AllergyAware: Personalized Recipe Recommendations and Tailored Recipes Solutions for Food Sensitivities

Anneketh Vij *, **Ayan Bhowmick***, **Changhao Liu***, **Edward Shi***, **Rahul Nair***, **Theo Ho***
Department of Computer Science
University of Southern California
Los Angeles, CA 90007
{anneketh, abhowmic, celiu, epshi, ranair, teho}@usc.edu

Abstract

Our project aims to develop a personalized recipe recommendation and generation system using natural language processing techniques, with a focus on accommodating food allergies and promoting healthy-alternatives. By leveraging user preferences, dietary restrictions, available ingredients and kitchenware, we seek to create an AI-powered assistant that can generate user-tailored recipes. We will be experimenting with different approaches to combine a large recipe dataset with user reviews to train a model capable of understanding culinary preferences and food allergens, then generating appropriate recipes. We expect this system to provide users with personalized, safe cooking suggestions, while enhancing their culinary experiences.

1 Introduction

In this project, we will be generating allergy-friendly recipes based on user preferences, dietary restrictions, and available ingredients and kitchenware. Cooking is a universal activity, but finding the right recipe can be challenging, especially for individuals with food allergies. Our project aims to address this challenge by creating an AI-powered recipe generator that takes these factors into account, with a particular emphasis on allergen awareness. We'll use natural language processing (NLP) techniques to analyze recipe data and user reviews, allowing our model to understand the relationships between ingredients, cooking methods, user preferences, and common food allergens. By doing so, we hope to create a system that can suggest recipes tailored to each user's unique background and dietary needs, making cooking more accessible and enjoyable for everyone.

2 Related Works

Some recent papers have explored similar ideas in recipe recommendation and generation, with some addressing allergen concerns. For example, Majumder et al. (Majumder et al., 2019) proposed a personalized recipe generation model that uses attention mechanisms to focus on a user's previously consumed recipes. Their approach showed promising results in generating recipes that align with user preferences, which could be extended to consider allergen avoidance.

Another interesting study by Aljbawi et al (Aljbawi, 2020) focused on completing partial ingredient lists. This system predicts additional ingredients based on an initial list provided by the user, which is similar to our goal of working with available ingredients while avoiding allergens.

Chen et al.(Chen et al., 2021) implemented a novel framework where constrained question answering over a large-scale knowledge graph was used to recommend food recipes to users after considering user's explicit requirements and crucial health factors. This helped to recommend healthy alternatives to users, which aligns with our goal of providing allergen-free options.

The FIRE system, by Chhikara et al (Chhikara et al., 2024) and the Nutrify AI by Han et al. (Han and Chen, 2024), both uses a multi-modal approach, generating recipes from food images and ingredients. While we're not using images, this shows the potential for incorporating different types of input in recipe generation, which could be useful for identifying potential allergens visually.

These works provide a foundation for our project, but we aim to combine multiple aspects (user preferences, available ingredients, and allergen awareness) in a unique way that caters specifically to individuals with food sensitivities.

3 Research Statement

Our core problem statement is: How can we develop an NLP-based system that generates personalized, allergy-friendly recipes by considering user preferences, dietary restrictions, and available in-

*All authors contributed equally.

gredients and kitchenware?

For personalization: How can we effectively capture and represent user preferences and allergen information in our model? This involves understanding not just what ingredients a user likes, but also their allergies and dietary restrictions, preferred cooking styles, and even cultural preferences. We need to explore methods of creating accurate user profiles based on these constraints.

Not only that, given that our model would be prompt-based, when asked for "something new", we would want our AI to generate a recipe that is outside of their usual preference but still safe for their dietary needs. This challenge involves balancing novelty with allergen safety. This goes directly with recipe quality and safety: the AI only has a fixed set of available ingredients, which means it may have to perform substitutions to a traditional recipe.We need to ensure that any substitutions made are not only sensible but also safe for users with specific allergies.

Finally, any recipe generated must be easy to understand and coherent. Steps should be in chronological order, and there should not be any jumps between steps that would be incomprehensible to a reader. Additionally, the recipe should include clear warnings about potential cross-contamination risks.

4 Methodology

In this section, we'll outline our approach to developing the recipe generation system. Our methodology involves several key steps, including data preparation, model architecture design, training process, and implementation of the recipe generation pipeline, all with a focus on allergen awareness.

4.1 Data

We will use a recipe dataset from Kaggle, which contains 180K+ recipes and 700K+ recipe reviews covering 18 years from Food.com. Each entry in the dataset contains the recipe name, ingredients, serving size, servings, an ordered list of strings for instructions, and a description of the recipe. The description of the recipe is essential because it allows our model to learn recipes based on what kind of recipe the user is looking for.

We will also use the Cookbook Reviews Dataset to help us classify user preferences and identify potential allergen concerns. The Cookbook Reviews dataset contains over 18,000 comments and ratings of 100 different recipes. Each item in the dataset has the user id, their comment text and rating, and the recipe title. This will allow us to learn what recipes a user will like based on their previous comments/ratings, and identify any allergen-related comments.

Additionally, we plan to augment our dataset with allergen information for common ingredients, creating a comprehensive allergen database that our model can reference when generating recipes.

4.2 Architecture

For our architecture, we plan to use several models to generate outputs and compare their performances with the evaluation metrics mentioned in the section below. The input to each of these models will be a combination of user preferences, dietary restrictions, and available ingredients and kitchenware. The output will be a generated allergy-friendly recipe with step-by-step instructions.

The models we are planning on using, for now, include a simple encoder-decoder model where the encoder will process the input (preferences, allergies, ingredients, tools), and the decoder will generate the recipe text, a transformer-based model like BERT or GPT-2 paired with a knowledge graph of ingredients and their allergen properties, similar to the one mentioned in Chen et al (Chen et al., 2021), and a small LLM like TinyLLama with guided profile generations, mentioned in Zhang et al (Zhang, 2024).

We will also be experimenting with different methods of representing the input to achieve the best results possible for each model, with a particular focus on effectively encoding allergen information. Figure 1 shows a simplified version of our architecture.

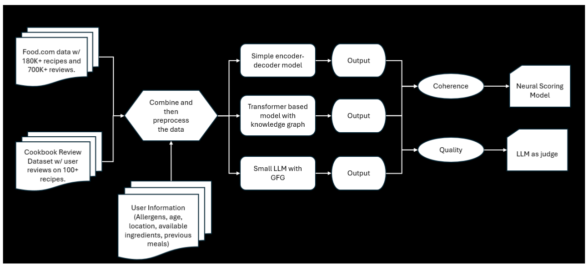

Figure 1: Simplified version of AllergyAware Architecture

5 Evaluation

To evaluate our models, we plan to use a combination of automatic metrics and human evaluation. For automatic metrics, we will use BLEU and ROUGE scores to compare the quality of our generated recipes to real ones. However, since recipes are subjective in nature, we think human evaluation will be necessary as well.

Since human surveys are unfeasible given the scope of the project, we're planning to implement an LLM as a judge that will rate our generated recipes. We are planning on doing this by training the LLM on user reviews of recipes, which would help the LLM understand what constitutes a good recipe. We will also fine-tune this LLM with allergen-specific information to ensure it can accurately judge the safety of generated recipes.

To evaluate for coherence of the recipes generated by the models, we plan to use the neural scoring model from Bosselut et al.(Bosselut et al., 2018). Additionally, we will implement a specific allergen-check module on top of the model to ensure that safety is accounted for.

5.1 Proposed Baseline

We have two models for our baseline: one is the trained model in Majumder et al.(Majumder et al., 2019). Another one is a simple encoder-decoder model with attention to ingredients which is one of our models. We will be using the evaluation techniques mentioned above to compare the performance of the models.

6 Expected Outcome

We're hoping our models will be able to generate recipes that are personalized, allergy-friendly, and actually usable. We expect it to be able to work with a wide range of ingredients and preferences, while strictly adhering to users' dietary restrictions and allergies.

We also hope that our model can demonstrate creativity, by substituting allergens with safe alternatives that are available, and substituting unavailable ingredients for availabe ones. This would demonstrate the model's ability to not only avoid allergens but also creatively adapt recipes to be inclusive.

Ultimately, we aim to develop a tool that could significantly improve the cooking experience for individuals, especially those with food allergies, making it easier for everyone to find and prepare safe, delicious meals. This project has the potential to make a real difference in people's lives, especially for college students who may be cooking for themselves and for those dealing with their food allergies.

References

Bushra Aljbawi. 2020. Health-aware food planner: A personalized recipe generation approach based on gpt-2.

Antoine Bosselut, Asli Celikyilmaz, Xiaodong He, Jianfeng Gao, Po-Sen Huang, and Yejin Choi. 2018. Discourse-aware neural rewards for coherent text generation. *Preprint*, arXiv:1805.03766.

Yu Chen, Ananya Subburathinam, Ching-Hua Chen, and Mohammed J. Zaki. 2021. Personalized food recommendation as constrained question answering over a large-scale food knowledge graph. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, WSDM '21. ACM.

Prateek Chhikara, Dhiraj Chaurasia, Yifan Jiang, Omkar Masur, and Filip Ilievski. 2024. Fire: Food image to recipe generation. *Preprint*, arXiv:2308.14391.

Michelle Han and Junyao Chen. 2024. Nutrifyai: An ai-powered system for real-time food detection, nutritional analysis, and personalized meal recommendations. *Preprint*, arXiv:2408.10532.

Bodhisattwa Prasad Majumder, Shuyang Li, Jianmo Ni, and Julian McAuley. 2019. Generating personalized recipes from historical user preferences. *Preprint*, arXiv:1909.00105.

Jiarui Zhang. 2024. Guided profile generation improves personalization with llms. *Preprint*, arXiv:2409.13093.