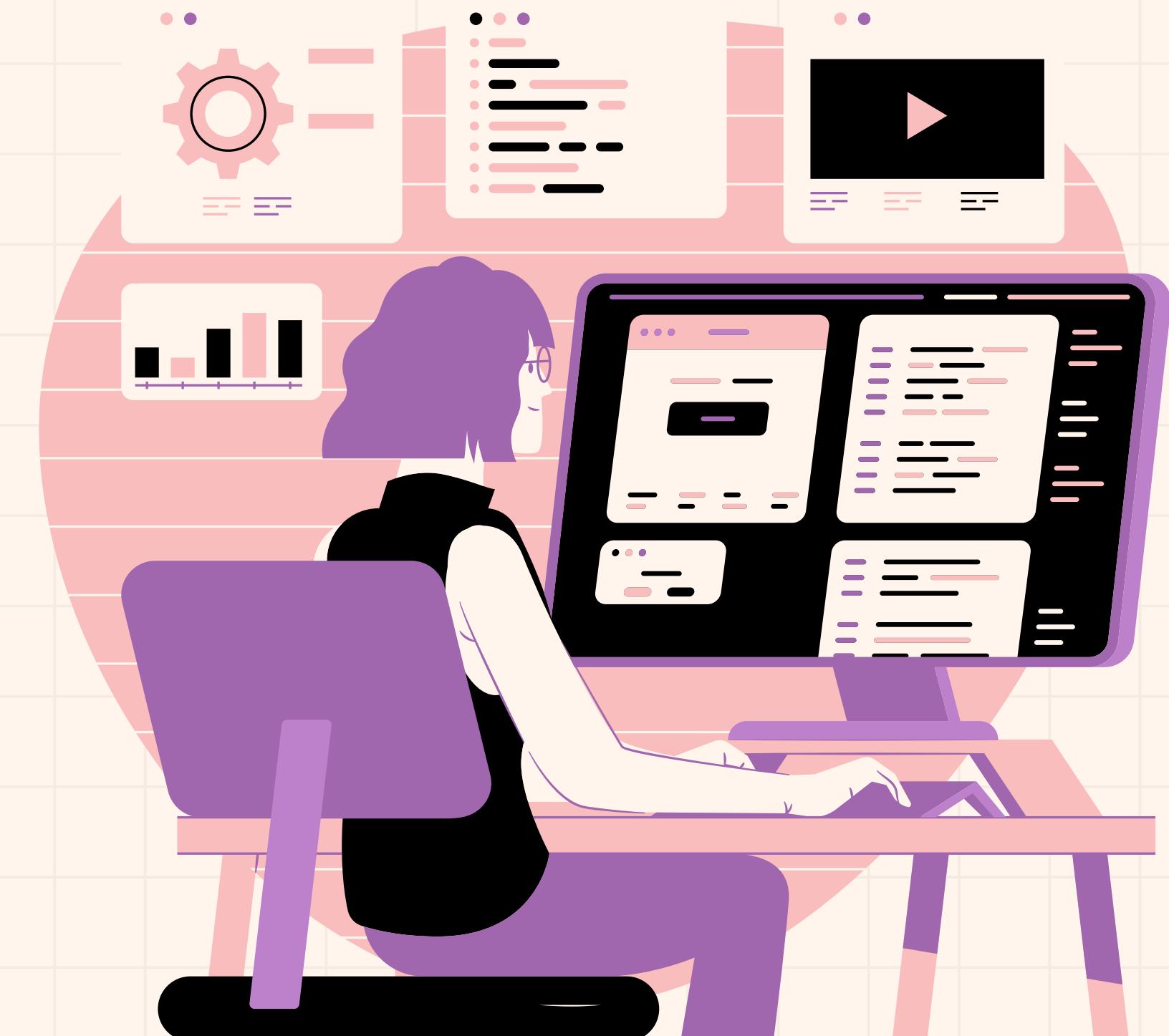
ARCHITECTURE

CONCLUSION

“CODE SMARTER, NOT HARDER”
POWERED BY DEEPSEEK-CODER

CODEGENIE

Presented By
Kanishka Gupta
(245523733027)
Group 413
Mentor: **Mrs. Sripooja**



ABOUT

In the rapidly evolving world of software development, Large Language Models (LLMs) are revolutionizing the way we write, understand, and maintain code. However, most high-performing models are closed-source and limited in accessibility.

Our project, CodeGenie, leverages DeepSeek-Coder, a powerful open-source LLM for code generation, to provide real-time, intelligent coding assistance within VS Code.

VISION

- ◆ **Real-Time Code Generation**

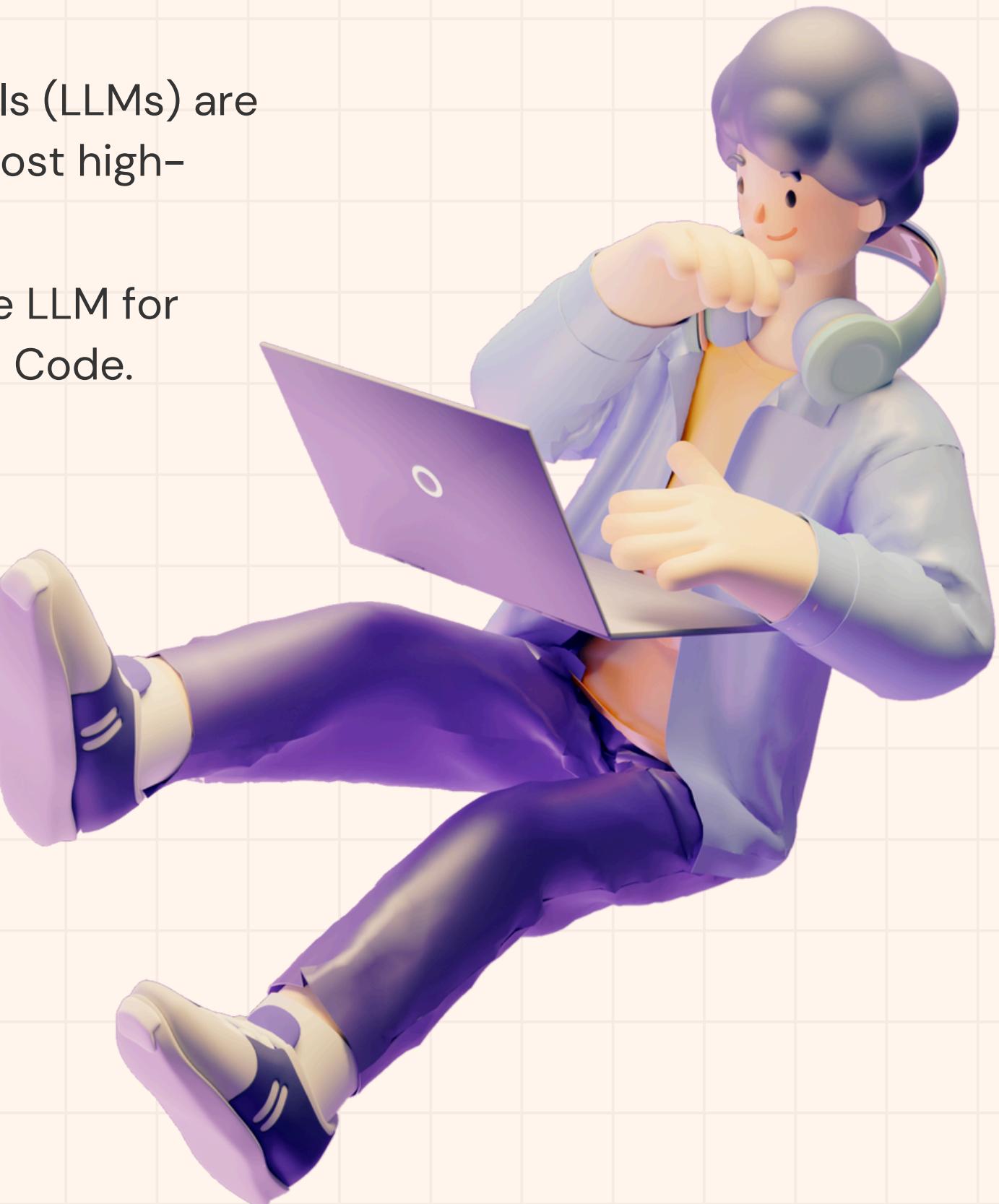
CodeGenie will transform natural language prompts into high-quality, functional code snippets instantly inside Visual Studio Code.

- ◆ **Intelligent Multilingual Support**

Built with DeepSeek-Coder, CodeGenie would support 87+ programming languages, providing developers flexibility across diverse tech stacks.

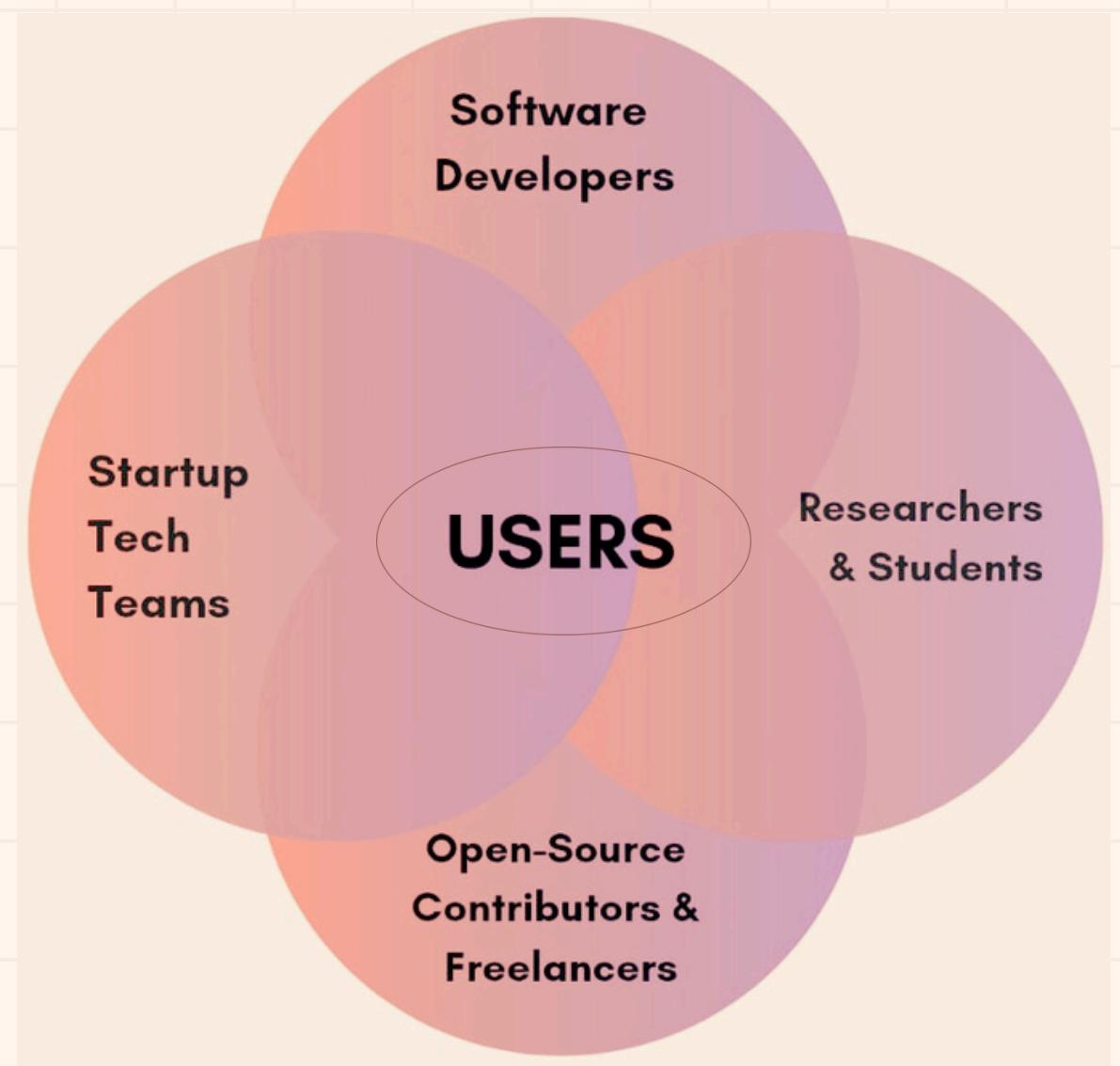
- ◆ **Long Context Understanding**

With a 16K token window, CodeGenie will be able to understand large files and multi-file projects, enabling accurate, context-aware code suggestions.



REAL-WORLD HURDLES CODING MADE EFFORTLESS

- 🔒 Restricted Access to Proprietary Tools
- ⚙️ Limited Capability in Open-Source Alternatives
- 🕒 Inefficient Development Workflows
- 💡 Barriers to AI Integration

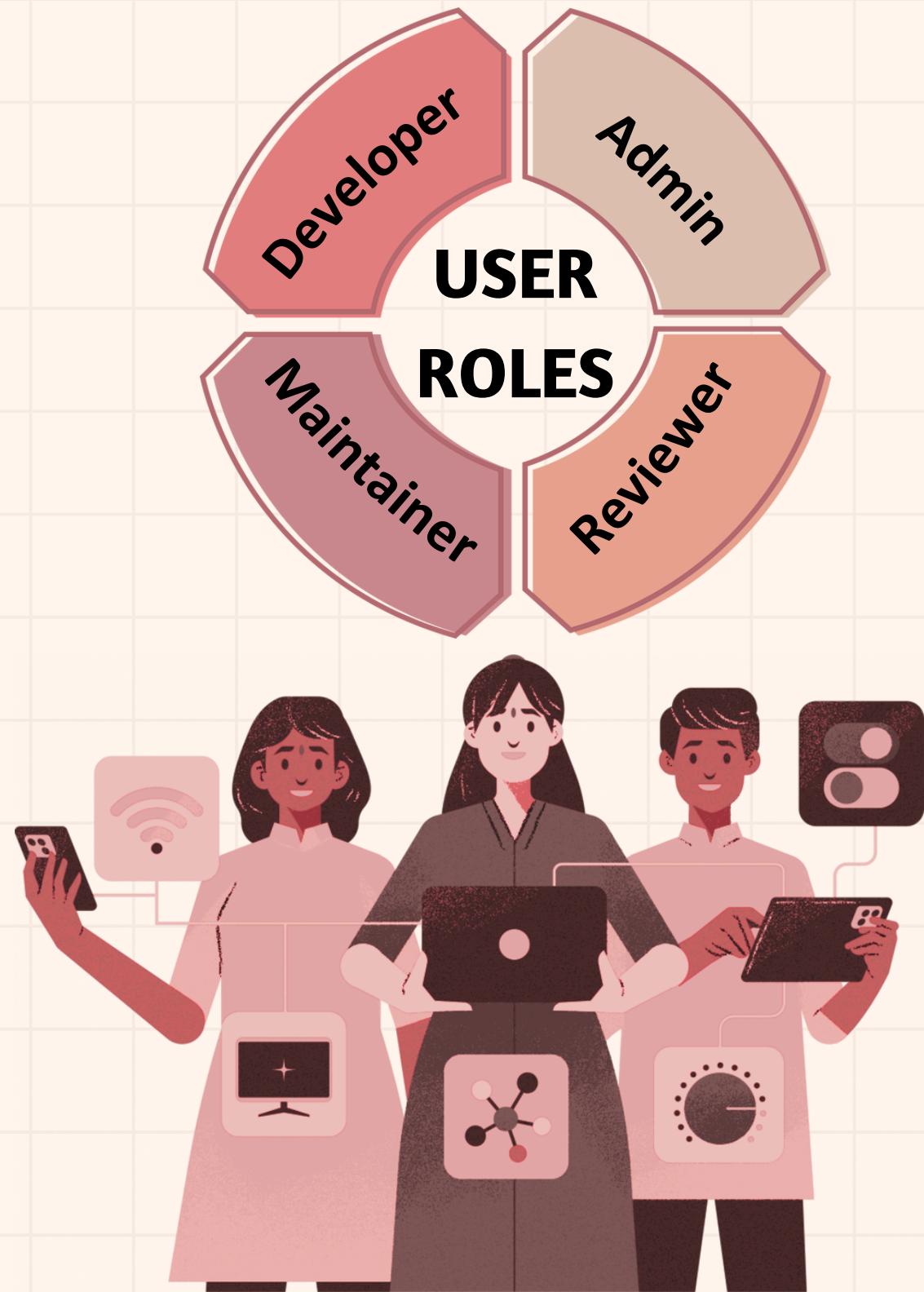


- Reduces time spent on boilerplate coding tasks.
- Helps developers by auto-completing code and offering intelligent snippets.
- Bridges knowledge gaps by offering code suggestions in unfamiliar languages.
- Allows focus on high-level problem-solving rather than low-level syntax.





USER ACCESS AND INTERACTION



- Natural language prompts or technical queries entered directly in the VS Code extension interface.
- Examples: “Create a responsive login page in React,” or “Suggest an optimized sorting algorithm in Python.”

Input

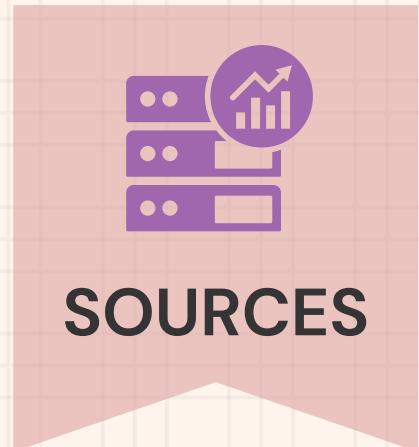
- AI-generated code snippets, completions, or suggestions are seamlessly displayed in the user’s active coding window.
- Users can review, modify, or directly integrate the code into their project.

Output

THE RISE OF CODE INTELLIGENCE

**DATASET**

- 2 trillion tokens of data.
- 87% source code across 87+ programming languages.
- 13% technical documents, API references, and coding Q&A

**SOURCES**

- Public GitHub repositories (open-source licensed).
- Open-source datasets like The Stack and CodeSearchNet.
- Organized at the project/repository level for better context learning

DATA PREPROCESSING



- › Data is first crawled from GitHub repositories using automated tools.
- › Rule-based filters are applied to remove low-quality or irrelevant content.
- › Parsed code is analyzed for dependencies to maintain logical structure.
- › Redundant or duplicate repositories are eliminated at the repo level.
- › The remaining data is screened for quality before being added to the final dataset.

CODEGENIE SYSTEM ARCHITECTURE

TECHNOLOGY STACK

Frontend:

- Language: TypeScript 
- Framework: VS Code Extension API
- UI Components: HTML, CSS, React



Backend:

- Model Hosting: Flask API
- Framework: VS Code Extension API
- API Integration: RESTful API



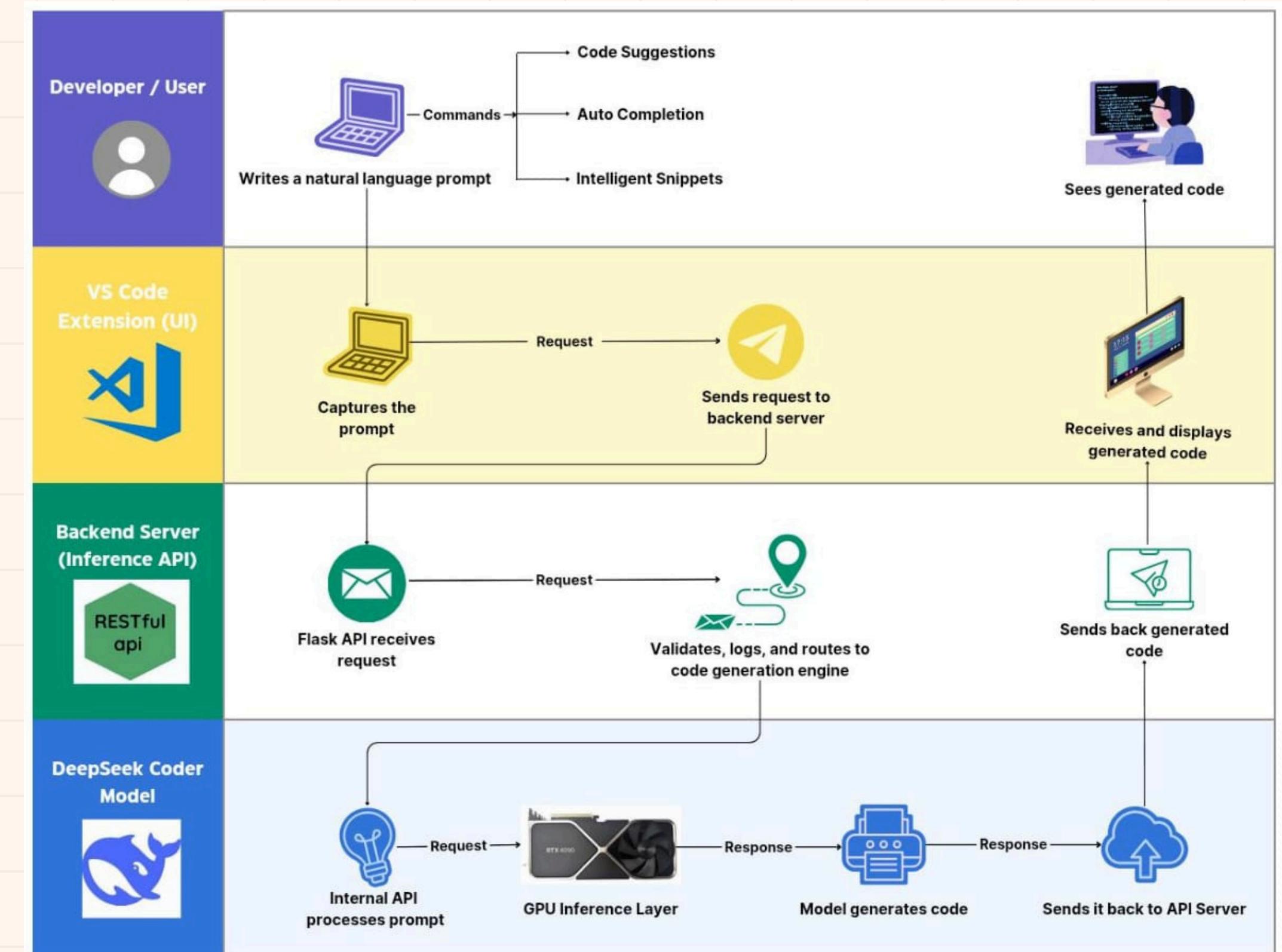
AI:

- Model: DeepSeek-Coder (LLM)
- Library: Hugging Face Transformers
(For loading and utilizing the model)
- GPU Acceleration: CUDA
(To leverage RTX 4090 GPU)



Deployment:

- Deploy: VS Code 
- Version Control: Git & GitHub

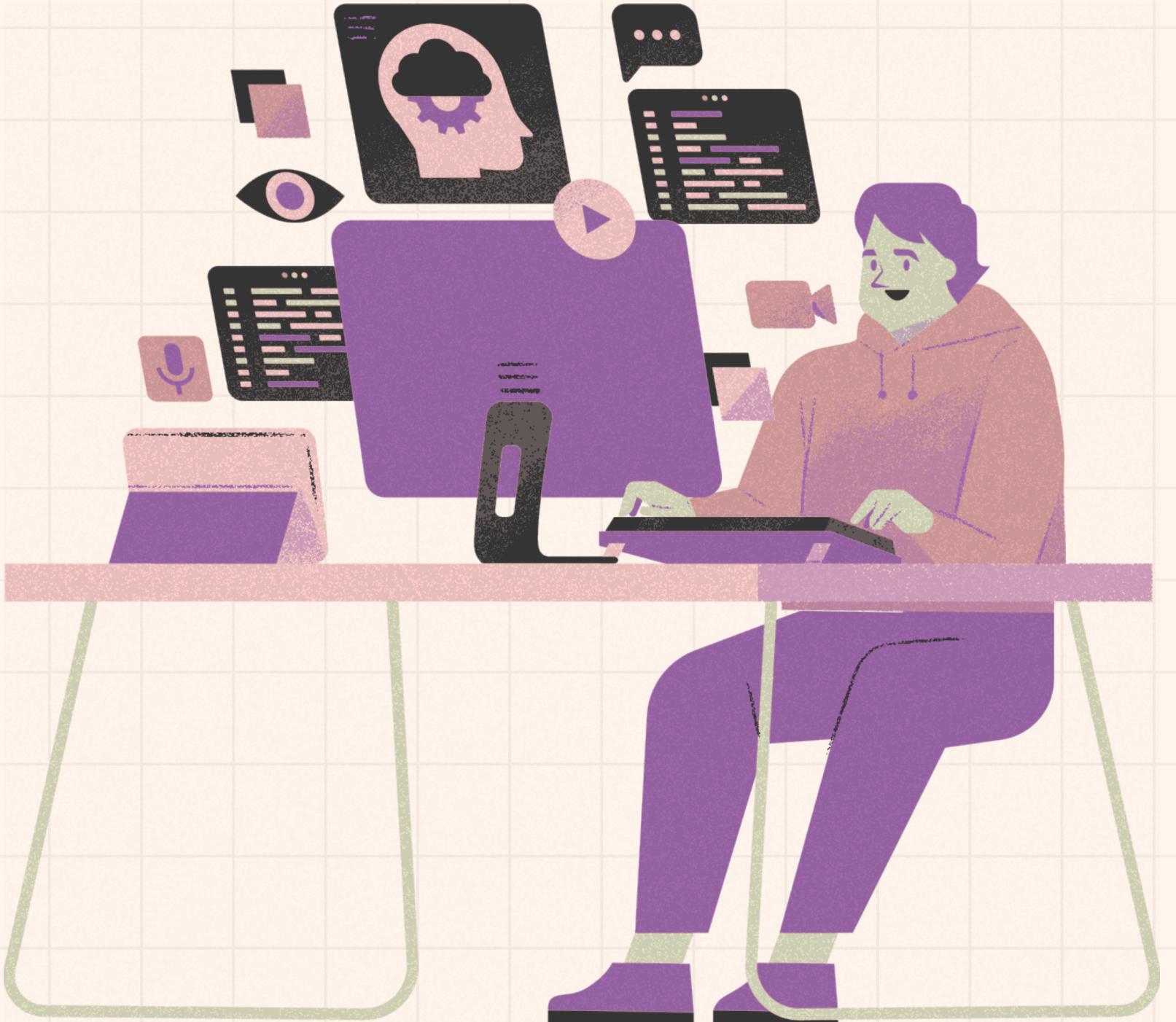


AI-POWERED CODE CREATION AT YOUR FINGERTIPS



DeepSeek-Coder marks a pivotal advancement in the world of open-source code generation. With its robust architecture, comprehensive multilingual training, and cutting-edge attention mechanisms, it offers capabilities that rival — and even surpass — many proprietary models.

By integrating DeepSeek-Coder into CodeGenie, we'll create more than just a coding assistant; we'll built an intelligent companion that understands, predicts, and collaborates with developers in real time. This fusion bridges the gap between AI research and practical development, bringing futuristic coding experiences into everyday workflows.



WHEN CODE WRITES ITSELF
“CODEGENIE MEETS DEEPSEEK-CODER”

THANK YOU!