Veganizing Recipes with Large Language Models

By Jonathan Sigh Musso
Funding By UPCH Green Solutions Center Supervised by Daniel Hershcovich and Morten A. Rasmussen

The case for adopting a plant-based diet is compelling. It offers both personal health benefits and broader environmental advantages, which are becoming increasingly essential. However, the widespread adoption of this diet faces challenges due to unfamiliarity with plant-based ingredients and their culinary uses.

In this project, we utilized Large Language Models (LLMs)—advanced machine learning models trained on billions of words—to transform traditional recipes into vegan versions, ensuring that the original dish's flavor, cultural essence, and aesthetic remained intact. To assess the versatility of this approach, we not only adapted recipes in English, where performance is anticipated to be optimal, but also in Danish, where limited resources pose challenges.

Image By Valeria Boltneva

Adaptions were well-received

The evaluation results indicate that the adaptations generated in English were well-received by the audience. In several metrics, the machine-generated adaptations surpassed the human adaptations, particularly in areas like appeal and creativity, by up to 14.7% and 15.6% on average[1]. Additionally, the machine-generated adaptations adhered well to the original recipe. The Danish recipes demonstrated slightly worse results, although still superior to the human adaptations, with 5.9% and 5.14%, respectively, in the same metrics. It faced challenges, particularly with complex or lengthy recipes.

This can be attributed to the LLMs' limited training in Danish. Both languages had issues with the model providing trivial substitutions. As such, while the initial outcomes are promising, further exploration on a larger evaluation scope—and perhaps even cooking the adaptations—is needed for a more definitive conclusion.

[1] Different setups Provide the best reults for different metrics

The technology behind the study

Traditionally, LLMs are refined for specific tasks by "fine-tuning" them on a smaller, task-specific dataset after initial training on a vast general dataset. This method, while effective, is resource-intensive and demands the creation of a comprehensive dataset—a task easier said than done.

Alternatively, using models like the GPT-3 variants, which are components of the renowned ChatGPT, enables "in-context learning" approaches. Here, the model is prompted with the task and, if needed, shown a few relevant examples. Both techniques were employed in this project.

Typically, natural language generation tasks are gauged using automatic metrics that compare outputs to a set of gold examples. However, this metric doesn't necessarily reflect the quality of an adaptation. Instead, we proposed metrics that assess various quality dimensions of the adaptations, evaluating the dish itself, as well as its fidelity as an adaptation.

These metrics were tested on a curated set of recipes that also included automatically matched human-made adaptations as an evaluation baseline, and assessed through crowdsourcing.

Examples of model outputs include adapting a chicken stew into a chickpea stew, chicken quesadillas into mushroom and spinach quesadillas, beef stuffed peppers into quinoa & vegetable stuffed peppers, etc.