



SummarFlex: Exploring Personalized News Filtering and Reading with Query-Focused Hierarchical Summarization

Zimo Xia*
Sun Yat-sen University
Zhuhai, China
xiazm5@mail2.sysu.edu.cn

Xichen Sun*
Sun Yat-sen University
Zhuhai, China
sunxch7@mail2.sysu.edu.cn

Xueyang Wu
The Hong Kong Polytechnic University
Hong Kong, China
xwuba@connect.ust.hk

Qiyu Pan
Sun Yat-sen University
Zhuhai, China
panqy27@mail2.sysu.edu.cn

Kaixiang Mo
Shopee
Singapore
nashmo@qq.com

Zhenhui Peng[†]
Sun Yat-sen University
Zhuhai, China
pengzh29@mail.sysu.edu.cn

ABSTRACT

News readers often need to filter articles of their interests and seek needed information in the selected article. However, they may face clickbait titles and redundant content, which could reduce their news filtering efficiency and reading experience. In this paper, we develop a technical prototype *SummarFlex* to explore news filtering and reading with query-focused hierarchical summarization. In *SummarFlex*, users can filter articles with titles generated by GPT that are relevant to the queried keywords. When reading a selected article, users can specify keywords to generate a high-level summary of the full content and click a sentence to get an expanded summary hierarchically. A within-subject study ($N = 20$) shows that compared to a baseline tool without the summarization features, *SummarFlex* significantly improves users' efficiency in finding interested articles and comprehension of needed information in the article. We conclude with design insights for supporting news filtering and reading with machine-generated content.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design**; • **Computing methodologies** → **Natural language processing**.

KEYWORDS

ChatGPT, LLMs, AI, Human-centered computing, News reading

ACM Reference Format:

Zimo Xia, Xichen Sun, Xueyang Wu, Qiyu Pan, Kaixiang Mo, and Zhenhui Peng. 2024. SummarFlex: Exploring Personalized News Filtering and Reading with Query-Focused Hierarchical Summarization. In *Companion of the 2024 Computer-Supported Cooperative Work and Social Computing (CSCW)*

*Both authors contributed equally to this research.

[†]The Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CSCW Companion '24, November 9–13, 2024, San Jose, Costa Rica

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-1114-5/24/11

<https://doi.org/10.1145/3678884.3681854>

Companion '24, November 9–13, 2024, San Jose, Costa Rica. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3678884.3681854>

1 INTRODUCTION AND RELATED WORK

Seeking needed information from news articles on social news sites or websites of news media is a daily activity for many people. A typical news reading process involves two stages [9]. First, readers **filter** interested news articles to reduce the quantity of news to process. Second, they **read** the selected articles in detail to seek needed information. Current news reading websites have offered many features to support readers in both the filtering and reading stages but may fall short in satisfying users with specific and impromptu reading interests. For example, they can actively sort the news based on the category of content (e.g., sport, politics), published time, trend, specified keywords, and so on. Readers can then skim through the titles of sorted articles to decide which one to read in detail. However, the title of an article is typically determined by the publishers and may not always accurately reflect the content that aligns with readers' interests. In fact, many titles are clickbait that aim at maximizing click-through rates rather than outlining the key ideas of the articles [10]. Consequently, readers would fail to efficiently find interested articles.

In the reading stage, users could often face difficulties associated with information overload [19, 27] in locating and memorizing their interested content of the filtered article. News reading websites widely enable users to navigate their interested information via the browser's keyword-search feature. Researchers in Human-Computer Interaction (HCI) also propose various reading support approaches, such as asking questions [22], providing summaries [4], and fading unimportant text [14], to help users efficiently grasp the article's ideas. However, few websites and HCI research seek to enable news readers to actively customize the article's text based on their interests, which would be beneficial for reducing readers' cognitive load [18].

In this paper, we develop a news reading support tool named *SummarFlex* that helps users filter interested articles with titles replaced by query-focused short summaries and allows users actively customize the content of the read article in a hierarchical structure, which means the multi-level presentation of content, from the title to the summarization, and then to the inter-sentence details.

Query-focused summarization, a technique that aims at generating summaries with different focuses for the same document based on different user-specified keywords, has been applied to digesting given materials like scientific papers [29], videos [25], and news articles [2]. However, it is under-investigated how the query-focused short summaries would affect the news readers' efficiency in filtering articles of interests. Furthermore, little work examines the user experience of reading news articles in an interactive hierarchical structure of query-focused summaries. Filling these gaps could shed light into customizing news reading interface with machine-generated content to satisfy users' information-seeking needs.

In *SummarFlex*, when users specify queried keywords in the filtering stage, it sorts the news articles based on BM25P score, which is an enhanced BM25 weighting model that scores query results by calculating a linear combination of term statistics from different sections of news articles [3], and generates short summaries conditioned to the keywords as a replacement for the articles' original titles. When users click a title of an article to proceed to the reading stage, *SummarFlex* supports reading news summary in a hierarchical structure. Specifically, users can first read a GPT-generated high-level summary of the article that focuses on the queried keywords. Then, they can click any sentence in the summary to view more details about it and locate the most related sentence in the original article.

We evaluate the effectiveness and user experience of *SummarFlex* in a within-subjects study with 20 university students. The results show that compared to a baseline tool without the features about summarization, *SummarFlex* significantly improves users' efficiency in filtering articles that contain interested content and aids them in recalling the the read article's ideas of about given keywords. Participants also report significantly less effort in filtering interested articles and rated that *SummarFlex* is significantly more useful than the baseline tool. We further discuss design insights for future news reading support tools and potential negative effects of these tools' generated content on journalism. Our work primarily contributes understandings of how query-focused hierarchical summaries, one type of generated content, would affect personalized news filtering and reading experience. We also contribute a news reading support tool *SummarFlex* and insights into how generated content can support consumption of online news articles.

2 SUMMARFLEX SYSTEM DESIGN

We design and develop *SummarFlex* system prototype to explore personalized news reading support with query-focused hierarchical summarization. As our focus is to probe the impact of proposed summarization features on keyword-search news filtering and reading process, *SummarFlex* implements necessary components of common news reading interfaces but excludes distractors like Advertisements and Related Topics.

News Filtering Stage. As shown in Figure 1, *SummarFlex* has a similar interface to the BBC website ¹ for searching news articles. Users can specify keywords (part a) and click a "Search" (b) button, which will return a sorted list of ten related articles. The articles are sorted based on their BM25P similarity scores [3] with the specified keywords. At the same time, for each returned article, *SummarFlex*

prompts GPT-3.5, a large language model (LLM) that achieves better performance than most non-LLM methods in text summarization tasks [24], to generate a title (part c) as a replacement to the original title (e). The prompt is "Find the sentence in the article that is most relevant to [keywords] and write a title of the article based on it. The title should have less than 20 words. Please directly return that title. The article is: [content of the article]". This feature could help to address users' challenges on falling for clickbait titles and encountering articles with redundant information, because the generated titles could provide less biased information that reveals users' interested content in the article. The generated titles are displayed letter by letter as does in the ChatGPT official interface to mitigate the issues of delayed responses. Users can click "View More" (part d) next to each article to read it in detail in the news reading stage.

News Reading Stage. When users click a "View More" button in the news filtering interface, they will proceed to a news reading interface as showed in Figure 2. On the right-hand side (part e) of the interface, users can view the full content as they normally can do in a news reading webpage. On the left-hand side, users can input keywords and click the "Search" button in the search box (a). They can then see a generated short summary (example in Figure 2b) under the search box. To enrich user personalization while preserving the simplicity of the summary as many users lack enough time for in-depth reading, *SummarFlex* employs a hierarchical structure. If users are interested in specific information in the short summary, they can click the corresponding sentence (e.g., the orange one in Figure 2b) and view its extended version (e.g., the highlighted part c). By clicking on the extended sentence, they can also see a highlighted sentence (d) in the original document that is most relevant to the clicked sentence ², which provides users context with further detail. These features could help users pinpoint essential information.

To support the interactive hierarchical summarization in the news reading stage, we experimented with several large language models for query-focused summarization, including open-sourced Randeng-Pegasus-523M-Chinese, Ziya-LLaMA-13B-v1.1³, and commercial GPT-3.5 and GPT-4. We observed that GPT-4 performed the best in our trials, followed by GPT-3.5 whose api is cheaper. Therefore, for query-focused high-level summary generation, we prompt GPT-4 with "Write a summary of a news article. The summary needs to focus on [keywords]. Please do not omit any detail related to [keywords] in this article. The summary should have less than 200 words. The article is: [content of the article]". Similarly, to generate an extended version of user's interested sentence, we prompt GPT-4 with "Please expand the sentence [sentence] with relevant details in a given news article. The expended sentence should have less than 50 words. The article is: [content of the article]". These prompts above were derived from our own preliminary testing and adjustments. Given our current level of prompt engineering expertise, their performance likely has room for improvement. We use different GPT models for different functions to balance performance and budget. GPT-3.5 is sufficient for generating titles since it

¹https://www.bbc.co.uk/search?d=SEARCH_PS

²The relevance is based on the cosine similarity between the sentences vectors encoded by text2vec.

³<https://fengshenbang-lm.com>



Figure 1: Screenshot of *SummarFlex*'s news filtering interface. (a) Search box. (b) Search button. (c) Sorted news articles based on the keywords in the search box. The titles of these articles are generated by GPT-3.5 with a focus on the user-specified keywords. (d) View More button to read an article in detail. (e) Original titles of the sorted articles used in the baseline tool.

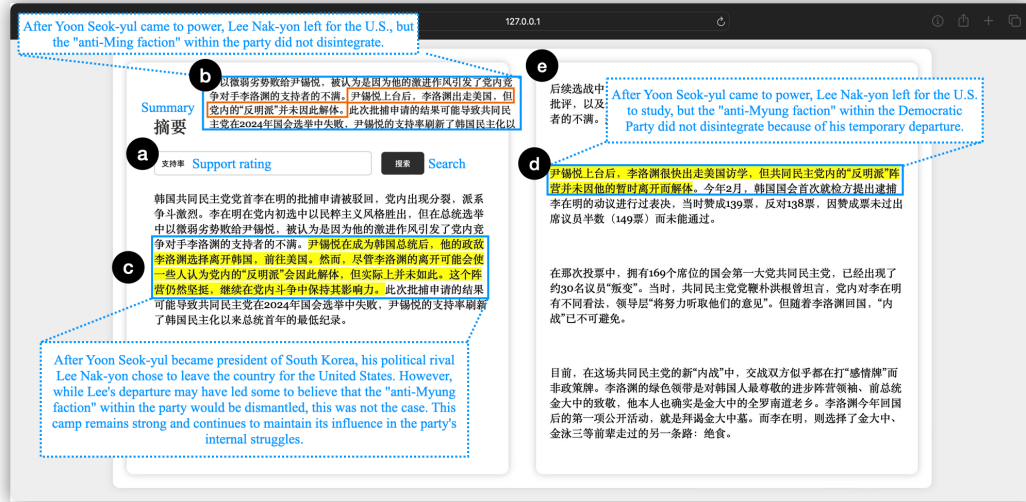


Figure 2: Screenshot of *SummarFlex*'s news reading interface. (a) Search box. (b) An example of the generated short summary of the news article with a focus on the user-specified keywords. (c) Expanded summary of the generated short one in (b) when users click an interested sentence. (d) Highlighted sentence in the news article that is most relevant to the clicked sentence.

is a simpler task compared to summarization, which requires GPT-4 for optimal performance.

3 USER STUDY

To evaluate the effectiveness and user experience of *SummarFlex* in the personalized news reading process, we conducted a within-subjects study with 20 university students (17 males, 3 females,

age: 18–25). We recruited them via the social media and word of mouth in a local university. To ensure the confidentiality of the data collected, all participant data were anonymized, guaranteeing that individual responses could not be traced back to any participant. All participants are native Chinese speakers, and their majors include artificial intelligence, software engineering, exhibition management, and information management. The majority of participants

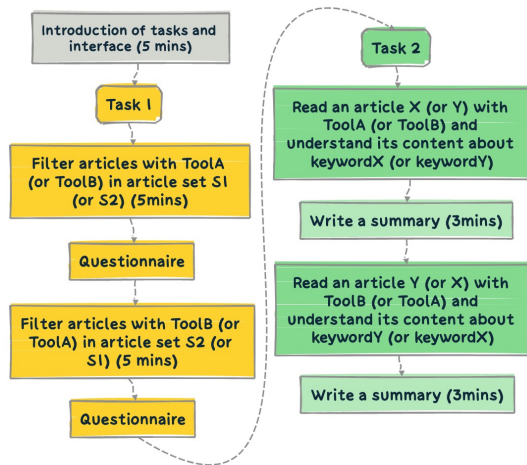


Figure 3: Procedure of the within-subjects user study. Participants need to conduct the news filtering task 1 and news reading task 2 with both *SummarFlex* and the baseline tool.

reported daily news consumption, while some engaged sporadically. Most participants tended to read news in fragmented sessions lasting less than 10 minutes.

Baseline. We control the features of a baseline news reading tool to examine the impact of *SummarFlex*’s unique features in the news filtering and reading stages. Specifically, in the filtering stage, the baseline tool will display the original titles (Figure 1e) of the sorted articles based on the keywords rather than the generated ones in *SummarFlex* (c). In the reading stage, the baseline tool only displays the original content of the news article (i.e., the part e in Figure 2), and the users can use the keyword-search feature provided by the web browser. Overall, the baseline tool simulates a traditional news web page for filtering and reading news articles.

Dataset of News Articles. To form the dataset, we collected 300 news articles about technology, politics, entertainment, and sports (each category has 75 articles) from Chinese social media platforms Weibo, WeChat Official Accounts, and Zhihu. While choosing articles for the four news categories, we follow three guidelines: 1) the articles should come from reputable news publishers or popular social media accounts known by at least two authors of this paper; 2) the articles should cover as many topics as they could, which could make it non-trivial for users in the news filtering task; and 3) either the articles do not contain images, or the removal of their images does not affect the completeness of the articles, such that we can better focus on the summarization of the text in this paper.

3.1 Procedure and Measures

We conducted the study offline following the procedure depicted in Figure 3. The procedure lasts about 40 minutes. Each participant gets 50 RMB (about 7 USD) for compensation. We conducted our study in December 2023 and spent about 57 USD invoking GPT APIs. After filling the consent form and getting instructions of the tasks and interface, each participant starts two tasks that serve different purposes.

3.1.1 Task 1: News Filtering. This task aims at examining whether *SummarFlex* can help users find articles that contain their interested content more efficiently than the baseline tool. Our team randomly splits the dataset of 300 news articles into set S1 and S2. For each set, we compile a list of 20 statements. Each statement reflects an idea of a randomly sampled article. Participants have two news filtering sessions, each with an article set and a news reading tool. Within 5 minutes in each session, they are required to find out and record as many articles that contain the idea of the statements as possible. With either tool, participants could enter keywords in the search box in the news filtering page (Figure 1) and click the “View More” button to read the article in detail if needed. To keep the focus on the news filtering stage in task 1, for both *SummarFlex* and the baseline tool, the news reading page only contains the article’s content and does not have the summarization features. After each session, participants fill a questionnaire about their perceived effort in completing this session. We counterbalance the orders of the experienced tools and article sets using Latin Square, i.e., ToolA (S1) → ToolB (S2), ToolA (S2) → ToolB (S1), ToolB (S1) → ToolA (S2), and ToolB (S2) → ToolA (S1), where ToolA refers to *SummarFlex*, and ToolB refers to the baseline tool.

In task 1, we recorded participants’ interaction with both tools. We count the number of correctly identified news articles that match to the given statements and the ratio of the correct articles to the total number of articles opened, i.e., the articles that users click “View More” next to their titles. In task 1’s questionnaire, we ask their perceived effort in the news filtering task on a 5-points Likert scale using an item adapted from Colligan et al. [5]: “How hard did you have to work to accomplish the task?” (1 / 5 - very easy / very hard).

3.1.2 Task 2: News Reading. This task aims at examining whether *SummarFlex* can help users memorize the article’s content from their interested aspects more effectively than the baseline tool. The first author selects two articles from the dataset and thinks of an interested keyword for each article (article X: “quality”; article Y: “difference”). Article X is about “Pinduoduo (Known as Temu in China)” (2641 words), while article Y is about “Instagram and Threads” (2096 words). Upon completion of task 1, participants proceed to task 2 which also has two sessions, each with an article and a news reading tool. In each session, they need to first read a given article in the tool’s reading interface (e.g., the *SummarFlex* one in Figure 2) within 3 minutes. Then, without referring back to the article, they need to write a summary less than 100 words that surrounds the given keyword within another 3 minutes. After each session, they fill a questionnaire about the perceived usefulness and easiness to use of the tool. We counterbalance the orders of the experienced tools and given articles using Latin Square, i.e., ToolA (Article X) → ToolB (Y), ToolA (Y) → ToolB (X), ToolB (X) → ToolA (Y), and ToolB (Y) → ToolA (X).

In task 2, we scored the summaries written by the participants on five dimensions adapted from Dang [6]: Non-redundancy: There should be no unnecessary repetition in the summary; Grammaticality: The summary should have no obviously ungrammatical sentences that make the text difficult to read; Referential clarity: It should be easy to identify who or what the pronouns and noun phrases in the summary are referring to; Structure & Coherence:

The summary should be well-structured and well-organized; Focus: The summary should have a focus on the given keyword. We followed [7, 23] to rate the dimension on a 7-points Likert scale (1 / 7 - very Poor / very Good). We invited two undergraduate students majored in Chinese language and literature to score the summaries in a blind order. We averaged their scores to obtain the final score for each dimension. Additionally, we recorded the time participants spent reading the articles and writing summaries during Task 2. In task 2's questionnaire, we ask the perceived usefulness ("The use of this tool enables me to understand the article from the aspect of the given keyword more efficiently") and easiness to use ("I would find this tool to be flexible to use") on a 5-points Likert scale, each with an item adapted from Yuan et al. [28] (1 / 5 - strongly disagree / strongly agree).

3.2 Analyses and Results

We performed Wilcoxon signed-rank tests to assess the difference between the *SummarFlex* and baseline conditions regarding participants' performance in the news filtering task 1, scored items in their summaries of the articles in the news reading task 2, and self-reported perceptions of the tools. The Wilcoxon signed-rank test is a non-parametric statistical test commonly used in HCI studies (e.g., [12, 22]) to compare two matched samples especially when the data do not meet the assumptions of normal distribution. Table 1 summarizes the statistical results.

News Filtering. Compared to the cases with baseline tool ($Median = 4$), participants with *SummarFlex* can identify significantly more news articles that correctly match to the given statements ($Mdn = 5$; $Z = -3.26, p < 0.001$). Moreover, the ratio of the correct articles to the total number of articles open is significantly higher in the *SummarFlex* condition ($Mdn = 0.67$) than that in the baseline condition ($Mdn = 0.55$; $Z = -3.62, p < 0.001$). These results indicate that *SummarFlex* significantly improved users' efficiency in filtering news articles that contain content of their interests.

News Reading. Table 1 shows the scores of summaries written by participants after reading given articles with *SummarFlex* and the baseline tool. In both conditions, participants have comparable performance in writing a short summary that overall has correct grammar, has clear reference about the pronouns and noun phrases, and is structured; $p > 0.05$. The written summaries after reading articles with *SummarFlex* ($Mdn = 5$) are significantly less redundant than those with the baseline tool ($Mdn = 4.5$; $Z = -4.05, p < 0.05$). Besides, after reading with *SummarFlex* ($Mdn = 5$), participants can write a summary that has significantly more focus on the given keyword than that in the baseline condition ($Mdn = 4.5$; $Z = -4.51, p < 0.01$). The result indicates that in the news reading stage, *SummarFlex* can improve users' impression of the articles from their interested aspects.

Perceptions. In general, participants feel that *SummarFlex* ($Mdn = 3$) significantly reduced their effort in the news filtering task when compared to the baseline tool ($Mdn = 3$; $Z = -2.01, p < 0.05$). Furthermore, they feel that *SummarFlex* is significantly more useful ($Mdn = 5$) and easier to use ($Mdn = 4.5$) than the baseline tool (usefulness: $Mdn = 2, Z = -3.91, p < 0.05$; ease of use: $Mdn = 2, Z = -3.54, p < 0.05$) in facilitating them in the news reading stage. Eight participants express their fondness for the query-focused

hierarchical summarization feature. "The summarization function helped me tailor the news content to my interests, which improves the efficiency of digesting needed information in this article" (P12, M, age:20). "The keyword-focused summaries helped me quickly navigate to the key information in long articles; otherwise, I would often miss relevant details of my interests" (P6, M, age:20).

4 DISCUSSION

In this work, we propose a technical prototype named *SummarFlex* to enhance personalized news reading experience via query-focused hierarchical summarization. Our user study demonstrates the effectiveness of *SummarFlex*'s query-focused generated titles in improving users' efficiency in filtering their interested news articles. This finding provides empirical evidence to the potentials of using short machine-generated content to support information filtering tasks, which may help to address the issues of clickbait titles in user-generated content [10], thereby helping users more quickly determine whether a news article provides the information they need. Participants with *SummarFlex* feel that it can significantly reduce their effort in news filtering task, which suggests that the generated titles could mitigate the information overload issue in social media. Our study shows that *SummarFlex*'s query-focused hierarchical summarization can improve users' recall of the articles with a focus on their interested keywords. These results align with Kalyanaraman and Sundar [11], which indicates that the customized content can increase user satisfaction.

Our design process of *SummarFlex* and findings of user study could have useful implications for future tools that support personalized news reading with generative AIs. For one thing, these tools could provide multi-topic news summarization in a hierarchy view. *SummarFlex* allows users to click a sentence in a summary to expand it with more detail and locate the context in the original article, which is explicitly appreciated by three participants in task 2. This feature of *SummarFlex* also serves the purpose of helping users understand the context and cross-check with the original text to mitigate potential errors in the summaries resulting from the hallucination phenomenon in LLMs. However, users still have to read the raw text around the sentence to get the contextual information, whose topics could be different from the specified keywords. We suggest that news reading support tools like *SummarFlex* should provide a more seamless interaction between the summarized content and its contextual information. For example, as inspired by Sensecape [26], the tools can organize the content of the news article in a hierarchy view which outlines its abstractive topics and their connections. For another, the reading support tools could support hierarchical summarization of multiple related news articles. *SummarFlex* provides generated titles and query-focused hierarchical summarization, which have been demonstrated useful in filtering and reading one news article. Nevertheless, users might need to read multiple related articles about a news topic to form a comprehensive understanding or recognize if it is fake news [15]. We envision that a similar type of hierarchical summarization support, customized to suit the scenarios of reading multiple related articles, would be beneficial for future news reading support tools. For example, these tools could aim to generate a coherent summary whose sentences are sourced from different articles. They could

Table 1: Participants’ performance scores and user perceptions towards *SummarFlex* and the baseline tool across both the Task 1 and the Task 2. Performance in the Task 1 and user perceptions are rated on a 5-point Likert scale (1 - the worst, 5 - the best). Scores for participants’ written summaries in the Task 2 are based on a 7-point Likert scale (1 - the worst, 7 - the best).

	Factor	<i>SummarFlex</i> Mean/S.D.	Baseline Mean/S.D.	Statistics p
Task1	Number of Correctly Identified Articles	5.45/1.56	3.85/1.53	< .001
	Ratio of Correct Articles to Total Articles Opened	0.69/0.17	0.47/0.15	< .001
Task2	Non-redundancy	4.90/1.06	4.38/0.69	0.027
	Grammaticality	4.83/1.12	4.53/1.07	0.341
	Referential clarity	5.03/1.25	5.08/1.23	0.947
	Structure & Coherence	4.50/1.41	4.28/1.53	0.383
	Focus	5.35/0.92	4.38/1.04	0.003
User perceptions	Perceived effort	2.65/1.14	3.35/1.23	0.044
	Usefulness	4.60/0.50	2.05/1.00	< .001
	Ease of use	4.20/1.06	2.30/0.86	< .001

use similar ‘citation jump’ interaction in Bing Chat ⁴ that navigates users to their interested article for detailed information.

The design concepts underlying *SummarFlex* can be adapted to other multi-topic reading scenarios, such as scientific papers and legal documents. For example, in the context of reading scientific papers, a tool like *SummarFlex* could generate customized titles or abstracts of the papers based on readers’ keywords to help them determine which paper to read in detail. It can also distill readers’ interested content such as measures in the user study to help readers quickly comprehend the information they need. For example, the generated titles based on user-specified keywords could reduce selective exposure behaviors in social media and increase the diversity of news exposure [16], while the hierarchical summarization could reduce user’s anxiety in the reading stage [1] and increase people’s willingness to use social media to read news [21]. For news creators, *SummarFlex* can help them draft a title and customize the content to match the interests of the article’s target audience.

However, using generated summaries in *SummarFlex* may bring potential negative impact on journalism. For example, it may reduce readers’ tendency to click on the articles that originally have eye-catching titles, leading to a decrease in click-through rates and page views on these news articles. Additionally, since the summaries and titles generated in *SummarFlex* are based on user-query keywords, there is a possibility that users may overlook some crucial background information, resulting in cognitive bias. Moreover, the generated content may make readers misinterpret the main ideas of the articles that the news creators want to convey. To mitigate these potential impacts, we should involve readers, authors of news articles, managers of news websites, and HCI researchers in the design process of future news reading tools that leverage machine-generated content.

Our work has several limitations. First, to focus on evaluating the impact of our summarization features, our *SummarFlex* and baseline tool simplify the interface without components such as advertisements [8] and number of likes and sharings [20] for each article. However, users in real-world scenarios may take these components into consideration when filtering and reading news articles. Second, we could not avoid the impact of the novelty effect on user perception towards *SummarFlex* in the lab study. Future work

needs to explore embedding *SummarFlex* (e.g., as an add-on to the browser) into the existing news websites or social media platforms and evaluate it with a long-term user study. Third, our collected news dataset for the user study only consists textual articles to keep our focus on text summarization, while images and videos of a news may also affect readers’ perception and experience [13]. The demonstrated benefits of *SummarFlex* can positively inform future work to leverage multi-modal summarization techniques to support reading of multi-modal news articles. Fourth, our participants in the user study are university students, and most of them are males. Future work should recruit other representative news readers to ensure broader applicability and understanding of *SummarFlex*’s effectiveness and user experience. Fifth, we have not incorporated the genres (e.g., factuality and formality [17]) of news articles into the summarization nor detected the harmfulness of generated content. We urge future work to examine these factors for personalized and safe news reading experience.

5 CONCLUSION

In this study, we developed an interactive tool named *SummarFlex* to explore the impact of generated query-focused summaries on personalized news filtering efficiency and news reading experience. It aids news filtering by replacing the titles of news articles with machined-generated ones based on the queried keywords and supports news reading with interactive hierarchical query-focused summaries. Our within-subject user study with 20 participants demonstrates the effectiveness of *SummarFlex* in improving users’ efficiency in filtering interested articles and memorizing interested content in a selected article. Our work exemplifies how machine-generated content can equip news readers with powerful tools for online information seeking and sense-making. Collectively, our work offers an exciting initial step towards powering personalized news reading process with generative models.

ACKNOWLEDGMENTS

This work is supported by the Young Scientists Fund of the National Natural Science Foundation of China (NSFC) with Grant No.: 62202509, NSFC Grant No.: U22B2060, and the General Projects Fund of the Natural Science Foundation of Guangdong Province in China with Grant No.: 2024A1515012226.

⁴<https://www.microsoft.com/en-us/edge/features/bing-chat?form=MA13FJ>

REFERENCES

- [1] Brooke E Auxier and Jessica Vitak. 2019. Factors motivating customization and echo chamber creation within digital news environments. *Social media+ society* 5, 2 (2019), 2056305119847506.
- [2] Seyed Ali Bahrainian, George Zerveas, Fabio Crestani, and Carsten Eickhoff. 2021. CATS: Customizable Abstractive Topic-Based Summarization. *ACM Trans. Inf. Syst.* 40, 1, Article 5 (oct 2021), 24 pages. <https://doi.org/10.1145/3464299>
- [3] Matteo Catena, Ophir Frieder, Cristina Ioana Muntean, Franco Maria Nardini, Raffaele Perego, and Nicola Tonello. 2019. Enhanced News Retrieval: Passages Lead the Way!. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) (SIGIR '19). Association for Computing Machinery, New York, NY, USA, 1269–1272. <https://doi.org/10.1145/3331184.3331373>
- [4] Xiang 'Anthony' Chen, Chien-Sheng Wu, Lidiya Murakhovska, Philippe Laban, Tong Niu, Wenhao Liu, and Caiming Xiong. 2023. Marvista: Exploring the Design of a Human-AI Collaborative News Reading Tool. *ACM Trans. Comput.-Hum. Interact.* (jul 2023). <https://doi.org/10.1145/3609331> Just Accepted.
- [5] Lacey Colligan, Henry W.W. Potts, Chelsea T. Finn, and Robert A. Sinkin. 2015. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. *International Journal of Medical Informatics* 84, 7 (2015), 469–476. <https://doi.org/10.1016/j.ijmedinf.2015.03.003>
- [6] Hoa Trang Dang. 2005. Overview of DUC 2005. <https://api.semanticscholar.org/CorpusID:61825275>
- [7] Thiago Castro Ferreira, Chris van der Lee, Emiel Van Miltenburg, and Emiel Krahmer. 2019. Neural data-to-text generation: A comparison between pipeline and end-to-end architectures. *arXiv preprint arXiv:1908.09022* (2019).
- [8] Jeffrey Gleason, Desheng Hu, Ronald E Robertson, and Christo Wilson. 2023. Google the Gatekeeper: How Search Components Affect Clicks and Attention. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 245–256.
- [9] Avery E Holton and Hsiang Iris Chyi. 2012. News and the overloaded consumer: Factors influencing information overload among news consumers. *Cyberpsychology, behavior, and social networking* 15, 11 (2012), 619–624.
- [10] Farnaz Jahanbakhsh, Amy X. Zhang, Karrie Karahalios, and David R. Karger. 2022. Our Browser Extension Lets Readers Change the Headlines on News Articles, and You Won't Believe What They Did! *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 530 (nov 2022), 33 pages. <https://doi.org/10.1145/3555643>
- [11] Sriram Kalyanaraman and S Shyam Sundar. 2006. The psychological appeal of personalized content in web portals: Does customization affect attitudes and behavior? *Journal of Communication* 56, 1 (2006), 110–132.
- [12] Youwen Kang, Zhida Sun, Sitong Wang, Zeyu Huang, Ziming Wu, and Xiaojuan Ma. 2021. MetaMap: Supporting visual metaphor ideation through multi-dimensional example-based exploration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [13] Alireza Karduni, Ryan Wesslen, Douglas Markant, and Wenwen Dou. 2023. Images, Emotions, and Credibility: Effect of Emotional Facial Expressions on Perceptions of News Content Bias and Source Credibility in Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 470–481.
- [14] Junpei Kobayashi and Toshio Kawashima. 2019. Paragraph-Based Faded Text Facilitates Reading Comprehension. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300392>
- [15] Sian Lee, Haeseung Seo, Dongwon Lee, and Aiping Xiong. 2023. Associative Inference Can Increase People's Susceptibility to Misinformation. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 530–541.
- [16] Sun Kyong Lee, Nathan J Lindsey, and Kyun Soo Kim. 2017. The effects of news consumption via social media and news information overload on perceptions of journalistic norms and practices. *Computers in human behavior* 75 (2017), 254–263.
- [17] Zilin Lin, Kasper Welbers, Susan Vermeer, and Damian Trilling. 2023. Beyond Discrete Genres: Mapping News Items onto a Multidimensional Framework of Genre Cues. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 542–553.
- [18] Zhieh Lor, Hae Jung Oh, and Jihyang Choi. 2022. Excluding and Including: News Tailoring Strategies in an Era of News Overload. *Digital Journalism* (2022), 1–19.
- [19] Rongjun Ma, Henrik Lassila, Leysan Nurgalieva, and Janne Lindqvist. 2023. When Browsing Gets Cluttered: Exploring and Modeling Interactions of Browsing Clutter, Browsing Habits, and Coping. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 861, 29 pages. <https://doi.org/10.1145/3544548.3580690>
- [20] Lillio Mok, Michael Inzlicht, and Ashton Anderson. 2023. Echo Tunnels: Polarized News Sharing Online Runs Narrow but Deep. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 662–673.
- [21] Chang Sup Park. 2019. Does too much news on social media discourage news seeking? Mediating role of news efficacy between perceived news overload and news avoidance on social media. *Social Media+ Society* 5, 3 (2019), 2056305119872956.
- [22] Zhenhui Peng, Yuzhi Liu, Hanqi Zhou, Zuyu Xu, and Xiaojuan Ma. 2022. CReBot: Exploring Interactive Question Prompts for Critical Paper Reading. *Int. J. Hum.-Comput. Stud.* 167, C (nov 2022), 17 pages. <https://doi.org/10.1016/j.ijhcs.2022.102898>
- [23] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. Anglekindling: Supporting journalistic angle ideation with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [24] Xiao Pu, Mingqi Gao, and Xiaojuan Wan. 2023. Summarization is (Almost) Dead. *arXiv:2309.09558* [cs.CL]
- [25] Aidean Sharghi, Boqing Gong, and Mubarak Shah. 2016. Query-focused extractive video summarization. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VIII 14*. Springer, 3–19.
- [26] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. *arXiv preprint arXiv:2305.11483* (2023).
- [27] Chance York. 2013. Overloaded by the news: Effects of news exposure and enjoyment on reporting information overload. *Communication Research Reports* 30, 4 (2013), 282–292.
- [28] Kangyu Yuan, Hehai Lin, Shilei Cao, Zhenhui Peng, Qingyu Guo, and Xiaojuan Ma. 2023. CriTrainer: An Adaptive Training Tool for Critical Paper Reading. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–17.
- [29] Xiaoyu Zhang, Jianping Li, Po-Wei Chi, Senthil Chandrasegaran, and Kwan-Liu Ma. 2023. ConceptEVA: Concept-Based Interactive Exploration and Customization of Document Summaries. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 204, 16 pages. <https://doi.org/10.1145/3544548.3581260>