

# Repo Analysis Chatbot

## Product Concept Description

|              |           |
|--------------|-----------|
| Shai Ballali | 207380379 |
| Yarin Barnes | 318851300 |
| Noy Keren    | 313307316 |

## Need

Developers often face various challenges when trying to understand and navigate complex codebases on GitHub, such as:

- Figuring out how to use a new library
- Understanding the purpose of the codebase
- Identifying the technologies used
- Pinpointing weaknesses and areas for improvement
- Gaining insights into specific functionalities or components of the codebase

These challenges can be time-consuming and require significant manual effort, hindering developers' productivity and efficiency.

## Innovation

The Repo Analysis Chatbot differentiates itself from other code analysis tools and methods by offering unique features and capabilities:

1. **Social and Cooperative Feature:** The chatbot is designed to be inherently social and cooperative. Once a user initiates the scanning of a repository, that repository becomes accessible to the broader user base, fostering a sense of community and optimizing resource utilization.
2. **Repository Marketplace:** The chatbot includes a "marketplace" feature that showcases the repositories currently embedded within the system. Users can browse this marketplace to discover and select repositories of interest, initiating inquiries and discussions around these pre-scanned codebases.
3. **Predefined Prompts & Social Ranking:** The chatbot offers well-tested prompts and the ability for users to upvote prompts based on the answer quality. The highest-ranked prompts for each repository are displayed as "Guaranteed and Tested Prompts," ensuring users can obtain great answers.

## Offering

The Repo Analysis Chatbot utilizes Retrieval-Augmented Generation (RAG) to ingest data from specified GitHub repositories. It transforms raw code and related documentation into a structured format using OpenAI Embeddings API and stores the data in a vector database. This enables the chatbot to retrieve relevant information efficiently when responding to user queries, ensuring accurate and contextually relevant responses.

The chatbot supports a wide range of programming languages, including JavaScript, TypeScript, Python, C++, Go, Java, PHP, Ruby, Rust, Swift, Markdown, HTML, LaTeX, Solidity, Protocol Buffers, reStructuredText, and Scala.

## Market

The primary target audience for the Repo Analysis Chatbot includes:

- Individual developers
- Development teams
- Companies

The chatbot is particularly valuable for developers, as it offers an unparalleled opportunity to dive deep into the intricacies of their code, uncovering insights that can lead to enhanced test coverage, optimized code quality, and a more holistic understanding of the codebase with minimal effort.

The potential market for this product is vast, as it caters to the needs of developers across various industries and programming languages. The demand for efficient code analysis tools is expected to be high, given the increasing complexity of codebases and the need for developers to quickly understand and navigate them.

Unlike existing solutions like LLM models, which may have outdated data and limited ability to answer specific questions about repositories, or GitHub Copilot, which requires an expensive annual subscription and can only answer questions about specific selected files, the Repo Analysis Chatbot offers a cost-effective, per-use pricing model (via API Key) and can answer questions regarding all repository files.