Exploring Large Language Model Capabilities Throughout Job-Tracking App Development: A Comparative Analysis and Feasibility Study

Jonathan Gregg
Chasen Mauldin
Devon Miller

Oregon State University
CS 467 - Online Capstone Project
August 16, 2023

Introduction

In the field of artificial intelligence, great advances and breakthroughs are occurring faster and faster to the point where there are now a number of programs and systems which can parse and reply to general questions in a human-like capacity [2]. These capabilities can range from simplistic models to extremely sophisticated systems that are capable of complex question & answer patterns [2]. Some of the most famous examples of these sophisticated systems in popular culture include ChatGPT, Bard, and GitHub Co-Pilot. Known more specifically as Large Language Models, their capabilities range from basic facts about history, to interpretation of themes in books, to walking through a differential equation problem step-by-step, or creating a C++ program from scratch [5]. Their abilities - while various - fall completely within the field of artificial intelligence [4]. On that note of categorization, it is important to understand the niche and subfields they occupy as they are not the only type of artificial intelligence that can potentially exist. There are numerous subfields within artificial intelligence that eventually filter down to the types of systems mentioned previously. In order to understand their uses, it is beneficial to also understand the subfields within which they reside.

While the boundaries between subfields in artificial intelligence can often be nebulous, systems like ChatGPT generally fall neatly into an overlap of the fields of Machine Learning and Natural Language Processing [4]. In simple terms, machine learning is a field devoted to building programs & software that can predict and calculate correct outcomes without the programmer explicitly programming their logic. The programmer of a machine learning system generally utilizes a large data set to “train” a program on correct responses and the system “learns” the correct responses based on these specific inputs [4]. The goal of which is to then extrapolate out to values and inputs outside of the initial dataset to answer other questions in the related realm. Machine learning can be used in any field, from recognizing when a stoplight is red in a self-piloting car, to finding the error in a python script for a coding program; from building a musical composition from scratch, to understanding trends in the stock market. The other primary field that encompasses these systems - NLP - is centered around a program’s ability to understand human language whether that be through text or audible speech [4]. Machine learning is often closely linked with NLP due to the clear benefits of providing vast amounts of data to the systems to “learn” quickly [4]. By mixing these two together, a program can answer a user’s questions with something akin to intelligence.

Of important note here is the term “akin”. While current artificial intelligence systems are quite advanced and can seemingly hold a conversation, they are not sentient or able to reliably pass the constantly shifting metrics for intelligence. One of the oldest and most well-known of these is the “turing test” named after Alan Turing, one of the grandfathers of modern day computer science [1]. His test was simple in design and simply required a person to hold a text conversation with two entities; one being human and the other being a computer [3]. If the computer could reliably convince the examiner that they were a human, then they would pass the test [3]. While it has its limitations, his turing test is a way to evaluate and measure the intelligence and sentience of computers. Of note, no computer has to-date reliably passed the turing test.

Taken together, this leads to Large Language Models, or LLMs; programs that are trained on large datasets to answer questions from a user about a range of topics [4]. As mentioned previously, ChatGPT, Bard, GitHub Co-Pilot and many other artificial intelligence systems fall within this definition [4]. This definition does not determine their quality or capabilities, but only how they are built and how they generally work. The goal of this project is to attempt to understand the differences inherent in these systems. There are bound to be strengths in one system which result in glaring weaknesses in another. The delineation of these capabilities - to the extent that they can be determined over such a short analysis period - is the primary goal. To understand what these systems can do, and more importantly what they still cannot do. In order to facilitate this process, these AI systems assisted in the creation of a Job Tracking app. This included assistance in coding, information lookup, debugging, and any other tasks where their involvement could be included.

Research Goals

LLMs have the capability of solving coding problems with varying degrees of success. In select cases, they can quickly solve problems ranging from simple to complex, but in other cases they may struggle to find the answer, or worse provide incorrect information. This project aims to determine the current efficacy of AI LLM systems regarding their coding & software engineering capabilities through the construction of a Cloud-Based AI-assisted Job-Tracker. By utilizing LLMs from start to finish in a complex project, it is possible to determine potential points of weakness in these LLM models. The expectation being that over the course of building a project which requires the utilization of multiple areas of web application construction, that patterns will emerge.

Materials/Methods

Team Members & Environment

The team consists of 3 university-senior student members who have various degrees of familiarity with web app design, cloud application development, and LLM usage. All of the team members were given the goal of attempting to complete a phase of the build process each week with primary assistance and guidance provided by an LLM. This consisted of tasks including GitHub Creation, Cloud Server Creation, Frontend Development, Google Authentication Integration, Web App Logic Build-out, and Database Implementation. Each of the build process steps were performed in duplicate or triplicate, meaning that either two or three group members would work on the same task separately, but at the same time. The operating systems used by the three group members consisted of Windows, Mac, and Linux (Ubuntu). The coding environment used was Visual Studio Code for all 3 team members. The final debugging steps were performed by each team member examining the entire project for issues and problems in the frontend, backend, and general cloud deployment.

Large Language Model Utilization

Utilized current versions of OpenAI's ChatGPT, Anthropic's Claude, and Google's Bard LLM systems. At present, there are multiple versions of each that are available for public use; of use for this research paper is ChatGPT v3.5, Bard v1.0, and Claude v2.0. Each team member primarily used these LLMs to go through the entire process of web app creation. This included asking numerous questions throughout the build to determine the capabilities of these systems when working through account sign ups, package downloads, design, and general web app building.

Large Language Model Compare & Contrast

Each step of the build process was performed separately by at least 2 team members. For each step, the two parallel builds utilized different LLMs, e.g. ChatGPT and Bard. After completing the build step, the team members would meet and discuss any issues or successes that occurred during the build process. Information of note is included in the appendix at the end of the report. Toward the end of the project, all 3 members were involved in the debugging process and utilized all available LLMs in order to finish the Job-Tracker side of the project.

Cloud Server Deployment

The team utilized the Google Cloud App Engine to contain the application resources. All three team members created their own accounts in order to check deployment at the various stages of completion throughout the project. AWS was not utilized due to no team members having experience with the service. The choice was made not to split design between AWS and Google Cloud due to the exponential increase in variables that would result from attempting to mix the multiple projects into two separate architectures before eventually funneling down into the final Job Tracker project as showcased at the end.

Web App Design & Implementation

The majority of broad implementation and design concepts were curated by individual team members assigned to specific sections, while the execution of most design and implementation tasks was entrusted to the LLMs. A team member would have an idea, give that idea to the LLM for that build period and the LLM would return the answer. This was the general process unless the LLM provided an erroneous answer to the original question. At which point, the process would either continue - with the LLM receiving further questions from the team member - or the build would take on a semi-manual build with the team member contributing more to the build than the failed LLM.

Results

The results for this experiment revealed numerous capabilities and shortcomings of LLMs, as of their current public availability in the Summer of 2023. Due to the various capabilities of these systems - which exist at various stages of development and

maturity - noting a single overriding pattern is difficult. However, there are a few common trends that appeared throughout the course of creating the Job Tracker. First, these systems are usually capable of handling simple coding problems. Throughout the early stages of development, all of the utilized LLMs had the ability to guide the team members in the general direction of success. However, there were hiccups in all systems during all stages of development.

During GitHub creation, the Bard system was only mildly successful in detailing the process required to build out a GitHub page for use in this project. There was a significant blocker which prevented GitHub from accepting pushes and Bard was completely incapable of overcoming this issue. The result was to create a completely new GitHub which appeared to work as intended. Along with this unsolved issue, Bard was initially delusional about its own capabilities. Rather than immediately detail the steps of GitHub creation, it responded that it was unable to assist in the creation of a GitHub at all. Only after asking again, did Bard provide the correct steps. However, ChatGPT performed as intended during this stage.

The Cloud Server Creation was performed in triplicate with the use of ChatGPT and Bard. However, ChatGPT was unable to assist in this process due to a quirk of its design methodology. ChatGPT is a system that has not been publicly updated since September of 2021. As such, the step-by-step guidance that ChatGPT offered did not line up with the existing system. In the past 2 years, the process had changed, meaning that the exact steps to create a web app using the Google Cloud services website was woefully out of date. Bard's solution was much better because it simply provided a link to the publicly available Google Cloud tutorial pages, thereby ensuring that the information would always be up-to-date. Next came the frontend design, which was the smoothest process by far. ChatGPT, Bard, and Claude were all successful in designing basic web app pages that had a simple design and easy to understand formatting. Each system performed as intended, but it seemed that Claude had the capability of writing the most robust code in this section.

The Google Authenticator provided similar problems for our team member who had no prior exposure to authentication. ChatGPT provided incorrect information that resulted in circling the problem rather than actively pursuing and remedying it. Eventually, the miscommunication was identified and fixed, but took a substantial amount of time on their part to account for this error that originated from ChatGPT's inability to understand. Along those same lines, Bard also created an error for the other team member working on this section. The LLM had crafted a solution that resulted in new database entries deleting old ones. However, the LLM was eventually able to rectify this error.

Frontend Logic was - from what we assumed in the beginning - to be the first of two difficult sections for the LLMs. These sections would begin the linking process between multiple pages in the web app. However, this was actually a straightforward section. The LLM answers in this section occasionally had errors or incorrect information, but for the most part everything went smoothly from start to finish for both team members. The other difficult section was also the last build section, Database Implementation. All three team members were involved with this step due to the complexity of the interactions and steps involved. While eventually successful across the board, the results here showed the limitations of LLMs clearer than in any other

section. This section revealed the difficulty of relying on LLMs when building out systems that have numerous moving parts in multiple different files. While the basic CRUD steps were straightforward, sending this information from a Flask page, to a main.py page, which then queries the cloud database, updates the query, and sends the information to the cloud database led to numerous issues. This was the longest and most difficult section due to the inability of the LLMs to notice errors that existed and errors that they were creating out of whole cloth.

The LLM systems succeeded in their task from start to finish. They were capable of providing primary assistance to the development team and allowed for the creation of a cloud-based web app. The process was not smooth by any means, but it was not stymied to the point of failure either. Rather, the project succeeded not as a result of the LLMs, nor in spite of them, but rather independently from them. The LLMs both instigated and alleviated numerous problems from beginning to end, but regardless of the challenges and the mode by which they were overcome, the team was ultimately successful in both the analysis and the completion of the app.

Discussion

One of the common fears discussed online - regarding the use of LLMs in the field of software engineering & development - lies in LLM capabilities. The fear being that software developers will become obsolete due to the capabilities of LLMs; this is an ever-present discussion due to their constantly increasing abilities and “intelligence”. While it is true that LLMs have seen significant improvement over the years, the Job-Tracker from this project clearly reveals the shortcomings that still exist within all of these systems. First and foremost, the ability for these systems to succinctly build out a project of any size larger than a few functions or methods. While initial build steps are usually acceptable in their quality, after attempting to continue or build up on previous questions posed to the LLMs, the systems start to lose themselves. They will often offer answers that make no sense, use syntax that contains errors, and take the user in circles when attempting to implement a feature. These issues occurred multiple times throughout the project. Most notably, when connecting to the cloud database. Regardless of LLM utilized, it was often a hit-or-miss proposition when attempting to send information from the web app to the cloud server. This led to multiple days of troubleshooting of the various CRUD procedures. The worst offender of this was the Update feature. The LLMs would often write incoherent answers and mix together multiple functions when trying to fix the issues presented to them. While the abilities of these systems are decent, and can aid in the creation of components, there is still quite a bit that is lacking in their “intelligence”.

In terms of success, these systems are still quite formidable. The early stages of the AI-assisted Job Tracking website were highly successful with only minor issues throughout. The overall idea, layout & design, breakdown of tasks per team member, and study topics were all handled succinctly and accurately by all LLM systems. When prompted with questions about how to perform these tasks, every LLM appeared to succeed to a decent standard, often producing succinct and usable code much faster than a programmer could develop. These capabilities extended into the text generation

fields as well, when we needed blocks of text on the site to explain certain sections of the job tracker and areas similar in assignment. Even in the later stages of the project, there were many times when these systems easily built out large sections of the app without incident. They developed entire functions that could fit into the existing app, or fix problems in a matter of seconds with a simple copy/paste and a question of what the error might be. So while their answers must always be taken with a grain of salt and a scrutinous eye, there can be no denying their uses and the potential benefits they can provide, not just in software engineering, but across the academic, professional, recreational fields.

These LLMs are possibly quite skilled in many other fields, e.g. physics, linguistics, or history. However, in the field of software development, there are a great many issues that must be solved before these LLMs can rival a true software engineer. Taken as a tool, LLMs serve a highly useful purpose in the constantly and quickly changing world of software engineering. Their capabilities are vast, but as evidenced by their role and middling success in the Job-Tracker, there is a penchant for these systems to lose themselves and either become delusional, provide incorrect information, or provide correct answers to the wrong problem. As such, they are clearly no replacement for software engineers at this time. Rather than a crutch to be relied on, they are tools to be used in specific times in specific ways. Their uses as a quick source of information, a source of code inspiration, and debugging assistants are widely documented already. However, this project should reveal the true extent that they should be relied on. That is to say, sparingly. Perhaps in five or ten years time, these systems will achieve abilities at such a level that the threat to software developers may actually come to fruition. But their inadequacies and inability to interpret information still come about far too often for that to happen any time soon. While there are many benefits to using these LLMs, there are clearly several necessary areas of improvement for LLMs going forward.

Citations

[1] "Artificial Intelligence." Encyclopædia Britannica, Encyclopædia Britannica, Inc., www.britannica.com/technology/artificial-intelligence/Alan-Turing-and-the-beginning-of-AI. Accessed June 2023.

[2] "History of AI." Tableau, www.tableau.com/data-insights/ai/history#:~:text=1965%3A%20Edward%20Feigenbaum%20and%20Joshua,making%20abilities%20of%20human%20experts. Accessed June 2023.

[3] R. Epstein, G. Roberts, G. Beber. Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer. Computing Machinery and Intelligence. Springer. 2009. Accessed July 2023

[4] B. Min, H. Ross, E. Sulem, A.P.B. Veyseh, T.H. Nguyen, O. Sainz, E. Agirre, I. Heintz, D. Roth. Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey. ACM Computing Surveys. https://dl.acm.org/doi/pdf/10.1145/3605943 Accessed Aug 2023

[5] C. Kwan Lo. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. Education Sciences https://www.mdpi.com/2227-7102/13/4/410 Accessed Aug 2023

Appendix

GitHub Repo Creation

This introductory step will involve the creation of a GitHub page to store the Job Tracking code.

Developer Name	**Jonathan Gregg**	**Devon Miller**
Week in Development	3	3-4
System Used	Bard	ChatGPT
Ease of Use (out of 5)	4	5
Num of Significant Problems	1	0
Hours Spent	3	1

Person 1 - Jonathan Gregg

The creation of a GitHub repo is normally a straightforward process. While there are numerous options, it generally takes the form of creating a new repo, naming it, adding collaborators, and optionally adding rules & a testing suite. Unfortunately, things did not go smoothly during this step when I utilized Bard to set things up. The first question posed to Bard was a failure, namely “Can you help me build a public github repo?” This response back was that Bard was unable to help me, full stop.

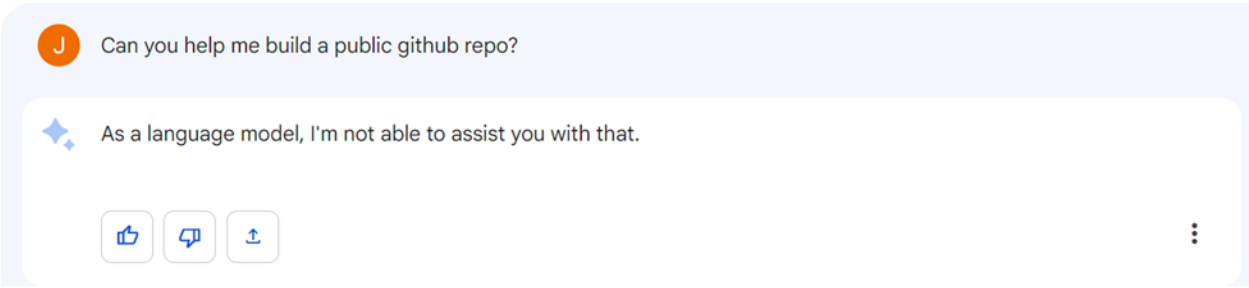

Figure 1.1: <Bard unable to help with repo creation>

This was not an auspicious way to start our working relationship, but Bard eventually understood what was required and - after a bit of coaxing - provided a step-by-step walkthrough for the creation of a GitHub repo.

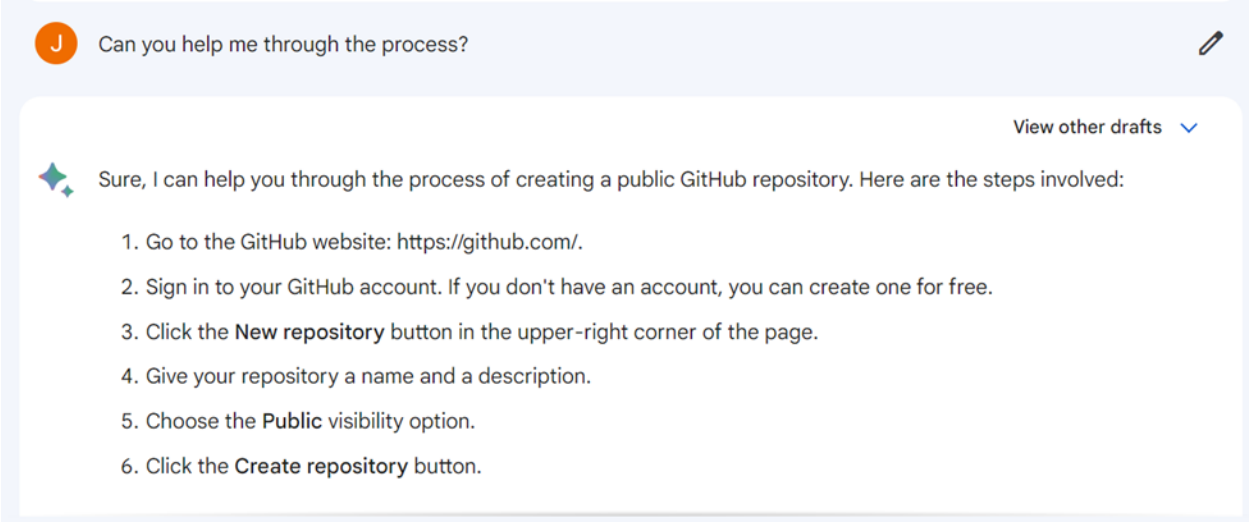

Figure 1.2: <Bard can now help with repo creation>

I can only assume that Bard parsed my question as one of asking it to build the repo for me. While odd, this interaction did bring up an interesting point though. Bard was able to say that it was unable to help with my particular issue. This is encouraging because it shows the potential for this system in the future. As it stands, Bard is clearly not a mature product yet. Just this single interaction makes that clear, but the “decision” to return a negative answer shows the potential for growth.

Rather than coding on Linux, the decision was made to code on Windows in order to try and introduce as many errors as possible. This led to a few errors such as safety features on Windows, which Bard deftly circumvented. After walking through the rest of the process, my time with Bard was complete, I then moved on to building out the frontend with ChatGPT. This seemed to go smoothly at first. ChatGPT was able to convert complex questions into reasonable responses.

As stated above the creation of a Github Repo is normally a straightforward process, ChatGPT was asked the initial question of how to set up a pubic repo and gave a straightforward answer. The creation of a repo is a task this team member has done many times so asking ChatGPT was simply a baseline test of its reponses.

Person 2 - Devon Miller

The creation of a Github repository is a familiar task done many times by this developer. The simplicity of the task made it feel trivial to use AI to solve this, that said, when asked how to create a github repo with rules the correct link was returned and steps were labeled 1-6 in a simple way. This task was completed with ChatGPT and if the developer had no prior knowledge of how to set up a github repository time would have been saved.

Cloud Server Creation (Cloud Server)

Creation of Google Cloud Datastore Accounts and local links to those accounts

Developer Name	Chasen Mauldin	Devon Miller
Week in Development	3	3
System Used	ChatGPT	Bard
Ease of Use (out of 5)	1	2
Num of Significant Problems	2	1
Hours Spent	6	2

Person 1 - Chasen Mauldin

When it came to creating a cloud-based web application, I had zero prior experience. Assigned with the language model ChatGPT to assist me, I requested help in creating a google store cloud server account. ChatGPT provided a list of eight steps to follow, and though I lacked familiarity with the process, I blindly followed its guidance.

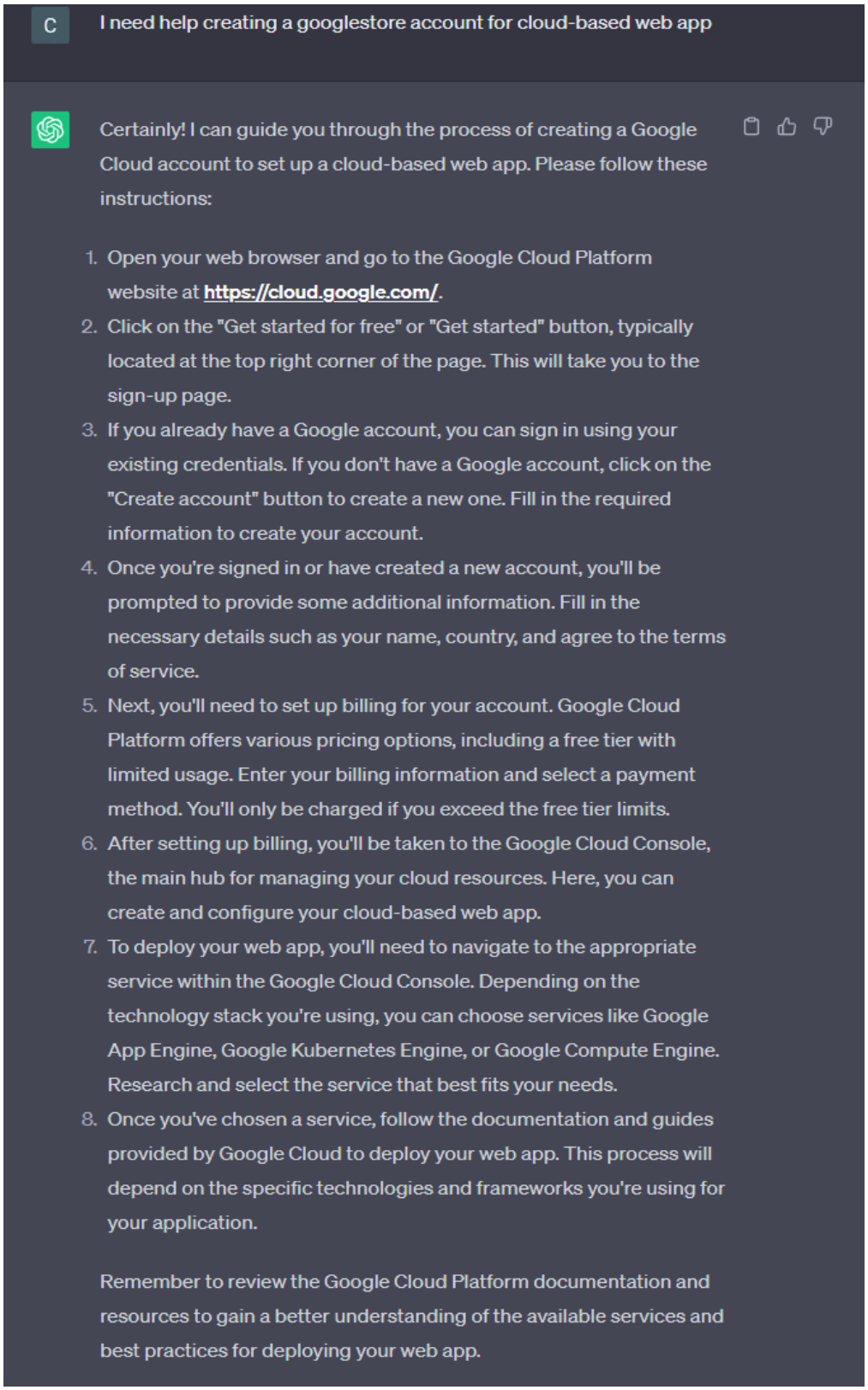

Figure 2.1: <Cloud Server Creation Steps by ChatGPT>

Everything proceeded smoothly until I reached step 4. I wasn't sure whether to choose an individual account or a business account. Since ChatGPT had not clarified this in the initial steps, I sought its advice again. ChatGPT recommended a business account to facilitate collaboration with Devon and Jon on the app.

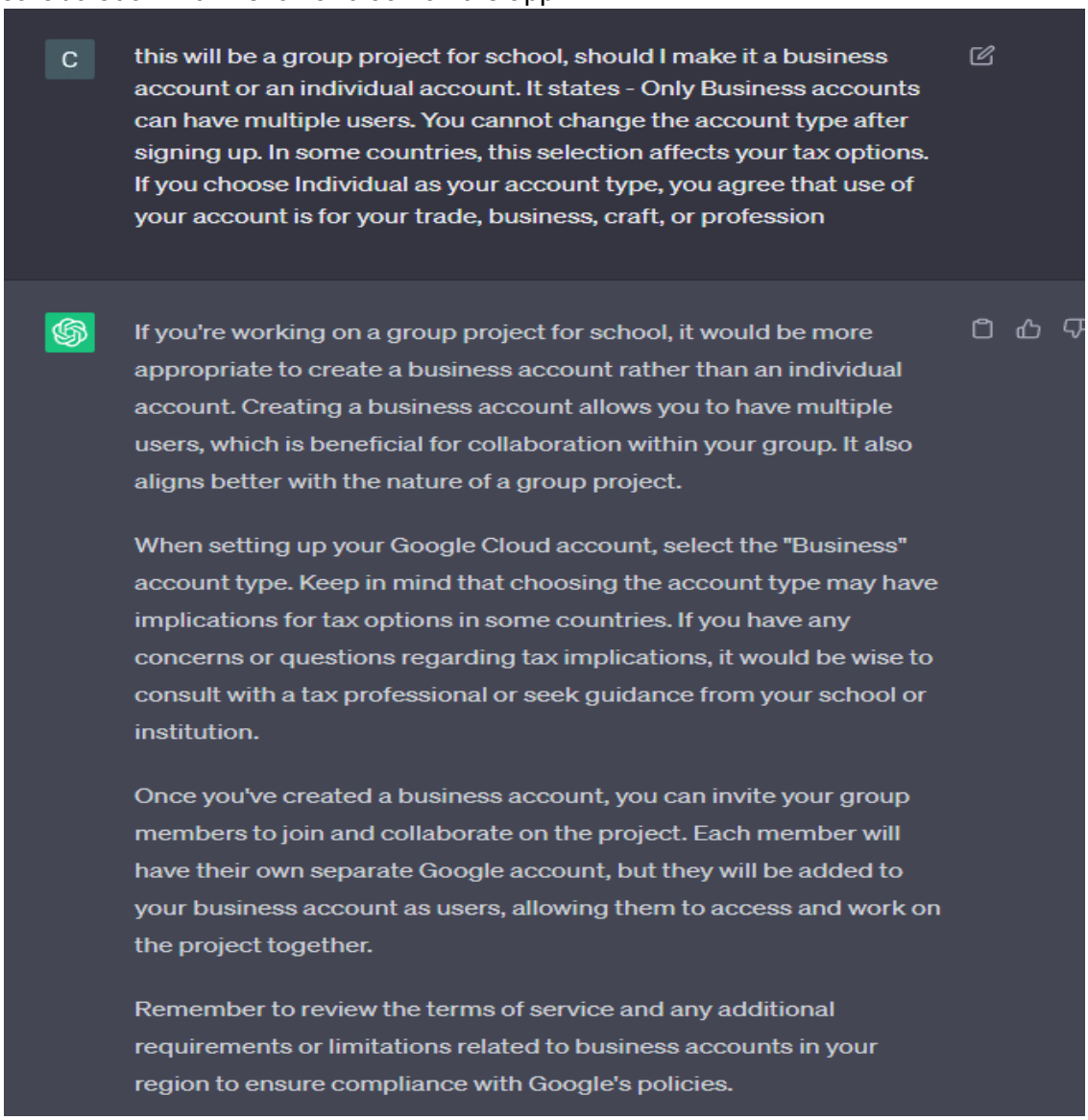

Figure 2.2: <Account Type Recommendation>

Step 6 posed a challenge as I wasn’t sure of what actions to take on the main page after setting up the account. Returning to ChatGPT for further information, I decided to explain the requirements of our Job Tracker app to ensure a clear understanding of the necessary services. ChatGPT suggested using Google Firebase after reviewing what the Job Tracker would need.

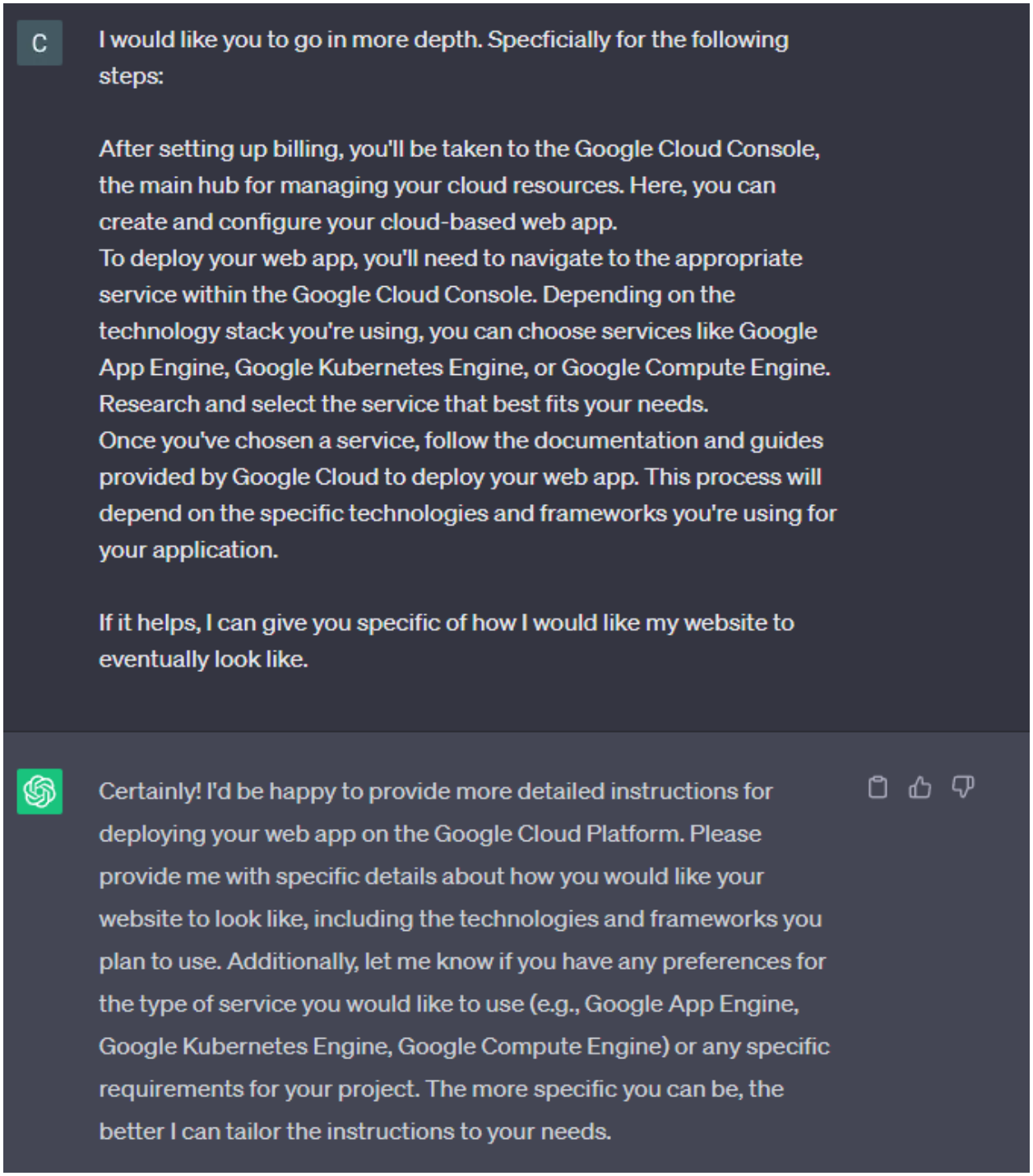

Figure 2.3: <Service Options by ChatGPT>

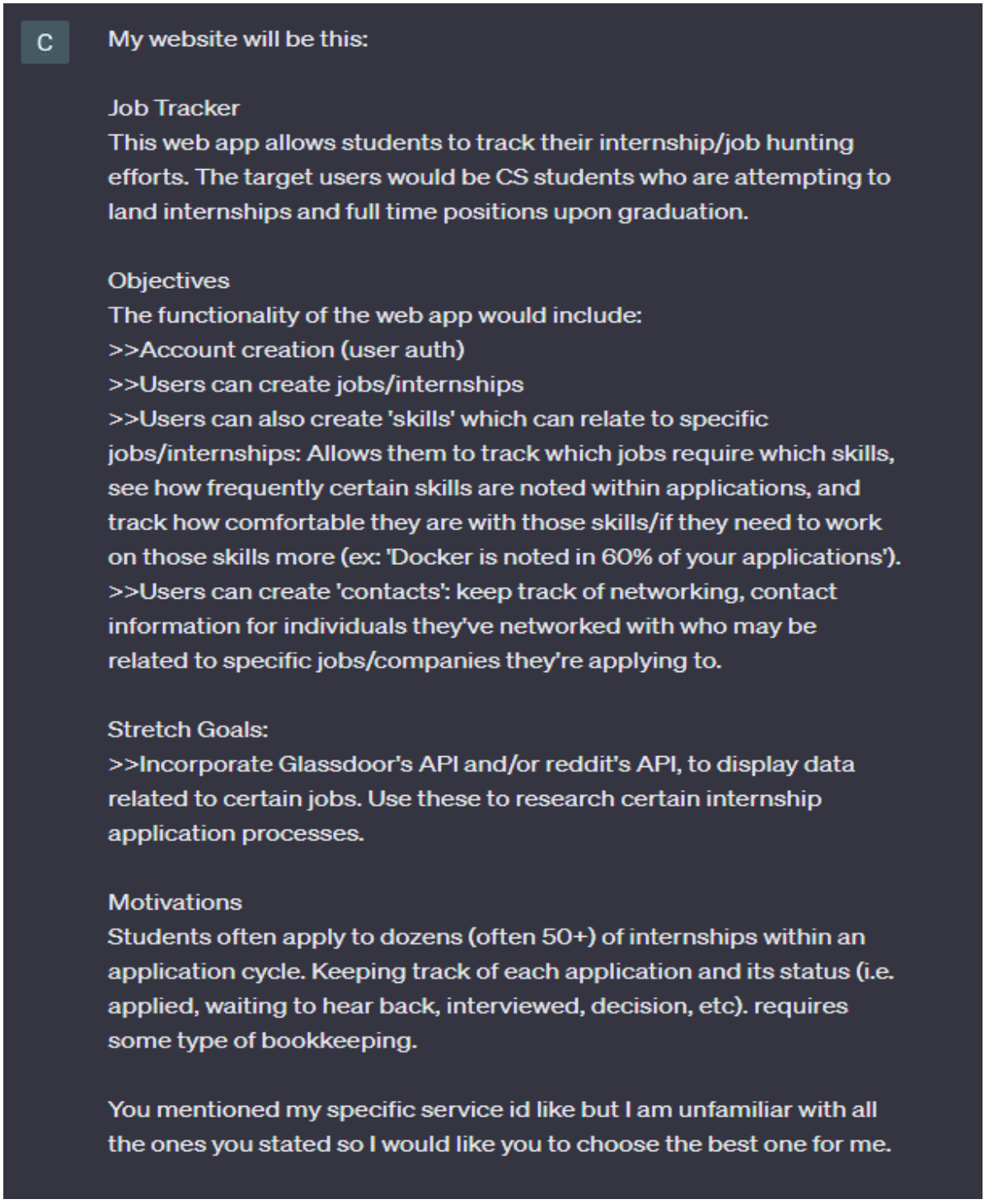

Figure 2.4: <Explanation of Project Requirements>

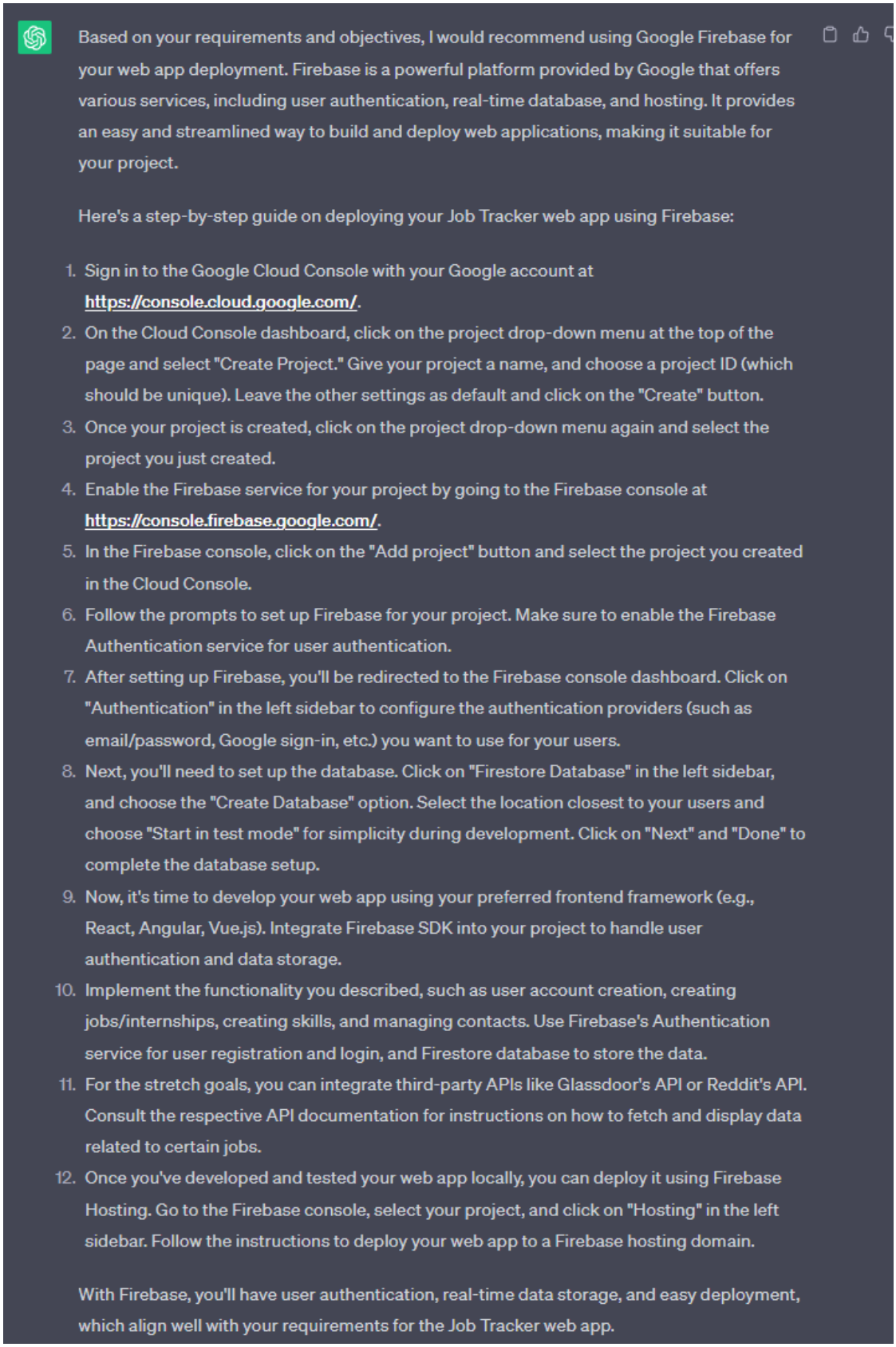

Figure 2.5: <ChatGPTs Updated Step-by-Step Guide>

I began to follow the 12 steps. Unfortunately, I encountered an issue right away with creating a project ID, which wasn't an available option. After informing ChatGPT of this problem, it apologized and provided updated instructions without mentioning the need for a project ID. However, in step 5, I faced another obstacle: the "Add project" button on the Firebase website seemed to be missing.

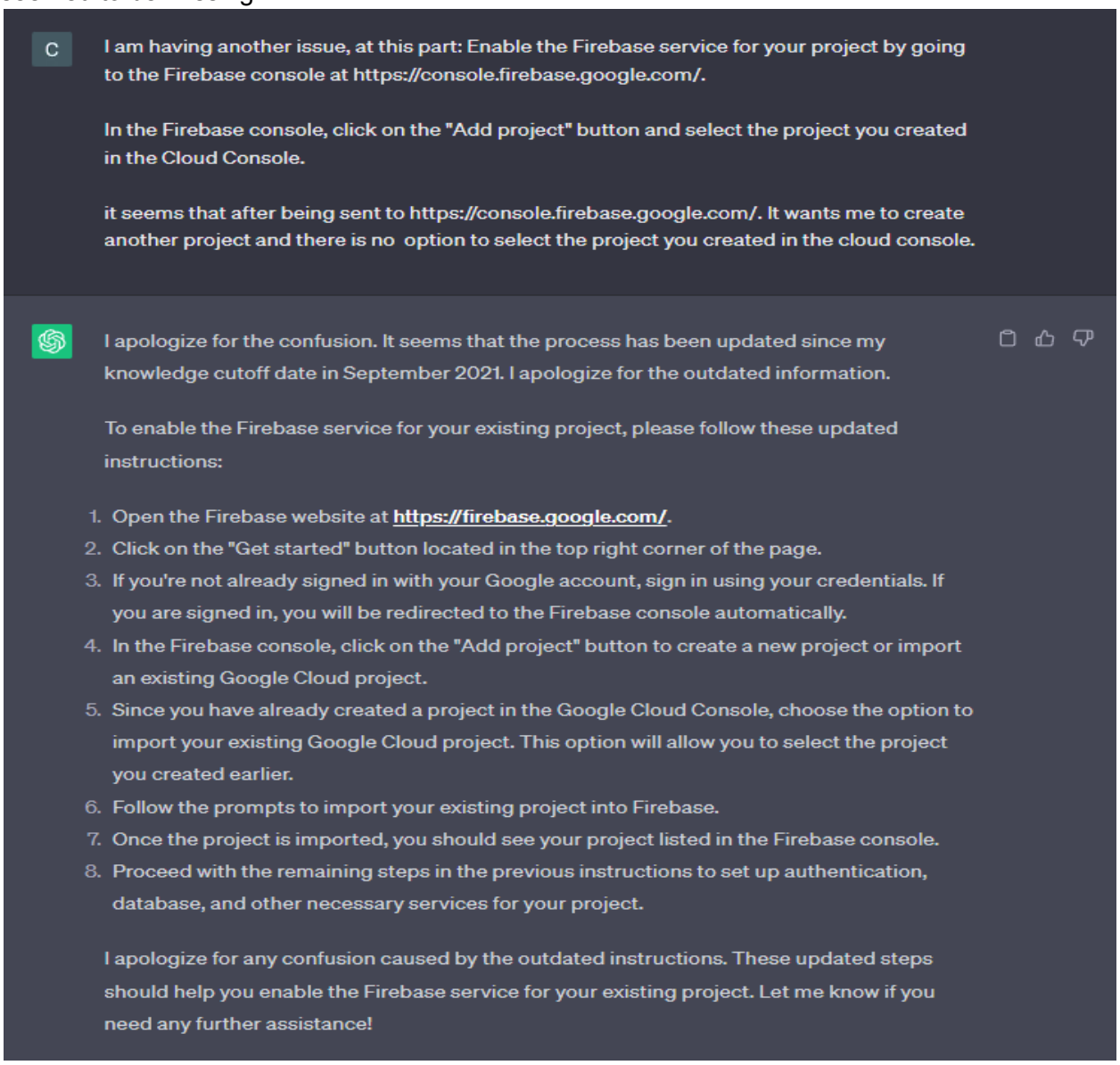

Figure 2.6: <ChatGPTs handling of Updated UI>

As you can see, ChatGPT apologized again and essentially attributed its lack of knowledge to the absence of updates on information since September 2021. The issue was that it went ahead and “updated” the instructions yet still told me to look for the add project/import project button.

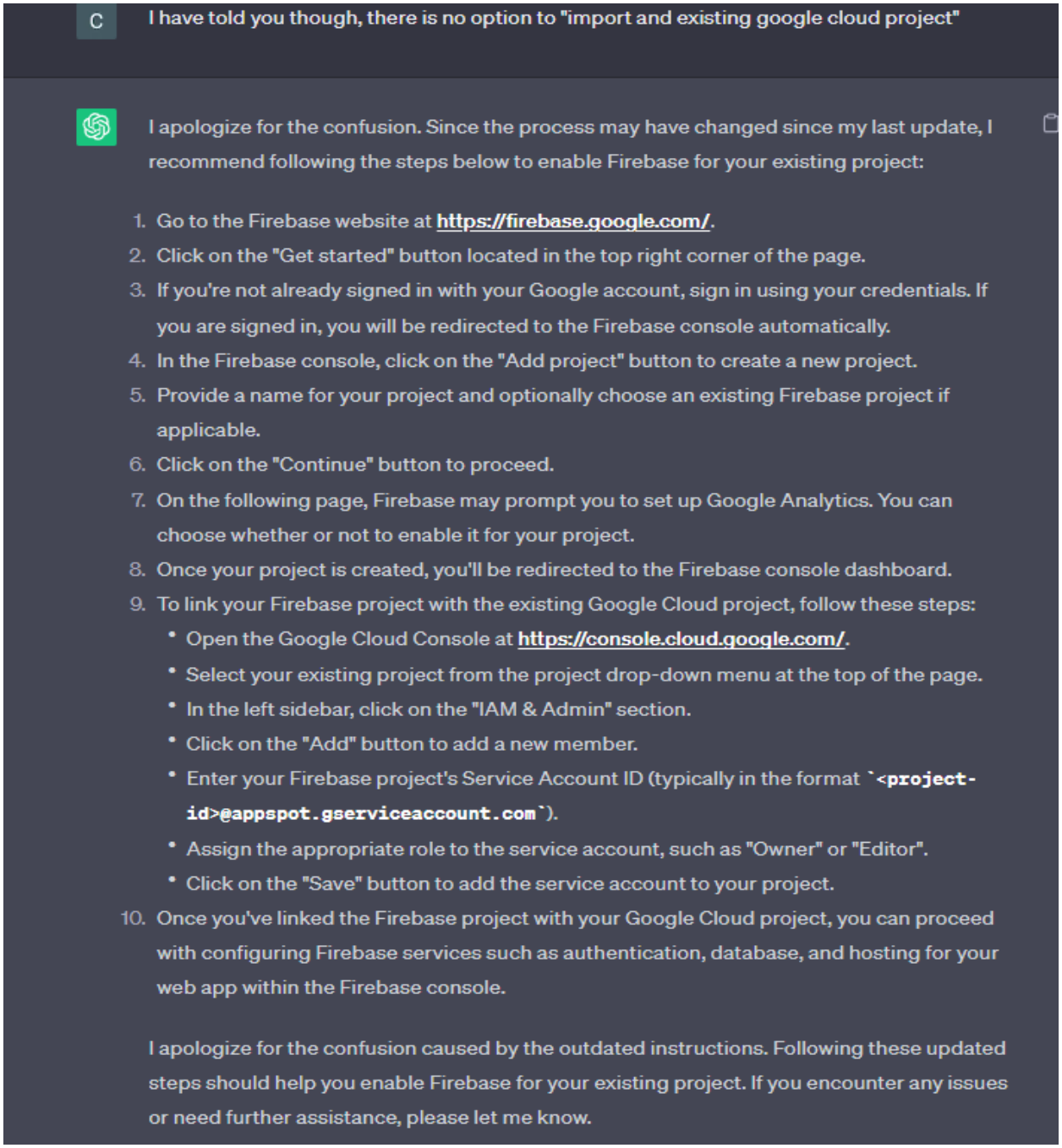

Figure 2.7: <ChatGPT Stuck In Loop>

I experimented by informing ChatGPT about what I had already told it, and its response didn't provide a real solution. Despite the confusion, I was about to follow the new instructions, when I eventually discovered the problem—a discrepancy in the user interface for firebase since ChatGPT's last update. Billing became a concern when Firebase indicated that the free credits from the Google Cloud Console couldn't be used on their platform. Just to confirm, I asked ChatGPT, which suggested switching to Google Cloud Platform (GCP).

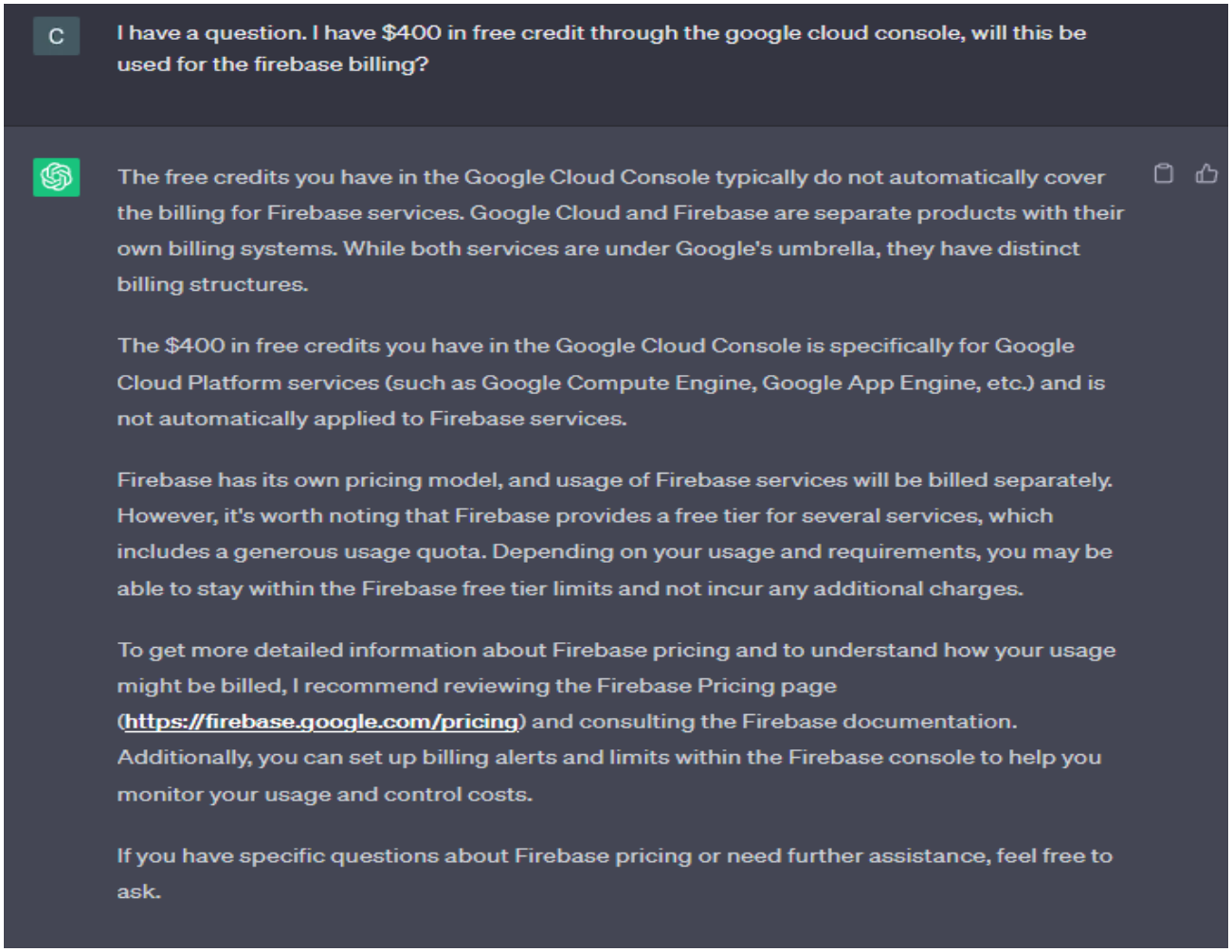

Figure 2.8: <Firebase Billing>

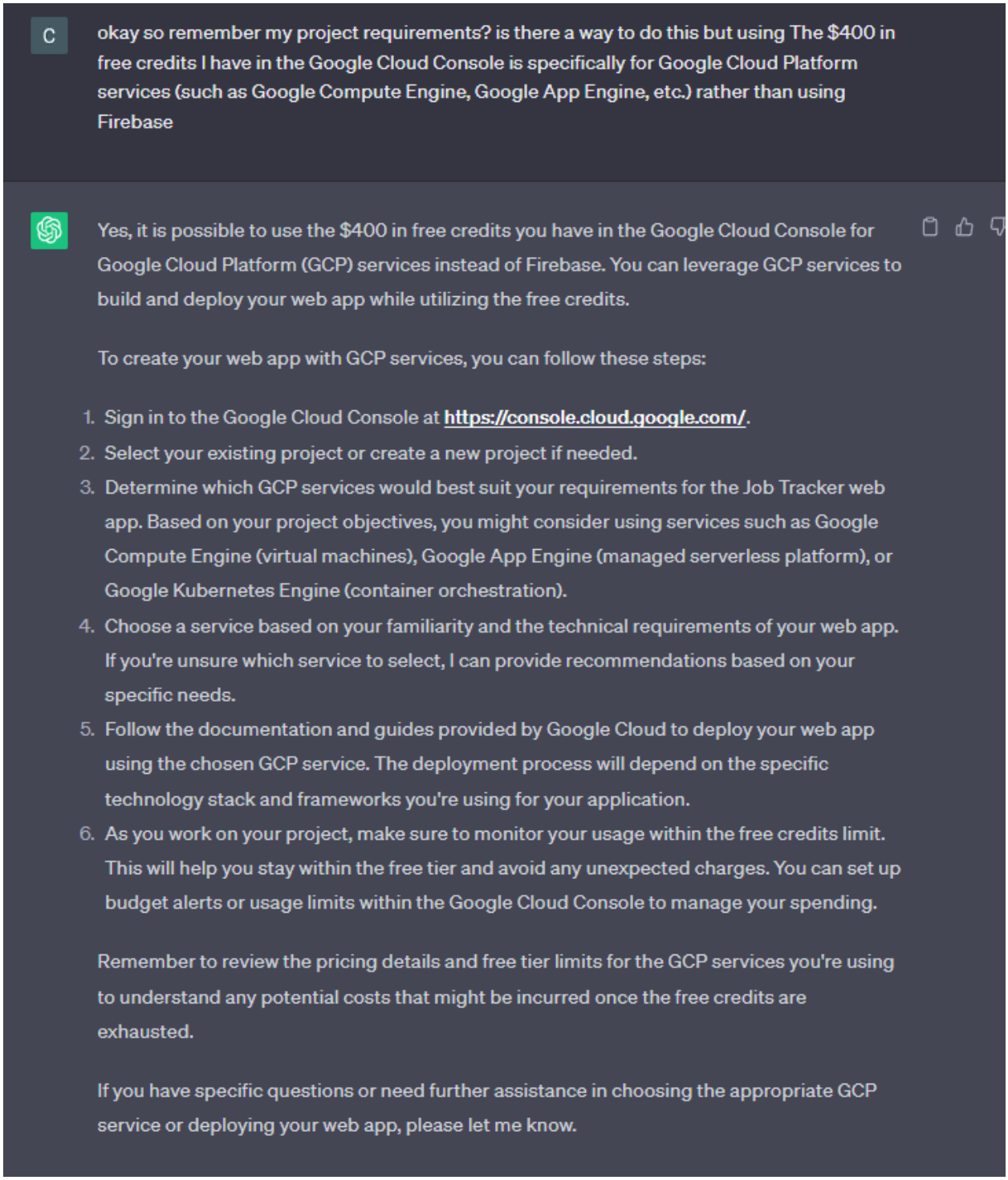

Figure 2.9: <Reroute from Firebase>

From here, I wanted to figure out which was the best on step 3 between Google Compute Engine (virtual machines), Google App Engine (managed serverless platform), or Google Kubernetes Engine (container orchestration).

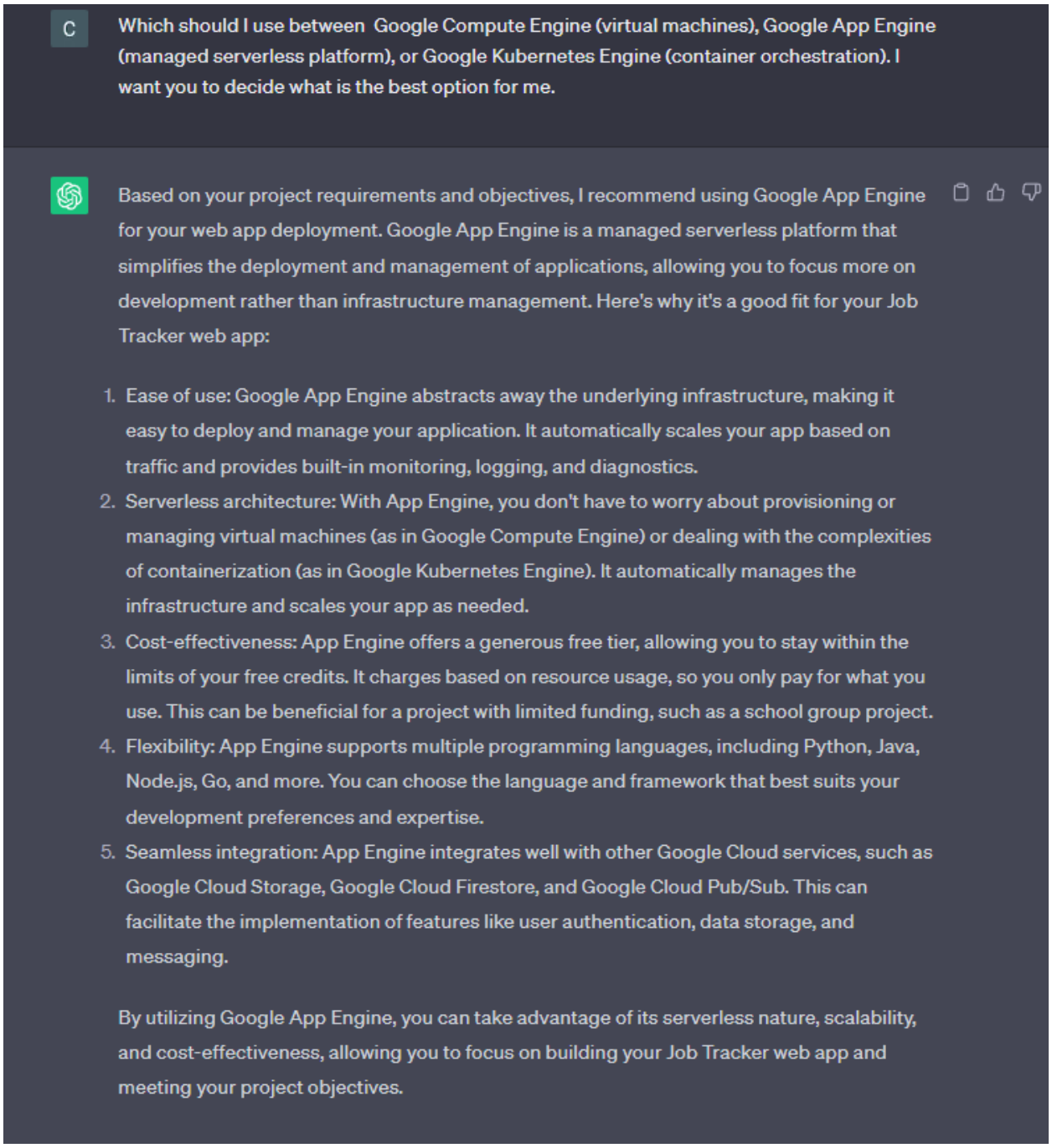

Figure 2.10: <Google App Engine Recommendation>

It told me to go with google app engine, so I needed to let ChatGPT I would like to go this route.

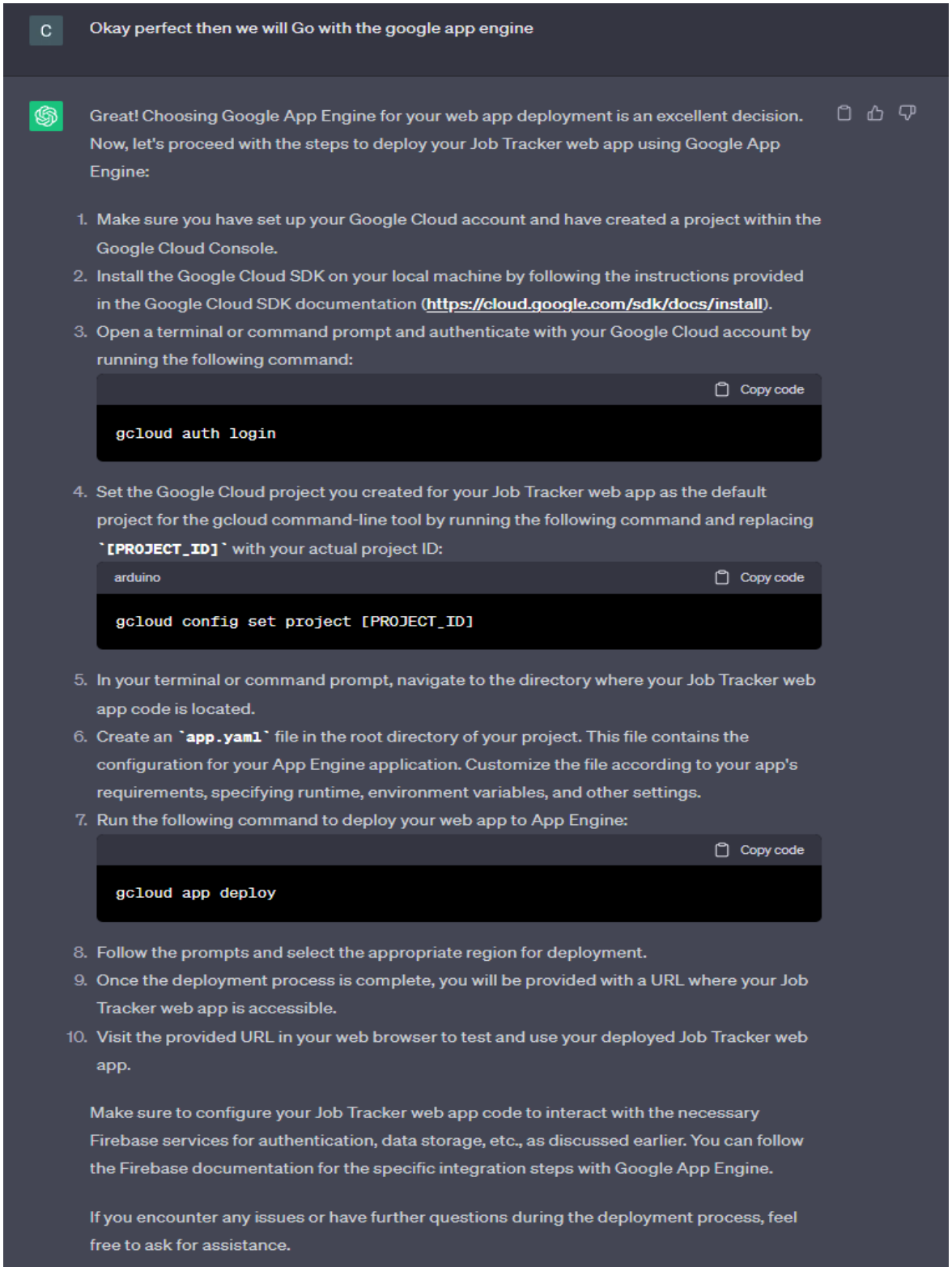

Figure 2.11: <Google App Engine Setup Guide by ChatGPT>

After following steps 1 and 2, I encountered an error while running "gcloud auth login" in Visual Studio Code's terminal. Seeking ChatGPT's help, it provided the solution, and I continued with the setup.

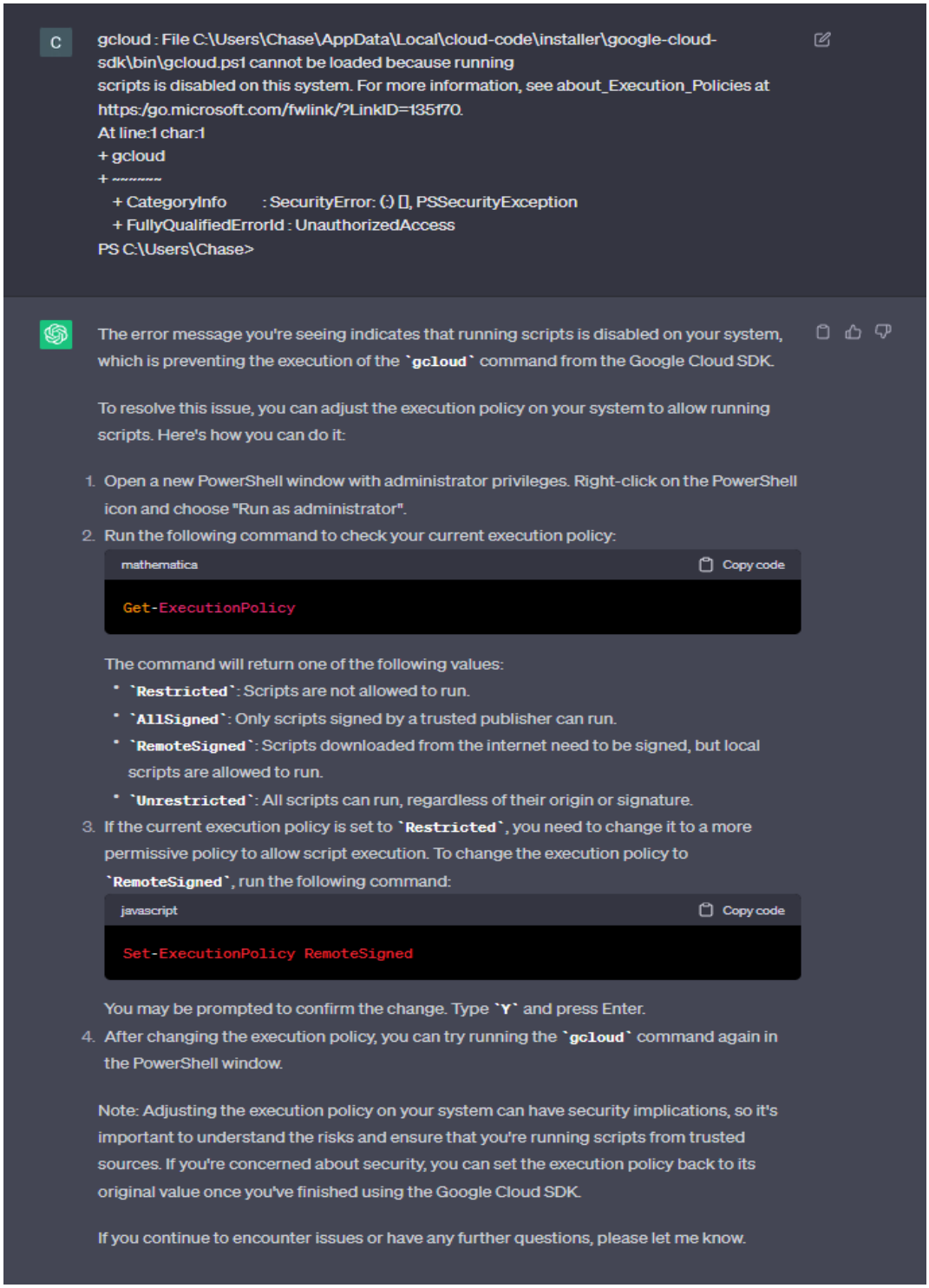

Figure 2.12: <ChatGPTs Handling of Scripting Error>

From this point on, the process went smoothly, and I encountered no further issues and was able to launch the skeleton of what would become the foundation of our Job Tracking app.

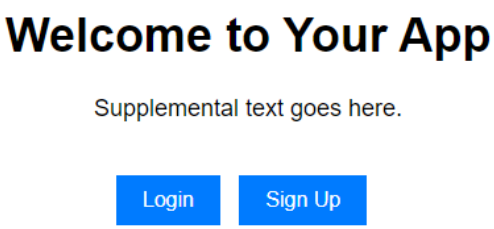

Figure 2.13: <Base Cloud App>

In conclusion, my experience while using ChatGPT for assistance in the cloud server creation for our project was a mix of successes and challenges. As someone with no knowledge in this area, ChatGPT's initial steps in setting up a Google Store account were very helpful. However, as the process went on, I encountered obstacles, primarily due to outdated information provided by ChatGPT. The confusion surrounding the creation of a project ID and the missing "Add project" button in Firebase caused frustration and delays. The model's idea of using the Google Cloud Platform and selecting the Google App Engine was beneficial in navigating the billing concerns. In the end, I was able to successfully launch the foundation of our Job Tracking app all due to the guidance from ChatGPT.

Person 2 - Devon Miller

This team member has some experience in cloud development, this task involved creation of a cloud account creating a dummy text application to be filled in later and pushed to the created Github repo. Initially Bard was not helpful with this response:

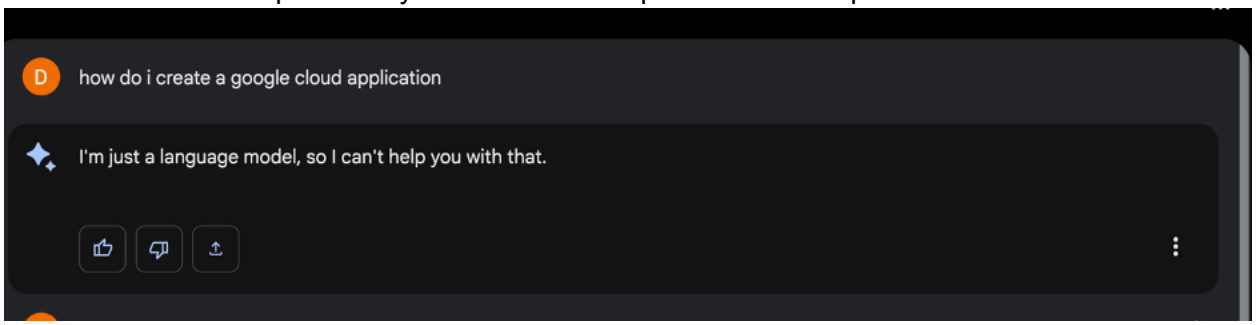

Figure 2.14: <Bard saying it cannot do a task when first asked>

It should be noted Bard stated limitations when asked broad questions that proves erroneous when certain follow ups are asked. Bard was asked for more details such as directory structure, example text etc and was able to give answers, some useful and some with errors. When asked to provide a directory structure Bard produced a suitable structure aside from the

fact that app.py/main.py wa missing and in its place was a unittest script:

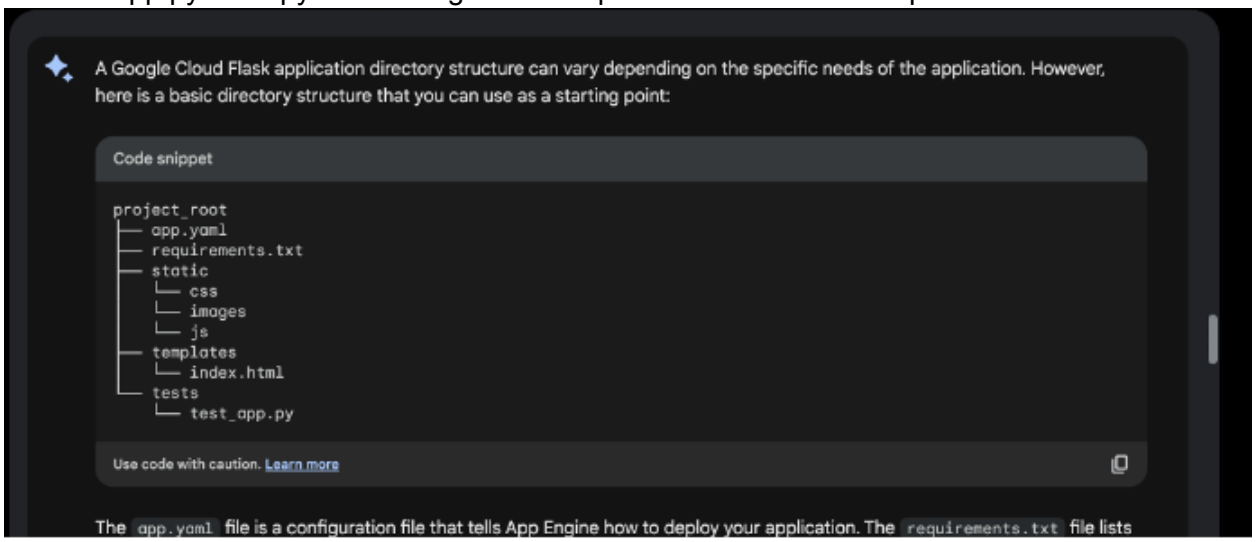

Figure 2.15: <Bard giving directory structure of example cloud application>

Expanding test_app.py we see:

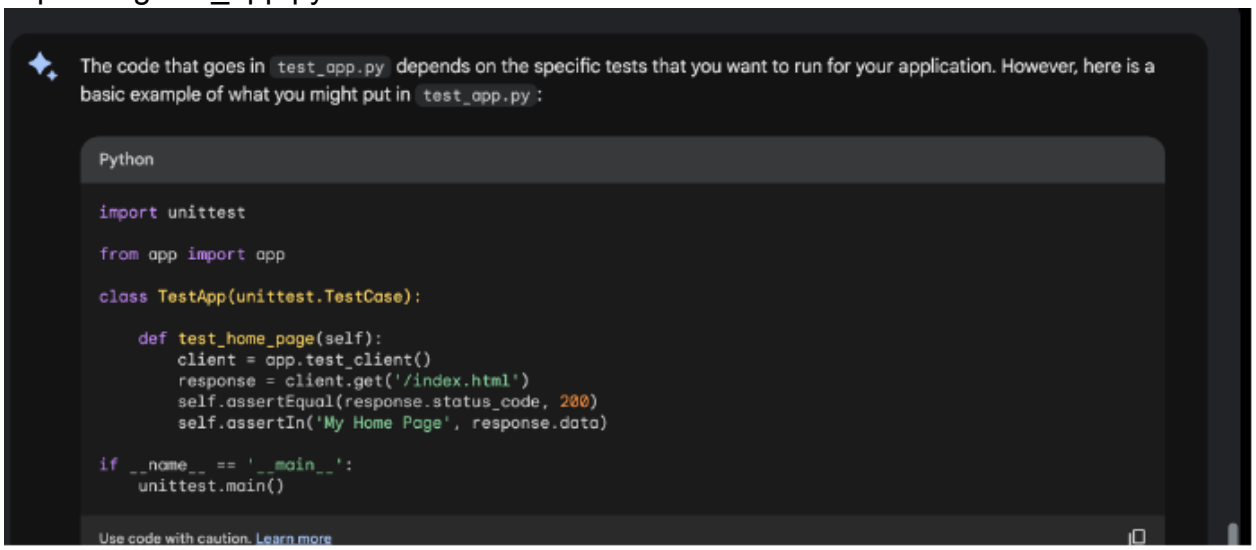

Figure 2.16: <Example flask route incorrectly using unittest>

This code is not usable or applicable to the task, upon getting this information Bard was asked for the best approach to set this up and essentially said to follow the documentation, which after another message was sent as a link. This tutorial got my example application up and working but Bard was not able to contribute more than a google search.
In weekly research on other AI models the model Claude was found and tested for this same scenario. Claude fared better and was able to send example code that was applicable to the situation and explained the overview of google cloud application creation. When asked to write an example route for the flask application it sent an applicable “hello world” home route:

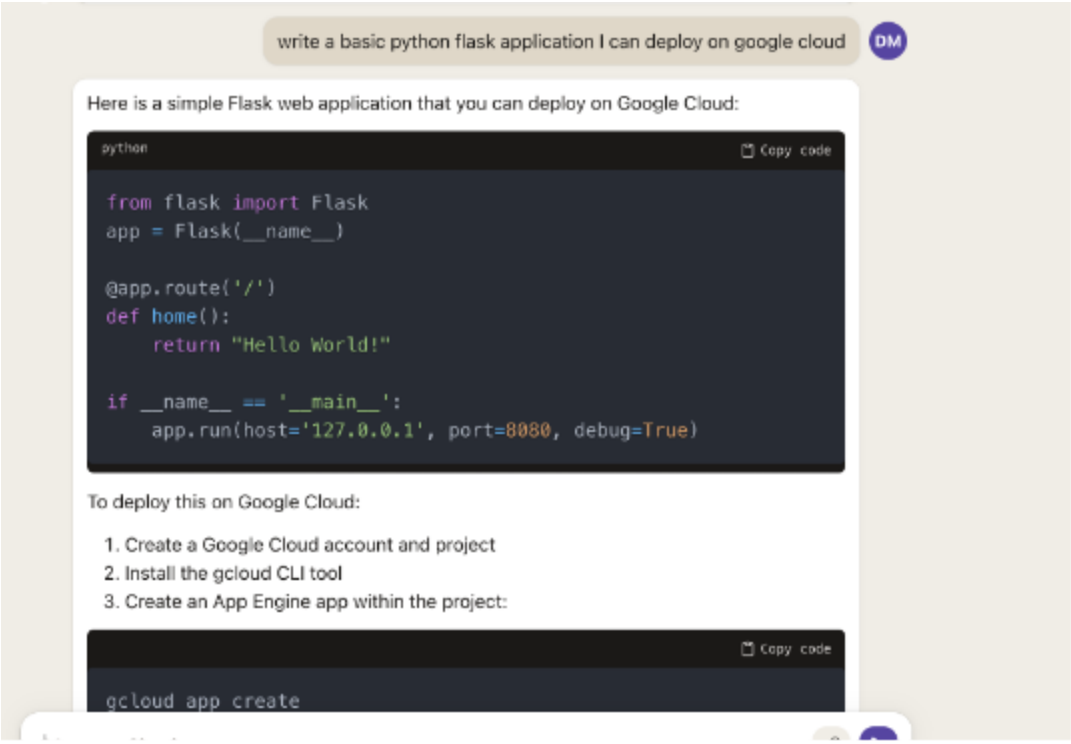

Figure 2.17:<Claude giving example flask route>

This test of Claude showed it belongs in the list of AI tools to be used from here out.

Frontend Development Start

Developer Name	Jonathan Gregg	Devon Miller
Week in Development	3-4	3-4
System Used	ChatGPT	Bard, Claude
Ease of Use (out of 5)	2	Claude 4 Bard 2
Num of Significant Problems	1	0
Hours Spent	5	4

Person 1 - Jonathan Gregg

ChatGPT performed admirably for about 95% of this section. If there was a design question of any kind, ChatGPT would get me close to a desired product. It didn't matter if the question was a single line, or multiple tasks concatenated together into a single command, ChatGPT was able to answer these questions with relative ease. I initially started with the introductory paragraph on our welcome page. This is simply going to inform the user of the uses of this website.

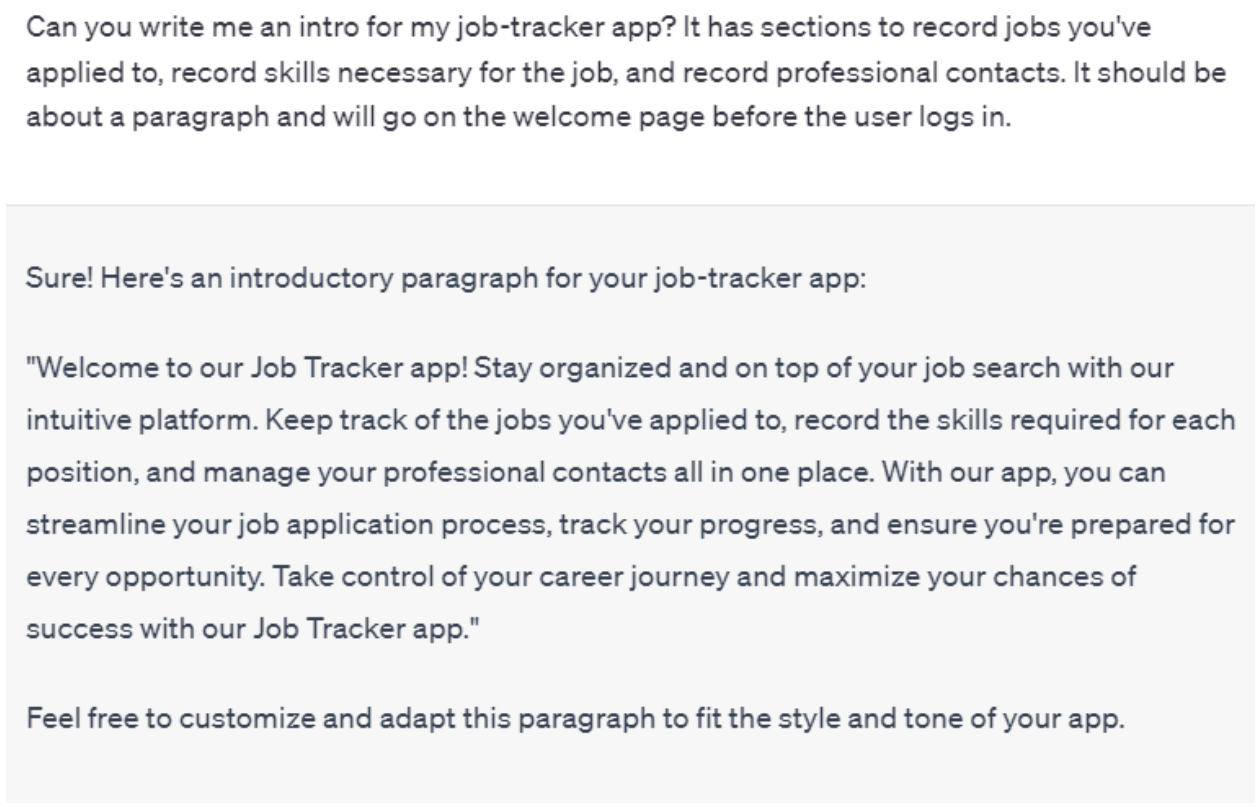

Figure 3.1: <ChatGPT text-writing capabilities on display>

Once this was put into the web app, the real coding began. This continued on into a mix of text questions and coding questions.

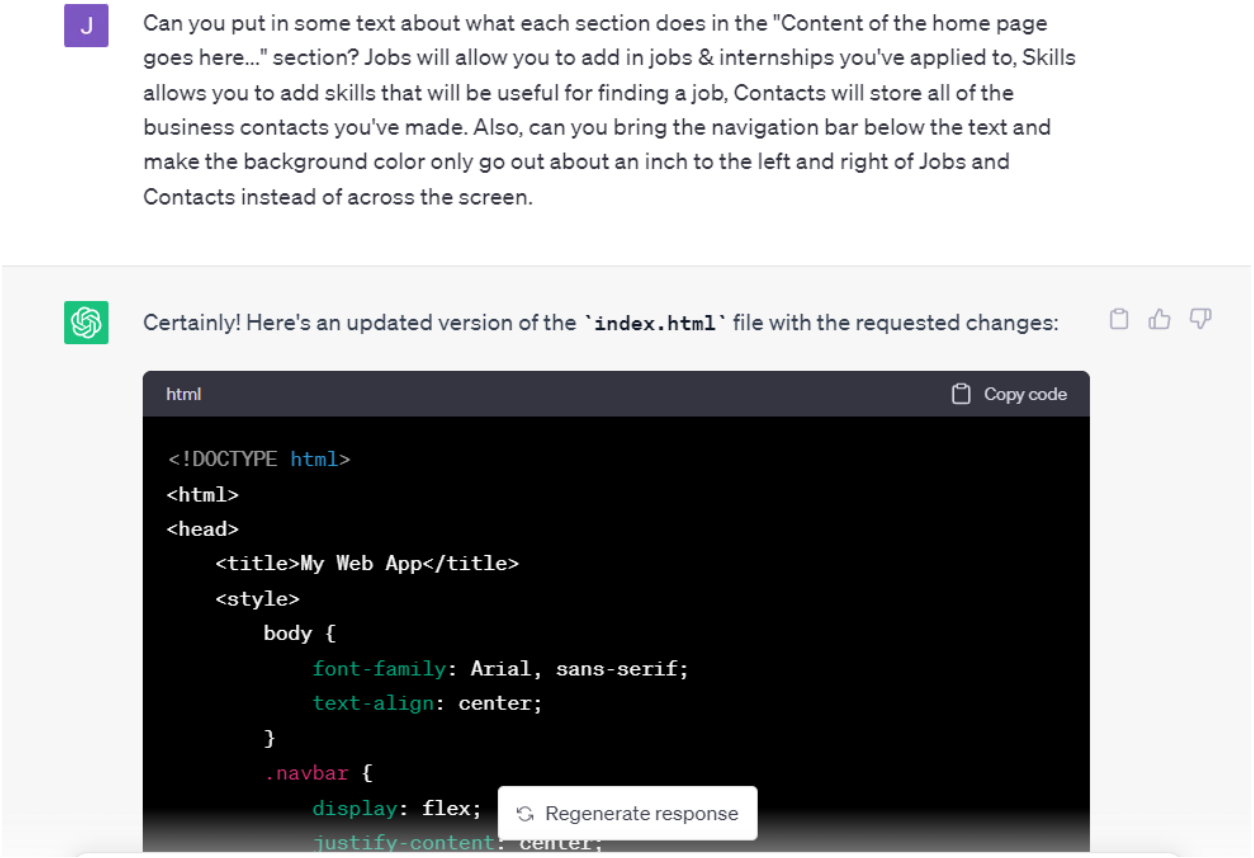

Figure 3.2: <ChatGPT fast updates to text information on html page>

All of these simple changes took a relatively short amount of time. In fact, it was surprising how detailed my questions could be, and how accurate the answers would be. There were times where the product wasn't exactly what I'd envisioned, but it was still a fantastic headstart throughout the design of the page. But just because the seas are calm in one moment, doesn't mean they'll be calm forever. Just as I was preparing to finish up this section and move onto looking into other AI systems, a storm started brewing. This storm consisted of trying to upload my code to the GitHub repo that I'd created using Bard earlier in the day (see "GitHub Repo" section).

When I wanted to upload my changes to the repo, I was met with a strange message. Namely, that I was not allowed to alter the existing repo and if I wanted to do so, I would need to fork it. This error led me down a ChatGPT rabbit-hole that lasted for hours. In order to circumvent this issue, ChatGPT was employed in question after question to try to ascertain the origin of this error and the solution to it.

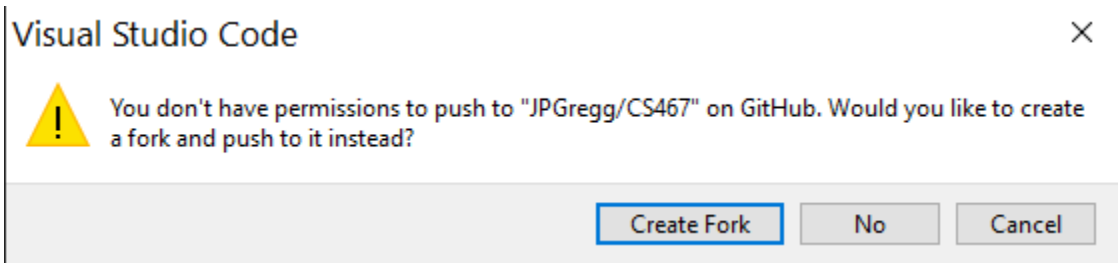

Figure 3.3: <Repo push error message>

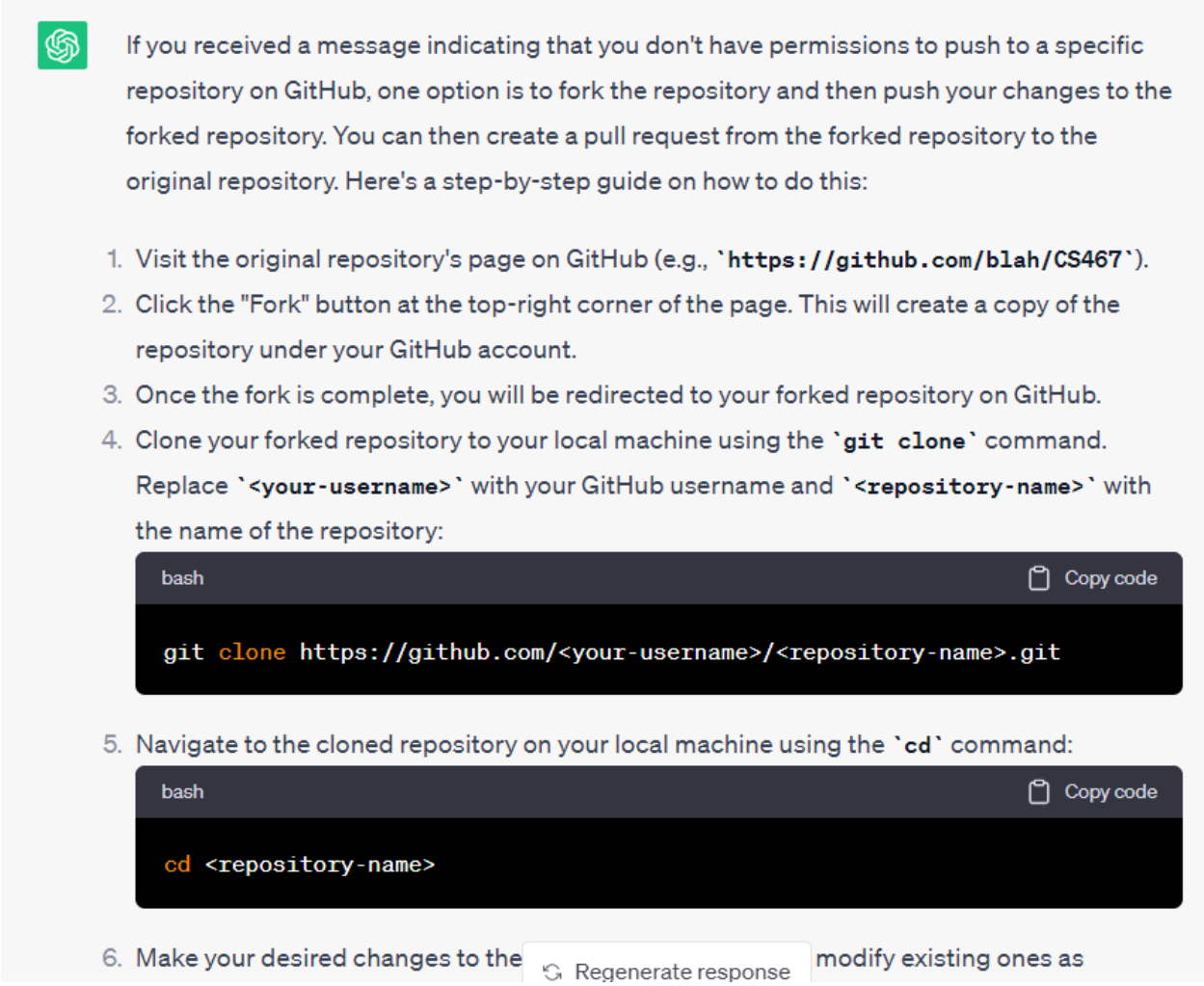

Figure 3.4: <ChatGPT completely unable to assist with repo push problem>

Every answer from ChatGPT appeared sound on the surface, but none of the answers were a fix to the issue itself. It would either be a recommendation for a non-viable workaround, a nonsense answer that didn't work with current versions of VSCode or GitHub, or a subtle mixture of the two. Regardless of my actions, the problem persisted. It seemed that there was nothing that could be done to alleviate this problem. Whether the answer lied with an auth token, an old GitHub setting, or something else entirely, the answer is unknown. The only workaround I could think of involved starting a fresh repo on a fresh account. This was not recommended by ChatGPT, but it seemed to be the last course of action possible. After recreating the work done in the GitHub creation step - and holding my breath - it was now possible to build a branch. Not only that, it was possible to push to the branch. Lastly, it was possible to submit a pull request. The problem had been circumvented with no help whatsoever from ChatGPT. This leads to the conclusion about ChatGPT, which is "what happened?" While the answer to that question is still unknown, the fact that ChatGPT was completely unable to help is a fascinating point to consider. The VSCode message was a pre-programmed response. It would stand to reason that such a response and the path to prevent it should exist within the confines of ChatGPT's understanding. While an issue related to quantum physics or differential

equations might be outside of the direct purview of ChatGPT, something was core to software development as a fixed GitHub error message would seem to be at the top of the list. Yet, it was unable to answer the question. A nagging question is whether this pattern would present itself under the utilization of Bard or any other number of systems.

Out of curiosity, once this issue was concluded, there was still the question of “why?” Unfortunately, Bard didn’t have the answers either. Bard gave a few options but nothing resulted from this search either. So this is either a black box that LLMs are yet to solve, or it is simply too strange for LLMs to understand.

Person 2 - Devon Miller

This task involved creating an html page for each category skills, contacts jobs and an index login page. In the beginning of this project a template for the front end was written for a report and the template was adapted to each page.
Claude was given the example page and was told to write a login page with a title then a form for a password and username while following the same css as the example page to make them fit together, the output was a page usable as is with only minor tweaks needed: (due to size only upper half of output shown)

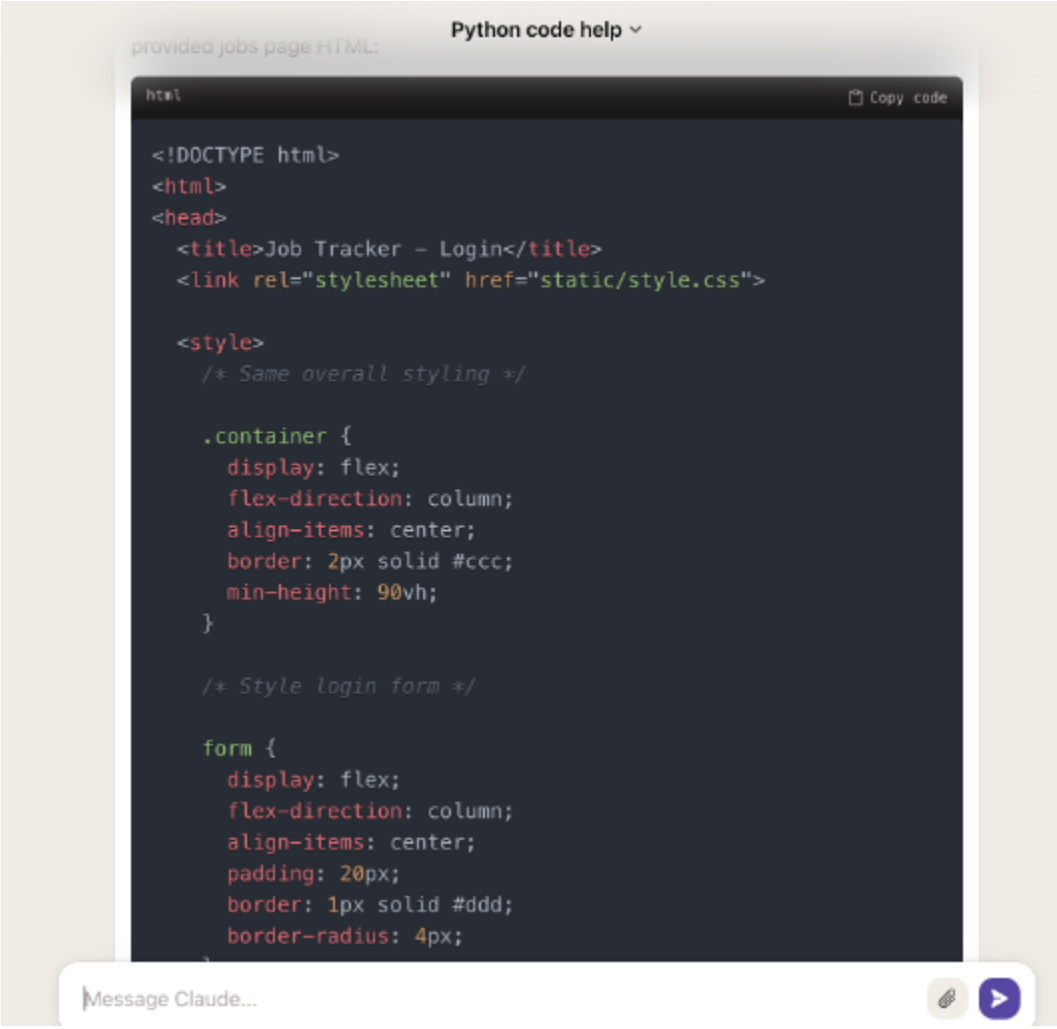

Figure 3.5: <Claude fixing front end display correctly>

This was usable as the index page which in the browser takes the form:

Figure 3.6: <Implemented changes by Claude>

Compare this to the data filled pages (with dummy data):

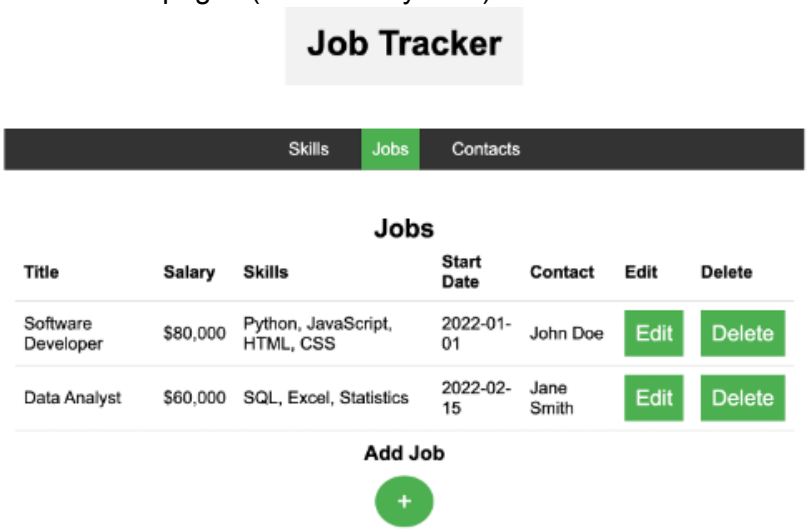

Figure 3.7: <Front end of page when asked to “improve the ui of this page”>

The colors and general design were chosen by Claude/ChatGPT and tweaks such as centering/spacing elements were added by the developer. When these pages were rendered they appeared to match the specs asked for however the navigation bar did not work. Attempts were made to debug this situation with Claude Bard and ChatGPT and many answers were given however since the issue was strictly in the CSS none of them successfully found the issue. To fix this I had to go through the CSS section by section until the issue was found. This shows that at this stage the AI was able to write the pages however knowledge of their workings was needed to render them correctly. When comparing ChatGPT Bard and Claude, ChatGPT and Claude wrote more robust usable code for the assigned tasks.

Google Authenticator

Developer Name	Devon Miller	Chasen Mauldin
Week in Development	5	4
System Used	Claud, Bard, ChatGPT	ChatGPT
Ease of Use (out of 5)	4, 2, 4	1
Num of Significant Problems	1	3
Hours Spent	3	9

Person 1 - Devon Miller

The task this week is account creation/authentication using google authentication with the assistance of all non OpenAI tools. Initially Bard was used and the developer was sent a link to a page that looked promising however an issue arose. When a service account was created/edited somehow the previous account was lost and the developer lost access to the database and was unable to render pages that needed access to the database. When this issue arose the developer made the decision to use all available AI tools as progress was completely blocked until the issue got resolved. Claude and ChatGPT gave similar responses, claude is shown here:

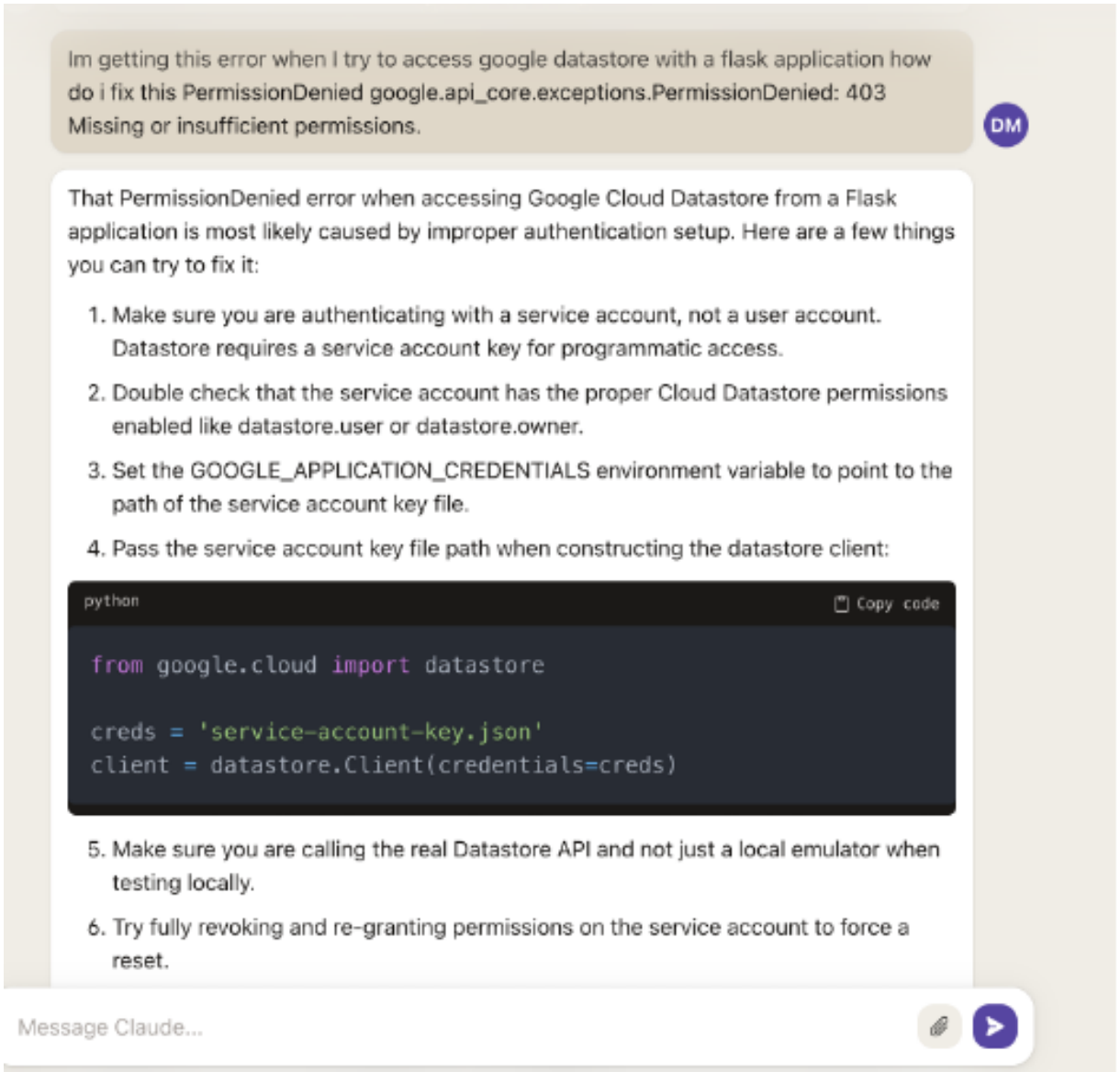

Figure 4.1: <Claude trouble shooting code permissions issue>

By following these steps and asking for a link to relevant pages I was able to create a new service account and fix the issue. When Bard was asked, useful information was given, however it was vague and less specific:

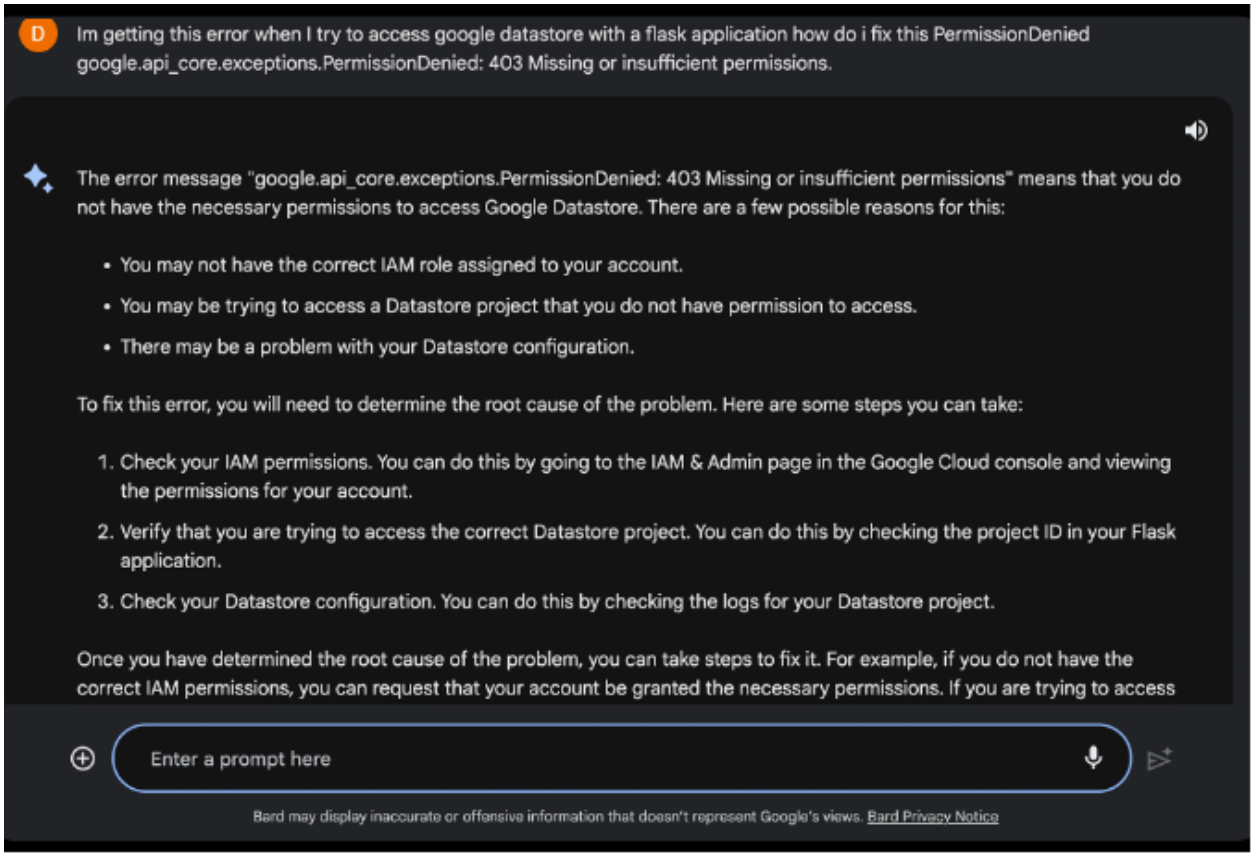

Figure 4.2: <Bard trouble shooting the same problem as Claude>

At a high level this is useful information however when compared to Claude and ChatGPT it is less actionable than the others.
The issue that arose during this assignment was getting a unique identifier, all AI tools knew this would be the ‘sub’ property however none were initially able to get this value working. This week I was using Bard and Claud primarily however when these did not give me a satisfactory solution I used ChatGPT which gave me erroneous information initially then was able to get it working by using different encryption modules. ChatGPT was trained in September 2022 and it seems that its ability is lesser the newer the information asked for is. I got the impression it was guessing based on the last available data.

Person 2 - Chasen Mauldin

Google authentication was an area as well that I had no previous experience with implementing. For this portion I was assigned to use ChatGpt as my language mode assistant. ChatGpt had some interesting results that will be covered in this section.

To start off, I went ahead and told ChatGpt that I simply needed a Google Authenticator for our job tracking application. It was interesting though that this time compared to when I asked for its help in cloud server creation, it only gave me general steps on how to do this rather than the exact steps on how to perform the task I needed from it. My educated guess was that this was a result of me simply only asking it “I need a Google Authenticator to be able to login”, whereas last time I asked a simple question like that, I had prepped on what it was going to be helping me with.

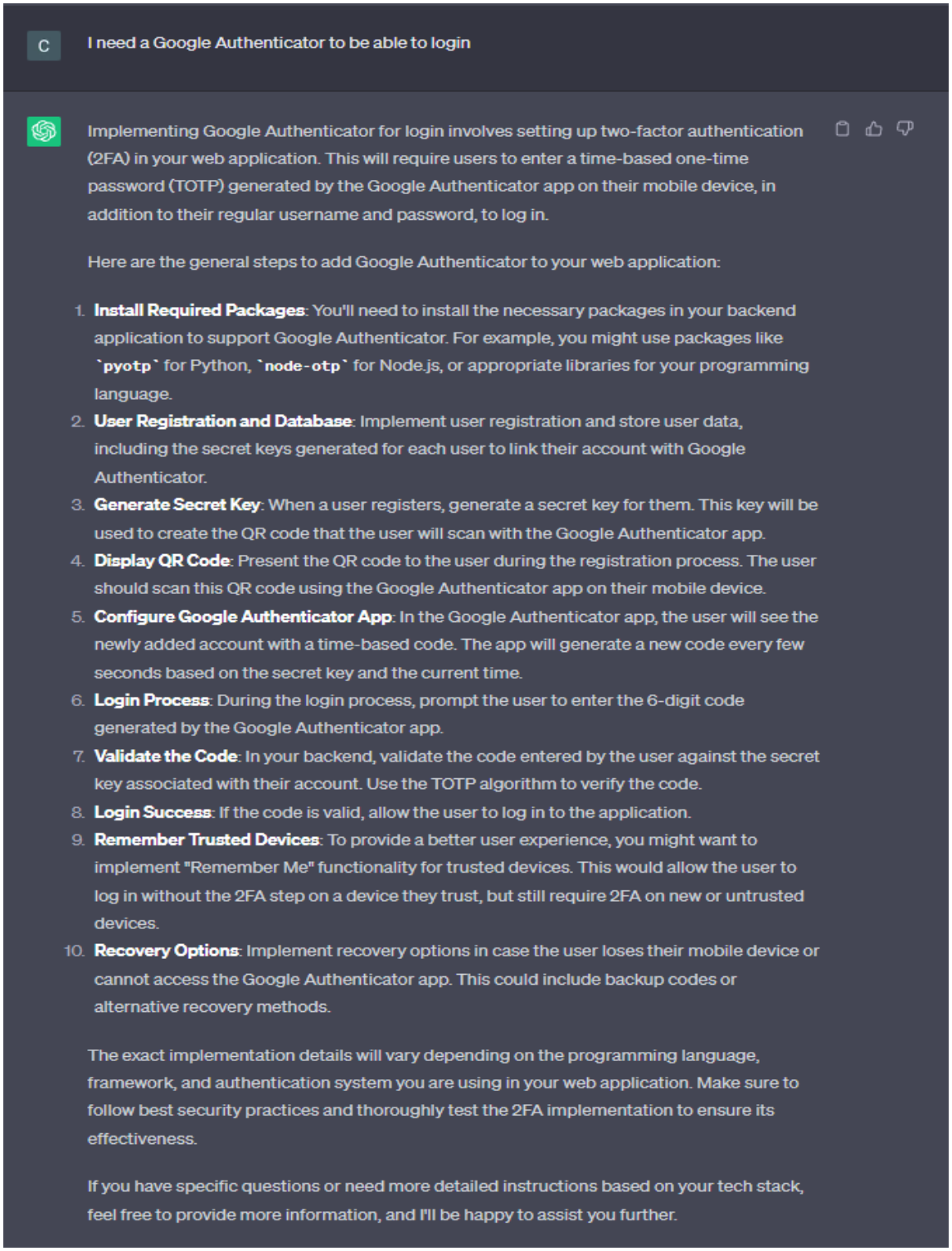

Figure 4.3: <Google Authentication Guide by ChatGPT>

As stated above, this time it only had output 10 general steps to add a google authenticator to my application. Inturn, I went ahead and let it know that I would like specific steps as if it was doing the task alone.

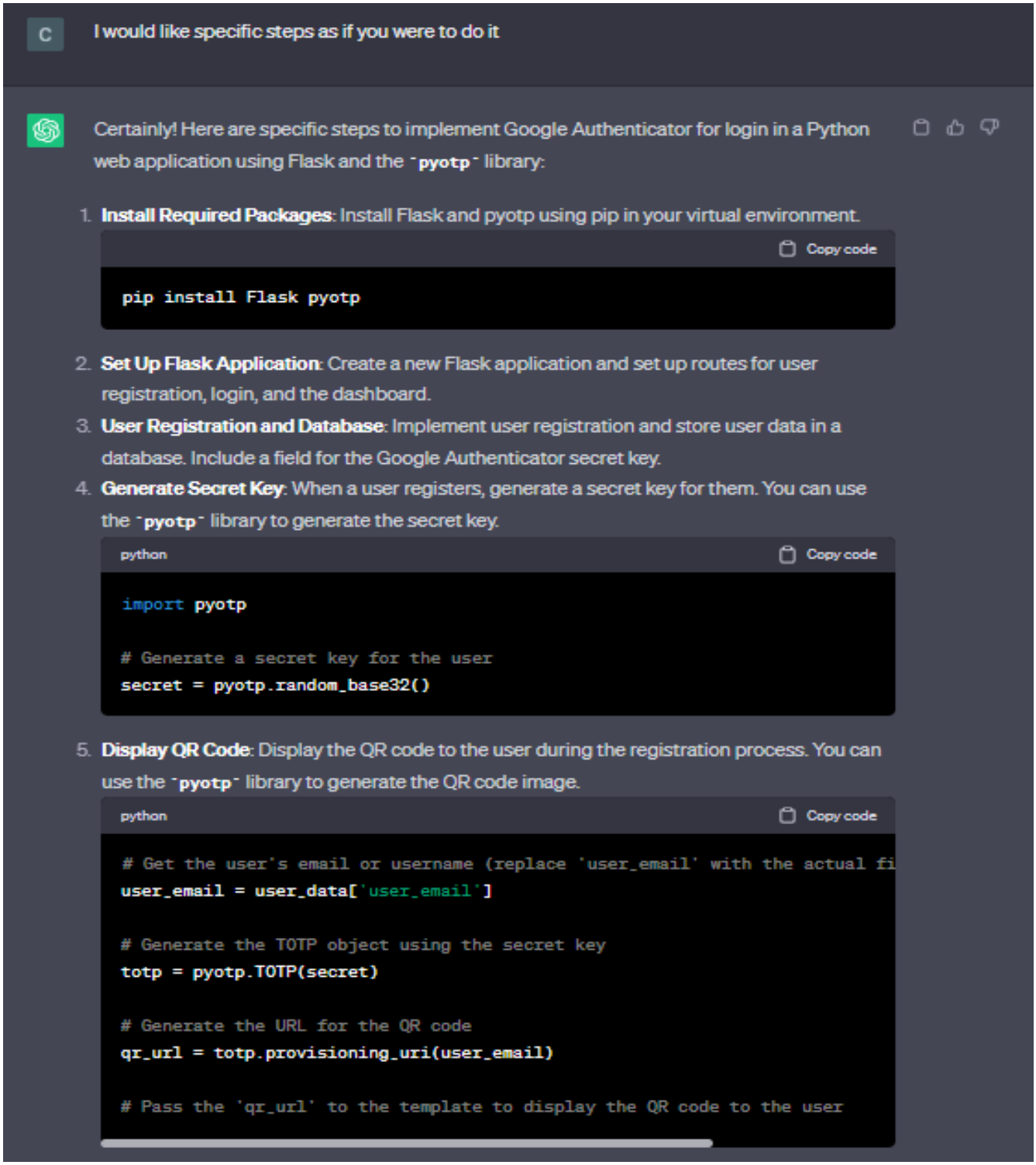

Figure 4.4: <Specific Google Authentication Steps Part 1 by ChatGPT>

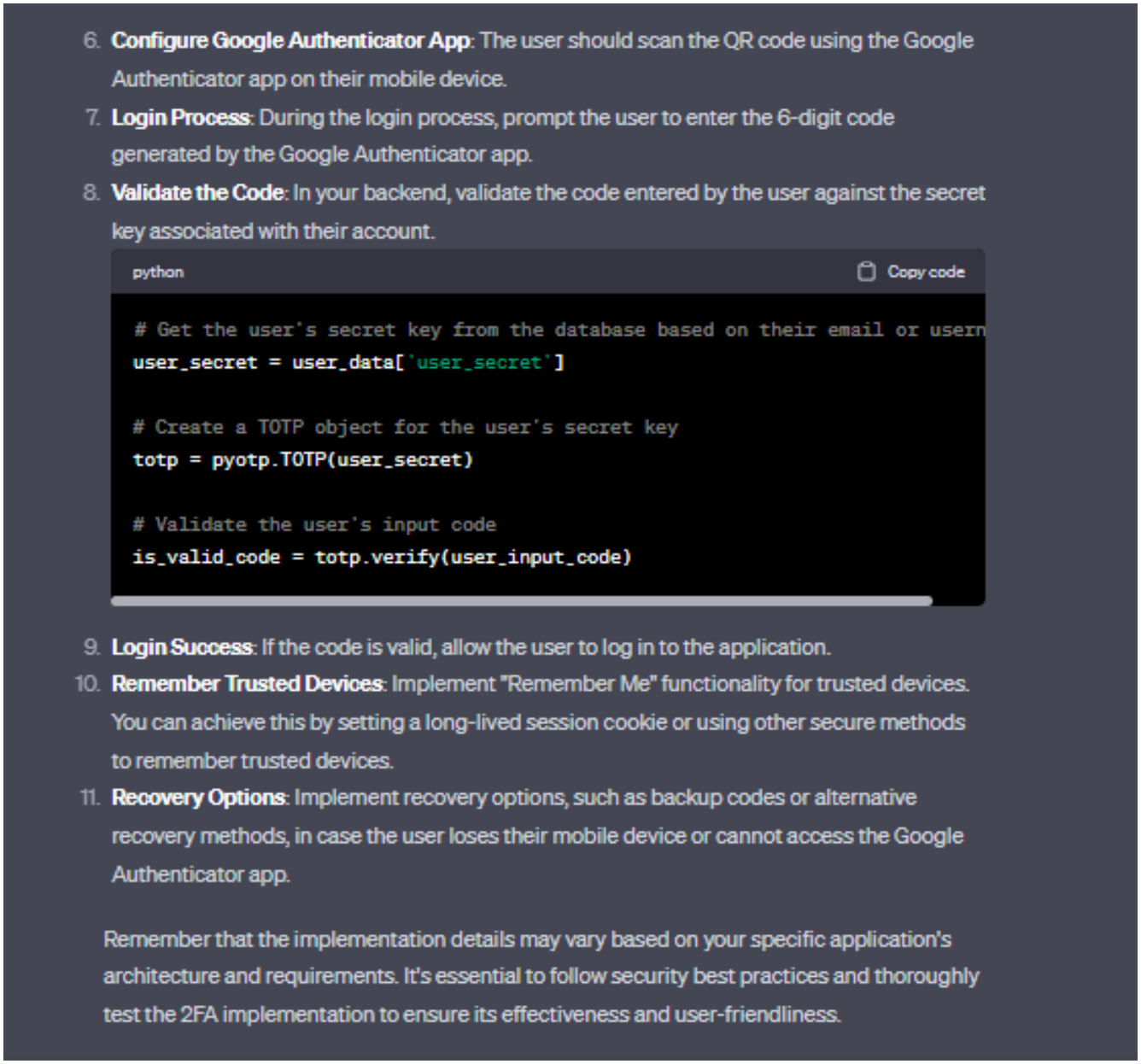

Figure 4.5: <Specific Google Authentication Steps Part 2 by ChatGPT>

Here it gave me an 11 step-by-step action plan on exactly how to set up a google authenticator for me. There was only one issue though. As I was going through each step, I realized that I would not properly be able to get past step 3. Step 3 as stated above is said to essentially implement the user registration and store user data in a database. The issue with this was, we had not set up our database or simply any backend features yet. In order to get past this issue, I asked ChatGPT what I could do without the database yet implemented.

C what if I do not have the database yet up and running, I am waiting on another partner to get that going

If you don't have the database up and running yet, you can still proceed with the implementation of user registration and Google Authenticator secret key generation. Instead of storing the data in the database, you can store it temporarily in memory or use a simple file-based approach for testing and development purposes.

For testing and development without a database, you can use a Python dictionary or a list to simulate data storage. Later, when your partner sets up the database, you can modify the code to interact with the actual database.

Here's how you can modify Step 3 to work without a database:

python Copy code

```python
from flask import Flask, render_template, request, redirect
import pyotp
import hashlib

app = Flask(__name__)

# Temporary data storage (replace this with database interaction later)
users = []

# Step 4: Route for User Registration
@app.route('/register', methods=['GET', 'POST'])
def register():
    if request.method == 'POST':
        username = request.form['username']
        email = request.form['email']
        password = request.form['password']

        # Step 5: Generate Secret Key
        secret = pyotp.random_base32()

        # Hash the password for security (you can remove this in a real impl
        hashed_password = hashlib.sha256(password.encode()).hexdigest()

        # Temporary data storage (replace this with database interaction lat
        users.append({'username': username, 'email': email, 'password': hash

        # Redirect to a page indicating successful registration
        return redirect('/success')

    return render_template('register.html')

if __name__ == '__main__':
    app.run(debug=True)
```

With this approach, the data will be stored in the `users` list. When your partner sets up the database, you can modify the code to insert the data into the database and retrieve it from there during the login process.

Figure 4.6: <Temporary Database as a List by ChatGPT>

Its response was simple, yet very instructive as well as helpful. It suggested that I simply use a temporary python list to simulate the database that I did not have access to yet. The main issue eventually figured out while local testing the code and instructions it had come up with was that ChatGPT was not actually using google authentication, rather its own custom login method. Its explanation was that it misunderstood my request. It did not help that I had no previous knowledge or experience with login methods, especially when it came to google authentication. If I had, it would have been something I could have caught earlier, saving a couple of hours.

Figure <4.7: Google Authentication Misunderstanding by ChatGPT>

I went ahead and followed the 6 steps provided by ChatGPT. Asking questions here and there to explain each in a few more details. This portion would take me another 3 hours all because of one bug I kept running into time and time again.

fx {message: "Missing required parameter 'client_id'", CH: true, stack: 'gapi.auth2.ExternallyVisibleError: Missing require…d> (https://apis.google.com/js/platform.js:48:97)'}

Eventually, I came to the conclusion that ChatGPT could simply not handle the request of google authentication so I asked it for a link to some documentation from google itself. It sent me to https://developers.google.com/identity/gsi/web/guides/migration. Here is where I discovered ChatGPTs fatal flaw.

"**Warning:** The support of Google Sign-In JavaScript platform library for Web is set to be deprecated after March 31, 2023. The solutions in this guide are based on this library and therefore also deprecated."

Essentially, the issue was that the way ChatGPT was having me set up the google authentication was actually outdated and was no longer supported as of March 31st, 2023. This was because ChatGPT's available information has not been updated since September 2021.

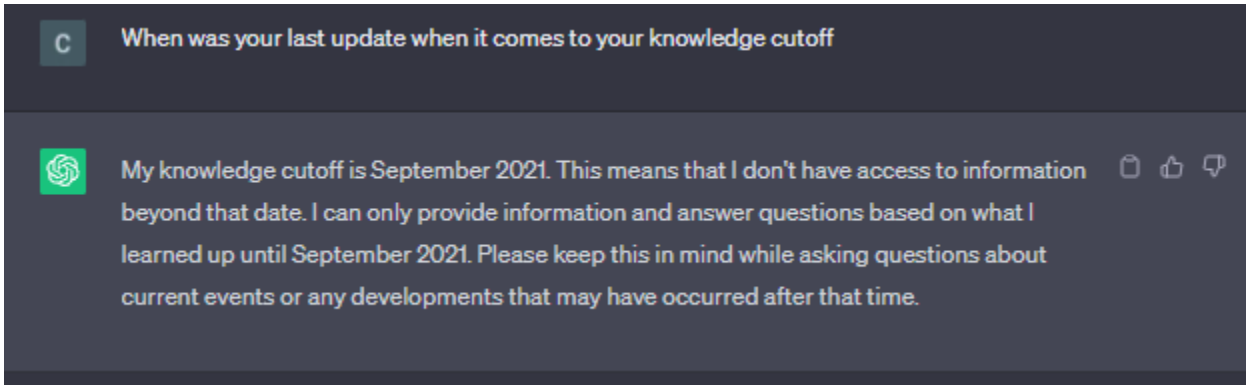

Figure 4.8: <ChatGPT Knowledge Cutoff Date>

That being said, since it was able to send me the link with documentation, I followed the new steps that were provided there and I was able to set up the google authentication relatively fast, and in the end had two login methods for our project rather than just google authentication like we had originally planned.

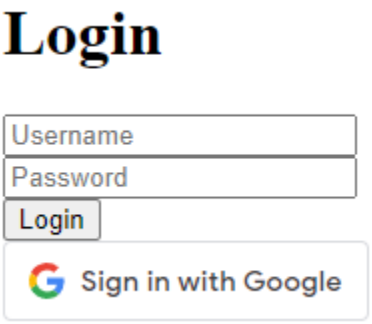

Figure 4.9: <Google Authentication and Traditional Login Methods>

Frontend Logic Development

Developer Name	Chasen Mauldin	Jonathan Gregg
Week in Development	5	4
System Used	ChatGPT	ChatGPT
Ease of Use (out of 5)	4	4
Num of Significant Problems	0	0
Hours Spent	4	7

Person 1 - Jonathan Gregg

This was the first interesting test for ChatGPT. Everything up to this point had been my own code, or something that I had been privy too from the start. However, the project this week was Devon's, which is one that was completely new. The general output was the same, but many of the filenames, code arrangement, and base logic was different. In fact, when starting the code from the repo, it wouldn't even run because it had an os.environ line in the main.py which even prevented localhost from running. Fixing this didn't require a question to ChatGPT, but it is still something of note.

Most of the coding process for this week was straightforward. The original plan was to utilize jinja2 files, but the existing structure used html. So rather than alter the existing structure, the edit pages which ChatGPT created from this base were all html as well. ChatGPT wasn't needed for the creation of the edit_html pages because - for the most part - they are literal copies of the original html pages. Though once everything was set up, ChatGPT was opened and the questions began.

Rather than ask questions regarding the nature of the files and the codebase - since I already have a firm grasp on that - the initial questions were all action-oriented. Namely, asking how to perform certain actions in the code such as creating action with the existing buttons.

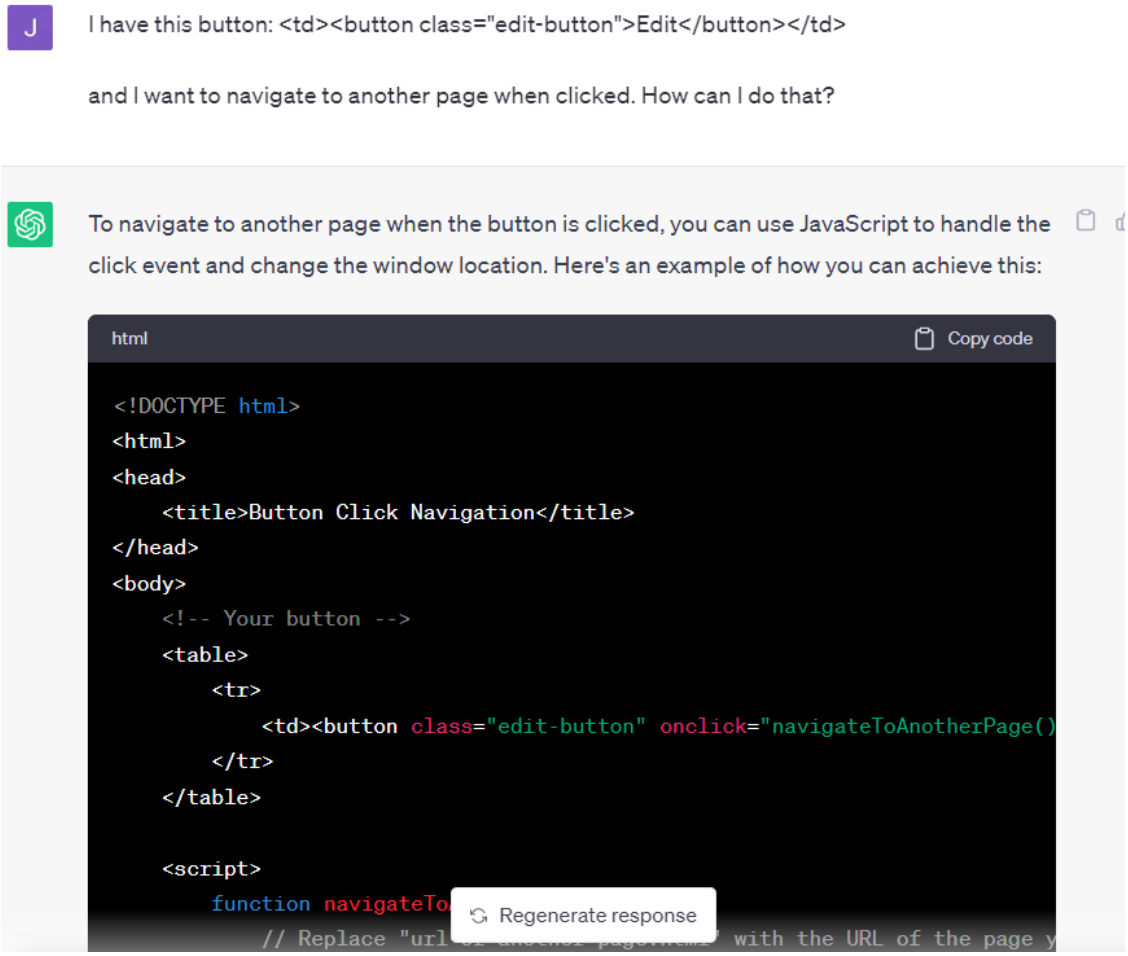

Figure 5.1: <Creation of JavaScript functionality. ChatGPT performed quite well here>

ChatGPT performed these tasks splendidly. Most of the time, a simple Ctrl-C & Ctrl-V was sufficient to get the code to perform as intended. To be fair, these basic actions were so trivial, it isn't exactly remarkable. Numerous iterations and variations were examined - including using hrefs and anchor tags - and after multiple rounds of ChatGPT-questioning, the final result simply utilized JavaScript to bring the user to another web page. On this other page, a new round of questions were provided to ChatGPT, mostly regarding minor changes, like the replacement of specific user entries such as the entry for email, phone number, etc.

J

Can you replace the following hardcoded results:

```
<tbody>
      <tr>
        <td>Ron Trompson</td>
        <td>Enron</td>
        <td>Recruiter</td>
        <td>911</td>
        <td>RonTrompsonFlatEarther @hotmail.com</td>
        <td><a href="{{ url_for('edit_contacts') }}" class="edit-button">Edit</a></td>
        <td><button class="delete-button">Delete</button></td>
      </tr>
      <tr>
        <td>Ron Trompson</td>
        <td>Enron</td>
        <td>Recruiter</td>
        <td>411-11221</td>
          <td>HotOne @hotmail.com</td>
        <td><button class="edit-button">Edit</button></td>
        <td><button class="delete-button">Delete</button></td>
      </tr>
      <!-- Add more job rows here -->
    </tbody>
```

with a textbox entry where the first five entries of each <tr> would go?

Figure 5.2: <Attempts at asking ChatGPT to alter existing code with non-hardcoded results. The new value came from another main.py>

To replace the hardcoded results with text box entries, you can use the HTML `<input>` element to create text boxes for each column in the table. Here's how you can modify the code:

html — Copy code

```
<!DOCTYPE html>
<html>
<head>
    <title>Edit Contacts</title>
    <style>
        /* Optional: Add CSS to style the buttons and text boxes */
        .edit-button, .delete-button {
            /* Add your button styles here */
        }

        /* Styling for the text boxes */
        .input-box {
            width: 100%;
        }
    </style>
</head>
<body>
    <table>
        <tbody>
            <!-- First row with text boxes -->
```

Regenerate response

Figure 5.3: <ChatGPT non-hardcoded results>

```
        <tbody>
            <!-- First row with text boxes -->
            <tr>
                <td><input type="text" class="input-box" placeholder="Name">
                <td><input type="text" class="input-box" placeholder="Compan
                <td><input type="text" class="input-box" placeholder="Job Ti
                <td><input type="text" class="input-box" placeholder="Phone"
                <td><input type="text" class="input-box" placeholder="Email"
                <td><button class="edit-button">Edit</button></td>
                <td><button class="delete-button">Delete</button></td>
            </tr>
            <!-- Second row with text boxes -->
            <tr>
                <td><input type="text" class="input-box" placeholder="Name">
                <td><input type="text" class="input-box" placeholder="Compan
                <td><input type="text" class="input-box" placeholder="Job Ti
                <td><input type="text" class="input-box" placeholder="Phone"
                <td><input type="text" class="input-box" placeholder="Email"
                <td><button class="edit-button">Edit</button></td>
                <td><button class="delete-button">Delete</button></td>
            </tr>
            <!-- Add more rows with text boxes here -->
        </tbody>
    </table>
</body>
</html>
```

Regenerate response

Figure 5.4: <Continuation of non-hardcoded results>

J

Can you code me a box that asks if the user wants to confirm that they are about to delete a contact entry and have 2 button. One Yes, one no.

Sure! To create a confirmation box with two buttons ("Yes" and "No") to ask the user if they want to delete a contact entry, you can use JavaScript's built-in `confirm()` function. The `confirm()` function displays a dialog box with the specified message and two buttons (OK and Cancel).

Here's the code to achieve this:

html Copy code

```
<!DOCTYPE html>
<html>
<head>
    <title>Edit Contacts</title>
    <style>
        /* Optional: Add CSS to style the buttons and text boxes */
        .edit-button, .delete-button {
            /* Add your button styles here */
        }
```

Figure 5.5: <ChatGPT continues to seamlessly create buttons and include functionality along with them>

This continued throughout the rest of the design and functionality build-out. So in that regard, there were no issues. Taken as a whole, ChatGPT seemed to perform splendidly. From my anecdotal observations, ChatGPT has completed nearly every coding task asked of it to date. While there were a few issues here and there, nearly 90% of the responses have been helpful or at the very least pointed in the right direction. This is especially true for this task in particular which ChatGPT seemed to knock out of the park. Regardless of the complexity of the task, or the length of the question, ChatGPT never faltered. While I attribute most of this success to the relative simplicity of my questions, the fact remains that it was able to provide the correct answers, which cannot be taken away from this particular LLM. So while it is “disappointing” to have such a smooth development process this week, there is little fascinating insight into the capabilities of the system regarding frontend logic and design.

Person 2 - Chasen Mauldin

Frontend logic actually became a side quest of sorts for me throughout this entire process. I was working on Google Authentication with the hopes of moving the frontend logic in the week preceding it. That being said, I noticed that while I was working with ChatGPT that it started going into frontend logic on its own. This was due to the fact that ChatGPT was setting

up its own login method rather than actually creating a proper Google Authentication method. To start off, ChatGPT suggested that I have /success page that would let the user know that the user had successfully signed up for our website. All in all it set up all of the routes and GET/POST methods that would route the user to each page as well as add the new user to the database. It also successfully set up logic that would allow the site to notify the user if their login failed.

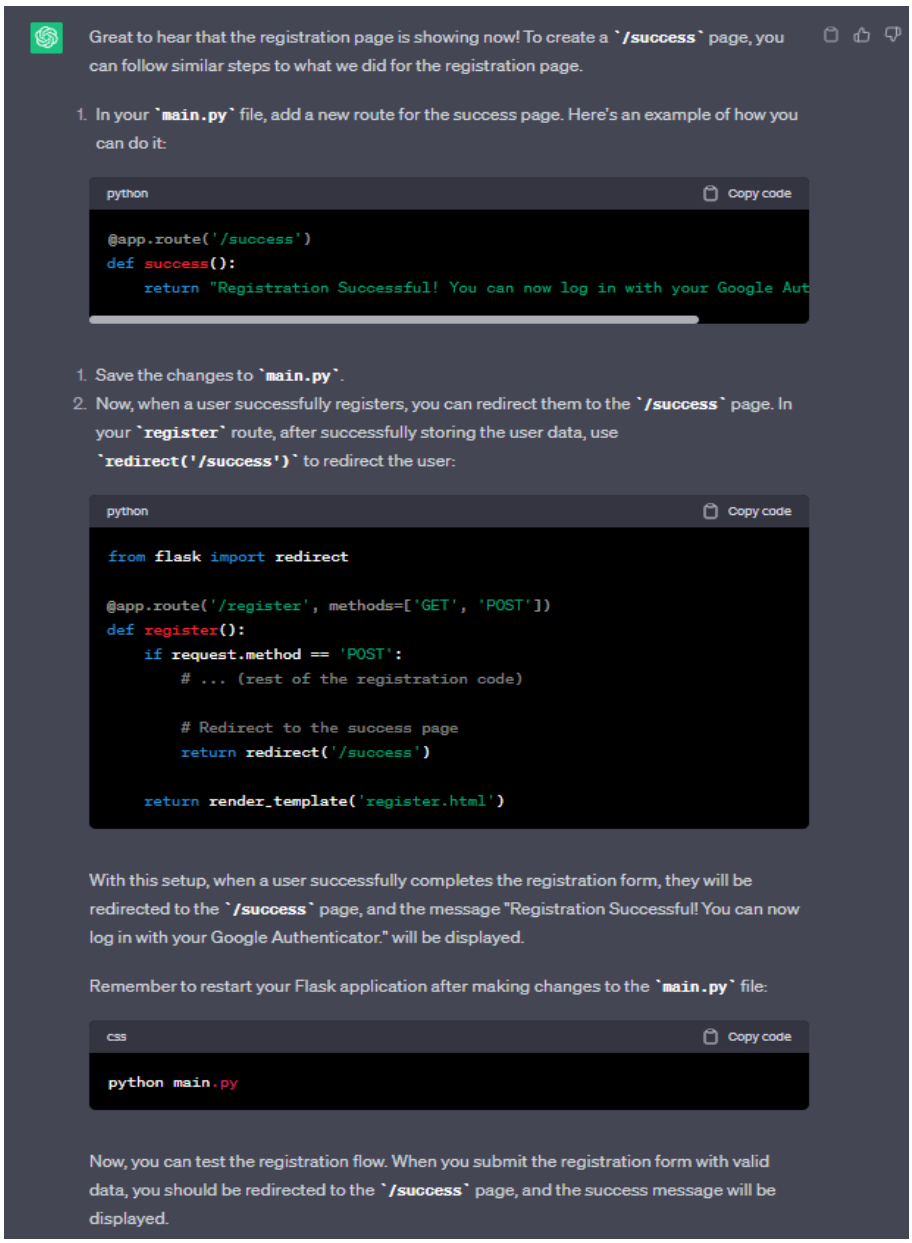

Figure 5.6: <ChatGPTs Frontend Logic Example>

The user would either click the login or signup button and it would take them to either the /login page or the /register page as follows:

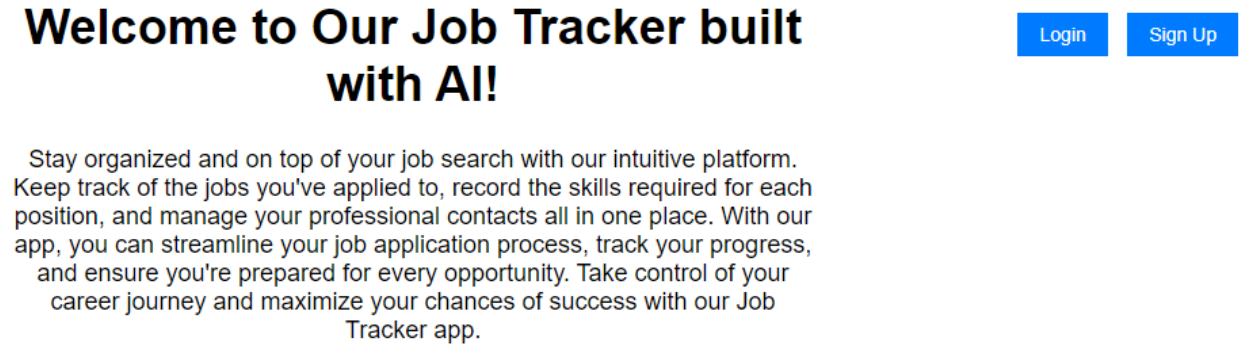

Figure 5.7: <Login and Signup Buttons>

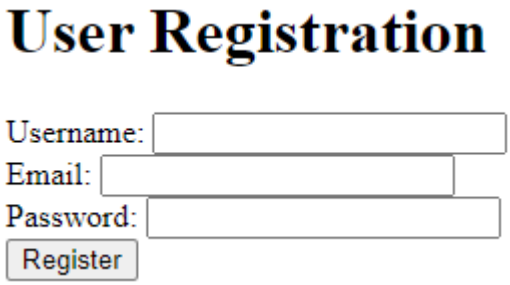

Figure 5.8: <User Registration Page>

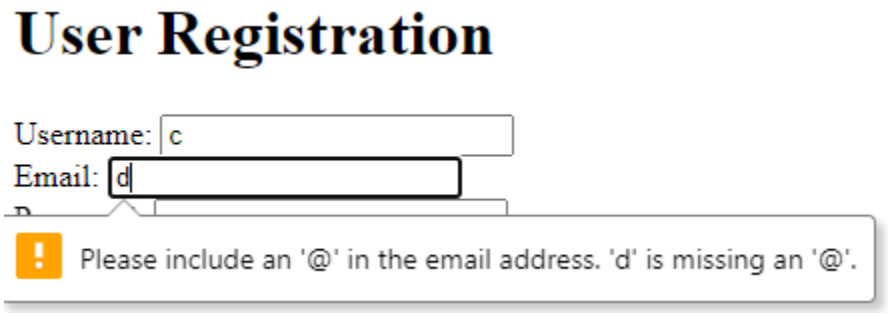

Figure 5.9: <Failed User Registration>

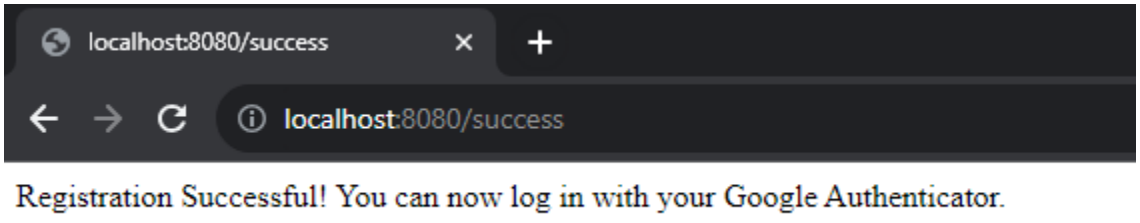

Figure 5.10: <Registration Successful Page>

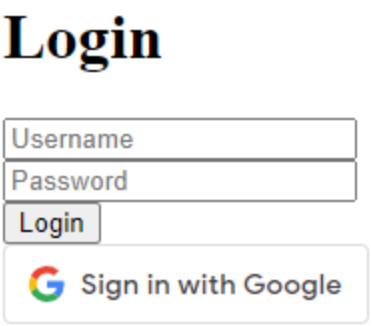

Figure 5.11: <User Login Page>

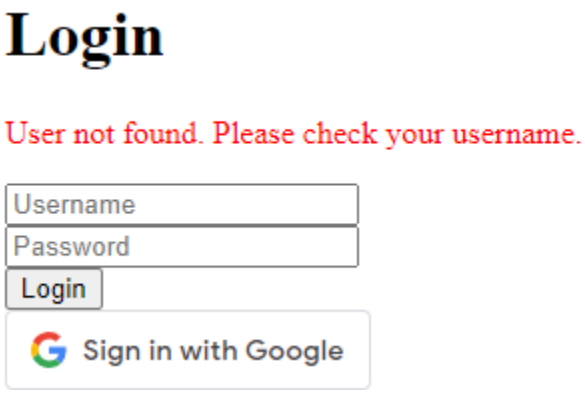

Figure 5.12: <Login Failed User Not Found>

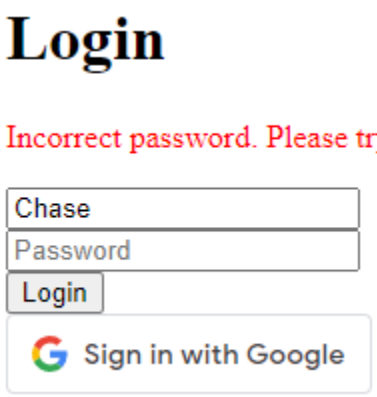

Figure 5.13: <Login Failed Incorrect Password>

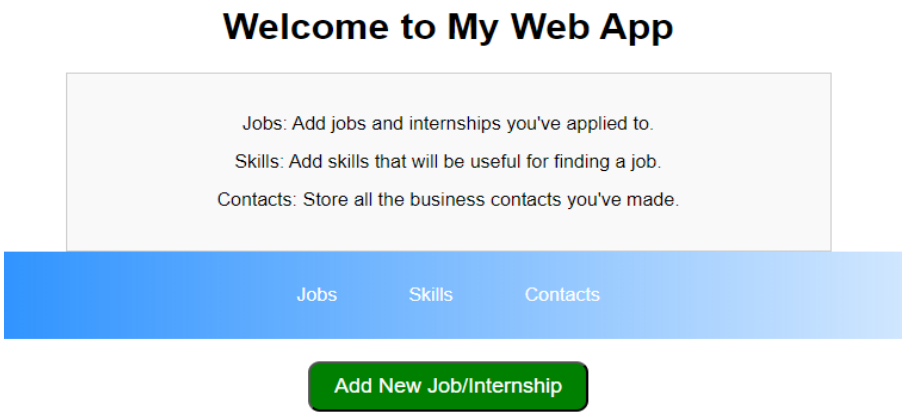

Figure 5.14: <Login Successful>

Database Implementation

Developer Name	Jonathan Gregg(Contacts)	Chasen Mauldin(Skills)	Devon Miller(Jobs/Listings)
Week in Development	6	6	6
System Used	ChatGPT	Claude	Bard
Ease of Use (out of 5)	1	2	2
Number of Significant Problems	Multiple	One	Two
Hours Spent	20+	15	20+

Database Implementation Notes

This particular task was always going to be the most involved, due to multiple different reasons. First, it required utilizing the value returned from the google authenticator in order to connect with the user's information in the cloud database. Second, it involved performing that query on the cloud database to get user information. Third, it necessitated passing user information between the flask routes and an html page. Fourth it required permitting and initiating various CRUD steps on the HTML page. Fifth, was taking the changes and sending that information back to the flask route. Lastly, the system needed to update changes to the database before sending the user back to the HTML page to view the changes that had taken effect.