

Please site the original publication: <https://doi.org/10.1177/02601060251408541>

# Nutrition and Health

## Accuracy of Polyphenol Content Information in Berries: A Comparative Analysis of ChatGPT and Phenol-Explorer

Journal:	<i>Nutrition and Health</i>
Manuscript ID	NAH-25-0294.R2
Manuscript Type:	Original Article
Keywords:	Artificial intelligence, ChatGPT accuracy, Phenolic explorer, Polyphenols, Berries
Abstract:	<p><b>Background</b> The objective of this study is to evaluate and compare the accuracy of polyphenol content information in berries, as provided by two different sources: ChatGPT and Phenolic Explorer.</p> <p><b>Aim</b> This analysis aims to assess the reliability of two models ChatGPT-4o mini (free version) and ChatGPT-4o (paid version) in predicting polyphenol compound concentrations and their potential use in nutritional research and health applications.</p> <p><b>Methods</b> Seven different berries were selected for the study, and their anthocyanins, flavonols, phenolic acids, lignans, and stilbenes were queried in three different sessions using both ChatGPT-4o mini (free version) and ChatGPT-4o (paid version). The responses were compared with those from Phenolic Explorer, and the evaluation was based on relative accuracy (%).</p> <p><b>Results</b> No significant difference in relative accuracy (%) was found between ChatGPT-4o mini (<math>41.36 \pm 34.74</math>) and ChatGPT-4o (<math>46.23 \pm 34.01</math>) models (<math>p &gt; 0.05</math>; Cohen's <math>d = -0.107</math>). In ChatGPT-4o mini, the highest mean accuracy was observed for total polyphenols (<math>68.01 \pm 25.00\%</math>), followed by anthocyanins (<math>58.95 \pm 32.68\%</math>), while in ChatGPT-4o, anthocyanins showed the highest accuracy (<math>65.36 \pm 38.17\%</math>), followed closely by total polyphenols (<math>65.72 \pm 20.93\%</math>). Accuracy for flavonols, phenolic acids, and stilbenes was lower than for other compounds.</p> <p><b>Conclusion</b> This study shows that ChatGPT-4o mini and ChatGPT-4o exhibit varying accuracy in predicting polyphenols, with higher accuracy for common compounds like polyphenols and anthocyanins, and lower accuracy for flavonols, phenolic acids, and stilbenes.</p>

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# Accuracy of Polyphenol Content Information in Berries: A Comparative Analysis of ChatGPT and Phenol-Explorer

## Polyphenol Accuracy of ChatGPT

### Abstract

### Background

Polyphenols are widely occurring bioactive compounds in fruits and are extensively investigated for their potential health effects. The growing prominence of artificial intelligence tools in nutrition science necessitates evaluating their capacity to provide accurate biochemical data.

### Aim

This analysis aims to assess the reliability of two models ChatGPT-4o mini (free version) and ChatGPT-4o (paid version) in predicting polyphenol compound concentrations and their potential use in nutritional research and health applications.

### Methods

Seven different berries were selected for the study, and their anthocyanins, flavonols, phenolic acids, lignans, and stilbenes were queried in three different sessions using both ChatGPT-4o mini (free version) and ChatGPT-4o (paid version). The responses were compared with those from Phenolic Explorer, and the evaluation was based on relative accuracy (%).

### Results

No significant difference in relative accuracy (%) was found between ChatGPT-4o mini ( $41.36 \pm 34.74$ ) and ChatGPT-4o ( $46.23 \pm 34.01$ ) models ( $p > 0.05$ ; Cohen's  $d = -0.107$ ). In ChatGPT-4o mini, the highest mean accuracy was observed for total polyphenols ( $68.01 \pm 25.00\%$ ; significantly higher than flavonols,  $p < 0.01$ ), followed by anthocyanins ( $58.95 \pm 32.68\%$ ). In ChatGPT-4o, anthocyanins showed the highest accuracy ( $65.36 \pm 38.17\%$ ; significantly higher than flavonols,  $p < 0.01$ , and stilbenes,  $p < 0.001$ ) followed closely by total polyphenols ( $65.72 \pm 20.93\%$ ). Accuracy for flavonols, phenolic acids, and stilbenes was lower than for other compounds.

### Conclusion

This study shows that ChatGPT-4o mini and ChatGPT-4o exhibit varying accuracy in predicting polyphenols, with higher accuracy for common compounds like polyphenols and anthocyanins, and lower accuracy for flavonols, phenolic acids, and stilbenes.

**Keywords:** Artificial intelligence, ChatGPT accuracy, Phenolic explorer, Polyphenols.

## 1 Introduction

The potential of plant-based antioxidant sources in disease prevention is increasingly gaining attention, with their use in alternative therapeutic approaches becoming progressively widespread (Sawicki et al., 2024). Polyphenols, bioactive compounds naturally found in these sources and known for their antioxidant properties, play a significant role in the prevention of chronic diseases. Particularly, “berries” such as blueberries, black chokeberries, raspberries, strawberries, and blackberries stand out nutritionally due to their high polyphenol content. Berries are rich in polyphenols such as flavonoids, phenolic acids, tannins, stilbenes and lignans (Pap et al., 2021). Berries contribute to the prevention of health issues such as cardiovascular diseases, diabetes, cancer, and obesity by providing various health benefits attributed to their polyphenol content, including antioxidant, anti-inflammatory, and antimicrobial effects (Pap et al., 2021; Olas, 2018).

The exploration of the beneficial effects and optimal therapeutic doses of berry polyphenols remains an active field of research (Olas, 2018). Estimating dietary polyphenol intake is a complex process, mainly affected by the data sources, analysis methods, and food composition databases used (Lanuza et al., 2022). One of the most frequently referenced databases for such analyses today, especially regarding polyphenol content, is the Phenol-Explorer (PE) database (<https://phenol-explorer.eu>), recognized for its reliability (Xu et al., 2021).

With the impact of the digital era on health research, artificial intelligence (AI) tools and information systems have begun to offer novel solutions (Qarajeh et al., 2023). ChatGPT, developed by OpenAI, is an advanced conversational AI model that utilizes deep learning and natural language processing (NLP) techniques to generate human-like text responses (<https://chat.openai.com/chat>). It has been trained on vast amounts of internet data, enabling it to engage in coherent conversations on a wide range of topics. The model employs sophisticated techniques from NLP, Supervised Learning, and Reinforcement Learning to comprehend and generate text comparable to human-generated content (Roumeliotis and Tselikas, 2023). While ChatGPT-4o mini is available for free, the enhanced ChatGPT-4o model, with improved performance and greater capability to handle complex tasks, is accessible only through subscription. Both models, with their extensive knowledge bases and advanced algorithms, can assist in the analysis of metabolomic data and dietary evaluations (Lo et al., 2024; Hieronimus et al., 2024; O’Hara et al., 2024; Qarajeh et al., 2023; Bayram and Ozturkcan, 2024; Hallal et al., 2023; Bayram et al., 2025). ChatGPT has the potential to inform users about dietary requirements and serve as a reliable food assessment tool for health professionals (Ponzo et al., 2024; Hieronimus et al., 2024; Haman et al., 2024; Bayram and Ozturkcan, 2024; Qarajeh et al., 2023). Such AI-based tools are increasingly applied in complex analyses within nutrition and health, including nutrient compounds (Hallal et al., 2023; Hieronimus et al., 2024; Ponzo et al., 2024; Kansaksiri et al., 2023). However, current AI tools also

1  
2  
3 exhibit certain limitations, such as lack of internet access and limited data updates, which can lead to  
4 informational gaps, particularly in scientific literature research, and potentially to inaccurate references  
5 (McGowan et al., 2023). Previous studies have shown that models like ChatGPT provide positive  
6 outcomes in basic nutritional analysis, such as energy and macronutrient data, but may provide  
7 incomplete or incorrect information regarding allergens, portion sizes, and meal planning (Bayram et  
8 al., 2025; Hoang et al., 2023; Niszczota and Rybicka, 2023; Papastratis et al., 2024).

9  
10  
11  
12  
13 In recent years, there has been a rise in human intervention studies exploring the anti-inflammatory  
14 properties of berries (Sarikaya et al., 2024; Prieto Martinez et al., 2024). Furthermore the use of the  
15 Phenol-Explorer database in research on dietary intake of polyphenols and flavonoids has significantly  
16 increased (Xu et al., 2021; Witkowska et al., 2015; Ivey et al., 2016; Anacleto et al., 2019; Yue et al.,  
17 2020). In this evolving field, the relevance of information on polyphenol content obtained through AI-  
18 driven models like ChatGPT is steadily increasing. Accurate predictions of polyphenol content in berries  
19 are essential for identifying effective doses in human intervention studies, thereby supporting  
20 sustainable nutrition practices. Nevertheless, a significant gap in the literature exists concerning the  
21 reliability and accuracy of AI-based models like ChatGPT in delivering information on polyphenols.  
22 This study aims to evaluate ChatGPT by comparing its outputs with the established Phenol-Explorer  
23 database to address this gap.

## 2 Materials and Methods

### 2.1 Study procedures

24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
This study evaluates the effectiveness of AI models in accurately gathering information on the  
polyphenol content of berries, a vital aspect of plant-based nutrition due to its positive effects on both  
health and the environment. Among the AI models analyzed, ChatGPT-4o mini and ChatGPT-4o were  
included. The obtained data were compared with the values in the Phenol-Explorer 3.6 database.

### 2.2 Selection of Berries and Polyphenols

The berries included in this study were selected based on a comprehensive review of human studies  
investigating the role of dietary polyphenols in inflammation, that used different forms such as extracts,  
fresh berries, juices, or dried powders (Joseph et al., 2014; Heneghan et al., 2018; Del Bo et al., 2015).  
In addition, a systematic search was conducted across major databases (PubMed, Web of Science, and  
Scopus) to determine the berries used in studies that evaluated their effects on inflammatory biomarkers  
and antioxidant enzyme activity in humans. Accordingly, well-characterized berries such as blueberries  
(Bowtell et al., 2017; Riso et al., 2013), blackcurrants (Okamoto et al., 2020), cranberries (Christman et  
al., 2024) and strawberries (Schell et al., 2019; Huang et al., 2021) which belong to the Ericaceae and  
Rosaceae families and are known to be rich sources of bioactive compounds, as well as less commonly

1  
2  
3 studied but biologically significant berries including bilberries (Chan et al., 2021), black chokeberries  
4 (Chamberlin et al., 2024; Xie et al., 2017) and black raspberries (Schell et al., 2019) were included due  
5 to their documented health benefits and common consumption (Skrovankova et al., 2015; Calvano et  
6 al., 2019). Furthermore, as comparative analyses utilized the Phenol-Explorer database, only berries  
7 with comprehensive and detailed polyphenol profiles including anthocyanins, flavonols, phenolic acids,  
8 lignans, and stilbenes were selected. Berries' key phenolic classes, including flavonoids “anthocyanins”  
9 and “flavonols”, “phenolic acids”, “lignans”, and “stilbenes”, were examined (Olas, 2018). Moreover,  
10 since comparative analyses utilized the Phenol-Explorer database, only berries with comprehensive and  
11 detailed profiles of these polyphenol classes were selected.  
12  
13  
14  
15  
16  
17

### 18 2.3 Data used from Phenolic Explorer

19  
20  
21 The polyphenol content was obtained from Phenol-Explorer under the "Fruits-Berries" category. Data  
22 for anthocyanins were derived using the pH differential method, while data for flavonols, phenolic acids,  
23 and stilbenes were obtained through chromatography. Lignan data were based on chromatography  
24 analysis following hydrolysis. Polyphenol data were derived using the Folin assay method. The values  
25 for flavonols, phenolic acids, stilbenes, and lignans were calculated by summing the compounds  
26 identified under specific subcompound headings (e.g., kaempferol, hydroxybenzoic acids, resveratrol,  
27 lariciresinol). In cases where multiple records were available for flavonoids, phenolic acids, stilbenes,  
28 and lignans within our dataset, values were chosen based on those derived from the largest sample size  
29 (N), characterized by lower variability (smaller standard deviation), and situated within the reported  
30 range (minimum–maximum).  
31  
32  
33  
34  
35  
36  
37

### 38 2.4 Evaluation of AI Model Performance and Repeated Analyses

39  
40 The prompt used to obtain responses from AI models was: “Provide the most precise values, rather than  
41 a range, for the anthocyanin, flavonoid, phenolic acid, total polyphenol, stilbene, and lignan contents of  
42 this berry separately (for 100 g): .....?” (For example, "Provide the most precise values, rather than a  
43 range, for the anthocyanin, flavonoid, phenolic acid, total polyphenol, stilbene, and lignan contents of  
44 this berry separately (for 100 g): Black chokeberry”).  
45  
46  
47  
48

49 ChatGPT-4o mini and ChatGPT-4o was used in three query sessions: the first on October 25, 2024, the  
50 second on November 9, 2024, and the third on November 25, 2024. The same prompt was used during  
51 all three query sessions.  
52  
53

54  
55 For stilbenes, when the content amount provided by the ChatGPT models was less than 0.1, this value  
56 was excluded from the accuracy analysis. In cases where ChatGPT models did not provide certain  
57  
58  
59  
60

polyphenol content during the query sessions, those values were marked as "NA" and were not queried again.

Due to frequent updates of AI models, a two-week interval was used to minimize the risk of significant model changes and account for short-term fluctuations in AI responses caused by temporary technical factors (Qarajeh et al., 2023). To reduce the possibility of variation in the results and analyze consistency, the entire methodology was repeated three times, with a two-week gap between each application.

## 2.5 Statistical analysis

All statistical analyses were conducted using SPSS Statistics, Version 30.00. For each type of berry, total polyphenols such as anthocyanins, flavonols, phenolic acids, lignans, and stilbenes were compared using the reference values from Phenolic Explorer and the values provided by ChatGPT models. This comparison was evaluated by calculating the relative accuracy percentage (Tofallis, 2015) using the following steps:

1. Mean Absolute Error =  $|\text{Chatgpt answer} - \text{phenolic Explorer data}|$
2. Mean Absolute Percentage Error (%) =  $\text{Mean Absolute Error} / \text{phenolic Explorer} \times 100$
3. Relative Accuracy (%) =  $100 - \text{Error} (\%)$

Values for relative accuracy percentage were expressed as mean and standard deviation. Negative relative accuracy percentages were winsorized to 0 to maintain interpretability. (Sharma and Chatterjee, 2021) Normality of accuracy percentages for phenolic compounds was assessed using the Shapiro-Wilk test, which indicated non-normal distribution for most variables ( $p < 0.05$ ). Therefore, accuracy values, expressed as percentages ranging from 0 to 100, were transformed using the arcsine square root transformation. After transformation, normality was reassessed and confirmed (Shapiro-Wilk,  $p > 0.05$ ).

Analysis of variance (ANOVA) was used to compare accuracy percentages across different polyphenol types for each model, with Tukey's post-hoc test used for post hoc analysis, adjusted using the Bonferroni correction to control for multiple comparisons. P-values were annotated according to significance levels as  $p < 0.05$ .

A post-hoc power analysis was conducted using G\*Power 3.1.9.4 software. Based on a point-biserial correlation model, with a small effect size ( $d = 0.18$ ),  $\alpha = 0.05$ , and a total sample size of 182, the achieved statistical power was calculated as 0.79.

### 3 Results

For this study, a total of 252 data points were obtained from seven berry species ( $n = 7$ ), six phenolic compounds ( $n = 6$ ), three query sessions ( $n = 3$ ) and two models ( $n = 2$ ), of which 191 were found in Phenol Explorer. Additionally, 23 data points were inaccessible in the analyses conducted with ChatGPT, and accuracy (%) evaluations were performed on a total of 182 data points. Polyphenol contents in berries according to phenolic explorer and chatgpt models (mg/100 g fw) is shown in Table 1.

[insert Table 1]

The relative accuracy (%) of the ChatGPT models' polyphenol content information on berries, according to the Phenolic Explorer, is presented in Table 2. In the ChatGPT-4o mini model, the highest mean accuracy was observed for the total polyphenols group ( $68.01 \pm 25.00\%$ ), followed by anthocyanins ( $58.95 \pm 32.68\%$ ) while the ChatGPT-4o model showed the highest accuracy for anthocyanins ( $65.36 \pm 38.17\%$ ), followed closely by total polyphenols ( $65.72 \pm 20.93\%$ ). In the ChatGPT-4o mini model, the accuracy for total polyphenols was significantly higher than that for flavonols ( $p < 0.01$ ). In the ChatGPT-4o model, anthocyanins were significantly more accurate than flavonols ( $p < 0.01$ ) and stilbenes ( $p < 0.001$ ).

[insert Table 2]

The relative accuracy (%) results for berry types are presented in Table 3. The highest mean accuracy was in ChatGPT-4o mini for bilberry at  $45.57 \pm 41.72$ , while for ChatGPT-4o, blueberry showed a mean of  $52.69 \pm 34.19$ . In ChatGPT-4o mini, although a statistically significant difference was found across berry types ( $F = 2.892$ ,  $p = 0.015$ ), this difference did not remain significant after Bonferroni-adjusted post-hoc comparisons. Similarly, for ChatGPT-4o, the overall differences between berry types were not statistically significant ( $F = 1.701$ ,  $p = 0.134$ ).

[insert Table 3]

The relative accuracy results (%) of ChatGPT models polyphenol content information on berries, according to the Phenolic Explorer and categorized by query sessions, are summarized in Table 4. For both the ChatGPT-4o mini and ChatGPT-4o models, there was no statistically significant difference in accuracy percentages across query sessions ( $p > 0.05$ ).

[insert Table 4]

1  
2  
3 No significant difference in relative accuracy (%) was found between ChatGPT-4o ( $46.23 \pm 34.01\%$ )  
4 and ChatGPT-4o mini ( $41.36 \pm 34.74\%$ ); the effect size was negligible (Cohen's  $d = -0.107$ ). The  
5 relative accuracy (%) of ChatGPT models in predicting polyphenol content in berries is shown in Figure  
6 1, while Figure 2 presents the relative accuracy (%) across different berry types.  
7  
8  
9

10 [insert Figure 1]  
11

12  
13 [insert Figure 2]  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

## 4 Discussion

This study evaluated the performance of ChatGPT-4o mini and ChatGPT-4o models in predicting polyphenols present in berries. No significant difference in relative accuracy (%) was found between ChatGPT-4o mini ( $41.36 \pm 34.74\%$ ) and ChatGPT-4o ( $46.23 \pm 34.01\%$ ) models. However, the average accuracy of approximately 40–46% indicates that the predictive reliability of both models remains low and suggests that ChatGPT cannot yet be considered a dependable tool for nutritional data extraction or composition estimation. The highest median accuracy (%) in ChatGPT-4o mini was found for total polyphenols ( $68.01 \pm 25.00\%$ ), while in ChatGPT-4o, anthocyanins showed the highest accuracy ( $65.36 \pm 38.17\%$ ). Although these values show relatively better performance for certain compounds, they should be interpreted with caution due to the overall low model accuracy. Accuracy for phenolic acids and flavonols was lower than that for the other compounds. Additionally, no significant difference was found between berry types in accuracy (%) for ChatGPT models' or query session ( $p > 0.05$ ). Additionally, these results emphasize that despite areas of relatively improved prediction, the overall performance remains insufficient for practical application.

ChatGPT, developed by OpenAI, is widely used for generating text across various domains, including nutrition-related inquiries (Chatelan et al., 2023; Kirk et al., 2023; Hoang et al., 2023). While previous research has shown its potential in predicting nutrient content, the present study reveals clear limitations and inconsistencies in its performance. The ChatGPT models examined in this study demonstrated relative accuracies of  $41.36 \pm 34.74\%$  for ChatGPT-4o mini and  $46.23 \pm 34.01\%$  for ChatGPT-4o; however, these rates were lower than those reported in studies assessing AI models' performance in nutrient predictions (Qarajeh et al., 2023; Bayram and Ozturkcan, 2024; Hieronimus et al., 2024). While its accuracy in potassium classification exceeds 80%, its performance in protein quality classification ranges between 50% and 54% (Qarajeh et al., 2023; Bayram and Ozturkcan, 2024). This suggests that the model's accuracy in nutritional predictions is influenced by food complexity and the analytical methods employed (Zhang et al., 2020; Aiumtrakul et al., 2024; Bayram and Ozturkcan, 2024; Naja et al., 2024; Qarajeh et al., 2023; Hieronimus et al., 2024). This contrast further strengthens the conclusion that ChatGPT's predictive capacity is highly variable and may not generalize well to phytochemical estimation.

Our study found both models showed higher relative accuracy in predicting total polyphenols and anthocyanins. This could be attributed to the broader and more inclusive classification of total polyphenols, which encompasses various compounds such as phenolic acids, flavonols, anthocyanins, stilbenes, and lignans (Ciupei et al., 2024). Additionally, the extensive research and availability of data on anthocyanins may have contributed to the improved accuracy of ChatGPT in predicting this compound (Zhao et al., 2023; Gonçalves et al., 2021). Nevertheless, even in these better-performing

1  
2  
3 categories, accuracy levels remain moderate rather than strong, indicating that ChatGPT is not yet robust  
4 enough for reliable biochemical profiling.  
5  
6

7 The relative accuracy of flavonoids and phenolic acids was statistically lower than other polyphenols,  
8 likely due to the differences in analytical methods, which influence model accuracy. Techniques such  
9 as the Folin-Ciocalteu method and the pH differential method can measure total polyphenol and  
10 anthocyanin content more homogeneously, while chromatographic methods require the separation of  
11 individual compounds and make the process more complex (Blainski et al., 2013; Lee et al., 2012).  
12 These simpler methods may provide more reliable and consistent results than more complex  
13 chromatographic analyses (Lee et al., 2012; Taghavi et al., 2022). Notably, flavonoids are often found  
14 as glycosides, while some, like flavonols, are present only in aglycone form (Dias et al., 2021). The  
15 varying reporting of these compounds across different databases, based on hydrolyzed (aglycone form)  
16 and non-hydrolyzed (glycosides, aglycones, esters) chromatographic methods, could contribute to  
17 inconsistencies in ChatGPT's predictions (Lanuza et al., 2023). These findings collectively indicate that  
18 molecular complexity and database inconsistency significantly restrict model performance.  
19  
20  
21  
22  
23  
24  
25  
26

27 Lignans and stilbenes are compounds with limited data availability in both the Phenolic Explorer and  
28 ChatGPT models. In instances where data was missing in Phenolic Explorer, ChatGPT's response of  
29 "limited" was treated as having a 100% relative accuracy (%). Although ChatGPT-4o mini showed  
30 relatively high accuracy for lignans, the results exhibited significant variability, suggesting  
31 inconsistencies in predicting lignans. For stilbenes, ChatGPT-4o mini showed higher accuracy ( $45.31 \pm$   
32  $12.25$ ) than ChatGPT-4o ( $16.07 \pm 15.19$ ), suggesting that the latter's poor performance may stem from  
33 insufficient data or limitations in the model's classification abilities. Similar findings in the literature  
34 suggest that accurately predicting specific molecular groups remains a challenge (Bayram and  
35 Ozturkcan, 2024; Hallal et al., 2023). Moreover, the lack of comprehensive data on complex and novel  
36 food compositions further limits the accuracy of AI systems (Fritsch et al., 2017). Taken together, these  
37 outcomes indicate that ChatGPT demonstrates instability in compound-specific predictions and cannot  
38 consistently differentiate between polyphenol subclasses.  
39  
40  
41  
42  
43  
44  
45  
46

47 ChatGPT tends to provide more accurate predictions when trained on large datasets. However, its  
48 accuracy decreases when dealing with smaller, heterogeneous datasets, leading to less reliable  
49 inferences. Our study observed similar inconsistencies in flavonoid content predictions, where the model  
50 produced inconsistent results. The literature emphasizes that combining glycoside and aglycone forms  
51 of flavonoids can create reliability issues (Anacleto et al., 2019; Yue et al., 2020). The differing reporting  
52 standards across databases may have contributed to the observed inconsistencies in ChatGPT's  
53 predictions. Thus, despite its potential, dataset limitations present a substantial barrier to achieving  
54 dependable performance in phytochemical estimation.  
55  
56  
57  
58  
59  
60

1  
2  
3 Our findings indicate variability in accuracy across query sessions, consistent with previous reports of  
4 inconsistencies in ChatGPT's performance (Qarajeh et al., 2023; Bayram and Ozturkcan, 2024). This  
5 suggests that model performance can change, possibly due to variations in data quality or model updates.  
6 Such instability reduces reproducibility, further limiting the tool's current applicability for nutrition  
7 science.  
8  
9

10  
11  
12 ChatGPT has shown higher consistency and accuracy compared to other AI models (Bayram and  
13 Ozturkcan, 2024; Hieronimus et al., 2024), offering reasonably accurate nutritional information (Haman  
14 et al., 2024; Kansaksiri et al., 2023). Nevertheless, it has notable limitations. ChatGPT is prone to  
15 generating plausible yet incorrect responses, a phenomenon commonly referred to as "hallucination"  
16 (<https://platform.openai.com/docs/models>) (Bayram and Ozturkcan, 2024). Moreover, it was stated that  
17 ChatGPT may fabricate or distort facts, leading to unreliable references (Bayram and Ozturkcan, 2024).  
18 Similarly, in our study, discrepancies in the sources cited by ChatGPT-4o mini during different query  
19 sessions were noted, with varying references to publishers and organizations. This raises questions about  
20 whether these sources were accessed from open-access databases, as no additional information was  
21 provided to verify these details. It lacks a hierarchy of evidence-based sources, leading to the  
22 propagation of inaccuracies and non-scientific views, particularly in nutrition-related topics, and it  
23 struggles with establishing causal relationships, such as in clinical reasoning (Chatelan et al., 2023;  
24 Cascella et al., 2023). Therefore, while AI models may assist in preliminary screening or hypothesis  
25 generation, they should not serve as primary tools for scientific decision-making.  
26  
27  
28  
29  
30  
31  
32  
33  
34

35  
36 The findings of this study provide insight into the strengths and limitations of ChatGPT-4o mini and  
37 ChatGPT-4o in predicting polyphenol content in berries. One of the key strengths of this study is its  
38 systematic evaluation of ChatGPT's predictive accuracy regarding polyphenols in berries. This  
39 contributes to the growing literature on AI applications in nutrition science. The study also utilized a  
40 well-established reference database, Phenolic Explorer, to assess ChatGPT's performance, ensuring a  
41 strong comparison. However, the limited accuracy observed — averaging only 40–46% — highlights  
42 that ChatGPT is not yet suitable as a reliable analytical tool for phytochemical prediction. Additionally,  
43 variability in ChatGPT's responses across different query sessions and the reliance on a single database  
44 for validation remain notable limitations. Genetic differences, geographic origin, and post-harvest  
45 processing can substantially influence phenolic concentrations (Josuttis et al., 2012; Yang et al., 2013;  
46 Kårlund et al., 2014). Variations in climate, latitude, temperature, solar radiation, and humidity shape  
47 phenolic profiles (Josuttis et al., 2012; Yang et al., 2013). This heterogeneity poses challenges for  
48 nutrient databases and AI training datasets, potentially limiting performance and introducing  
49 bias into accuracy estimates (Busato et al., 2023; Li et al., 2021). Although Phenol-Explorer applies  
50 rigorous aggregation and quality control procedures to ensure data accuracy (Neveu et al., 2010;  
51 Rothwell et al., 2012, 2013), both biological and technical variability must be considered when  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 interpreting predictive outcomes. These factors further support the conclusion that current LLM-based  
4 models lack consistency for high-precision nutritional characterization.  
5  
6

7 Although the analysis of the ChatGPT-4o mini model indicated an overall signal of difference, these  
8 patterns were not supported by more detailed comparisons, likely due to limited sample size and high  
9 within-group variability. This mirrors previous findings demonstrating that initial statistical significance  
10 does not necessarily translate into meaningful group-level separation, particularly when assumptions or  
11 methodological constraints are not fully met (Sawyer, 2009). Likewise, the literature notes that different  
12 post-hoc procedures may yield inconsistent outcomes, further indicating the need for careful  
13 interpretation (Hilton and Armstrong, 2006). Taken together, the findings suggest a possible trend in  
14 performance variation across berry types; however, the evidence remains too weak to support confident  
15 discrimination between them. Accordingly, the results should be interpreted with caution, as the model  
16 does not yet provide the level of reliability required for precise classification or compound-specific  
17 differentiation.  
18  
19  
20  
21  
22  
23  
24  
25

26 A further limitation of this study is that the potential influence of prompt phrasing on output accuracy  
27 was not examined. This restricted sampling may limit the robustness of the reproducibility assessment.  
28 Additionally, the evaluation was limited to two LLM tools (ChatGPT-4o mini and ChatGPT-4o),  
29 without the inclusion of other large language models. While AI-driven models offer a promising tool  
30 for nutrition science, their accuracy currently below reliable analytical thresholds suggests that they  
31 should not be used as standalone sources of compositional data. Enhancing AI models with more  
32 comprehensive and standardized food composition databases could improve performance and reduce  
33 variability in predictions. While AI-powered models like ChatGPT show promise in generating dietary  
34 recommendations and nutrient classifications, their application in high-precision chemical analysis  
35 remains limited, and caution is required when interpreting AI-generated nutritional outputs.  
36  
37  
38  
39  
40  
41  
42

## 43 **5 Conclusion**

44  
45 This study reveals that the ChatGPT-4o mini and ChatGPT-4o models exhibit varying levels of accuracy  
46 in predicting polyphenols in berries. Higher accuracy rates were observed for compounds such as total  
47 polyphenols and anthocyanins, while accuracy issues were encountered with less common compounds  
48 like flavonoids and stilbenes. These findings indicate that AI-based tools may show variability in  
49 accuracy, depending on factors such as data quality and the chemical structure of the compounds.  
50 ChatGPT's performance is notably more successful with compounds that are better researched and have  
51 more data available, although accuracy issues persist for lesser-known compounds.  
52  
53  
54  
55  
56  
57

## 58 **Author Contributions/ CRediT**

59  
60 Conceptualization: BS, HKK

1  
2  
3 Data curation: Buse SARIKAYA  
4

5 Methodology: Buse SARIKAYA, Hüsna KAYA KAÇAR  
6

7 Formal Analysis: Buse SARIKAYA, Hüsna KAYA KAÇAR  
8

9 Investigation: Buse SARIKAYA, Hüsna KAYA KAÇAR  
10

11 Writing – Original Draft Preparation: Buse SARIKAYA  
12

13 Writing – Review & Editing: Hüsna KAYA KAÇAR  
14

15 Methodology: Buse SARIKAYA, Hüsna KAYA KAÇAR  
16

17 Supervision: Hüsna KAYA KAÇAR  
18

19 Visualization: Hüsna KAYA KAÇAR  
20

## 21 **Acknowledgments** 22

23 None.  
24

## 25 **Ethics Statement** 26

27 This research does not require ethical approval.  
28

## 29 **Conflicts of Interest** 30

31 None.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

- 1  
2  
3  
4  
5  
6 Aiumtrakul N, Thongprayoon C, Arayangkool C, et al. (2024) Personalized medicine in urolithiasis: AI  
7 chatbot-assisted dietary management of oxalate for kidney stone prevention. *Journal of*  
8 *Personalized Medicine* 14(1): 107.
- 9 Anacleto SL, Lajolo FM and Hassimotto NM (2019) Estimation of dietary flavonoid intake of the  
10 Brazilian population: a comparison between the USDA and Phenol-Explorer databases. *Journal*  
11 *of Food Composition and Analysis* 78: 1-8.
- 12 Bayram HcM and Ozturkcan A (2024) AI showdown: info accuracy on protein quality content in foods  
13 from ChatGPT 3.5, ChatGPT 4, bard AI and bing chat. *British Food Journal* 126(9): 3335-3346.
- 14 Bayram HM, Çelik ZM and Barcın Güzeldere HK (2025) Can artificial intelligence (AI) chatbot tools be  
15 used effectively for nutritional management in obesity? *Nutrition and Health*.  
16 02601060251329070.
- 17 Blainski A, Lopes GC and De Mello JCP (2013) Application and analysis of the folin ciocalteu method for  
18 the determination of the total phenolic content from *Limonium brasiliense* L. *Molecules* 18(6):  
19 6852-6865.
- 20 Bowtell JL, Aboo-Bakkar Z, Conway ME, et al. (2017) Enhanced task-related brain activation and resting  
21 perfusion in healthy older adults after chronic blueberry supplementation. *Applied Physiology,*  
22 *Nutrition, and Metabolism* 42(7): 773-779.
- 23 Busato S, Gordon M, Chaudhari M, et al. (2023) Compositionality, sparsity, spurious heterogeneity, and  
24 other data-driven challenges for machine learning algorithms within plant microbiome studies.  
25 *Current opinion in plant biology* 71: 102326.
- 26 Calvano A, Izuora K, Oh EC, et al. (2019) Dietary berries, insulin resistance and type 2 diabetes: An  
27 overview of human feeding trials. *Food & function* 10(10): 6227-6243.
- 28 Cascella M, Montomoli J, Bellini V, et al. (2023) Evaluating the feasibility of ChatGPT in healthcare: an  
29 analysis of multiple clinical and research scenarios. *Journal of medical systems* 47(1): 33.
- 30 Chamberlin ML, Peach JT, Wilson SM, et al. (2024) Polyphenol-rich *Aronia melanocarpa* fruit  
31 beneficially impact cholesterol, glucose, and serum and gut metabolites: A Randomized Clinical  
32 Trial. *Foods* 13(17): 2768.
- 33 Chan SW, Chu TT, Choi SW, et al. (2021) Impact of short-term bilberry supplementation on glycemic  
34 control, cardiovascular disease risk factors, and antioxidant status in Chinese patients with  
35 type 2 diabetes. *Phytotherapy Research* 35(6): 3236-3245.
- 36 Chatelan A, Clerc A and Fonta P-A (2023) ChatGPT and future artificial intelligence chatbots: what may  
37 be the influence on credentialed nutrition and dietetics practitioners? *Journal of the Academy*  
38 *of Nutrition and Dietetics* 123(11): 1525-1531.
- 39 Christman L, De Benedetto A, Johnson E, et al. (2024) Polyphenol-Rich Cranberry Beverage Positively  
40 Affected Skin Health, Skin Lipids, Skin Microbiome, Inflammation, and Oxidative Stress in  
41 Women in a Randomized Controlled Trial. *Nutrients* 16(18): 3126.
- 42 Ciupei D, Colișar A, Leopold L, et al. (2024) Polyphenols: From Classification to Therapeutic Potential  
43 and Bioavailability. *Foods* 13(24): 4131.
- 44 Del Bo C, Martini D, Porrini M, et al. (2015) Berries and oxidative stress markers: an overview of human  
45 intervention studies. *Food & function* 6(9): 2890-2917.
- 46 Dias MC, Pinto DC and Silva AM (2021) Plant flavonoids: Chemical characteristics and biological activity.  
47 *Molecules* 26(17): 5377.
- 48 Fritsch C, Staebler A, Happel A, et al. (2017) Processing, valorization and application of bio-waste  
49 derived compounds from potato, tomato, olive and cereals: A review. *Sustainability* 9(8): 1492.
- 50 Gonçalves AC, Nunes AR, Falcão A, et al. (2021) Dietary effects of anthocyanins in human health: A  
51 comprehensive review. *Pharmaceuticals* 14(7): 690.
- 52 Hallal K, Hamdan R and Tlais S (2023) Exploring the potential of AI-Chatbots in organic chemistry: An  
53 assessment of ChatGPT and Bard. *Computers and Education: Artificial Intelligence* 5: 100170.
- 54 Haman M, Školník M and Lošťák M (2024) AI dietitian: Unveiling the accuracy of ChatGPT's nutritional  
55 estimations. *Nutrition* 119: 112325.
- 56  
57  
58  
59  
60

- 1  
2  
3 Heneghan C, Kiely M, Lyons J, et al. (2018) The effect of berry-based food interventions on markers of  
4 cardiovascular and metabolic health: A systematic review of randomized controlled trials.  
5 *Molecular Nutrition & Food Research* 62(1): 1700645.  
6 Hieronimus B, Hammann S and Podszun MC (2024) Can the AI tools ChatGPT and Bard generate energy,  
7 macro-and micro-nutrient sufficient meal plans for different dietary patterns? *Nutrition*  
8 *Research* 128: 105-114.  
9 Hilton A and Armstrong RA (2006) Statnote 6: post-hoc ANOVA tests. *Microbiologist* 2006: 34-36.  
10 Hoang YN, Chen Y-L, Ho DKN, et al. (2023) Consistency and accuracy of artificial intelligence for  
11 providing nutritional information. *JAMA network open* 6(12): e2350367-e2350367.  
12 Huang L, Xiao D, Zhang X, et al. (2021) Strawberry consumption, cardiometabolic risk factors, and  
13 vascular function: a randomized controlled trial in adults with moderate  
14 hypercholesterolemia. *The Journal of Nutrition* 151(6): 1517-1526.  
15 Ivey KL, Croft K, Prince RL, et al. (2016) Comparison of flavonoid intake assessment methods. *Food &*  
16 *function* 7(9): 3748-3759.  
17 Joseph SV, Edirisinghe I and Burton-Freeman BM (2014) Berries: anti-inflammatory effects in humans.  
18 *Journal of agricultural and food chemistry* 62(18): 3886-3903.  
19 Josuttis M, Carlen C, Crespo P, et al. (2012) A comparison of bioactive compounds of strawberry fruit  
20 from Europe affected by genotype and latitude. *Journal of Berry Research* 2(2): 73-95.  
21 Kansaksiri P, Panomkhet P and Tantisuwichwong N (2023) Smart Cuisine: Generative recipe & ChatGPT  
22 powered nutrition assistance for sustainable cooking. *Procedia Computer Science* 225: 2028-  
23 2036.  
24 Kårlund A, Moor U, Sandell M, et al. (2014) The impact of harvesting, storage and processing factors  
25 on health-promoting phytochemicals in berries and fruits. *Processes* 2(3): 596-624.  
26 Kirk D, van Eijnatten E and Camps G (2023) Comparison of answers between ChatGPT and human  
27 dieticians to common nutrition questions. *Journal of Nutrition and Metabolism* 2023(1):  
28 5548684.  
29 Lanuza F, Bondonno NP, Zamora-Ros R, et al. (2022) Comparison of Flavonoid Intake Assessment  
30 Methods Using USDA and Phenol Explorer Databases: Subcohort Diet, Cancer and Health-Next  
31 Generations—MAX Study. *Frontiers in Nutrition* 9: 873774.  
32 Lanuza F, Zamora-Ros R, Petermann-Rocha F, et al. (2023) Advances in Polyphenol Research from Chile:  
33 a literature review. *Food Reviews International* 39(6): 3134-3171.  
34 Lee SG, Nam T-G, Kim D-O, et al. (2012) Comparison of analytical methods for anthocyanin  
35 quantification in berries: HPLC and pH differential methods.  
36 Li G, Zrimec J, Ji B, et al. (2021) Performance of regression models as a function of experiment noise.  
37 *Bioinformatics and Biology Insights* 15: 11779322211020315.  
38 Lo FP-W, Qiu J, Wang Z, et al. (2024) Dietary assessment with multimodal ChatGPT: A systematic  
39 analysis. *IEEE Journal of Biomedical and Health Informatics*.  
40 McGowan A, Gui Y, Dobbs M, et al. (2023) ChatGPT and Bard exhibit spontaneous citation fabrication  
41 during psychiatry literature search. *Psychiatry Research* 326: 115334.  
42 Naja F, Taktouk M, Matbouli D, et al. (2024) Artificial intelligence chatbots for the nutrition  
43 management of diabetes and the metabolic syndrome. *European Journal of Clinical Nutrition*  
44 78(10): 887-896.  
45 Niszczota P and Rybicka I (2023) The credibility of dietary advice formulated by ChatGPT: Robo-diets  
46 for people with food allergies. *Nutrition* 112: 112076.  
47 O'Hara C, Kent G, Flynn A, et al. (2024) Using artificial intelligence to estimate the nutritional content  
48 of meal photos: an evaluation of ChatGPT-4. *Proceedings of the Nutrition Society* 83(OCE4):  
49 E336.  
50 Okamoto T, Hashimoto Y, Kobayashi R, et al. (2020) Effects of blackcurrant extract on arterial functions  
51 in older adults: A randomized, double-blind, placebo-controlled, crossover trial. *Clinical and*  
52 *Experimental Hypertension* 42(7): 640-647.  
53 Olas B (2018) Berry phenolic antioxidants—implications for human health? *Frontiers in pharmacology*  
54 9: 78.  
55  
56  
57  
58  
59  
60

- 1  
2  
3 Pap N, Fidelis M, Azevedo L, et al. (2021) Berry polyphenols and human health: Evidence of antioxidant,  
4 anti-inflammatory, microbiota modulation, and cell-protecting effects. *Current Opinion in Food*  
5 *Science* 42: 167-186.
- 6 Papastratis I, Konstantinidis D, Daras P, et al. (2024) AI nutrition recommendation using a deep  
7 generative model and ChatGPT. *Scientific Reports* 14(1): 14620.
- 8 Ponzo V, Goitre I, Favaro E, et al. (2024) Is ChatGPT an Effective Tool for Providing Dietary Advice?  
9 *Nutrients* 16(4): 469.
- 10 Prieto Martinez A, Coutiño Diaz M, Anaya Romero L, et al. (2024) Effects of Vaccinium berries  
11 (blueberries, cranberries and bilberries) on oxidative stress, inflammation, exercise  
12 performance, and recovery—a systematic review. *Food & function*.
- 13 Qarajeh A, Tangpanithandee S, Thongprayoon C, et al. (2023) AI-Powered Renal Diet Support:  
14 Performance of ChatGPT, Bard AI, and Bing Chat. *Clinics and Practice* 13(5): 1160-1172.
- 15 Riso P, Klimis-Zacas D, Del Bo' C, et al. (2013) Effect of a wild blueberry (*Vaccinium angustifolium*) drink  
16 intervention on markers of oxidative stress, inflammation and endothelial function in humans  
17 with cardiovascular risk factors. *European Journal of Nutrition* 52: 949-961.
- 18 Roumeliotis KI and Tselikas ND (2023) Chatgpt and open-ai models: A preliminary review. *Future*  
19 *Internet* 15(6): 192.
- 20 Sarıkaya B, Kolay E, Guney-Coskun M, et al. (2024) The Effect of Black Chokeberry (*Aronia melanocarpa*)  
21 on Human Inflammation Biomarkers and Antioxidant Enzymes: A Systematic Review of  
22 Randomized Controlled Trials. *Nutrition Reviews*. nuae143.
- 23 Sawicki T, Jabłońska M, Danielewicz A, et al. (2024) Phenolic Compounds Profile and Antioxidant  
24 Capacity of Plant-Based Protein Supplements. *Molecules* 29(9): 2101.
- 25 Sawyer SF (2009) Analysis of variance: the fundamental concepts. *Journal of Manual & Manipulative*  
26 *Therapy* 17(2): 27E-38E.
- 27 Schell J, Betts NM, Lyons TJ, et al. (2019) Raspberries improve postprandial glucose and acute and  
28 chronic inflammation in adults with type 2 diabetes. *Annals of Nutrition and Metabolism* 74(2):  
29 165-174.
- 30 Sharma S and Chatterjee S (2021) Winsorization for robust bayesian neural networks. *Entropy* 23(11):  
31 1546.
- 32 Skrovankova S, Sumczynski D, Mlcek J, et al. (2015) Bioactive compounds and antioxidant activity in  
33 different types of berries. *International journal of molecular sciences* 16(10): 24673-24706.
- 34 Taghavi T, Patel H, Akande OE, et al. (2022) Total anthocyanin content of strawberry and the profile  
35 changes by extraction methods and sample processing. *Foods* 11(8): 1072.
- 36 Tofallis C (2015) A better measure of relative prediction accuracy for model selection and model  
37 estimation. *Journal of the Operational Research Society* 66(8): 1352-1362.
- 38 Witkowska AM, Zujko ME, Waśkiewicz A, et al. (2015) Comparison of various databases for estimation  
39 of dietary polyphenol intake in the population of Polish adults. *Nutrients* 7(11): 9299-9308.
- 40 Xie L, Vance T, Kim B, et al. (2017) Aronia berry polyphenol consumption reduces plasma total and low-  
41 density lipoprotein cholesterol in former smokers without lowering biomarkers of  
42 inflammation and oxidative stress: A randomized controlled trial. *Nutrition Research* 37: 67-  
43 77.
- 44 Xu Y, Le Sayec M, Roberts C, et al. (2021) Dietary assessment methods to estimate (poly) phenol intake  
45 in epidemiological studies: A systematic review. *Advances in Nutrition* 12(5): 1781-1801.
- 46 Yang B, Zheng J, Laaksonen O, et al. (2013) Effects of latitude and weather conditions on phenolic  
47 compounds in currant (*Ribes spp.*) cultivars. *Journal of agricultural and food chemistry* 61(14):  
48 3517-3532.
- 49 Yue Y, Petimar J, Willett WC, et al. (2020) Dietary flavonoids and flavonoid-rich foods: validity and  
50 reproducibility of FFQ-derived intake estimates. *Public health nutrition* 23(18): 3295-3303.
- 51 Zhang J, Oh YJ, Lange P, et al. (2020) Artificial intelligence chatbot behavior change model for designing  
52 artificial intelligence chatbots to promote physical activity and a healthy diet. *Journal of*  
53 *medical Internet research* 22(9): e22845.
- 54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Zhao Y, Jiang C, Lu J, et al. (2023) Research progress of proanthocyanidins and anthocyanidins.  
*Phytotherapy Research* 37(6): 2552-2577.

For Peer Review

**Table 1** Polyphenol Contents in Berries According to Phenolic Explorer and ChatGPT Models  
(mg/100 g FW)

Polyphenols	Berries	Phenolic Explorer	ChatGPT-4o mini			ChatGPT-4o		
			Prompt I	Prompt II	Prompt III	Prompt I	Prompt II	Prompt III
<b>Anthocyanins</b>	Bilberry	299.00	260.00	300.00	400.00	300.00	300.00	300.00
	Black chokeberry	444.25	300.00	1155.00	170.00	529.00	400.00	450.00
	Blackcurrant	225.04	250.00	287.91	200.00	361.00	250.00	150.00
	Blackraspberry	589.00	250.00	385.00	160.00	321.00	250.00	150.00
	Blueberry	164.37	100.00	82.30	163.00	163.00	150.00	164.00
	Cranberry	32.00	60.00	0.39	30.00	101.00	101.00	91.00
	Strawberry	NA	110.00	57.74	40.00	24.00	40.00	35.00
<b>Flavonols</b>	Bilberry	1.27	8.00	5.00	20.00	18.00	7.00	8.00
	Black chokeberry	88.41	6.00	5.23	25.00	30.00	10.00	12.00
	Blackcurrant	14.85	13.00	10.92	50.90	65.00	15.00	15.00
	Blackraspberry	19.00	4.00	NA	40.70	11.00	8.00	10.00
	Blueberry	39.80	5.00	13.10	2.50	9.00	5.00	7.00
	Cranberry	28.09	5.00	0.14	3.00	23.00	6.00	5.00
	Strawberry	11.46	2.00	16.99	6.00	6.00	2.00	5.00
<b>Phenolic acids</b>	Bilberry	2.90	20.00	50.00	25.00	70.00	15.00	40.00
	Black chokeberry	141.14	120.00	271.30	60.00	95.00	35	70.00
	Blackcurrant	14.80	90.00	85.53	31.70	116.00	20.00	25.00
	Blackraspberry	38.0	40.00	NA	116.30	89.00	15.00	35.00
	Blueberry	136.47	40.00	40.90	22.00	52.00	30.00	35.00
	Cranberry	56.91	40.00	15.58	20.00	80.00	20.00	25.00
	Strawberry	12.74	30.00	105.76	30.00	45.00	15.00	25.00
<b>Lignans</b>	Bilberry	0.42	1.00	0.50	0.50	0.50	0.20	0.10
	Black chokeberry	NA	0.60	0.40	0.05	1.00	NA	4.00
	Blackcurrant	0.16	1.00	NA	3.10	1.50	0.50	0.20
	Blackraspberry	NA	0.30	NA	NA	0.40	0.30	0.05
	Blueberry	NA	0.20	2.30	0.10	0.20	0.30	NA
	Cranberry	NA	0.50	NA	0.20	0.30	0.20	0.10
	Strawberry	0.35	0.10	NA	1.00	0.20	0.10	0.03
<b>Stilbenes</b>	Bilberry	0.67	0.30	0.50	1.00	0.20	0.20	0.02
	Black chokeberry	NA	0.20	2.00	2.00	0.20	0.05	0.01
	Blackcurrant	0.00	0.50	NA	3.50	0.50	0.05	0.10
	Blackraspberry	NA	0.10	NA	NA	0.30	0.05	0.01
	Blueberry	NA	0.50	0.30	0.10	0.30	0.05	0.05
	Cranberry	NA	0.50	NA	0.10	0.40	0.05	0.03
	Strawberry	0.35	0.20	NA	0.10	0.10	0.05	0.01
<b>Polyphenols</b>	Bilberry	525.00	500.00	400.00	500.00	525.00	500.00	550.00
	Black chokeberry	1752.07	1500.00	1850.00	1200.00	725.00	700.00	1000.00
	Blackcurrant	820.64	800.00	192.06	410.00	757.00	450.00	500.00
	Blackraspberry	980.00	600.00	600.00	522.20	563.00	500.00	600.00
	Blueberry	223.41	400.00	168.60	200.00	366.00	300.00	280.00
	Cranberry	315.00	380.00	85.00	200.00	475.00	400.00	500.00
	Strawberry	289.20	300.00	105.76	225.00	275.00	250.00	300.00

NA: Not Available

**Table 2** Relative Accuracy (%) of ChatGPT Models Polyphenol Content Information on Berries According to the Phenolic Explorer

Model	Polyphenols	Mean Accuracy (%) $\pm$ SD	F-value (df)	<i>p</i>	Significant Differences (Post-hoc)
ChatGPT-4o mini					
	Anthocyanins	58.95 $\pm$ 32.68			Total Polyphenols > Flavonols(*)
	Flavonols	21.27 $\pm$ 25.94			
	Phenolic acids	21.91 $\pm$ 30.27			
	Lignans	26.90 $\pm$ 37.66	4.476	0.001	
	Stilbenes	45.31 $\pm$ 12.25			
	Total Polyphenols	68.01 $\pm$ 25.00			
ChatGPT-4o					
	Anthocyanins	65.36 $\pm$ 38.17			Anthocyanins > Flavonols(*), >Stilbens(**)
	Flavonols	33.17 $\pm$ 31.02			
	Phenolic acids	32.43 $\pm$ 31.53			
	Lignans	35.59 $\pm$ 30.64	6.734	0.000	
	Stilbenes	16.07 $\pm$ 15.19			
	Total Polyphenols	65.72 $\pm$ 20.93			

\*  $p < 0.01$ \*\*  $p < 0.001$

**Table 3** Relative Accuracy (%) of ChatGPT Models Polyphenol Content Information on Berries According to the Phenolic Explorer, Categorized by Berry Types

Polyphenols	Mean Accuracy (%) $\pm$ SD	F-value (df)	<i>p</i>	Significant Differences (Post-hoc)
ChatGPT-4o mini				
Bilberry	45.57 $\pm$ 41.72	2.892	0.015	Not significant between groups
Black chokeberry	44.22 $\pm$ 35.00			
Blackcurrant	41.56 $\pm$ 41.55			
Blackraspberry	42.65 $\pm$ 30.51			
Blueberry	43.60 $\pm$ 31.16			
Cranberry	36.60 $\pm$ 32.04			
Strawberry	34.35 $\pm$ 31.55			
ChatGPT-4o				
Bilberry	45.54 $\pm$ 45.31	1.701	0.134	Not significant between groups
Black chokeberry	50.82 $\pm$ 30.28			
Blackcurrant	51.41 $\pm$ 37.57			
Blackraspberry	48.03 $\pm$ 22.01			
Blueberry	52.69 $\pm$ 34.19			
Cranberry	35.25 $\pm$ 28.15			
Strawberry	40.75 $\pm$ 33.71			

**Table 4** Relative Accuracy (%) of ChatGPT Models Polyphenol Content Information on Berries According to the Phenolic Explorer, Categorized by Query Sessions

Prompts	Relative Accuracy (%)	Mean rank	$X^2$	<i>p</i> -value
ChatGPT-4o mini	Median (25th-75th percentile)			
Query session I	44.7 (12.5 – 85.6)	2.22	2.787	0.248
Query session II	32.9 (1.2 – 73.5)	1.81		
Query session II	50.0 (10.7 – 77.8)	1.96		
ChatGPT-4o				
Query session I	45.3 (0.0 - 80.1)	2.20	2.480	0.289
Query session II	42.3 (16.7 - 87.1)	1.83		
Query session II	43.8 (12.3 - 74.2)	1.97		
Freidman test				