ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1460

Innovations in Recipe Generation and Ingredient Substitution: A Survey on RAG and Generative AI Approaches

Swati Jakkan[1]; Rupali Susar Patil[2]; Shantanu Ekad[3]; Pranjali Shinde[4]

[1]Assistant Professor
[1,2,3,4]Artificial Intelligence and Data Science, Marathwada Mitramandal's College of Engineering
Savitribai Phule Pune University Pune, India

Publication Date: 2025/04/05

Abstract: The Recipe Generation and Ingredient Substitution System redefines the cooking experience by offering personalized recipe suggestions and intelligent ingredient substitutions tailored to user preferences, dietary needs, and available ingredients. Leveraging advanced Large Language Models (LLMs) integrated with Retrieval-Augmented Generation (RAG)' the system processes natural language inputs to retrieve relevant recipes or generate customized recipes with precision. Its robust ingredient substitution mechanism evaluates alternatives through a scoring framework that considers flavor profiles, texture, functionality, and compatibility, ensuring that substitute ingredient maintain the dish's integrity and quality. By incorporating flavor profile analysis, the system ensures substitutions preserve the harmony of tastes, enhancing both the quality and enjoyment of meals. The system is designed to empower novice and experienced cooks, the system transforms kitchen challenges into opportunities for exploration, enabling confident and innovative meal preparation.

Keywords: *Recipe Generation, Ingredient Substitution, Large Language Models, (LLMs), Retrieval-Augmented Generation (RAG), Natural Language Processing (NLP).*

How to Cite: Swati Jakkan; Rupali Susar Patil; Shantanu Ekad; Pranjali Shinde. (2025). Innovations in Recipe Generation and Ingredient Substitution: A Survey on RAG and Generative AI Approaches. *International Journal of Innovative Science and Research Technology*, 10(3), 2117-2122. https://doi.org/10.38124/ijisrt/25mar1460.

I. INTRODUCTION

Cooking is often a creative and enjoyable activity, but it can also become challenging when certain ingredients are missing from your kitchen. In such situations, people either try to find alternatives for the missing ingredients or look for different recipes that match their available supplies. This research focuses on building a recipe generation and ingredient substitution system that assists individuals in making the most out of their available ingredients by suggesting appropriate recipes and substitutions when necessary. Leveraging Large Language Models (LLMs) [2] combined with Retrieval-Augmented Generation (RAG) [1], this system will intelligently match the user requests to relevant recipes and suggests high-quality ingredient substitutions when certain ingredients are unavailable. Additionally, the system incorporates a scoring mechanism to ensure that the suggested substitutes are appropriate in terms of flavor, texture, and functionality. The system generates personalized recipes based on user preferences, such as dietary restrictions, using RAG to ensure alignment with specified constraints. This helps ensure that the dish will still turn out well even with substitutes. This research aims to simplify the cooking experience for individuals by allowing individuals to prepare their desired meals with the ingredients they have at hand, all while ensuring that any substitutions made do not compromise the quality or flavor of the final dish.

II. LITERATURE SURVERY

Pre-trained language models like BERT and GPT store factual knowledge implicitly but face challenges in updating knowledge, ensuring interpretability, and avoiding hallucinations. Hybrid models, such as REALM and ORQA, address these issues by integrating parametric models with non-parametric memory for tasks like extractive question answering. Retrieval- Augmented Generation (RAG) [1] takes this further by combining a seq2seq model (e.g., BART) with a dense retriever (e.g., DPR) for knowledge-intensive NLP tasks. RAG achieves state-of-the-art results in open-domain QA and improves factuality and diversity in language generation. Its non-parametric memory allows updating knowledge dynamically, but challenges remain in retriever dependency, computational costs and broader multilingual applications.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1460

Large Language Models (LLMs) [2] have emerged as powerful tools capable of handling diverse NLP tasks through large- scale architectures and extensive data training. Their evolution from traditional language models to self-supervised PLMs, and finally to LLMs with billions of parameters, has been transformative. Models like T5 and GPT-3 demonstrated that LLMs can generalize effectively across tasks, especially with fine-tuning and instruction-based training. These models display emergent capabilities like reasoning, planning, and multi-modal integration, making them adaptable to various fields, including robotics and autonomous systems. Several papers have examined the association between ingredients using common flavor compounds or common food ontologies such as food groups or growing region of the world [17]. Despite impressive generalization, LLMs come with high computational costs and hardware requirements, prompting innovations in architecture, training efficiency, and parameter optimization. The research community has responded with surveys, benchmarking efforts, and numerous contributions exploring efficiency, context handling, and more.

The integration of regional flavor profiles into personalized recipe recommendations builds on advancements in flavor science, data-driven culinary systems, and personalization algorithms. Previous studies have utilized statistical methods, such as TF-IDF, to analyze ingredient importance across diverse cuisines, enabling a quantitative approach to regional flavor characterization. Flavor similarity networks have been extensively researched to understand ingredient pairings and culinary trends, highlighting the role of graph-based models in capturing complex relationships. Recommendation systems in food technology, have shown potential in aligning suggestions with user preferences but are limited in addressing cultural and regional nuances. This work combines TF-IDF-based ingredient scoring and graph-based flavor networks [15] for a more tailored approach. By synthesizing user preferences with regional flavor profiles, it offers an innovative framework for enhancing culinary exploration, addressing gaps in personalization and cross-regional flavor integration in existing systems.

The EvoRecipes framework builds upon existing research in generative AI, genetic algorithms, and knowledge-based systems for recipe creation. Previous works have explored genetic algorithms for culinary innovation, emphasizing optimization of taste, nutrition, and preparation time. Generative AI models like GPT have demonstrated efficacy in generating creative and personalized content, including recipes, although challenges in logical coherence persist. Ontology-based frameworks [14], such as RecipeOn, and knowledge graph [7] like FoodKG [15], provide structured data representation for better contextual understanding of ingredients and cooking actions. Existing studies highlight the importance of evaluating novelty, simplicity, and feasibility in generated recipes. However, few approaches integrate these diverse methodologies into a cohesive system. The EvoRecipes framework uniquely combines genetic algorithms with generative AI and in the ontology-based knowledge to deliver the best context-aware, user-personalized recipes while addressing the shortcomings of standalone techniques, such as excessive complexity, error-prone substitutions, and lack of practicality in real-world recipe creation.

Ingredient substitution systems have been a focal point in culinary informatics, particularly for preserving the integrity of traditional dishes while adapting to ingredient availability. Existing research leverages graph-based approaches and correlation analysis to model ingredient co-occurrence and flavor compatibility. Studies in food pairing [15] have shown that ingredient substitutions can be optimized using statistical models and similarity metrics, ensuring that the substitutes maintain the dish's flavor and composition. Thai cuisine, known for its intricate balance of flavors, has been less explored in this domain, presenting unique challenges and opportunities. The proposed system combines smoothed correlation weight functions with graph-based modeling to analyze ingredient relationships in Thai recipes. By integrating data-driven insights with user-centric design, it provides practical and culturally sensitive substitution recommendations. This approach bridges the gap between traditional recipe preservation and modern ingredient accessibility, advancing the field of personalized recipe adaptation.

Ingredient substitution systems play a critical role in personalized cooking and dietary adjustments, addressing issues like allergies, unavailability, and personal preferences. Previous research has highlighted the importance of maintaining flavor integrity and culinary context during substitutions. Techniques such as ingredient categorization based on attributes like taste, texture, and nutritional value have been employed to identify suitable replacements. Co-occurrence analysis, widely used in recipe mining, has proven effective in understanding ingredient pairings and enhancing substitution relevance. Existing systems often struggle to balance personalization with preserving the dish's original flavor profile. The proposed system integrates ingredient categorization, single-ingredient frequency analysis, and co-occurrence data to suggest contextually appropriate alternatives. By emphasizing compatibility with seasoning and culinary intent, it offers a comprehensive solution for maintaining the authenticity of recipes while catering to individual needs, advancing the state-of-the-art adaptive recipe recommendation systems.

Ingredient substitution systems increasingly incorporate semantic analysis to enhance the accuracy and relevance of recommendations. Traditional approaches rely on explicit semantics, such as ingredient categories (e.g., meats, spices), and implicit semantics, capturing nuanced relationships between ingredients through data-driven techniques. In recent years, word embeddings have come in vogue in the Natural Language Processing community. Their ability to capture meaning based on context has allowed linguists to qualitatively define the meaning of words [16]. Word embeddings like Word2Vec have been used to quantify ingredient similarity based on linguistic and contextual usage patterns in large recipe datasets. Tools like spaCy extend this by offering semantic similarity scores, facilitating deeper

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1460

contextual understanding. Co-occurrence analysis and Pointwise Mutual Information (PMI) have demonstrated effectiveness in uncovering implicit relationships, such as frequent pairings and their strength in recipes. The proposed approach uniquely combines these methodologies to integrate explicit and implicit semantics, ensuring substitutions maintain both flavor integrity and culinary context. By leveraging multi-faceted similarity measures, this system advances ingredient substitution frameworks, addressing challenges in balancing semantic accuracy with practical applicability.

Advances in natural language processing (NLP) have opened new avenues for culinary applications, including ingredient substitution. Pre-trained language models like BERT and Word2Vec have proven effective in understanding contextual and semantic relationships in textual data. Food-specific adaptations, such as FoodBERT and Food2Vec, enhance these models by fine-tuning them on culinary datasets, capturing domain-specific nuances like ingredient pairing and substitution patterns. Recent studies emphasize the importance of multimodal approaches, integrating textual, visual, and contextual data to improve model performance. User reviews from online platforms provide valuable insights into practical ingredient usage and substitutions, reflecting real-world preferences and feedback. The proposed system combines domain-adapted language models with user-generated data and multimodal techniques to enhance substitution accuracy. This innovative integration of NLP and user insights represents a significant step toward personalized & context-aware ingredient substitution, addressing gaps in adaptability and user-centric design in existing systems.

Recent advancements in computational gastronomy focus on leveraging ingredient embeddings and graph-based models to capture the intricate relationships between ingredients, recipes, and their chemical compounds. Embedding techniques, which encode the semantic and contextual properties of ingredients, have shown promise in preserving nuanced relationships such as flavor compatibility, texture, and culinary roles. Studies that incorporate chemical compound data have demonstrated enhanced performance in predicting ingredient pairings, as shared compounds often underpin flavor harmony. Graph-based models, particularly those with structured relationships like Ingredient-Recipe-Ingredient and Ingredient-Chemical-Ingredient, effectively represent these complex networks, enabling a deeper understanding of ingredient interactions.

Existing research typically considers either recipe data or chemical compound relationships but rarely both. This approach integrates these dimensions, using graph embeddings to capture multi-layered connections and produce richer, more accurate ingredient representations. Such a system offers significant potential for improving applications like ingredient substitution, personalized recipe recommendations, and flavor discovery, addressing gaps in multi-modal culinary knowledge representation.

Graph-based approaches have become a cornerstone in computational gastronomy, particularly for understanding ingredient relationships and enhancing food pairing and substitution systems. Co-occurrence graphs, which represent the frequency with which ingredients appear together in recipes, have been extensively studied for identifying culinary trends and flavor compatibility. Similarly, chemical compounds overlapping between ingredients has been recognized as a key factor in food pairing, as shared compounds often contribute to complementary flavors. Previous research, such as flavor pairing theory, has validated the importance of these connections in enhancing recipe design and substitutions.

The proposed methodology combines ingredient co-occurrence, and chemical compound overlaps into a unified graph structure, offering a multi-faceted representation of ingredient relationships. This graph not only facilitates accurate food pairing but also supports context-aware ingredient substitution by ensuring that replacements align with both culinary usage and flavor chemistry. By integrating these dimensions, the approach advances the state-of-the-art in data-driven culinary applications.

FlavorGraph has emerged as a pioneering framework for understanding ingredient relationships by combining co-occurrence, chemical compound similarity, and flavor compatibility in a graph-based structure. Building on this foundation, GISMo Graph introduces an enhanced representation that incorporates not only the ingredient relationships captured in FlavorGraph but also ingredient and context embeddings. Ingredient embeddings leverage semantic relationships derived from recipe and ingredient data, while context embeddings encode the situational or usage context, such as dietary preferences or regional cuisine.

The integration of these embeddings with GISMo Graph represents a significant advancement, enabling more nuanced ingredient substitution that respects both flavor integrity and culinary context. This approach addresses limitations in traditional systems by combining deep semantic understanding with graph-based relational insights. By enhancing the adaptability and accuracy of substitutions, GISMo Graph sets a new standard for personalization and user-centric recipe recommendations in computational gastronomy.

FlavorDB2 is an advanced and updated database of flavor molecules, designed to enhance understanding of flavor compounds through a comprehensive approach. The database expands upon its predecessor, FlavorDB, by integrating molecular properties, flavor profiles, regulatory details, natural occurrences, and applications in food categories. Data is compiled from diverse resources, including FooDB, BitterDB, and SuperSweet, with enriched fields such as nomenclature, taste/aroma thresholds, and synthesis details. By combining molecular and entity-level information, FlavorDB2 provides a structured repository, facilitating advanced search mechanisms and offering applications in food pairing, molecular gastronomy, and sensory analysis. Its robust framework serves as a standard tool for researchers and professionals in computational gastronomy, bridging gaps in flavor science.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1460

This study explores the application of large language models (LLMs) in ingredient substitution to enhance the phytochemical content of meals. Phytochemicals, plant-derived bioactive compounds, are known for their potential health benefits, including antioxidant and anti-inflammatory properties. By fine-tuning models such as OpenAI's GPT-3.5 and Meta's TinyLlama using a curated ingredient substitution dataset, we aimed to generate substitutions that align with nutritional goals and produce enriched recipe datasets.

The methodology improved the Hit@1 accuracy for predicting ingredient substitutions, raising it from 34.53% to 38.03% on the GISMo dataset and from 40.24% to 54.46% on its refined version. This led to the creation of 1,951 phytochemically enriched ingredient pairings and 1,639 unique recipes. The study builds on the foundation of prior work in computational gastronomy, advancing from statistical co-occurrence and graph neural network (GNN) approaches to LLMs. These models leverage semantic understanding, providing contextually appropriate and nutritionally optimized substitutions.

While this work demonstrates the potential of LLMs in enhancing dietary practices, it emphasizes caution in interpreting health claims due to the reliance on preclinical evidence.

III. SCOPE AND OBJECTIVE

A. *Scope*

Our System aims to create an intelligent recipe generation and ingredient substitution system that is capable of generating personalized recipes tailored to their taste preferences, and ingredient availability. The system utilizes large language models (LLMs) and data-driven techniques such as embeddings, co-occurrence algorithms, and flavor compound analysis to generate recipes and suggest meaningful substitutions without compromising the quality or flavor of the dish. It will support various user inputs, including ingredient lists, recipe names, or taste profiles, and provide real-time feedback to refine its recommendations.

B. *Objective*

- Develop a robust system to generate recipes based on user inputs, including ingredient lists, recipe names, and taste preferences.
- Implement ingredient substitution algorithms that preserve the flavor profile and integrity of recipes with the help of robust scoring mechanism.
- Ensure compatibility with dietary restrictions (e.g., vegan, gluten-free) for safe and inclusive recipe planning.
- Optimize system performance to provide recipe suggestions and substitutions in real-time.
- Integrate external culinary databases for accurate and comprehensive recipe generation.
- Provide a user-friendly interface accessible across devices, ensuring seamless user interaction and satisfaction.

IV. SYSTEM ARCHITECTURE

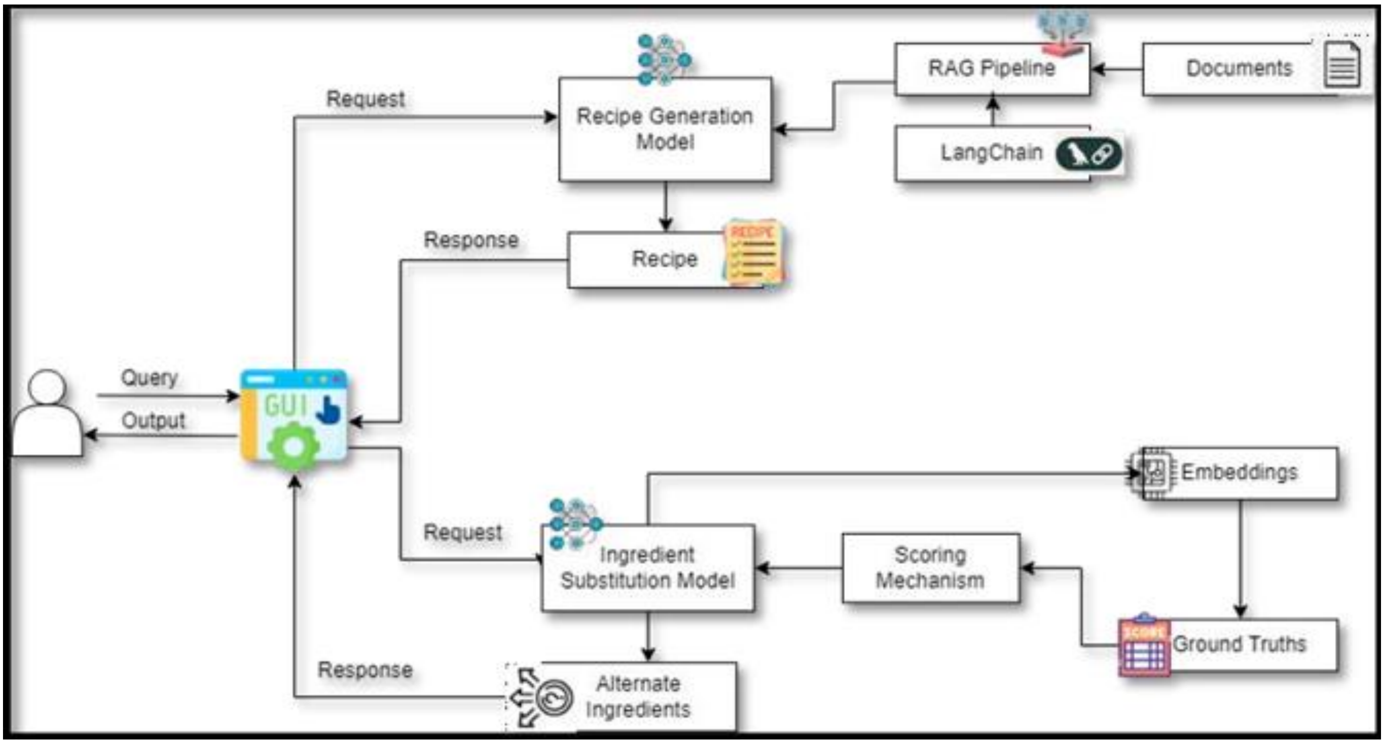

Fig 1: System Architecture

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1460

The above diagram illustrates our system consisting of following main modules:

➢ *User Input (Prompt):*

- The user submits a query or prompt through a GUI (Graphical User Interface), likely specifying preferences for a recipe.

➢ *Recipe Generation Model:*

- The system sends a request to the Recipe Generation Model, which returns a tailored recipe based on user preferences. The model incorporates a RAG (Retrieval-Augmented Generation) Pipeline and uses LangChain to integrate with external knowledge sources, processing Ground Truths (likely databases or verified recipe information).

➢ *Ingredient Substitution Model:*

- If the user wants to substitute a particular ingredient, a request is sent to the Ingredient Substitution Model, which suggests alternate ingredients. This model leverages Embeddings (for semantic understanding) and Ground Truths to offer suggestions. It uses a Scoring Mechanism to rank or evaluate alternate ingredients based on user preferences or constraints.

➢ *Response:*

- The system sends back the recipe or ingredient substitutions to the user via GUI based on the prompt, generating the final output.

This architecture efficiently addresses both recipe generation and ingredient substitution based on ingredient availability and user preferences.

V. CONCLUSION

The system aims to tackle the challenge of generating personalized recipe recommendations based on user-provided ingredients, dietary preferences, and culinary choices. By integrating advanced techniques such as language models and retrieval-augmented generation (RAG), the system will be able to generate customized recipes while suggesting ingredient substitutions that considers critical factors like flavor, texture, and functionality. The personalization aspect of the system allows users to input their available ingredients, enabling tailored recipe suggestions that fit individual needs. Moreover, the incorporation of a scoring mechanism for ingredient substitutions ensures that alternatives are compatible with the recipe, maintaining its original intent while offering flexibility. Overall, the system aims to combine multiple advanced methodologies to create a powerful system for personalized recipe generation and ingredient substitution, demonstrating its potential in both culinary and machine learning applications.

REFERENCES

[1]. Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." Advances in Neural Information Processing Systems 33 (2020): 9459-9474.

[2]. Naveed, Humza, et al. "A comprehensive overview of large language models." arXiv preprint arXiv:2307.06435 (2023).

[3]. Mao, Xuehui, et al. "Recipe recommendation considering the flavor of regional cuisines." 2016 International conference on progress in informatics and computing (PIC). IEEE, 2016.

[4]. Razzaq, Muhammad Saad, et al. "EvoRecipes: a generative approach for evolving context-aware recipes." IEEE Access(2023).

[5]. Nadee, Wanvimol, and Sayan Unankard. "Alternative-ingredient recommendation based on correlation weight for thai recipes." 2021 joint international conference on digital arts, media and technology with ECTI northern section conference on electrical, electronics, computer and telecommunication engineering. IEEE, 2021.

[6]. Shino, Naoki, Ryosuke Yamanishi, and Junichi Fukumoto. "Recommendation system for alternative-ingredients based on co-occurrence relation on recipe database and the ingredient category." 2016 5th IIAI international congress on advanced applied informatics (IIAI-AAI). IEEE, 2016.

[7]. Shirai, Sola S., et al. "Identifying ingredient substitutions using a knowledge graph of food." Frontiers in Artificial Intelligence 3 (2021): 621766.

[8]. Pellegrini, Chantal, et al. "Exploiting Food Embeddings for Ingredient Substitution." HEALTHINF 5 (2021): 67-77.

[9]. Yoshimaru, Naoki, et al. "Construction of Ingredient Embedding Considering Both Cooking Recipes and Their Ingredients." 2024 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2024.

[10]. Park, Donghyeon, et al. "FlavorGraph: a large-scale food-chemical graph for generating food representations and recommending food pairings." Scientific reports 11.1(2021):931.

[11]. Fatemi, Bahare, et al. "Learning to substitute ingredients in recipes." arXiv preprint arXiv:2302.07960(2023).

[12]. Goel, Mansi, et al. "FlavorDB2: an updated database of flavor molecules." Journal of Food Science (2024).

[13]. Rita, Luís, et al. "Optimizing ingredient substitution using large language models to enhance phytochemical content in recipes." Machine Learning and Knowledge Extraction 6.4 (2024): 2738-2752.

[14]. Haussmann, S. et al. (2019). FoodKG: A Semantics-Driven Knowledge Graph for Food Recommendation. In: Ghidini, C., et al. The Semantic Web – ISWC 2019. ISWC 2019. Lecture Notes in Computer Science(), vol 11779. Springer, Cham.

[15]. Ahn, Yong-Yeol, et al. "Flavor network and the principles of food pairing." Scientific reports 1.1 (2011): 196.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1460

[16]. Sauer, Christopher, Alex Haigh, and Jake Rachleff. "Cooking up food embeddings." (2017)
[17]. Sozuer Zorlu, Sibel, Oded Netzer, and Kriste Krstovski. "A Recipe for Creating Recipes: An Ingredient Embedding Approach." Available at SSRN 4686749 (2024).