Ruben Dennis

I've gotten the Open Evolve environment running on my personal system. The initial setup wasn't working right away though, Open Evolve repeatedly failed to properly call the Gemini API and entered a loop where it kept trying to ping the service. After some troubleshooting, we found out that the issue came down to how the API key was being referenced. On Gemini's website, the key is provided under its own variable name, but Open Evolve expects the variable to be named OPENAI_API_KEY. When I changed it to the proper name it ran smoothly.

We also ran into another issue with Gemini's student tier, the API doesn't properly use the tokens. This limits how well Open Evolve can interact with it. Since I personally have a paid subscription for ChatGPT, I've been exploring whether there is a way to leverage that access to ping the API for the paid version instead. This may become important if we want consistent and some higher quality responses for larger test cases.

On the experimentation side, I have been able to run the provided example problems, including Function Minimization and AlgoTune Optimization, across several different AI models: Gemini 2.5 Pro, Gemini 2.5 Flash Lite, and GPT-5. At this early stage, I haven't noticed significant performance differences between the models, though that may change once we feed them larger or more complex datasets. For now the outputs look fine and stable, but the lack of variation leads me to wanting to make some more complex problems.

Another thing we brought up to the sponsor is to create a script that automatically documents the different iterations of the "evolved" algorithms and their evaluation results. Open Evolve already produces these iterations as part of its workflow, but right now they are not logged in a structured way. The goal is to checkpoint each version of the algorithm, along with its evaluation metrics, and store them in a CSV file for analysis. This will allow us to compare results across multiple runs, visualize improvement trends, and benchmark performance over time.