УДК 004.832.22

COMPARATIVE ANALYSIS OF LARGE LANGUAGE MODEL-BASED SYSTEMS AND BARCODE-DRIVEN MOBILE APPLICATION FOR INGREDIENT-BASED FOOD QUALITY ASSESSMENT AND ITS HEALTH-RELATED IMPLICATIONS

Golikova V.A. [1] (student), **Kurmanova V.A.** [1] (student), **Boldina D.A.** [1] (student)
Scientific supervisor – PhD, Liliane Said[1]
[1]ITMO University
geometry345@mail.ru

Abstract
The increasing complexity of food compositions and misleading marketing practices make it difficult for consumers to assess product quality based on ingredient lists alone. While all necessary information is formally present on packaging, its proper interpretation requires specialized nutritional knowledge. This study compares the effectiveness of a traditional barcode-scanning application (Mira) and three large language models (Qwen, DeepSeek, and ChatGPT) in evaluating food product quality. Nine products from three categories (dairy products, beverages, canned goods) were analyzed using both approaches. The results demonstrated high consistency among the LLMs, which performed context-aware analysis, identified marketing discrepancies, and evaluated ingredient functionality approaching expert-level assessment. The Mira app showed some limitations, including database scarcity, recognition errors, and template-based analysis without contextual understanding. Each LLM exhibited functional specialization. The findings confirm that LLM-based systems can serve as effective alternatives to barcode-scanning applications. Practical implications include integrating AI-based photo analysis into existing apps or developing a Telegram bot for real-time product evaluation, which could be pilot-tested in university settings.
Keywords
large language models (LLMs), consumer health, food quality assessment, ingredient analysis, barcode-based mobile applications

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СИСТЕМ НА ОСНОВЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ И МОБИЛЬНОГО ПРИЛОЖЕНИЯ С ШТРИХКОД-СКАНИРОВАНИЕМ ДЛЯ ОЦЕНКИ КАЧЕСТВА ПИЩЕВЫХ ПРОДУКТОВ ПО СОСТАВУ И ИХ ВЛИЯНИЯ НА ЗДОРОВЬЕ

Голикова В.А. [1] (магистрант)**, Курманова В.А.** [1] (магистрант)**, Болдина Д.А.** [1] (магистрант)
Научный руководитель – доктор наук по Public Health Саид Л.И. [1]
[1]Университет ИТМО
geometry345@mail.ru

Аннотация
Возрастающая сложность состава пищевых продуктов и вводящие в заблуждение маркетинговые практики затрудняют для потребителей оценку качества продукции исключительно на основе списка ингредиентов. Хотя вся необходимая информация формально представлена на упаковке, её корректная интерпретация требует специализированных знаний в области питания. В данном исследовании сравнивается эффективность традиционного приложения для сканирования штрихкодов (Mira) и трёх больших языковых моделей (Qwen, DeepSeek и ChatGPT) при оценке качества пищевых продуктов. Девять продуктов из трёх категорий (молочная продукция, напитки, консервированные товары) были проанализированы с использованием обоих подходов.

Результаты продемонстрировали высокую согласованность между большими языковыми моделями, которые выполняли контекстно-ориентированный анализ, выявляли маркетинговые несоответствия и оценивали функциональную роль ингредиентов, приближаясь к экспертной оценке. Приложение Mira показало существенные ограничения, включая устаревание базы данных, ошибки распознавания и шаблонный анализ без учёта контекста. Каждая большая языковая модель продемонстрировала функциональную специализацию. Полученные результаты подтверждают, что системы на основе больших языковых моделей могут служить эффективной альтернативой приложениям для сканирования штрихкодов. Практические перспективы включают интеграцию ИИ-анализа фотографий в существующие приложения или разработку Telegram-бота для оценки продуктов в режиме реального времени, который может быть протестирован внутри университета.

Ключевые слова
большие языковые модели, здоровье потребителей, оценка качества пищевых продуктов, анализ состава продуктов, мобильные приложения на основе штрихкод-сканирования

Introduction

The issue of food quality and its impact on human health is well known. Despite the growing availability of information about ingredients, the compositions themselves are becoming more complex: new food additives and substitutes are appearing, the purpose and safety of which often remain unclear to consumers [1]. An additional risk factor is the marketing practices of manufacturers, where attractive positioning does not correspond to the actual content, which makes it difficult to make informed choices and can have a negative impact on health.

Although all the necessary information is formally present on the label, its proper interpretation requires special knowledge that not every consumer has. In this regard, the question of the reliability and completeness of the analysis offered by existing digital services designed to simplify the assessment of product quality for consumers becomes particularly relevant.

The development of reliable product evaluation systems based on evidence-based approaches is crucial; foreign experience shows that large language model (LLM)-based systems capable of analyzing photographs of ingredient lists in real time can become a promising alternative to traditional barcode-based database retrieval applications, which requires experimental verification [2].

Main part

The study was aimed at testing the hypothesis that LLM-based systems could effectively substitute traditional barcode-scanning applications in food quality assessment. As part of the experiment, nine products from three categories (dairy products, beverages, and canned goods) were analyzed using four tools: a Russian barcode-scanning application “Mira” and three large language models (Qwen, DeepSeek, and ChatGPT). Methodologically, two fundamentally different approaches were compared: the application's work based on database-driven barcode matching and the LLMs semantic interpretation of photographed labeling.

The results demonstrated a high degree of consistency between the LLMs: their assessments of most products coincided, especially with regard to products rated as high- or low-quality. This indicates the strong capacity of large language models to perform context-aware analysis of ingredients when provided with identical input data. The LLMs did not limit themselves to listing ingredients, but identified marketing plans, analyzed technological features, and assessed the conformity of the product name to its actual composition, which brings their conclusions closer to expert-level evaluation.

At the same time, the study revealed significant limitations in the approach used in the Mira app. In a number of cases, its assessments differed significantly from the conclusions of the LLMs, which may be associated with identification errors, outdated information about the

ingredients, and template-based analysis logic that does not take into account product context. The app demonstrated high speed and convenience for quick checks, but the formal classification of additives as harmful or safe and the lack of analysis of the functional role of ingredients reduced the reliability of the results.

A comparison of the LLMs allowed us to identify their functional specialization: Qwen showed the maximum detail and transparency of evaluation criteria, DeepSeek provided the most in-depth analysis of ingredient functions and potential risks, and ChatGPT demonstrated the optimal balance between analytical depth, structure clarity, and accessibility of explanations for the mass user. The operational principle of LLM-based systems relies on semantic parsing of ingredient lists using structured knowledge about food additives, production technologies, and regulatory frameworks, enabling evaluation of both individual components and their cumulative influence on the health-related characteristics of the final product.

Conclusion

A comparative study of the Mira application and the large language model-based systems (Qwen, DeepSeek, and ChatGPT) has revealed the following:

Firstly, using LLM-based systems can be an effective alternative to barcode-driven applications. The LLMs demonstrated a high degree of consistency in their results, as well as the ability to conduct contextual analysis of ingredients, including identifying manufacturer's marketing strategies and evaluating the technological role of ingredients.

Secondly, systemic limitations of the Mira application, which operates on the basis of barcode-database matching, were identified. The main drawbacks include: potential limits of the database restricting the analysis without consideration of the functional purpose of the ingredients; formal classification of additives that ignores the technological and contextual aspects of their use in a specific product.

Thirdly, the analysis indicates the presence of functional differentiation among the evaluated LLM-based systems. Qwen was characterized by comparatively higher transparency of evaluation criteria. DeepSeek demonstrated greater analytical depth in assessing ingredient roles and potential health-related risks. ChatGPT showed a balanced integration of structured reasoning and accessibility for mass users.

The results obtained can already be used in practice. For example, a function for uploading photographs of ingredient lists could be integrated into the Mira application for independent verification through an LLM-based system, thereby reducing the risk of outdated database information. Additionally, a free Telegram bot could be developed: users would photograph the ingredient list, and the bot (based on ChatGPT or DeepSeek) would provide a clear and accessible health-oriented evaluation of the product based on its ingredient profile. Such a system could be pilot-tested within a university environment with the involvement of students and staff.

References

1. Kotb R.E., Hallit S., Asmar F., et al. Socio-demographic influence of food additives acceptance and avoidance among the Lebanese population: A cross-sectional study // PLoS ONE. - 2025. - Vol. 20, № 6. - e0326629. - DOI: 10.1371/journal.pone.0326629.
2. Theodore Armand T.P., Nfor K.A., Kim J-I., Kim H.-C. Applications of Artificial Intelligence, Machine Learning, and Deep Learning in Nutrition: A Systematic Review // Nutrients. - 2024. - Vol. 16, № 7. - 1073. - DOI: 10.3390/nu16071073.