Semantics in Intelligent Information Access

2023/2024

SUSTAINAMEAL

AI-Powered Recipe Recommendations for Sustainable Eating

Giovanni Tempesta
Michele Di Carlo

Content

Introduction to the problem

In the face of escalating health issues and environmental concerns, our proposal confronts the critical need for a dietary shift that prioritizes both personal health and ecological well-being.

Health & Diet Connection

Sustainable Eating

The environmental cost of food production is unsustainable, prompting a movement towards diets that have a lower ecological footprint.

TASK

In the face of escalating health issues and environmental concerns, our project confronts the critical need for a dietary shift that prioritizes both personal health and ecological well-being.

“ We need to build a system which given a recipe name suggest an alternative recipe which is more healthier and sustainable”

TASK

Data Analysis and Preprocessing

Starting with a diverse web-sourced recipe dataset we analyzed and reduced them through a preprocessing process

Healthiness and Sustanability Approach

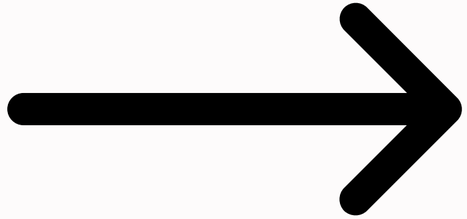
For each recipe we've applied data processing to calculate each ingredient's CFP (carbon foot print) and WFP (water foot print)

AI-Powered

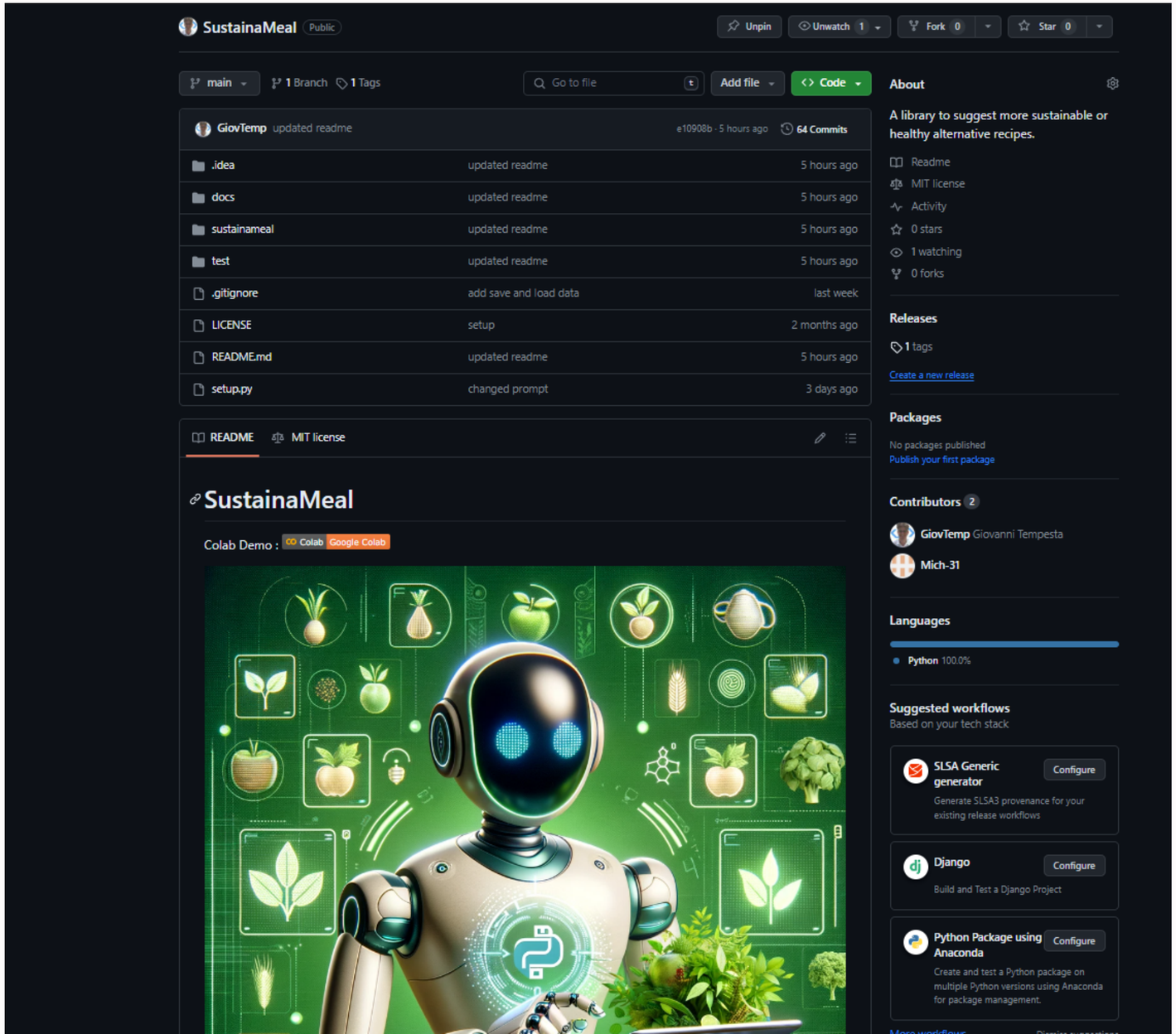
we will use the power of artificial intelligence tools (such as transformers and large language models) to suggest alternative recipes

STARTING POINT

“ We need to build a system which given a recipe name suggest an alternative recipe which is more healthier and sustainable ”

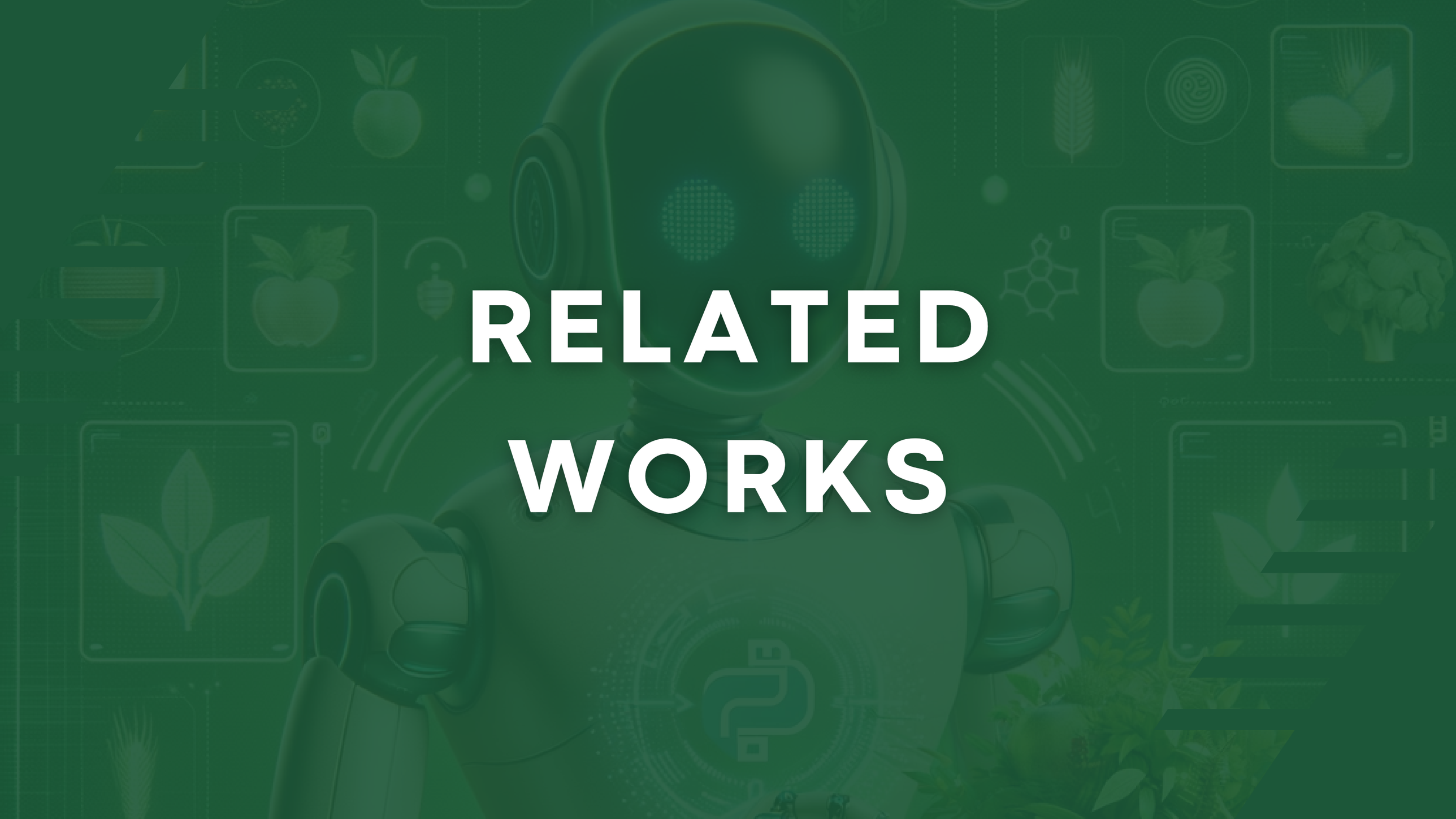


PROPOSAL



PIP

LangChain


RELATED
WORKS

FoodPrintDB

The research "FoodPrintDB" by Gigantelli Alberto and Iacovazzi Antonio Raffaele aimed to create a comprehensive database of ingredients and recipes for future food recommendation systems, focusing on high-quality, reliable information with a special emphasis on sustainability. The database offers valuable insights into sustainability, facilitating the identification of sustainable alternatives for specific recipes. The parameters considered
to evaluate sustainability were the Carbon Foot Print and the Water Foot Print.

First Version

The first version of the database was originally released by two other colleagues, Matteo Fusillo and Salvatore Amoruso, as support for their recommendation system

Second Version

Subsequently, additional information from the SU-EATABLE Life database was integrated into this initial version. In particular, the Carbon Foot Print and Water Foot Print values were added

DATASETS

Datasets Description

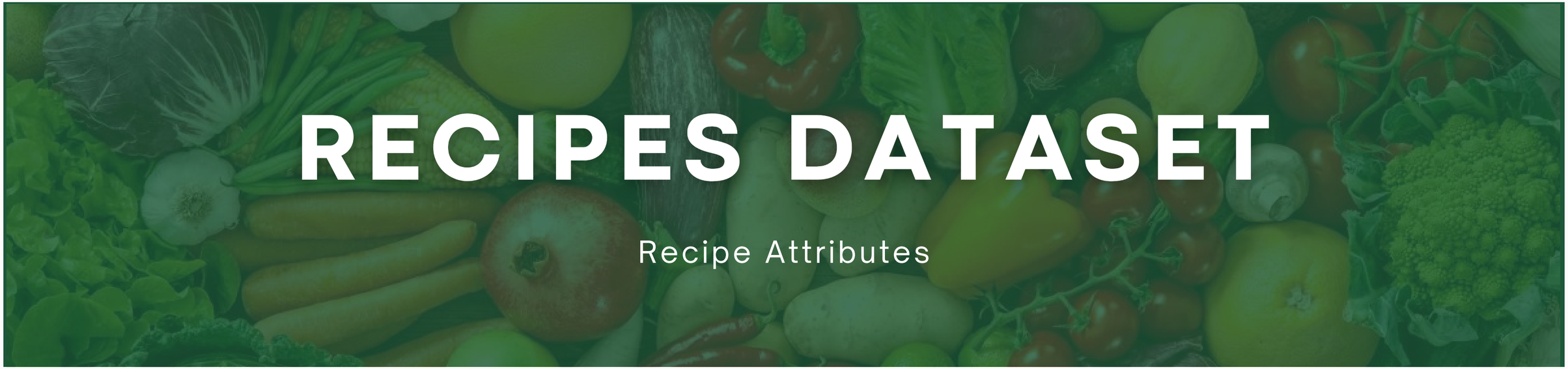
Recipes Dataset

The recipes in our dataset were obtained through a comprehensive web scraping process. This involved systematically collecting data from various online culinary websites and recipe databases.

CSEL Dataset

The dataset comes from an elaboration of the SU-EATABLE Life(SEL) database, which is a multilevel database on the carbon (CF) and water footprint (WF) values of food raw materials


EDA



RECIPES DATASET

Recipe Attributes

recipe_id

A unique identifier for each recipe

title

The name of the recipe.

ingredients

List of ingredients used in the recipe

tags

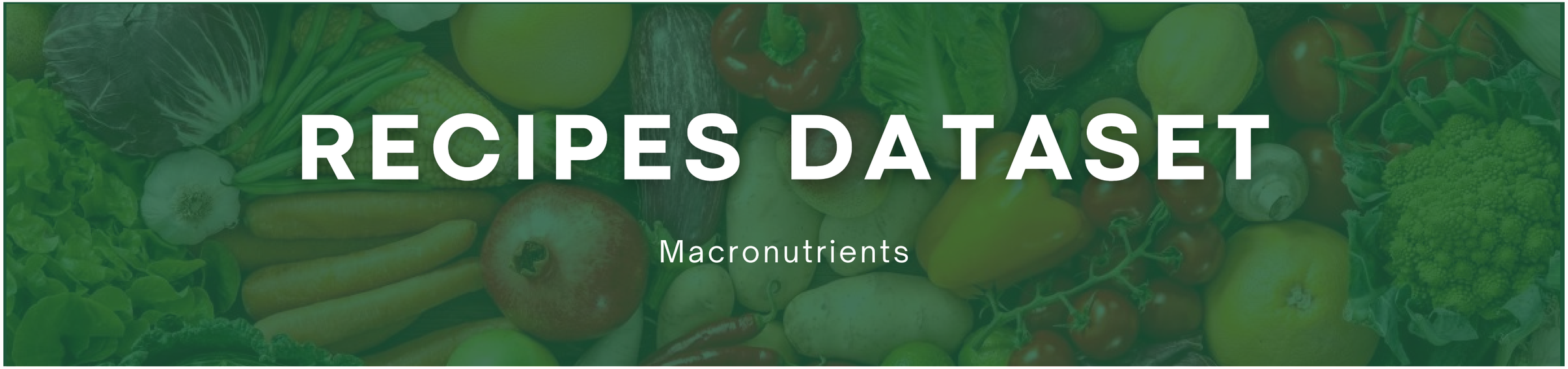
Categorization tags associated with each recipe

= attributes used in the workflow

Tags

The dataset has a total of 902 tags, for example we have:

- Main dish = 71285
- Desserts = 42807
- Breakfast = 13384
- Appetizers = 20187
- Vegetables = 53485
- Meat = 50740
- Seafood = 14722
- Vegetarian = 35599
- Fruit = 31245
- Pasta rice and grains = 23924
- Chicken = 20304
- Pork = 12664
- .....



RECIPES DATASET

Macronutrients

calories [cal]

The total calorie content of the recipe

caloriesFromFat [cal]

The amount of calories derived from fat

totalFat [g]

Total fat content in grams

cholesterol [mg]

Cholesterol content in milligrams

sodium [mg]

Sodium content in milligrams

totalCarbohydrate [g]

Total carbohydrates in grams.

dietaryFiber [g]

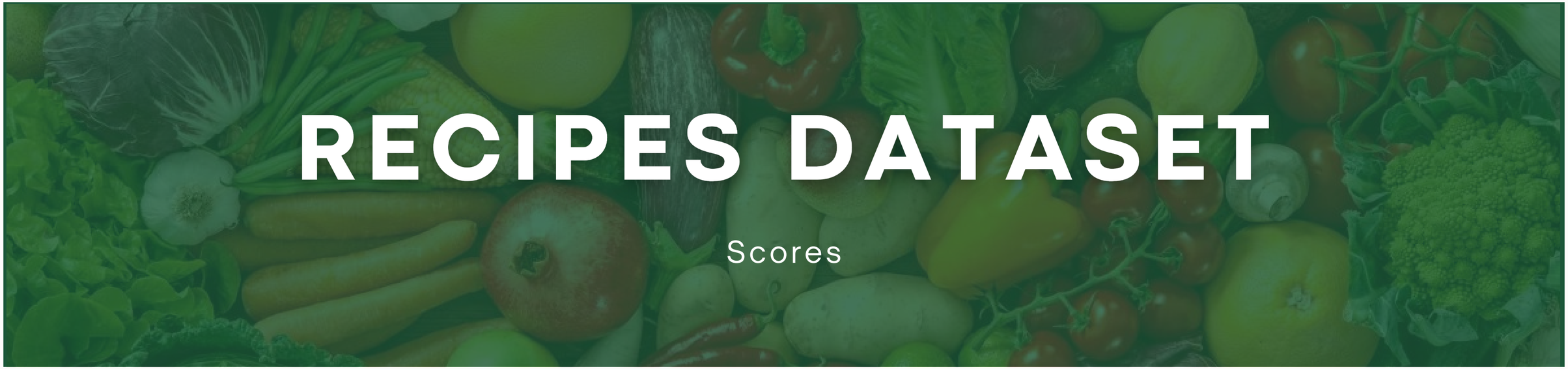
Dietary fiber content in grams

sugars [g]

Total sugars in grams

protein [g]

Protein content in grams



RECIPES DATASET

Scores

who_score

A healthiness score for each recipe based on the World Health Organization (WHO) methodology, ranging from 0 to 14, with 14 being the best

fsa_score

A healthiness score based on the UK Food Standards Agency (FSA) nutrient profiling system, ranging from 0 to 8, with 8 being the best.

nutri_score

A nutritional score for each recipe, graded from A (best) to E (worst)

Limitations

The structure of recipes dataset presents some important limitations that requires some strategies to reduce its impact on our analisis

1

Dataset noise

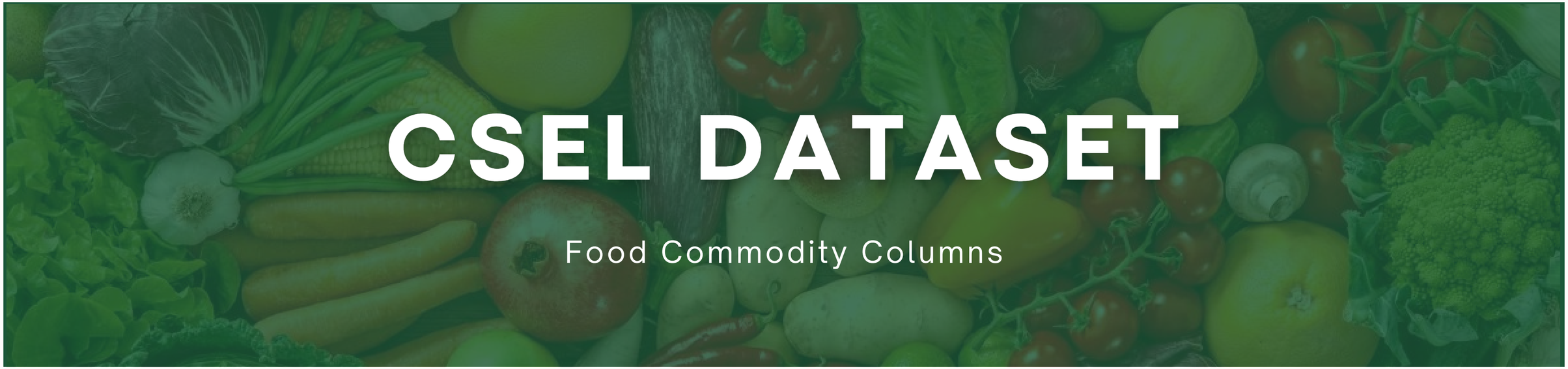
the unstructured nature of web-sourced recipe data

('dried cranberries', '1 time(s) cup ')

2

Null tags and ingredients

On 507 000 recipes only 214 000 were suitable for our use case because they had ingredients and tags other than null



CSEL DATASET

Food Commodity Columns

Food Commodity GROUP

The general category to which food goods belong

Food commodity ITEM

Specific ingredient within the category

Food commodity TYPOLOGY

Subcategory within a food group, particular typology

Food commodity sub-TYPOLOGY

Subcategory of the subcategory within a food group

CSEL DATASET

Foot Prints

final_co2

Final value of Carbon Foot Print

final_wfp

Final value of Water Foot Print

Limitations

The CSEL dataset faces some important limitations that are challenging for us. To address these challenges, we've resorted to employing methodologies outlined in the data preprocessing section.

1

Dataset noise

Presence of additional word and character

2

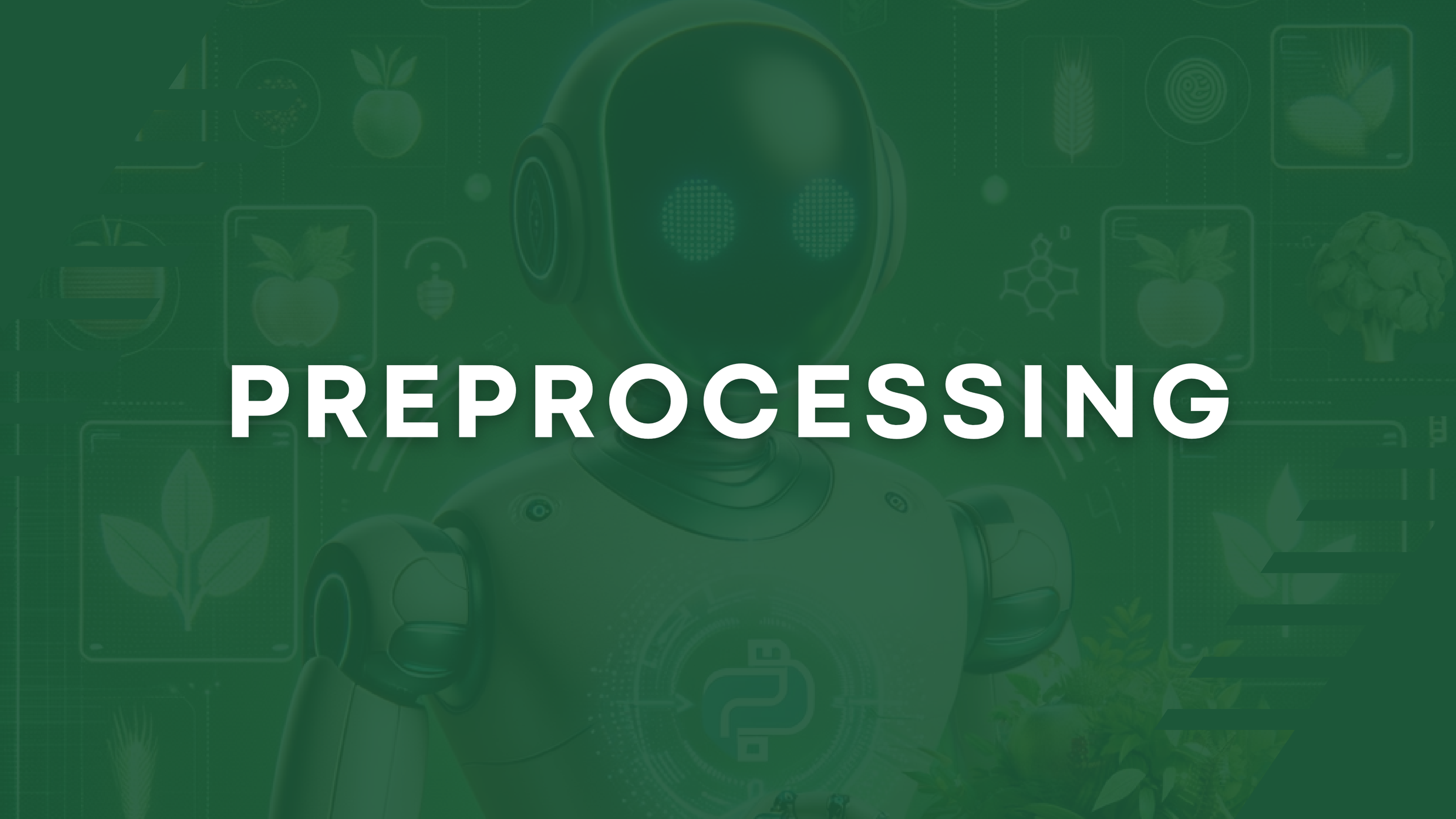
Limited ingredients

Only 471 items

3

Absence of common ingredients

Some important ingredients are missing


PREPROCESSING

Recipe Dataset Reduction

01 - Removing duplicates

Recipes with identical titles are dropped

02 - Removing Null Tags Recipes

Recipes with ’tags’ property as null are dropped

03 - Removing Null Ingredients Recipes

Recipes with ’ingredients’ property as null are dropped

This step reduced recipes from 507.335 to 214.800

Recipe Dataset Cleaning

01 - Text Cleaning

Removal of adjectives, stopwords, verbs, numbers, special characters and content within parentheses

Caramel Sweetened Condensed Milk

Caramel Milk

CSEL Dataset Cleaning

01 - Removing no cfp and wfp ingredients

Removal of ingredients lacking both cfp and wfp values

02 - Text Cleaning

Elimination of adjectives, stopwords, and special characters like asterisks (*).

Dark Chocolate

Chocolate

03 - Removing Parenthesis

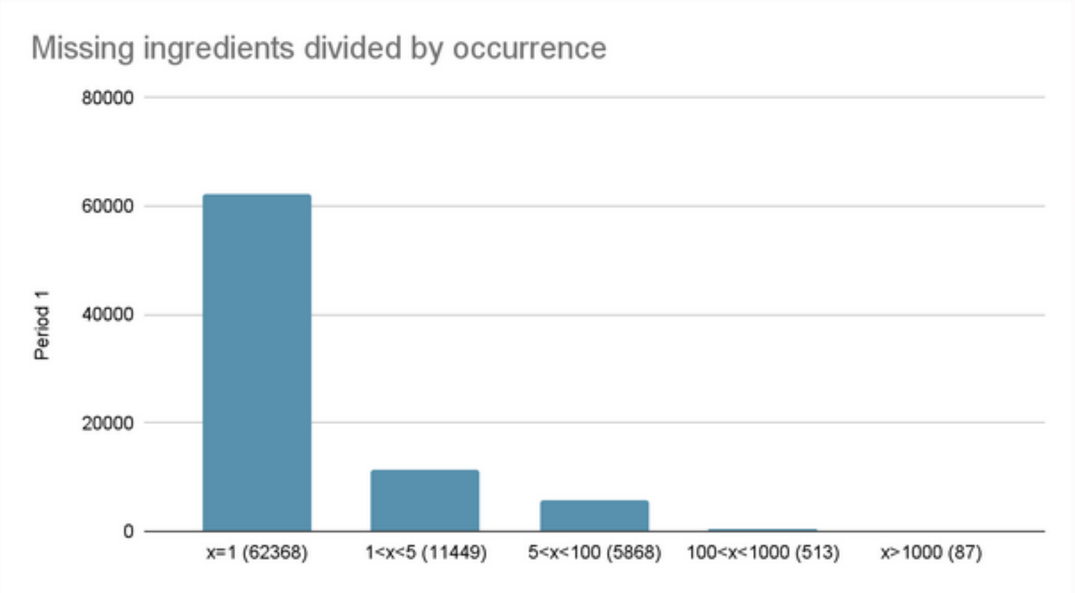
Removal of all parentheses and their contents, e.g., ’(F)’, ’(G)’, etc

Carrots (F)

Carrots

This step reduced igredients from 471 to 448

Ingredient Matching



01 - Ingredients Checking(CSEL -> Recipe Dataset)

Checking if the CSEL dataset ingredient name is contained within our ingredient name.

Milk ⟶ low-fat Milk

02 - Ingredients Checking(Recipe Dataset -> CSEL)

Checking if the ingredient name is contained within the CSEL ingredient name

Beef ⟶ Beef Bone Free Meat

This matching process assigned 78.998 out of 159.284 ingredients

Transformer Usage for Missing Ingredients

01 - Transformer

Used a transformer to calculate similarity between missing ingredients (occurring more than 1 time) and matched ingredients

02 - Threshold

Set a threshold of 0.98 as the minimum similarity

03 - Filtering results

Obtained around 240 possible similarities, manually reviewing and eliminating the inconsistent ones, resulting in 178 similarities.

Turkey neck bone ⟶ Turkey meat bone ⟶ Turkey Meat

Manual Assignment of Most Relevant Missing Ingredients

01 - Assignments

After automated analysis, further manual intervention addressed 87 missing ingredients with more than 1000 occurrences, finding 19 possible associations.

gorgonzola ⟶ cheese

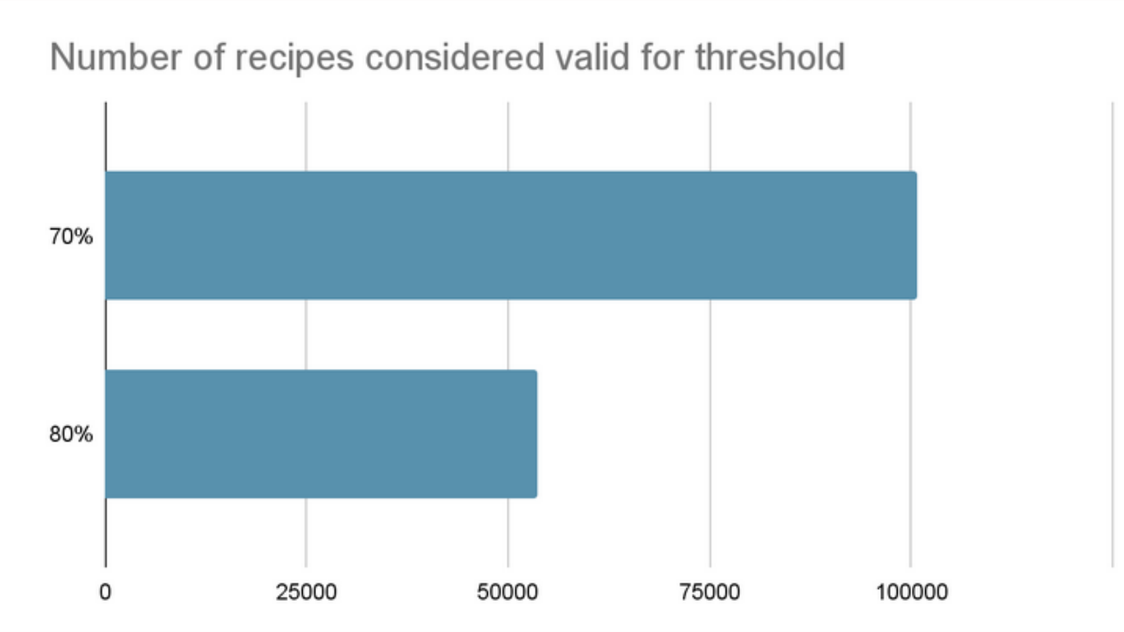
Creation of Dictionary

01 - Dictionary

Created a dictionary containing names of all processed ingredients with their respective cfp and wfp values

Recipes dataset reduction

We employ the same approach as the previous work ’Food Print DB’ for managing recipes with missing ingredients. Thus, we proceed to reduce our dataset by eliminating all recipes that we do not deem valid



Valid Recipe

The percentage of known and valid ingredients is greater than a specified threshold.

An ingredient is considered valid if we have the values for both cfp and wfp

Currently we have decided to use recipes that respect the threshold constraint of 70% (100.870 out of 214.800). Therefore the final number of recipes is 100.870

Sustainability Calculation Procedure

ISS(Ingredient Sustainability Score)

The ISS (Ingredient Sustainability Score) was computed for each ingredient within the recipe employing the ’iss score’ function.

$$\mathrm{ISS} = \alpha \times \mathrm{Ncfp}(x) + \beta \times \mathrm{Nwfp}(x)$$

alpha

A constant value setted to 0.8

Ncfp(x)

Ncfp(x) represents the normalized Carbon Foot Print value

beta

A costant value setted to 0.2

Nwfp(x)

Nwfp(x) denotes the normalized Water Foot Print value for the ingredient

Sustainability Calculation Procedure

DSS(Dish Sustainability Score)

After obtaining the ISS for each ingredient within the recipe, the DSS (Recipe Sustainability Score) was calculated using the formula:

$$\mathrm{DSS} = \sum_{i=0}^{|K|-1} \mathrm{ISS}(K_i)e^i$$

$ISS(K_i)$

The ISS value of the i-th ingredient of K, where K is a set of ingredients oredered descending by the ISS score

e^i

Numeric costant with a value of 2.71

Sustainability Calculation Procedure

Normalization of the DSS

Finally, the ultimate sustainability score was computed using the formula:

$$\text{Sustainability Score} = \frac{\text{DSS(R)} - \text{MinDSS}}{\text{MaxDss} - \text{MinDss}}$$

DSS(R)

The DSS value of the recipe R

MinDSS

Minimum value between all the DSS scores

MaxDSS

Maximal value between all the DSS scores

Sustainability Calculation Procedure

Label Assignment

Employing a similar methodology as in the previous study, we assign labels to each recipe based on the computed score. Recipes are categorized into three labels:

Low(0)

Indicating recipes with low sustainability(score >= 0.5)

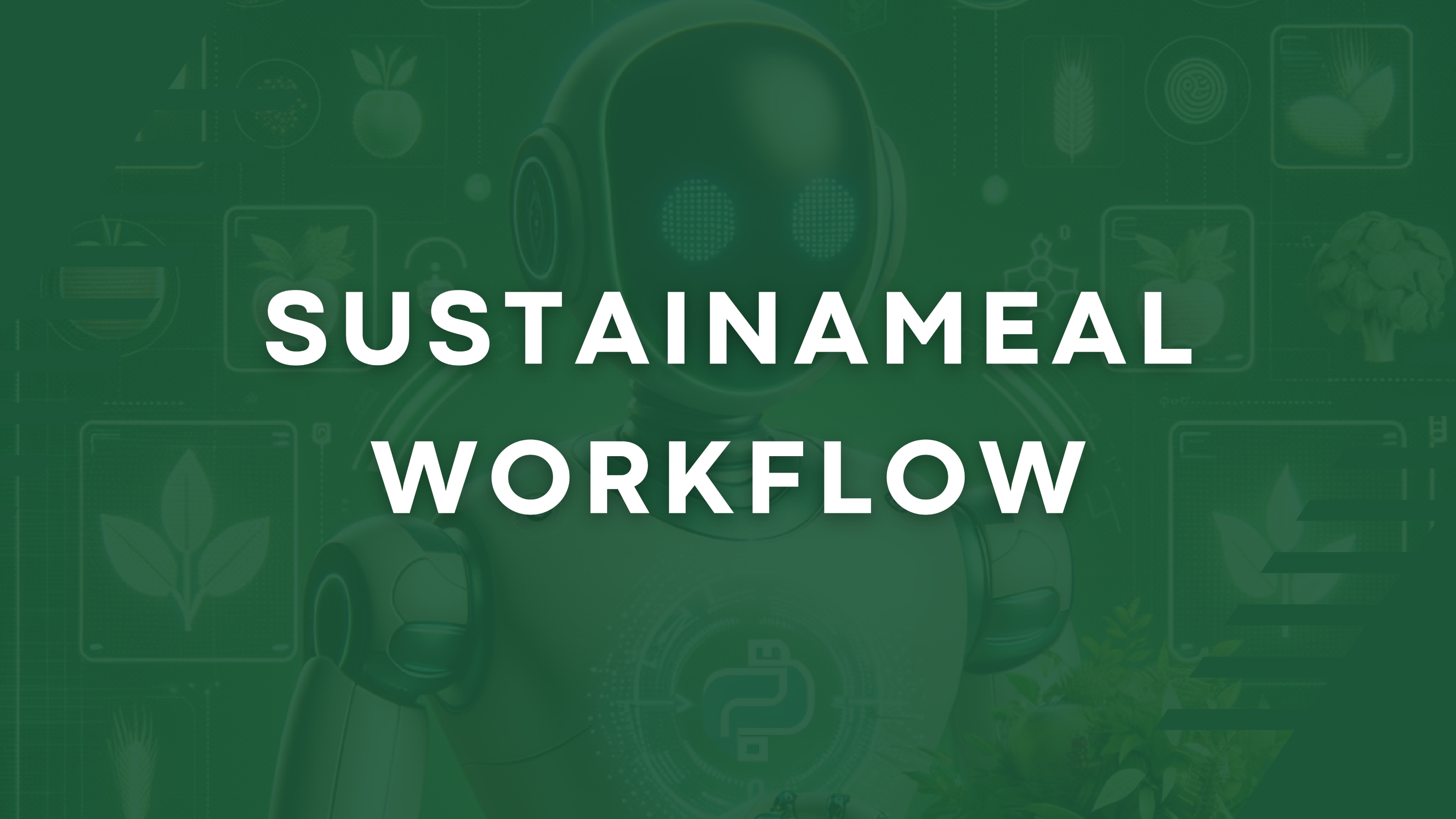
Medium(1)

Representing moderately sustainable recipes(0.1 < score < 0.5)

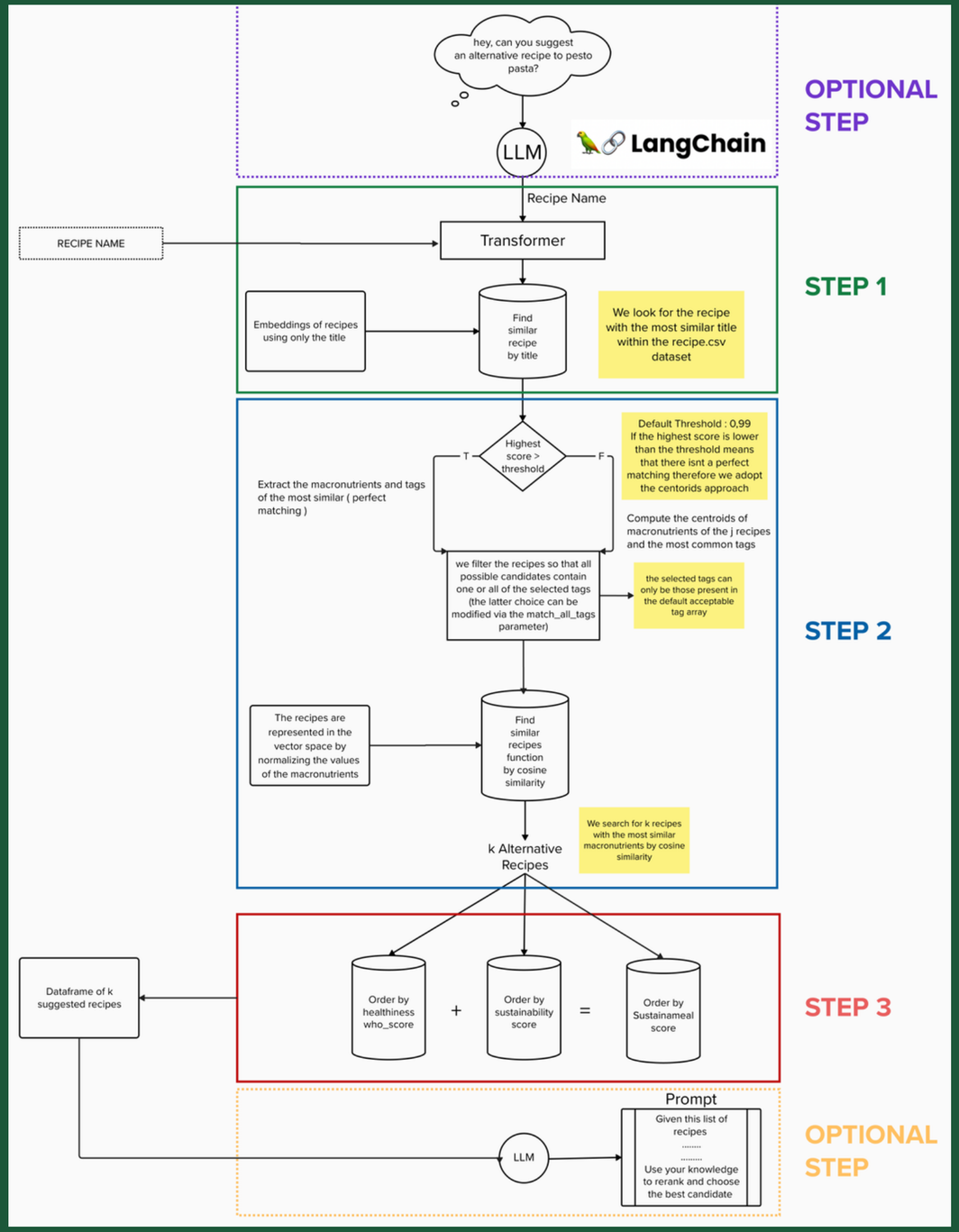
High(2)

Indicating highly sustainable recipes(score<=0.1)

*In order to guarantee uniformity with the other scores in our library , we inverted this score (1-score) in the initialization phase of the library


SUSTAINAMEAL
WORKFLOW

WORKFLOW



LET'S ANALYZE IT STEP BY STEP

INITIALIZATION

“WHAT DO I NEED?”

INITIALIZATION

01

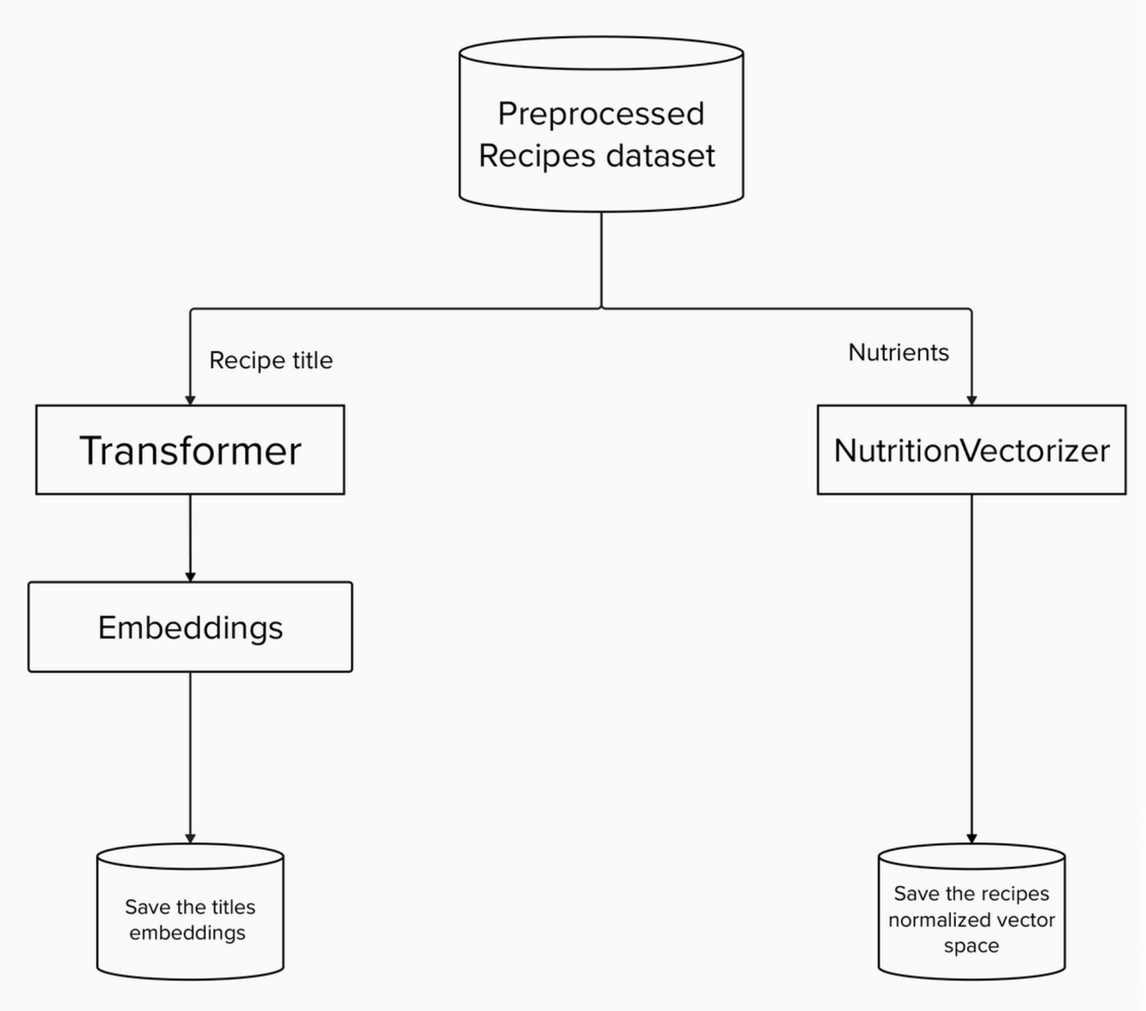
We compute the title embeddings

02

We compute the recipes vector space through the nutrients

03

*In order to guarantee uniformity with the other scores in our library , we inverted this score (1-score) in the initialization phase of the library



STEP 1

“DID I KNOWN THE RECIPE?”

STEP 1

01

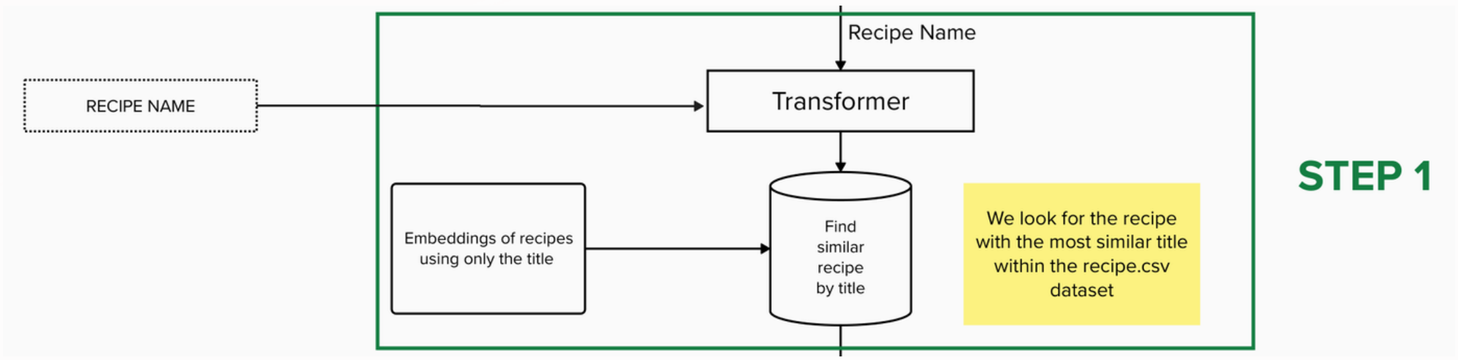
We get the recipe title

02

We use the transformer to obtain the title embedding

03

We look for k the recipe with the most similar embedding within our dataset



STEP 2

“LET’S FIND SOME SIMILAR RECIPE”

STEP 2

01

Check if we already known the recipe

02

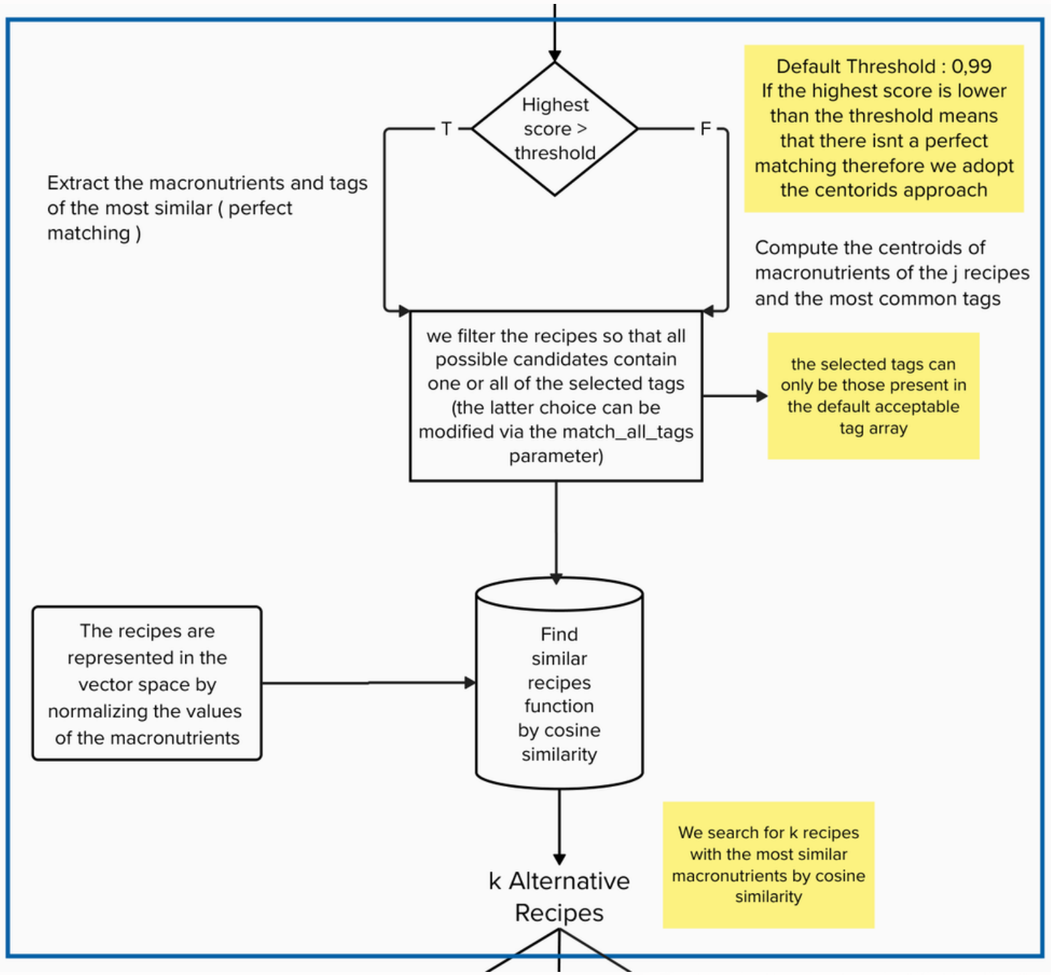
We get the macronutrients and the tags

03

We filter the recipes by tags

04

We get the k most similar recipes by macronutriens



STEP 2

STEP 3

“LET’S ORDER THE RECIPES FOR SUSTAINABILITY AND HEALTHINESS”

STEP 3

01

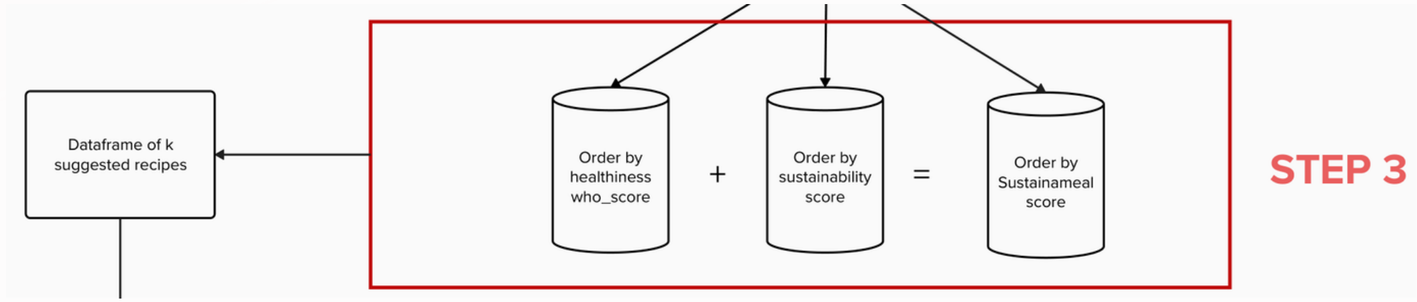
We reorder the recipes by healthiness

02

We reorder the recipes by sustanability scores

03

We reorder the recipes by Sustainameal score



$$\text{sustainameal\_score} = \text{sustainability\_score}(*) \cdot \alpha + \text{who\_score} \cdot \beta$$


OPTIONAL STEP
“WHICH RECIPE IS THE BEST?”

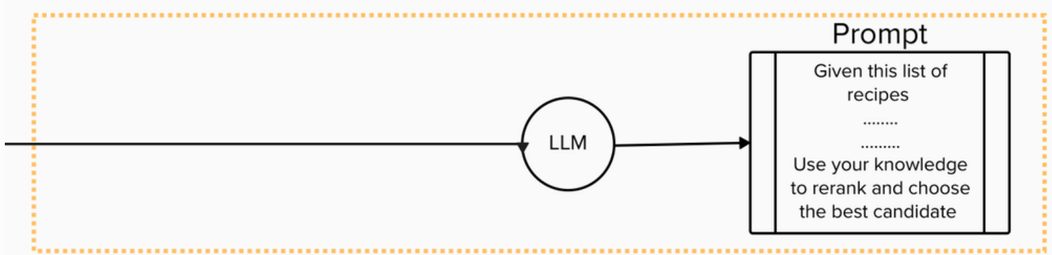
OPTIONAL STEP

01

Given a Dataframe of suggested recipes order by sustainameal score

02

We use the **Gpt-3.5-turbo** to rerank the recipes and to get the top 1



OPTIONAL STEP

```
{
Prompt used:
Using your knowledge, please rank (if necessary) the following
recipes from most to least recommended based on a balance
of sustainability and healthiness:

1. Recipe:Healthy Salad
2. Recipe: Quinoa Bowl
3. Recipe: Veggie Stir-Fry
}
Which one should I choose? Return just the name.
```


OPTIONAL STEP
“PRIMOARDIAL AGENT”

OPTIONAL STEP

As optional step we can initialize a primordial version (developed with LangChain) of an agent which can handle a generic question such as “ Can you suggest an alternative recipe to pesto pasta?”

OPTIONAL STEP

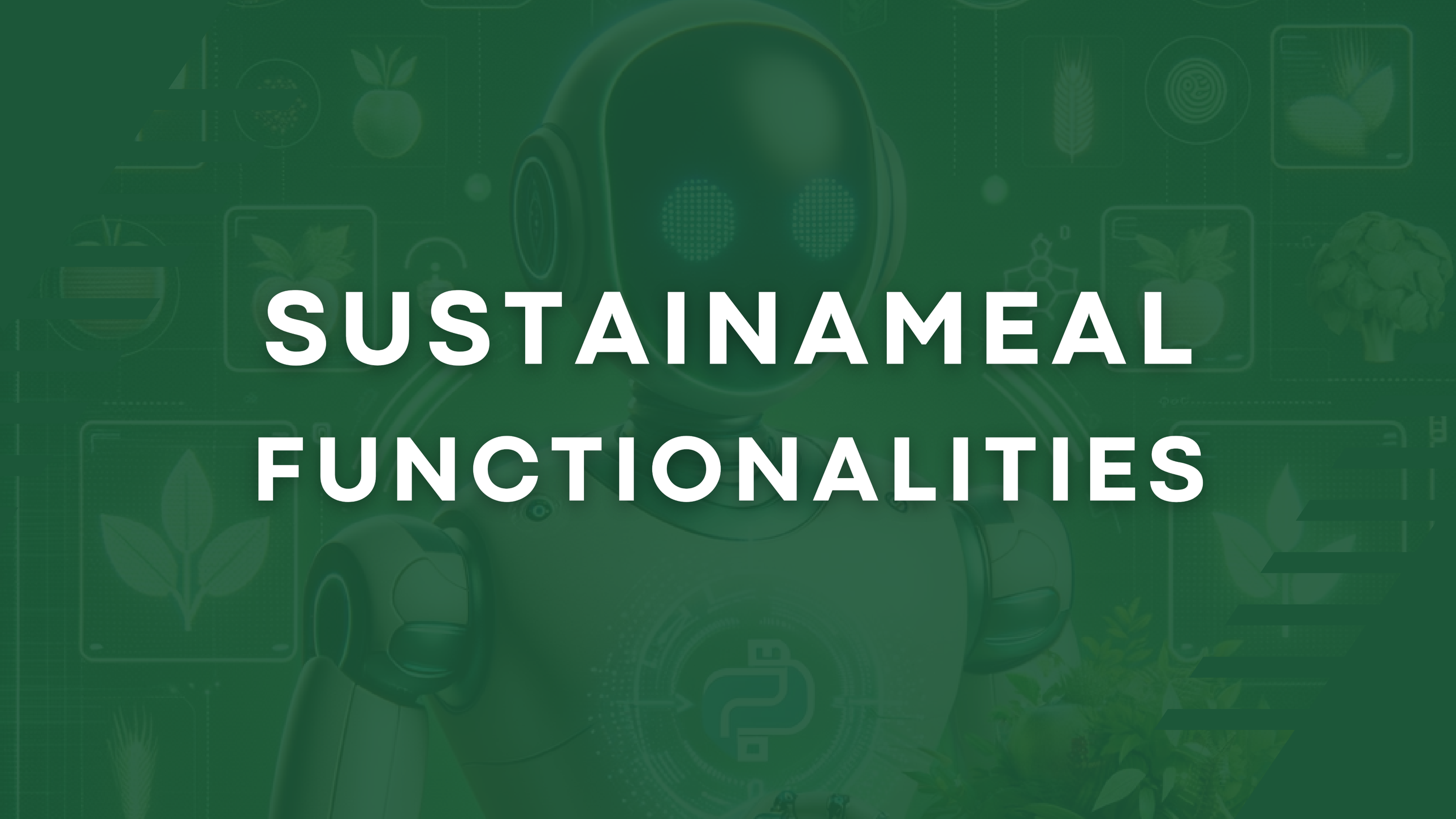
LANGCHAIN

LangChain's flexible abstractions and extensive toolkit unlocks developers to build context-aware, reasoning LLM applications.

Tools are interfaces that an agent can use to interact with the world. They combine a few things:

1. The name of the tool
2. A description of what the tool is
3. Schema of what the inputs to the tool are
4. The function to call
5. Whether the result of a tool should be returned directly to the user

```python

class CurrentRecipeInput(BaseModel):
    """Inputs for AlternativeSustainableRecipeTool"""
    recipe: str = Field(description="Name of the recipe")


class AlternativeSustainableRecipeTool(BaseTool):
    name = "AlternativeSustainableRecipeTool"
    description = """
        Useful when you want to get alternative recipes to a given recipe. This tool return a an alternative recipe.
        """

    args_schema: Type[BaseModel] = CurrentRecipeInput


    get_alternative_recipe: Callable = None

    def __init__(self, **data):
        super().__init__(**data)
        self.get_alternative_recipe = data.get('get_alternative_recipe_func')

    def _run(self, recipe: str):
        recipes_response = self.get_alternative_recipe(recipe)
        return recipes_response
```



SUSTAINAMEAL FUNCTIONALITIES

Functionalities

- Find similar recipes by title
- Find similar recipes by nutriens
- Order by Healthiness
- Order by Sustainability
- Order by SustainaMeal score
- Gpt ReRank
- Primordial Conversational Agent
- CMD scripts
- Save and Load Data

EXPERIMENTS

Experiments

We conducted experiments on 4 distinct scenario

1

Scenario 1
100 recipes with a WHO score above average

2

Scenario 2
100 recipes labeled as 1 (moderately sustainable)

3

Scenario 3
100 recipes labeled as 2 (highly unsustainable)

4

Scenario 4
Unknown recipes

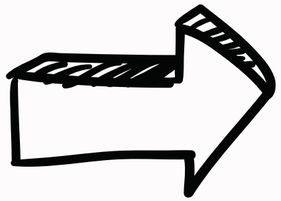
Different Configurations

K=1 , 10 , 50 , 100

Number of candidates

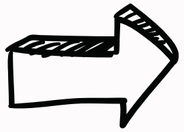
nutrients (A) = [Calories [cal] , Total Fat [g], Sodium [mg],Dietary Fiber [g] , Sugars [g], Protein [g]]

nutrients (B) = Calories [cal] , Total Fat [g] , Saturated Fat [g] ,Cholesterol [mg] , Sodium [mg] , Dietary Fiber [g] , Sugars [g] , Protein [g]



Nutrients for the vector space computation

match_all_tags=True , False



Filtering Strategy

5000

Experiments

Scenario 1

In this first scenario, we are working with consists of recipes that are already moderately healthy.

01

With a low value of k, in terms of healthiness, our algorithm tends to suggest recipes that do not result in a real increase in nutritional values but rather a decrease, unlike the sustainability

02

Increasing k the results improved

03

k = 1 makes the library unusable

	k	nutrients	match_all_tags	H_inc_mean	S_inc_mean	SM_inc_mean
0	1	B	False	-0.61	2.38	1.70
1	1	B	True	-0.33	1.17	0.75
2	1	A	False	-2.94	0.47	-0.22
3	1	A	True	-2.49	1.04	0.23
4	10	B	False	-2.20	1.42	0.55
5	10	B	True	-2.07	0.11	-0.50
6	10	A	False	-3.82	-0.75	-1.49
7	10	A	True	-3.54	-1.13	-1.74
8	50	B	False	3.41	17.72	15.27
9	50	B	True	4.27	16.60	14.44
10	50	A	False	3.19	17.66	15.20
11	50	A	True	3.58	17.13	14.78
12	100	B	False	6.38	19.92	17.53
13	100	B	True	6.94	19.34	17.08
14	100	A	False	5.51	20.19	17.67
15	100	A	True	6.08	19.50	17.12

Scenario 2

In this second scenario, where we are considering recipes that are moderately sustainable

01

There is an increase in all possible configurations, except for the configurations with k = 1 and nutrients of type B

	k	nutrients	match_all_tags	H_inc_mean	S_inc_mean	SM_inc_mean
0	1	B	False	-4.08	0.05	-0.52
1	1	B	True	-4.69	1.67	0.85
2	1	A	False	1.16	0.67	0.10
3	1	A	True	2.02	1.90	1.26
4	10	B	False	18.29	2.93	2.34
5	10	B	True	18.41	3.51	2.82
6	10	A	False	7.86	1.69	1.13
7	10	A	True	7.39	2.43	1.72
8	50	B	False	50.94	18.90	17.55
9	50	B	True	52.40	18.96	17.72
10	50	A	False	52.49	19.68	18.33
11	50	A	True	52.88	19.14	17.87
12	100	B	False	64.54	21.98	20.70
13	100	B	True	66.81	21.93	20.90
14	100	A	False	69.27	22.70	21.38
15	100	A	True	70.76	22.29	21.18

Scenario 3

In this scenario we used only unsustainable recipes

01

Employing solely unsustainable recipes, we will consistently experience increments in terms of sustainability across all conceivable parameter configurations

	k	nutrients	match_all_tags	H_inc_mean	S_inc_mean	SM_inc_mean
0	1	B	False	2.77	68.41	54.27
1	1	B	True	1.11	42.35	33.05
2	1	A	False	2.76	78.42	62.05
3	1	A	True	0.36	42.33	32.85
4	10	B	False	2.44	67.28	53.53
5	10	B	True	5.03	41.54	33.02
6	10	A	False	4.19	74.69	59.55
7	10	A	True	3.97	44.34	34.87
8	50	B	False	8.04	127.85	103.21
9	50	B	True	9.08	112.34	90.24
10	50	A	False	8.47	132.48	106.90
11	50	A	True	8.97	115.75	92.97
12	100	B	False	10.84	137.63	111.52
13	100	B	True	13.43	130.75	105.69
14	100	A	False	12.70	139.03	112.89
15	100	A	True	12.47	132.46	106.86

Scenario 4

In this scenario we considered only recipes which are not present in our dataset

01

As observed from the table, when considering recipes not present in our dataset, with k=1, there are decreases in terms of healthiness in almost all combinations, unlike sustainability.

02

With a k = 10, there is an increment for the healthiness, unlike the sustainability.

03

With a k > 10, there is an increment in all configurations both in the healthiness and in the sustainability

	k	nutrients	match_all_tags	H_inc_mean	S_inc_mean	SM_inc_mean
0	1	B	False	-3.00	2.58	1.83
1	1	B	True	-2.91	-0.20	-0.56
2	1	A	False	-3.00	2.58	1.83
3	1	A	True	2.64	0.46	0.76
4	10	B	False	2.96	-3.25	-2.69
5	10	B	True	3.93	-1.43	-0.98
6	10	A	False	1.51	-4.88	-4.36
7	10	A	True	3.20	-4.56	-3.85
8	50	B	False	13.12	14.95	14.47
9	50	B	True	13.79	14.93	14.54
10	50	A	False	12.98	14.84	14.50
11	50	A	True	12.67	15.07	14.66
12	100	B	False	15.77	18.18	17.64
13	100	B	True	17.66	18.06	17.76
14	100	A	False	16.43	17.87	17.51
15	100	A	True	17.78	17.90	17.71

Best Configuration

K=50

nutrients = [Calories [cal] , Total Fat [g], Sodium [mg],Dietary Fiber [g] , Sugars [g], Protein [g]]

tags = [’appetizers’, ’main-dish’, ’side-dishes’, ’drinks’, ’beverages’, ’fruits’, ’desserts’,’breakfast’, ’pasta-rice-and-grains’, ’rice’, ’pasta’,’pizza’, ’breads’, ’meat’, ’fish’,’seafood’, ’beef’, ’chicken’, ’vegetarian’]

match_all_tags=false

+19%

average increase in healthiness

+45%

average increase in sustanability

+38%

average increase on sustainameal score


EXPERIMENTS
WITH RERANKING

GPT Reranking

The following table illustrates the average increases in metrics for the top-ranked recipe.

We conducted two sets of comparisons:
the first set with k = 10 for both (to assess the extent of improvement lost when GPT is asked to rerank),

and the second set comparing the

results obtained without reranking at k = 1 against reranking with k = 10.

01

The data in the table shows that applying GPT reranking with k = 10 results in a significant increase in the average healthiness , sustainability scores and the sustainameal score, but certainly lower than the average increases of the top 1 recipes of the scenarios where reranking is not used

Top_H_inc_mean	Top_S_inc_mean	Top_SM_inc_mean	gpt_rerank	k
3.26	71.33	56.07	True	10
6.22	129.08	103.87	False	10
2.77	68.41	54.27	False	1

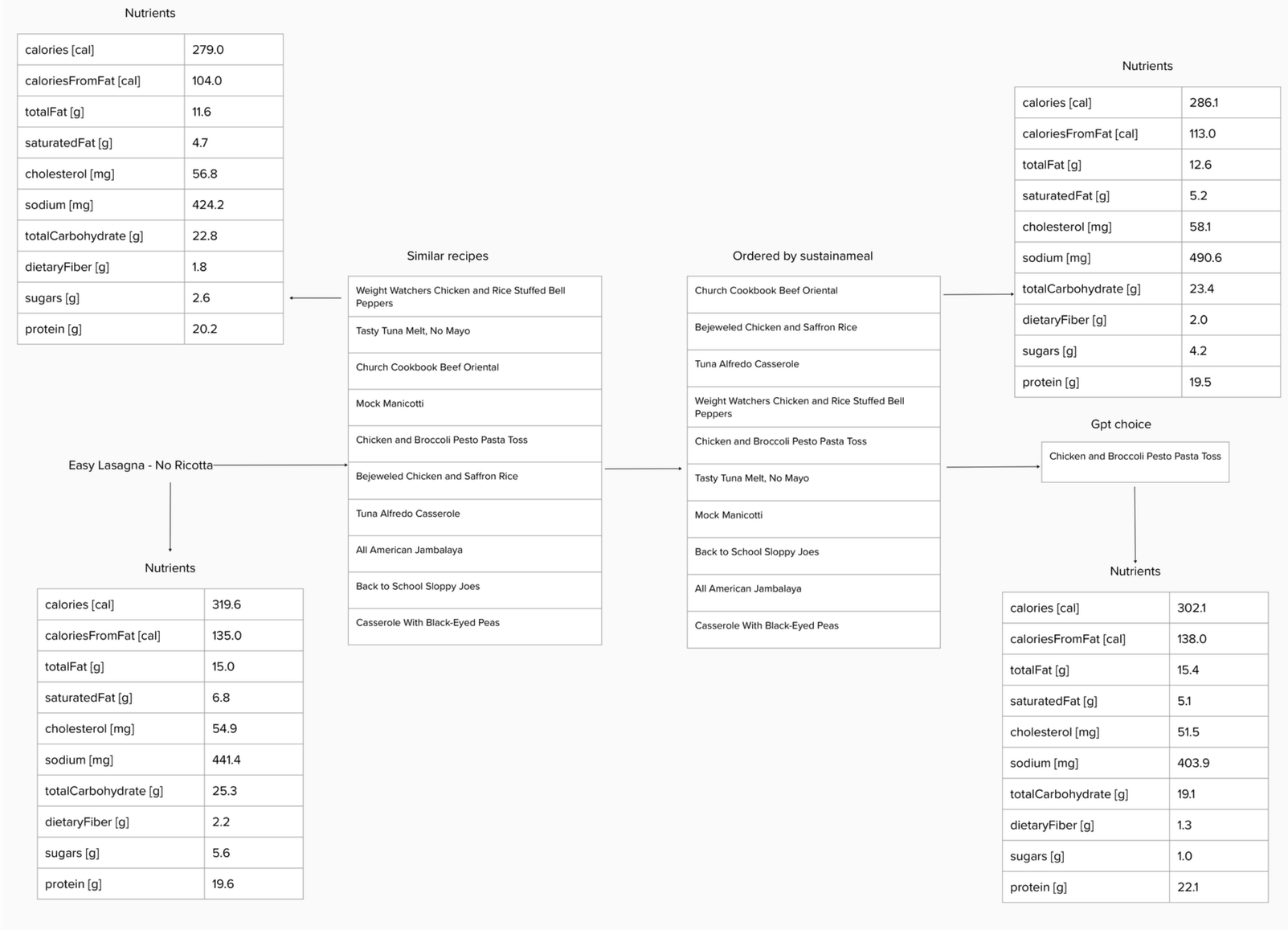
02

In contrast, without reranking at k = 1, the average increases
for healthiness and sustainability are lower than when reranking with k = 10, suggesting
that as we expected, reducing the number of candidates is not a recommended practice


QUALITATIVE
ANALYSIS

Qualitative Analisis

1 Easy Lasagna - No Ricotta



Qualitative Analisis

2 Spicy Homemade Hamburger Helper

Nutrients

calories [cal]	228.8
caloriesFromFat [cal]	77.0
totalFat [g]	8.6
saturatedFat [g]	3.3
cholesterol [mg]	71.4
sodium [mg]	723.0
totalCarbohydrate [g]	6.7
dietaryFiber [g]	0.1
sugars [g]	1.3
protein [g]	29.9

Similar recipes

- Fillet Bubbly Bake
- Korean Fire Pork- Toeji Kogi Kui
- Quick and Easy Crock Pot Beef Tips
- Tasty 2-Step Chicken
- Lemon Pepper Chicken Stew
- Round Steak With Beer Marinade for the Grill
- Chinese Pork With Shallots
- Skillet Herb Roasted Chicken
- Zaney's Sesame Chicken
- Lady's Italian Baked Tilapia

Spicy Homemade Hamburger Helper

Nutrients

calories [cal]	297.4
caloriesFromFat [cal]	128.0
totalFat [g]	14.3
saturatedFat [g]	5.0
cholesterol [mg]	73.7
sodium [mg]	643.8
totalCarbohydrate [g]	15.7
dietaryFiber [g]	1.1
sugars [g]	11.5
protein [g]	25.4

Ordered by sustainameal

- Round Steak With Beer Marinade for the Grill
- Chinese Pork With Shallots
- Korean Fire Pork- Toeji Kogi Kui
- Zaney's Sesame Chicken
- Skillet Herb Roasted Chicken
- Tasty 2-Step Chicken
- Lemon Pepper Chicken Stew
- Lady's Italian Baked Tilapia
- Fillet Bubbly Bake
- Quick and Easy Crock Pot Beef Tips

Nutrients

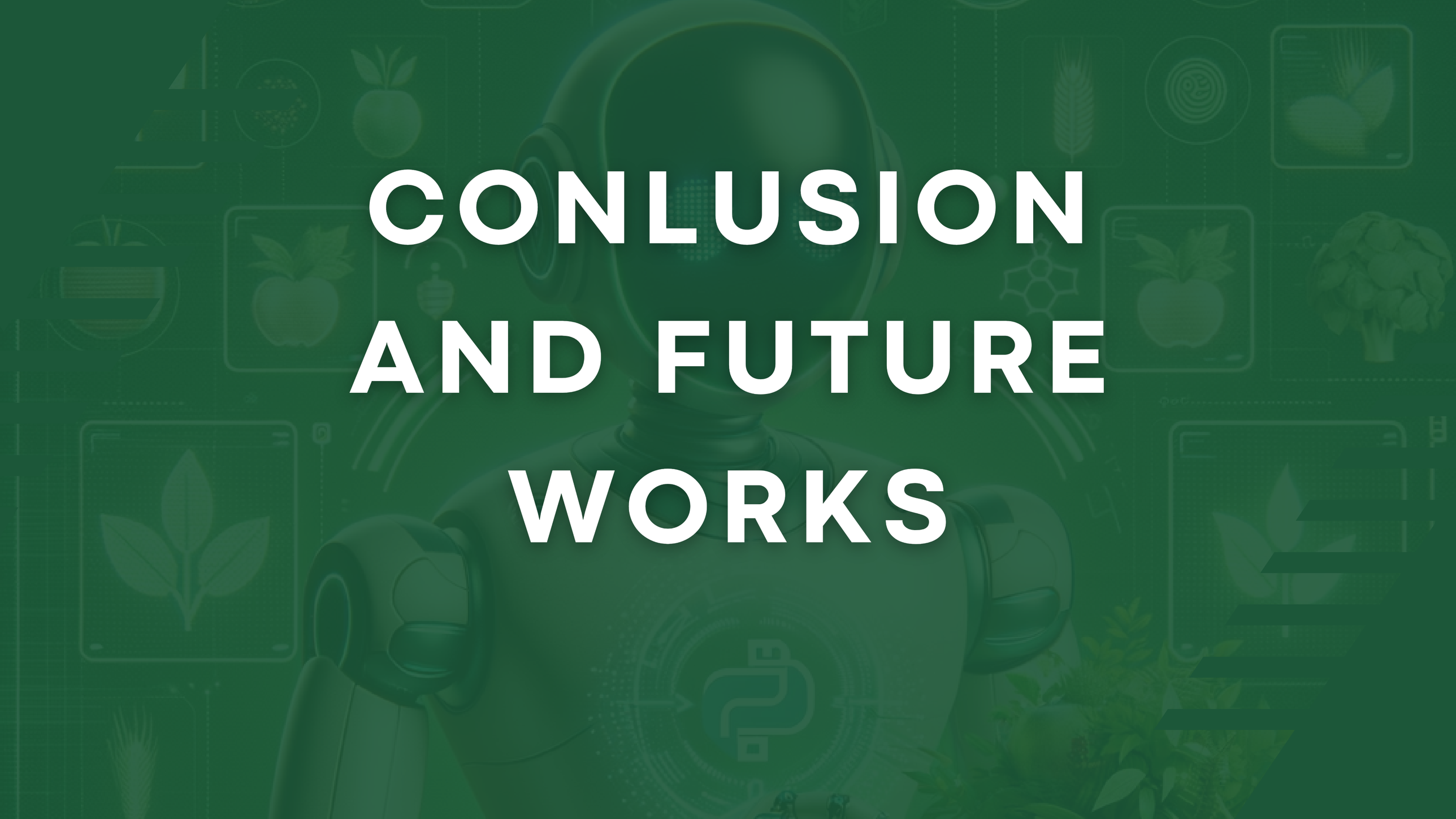
calories [cal]	262.1
caloriesFromFat [cal]	96.0
totalFat [g]	10.7
saturatedFat [g]	4.2
cholesterol [mg]	69.2
sodium [mg]	563.1
totalCarbohydrate [g]	10.8
dietaryFiber [g]	0.5
sugars [g]	6.8
protein [g]	25.9

Gpt choice

Chinese Pork With Shallots

Nutrients

calories [cal]	266.5
caloriesFromFat [cal]	110.0
totalFat [g]	12.2
saturatedFat [g]	2.7
cholesterol [mg]	78.8
sodium [mg]	570.6
totalCarbohydrate [g]	9.0
dietaryFiber [g]	0.4
sugars [g]	2.9
protein [g]	28.2

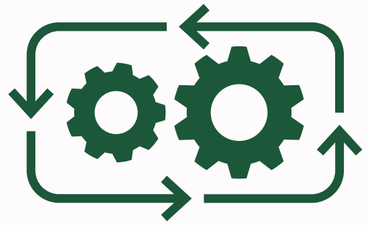


CONLUSION AND FUTURE WORKS

Conlusion and Future Works

01

Improve dataset preprocessing with a more refined data cleaning process to achieve
even more accurate and reliable results

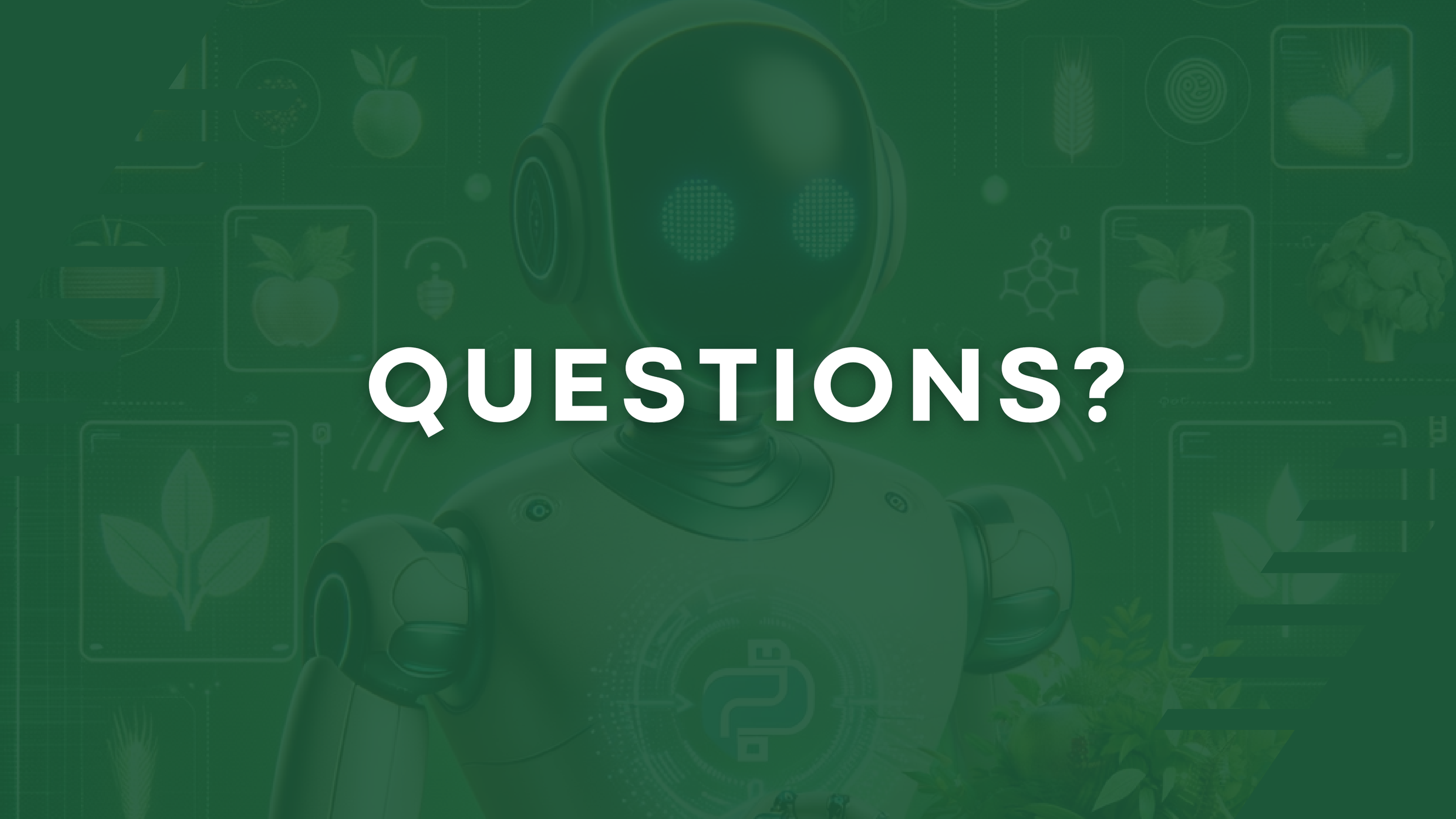


Implement a query engine to justify the results obtained by the system, enhancing transparency and understanding of the decision-making process

02

03

Explore further insights into alternative research, including new algorithmic approaches or the integration of additional data sources to further improve the diversity
and quality of generated proposals.


QUESTIONS?

THANK YOU

Giovanni Tempesta
Michele Di Carlo