

HACKATHON IBM Granite

CodeMyth: A guild that builds



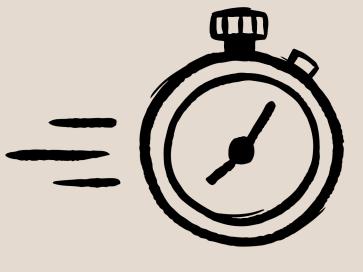


CODEMYTH





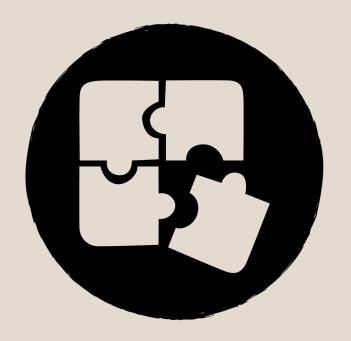




"MAKING LIVES EASIER FOR DEVS"

DEVDOCS: DOCUMENTATION CENERATOR

PROBLEM STATEMENT



The Challenge of Keeping GitHub Documentation Up-to-Date

Writing and maintaining up-to-date documentation is often neglected because it's time-consuming, repetitive, and, let's face it—less exciting than coding



Poor Documentation Practices

Poor documentation leads to confusion, onboarding delays, and more time spent deciphering code instead of building features.





INTRODUCING

AN AI POWERED GITHUB DOCUMENTATION GENERATOR:

IT CAN HELP DEVELOPERS WITH THE DOCUMENTATION, THEY CAN FOCUS ON CODING AND SHIPPING MORE.

What our application offers?



Just Plug and Play

35%

Users can just login with their Github



Less effort



Select the repository they want the documentation for



Less time

90%

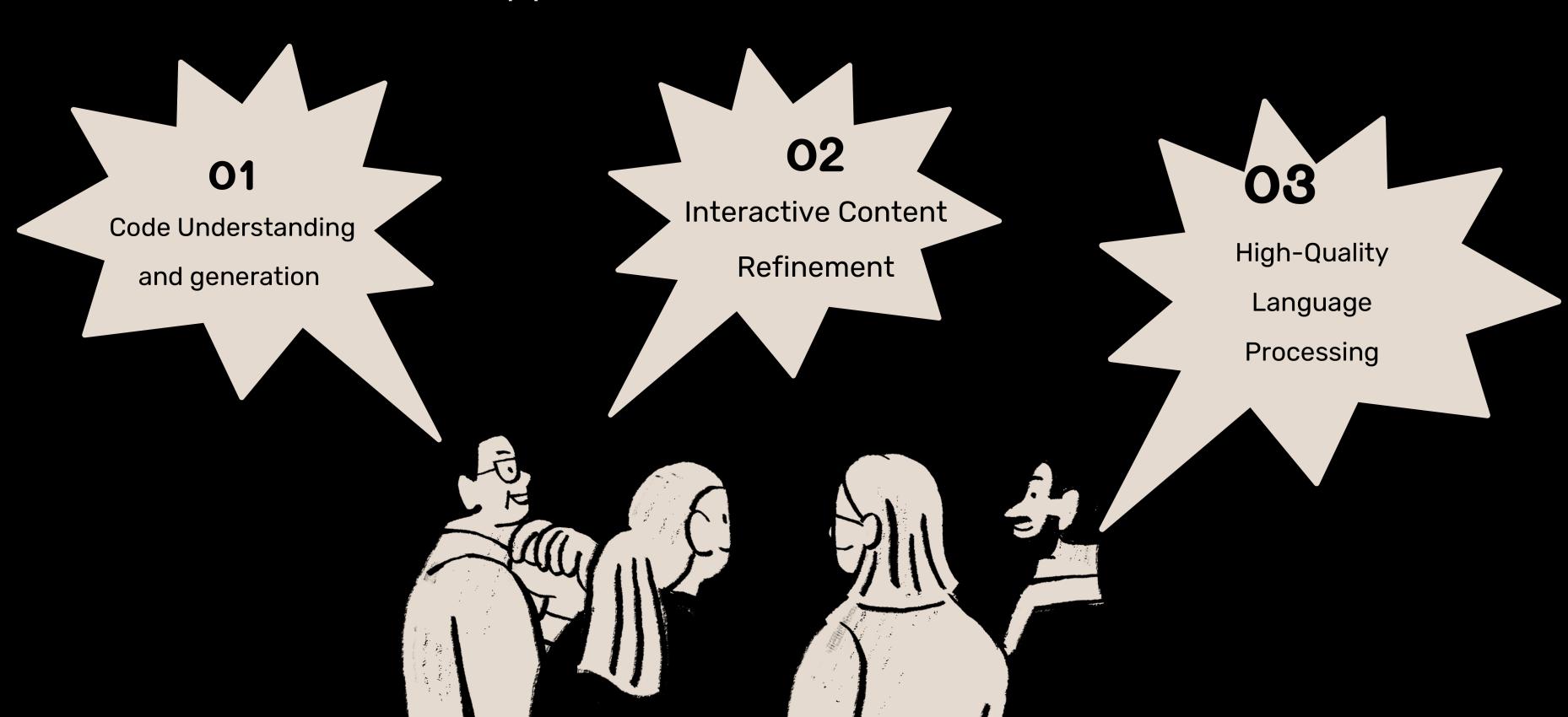
Review and give feedback and when they're happy,
VOILA!!

it'd be done in the click of a button.



IBM GRANITE LLMS

At the heart of our application is the IBM Granite 3.1 Dense 2B model.



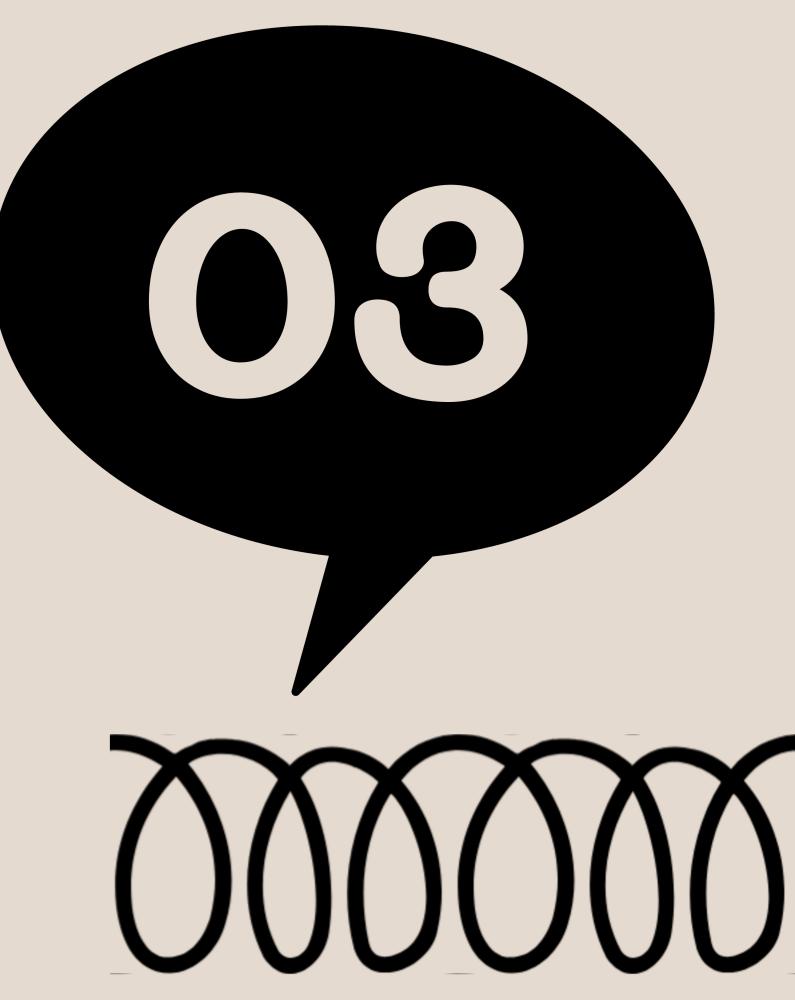
TECH STACK USED:



Backend	FastAPI
Frontend	NextJS
AI	Langchain_ollama
Model	IBM Granite Dense 3.1:2B

APPLICATION FLOW



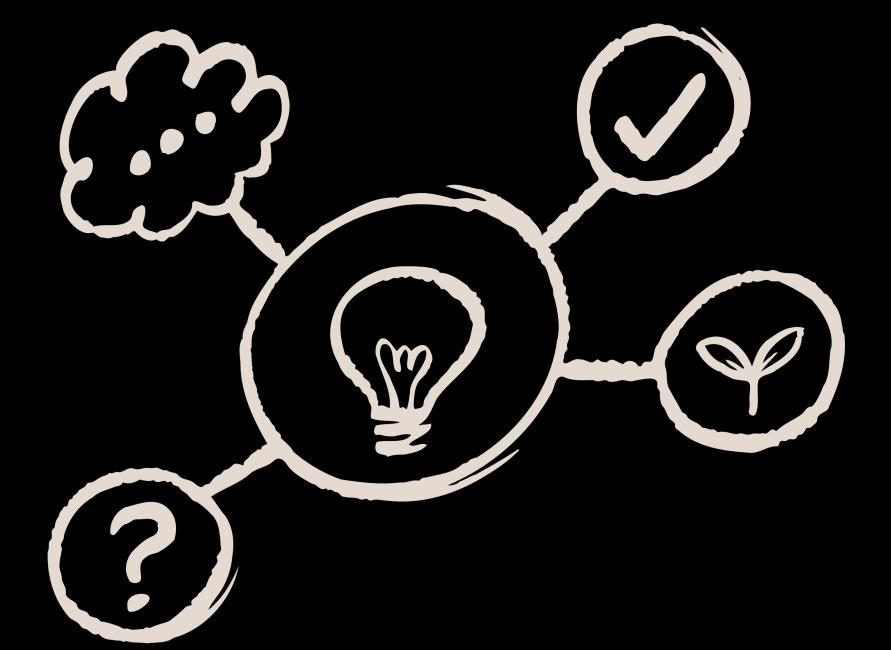


Important Ideas



Github Login and Authentication

Secure GitHub OAuth login for seamless access to user repositories.





User selects the Repositories

Users choose desired repositories; relevant code files are fetched via the GitHub API.



Al Generates documentation

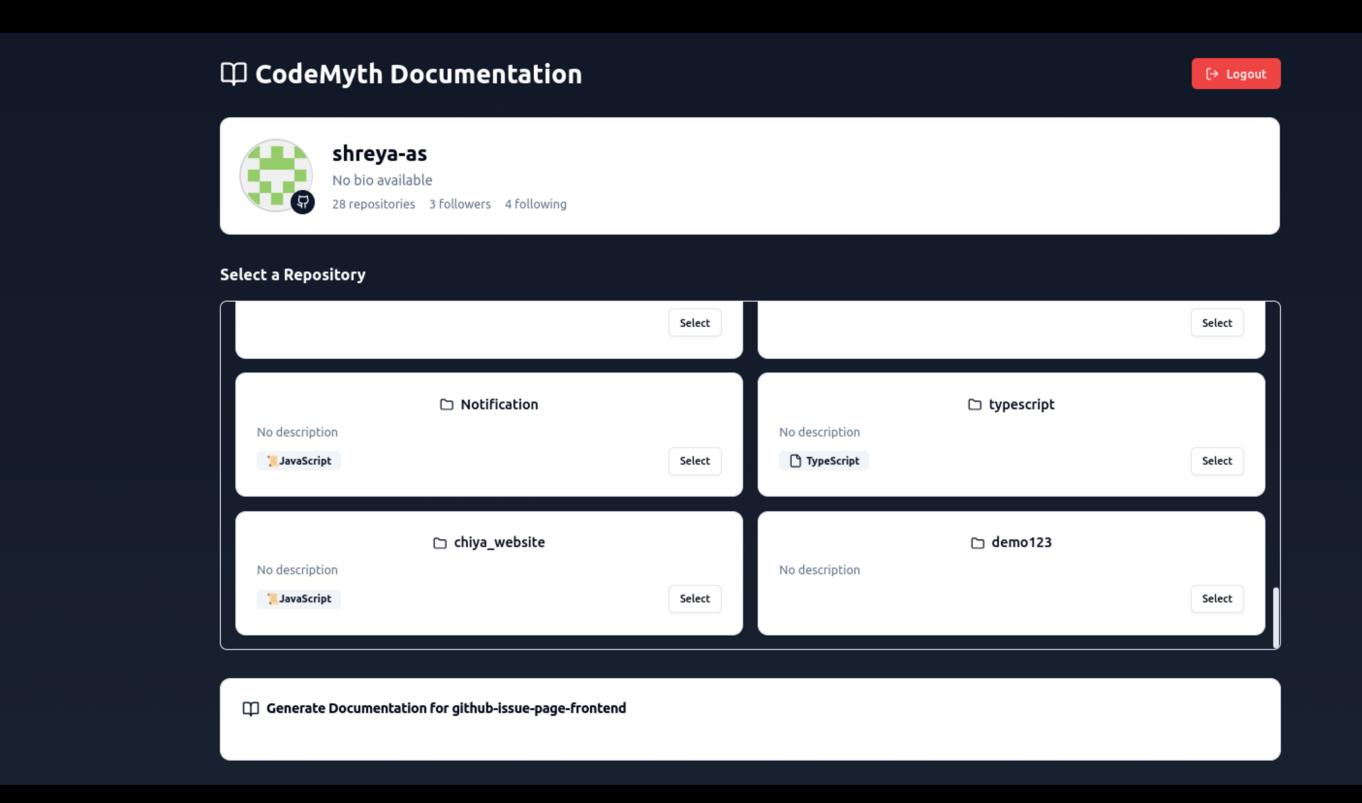
IBM Granite 3.1 Dense model generates accurate, context-aware documentation for selected repositories



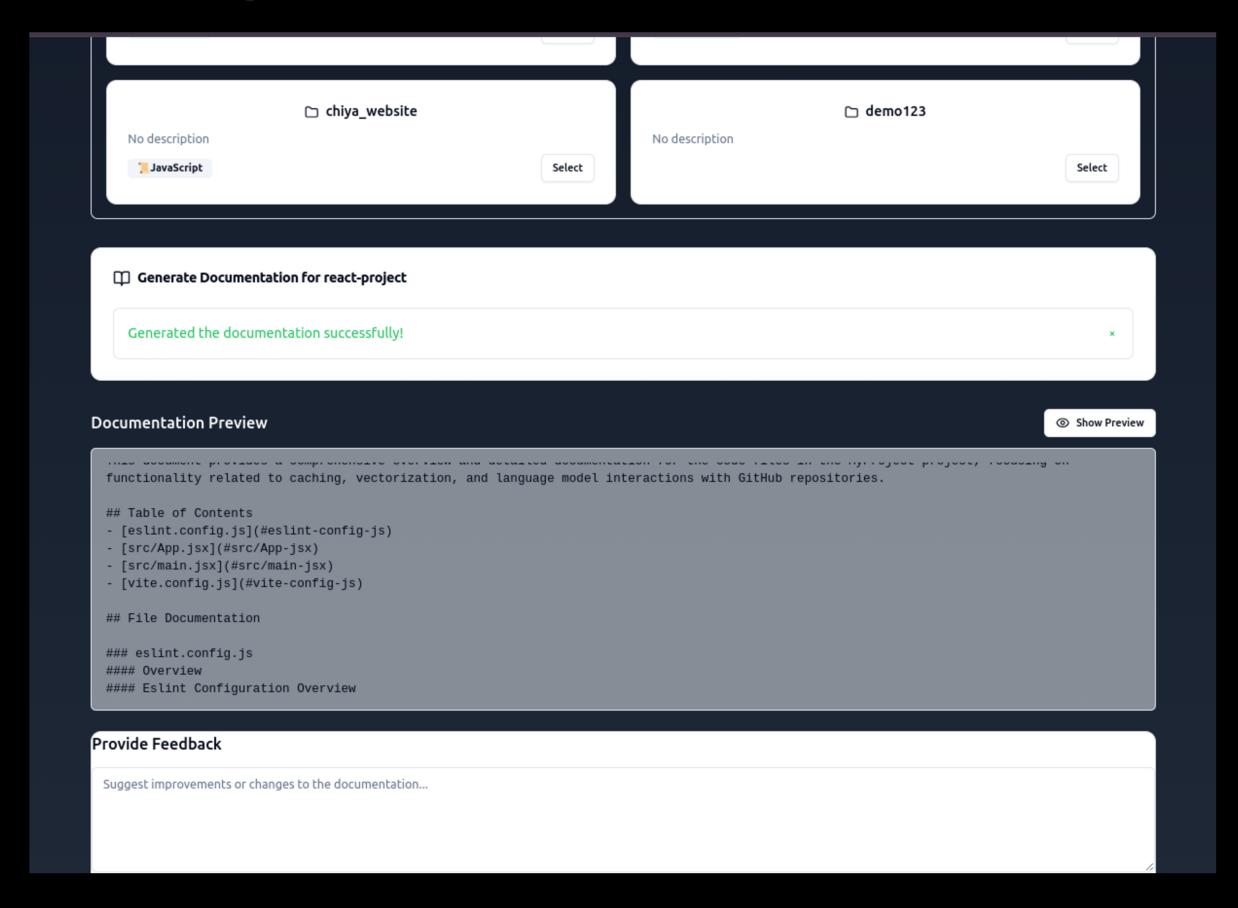
User Feedback and refinement

Users review documentation, provide feedback, and AI refines content in real time and commit when satisfied

1. User Dashboard

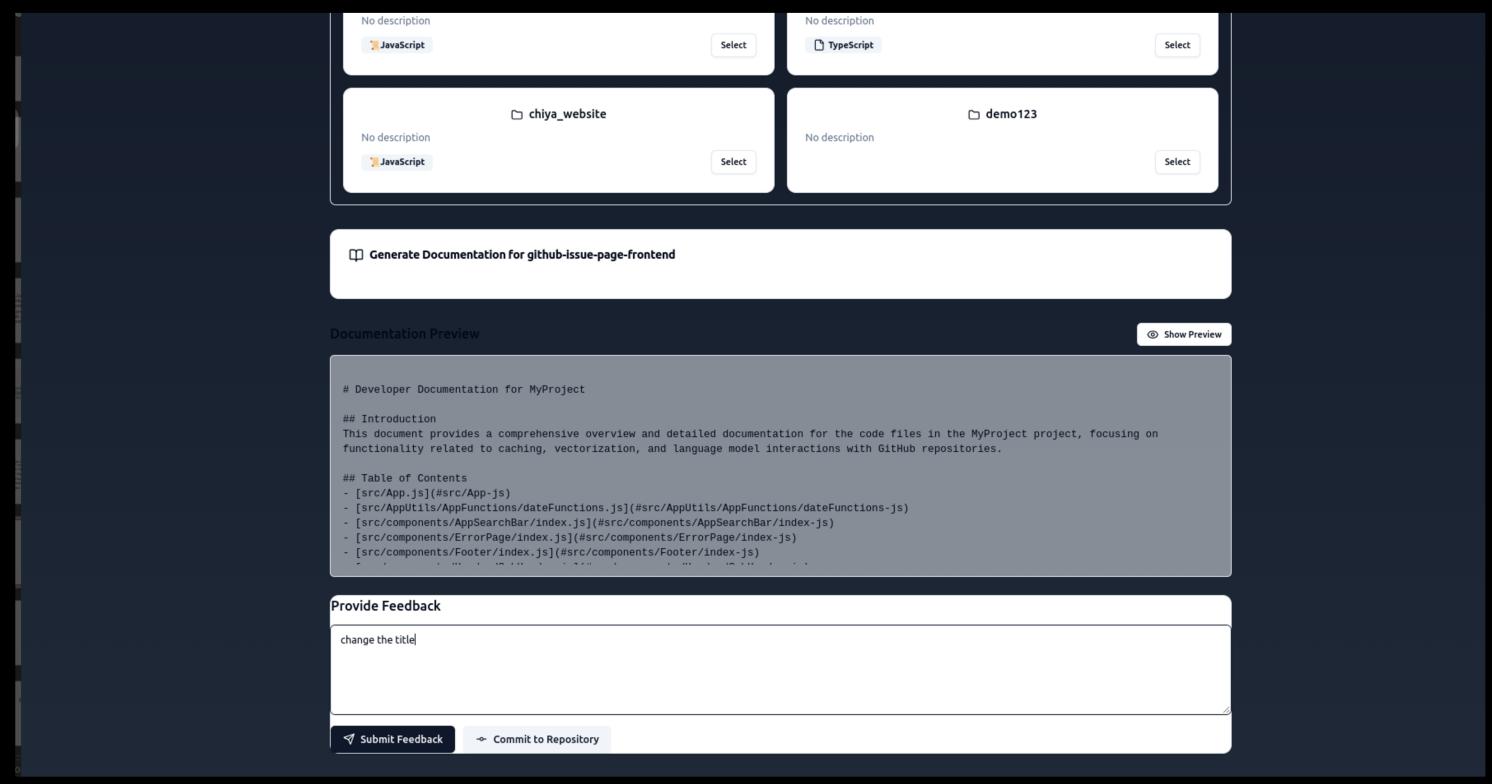


2. Generating documents for selected repo

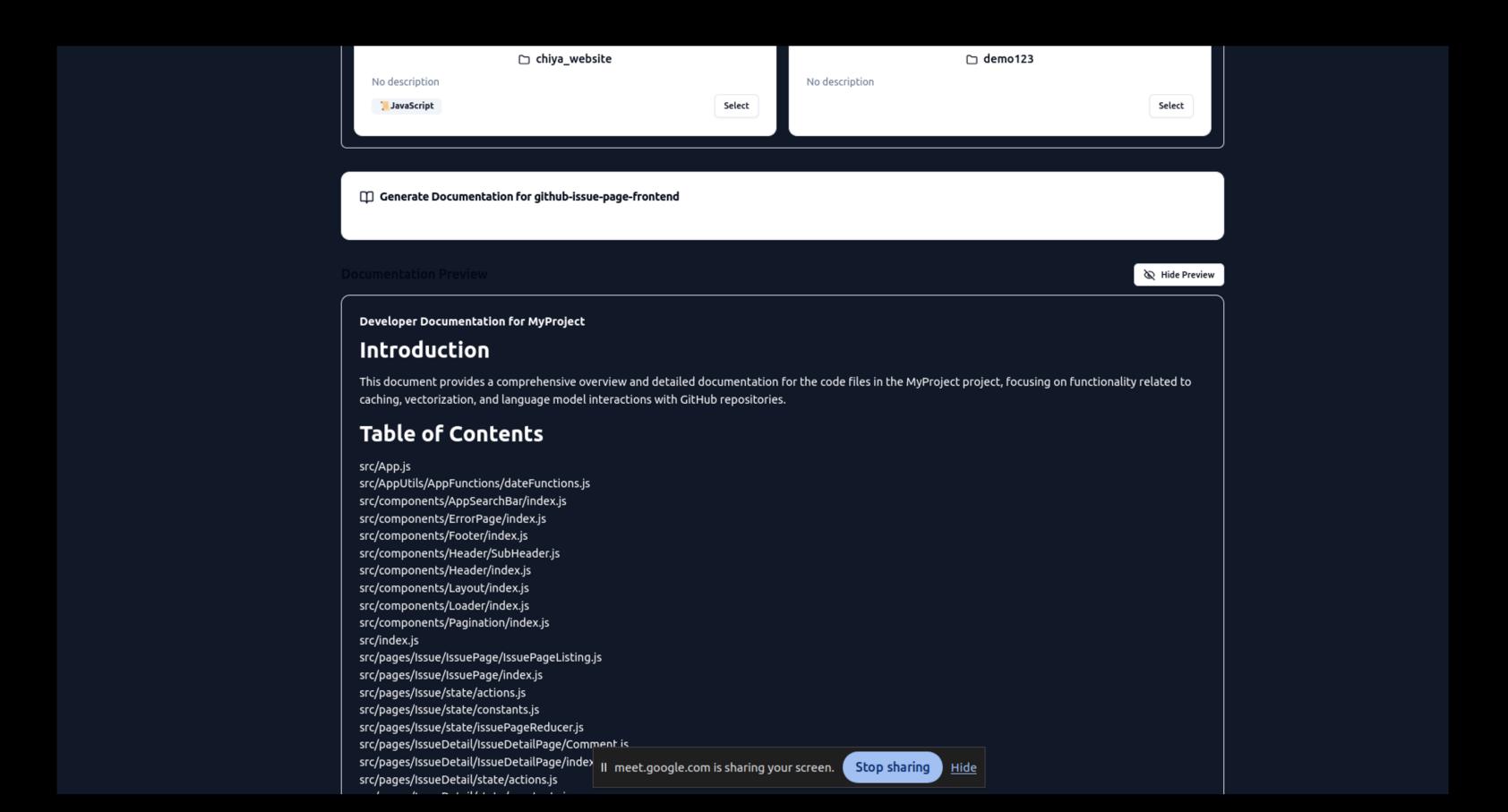


3. User feedback interface for the generated

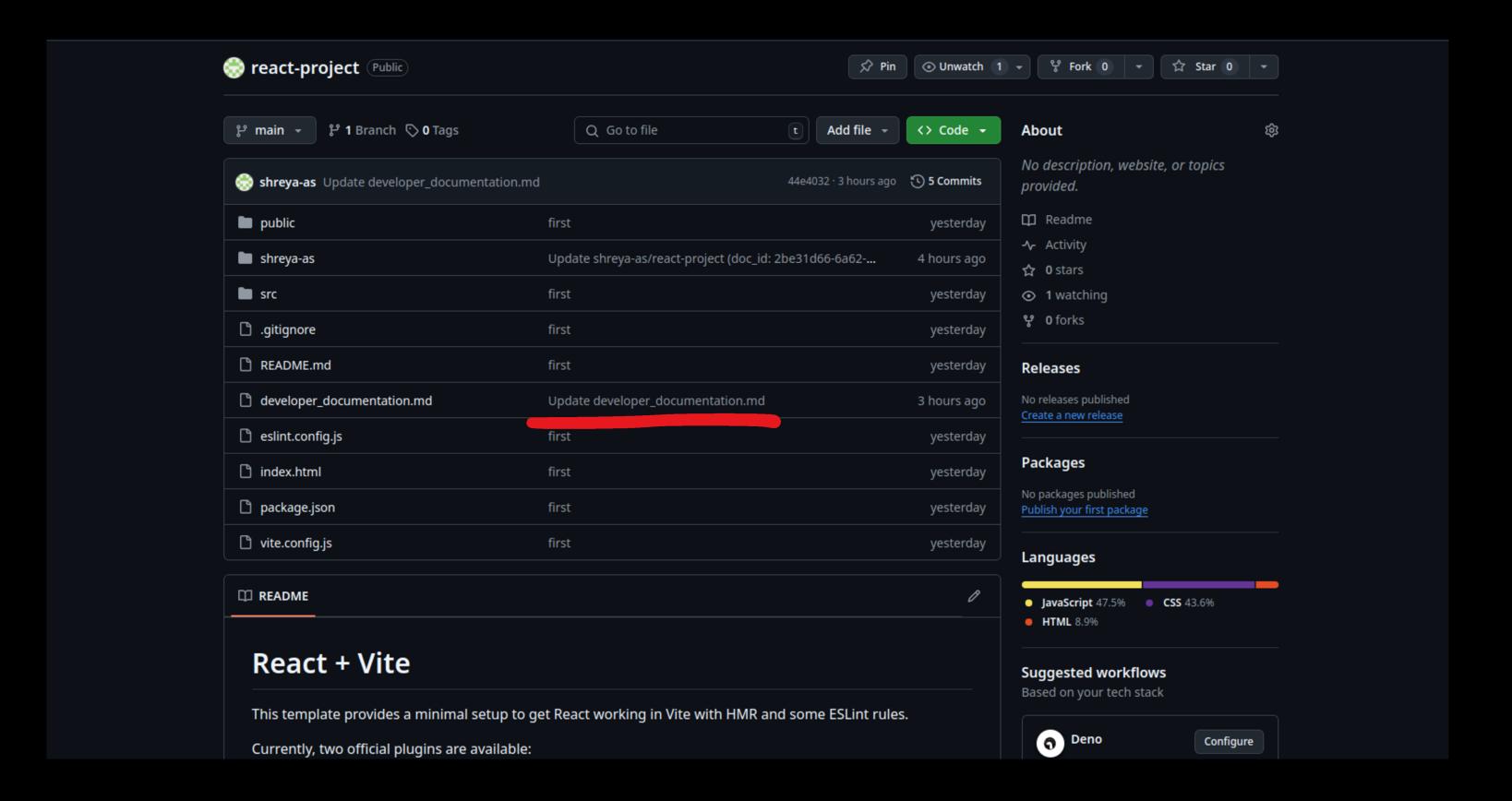
docs



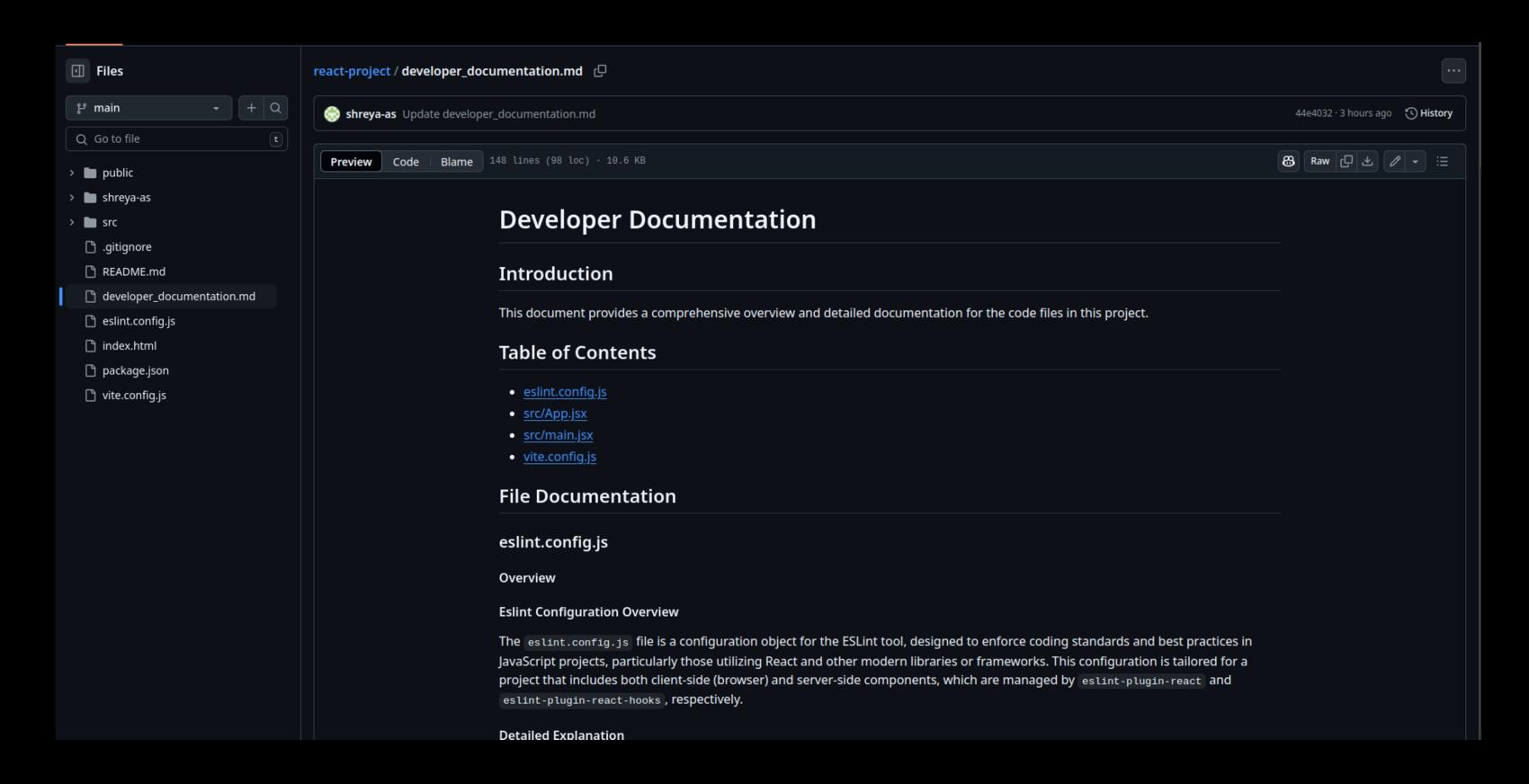
4. Preview of documenation



5. Commit Message



6. Documenation on github



Key Strategies





Token-Based Chunking for Large Files:

The chunk_code function uses RecursiveCharacterTextSplitter with a token-based length function (tiktoken) to split large code files into manageable chunks. This ensures that each chunk stays within the model's maximum token limit (4000 tokens), allowing efficient processing without truncation.



Incremental Documentation Generation:

For large files, documentation is generated chunk by chunk using the generate_doc_chunk function and then merged with clear chunk indicators. This approach enables the model to handle large files effectively while providing granular documentation per chunk.



GitHub API Integration for Seamless Deployment:

The app integrates with GitHub's API to push final documentation directly to repositories. It handles token validation, repository/branch checks, and file updates (with SHA validation), ensuring a secure and streamlined deployment process.

