# Project Development Phase

# Data Collection & Preparation

| | |
|---|---|
| Date | 10-02-2026 |
| Team ID | LTVIP2026TMIDS65437 |
| Project Name | Flavour Fusion: -Ai-Driven Recipe Blogging |
| Maximum Marks | 5 marks |

## 1.Data source identification (datasets, APIs, user inputs, CSV, etc.)

Data source identification for a "Flavour Fusion" AI-driven recipe blogging platform is the crucial process of gathering, cleaning, and structuring culinary information to train models that can generate creative, hybrid, and, in some cases, image-based recipes. The goal is to blend different culinary traditions (e.g., Italian Indian fusion) based on inputs like ingredient availability, visual images, or user preferences.

| Data Source | Type | Purpose for Flavour Fusion |
|---|---|---|
| **Recipe NLG/Recipe1M+** | Dataset | Teaching the AI structure of recipes |
| **Spoonacular/Edamam API** | API | Nutrition, substitutions, and accuracy |
| **User Input (Image)** | User | Identifying ingredients in a photo |
| **User Input (Pantry)** | User | Generating recipes from available food |
| **Custom CSVs** | Structured | Mapping ingredient taxonomy |
| **Web Scraping** | Unstructure | Discovering new fusion trends |

## 2.Data collection method

Data collection for AI-driven, "flavour fusion" recipe blogging involves using machine learning to analyze vast, disparate datasets—culinary, chemical, and social—to identify unique, high-probability ingredient combinations that humans might not immediately pair.

| Method | Data Source | Application in Flavour Fusion |
|---|---|---|
| **API/Web Scraping** | Recipe Sites (All Recipes) | Gathering common ingredient pairings. |
| **Scientific Database** | Flavour DB, FooDB | Identifying chemical/molecular matches. |
| **Social Listening** | Instagram, TikTok | Detect trending flavors/styles. |
| **Computer Vision** | Food Images/Camera | Identify ingredients visually. |
| **NLP** | Blogs/Reviews | Extracting unstructured data/trends. |

## 3.Data cleaning (remove nulls, duplicates, noise):

Data cleaning is the foundational step in building an AI-driven recipe blog, especially for "flavour fusion," where the AI must learn to creatively combine distinct culinary traditions (e.g., Thai Mexican). Raw recipe data is often scraped from diverse, messy sources, resulting in inconsistencies that can cause AI models to produce incorrect measurements or nonsensical, unappetizing combinations.

| Issue | Description | Technique |
|---|---|---|
| **Nulls** | Missing ingredient amounts or steps. | Drop rows or impute with median value. |
| **Duplicates** | Multiple instances of "Kimchi Quesadilla". | Normalize titles, then drop duplicates. |
| **Noise** | "1 cup" vs "200g" of ingredient. | Standardize units (e.g., to grams). |
| **Structure** | Varying formats (List vs Paragraph). | Tokenize and parse ingredient entities. |

## 4.Data preprocessing & formatting for GenAI input:

Data preprocessing and formatting for Generative AI (GenAI) in flavor fusion recipe blogging involves transforming unstructured, heterogeneous culinary data (images, ingredient lists, blog text) into structured, machine-readable, and semantically rich formats. This ensures that GenAI models can understand culinary relationships to generate new, innovative, and functional recipes.

### Summary of Workflow

- **Gather:** Scrape/upload recipes and images.
- **Clean:** Remove junk, fix typos.
- **Normalize:** Standardize units and in
- **Inject:** Use as context in a prompt for AI (e.g., "Using these ingredients, generate a Indian-Italian fusion recipe...").
- **Structure:** Convert to JSON.

## 5.Data validation and quality check:

Data validation and quality checks for Generative AI in an AI-driven recipe blog like "Flavour Fusion" are critical to ensure that generated recipes are safe, accurate, and delicious. Since GenAI can "hallucinate" (invent) ingredients or cooking methods, implementing rigorous input validation—before the data reaches the LLM—and output checks is crucial for building trust, safety, and high-quality content.

### Key Metrics for Flavour Fusion

| Metric | Description |
|---|---|
| **Relevance** | Does the recipe match the user's available ingredients? |
| **Completeness** | Does it include necessary details like cooking times and serving sizes? |
| **Safety/Sanity** | Are the ingredient combinations and cooking methods safe for consumption? |
| **Readability** | Is the complexity of the language suitable for the target audience (e.g., Flesch-Kincaid score)? |

| | |
|---|---|
| 1. Input Data Validation | • Image Preprocessing & Validation<br>• Structured Data Checks<br>• Entity Resolution |
| 2. Quality Checks for AI Generation | • Faithfulness Score<br>• Coherence & Flow<br>• Constraint Validation |
| Continuous Improvement | To maintain high standards, Flavour Fusion utilizes **Feedback-Driven Development**. User ratings and reviews are analyzed to refine the machine learning models over time, reducing errors and improving the "deliciousness" of suggestions. |

## 6.Data privacy & ethical consideration:

### Table of Ethical & Privacy Concerns

| Area | Key Concern | Mitigation Strategy |
|---|---|---|

| | | |
|---|---|---|
| **Privacy** | Unauthorized data mining of user preferences | Data minimization & explicit consent |
| **Security** | Data breaches of personal info | Encryption & secure storage |
| **Ethics** | Cultural insensitivity/misappropriation | Human oversight/culinary experts |
| **IP** | Unknown ownership of AI recipes | Clear licensing/contracts |
| **Bias** | Narrow, Western-centric recipes | Diverse dataset training |

## 7.Input-output examples for prompting:

| | |
|---|---|
| The Core Components of a Fusion Prompt | • Target Cuisines<br>• Key Ingredients<br>• Format Constraints<br>• Style Notes |
| Input (Prompt): | 1. Create a unique, 30-minute fusion recipe combining Italian and Thai cuisines. The recipe should be a pasta dish featuring coconut milk and basil. Give it a creative name, detailed instructions, and an engaging blog post introduction." |
| Output (AI Generation) | • Title<br>• Intro<br>• Key Ingredients<br>• Steps |

## 8.Dataset documentation:

Dataset documentation for a Generative AI-driven recipe blogging platform, such as "Flavour Fusion," is critical for ensuring the AI understands culinary relationships, produces accurate recipes, and generates engaging content. A well-documented dataset transforms raw data into structured, actionable insights for the model, enabling it to suggest unique flavor combinations.