The Quest to Improve Online Search

The impact of a simple query reformulation on result quality

Alexandra Darie
Supervisor: Hrishita Chakrabarti
Responsible Professor: Dr. Maria Soledad Pera
EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Alexandra Darie
Final project course: CSE3000 Research Project
Thesis committee: Dr. Maria Soledad Pera, Hrishita Chakrabarti, Catholijn Jonker

Abstract

As children increasingly turn to general-purpose search engines for learning and exploration, concerns arise about the relevance and safety of the content they encounter. These platforms are designed primarily for adults, often failing to deliver child-friendly results. One existing strategy is to append the phrase "for kids" to search queries, a technique previously shown to improve content relevance. However, those findings are outdated and based on limited evaluation methods. Since then, web content and search engine algorithms have evolved significantly, leaving an open question: how does this simple strategy impact the search quality for young users today? Since children continue encountering complex or unsafe content online, reassessing this approach is important. We aim to evaluate the impact of this query modification across three key dimensions: readability, language safety, and domain trustworthiness. Our results show that, while appending "for kids" modestly improves readability, it does not have a positive impact on the profanity rates or presence of unsafe URLs. These findings suggest that although the strategy remains partially effective, it is not a one-size-fits-all solution. This work highlights the need for more adaptive and child-aware search interventions in modern search environments.

1 Introduction

Children increasingly turn to web search engines for learning, completing homework, and satisfying their curiosity [1]. Internet usage among children has grown rapidly in recent years. For example, in the U.K., 63% of children aged 5–7, 76% of those aged 8–11, and 83% of those aged 12–15 regularly access the internet at home. In the U.S., children under 18 made up nearly 19% of the online population in 2008, totalling over 32 million users [2].

However, most search technologies are not developed with young users in mind, often producing results that are too complex, irrelevant, or inappropriate [3]. The way queries are formulated plays a crucial role in the search engine's outcomes [4]. Even subtle changes in phrasing can significantly influence the ranking of the results, yet children are not always instructed how to take advantage of that [5]. Children's queries often contain grammatical mistakes, vague phrasing, or overly broad scopes [2]. As a result, the search engine may return responses that are difficult to read, irrelevant to their intent, or potentially disturbing. In some cases, such results may pose safety risks for young users online.

Therefore, it is essential to evaluate the techniques we use to ensure that the online environment meets their unique needs. In response, prior research has explored ways to adapt search experiences for children. One existing technique is appending the phrase "for kids" to search queries. Earlier studies, including the work in [6], show that such child-oriented reformulations could return more relevant results than generic query suggestions. Yet, much has changed since those findings. Modern search engines now use more advanced algorithms, personalised rankings, and evolving content landscapes [7], raising the question of whether this strategy still works effectively today.

The main research question guiding this study is: "What is the impact of appending 'for kids' to children's search queries on the quality of search engine results?" To address this question in a more focused way, we broke it down into specific subquestions that examine key aspects of how suitable search results are for children:

- In what ways does appending "for kids" change the readability of retrieved search results?
- How does the presence of profanity or offensive content in search results vary when queries are appended with "for kids"?
- What is the impact of appending "for kids" on the distribution of trusted versus non-trusted domain types among search results?

By answering these subquestions, we aim to comprehensively assess whether this strategy continues to be a practical way to enhance search experiences for young users today.

Using a dataset of queries written by children in grades K-5, we compare search results generated with and without the modification. These results are assessed across three core dimensions of child-appropriate search quality: readability, language safety, and domain trustworthiness, and the resulting metric values are compared to determine if the strategy still improves the child-friendliness of search content.

Our findings show that while this modification modestly improves the readability of search results, it does not reduce exposure to profanity or unsafe websites. These mixed results suggest that although simple query reformulations can offer some benefits, they are insufficient on their

own. This underscores the importance of developing more sophisticated, child-aware search strategies to better support young users in today's complex digital landscape, while also re-evaluating older approaches to check their current effectiveness.

The remainder of this report is structured as follows: we begin with a Related Work section that reviews previous studies on child-oriented search strategies and highlights the limitations of past approaches. The Methodology section outlines our dataset, the technical tools used, the experimental setup, and the evaluation metrics chosen. The Results and Discussion section presents the quantitative findings and analyses them. This is followed by a Responsible Research discussion, reflecting on the ethical considerations and transparency of our process. Then, we offer conclusions and suggest directions for Future Work. Finally, we share our code for reproducibility. [1]

2 Related Work

One of the most influential studies in this area demonstrates that appending "for kids" to search queries significantly improves the relevance of search suggestions for children [6]. Their approach combined a biased random walk on a bipartite graph of web resources and tags with a learning-to-rank model, showing promising results in tailoring search experiences for young users. However, this study focused primarily on relevance (measured through NDCG and recall), relying on a single metric to assess improvement. Moreover, their evaluation was conducted using search logs from over a decade ago, and the web landscape, user expectations, and search algorithms have evolved significantly since then [7].

In addition to work on query reformulation [1], this study's approach is inspired and informed by a broader set of studies aimed at understanding children's search behaviour. Further research conducted large-scale log analyses to uncover challenges children face online, such as accidental clicks on ads, shorter and more ambiguous queries, and difficulty with reading levels [8]. These insights directly shaped the dimensions chosen to evaluate in this study—namely, readability, safety, and trustworthiness. Their complementary work on developmental trends in children's query formulation and reading ability further reinforced the importance of designing age-aware search support strategies [2].

Parallel research has proposed systems like ReQuIK, which tailors query suggestions for children using a combination of linguistic traits and readability metrics [9]. Beyond child-specific search adaptations, prior work in the tagging and personalisation literature has shown that incorporating target audience information can meaningfully shape how users describe or retrieve content. One such study demonstrates that users consciously adjust the type and number of tags they assign depending on whether the intended audience is themselves, peers, or the public [10].

While this work does not introduce a new suggestion algorithm, it extends this line of research by reassessing a strategy that was proven in the past to improve the search results under modern web conditions and with complementary evaluation criteria. Specifically, the findings in [6] motivate the core intervention used in this study: appending audience-specific cues to queries. Analysis of children's online behaviour from [8], [2] informs the selection of evaluation dimensions, particularly readability and content safety, as key indicators of suitability. The ReQuIK system presented in [9] demonstrates that linguistic and readability-based features can effectively guide search tailoring for children, further justifying our choice of readability metrics in the evaluation. Together, these studies validate our intervention and shape our definition of child-friendliness: search results that are age-appropriate in readability, free from unsafe language, and come from trustworthy sources. By synthesising these insights, this study builds on established techniques and applies them in a contemporary context to examine not only relevance, but also the appropriateness and safety of search results for young users.

3 Methodology

To address our research question, we first outline the methodology used in our experiment. This section begins by describing the data, search system and evaluation metrics that form the basis of our approach. Lastly, we explain how the results will be interpreted in the subsequent analysis.

3.1 Software Environment

All experiments were conducted on a MacBook Pro (Retina, 15-inch, Mid 2015) running macOS

[1] `https://github.com/alexandra-darie/OnlineSearch-Children`

Monterey (version 12.7.6). We used Visual Studio Code as the integrated development environment for running and managing the project scripts.

3.2 Data

For this study, we use a dataset of 301 queries written by children between kindergarten and 5th grade levels, collected during search tasks conducted at Boise State University between July 2016 and April 2017 [11]. It also contains 302 queries written by adults to serve as a control group. The queries cover common topics relevant to children's information needs, such as animals, science, hobbies, and general knowledge. Collected in an educational context with real child participants, the dataset is designed specifically to support research in child-oriented information retrieval. It captures naturally occurring vocabulary, intent patterns, and topical interests across early grade levels, making it suitable for analysing the impact of query reformulation on search result quality [11]. Moreover, we clean the dataset by removing duplicates, leading to 294 unique children's queries and 302 unique adult queries. To systematically evaluate the impact of query reformulation, we create paired versions of each of the 294 child-written queries: the original and a modified version with "for kids" appended. We also evaluate the metrics for queries written by adults in the same dataset to ensure a robust comparative analysis. While the "for kids" modification is not applied to these adult queries, they serve as a baseline to highlight differences in search result quality and safety between child and adult queries in general. This comparison helps isolate whether observed improvements are specific to the modification or reflect broader disparities in how search engines handle queries from different user groups.

3.3 Search System

Each query is automatically submitted to the Brave Search API [12], both in its original form and with the "for kids" modification appended. Prior research shows that children tend to interact primarily with the first few links on the search engine results page, often avoiding deeper exploration [2]. To ensure consistent analysis, we retrieve the top 5 search results for each query using the default ranking provided by the Brave API. For each result, we analyse the snippet (description) and destination URL, as these are the elements most visible to users, particularly children [2], without requiring additional interaction such as scrolling or refining the query.

3.4 Evaluation Metrics

In this study, we define *child-friendliness* as the extent to which search results are appropriate, understandable, and safe for children's cognitive and emotional development. Specifically, we evaluate child-friendliness across three dimensions: readability (language complexity), language safety (absence of offensive or harmful language), and domain trustworthiness (reliability of information sources). These dimensions collectively ensure that online search experiences are suitable and beneficial for young users:

- **Readability:** Young users often struggle with complex language and unfamiliar vocabulary [2]. Assessing readability helps determine whether the "for kids" modification effectively tailors content to be more accessible for children's reading levels. Research has highlighted the importance of adapting content to different reading levels for better learning outcomes and safer online experiences [13].

  Firstly, Flesch-Kincaid Grade Level (1) and Dale-Chall (2) formulas are selected because they are among the most established and interpretable readability measures: Flesch-Kincaid provides a widely used estimate of U.S. school grade level [14], while Dale-Chall accounts for vocabulary familiarity, making it particularly relevant when assessing content for children [15]. To further strengthen our analysis, we also include two additional metrics: the Coleman-Liau Index (3), which measures readability based on characters and sentence length and provides a complementary perspective to syllable-based methods [16], and the Child Digital Readability (Spache Index) (4), which is designed to assess texts for younger readers in digital environments [17].

  Regarding the actual score values, Flesch-Kincaid and Coleman–Liau return approximate U.S. grade levels, where scores above 9 typically correspond to high school or above, and scores above 12 indicate college-level complexity. Spache and Dale-Chall are tailored for assessing younger readers, with higher values indicating more difficult

vocabulary and sentence structures. For each retrieved result, we calculate readability scores using the "textstat" Python library [18].

$$\text{Flesch-Kincaid} = 0.39 \left(\frac{\text{Total Words}}{\text{Total Sentences}} \right) + 11.8 \left(\frac{\text{Total Syllables}}{\text{Total Words}} \right) - 15.59 \quad (1)$$

$$\text{Dale-Chall} = 0.1579 \left(\frac{\text{Difficult Words}}{\text{Total Words}} \times 100 \right) + 0.0496 \left(\frac{\text{Total Words}}{\text{Total Sentences}} \right) \quad (2)$$

$$\text{Coleman-Liau} = 0.0588L - 0.296S - 15.8$$
$$\text{where } L = \text{avg letters per 100 words}$$
$$S = \text{avg sentences per 100 words} \quad (3)$$

$$\text{Spache} = 0.659A + 0.121P + 0.810$$
$$\text{where } A = \text{average sentence length}$$
$$P = \text{percentage of difficult words} \quad (4)$$

- **Language Safety:** Children's online environments should be free from offensive or inappropriate language because exposure to such content can be harmful or distressing for young users, and may negatively impact their emotional development [3]. Evaluating this dimension helps us understand whether the modification reduces exposure to harmful language. Recent work has shown the importance of accurately detecting harmful or abusive language in online spaces and mitigating its effects [19].

  Offensive or inappropriate language is detected using an ML-based profanity detector called "better-profanity" [20], which is widely used for fast and reproducible text filtering in research settings due to its ease of use and minimal dependencies [21], [22]. This framework provides functions that process raw text and return a boolean value indicating whether profane language was found in the description of the search result. Internally, the tool applies normalisation and pattern matching to reveal profane words that may be obfuscated.

  In addition, we use the LDNOOBW public offensive keyword list [23]. This publicly available resource contains a community-curated collection of known profanities across multiple languages. For consistency with the query set and other tools used in this study, we restrict our use to the English version. In our implementation, each search result's text is tokenised and compared against the list through direct word matching. This approach offers a customizable layer of profanity detection, complementing the learning-based method with a simple rule-based check.

- **Domain Trustworthiness:** The reliability and safety of the websites themselves are critical when children seek information online. Previous studies have underscored the ethical considerations of using search engines in educational settings and the need to prioritise trusted content [3].

  To assess the trustworthiness of each retrieved search result URL, we submit the destination URL to the Google Safe Browsing Lookup API v4. This API checks the URL against Google's constantly updated list of unsafe web resources and returns a threat classification if the URL is found. Threats include malware, phishing (social engineering), unwanted software, and potentially harmful applications, as defined by Google's Safe Browsing policies.

  For example, for any URL used as input, the API returns a JSON object that is either empty, if the URL appears to be safe, or contains an object called "matches" that lists the found threat types. The API categorises threats into several kinds: malware, social engineering, unwanted software, and potentially harmful applications. We record whether each URL is flagged as unsafe (binary classification) and calculate the percentage of unsafe URLs per query set. URLs that are not found in the threat database are considered safe for our analysis.

3.5 Experimental Protocol

This experiment compares the quality of search results between two sets of child-generated

queries: Children (Original), which contains unmodified queries, and Children (For Kids), where the phrase "for kids" is appended to encourage more child-friendly results. The goal is to evaluate whether this modification leads to improvements across readability, safety, and trustworthiness metrics. A third group, Adults (Original), includes queries written by adults and serves as a control group, providing a contrast to illustrate the types of content children might encounter if their search behaviour aligns more closely with that of adults.

For each query, we retrieve the top 5 results from the Brave Search API. Prior work shows that children typically interact with only the top few search results and rarely scroll further [2]. We extract the snippet and URL from each result and evaluate it using six metrics: four readability formulas (Flesch-Kincaid, Dale-Chall, Coleman–Liau, and Spache), profanity rate, and unsafe URL presence.

We compute the mean score per query across its top 5 results. This ensures our final values reflect the average user experience for a single query while preserving natural variation across queries.

To assess whether the differences between the original and "for kids" modified queries for children are statistically significant, we apply the Shapiro–Wilk normality test to the paired differences [24]. Since all the differences appear not to be normally distributed, we use the non-parametric Wilcoxon signed-rank test [25]. In all cases, we consider results to be statistically significant at the $p < 0.05$ level.

4 Results and Discussion

In this section, we present the findings of our study. We begin by reporting the quantitative results and then reflect on these findings in a broader context.

4.1 Results of Child-Oriented Query Reformulation

We evaluated readability using four standard metrics: Flesch-Kincaid, Dale-Chall, Coleman–Liau, and Spache, with summary scores visualised in Table 2. An overview of the query groups and total retrieved search results is provided in Table 1.

As shown in the Flesch-Kincaid box plot (Figure 1), the average readability score decreased from 10.63 for Children (Original) to 9.72 for Children (For Kids), suggesting a modest improvement in reading level.

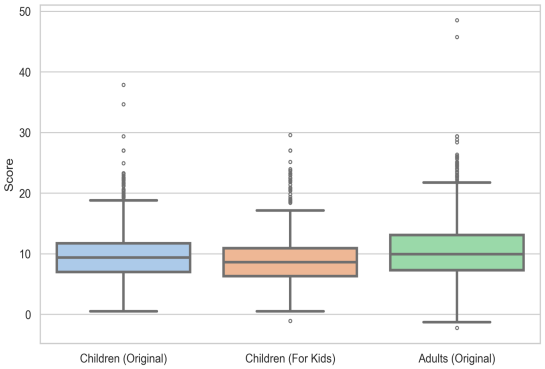

Figure 1: Flesch-Kincaid scores across Children (Original), Children (For Kids), and Adults (Original) queries.

A similar trend is observed in the Dale-Chall (Figure 3) and Coleman-Liau (Figure 2) metrics, where the "for kids" queries scored slightly lower (11.31 and 15.30, respectively) than the original children's queries (11.80 and 15.77). However, all scores for the Children queries remained lower than those of the Adult (Original) queries, which scored 11.89 (Flesch-Kincaid), 12.41 (Dale-Chall), and 18.82 (Coleman-Liau) on average, suggesting that child-intended queries tend to elicit simpler, more readable results.

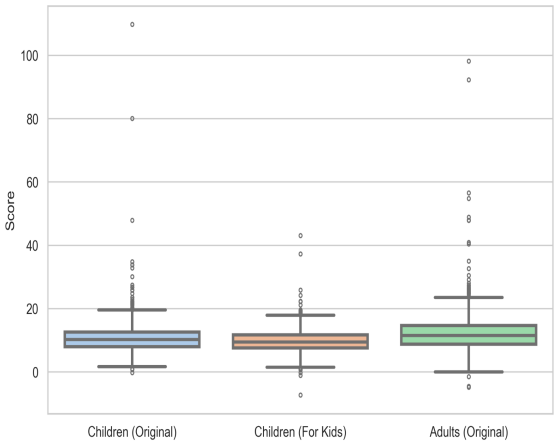

Figure 2: Coleman-Liau scores across Children (Original), Children (For Kids), and Adults (Original) queries.

Spache scores (Figure 4) show a similar trend: Children (For Kids) queries score 5.81 on average compared to 6.14 for Children (Original), and 6.38 for Adult queries.

Query Group	Number of Queries	Total Search Results
Children (Original)	294	1470
Children (For Kids)	294	1470
Adults (Original)	302	1510

Table 1: Query and search result counts across all groups. Each query retrieved 5 results.

Label	Flesch-Kincaid	Dale-Chall	Coleman-Liau	Spache
Children (Original)	10.63	11.80	15.77	6.14
Children (For Kids)	9.72	11.31	15.30	5.81
Adults (Original)	11.89	12.41	18.82	6.38

Table 2: Readability scores and result counts for different query sets.

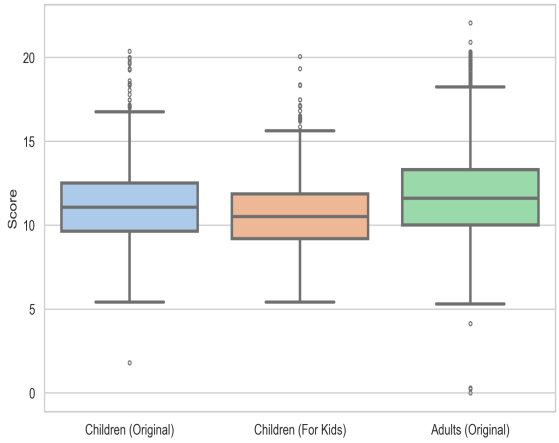

Figure 3: Dale-Chall scores across Children (Original), Children (For Kids), and Adults (Original) queries.

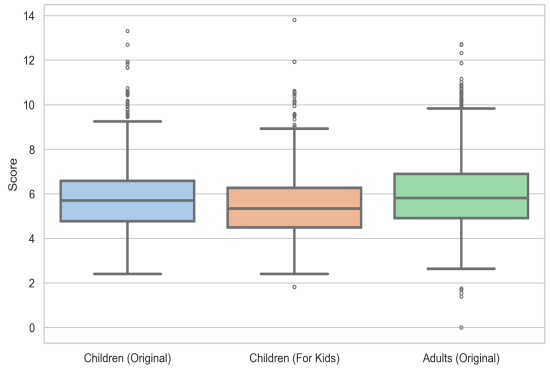

Figure 4: Spache scores across Children (Original), Children (For Kids), and Adults (Original) queries.

Language safety results are presented in Figures 5 and 6. Profanity rates slightly increased from 2.52% in the original children's queries results to 3.27% in the "for kids" version, as seen in Table 3. In contrast, adult queries exhibited the highest profanity rate at 4.30%. Moreover, no unsafe URLs were detected in any query set (0.00% across all), as seen in Table 3.

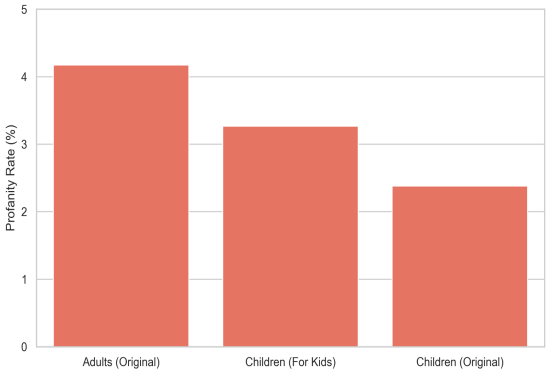

Figure 5: Profanity rates across Children (Original), Children (For Kids), and Adults (Original) queries.

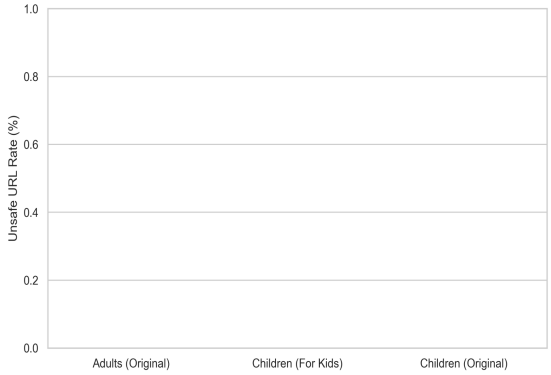

Figure 6: Unsafe URL percentages across Children (Original), Children (For Kids), and Adults (Original) queries.

Appended "for kids" queries resulted in modest improvements in readability, with consistently lower average grade-level scores across all four metrics. Profanity rates, however, were

Query Group	Profanity Rate	Unsafe URL Rate
Children (Original)	2.52%	0.0%
Children (For Kids)	3.27%	0.0%
Adults (Original)	4.30%	0.0%

Table 3: Content safety metrics (profanity and unsafe URLs) across query groups.

slightly higher in the modified queries, and no unsafe URLs were observed in any query set. In contrast, adult queries consistently produced more complex and less filtered results, highlighting the potential risks of unmoderated input. For robustness, we note that readability improvements were statistically significant, while the difference in profanity rates was not. No statistical test is applied to unsafe URLs, as all values were zero across datasets. [2]

4.2 Discussion

Results show nuanced effects of appending "for kids" to children's search queries. Across all four readability metrics: Flesch-Kincaid, Dale-Chall, Coleman–Liau, and Spache, queries with the added phrase consistently yielded content with lower linguistic complexity than their original counterparts. However, despite these statistically significant improvements, the overall readability scores remain high: the average Flesch-Kincaid score for modified queries is 9.72, corresponding to a ninth-grade reading level [14]. This is well above the reading ability of the primary school students the queries represent. Similar gaps appear across the other metrics, suggesting that even with the "for kids" modification, the returned content is still too advanced for young users.

While the modification modestly improves readability, it does not significantly reduce the presence of profanity in search results. Profanity rates were slightly higher in the "for kids" queries (3.27%) compared to the original children's queries (2.52%). One possible explanation for this unexpected result is the nature of the content that tends to surface when "for kids" is appended. Much of this content falls under the category of entertainment or pop culture, such as video game reviews, animated shows, or YouTube content aimed at younger audiences, which may include mild or humorous profanity as part of its tone or branding. Prior research has noted that media targeting younger demographics sometimes incorporates light swearing to appear relatable or comedic without crossing into adult territory [26, 27]. In this context, the use of mild profane language may not always be intended to harm, but it still raises concerns about how content labelled or marketed as "for kids" is moderated by search engines.

Domain trustworthiness was consistently high, with no unsafe URLs detected across any query set. This potentially reflects the effectiveness of Brave Search's internal filtering mechanisms, which appear to provide strong baseline protection regardless of query phrasing. This shift in responsibility from the user's query to the platform's backend protections is encouraging, as it reduces the level to which we rely on query formulation and helps ensure safer experiences by default. In this context, system-level defence serves as an essential protection layer that can work in parallel to reformulation techniques.

These findings highlight that simple query reformulations, such as appending "for kids," are insufficient to ensure a child-friendly search experience. From a system design perspective, this suggests that search engines should incorporate more advanced content filtering mechanisms and develop ranking algorithms specifically designed for child-appropriate content, rather than relying solely on simple query text manipulation.

At a broader societal level, our study underscores the continuing challenges in ensuring safe information access for children online. Children's reliance on general-purpose search engines for education and curiosity can inadvertently expose them to inappropriate or complex content that undermines their development and trust in online information. As digital literacy becomes increasingly relevant in early education, search systems and policymakers must prioritise child-centred information access design, balancing children's agency and safety while supporting their educational growth and curiosity. Addressing these limitations, such as their potential to expose children to inappropriate content or overlook their developmental needs, requires a holistic approach that combines system-level safeguards with societal efforts to foster safe online environments for young users.

[2] All search results retrieved on June 2, 2025.

5 Responsible Research

This study recognises the ethical and societal dimensions of conducting research on search experiences for children. The dataset of children's queries is collected under Institutional Review Board (IRB) approval, ensuring appropriate oversight and privacy safeguards.

From a privacy perspective, all search results were retrieved via public APIs, and no personal user data is collected or stored during our experiments. To prioritise reproducibility, the scripts and code used to process the data and perform the analyses are available in a public repository, supporting future research.

Beyond methodological ethics, it is also important to consider the broader societal implications of children's search experiences. Search engines play an increasingly central role in how children explore the world, form beliefs, and satisfy curiosity [28]. Inaccurate or biased results can shape their understanding in problematic ways, reinforcing stereotypes [29] or exposing them to inappropriate material. Ensuring that young users can search safely is not just a technical challenge, but a societal responsibility.

6 Conclusion and Future Work

This study investigated the effectiveness of a simple query reformulation strategy, appending "for kids" to search queries, in improving the quality of search results for children. We evaluated the impact of this modification across readability, language safety, and domain trustworthiness. Our findings show that the strategy modestly improves readability, but does not reduce the presence of unsafe URLs and negatively changes the profanity rates. These results suggest that while intuitive and easy to implement, the "for kids" reformulation is not a comprehensive solution to ensuring child-friendly search outcomes.

Our analysis has several limitations that also offer directions for future work. We focus on a single reformulation strategy and evaluate it using the Brave Search API. While this provides a concrete case study, results may vary across other search engines. Additionally, language safety analysis is limited to English terms and content. As such, results may not reflect the effectiveness of this strategy in non-English contexts. We also recognise that our readability and language safety metrics may have inherent biases. Readability formulas, for example, rely on standardised language models and may not capture the full cultural diversity of children's experiences [30]. Similarly, static keyword lists for offensive language detection may miss context-dependent terms. Finally, our dataset of children's queries was collected in 2016–2017. While it remains a valuable benchmark, children's search behaviour and the web ecosystem may have changed, potentially affecting the relevance of our findings.

Future work can build on this by re-evaluating a broader range of reformulation strategies. It would also be valuable to assess these strategies using more recent child-generated data and to consider adaptive or learning-based query suggestions that better align with children's evolving language and search behaviours. Combining simple modifications with child-aware design features could offer more robust support for young searchers navigating search environments.

References

[1] M. Kalsbeek, J. Wit, R. Trieschnigg, P. Vet, T. Huibers, and D. Hiemstra, "Automatic reformulation of children's search queries," 01 2010.

[2] S. Duarte Torres and I. Weber, "What and how children search on the web," in *Proceedings of the 18th ACM Conference on Information and Knowledge Management*. Hong Kong, China: ACM, 2009, pp. 393–402.

[3] M. Landoni, T. Huibers, E. Murgia, and M. S. Pera, "Ethical implications for children's use of search tools in an educational setting," *International Journal of Child-Computer Interaction*, vol. 32, p. 100386, 2022. [Online]. Available: https://doi.org/10.1016/j.ijcci.2021.100386

[4] M. Alaofi, L. Gallagher, D. Mckay, L. L. Saling, M. Sanderson, F. Scholer, D. Spina, and R. W. White, "Where do queries come from?" ser. SIGIR '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 2850–2862. [Online]. Available: https://doi.org/10.1145/3477495.3531711

[5] N. Vanderschantz and A. Hinze, "A study of children's search query formulation habits," in *Proceedings of the 31st*

British Computer Society Human Computer Interaction Conference, ser. HCI '17. Swindon, GBR: BCS Learning & Development Ltd., 2017. [Online]. Available: https://doi.org/10.14236/ewic/HCI2017.7

[6] S. D. Torres and D. Hiemstra, "Query recommendation in the information domain of children," in *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 2014, pp. 865–868.

[7] A. A. Mahajan, "Evolution of search engines: A comprehensive review," *International Research Journal of Modernization in Engineering, Technology and Science*, vol. 5, no. 6, 2023. [Online]. Available: https://www.irjmets.com/uploadedfiles/paper/issue_6_june_2023/42435/final/fin_irjmets1687416649.pdf

[8] S. Duarte Torres, I. Weber, and D. Hiemstra, "Analysis of search and browsing behavior of young users on the web," *ACM Transactions on the Web (TWEB)*, vol. 8, no. 2, pp. 1–54, 2014.

[9] I. Madrazo Azpiazu, N. Dragovic, O. Anuyah, and M. S. Pera, "Looking for the movie seven or sven from the movie frozen? a multi-perspective strategy for recommending queries for children," in *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR)*. ACM, 2018, pp. 91–100.

[10] H. Alsarhan, "The effects of target audience on social tagging," Ph.D. dissertation, Indiana University, 2013.

[11] S. Pera and O. Yilmazel, "Scripts and query logs dataset for child-oriented information retrieval," Boise State University ScholarWorks, 2016, accessed: 2025-04-29. [Online]. Available: https://scholarworks.boisestate.edu/cs_scripts/5/

[12] Brave Software, "Brave search api," 2025, accessed: 2025-05-30. [Online]. Available: https://api.search.brave.com

[13] K. Collins-Thompson, P. N. Bennett, R. W. White, S. de la Chica, and D. Sontag, "Personalizing web search results by reading level," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011, pp. 403–412.

[14] R. Flesch, "A new readability yardstick," *Journal of Applied Psychology*, vol. 32, no. 3, p. 221, 1948.

[15] E. Dale and J. S. Chall, "A formula for predicting readability," *Educational Research Bulletin*, pp. 11–20, 1948.

[16] M. Coleman and T. L. Liau, "A computer readability formula designed for machine scoring," *Journal of Applied Psychology*, vol. 60, no. 2, p. 283, 1975.

[17] G. D. Spache, "A new readability formula for primary-grade reading materials," *The Elementary School Journal*, vol. 53, no. 7, pp. 410–413, 1953.

[18] S. Bansal, "textstat: Python library for calculating readability scores," https://github.com/shivam5992/textstat, 2021, accessed: 2025-04-30.

[19] A. Balayn, J. Yang, Z. Szlavik, and A. Bozzon, "Automatic identification of harmful, aggressive, abusive, and offensive language on the web: A survey of technical biases informed by psychology literature," *ACM Transactions on the Web (TWEB)*, vol. 15, no. 4, pp. 1–40, 2021.

[20] T. S. Yaman, "better-profanity: Fast and efficient profanity filter for python," https://github.com/snguyenthanh/better-profanity, 2020, accessed: 2025-04-30.

[21] S. Saseendran, R. Sudharshan, V. Sreedhar, and S. Giri, "Classification of hate speech and offensive content using an approach based on distilbert." in *FIRE (Working Notes)*, 2021, pp. 142–153.

[22] M. A. Orme, Y. Yu, and Z. Tan, "How much do robots understand rudeness? challenges in human-robot interaction," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, and N. Xue, Eds. Torino, Italia: ELRA and ICCL, May 2024, pp. 8247–8257. [Online]. Available: https://aclanthology.org/2024.lrec-main.723/

[23] L. Contributors, "List of dirty, naughty, obscene, and otherwise bad words," https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words, 2021, accessed: 2025-04-30.

[24] SciPy Developers, *scipy.stats.shapiro — SciPy v1.11.3 Manual*, 2023, accessed: 2025-06-10. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html

[25] ——, *scipy.stats.wilcoxon — SciPy v1.11.3 Manual*, 2023, accessed: 2025-06-10. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wilcoxon.html

[26] Q. A. Phan and V. Tan, "Play with bad words: A content analysis of profanity in video games," 06 2017.

[27] S. M. Coyne, L. A. Stockdale, D. A. Nelson, and A. Fraser, "Profanity in media associated with attitudes and behavior regarding profanity use and aggression," *Pediatrics*, vol. 128, no. 5, pp. 867–872, 11 2011.

[28] J. Danovitch, "Growing up with google: How children's understanding and use of internet-based devices relates to cognitive development," *Human Behavior and Emerging Technologies*, vol. 1, pp. 81–90, 04 2019.

[29] M. S. Pera, K. L. Wright, C. Kennington, and J. A. Fails, "Children and information access: Fostering a sense of belonging," in *CEUR Workshop Proceedings*, vol. 3359. CEUR, 2023, pp. 254–257, final published version. [Online]. Available: http://ceur-ws.org/Vol-3359/

[30] B. Bruce, A. Rubin, and K. Starr, "Why readability formulas fail," *IEEE Transactions on Professional Communication*, vol. PC-24, no. 1, pp. 50–52, 1981.

A Appendix - Use of LLMs

To enhance the readability and flow of the paper, a couple of tools were used: Grammarly, which signals grammatical errors and provides suggestions on how to fix them, and ChatGPT, which was used to help reformulate certain sentences. Since English is my second language, obtaining a pleasant flow while keeping an academic style can sometimes be challenging, so to speed up the process, I used the following prompt: "I need to reformulate the following sentence to improve its readability".