### **AI Health Coach Risk Analysis**

The Health Coach project leverages AI to provide affordable and personalized nutrition guidance, utilizing models such as GPT-4o-mini and Claude-3.5-sonnet. Given the critical nature of nutrition-related recommendations, this project will be assessed using the framework outlined in Chapter 6 of *The AI Conundrum*. This framework considers three key risk factors: the need for precision, control over inputs, and the requirement for rationale.

#### **Need for Precision**

The effectiveness of the Health Coach application depends heavily on the accuracy of its dietary recommendations. Nutrition plays a direct role in an individual's well-being, meaning that errors in meal planning, dietary restrictions, or nutritional balance can have serious consequences. The experimental results showed a variance in performance between AI models, with OpenAI's ChatGPT demonstrating higher accuracy and readability than Anthropic's Claude. However, even the best-performing model may occasionally generate misleading or incomplete information, raising concerns about AI reliability in health-related applications. The project must assess whether these risks outweigh the benefits and whether alternative solutions could provide more consistent results.

#### **Control Over Inputs**

One of the main risk factors in AI-based applications is the ability to manage and control input data. The Health Coach application generates personalized meal plans based on user-provided details, such as height, weight, dietary restrictions, and medical conditions (e.g., anemia). The reliability of AI-generated advice is directly influenced by the quality and specificity of user inputs. However, users may provide incomplete or incorrect information, leading to inaccurate recommendations. Additionally, the AI model’s training data plays a crucial role in its ability to generate accurate meal plans. If the model has not been trained on diverse, up-to-date, and medically verified nutritional data, it may offer recommendations that are not optimally suited for specific health conditions. The extent to which the Health Coach project can control and verify these inputs is a crucial risk factor that must be continuously evaluated.

#### **Requirement for Rationale**

Since the Health Coach provides personalized dietary recommendations, users need to understand why specific meal plans or nutritional choices are suggested. Many AI models function as "black boxes," offering outputs without clear explanations. This lack of transparency can make it difficult for users to trust or verify AI-generated suggestions. While the Health Coach project partially addresses this by including general dietary advice (e.g., explaining how vitamin C enhances iron absorption for anemia patients), it does not provide full explanations of why specific meal combinations are recommended. The inability to justify recommendations clearly could reduce user confidence and limit the app’s effectiveness in promoting long-term adherence to dietary plans.

### **Reliance on AI Provider**

Another critical factor to analyze is the extent to which the Health Coach project depends on external AI providers. The project’s experiments compared OpenAI’s GPT-4o-mini and Anthropic’s Claude-3.5-sonnet, ultimately selecting ChatGPT due to its superior accuracy and personalization. This reliance on a third-party AI provider introduces several risks. First, the project is dependent on OpenAI’s continued support and development of its AI models. If OpenAI discontinues GPT-4o-mini or makes significant modifications that reduce its effectiveness for nutrition-related queries, the project may face disruptions. Second, AI models process user health-related information, which may fall under regulatory frameworks such as HIPAA (for U.S. users). Since OpenAI controls the AI model, we must be aware of potential compliance challenges and how data privacy is managed by the provider. Lastly, since the AI models are externally developed, there is limited control over their inner workings. While prompt engineering can improve output quality, the lack of direct model customization limits the ability to refine recommendations or correct inaccuracies beyond surface-level tweaks.

These challenges highlight the complexities of integrating AI into health-related applications. The reliance on a third-party AI provider introduces vulnerabilities that must be considered when assessing the project's long-term sustainability and effectiveness. Without control over the AI models, we could be faced with provider changes, pricing adjustments, and policy updates that could impact the application’s performance and reliability. The risks associated with external AI dependencies must be carefully weighed against the benefits of AI-powered customization and accessibility. If these risks become too great, the project may need to reconsider its reliance on external providers and explore alternative methods for delivering personalized health recommendations. Assessing these risks in an ongoing manner will be crucial to ensuring the success and sustainability of the Health Coach project.

**If the provider changed API**

One of the most significant risks of our AI-powered health coach app is its reliance on external AI providers like OpenAI’s GPT -4o-mini or Anthropic’s Claude-3.5-sonnet. If the provider decides to change the API or alter the output of their models, the Health coach could experience disruptions that affect both functionality and user trust. These disruptions could come from several areas, including API changes, modifications to the underlying models, and adjustments in prices or service terms.

The API serves as a bridge between the health coach application and the underlying AI models. If the API were to change or if OpenAI or Anthropic were to create significant modifications, it could disrupt the Health Coach’s ability to function effectively. Changes in API endpoints, data formats, or functionality might require the Health Coach team to adapt or rewrite portions of the application code. If this happens it could result in downtime or poor service quality, and could negatively impact the user's experience.

These changes could create a scenario where the Health Coach’s features are suddenly incompatible with the new API, forcing the project team to find a solution. Even if the Health Coach application could adapt to API changes, this process would likely require a lot of technical effort and could lead to temporary loss of features, broken user interfaces, or the inability to process user inputs correctly. API changes could affect the accuracy and consistency of the AI’s recommendation. If the AI model's behavior is modified by an update in the API or model design, users may notice that meal plans are less accurate or less personalized. This could create a reduction in user trust, as individuals may receive dietary recommendations that no longer meet their specific needs, such as the AI failing to provide adequate solutions for health conditions like anemia or diabetes.

Another major risk is the possibility of changes to the output generated by the AI models. Open AI or Anthropic may modify the model's behavior or adjust their training data, which could result to be less accurate, less personalized, or even potentially harmful recommendations. A model update might reduce the attention given to medical dietary guidelines, leading to meal plans that are poorly suited for individuals with certain health conditions. The book “Fairness and Machine Learning: Limitations and Opportunities” by Barocas, S., Hardt, M., & Narayanan, A, explores the challenges of fairness in machine learning, focusing on how AI models can be biased and the impact this can have in the areas like health care where decisions can affect individuals well-being. If the new model outputs lower-quality recommendations, the Health Coach would need to quickly evaluate the impact on user outcomes and possibly go back to an older version of the model or seek out a different AI provider.

During an output change, the Health Coach team would need to quickly determine whether the new model still meets the needs of their target audience. An example would be if the AI’s ability to recommend iron-rich foods for users with anemia decreases, it would directly reduce the functionality of the product. The changes could lead to customer satisfaction, increased support requests, and create reputational damage.

To reduce these risks, the Health Coach project should maintain a close relationship with the AI provider to stay updated on any upcoming changes and prepare for potential disruptions. They may want to implement the use of different AI models or have manual override features to ensure continuous service in case of unexpected API or model changes.

When using AI models for health-related recommendations, one of the biggest concerns is the handling of sensitive personal health data. If AI models process data such as user health conditions, there are legal and ethical considerations regarding data privacy. The Health Coach application would need to ensure that the AI provider follows HIPAA or other data privacy regulations. If OpenAI or Anthropic does not meet these standards or if they change their data policy, the Health Coach project could be exposed to legal and financial risks. The article “Artificial Intelligence and the Ethics of Health Data” by Binns, R. discusses the ethical considerations of using AI in health-related applications, including issues of data privacy, bias, and the reliability of AI-generated medical advice. If data security is compromised the Health Coach users could face the potential theft of sensitive health information, leading to serious violations and loss of trust.

The Health Coach project stands to offer personalized, affordable nutrition guidance through AI. The Health Coach must develop strong strategies to lower these risks, including maintaining flexibility to switch providers if necessary, ensuring compliance with data privacy laws, and regularly auditing AI recommendations for fairness and accuracy. By addressing these challenges, the Health Coach can help ensure that it remains a reliable, effective tool for users seeking personalized dietary advice. By doing this the Health Coach would have more users.