An empirical study on the use of Large Language Models (LLMs) to translate recipes into a semi-structured data format

Vansh Nagpal[1], Siva Likitha Valluru[1], Biplav Srivastava[1]
[1]University of South Carolina, Computer Science and Engineering, AI 4 Society Group

Funders: University of South Carolina, South Carolina Honors College, SCRA

https://github.com/vnagpal25/BEACON/tree/main/items_

Introduction – Recipe Translation

Background:

- Previous works have introduced the *Rich Recipe Representation(**R3**)* to synthesize recipe information Converting recipes from plain text to **R3** is a tedious process and involves a human to ensure correctness and completeness of the translation

Objective:

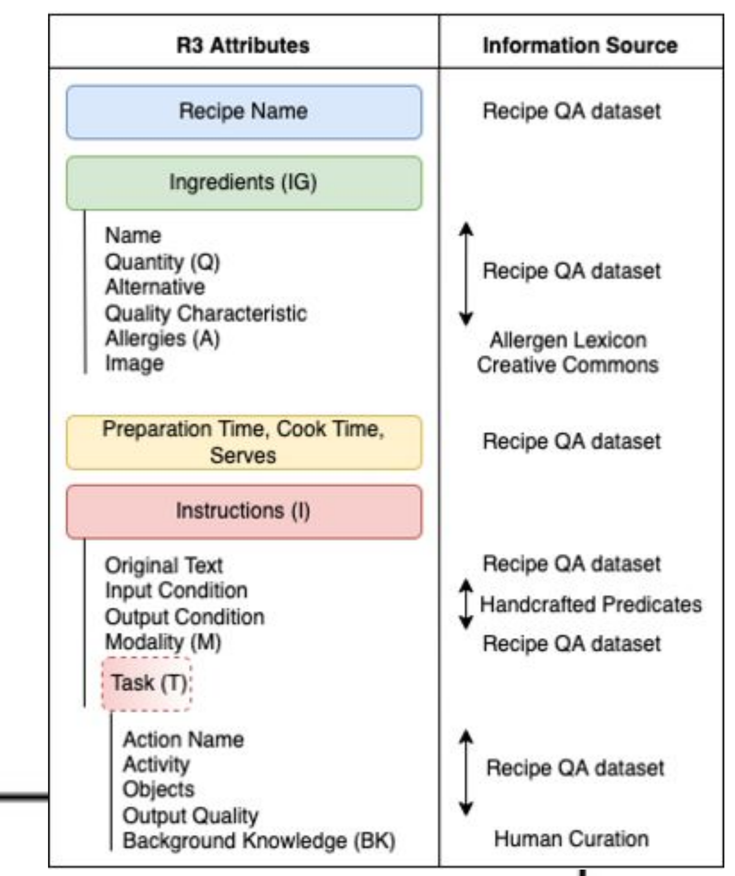
- Build novel methods to automate the conversion of recipes from plain text to **R3** using a combination of rule-based and Large Language Model (LLM)-based approaches

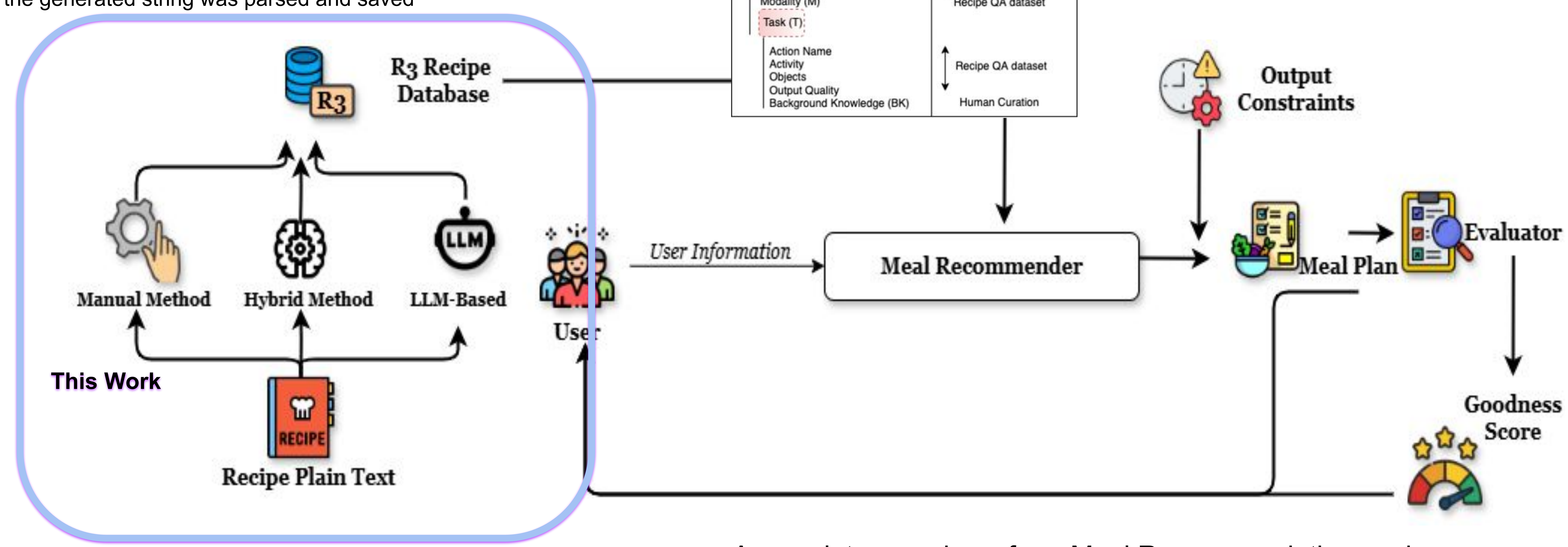
Significance:

- Aids in providing personalized meal recommendations
- Solves unstructured language to domain-specific language problem

Approach and System Overview

- *RC0* (Fully manual): 25 egg-based recipes were converted manually to **R3** representation
- *RC1 (Hybrid LLM-based)*: Ingredients and Instructions were extracted separately using Chain of Thought (CoT) prompting with GPT-3 and GPT-4o and manually collated into **R3** format
- *RC2 (Automatic LLM-based):*
    - Initial experiments were carried out to determine optimal configuration (n-shot prompting, atomicity of example, temperature) -> **Only one config. yielded a valid string**
    - Entire recipe text was passed to the LLM (Mixtral-8x7B Instruct) and the generated string was parsed and saved





A complete overview of our Meal Recommendation work

Findings

Metric	Definition
JSON Error Counts	Assesses number of errors(missing commas, extra braces, …) in the **R3** representation
Syntactic Similarity	Assesses the JSON structure of the **R3** representation to see if necessary tags are present
Semantic Similarity	Assesses if semantic information (meaning) is preserved

Config	JSON Error Counts	Syntactic Similarity	Semantic Similarity
RC0	0	0.584	**0.763**
RC1	0	**0.778**	0.733
RC2	26	0.376	0.758

Discussion

- Results show that *RC0* achieves the highest semantic accuracy but a lower syntactic score due to fewer JSON keys.
- *RC1* introduced extra JSON keys in the R3s, thereby inflating syntactic scores.
- *RC2* proved infeasible as LLMs frequently produced invalid JSONs with multiple errors.
- Future Work:
    - Making *RC2* more robust by refining prompting strategies and rule-based post-processing
    - Experimenting with larger LLMs with more parameters and larger context lengths

Resources

Resources

Translated Recipe Data:

- Checkout the resulting recipes from our recipe translation efforts here

Group Recommendation Effort:

- Checkout information about our broader group recommendation (including meal recommendation) effort here

Papers:

- *Nagpal, Vansh, et al. "BEACON: Balancing Convenience and Nutrition in Meals With Long-Term Group Recommendations and Reasoning on Multimodal Recipes." arXiv preprint arXiv:2406.13714 (2024).*
- *Nagpal, Vansh, et al. "A Novel Approach to Balance Convenience and Nutrition in Meals With Long-Term Group Recommendations and Reasoning on Multimodal Recipes and its Implementation in BEACON." arXiv preprint arXiv:2412.17910 (2024).*

https://ai4society.github.io/projects/group_rec/index.html

An empirical study on the use of Large Language Models (LLMs) to translate recipes into a semi-structured data format

Vansh Nagpal[1], Siva Likitha Valluru[1], Biplav Srivastava[1]
[1]University of South Carolina, Computer Science and Engineering, AI 4 Society Group

Funders: University of South Carolina, South Carolina Honors College, SCRA

Introduction – Recipe Translation

Background:

- Previous works have introduced the *Rich Recipe Representation(**R3**)* to synthesize recipe information Converting recipes from plain text to **R3** is a tedious process and involves a human to ensure correctness and completeness of the translation

Objective:

- Build novel methods to automate the conversion of recipes from plain text to **R3** using a combination of rule-based and Large Language Model (LLM)-based approaches

Significance:

- Aids in providing personalized meal recommendations
- Solves unstructured language to domain-specific language problem

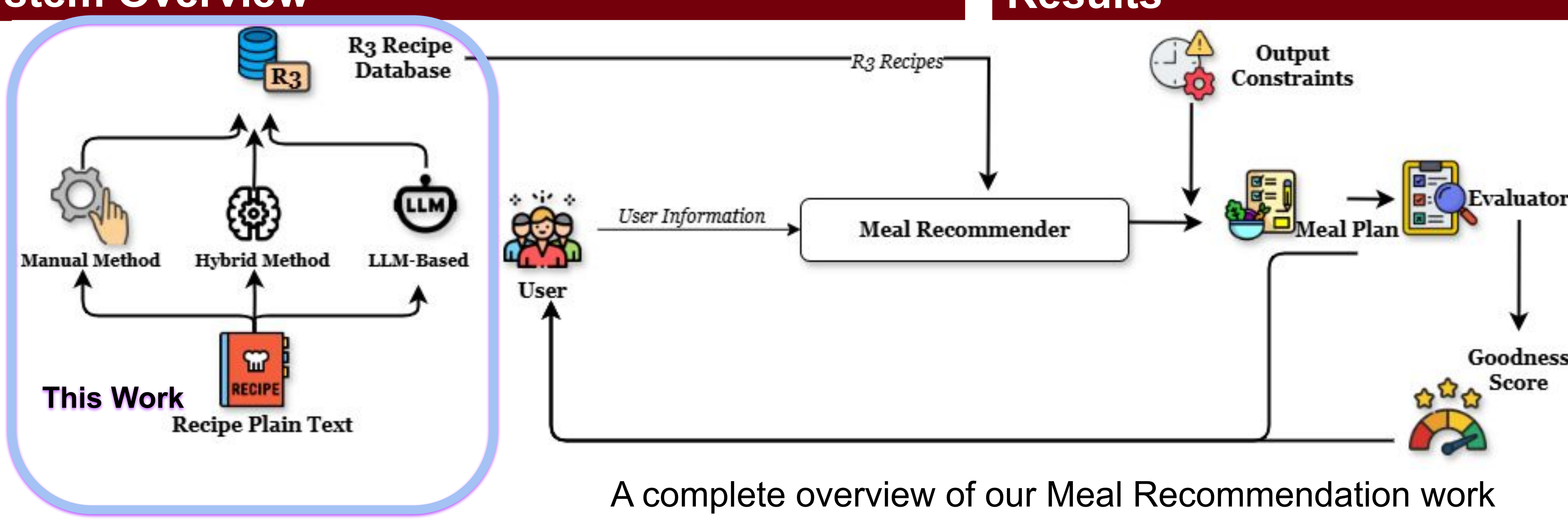
Methodology & Experimental Design

Implementation: We consider the following methods and metrics to evaluate the quality of **R3** representations

- *RC0* (Fully manual): 25 egg-based recipes were converted manually to **R3** representation
- *RC1 (Hybrid LLM-based)*: Ingredients and Instructions were extracted separately using Chain of Thought (CoT) prompting with GPT-3 and GPT-4o and manually collated into **R3** format
- *RC2 (Automatic LLM-based):*
  - Initial experiments were carried out to determine optimal configuration (n-shot prompting, atomicity of example, temperature) -> **Only one config. yielded a valid string**
  - Entire recipe text was passed to the LLM (Mixtral-8x7B Instruct) and the generated string was parsed and saved

Metric	Definition
JSON Error Counts	Assesses number of errors(missing commas, extra braces, …) in the **R3** representation
Syntactic Similarity	Assesses the JSON structure of the **R3** representation to see if necessary tags are present
Semantic Similarity	Assesses if semantic information (meaning) is preserved

System Overview



A complete overview of our Meal Recommendation work

Results

Config	JSON Error Counts	Syntactic Similarity	Semantic Similarity
RC0	0	0.584	**0.763**
RC1	0	**0.778**	0.733
RC2	26	0.376	0.758

Comparison of three **R3** translation methods

Discussion and Future Work

Discussion:

- Results show that *RC0* achieves the highest semantic accuracy but a lower syntactic score due to fewer JSON keys.
- *RC1* introduced extra JSON keys in the R3s, thereby inflating syntactic scores.
- *RC2* proved infeasible as LLMs frequently produced invalid JSONs with multiple errors.
- Future Work:
  - Making *RC2* more robust by refining prompting strategies and rule-based post-processing
  - Experimenting with larger LLMs with more parameters and larger context lengths

Resources

Translated Recipe Data:

- Checkout the resulting recipes from our recipe translation efforts here

Group Recommendation Effort:

- Checkout information about our broader group recommendation (including meal recommendation) effort here

https://github.com/vnagpal25/BEACON/tree/main/items_data

https://ai4society.github.io/projects/group_rec/index.html

Related preprints:

- *Nagpal, Vansh, et al. "BEACON: Balancing Convenience and Nutrition in Meals With Long-Term Group Recommendations and Reasoning on Multimodal Recipes." arXiv preprint arXiv:2406.13714 (2024).*
- *Nagpal, Vansh, et al. "A Novel Approach to Balance Convenience and Nutrition in Meals With Long-Term Group Recommendations and Reasoning on Multimodal Recipes and its Implementation in BEACON." arXiv preprint arXiv:2412.17910 (2024).*