

From Overwhelmed to Overview: Understanding Smartphone Users' Preferences and Expectations in Relieving Notification Overload via Text Summarization

UEI-DAR CHEN, National Yang Ming Chiao Tung University, Taiwan

PENG-JUI WANG, National Yang Ming Chiao Tung University, Taiwan and University of California San Diego, USA

YI-CHI LEE, National Yang Ming Chiao Tung University, Taiwan and The University of Texas at Austin, USA

YONG-HAN LIN, National Yang Ming Chiao Tung University, Taiwan

YU-LING CHOU, University of California Santa Cruz, USA and National Yang Ming Chiao Tung University, Taiwan

YUNG-JU CHANG, National Yang Ming Chiao Tung University, Taiwan

To help users manage the overwhelming influx of smartphone notifications, this study explores how large language models (LLMs) can be leveraged to generate notification summaries. We developed an Android application that integrates ChatGPT to summarize notifications and conducted an in-the-wild deployment to examine how users guided the model. To further understand user expectations for LLM-generated summaries, we interviewed 20 participants following a week-long engagement with the app. Our findings reveal five main strategies that users employed in their prompts for generating summaries. Additionally, interviewees expected summaries to prioritize three types of notifications, preferred three levels of information disclosure influenced by content anticipation and perceived criticality, and used three different approaches to synthesizing notifications based on their interrelationships. Finally, interviewees envisioned notification summarization functioning like a virtual assistant, desiring capabilities beyond simple information condensation, including support for task and information management, revisiting archived content, and tracking activities for reflection.

CCS Concepts: • **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**.

Additional Key Words and Phrases: Notification Management; Notification Overload; Text Summarization

ACM Reference Format:

Uei-Dar Chen, Peng-Jui Wang, Yi-Chi Lee, Yong-Han Lin, Yu-Ling Chou, and Yung-Ju Chang. 2025. From Overwhelmed to Overview: Understanding Smartphone Users' Preferences and Expectations in Relieving

Authors' Contact Information: **Uei-Dar Chen**, Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, udchen.cs08@nycu.edu.tw; **Peng-Jui Wang**, Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, Department of Computer Science and Engineering and University of California San Diego, La Jolla, CA, USA, james900425.cs08@nycu.edu.tw; **Yi-Chi Lee**, Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, Department of Computer Science and The University of Texas at Austin, Austin, TX, USA, yichi170.cs08@nycu.edu.tw; **Yong-Han Lin**, Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, alberta.cs07@nycu.edu.tw; **Yu-Ling Chou**, Computational Media, University of California Santa Cruz, Santa Clara, CA, USA, Department of Computer Science and National Yang Ming Chiao Tung University, Hsinchu, Taiwan, yhou14@ucsc.edu; **Yung-Ju Chang**, Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, armuro@nycu.edu.tw.

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.

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2573-0142/2025/9-ARTMHCI011

<https://doi.org/10.1145/3743703>

Notification Overload via Text Summarization. *Proc. ACM Hum.-Comput. Interact.* 9, 5, Article MHCI011 (September 2025), 28 pages. <https://doi.org/10.1145/3743703>

1 Introduction

Smartphone users face an ever-increasing volume and variety of notifications, including instant messages, emails, news, and social media updates. While these notifications are designed to keep users informed and connected, their continuous influx often leads to notification overload [16, 51, 65, 68], which can strain users' cognitive capacities [1, 4, 68], time management [27, 51], and overall well-being [16, 29, 65]. Past research has explored strategies to help users control when and how they handle notifications. These include finding opportune times to deliver alerts [13, 32, 38, 40, 47, 49], allowing users to defer notifications [3, 47, 64], sorting notifications to prioritize them [33, 34], and pinning notifications for subsequent review [34]. While these solutions prove useful in helping users avoid unwanted interruptions and manage their time, they fall short in reducing the overall volume of information requiring attention.

Summarization has long been used to condense large volumes of content into more digestible forms [25, 37, 42, 44]. By presenting information from notifications in a condensed overview, summaries can potentially alleviate cognitive burden [2, 5, 44, 57], enabling users to swiftly identify important information without reviewing each notification in full. Recent advancements in large language models (LLMs) [7, 46] have made automated notification summarization increasingly viable. State-of-the-art LLMs demonstrate strong capabilities in contextual understanding, information extraction, and coherent summary generation across diverse domains [61, 71]. They are also readily accessible via web APIs, allowing seamless integration into mobile applications.

However, while the underlying technology for notification summarization is available, the effectiveness of such systems depends not only on technical advancements but also on a clear understanding of users' expectations and needs. It remains unclear how users prefer their notifications to be summarized, what information they find most valuable, and what characteristics define a useful and effective summary from their perspective. Without these insights, summarization systems risk failing to support users in efficiently processing the notifications that accumulate on their devices.

To address this research gap, we conducted the first study that explores this issue. Specifically, we seek to address the following research questions:

- RQ1: How would smartphone users like their notifications to be summarized?
- RQ2: What information do users expect and desire to see in a notification summary?

To answer these research questions, we developed NotiSummary, an Android app that uses GPT to generate notification summaries. NotiSummary employs a default summarization prompt but also allows users to edit and customize their own prompts, enabling us to observe how users preferred their notifications to be summarized. We deployed NotiSummary on Google Play Store and observed real-world usage over a three-month period, focusing on how users guided GPT in summarizing their notifications. The results show that users' created prompts reflected diverse prompting strategies, capturing a broad range of needs and expectations for notification summarization. However, sustained adoption and engagement were relatively low, suggesting a potential mismatch between the system-generated summaries and user expectations. This observation further strengthened our motivation to answer our research questions—specifically, to understand where the generated summaries fell short in meeting users' needs and expectations. To explore this, we conducted a follow-up interview study after the in-the-wild deployment, aimed at uncovering the limitations of the GPT-generated summaries users received and identifying their expectations and preferences for effective notification summarization.

This study makes three primary contributions to the field of HCI and mobile notification systems:

- It identifies users' expectations and preferences for notification summaries, including three types of information to prioritize, three types of details to extract and disclose, and three methods for synthesizing notifications.
- It uncovers factors shaping how users expect notifications to be summarized, including their familiarity with the notifications, prior experiences with these notifications, and the interrelationships among notifications.
- It shows that users perceived notification summarization as more than just a means of content reduction—they envisioned it as functioning like a virtual assistant that supports personal information management, including managing tasks, revisiting archived content, and tracking activity trends for self-reflection.

2 Related Work

The rapid increase in the number and variety of smartphone notifications has led to a significant problem of notification overload [16, 51, 65, 68]. The sheer volume of notifications, coupled with frequent alerts and large amounts of accompanying information, has been shown to cause negative emotions such as annoyance, stress, and anxiety [4, 14, 16, 18, 23, 33, 50, 54, 55, 68], ultimately undermining users' well-being.

To mitigate the perceived disturbance caused by incoming notifications and the adverse effects of task interruption, prior research has explored a range of strategies. A common approach is to identify opportune moments for delivering notifications [8, 21, 27, 40, 45, 48, 49, 53, 58]. In particular, machine learning and statistical modeling techniques have proven effective in reducing cognitive load and perceived interruption by tailoring delivery times to users' context [13, 27, 32, 38, 40, 45, 47, 49]. Beyond timing, other strategies have focused on altering the mode of notification alerts or suppressing them entirely to reduce disruption [29, 36]. For instance, one study found that participants preferred silent delivery over deferred delivery to minimize unwanted disturbance [32], while another showed that completely suppressing alerts reduced inattention compared to allowing them [29]. Batching multiple notifications into a single, consolidated delivery has also been shown to reduce perceived interruption and alleviate anxiety [22].

Delivering notifications at opportune times not only reduces interruptions but also increases the likelihood that users have sufficient bandwidth and motivation to engage with them. When numerous notifications arrive at inconvenient moments, users may lack the time or incentive to process them fully. Therefore, besides optimizing delivery timing through machine learning techniques, prior work has also introduced deferral features that allow users to postpone notifications until they are ready to engage [3, 34, 64]. However, even when deferred to more convenient times, users may still lack the capacity or willingness to review all incoming content. While a recent study suggests that users can be incentivized to view relatively trivial notifications [10], other studies found that such messages are often dismissed without being read [43, 62, 65]. Similarly, Lin et al. [34] introduced a pinning feature for deferring notifications but observed that participants often merely skimmed messages, pinned those they considered valuable, and dismissed the rest to reduce clutter. While this strategy can efficiently manage large volumes of notifications, it also risks overlooking important information buried among dismissed content [43, 62].

Given the variety of notifications, researchers have also investigated users' notification preferences and how these preferences shape engagement. Generally, users prefer notifications that are urgent, important, or attractive [9, 28, 31, 33, 41, 52, 55, 58, 65]. Notifications from communication-related apps are especially favored [26, 31, 33–35, 41, 55, 65]. This preference is often shaped by the relationship users have with the sender [12, 15, 30, 33, 40, 41, 69], which further shapes their

attentiveness and even their speculations [12] about the notifications. To address these preferences, several studies have introduced features that allow users to categorize notifications [9, 26, 34]. Notably, rather than classifying notifications solely by app or sender, studies indicate that users may prefer grouping them by related topics [9, 34].

In addition to user preferences, various contextual factors have also been shown to influence notification engagement [10, 32, 33, 58]. These include location [19, 20, 33, 39, 66], social context [8, 31, 33, 47, 59], and ongoing activity [17, 18, 31, 39, 41]. For example, Lee et al. [31] examined how 20 contextual factors influenced whether users considered it an opportune moment to proceed through four stages of notification interaction: delivering the alert, glancing at the notification, reading its full content, and acting on it. Their findings showed that sender-related and content-related factors were most critical in determining whether users felt the timing was appropriate for further engagement, while environmental factors were more influential during the initial alert-delivery stage. Chang et al. [10] further explored how users' activity context shapes their motivation to read notifications.

Despite extensive research on mitigating notification overload and supporting notification management, relatively little work has examined how users want notification content to be distilled and presented in more concise and informative formats to support efficient information processing. We aim to fill this research gap through the present study.

3 Methodology

3.1 NotiSummary

We developed NotiSummary, an Android application that allows users to summarize their own smartphone notifications via GPT-3.5. Available on Google Play¹, the app supports both Chinese and English, catering to a diverse user base. Beyond its core summarization function, NotiSummary enables users to customize prompts, configure app and notification filters, and schedule summary generation, each designed for flexibility to meet individual user needs and preferences.

3.1.1 Notification Summarization. NotiSummary integrates the *gpt-3.5-turbo-16k* model from OpenAI's API services², using default model parameters. The data flow is integrated with Android's Notification Listener Service³, which collects and manages active notification data. To ensure consistency across disparate notifications, NotiSummary uniformly formats all notifications as: "App: [Source Application Name], Time: [Notification Post Time], Title: [Notification Title], Content: [Notification Content]." The formatted notifications are then concatenated into a single string, with each entry separated by a newline character. This organized string is then combined with a summarization prompt—defaulted to "Summarize these notifications in a brief statement"—which users can customize in the app settings.

When a summary is triggered, the combined data is transmitted to our backend server, which handles communication with the OpenAI API. If the data exceeds the token limit (16,385 tokens for *gpt-3.5-turbo-16k*), the server segments it into smaller chunks while ensuring that each notification remains whole, as shown in Figure 1. Each chunk is processed separately, and the resulting summaries are concatenated and refined through a final summarization step using the ```t1;dr``` prompt—the default summarization example prompt provided by OpenAI at the time of development. The final summary is then directly accessible through the main interface upon opening the app. As illustrated in Figure 2a, the bottom half of the screen shows the generated summary along with the prompt used to produce it, while the top half displays the current notifications mirrored

¹<https://play.google.com/store/apps/details?id=org.muilab.noti.summary>

²<https://platform.openai.com/docs/guides/gpt/chat-completions-api>

³<https://developer.android.com/reference/android/service/notification/NotificationListenerService>

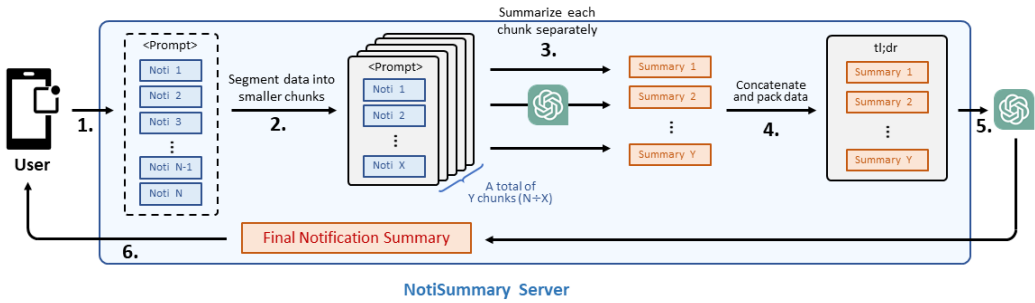
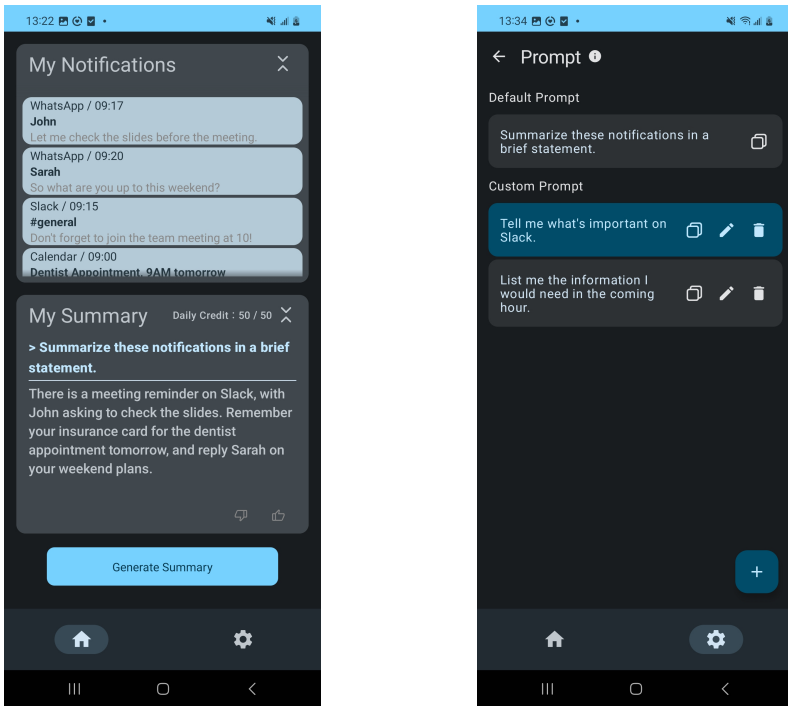


Figure 1. Flow Diagram of NotiSummary's Notification Summary Generation

from the system drawer as non-interactive cards, serving as a reference to the original content submitted for summarization.



(a) Main Interface

(b) Prompt Editing Interface

Figure 2. Interface of NotiSummary.

3.1.2 Customization Features. NotiSummary allows users to customize notification summaries based on their preferences. As shown in Figure 2b, the settings page lets users manage their custom prompts. A default prompt is provided, and additional prompts can be created through an editing dialog. Each prompt can be edited, copied, or deleted by clicking on corresponding icons. Prompts are displayed as a list, allowing users to easily select their preferred option for summary generation.

Users can also specify which notifications to include or exclude based on the source application and select which attributes—namely, source application name, post time, title, and content—to include in the summarization process. By default, all notifications and attributes are included for summary generation. In addition, the system allows users to schedule automatic summaries at specific times. Users can choose the days of the week for scheduled summaries, with daily summarization as the default, and enable push notifications for active delivery.

3.2 In-The-Wild Deployment of NotiSummary

To examine how users prefer LLMs to summarize their notifications in real-world settings, we launched NotiSummary on Google Play Store. Upon installation, an in-app survey gathered demographic data, including age, gender, country of residence, and how users' engaged with the app. Users consented to data collection via the in-app Privacy Policy. We logged each summary request and in-app interaction. Collected data included timestamps of summary requests and responses, prompts used for summarization, prompt editing records, word counts of the original notifications and eventual summaries, and the status of each summary request (e.g., success, network failure, or OpenAI API issues). To protect user privacy, actual notification content was not collected.

3.3 Interview Study

To gain insights into users' experiences with NotiSummary and their expectations for notification summarization, we conducted semi-structured interviews. The interview questions were designed to address both of our research questions, with a particular focus on identifying where GPT-generated summaries may have fallen short and how they could be improved. This inquiry was especially motivated by the low engagement observed during the in-the-wild deployment. We were interested in exploring whether limited usage stemmed from users not perceiving value in notification summarization, or from the generated summaries failing to meet their needs and expectations. We used participants' reported shortcomings and suggested improvements as a starting point, prompting them to reflect on how they would ideally like their notifications to be summarized and what information they would expect and desire in such summaries.

3.3.1 Recruitment and Participants. Participants were recruited through online forums and social media platforms in our home country. We recruited 20 participants, including both current and potential NotiSummary users. Seven (35%) were existing users, while the remaining were new users who interacted with the app as part of the study. Participants ranged in age from 21 to 52 years ($M = 28.35$, $SD = 8.39$), were gender-balanced (10 female, 10 male), and included 8 students and 12 professionals from various industries. Demographics are detailed in Table 1.

3.3.2 Study Procedure. We began with a briefing session for each participant, during which we outlined their rights, the features of NotiSummary, and the study procedures to ensure consistency across participants, regardless of their prior experience with the app. Before participation, all individuals provided informed consent and were instructed to use NotiSummary for a minimum of seven days, with prior usage being factored into this requirement. Participants were advised to use the app as if they had a notification summarization tool integrated into their daily lives, with no obligations or minimum requirements for summary requests or prompt edits during this period. Following this period of active engagement, we conducted semi-structured interviews, each lasting 60 to 90 minutes, which were held online according to participants' preferences. The interviews were recorded, transcribed, and reviewed for qualitative analysis. Participants were compensated with US\$16 for their engagement with the app and participation in the interview.

Table 1. Participant Profiles for Interview Study

Participant ID	Age Range	Gender	Occupation	Previous Engagement with NotiSummary	Reported Average Count of Present Notifications
P1	20-24	Male	Student	Y	11-15
P2	50-54	Male	Medical	Y	11-15
P3	25-29	Male	Software Engineer	Y	6-10
P4	20-24	Female	Student	Y	6-10
P5	20-24	Male	Student	Y	6-10
P6	20-24	Male	Student	Y	1-5
P7	35-39	Female	NGO	N	>21
P8	20-24	Male	Civil Servant	N	6-10
P9	35-39	Male	Business Support	N	6-10
P10	30-34	Male	Technicians	N	>21
P11	20-24	Female	Freelancer	N	6-10
P12	20-24	Female	Student	N	11-15
P13	40-44	Female	Business Support	Y	1-5
P14	20-24	Male	Student	N	11-15
P15	25-29	Female	Technicians	N	6-10
P16	20-24	Female	Salesperson	N	1-5
P17	20-24	Female	Craft & Related	N	6-10
P18	20-24	Male	Student	N	11-15
P19	30-34	Female	Technicians	N	16-20
P20	20-24	Female	Student	N	1-5

3.3.3 Data Analysis. We conducted inductive thematic analysis [6], refining our codes as new interviews were analyzed. The team collaboratively reviewed five initial transcripts to develop a preliminary codebook. Subsequent interviews were independently coded using this codebook, with regular meetings to ensure alignment. Discrepancies were resolved collaboratively, new insights were integrated, and the codebook was iteratively expanded. New or revised codes were cross-checked across transcripts for consistency. We used ATLAS.ti⁴, a qualitative research software, to manage the dataset throughout the analysis.

4 Observing In-the-Wild Interactions with NotiSummary

4.1 Overview

After releasing NotiSummary on Google Play Store, we promoted the app across multiple platforms to increase visibility, focusing on Facebook groups and forums related to Android applications, ChatGPT, and generative AI. Within three months after the app's launch, 130 users installed NotiSummary, of whom 107 initiated at least one summary request. Our data analysis focuses on the 79 users who both successfully received at least one summary and were at least 18 years old⁵. Together, these users contributed 87 prompts, offering insights into how they wanted their notifications summarized. Among the 79 users, 71 were located in Taiwan, while the remaining users were based in the United States, Austria, Hong Kong, Malta, and Poland. Most users were male (68), with seven identifying as female, one as non-binary, and three preferring not to disclose their gender. The self-reported age distribution ranged from 18 to 65 ($M = 33.03$, $SD = 10.97$), with a notable two-fifths (43.04%) in their 20s.

During NotiSummary's in-the-wild deployment, 21.52% (17) of these 79 users either created custom prompts or modified the default one, 13.92% (11) adjusted app exclusion settings, 7.59% (6) used the scheduler, and only one user employed the information exclusion settings. Overall, engagement with NotiSummary was limited: 72.15% of users (57) stopped using the app within the

⁴<https://atlasti.com>

⁵Ethical regulations from our university's Institutional Review Board.

first three days, while only 27.85% (22) continued beyond that point. Excluding failed attempts due to system or network errors, 499 summaries were successfully generated during the deployment period, with users receiving an average of 6.32 summaries (SD = 16.04, Min = 1, 25th = 1.0, Md = 2.0, 75th = 4.5, Max = 115). Of these, nearly three-quarters (73.95%) were manually triggered, while the remainder (26.05%) were scheduled in advance.

Although overall engagement was low, summary requests were relatively common among the 22 users who remained active beyond the third day. On average, 41.26% of NotiSummary app sessions—instances where users opened the app to view notifications—resulted in a summary request. This suggests that for these users, a substantial portion of their notification interactions involved a preference for viewing summaries rather than processing notifications individually.

Due to the low overall engagement with NotiSummary, we were particularly interested in understanding the causes of early disengagement. One plausible explanation is that users did not find the summaries helpful when the volume of notifications was low—a likely scenario when users frequently check and dismiss notifications, limiting their accumulation between summary requests. In such cases, users may have perceived the summaries as offering limited value in condensing information, thereby reducing their motivation to continue using the app. Supporting this explanation, we found that among users who disengaged within the first three days, 36.25% of their manually requested summaries involved only five or fewer notifications.

Another possible explanation emerged from our observation that generated summaries may not have aligned with users' needs or expectations. In response—and in parallel with our broader goal of answering our research questions—we were further motivated to examine how users preferred their notifications to be summarized by analyzing the prompts they created. In addition, we examined whether there were notable differences in prompting editing patterns between users who continued using the app and those who discontinued. These findings are presented in the sections that follow.

4.2 Observation of Users' Prompting Strategies for Notification Summarization

We found that the majority of users relied on the default prompt. Among the rest, 17 users either created new prompts or modified existing ones, resulting in 87 textually distinct prompts. Recognizing that some of these prompts differed only in phrasing while expressing the same intent, we removed duplicates with minor wording variations and merged those likely refined through trial-and-error experimentation, ultimately identifying 54 unique user-created prompts. These prompts fell into five main themes: filtering notifications, extracting details, querying essentials, synthesizing information, and tailoring the presentation of summaries, as illustrated in Table 2.

4.2.1 Filtering Notifications. Filtering notifications emerged as the most common strategy, reflecting a desire for the system to focus on specific content in its summaries. In particular, users' prompts often targeted notifications by their source apps or app types, aiming to include communication-related updates and exclude those deemed less important. For example, they issued requests such as: *"Please summarize notifications from LINE, Messenger, and Instagram in one brief statement"* and *"Exclude weather notifications."* Beyond filtering by app type, users also leveraged the LLM's semantic understanding to exclude notifications by topic. For instance, one prompt specified: *"Ignore notif about genshin impact from Discord,"* indicating a preference to eliminate only certain content from an app, rather than excluding all notifications from that app. This emphasis on selective filtering demonstrates users' desire to control precisely which notifications should (and should not) be included in their summaries.

4.2.2 Extracting Details. Beyond selecting which notifications to focus on, certain users wanted to ensure that specific details were always included in their summaries. Commonly requested details included the source or sender of the notifications, as illustrated by prompts such as: *"make sure*

Table 2. Observed Prompting Strategies of NotiSummary Users. Each concept in this table was observed in prompts of at least two users.

Objectives	Concepts	Example Prompts
Filtering Notifications	Type of App	Exclude weather notifications *
	Source Application	Please summarize notifications from LINE, Messenger, and Instagram in one brief statement. *
	Notification Topic	[...] Ignore notif about genshin impact from Discord. [...]
Extracting Details	Notification Source	[...] make sure to include who sends the messages to me
	App-Specific Content	[...] summarize the sentiment about the message
Synthesizing Information	Source	[...] If the notifications are from the same app, merge them together *
	Sender	Please summarize related content into key points by groups/official channels/communities *
Querying Essentials	General Information	Provide me the essential information I'll need in the near future. *
	Specific Needs	What are the current traffic issues? *
Tailoring Presentation	Listing	Briefly summarize notifications for each app in a bullet-point format, one app per line *
	Keyword	please collect all messages, write a summary and <i>list hot words like word cloud</i> . [...]
	Format Specification	[...] The sentence is in this form: app name: who sends the message
	Removing Irrelevance	[...] Do not tell me they are notifications, directly tell me the summary.

Note: Prompts marked with an asterisk (*) were translated from Chinese; others were originally created in English.

to include who sends the messages to me.” Some prompts also called for a degree of interpretation rather than mere extraction, as evident in requests like “summarize the sentiment about the message.” This indicates that users not only valued extracting key information but also appreciated deeper contextual or semantic meanings.

4.2.3 Querying Essentials. While the previous categories of prompts focused on filtering notifications or extracting specific details from individual messages, some users posed *queries* in the prompts, aiming to address immediate informational needs. These queries were often tailored to specific scenarios. For instance, users requested timely, actionable information with prompts like “Provide me the essential information I’ll need in the near future” or posed direct, purpose-oriented queries such as “How’s the traffic?”

4.2.4 Synthesizing Information. Rather than specifying which details to include, several users also sought to guide how the information should be combined into more concise forms. This need is evident in prompts such as “If there are many similar notif, aggressively summarize them together.” Strategies for consolidation included grouping notifications by the originating app, sender, or shared topic. For example, one user requested: “Please summarize related content into key points by groups/official channels/communities.”

4.2.5 Tailoring Presentation. Several users focused not only on *what* information should be included in their summaries, but also *how* it should be presented. Some requested structured formats, such as bullet points, to distinguish notifications by app or characteristic for enhanced readability. For example, one prompt specified: “Briefly summarize notifications for each app in a bullet-point format,

one app per line.” Others aimed to reduce the total amount of displayed information, opting for keywords rather than complete statements, as in: *“list hot words like word cloud.”* In addition, users often specified details to optimize the presentation of summaries. Examples included enforcing a particular format—*“The sentence is in this form: app name: who sends the message”*—or removing unnecessary or trivial elements—*“Do not tell me they are notifications, directly tell me the summary.”*

4.3 Disengaged Users Showed Minimal Prompt Editing Effort

In exploring differences in prompt editing patterns between users who continued using the app (5 users) and those who disengaged early (12 users), we observed that users who disengaged early also tended to invest little effort in modifying or refining their prompts, mirroring their limited engagement with the app overall. Specifically, the 12 users who disengaged early either created a single prompt with no subsequent modifications or made only minor, surface-level revisions. Half of them created just one prompt, with their prompts typically focused on *filtering notifications* of interest or *querying essentials*, without further refinement. The other six users either replaced their prompts entirely—creating new ones based on a different strategy rather than refining earlier versions—or made light edits, such as rephrasing instructions related to *synthesizing information* or *tailoring presentation*. However, these edits also remained superficial, primarily polishing the wording of earlier requests rather than introducing new requirements or specifying more clearly how notifications should be summarized.

In contrast, among the five users who continued using NotiSummary, three invested considerable effort in refining their prompts, iteratively developing instructions that incorporated multiple summarization strategies. For example, one user progressively enhanced his prompt by layering instructions to *synthesize information*, *extract details*, and *tailor presentation*. His final prompt read: *“Summarize the notifications for every app with a sentence. When the notifications are from a chatting app, make sure to include who sends the messages to me. The sentence is in this form: app name: who sends the message.”* This reflected a deliberate and evolving process to better align the summaries with his needs.

In another example, a different user iteratively refined a prompt for weather-related notifications, instructing the LLM to *filter out notifications* and *extract details*. After four iterations, the instruction evolved into: *“Ignore outdated weather forecasts, but describe the latest weather forecast in detail.”* Notably, this statement represented only a portion of a longer prompt that incorporated several other needs. Compared to users who disengaged early, these participants showed greater willingness to invest effort in iteratively refining and combining strategies. All three who engaged in this type of prompt development remained active for over a week, with two continuing to use the app for more than two months.

The remaining two users who did not iteratively refine their prompts each created only a single prompt or made minimal edits. However, unlike early disengagers, they continued using the app—likely because their initial prompts sufficiently aligned with their needs. Notably, both prompts were targeted at specific applications: *“Please summarize notifications from LINE, Messenger, and Instagram in one brief statement”* and *“Organize email content from Gmail.”*

Taken together, our prompt analysis reveals that users expressed diverse needs and employed multiple strategies for how they wanted their notifications to be summarized. These prompts offer initial insights into user preferences for summary content and structure. However, our interview findings suggest that users’ expectations and desires were more nuanced and complex than what their initial prompts alone revealed. This may help explain why many users were not satisfied with the summaries generated from their initial prompts and disengaged shortly thereafter. The following section presents our interview findings, offering deeper insight into how participants

wanted their notifications to be summarized and what types of information they expected and desired in a notification summary.

5 Qualitative Findings

Through our interviews, we found that while some usability issues—such as the need to leave the notification drawer to access summaries or delays in LLM-generated output—had caused inconvenience and contributed to participants’ dissatisfaction, the primary reason for disengagement stemmed from the summaries themselves. Participants held high expectations for notification summarization, anticipating substantial benefits such as saving time and reducing cognitive effort. When the generated summaries failed to meet these expectations—by being insufficiently informative or not meaningfully reducing the effort required to process notifications—participants felt less motivated to continue using the app.

During the interviews, participants reflected on specific instances where summaries fell short and offered concrete examples of how they would ideally like their notifications to be summarized. Their feedback revealed the types of situations in which they particularly desired summarization support, along with three types of notifications to prioritize, three types of information extracted and disclosed, and three preferred approaches to synthesizing related notifications. In addition, participants articulated expectations that extended beyond summarizing active notifications—envisioning notification summaries as part of a broader information management tool, akin to a virtual assistant that supports their daily routines. We present these findings in the sections that follow.

5.1 When Notification Summaries are Particularly Deemed Helpful and Critical

Participants described several scenarios where they found notification summaries helpful, or believed they would have benefited from them. Three key situations emerged: encountering a large accumulation of notifications, having limited time and mental or physical capacity to process them, and encountering notifications that were lengthy.

Participants often found summaries especially useful when notifications had piled up. While some noted they would check individual messages if only a few were present, a backlog of notifications led to a stronger desire for a summary to help filter and prioritize information. P4 described the overwhelming feeling after a long period of phone inactivity: *“If I haven’t checked my phone all afternoon, and then I open it, it’s like, ‘Wow, there’s so much!’ and I just get too lazy to check each one, so I delete everything and end up forgetting to reply or read something important. But with the summary, I can read through it and then clear everything out without having to go through each notification.”*

This reluctance to process notifications one by one was often more pronounced after activities that left participants mentally or physically drained. In such situations, notifications had accumulated, but participants lacked the energy to go through them individually. As P6 shared, *“When I’m really tired after working out and don’t feel like checking notifications one by one, I hope it can just pop up a summary for me. That way, I can quickly see what notifications I missed during the time I wasn’t looking at my phone.”*

The urgency to process notifications quickly was also felt during short breaks, when time and attention were especially limited. In these moments, participants aimed to maximize efficiency, viewing summaries as a way to quickly sift through messages. P10 explained, *“When I’m busy, with only ten minutes of break, I don’t want to scroll through notifications and open Gmail or LINE. I think summaries help me quickly get the information I need and filter out unnecessary messages like LINE notifications or other irrelevant things. It’s faster and more convenient to check a summary.”*

Finally, participants hoped summaries could assist them in processing lengthy notifications, which they often found overly detailed and inefficient to read in full. As P5 noted, *“It’s hard to*

quickly grasp the meaning of a lengthy notification, but if it's summarized into two or three sentences, I can get through it faster and understand its core message."

5.2 Types of Notifications to Include in a Summary

There was a broad consensus that a notification summary should mainly consist of three types of information, namely time-sensitive information, task-related content, and personally significant notifications.

5.2.1 Time-Sensitive Information. Participants emphasized that summaries should highlight time-sensitive content, especially notifications involving approaching deadlines or expiration dates. They felt that critical time-related details were often buried within a large volume of notifications or lengthy messages, making them difficult to spot and easy to miss. This concern was especially critical when participants were dealing with a large backlog of notifications, as reading through them one by one could result in delayed awareness of urgent matters. As P5 explained, *"If I read through it one by one, [...] I might realize halfway through that there was something important I should have responded to earlier, but by then it's too late."*

Several participants also noted that time sensitivity extended to seemingly less critical notifications, such as promotional offers with limited durations. For example, P18 suggested sorting summaries by urgency: *"If it's a promotional ad [...] it could go into the less urgent summary. But if it says it only lasts three hours, it should be in the important summary."* This concern stemmed from the perception that while more important notifications might not require immediate attention, those with tight deadlines could present limited-time opportunities they did not want to miss due to delayed awareness.

5.2.2 Task-Oriented Information. Participants also underscored the importance of surfacing task-related content in summaries. These tasks were often tied to ongoing projects or personal and professional responsibilities. While not always urgent, participants felt that such tasks warranted attention, as they involved follow-up actions and could lead to negative consequences if overlooked. As P1 noted, *"After a meeting, it would be especially helpful to review the notification summary to see what tasks require my attention and need to be completed."* Highlighting these responsibilities in a summary was viewed as a way to ensure that critical responsibilities were not missed amidst daily activities.

5.2.3 Personally and Societally Significant Information. Participants expressed a desire for summaries to include notifications that held personal or societal significance. Personally relevant notifications included those aligned with their preferences, opportunities, or events that directly affected their daily lives. For instance, P18 noted, *"If there's a notification about, like, a water shutdown in the library, something that would directly affect me, then it should be included in the summary."* Several participants also wanted summaries to highlight societally important notifications, such as breaking news or major public events. As P19 shared, *"If it's news that everyone agrees is significant, it should be flagged for me so I know to check it—otherwise, I might not open it at all."*

5.3 Types of Information Extracted and Disclosed from a Notification

Participants expressed distinct preferences regarding the type of information that should be extracted from notifications and disclosed in a summary. We observed that these preferences were influenced by their anticipation of the notification's content and criticalness, which were shaped by their familiarity with the notification source and their prior experience with similar notifications. Our analysis revealed three core disclosure types: 1) merely indicating a notification's presence, 2)

offering a concise “aboutness” to decide whether further engagement is necessary, and 3) providing a conclusion or key takeaway. Below, we examine each type in detail.

5.3.1 Presence of Notifications. At the most basic level, participants wanted to be simply informed of the *presence* of a notification—particularly for routine or less critical notifications. In these cases, the content was often familiar or repetitive, which participants could already anticipate without needing any further details. P5, for instance, shared, *“The notification from my cloud invoice app used to be quite long, but most of the details weren’t important. I just need to know I received an invoice, not the specific date or time.”* Similarly, P6 stated, *“If it’s a friend who always sends random stuff or pictures, it’s enough to just know they sent a message without seeing the details.”*

5.3.2 Aboutness of Notifications. The second type of information disclosure involved providing the *aboutness* of the notification. Participants expressed a need for this level of information to help them quickly assess whether and when to engage with the notification. Unlike the first level, where simply knowing a notification’s presence was sufficient, participants sought the “aboutness” to obtain enough detail to make an informed decision about further engagement. The presentation of “aboutness” varied in form, ranging from keywords and topics to short sentences. The preferred format depended on how easily participants could associate these pieces of information with the notification’s content. For less familiar notifications, participants expressed a preference for more detailed sentence-level summaries to establish context. P18 highlighted this need, saying, *“I get so many emails—some are important, some aren’t—so if it can summarize the email in a sentence to let me know whether it’s urgent, that would be great.”* However, for notifications with more familiar content, participants often found that a few keywords sufficed to convey the essence and inform their decision-making process. In this case, more information were perceived as unnecessary. P19, for instance, stated, *“For a promotion about beauty products, it included details about the product line and discount information. It felt a bit too long. I’d prefer it just shows the store name and discount, not the entire product line. Some information is too detailed and unnecessary. If it could simplify it down to the most important points, I could decide whether to click and see more.”*

Participants noted that receiving this aboutness information not only aided in decision-making but also helped them process notifications more efficiently. P10 explained, *“This [information] not only saves time but also helps me organize the information, including the people, time, and place, reducing the need to read through the entire chat. It helps me prioritize my work as well.”*

5.3.3 Conclusion of Notifications. The third type of information disclosure participants desired was a *conclusion* or key takeaway from the notification. This preference was particularly pronounced when participants felt reluctant to read through the entire notification and therefore wanted a quick conclusion that covers the key takeaway message. As P15 explained, *“When someone sends a lot of messages, I want it to let me quickly understand what they are trying to say without having to scroll through each message.”* Similarly, P6 highlighted the need to streamline lengthy emails: *“I’d prefer to skip all the formalities and just see the actual tasks and deadlines [...] I just need to know what to do and by when, rather than reading through a long, tedious work email.”*

For these types of notifications, participants felt that simply indicating their presence or providing an “aboutness” was insufficient. Based on prior experience, they expected that critical information was likely included and therefore desired key details to be disclosed as a conclusion to inform their decisions. Participants believed the LLM’s language understanding capability should allow such information to be succinctly presented, eliminating the need for further reading. As P18 noted, *“If the school sends an email but it’s condensed into ‘Typhoon Precaution Notice’ [in the summary], that’s not helpful to me. I feel like something important might get left out, so I would still have to read the full notification.”* Similarly, P12 recalled an instance where the summary failed to present specific,

actionable details, emphasizing the need for such information, explaining, *"My notification says '[city name] feels like 30 degrees,' and the summary tells me it's '[city name] weather and feels-like temperature,' but it doesn't give me the actual number. The number is what I really need to know."*

5.4 Approaches for Synthesizing Notifications to Efficiently Process Related Notifications

Participants desired different methods to synthesize information across multiple notifications. This synthesis, they believed, would enable them to process large volumes of notifications more efficiently and comprehend their collective meaning. Three distinct synthesis approaches emerged, as described below.

5.4.1 Grouping Notifications from the Same Source. The most frequently mentioned synthesis approach involved grouping and organizing notifications by their source, such as by application or sender. Participants consistently reported that this approach allowed for more efficient batch-processing of notifications compared to handling them individually. P4 articulated this preference, stating, *"It's difficult for me to find out which app a notification belongs to when they're all scattered [...] Grouping by app makes it easier to respond directly within that app."* This sentiment was echoed by P3, who subscribed to multiple merchant notifications, *"I receive notifications from over ten different merchants, and it's hard to keep track of all the promotions [...] I would prefer to know which merchant the ad is from before diving into the details."*

Beyond improving efficiency and convenience, many participants also valued the ability to track the *quantity* of notifications when they were grouped by source. This approach provided an overview of how many notifications had accumulated over a certain period, aiding broader understanding of notification activity. For instance, P20 highlighted, *"If I receive multiple Gmail notifications, I'd prefer them to be grouped together. This way, I can easily see how many messages I have and who sent them."* P1 also suggested that the frequency of communication-related notifications could offer insight into the sender's emotional state: *"If it tells me who sent the message and how many they sent [...] like if someone sends 20 messages in 10 minutes, I might infer that the person is angry or upset."* Similarly, P10 connected notification quantity/frequency to economic activity success, *"I use a system like a trading bot, and I set a price target. When the stock reaches that price, I get a notification indicating I might have sold. The frequency of these notifications helps me estimate how many times I've made a sale. [...] The higher the frequency, the more likely I am to have completed a transaction."*

Notably, several participants preferred providing only the quantity of synthesized notifications rather than a combined textual summary. They believed that forcing a synthesis of disparate details could be ineffective due to the varied nature of the notifications. As P1 noted, *"Even if all the notifications are about the stock market, each one has different details, so you can't really turn them into a single sentence. In that case, just saying 'four stock market updates' would be better."*

5.4.2 Grouping Notifications under the Common Themes. The second commonly desired synthesis approach was grouping notifications by overarching themes, regardless of their source. Many participants viewed this method as a more efficient way to process related content, particularly when notifications from different apps shared the same topic. For instance, P5, who received discount notifications from various apps, explained, *"If several apps are sending notifications about discounts, grouping them together would make it easier. That way, I could see all the discount notifications in one place, summarized in a single sentence, and quickly understand which apps are offering what discounts."*

Grouping by themes was also preferred by participants whose notifications from the same app could cover different topics, but they only wanted to selectively attend to certain types. In these

cases, organizing by theme rather than app allowed them to filter out irrelevant notifications. As P5 elaborated, *"Sometimes the job searching app sends notifications saying 'a company viewed your profile,' but I don't find that useful. However, if a company sends me a message, that's important [...] If all the notifications are grouped by that app, I'd have to go through each one. But if they're organized by content, I could see 'several companies viewed your profile today,' and I'd know they're less important. If it says 'one company sent you a message,' I'd know to pay attention."*

5.4.3 Consolidating Temporally Connected Notifications. Thirdly, participants expressed a need for a synthesis approach that could handle temporally connected notifications, particularly those stemming from conversations. These notifications often belonged to a longer thread, with messages logically connected over time. Rather than receiving each individual notification, participants preferred a summary that consolidated key highlights or takeaway messages from the entire conversation. This approach, they felt, would provide more meaningful insights while reducing the cognitive load of scanning through every message. As P4 explained, *"Some people tend to talk a lot and don't get straight to the point. I just need to know the main idea, not all the details."* Similarly, P12 emphasized, *"In chat messages, there's usually a lot of filler words, emotions, or details. For example, my friend was talking about her doll missing a mouth and used a lot of symbols and emojis [...] The summary should just tell me she's talking about the 'doll missing a mouth'"*

Participants also stressed the importance of maintaining context in temporally connected notifications. They felt that receiving a message without understanding its connection to previous ones could make the conversation feel fragmented and confusing. For instance, P8 stated, *"If I only get the current message, it might seem out of context. For example, my classmate mentioned 'using inappropriate language,' but without the context, it's unclear why. If the summary included previous messages to provide a fuller picture, it would make more sense."*

In addition, participants emphasized that failing to link current notifications with earlier conversations might lead the system to make incorrect assumptions about the content. For example, P18 mentioned an occasion where NotiSummary misunderstood a conversation, sharing, *"Sometimes it doesn't realize we're joking, so it totally misunderstands. Like when my friend said he feels old and can't keep up anymore, ChatGPT thought we were discussing [country name] tourism and aging issues."* P5 similarly observed that casual conversations often omit key details, which can result in inaccurate summaries when prior context is not considered: *"We might be talking about something that happened to someone, but since we don't mention their name—because we all know who we're talking about—the system could mistakenly think the person is talking about themselves, leading to inaccurate summaries."* Therefore, participants expected the system to reference earlier notifications or conversations when generating summaries. They hoped that by linking the current notification with its preceding context, the system could produce more accurate, cohesive summaries that avoid misinterpretations and reflect the true meaning of ongoing conversations.

5.5 Additional Expected Roles of Notification Summaries: Supporting Assistant-like Functions

Lastly, in addition to reflecting on how summaries should be generated, participants frequently linked their ideal notification summaries to broader functions beyond simple information extraction and synthesis, envisioning them functioning like virtual assistants that help manage, retrieve, and reflect on personal information. Our analysis revealed three additional roles that participants envisioned notification summaries could serve: assisting with task and information management, allowing them to revisit archived or dismissed notifications, and providing an overview of their activity trends.

5.5.1 Assisting Task and Information Management. Many participants saw notification summaries as tools to help them manage tasks and organize information. Given that some notifications contained task-related details, summaries were seen as a useful way to keep track of these tasks. P12 shared that the summary function acted as an unexpected reminder: *"I did not anticipate this, but the summary helped me review my conversations with others because I tend to forget what I've seen after reading their notifications or may not have read them at all. [...] But when the summary tells me the content, it reminds me that I haven't responded yet."*

Several participants expressed a desire for the summary system to proactively remind them of tasks or retain important information. P2 suggested, *"I might forget about a lecture I need to attend at the hospital, or I might have glanced over a notification without realizing how to handle it. [...] It would be useful if I didn't need to go into Gmail and could instead be reminded by the summary system to complete the document."* Similarly, P9 noted, *"I imagine a scenario where the summary shows something like a notification from Gmail saying that: 'your bank bill is due'. These pieces of information could be combined into a sentence, so I know I need to take care of it, especially since I might have accidentally swiped the notification away earlier."*

A few participants further envisioned the summary system acting as a more integrated task-management tool, preserving to-do items and helping them track their progress. P20, for example, noted, *"The system could add a 'check-off' feature. If I finish something, I could check it off. [...] For example, you could have a 'completed' section, so I know what I've already done."*

5.5.2 Retaining and Revisiting Archived or Dismissed Notifications. Many participants expected the notification summary system to retain a history of all notifications, assuming the system would preserve this archive rather than delete information after generating a summary. They envisioned this archive as a way to revisit notifications that may have been missed or overlooked. As P3 said, *"I want to check what is being missed in the past couple of days, or what offers I might have overlooked."*

This revisit possibility was emphasized by many participants, who found it particularly useful for retrieving notifications they had dismissed in the moment. While they initially cleared these notifications to reduce immediate clutter, they still wanted the flexibility to access them later. P17 explained, *"Sometimes I accidentally swipe away notifications, and other times I just don't like seeing a lot of them at once. I tend to skim through and then swipe them away quickly, but later I realize some of them might have contained important information, and I want to get them back."* Similarly, P19 noted, *"Sometimes there are important news notifications that I find overwhelming, so I clear them. But later, I wish I could know what those news notifications were about. If the summary feature could include notifications I dismissed but want to revisit, that would be great."*

5.5.3 Reviewing Notification and Activity Trends. Participants also envisioned the summary system to provide a comprehensive overview of their activity. Specifically, given that the summary system inherently tracks notifications, participants saw potential for it to offer insights into notification trends and activity patterns. P3 described this expectation, noting, *"I expect this app would track all my notifications over 24 hours and tell me things like which ads were most frequent, which chat rooms were most active, or which apps sent the most notifications, helping me understand how active they are."*

In addition to tracking notification activity, several participants saw the potential for the system to assist in tracking their own daily activities. For instance, P8 expressed interest in using the summary to organize financial activities, explaining, *"If, at the end of the day, it tells me how many notifications I received about transfers, that would be very useful for organizing my activities [...] I keep track of my expenses, so if it could summarize what I did today, it would help me record things faster."*

Some participants hoped that the system could provide insights over an even longer period, helping them evaluate their behavior and habits. For example, P9 reflected on how such summaries could help monitor spending, saying, *"If I often shop online and maybe I'm a little addicted to it, this could serve as a deterrent [...] For example, if I set it to 'track my credit card spending over 24 hours,' it could tell me something like, 'Yesterday you spent \$10,000.' If it keeps this data and shows me a summary over a week, I'd realize that I've spent way too much [...] It would help me reflect on whether my spending is too much and if these purchases are really necessary."*

The aforementioned desires suggest that participants envisioned the summary system as a tool not only for managing notifications but also for gaining insights into their broader activities and habits, promoting self-reflection and organization.

6 Discussion

Historically, the overwhelming volume and variety of notifications have posed challenges for users in processing information efficiently [16, 51, 65, 68]. Recent advancements in LLMs, recognized for their robust performance across language tasks, including text summarization [11, 70], now offer the potential for real-time notification summarization at scale. Our study provides initial insights into how LLMs can be leveraged to summarize notifications effectively that meet users' expectations, making a timely and relevant contribution to the field. In the following sections, we discuss these findings and their implications for future notification systems in detail.

6.1 Key Factors Underlying Users' Expectations for Notification Summarization

The prompts created and edited by users during the in-the-wild deployment provide initial insights into how users prefer LLMs to summarize their notifications. These prompts addressed various dimensions, including filtering, extracting, synthesizing, querying, and tailoring the forms of expression or presentation. Across these aspects, we observed a wide range of instructions and strategies, underscoring the diversity in user needs for notification summaries.

However, summaries generated using users' initial prompts often failed to meet their expectations. As a result, those who did not iteratively modify and refine their prompts typically disengaged from the app shortly after initial use. The combination of diverse needs and specific expectations likely contributes to the difficulty of designing a generic, one-size-fits-all prompt capable of satisfying the wide range of user goals for notification summarization.

Building on the initial observations of users' prompting strategies and editing behaviors, our in-depth interviews further uncovered key underlying factors shaping users' preferences and expectations for notification summaries. Specifically, the findings revealed nuanced needs regarding the types of information users wish to extract from individual notifications and how they expect information across multiple notifications to be synthesized into a coherent summary. Importantly, these factors significantly influence whether a summary is perceived as excessive, insufficient, or unhelpful. However, these critical considerations were not explicitly articulated in the prompts provided during deployment—and arguably, may be difficult for users to express clearly in such prompts.

The absence of these factors in user-provided prompts likely contributed to the observed gap between users' stated preferences and the summaries generated by the system, potentially explaining why users, even after customizing their prompts, often found the results unsatisfactory. This gap may also help to account for the limited sustained use of NotiSummary. Below, we discuss these key factors and their implications for the design of future notification summarization systems.

6.1.1 Desired Detail Disclosure Level Depends on Users' Anticipations of Notification Content and Criticalness. Our findings reveal how users would like to balance brevity and informativeness in

notification summaries. Participants preferred three distinct levels of information disclosure—*presence*, *aboutness*, and *conclusion*—and we found that these preferences were closely tied to two key factors: users' ability to anticipate the content of notifications and their perception of how critical the information might be.

Specifically, for routine updates or notifications whose content participants could anticipate, they typically preferred only indicating *presence* without further detail. This was often the case when no further engagement was necessary. However, when participants were less certain about the content but anticipated its potential criticalness for taking actions, they expressed a desire for *aboutness*—a brief glimpse into the notification's content to help them determine whether and when further attention was required.

In contrast, participants favored receiving a *conclusion* when they anticipated that a notification might contain critical information, but believed that the essential content could be concisely conveyed. In such cases, they preferred a distilled takeaway that allowed them to grasp the key point without processing the full notification.

Identifying these factors and their relationship to the types of information users seek to quickly assess notifications is crucial, as it extends prior work on notification attendance and management behaviors. Existing studies have demonstrated how users' actions, decisions, and preferences regarding notification visibility are shaped by the inherent characteristics of notifications, including their importance, urgency, and attractiveness [33, 41, 52, 55, 65]. Building on these studies, our study highlights that to support efficient scanning and processing of notifications—such as through summaries—future systems should account for two critical factors: (1) the extent to which users can anticipate the content of notifications, and (2) whether critical information can be effectively summarized in a concise format or requires further detail. These factors significantly influence how much information users wish to extract. Moreover, we observed that these factors are shaped by users' familiarity with the notification source and their prior experiences interacting with similar notifications, underscoring the importance of incorporating users' prior knowledge and expectations into notification system design.

In addition, these findings offer a nuanced perspective on notification non-engagement—that is, the decision not to open a notification. While prior research has linked non-engagement to specific contexts [10, 33] or notification types (often categorized by app) [12, 33], our results suggest that non-engagement may not solely stem from a lack of interest. Instead, it may reflect users' anticipation of the content or their perception that the notification title has already provided sufficient information, making further interaction unnecessary. Differentiating these reasons could be valuable for identifying notifications that may no longer need to be delivered, versus those that should continue to be presented despite users not engaging immediately.

Furthermore, our findings help explain why traditional text summarization techniques, such as extractive summarization—which condenses content by selecting sentences or phrases based on statistical and linguistic features [42]—may fail in producing a summary that meets user expectations. Participants in our study expressed specific and concrete desires for the types of information they wished to extract from notifications, shaped by the aforementioned factors. These findings underscore the need to tailor information extraction tasks to align with users' prior knowledge and experiences. Incorporating user-specific contextual information into prompts for LLMs is essential to generating summaries that effectively support user needs.

6.1.2 Synthesizing Notifications Based on Their Interrelationships. In contrast to the factors affecting the desired level of disclosure, another key aspect of notification summarization lies in users' desire to synthesize information from notifications that are related in various ways, such as originating from the same source, sharing common themes, or being temporally connected. This task goes

beyond traditional text summarization methods, including extractive approaches that concatenate key sentences [42] and abstractive methods that rephrase main concepts [25]. In the context of notification summarization, we observed that users' expectations for how notifications should be synthesized were influenced by the interrelationships among notifications. The most straightforward approach participants favored involved grouping notifications from the same source or sender, often to reveal quantity information that carried specific meanings for them, such as frequency or intensity. A more nuanced synthesis method involved merging notifications that shared a common theme or topic, regardless of the app from which they originated. This cross-app integration enabled participants to filter notifications based on relevance or interest in the topic, making it easier to selectively attend to what mattered most.

Prior research has documented similar user preferences for grouping notifications by source app or content [33, 34]. Our findings build on this by highlighting an additional preference for synthesizing *temporally connected* notifications—those that arrive sequentially and build upon previous messages in a logical progression. Participants expressed a desire for such notifications to be synthesized coherently, taking into account their temporal and contextual relationships. Rather than treating each message as an isolated update, they preferred a unified flow of information that preserved references and key details. This approach ensures that summaries accurately reflect the ongoing development of events, enabling users to grasp broader context without missing crucial connections. Previous studies have also suggested that users wish to batch-process similar notifications within a notification center [33, 34, 52, 65]. Our findings extend this by suggesting that systems can go beyond merely clustering notifications: they can provide synthesized summaries that distill the essential information from multiple related notifications.

Taken together, these findings suggest that notification summarization cannot rely on a one-size-fits-all approach. Instead, it should account for various interrelationships among notifications. The dissatisfaction some participants expressed with summaries they received—often feeling they resembled article summarizations—may reflect a failure to capture the nuanced connections between notifications. This indicates a need for more adaptive summarization methods that can reflect the diverse ways notifications are interrelated. However, the effectiveness of combining multiple synthesis methods warrants further investigation. Notifications frequently exhibit more than one type of relationship—some may be both temporally and thematically related—and determining the optimal synthesis approach becomes more complex when these multidimensional relationships exist. In such cases, the system may need to prioritize certain relationships over others. This raises important questions about how to effectively manage and synthesize notifications when relationships overlap or conflict.

6.2 Beyond Efficient Information Processing: Linking Notification Summaries to Task Management, Historical Retrieval, and Activity Tracking

Our findings indicate that some participants' expectations extended beyond the basic summarization of notifications, aligning the role of notification summaries more closely with that of virtual assistants supporting broader information management. Participants envisioned future summarization systems not only helping them quickly assess incoming notifications, but also assisting with retrieving past messages, tracking pending tasks, and identifying patterns or trends over time.

Many assumed that a notification summary system would inherently retain an archive of notifications and their summaries, making it both logical and valuable for accessing historical information when needed. A study by Chang et al. [10] found that users often revisit lower-priority notifications during breaks. Similarly, participants in our study expected summaries to prioritize urgent notifications while preserving access to less urgent ones for later retrieval. This behavior suggests that dismissing a notification does not always indicate disinterest but may reflect a preference to

delay engagement or a response to notification overload [43, 62]. A repository preserving dismissed notifications could help users avoid missing critical updates, offering greater control over clutter without added anxiety.

The potential for notification summaries to support task and information management also emerged as a significant theme. Participants frequently viewed summaries as a way to organize and track pending tasks, especially when notifications contained task-related content. This expectation aligns with prior research showing that users often rely on notifications as reminders [33, 34]. Our findings extend this by revealing that some participants imagined a more proactive role for summarization systems. Specifically, they envisioned a dashboard-like interface that offers a task overview and monitors progress, suggesting rising expectations for notification systems to play an active role in task management.

Additionally, several participants viewed summaries as valuable tools for tracking notification activity and broader daily routines, akin to fitness or screen-time apps. Some expressed interest in summaries that revealed patterns in the frequency or timing of specific types of notifications. Others imagined the system extending beyond notifications to track non-notification activities, enabling reflection on daily behaviors. This suggests that notification summaries could also support self-reflection, helping users identify behavioral trends and manage their engagement with notifications more mindfully.

Our findings suggest that participants envisioned notification summarization as more than just a tool for streamlining information processing. Instead, they saw it as a service that could support a range of broader goals, including task management, historical retrieval, and activity tracking. This perspective highlights an opportunity for future notification summarization systems to go beyond mere content reduction and instead serve as proactive tools that assist users in managing their digital interactions more effectively. However, these expanded expectations introduce design challenges. Adding functionality may enhance utility but also risks increasing complexity and reducing the simplicity users value in text summarization. Striking the right balance between feature richness and usability remains a key design consideration. Future research should explore how to integrate advanced capabilities while keeping summarization lightweight and intuitive to meet users' evolving needs.

6.3 Design Suggestions

Our findings offer several design implications for both developers of future notification systems and practitioners designing notifications for smartphone apps.

6.3.1 Balancing Glanceability and Informativeness in Notification Titles. For app notification designers, our results underscore the importance of designing notification titles that enhance glanceability. In many cases, the notification title itself can serve as a brief summary. Practitioners should consider the three types of information disclosure—*presence*, *aboutness*, and *conclusion*—when creating titles. Achieving a balance between brevity and informativeness is key. For notifications where brief keywords suffice to provide context, overly detailed titles might overwhelm users and reduce engagement. Conversely, for notifications where users may seek further details, providing enough information in the title can help users decide when to engage. For less critical notifications, the goal should be to allow users to quickly process the titles within the notification center, without feeling the need to open the app. There may be a trade-off between providing informative titles and driving app engagement, as more informative titles may reduce the need to open the app. However, considering the growing issue of notification overload, which can lead to users dismissing notifications altogether [34, 65], designing more informative titles could better meet users' information needs without requiring full interaction with every notification.

6.3.2 *Determining Suitable Detail Levels for a Notification Summary.* For developers of notification systems, we recommend enhancing the system’s ability to determine the appropriate level of detail for each notification. This improvement is relevant for both summary generation and notification title design. A potential approach could involve assessing users’ familiarity with and anticipated criticalness of notifications by tracking their notification interaction or other accessible content. Additionally, the system could employ human-in-the-loop methods [56, 60, 63, 67], such as directly asking users about their ability to anticipate notification content—a method previously used to gather user speculation about notifications [12]—their familiarity with the content, and/or reasons for non-engagement.

Furthermore, notification systems could prompt users to specify their preferred level of detail for specific notification types. While potentially challenging, collecting such data could help prevent inappropriate levels of disclosure. Overly brief titles or summaries may fail to help users make informed decisions, while overly detailed ones may be too time-consuming to read, thus reducing their utility.

6.3.3 *Structured Prompts and Procedures to Refine Summaries.* To improve the quality and relevance of summaries generated by LLMs, we propose moving beyond generic prompts by incorporating structured, specific information into the instructions for summarization. This approach would guide the LLM on which details to extract from each notification and how to synthesize them effectively. For example, the initial prompt could direct the system to assess relationships among notifications, such as whether they originate from the same app or sender, share a common topic or theme, or are temporally related. Identifying these relationships would enable the system to determine which notifications should be synthesized together.

If data on users’ familiarity with notifications, their anticipation of content, or their preferred levels of detail are available—either collected explicitly or inferred from prior interactions—such context could also be integrated into the prompt. This approach aligns with the Retrieval-Augmented Generation (RAG) [24] technique, where LLMs are equipped with relevant external context retrieved from knowledge sources. The prompts can then incorporate this guidance to produce more tailored and meaningful summaries based on user-specific needs.

Given the inherent complexity of summarizing notifications, a multi-step process may be necessary to ensure accuracy and coherence. This process could involve guiding the LLM through a sequence of subtasks, such as identifying relevant relationships between notifications and synthesizing content in stages. By breaking the summarization task into manageable steps, the system may be more likely to produce summaries that are both precise and aligned with user expectations.

6.3.4 *Facilitate Prompt Tuning By Users to Support Summary Customization.* We also recommend providing users with settings that allow them to specify their preferred levels of disclosure and how notifications should be synthesized, with options based on our findings. This flexible approach would cater to both users who prefer minimal information and those who seek more comprehensive notification details. Additionally, for advanced users with a deeper understanding of how to prompt LLMs, providing direct prompt customization allows greater control over summary generation. To support this, the system should offer structured guidance on refining prompts for different disclosure levels and specific synthesis approaches. Offering prompt templates and providing example configurations would further assist users in tailoring outputs to their expectations. This approach could strike a balance between offering out-of-the-box functionality for general users and enabling greater customization for those seeking more control over notification summarization.

6.3.5 *Integrating Task Management, Notification Archives, and Activity Trends.* Lastly, we suggest that future notification systems take a more active role in helping users track and manage tasks

that arise from notifications—functioning more like virtual assistants that support users in handling everyday responsibilities. This could be achieved by integrating task management features directly into notification summaries, including extracting task-related information, creating to-do lists, setting reminders, and monitoring task progress. Such features would transform the system into an active support tool for task and time management. In addition, we suggest that systems store notification histories and summaries in a retrievable archive, allowing users to revisit notifications easily. Implementing features such as search, filter, sorting options, and a “recently dismissed” section could help users quickly locate specific notifications, whether they were missed or dismissed in the moment. The system could also provide users with insights into their activity trends. This could involve generating periodic summaries that reveal patterns in notification activity, allowing users to review trends across specific periods, organized by app or notification type.

Given that notifications are a rich source of information related to users’ daily routines and task management, incorporating these features would position notification summary systems as effective personal assistants. Such systems could not only help users manage tasks more efficiently but also support their self-reflection on digital habits and overall productivity.

6.4 Study Limitations

This research has several limitations that should be considered when interpreting the findings. First, the sample size of users who consistently engaged with NotiSummary was relatively small and regionally biased, with the majority of participants being male and from the researchers’ home country. As a result, our findings may not fully capture the preferences or long-term engagement patterns of a broader population. There may be more diverse prompting strategies for LLM-based summarization and different expectations for notification summaries that were not fully represented by our participants.

Second, due to the limited usage of NotiSummary and the absence of contextual data collection, we were unable to conduct additional quantitative analyses on behavioral patterns. For example, we could not examine when users were most likely to request a summary based on contextual factors such as time of day and phone status. Although we collected usage logs (e.g., summary requests and prompt editing behaviors) during the in-the-wild deployment, we did not collect corresponding data from interview participants. Future research could incorporate richer contextual data to examine behavioral trends and align usage behavior logs with interview feedback to better understand the motivations behind user actions, enabling a deeper exploration of user interactions with notification summarization systems.

Third, NotiSummary operated as an external application rather than being embedded into the native notification drawer. As a result, the user experiences shared in interviews may not fully reflect how participants would perceive and interact with summarization features if they were directly integrated into the system-level notification interface.

Fourth, while our findings underscore that the quality of summary content plays a critical role in meeting user expectations, it remains unclear how the perceived value of summaries varies with notification frequency. Specifically, users who frequently check their phones may encounter fewer accumulated notifications, potentially reducing the perceived benefit of summarization. However, in our study, it was difficult to isolate the influence of frequency of phone-checking behavior from the role of summary quality, as both factors likely impacted user engagement. Therefore, as future systems improve summary quality informed by these findings, we suggest that further research investigate how notification volume and phone-checking behavior shape users’ perceived value of notification summarization.

Fifth, while this study focused on how users want their notifications to be summarized, it did not examine how such summarization should be implemented from a technical standpoint. We assumed

that state-of-the-art LLMs already possess sufficient capabilities to support effective summarization, and that realizing the proposed design primarily entails careful prompt engineering and system-level design. However, future work could move beyond this assumption to investigate whether model-level tuning or optimization is necessary to fully meet users' expectations and implement the proposed design suggestions more effectively.

Finally, our deployment utilized the GPT-3.5 model, which was widely available at the time of the study. While it was suitable for this study's goals, its performance may not have matched more recent models, such as GPT-4o. This may have contributed to less satisfactory summaries and lower engagement during the in-the-wild deployment, limiting the sample size for quantitative analysis. Additionally, the relatively lower quality of the summaries likely influenced participants' dissatisfaction during interviews. However, this limitation inadvertently provided an opportunity to elicit detailed feedback on where summaries fell short and what improvements users desired, offering valuable insights into their expectations for an ideal summarization system.

7 Conclusion

This study examined the potential of LLMs in addressing the ongoing challenge of notification overload by leveraging their capabilities to generate concise summaries, thereby reducing cognitive burden. Through an in-the-wild deployment and interviews, we explored how users prompted LLMs to generate notification summaries, as well as their expectations regarding how notifications should be summarized and what information should be included. Our findings contribute to the HCI literature by identifying three key factors that shape effective notification summarization: (1) the desired level of information disclosure, which depends on users' anticipation of notification content and perceived criticalness; (2) the methods for synthesizing notifications, informed by the interrelationships among notifications; and (3) users' broader expectations for summaries to assist with task and information management, revisiting archived notifications, and tracking activity patterns for reflection. These insights underscore the unique complexities of notification summarization, distinguishing it from conventional text summarization by emphasizing the need to account for notification context, user familiarity, and prior experiences. They also reveal evolving user expectations for notification systems, as participants increasingly envision these tools as virtual assistants that go beyond simple information delivery to support broader information management goals. This study not only demonstrates how LLMs can be leveraged to accommodate the multifaceted nature of notifications and meet diverse user needs but also suggests the emerging functions that future notification systems should incorporate to better integrate into users' daily lives.

Acknowledgments

We gratefully acknowledge the support provided by the National Science and Technology Council, Taiwan (110-2222-E-A49-008-MY3). We also deeply thank all the study participants, anonymous reviewers, and our colleagues, whose support and feedback contributed to the completion of this research.

References

- [1] Piotr D. Adamczyk and Brian P. Bailey. 2004. If not now, when? the effects of interruption at different moments within task execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. Association for Computing Machinery, New York, NY, USA, 271–278. doi:10.1145/985692.985727
- [2] Sumit Asthana, Sagi Hilleli, Pengcheng He, and Aaron Halfaker. 2025. Summaries, Highlights, and Action Items: Design, Implementation and Evaluation of an LLM-powered Meeting Recap System. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW176 (May 2025), 29 pages. doi:10.1145/3711074

- [3] Jonas Auda, Dominik Weber, Alexandra Voit, and Stefan Schneeeggass. 2018. Understanding User Preferences towards Rule-Based Notification Deferral. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI EA '18*). Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3170427.3188688
- [4] Brian P. Bailey and Joseph A. Konstan. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior* 22, 4 (2006), 685–708. doi:10.1016/j.chb.2005.12.009 Attention aware systems.
- [5] Harold Borko and Charles L Bernier. 1975. *Abstracting Concepts and Methods*. Academic Press, New York, NY, USA.
- [6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa arXiv:https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa
- [7] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. arXiv:2005.14165 [cs.CL]
- [8] Seyma Kucukozer Cavdar, Tugba Taskaya-Temizel, Mirco Musolesi, and Peter Tino. 2020. A Multi-Perspective Analysis of Social Context and Personal Factors in Office Settings for the Design of an Effective Mobile Notification System. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 15 (mar 2020), 38 pages. doi:10.1145/3381000
- [9] Tang-Jie Chang, Li-Ting Su, Yong-Han Lin, Jie Tsai, Zi-Xun Tang, and Yung-Ju Chang. 2024. "I Want Lower Tone for Work-Related Notifications": Exploring the Effectiveness of User-Assigned Notification Alerts in Improving User Speculation of and Attendance to Mobile Notifications. *Proceedings of the ACM on Human-Computer Interaction* 8, MHCI (2024), 1–25.
- [10] Xi-Jing Chang, Fang-Hsin Hsu, En-Chi Liang, Zih-Yun Chiou, Ho-Hsuan Chuang, Fang-Ching Tseng, Yu-Hsin Lin, and Yung-Ju Chang. 2023. Not Merely Deemed as Distraction: Investigating Smartphone Users' Motivations for Notification-Interaction. 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 650, 17 pages. doi:10.1145/3544548.3581146
- [11] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxian Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. 2024. A Survey on Evaluation of Large Language Models. *ACM Trans. Intell. Syst. Technol.* 15, 3, Article 39 (mar 2024), 45 pages. doi:10.1145/3641289
- [12] Yung-Ju Chang, Yi-Ju Chung, and Yi-Hao Shih. 2019. I Think It's Her: Investigating Smartphone Users' Speculation about Phone Notifications and Its Influence on Attendance. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) (*MobileHCI '19*). Association for Computing Machinery, New York, NY, USA, Article 14, 13 pages. doi:10.1145/3338286.3340125
- [13] Kuan-Wen Chen, Yung-Ju Chang, and Liwei Chan. 2022. Predicting Opportune Moments to Deliver Notifications in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 186, 18 pages. doi:10.1145/3491102.3517529
- [14] Kuan-Yin Chen, Hao-Ping Lee, Chih-Heng Lin, and Yung-Ju Chang. 2017. Who matters: a closer look at interpersonal relationship in mobile interruptibility. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) (*UbiComp '17*). Association for Computing Machinery, New York, NY, USA, 910–915. doi:10.1145/3123024.3124569
- [15] Yu-Ling Chou, Yu-Ling Chien, Yu-Hsin Lin, Kung-Pai Lin, Faye Shih, and Yung-Ju Chang. 2022. Because I'm Restricted, 2 – 4 PM Unable to See Messages: Exploring Users' Perceptions and Likely Practices around Exposing Attention Management Use on IM Online Status. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>New Orleans</city>, <state>LA</state>, <country>USA</country>, </conf-loc>) (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 529, 18 pages. doi:10.1145/3491102.3517616
- [16] Karen Church and Rodrigo de Oliveira. 2013. What's up with Whatsapp? Comparing Mobile Instant Messaging Behaviors with Traditional SMS. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Munich, Germany) (*MobileHCI '13*). Association for Computing Machinery, New York, NY, USA, 352–361. doi:10.1145/2493190.2493225
- [17] Fulvio Corno, Luigi De Russis, and Alberto Monge Roffarello. 2018. AwareNotifications: Multi-device semantic notification handling with user-defined preferences. *J. Ambient Intell. Smart Environ.* 10, 4 (Jan. 2018), 327–343. doi:10.3233/AIS-180492
- [18] Jon D. Elhai, Dmitri Rozgonjuk, Ahmad M. Alghraibeh, and Haibo Yang. 2021. Disrupted Daily Activities From Interruptive Smartphone Notifications: Relations With Depression and Anxiety Severity and the Mediating Role of Boredom Proneness. *Social Science Computer Review* 39, 1 (Feb. 2021), 20–37. doi:10.1177/0894439319858008 Publisher:

SAGE Publications Inc.

- [19] Anja Exler, Marcel Braith, Kristina Mincheva, Andrea Schankin, and Michael Beigl. 2018. Smartphone-Based Estimation of a User Being in Company or Alone Based on Place, Time, and Activity. In *Mobile Computing, Applications, and Services (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering)*, Kazuya Murao, Ren Ohmura, Sozo Inoue, and Yusuke Gotoh (Eds.). Springer International Publishing, Cham, 74–89. doi:10.1007/978-3-319-90740-6_5
- [20] Anja Exler, Marcel Braith, Andrea Schankin, and Michael Beigl. 2016. Preliminary investigations about interruptibility of smartphone users at specific place types. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 1590–1595. doi:10.1145/2968219.2968554
- [21] Joel E. Fischer, Chris Greenhalgh, and Steve Benford. 2011. Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (Stockholm, Sweden) (MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 181–190. doi:10.1145/2037373.2037402
- [22] Nicholas Fitz, Kostadin Kushlev, Ranjan Jagannathan, Terrel Lewis, Devang Paliwal, and Dan Ariely. 2019. Batching smartphone notifications can improve well-being. *Computers in Human Behavior* 101 (2019), 84–94. doi:10.1016/j.chb.2019.07.016
- [23] Hector Fuster, Andres Chamarro, and Ursula Oberst. 2017. Fear of Missing Out, online social networking and mobile phone addiction: A latent profile approach. *Aloma* 35 (06 2017), 30. doi:10.51698/aloma.2017.35.1.22-30
- [24] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. 2024. Retrieval-Augmented Generation for Large Language Models: A Survey. arXiv:2312.10997 [cs.CL] <https://arxiv.org/abs/2312.10997>
- [25] Som Gupta and Sanjai Kumar Gupta. 2019. Abstractive summarization: An overview of the state of the art. *Expert Systems with Applications* 121 (2019), 49–65.
- [26] Guhyun Han, Jaehun Jung, Young-Ho Kim, and Jinwook Seo. 2023. DataHalo: A Customizable Notification Visualization System for Personalized and Longitudinal Interactions. 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 648, 21 pages. doi:10.1145/3544548.3580828
- [27] Shamsi T. Iqbal and Brian P. Bailey. 2008. Effects of Intelligent Notification Management on Users and Their Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Florence, Italy) (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 93–102. doi:10.1145/1357054.1357070
- [28] Isabelle F Janzen and Joanna McGrenere. 2022. Reflective Spring Cleaning: Using Personal Informatics to Support Infrequent Notification Personalization. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 179, 16 pages. doi:10.1145/3491102.3517493
- [29] Kostadin Kushlev, Jason Proulx, and Elizabeth W. Dunn. 2016. "Silence Your Phones": Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (San Jose, California, USA) (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1011–1020. doi:10.1145/2858036.2858359
- [30] Hao-Ping Lee, Kuan-Yin Chen, Chih-Heng Lin, Chia-Yu Chen, Yu-Lin Chung, Yung-Ju Chang, and Chien-Ru Sun. 2019. Does Who Matter? Studying the Impact of Relationship Characteristics on Receptivity to Mobile IM Messages. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300756
- [31] Yu-Jen Lee, Meng-Hsin Wu, Chung-Chiao Chang, Xi-Jing Chang, and Yung-Ju Chang. 2024. Investigating User-perceived Impacts of Contextual Factors on Opportune Moments. *Proceedings of the ACM on Human-Computer Interaction* 8, MHCI (2024), 1–28.
- [32] Tianshi Li, Julia Katherine Haines, Miguel Flores Ruiz De Eguino, Jason I. Hong, and Jeffrey Nichols. 2023. Alert Now or Never: Understanding and Predicting Notification Preferences of Smartphone Users. *ACM Trans. Comput.-Hum. Interact.* 29, 5, Article 39 (jan 2023), 33 pages. doi:10.1145/3478868
- [33] Tzu-Chieh Lin, Yu-Shao Su, Emily Helen Yang, Yun Han Chen, Hao-Ping Lee, and Yung-Ju Chang. 2021. "Put It on the Top, I'll Read It Later": Investigating Users' Desired Display Order for Smartphone Notifications. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 520, 13 pages. doi:10.1145/3411764.3445384
- [34] Yong-Han Lin, Li-Ting Su, Uei-Dar Chen, Yi-Chi Lee, Peng-Jui Wang, and Yung-Ju Chang. 2024. Pinning, Sorting, and Categorizing Notifications: A Mixed-methods Usage and Experience Study of Mobile Notification-management Features. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 115 (sep 2024), 27 pages. doi:10.1145/3678579

- [35] Yong-Han Lin, Li-Ting Su, Uei-Dar Chen, Peng-Jui Wang, Yi-Chi Lee, and Yung-Ju Chang. 2023. Automatic, Manual, or Hybrid? A Preliminary Investigation of Users' Perception of Features for Supporting Notification Management. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* (Cancun, Quintana Roo, Mexico) (*UbiComp/ISWC '23 Adjunct*). Association for Computing Machinery, New York, NY, USA, 98–102. doi:10.1145/3594739.3610699
- [36] Hugo Lopez-Tovar, Andreas Charalambous, and John Dowell. 2015. Managing Smartphone Interruptions through Adaptive Modes and Modulation of Notifications. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta, Georgia, USA) (*IUI '15*). Association for Computing Machinery, New York, NY, USA, 296–299. doi:10.1145/2678025.2701390
- [37] Inderjeet Mani and Mark T. Maybury. 1999. *Advances in Automatic Text Summarization*. MIT Press, Cambridge, MA, USA.
- [38] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2016. PrefMiner: Mining User's Preferences for Intelligent Mobile Notification Management. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) (*UbiComp '16*). Association for Computing Machinery, New York, NY, USA, 1223–1234. doi:10.1145/2971648.2971747
- [39] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2019. NotifyMeHere: Intelligent Notification Delivery in Multi-Device Environments. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (*CHIIR '19*). Association for Computing Machinery, New York, NY, USA, 103–111. doi:10.1145/3295750.3298932
- [40] Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. 2015. Designing Content-Driven Intelligent Notification Mechanisms for Mobile Applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) (*UbiComp '15*). Association for Computing Machinery, New York, NY, USA, 813–824. doi:10.1145/2750858.2807544
- [41] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 1021–1032. doi:10.1145/2858036.2858566
- [42] N. Moratanch and S. Chitrakala. 2017. A survey on extractive text summarization. In *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*. IEEE, Piscataway, NJ, USA, 1–6. doi:10.1109/ICCCSP.2017.7944061
- [43] Alexander Neff and Philipp Wintersberger. 2022. An Experience Sampling Study to Evaluate Why Users Dismiss Smartphone Notifications. In *Adjunct Publication of the 24th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vancouver, BC, Canada) (*MobileHCI '22*). Association for Computing Machinery, New York, NY, USA, Article 18, 5 pages. doi:10.1145/3528575.3551446
- [44] Ani Nenkova, Kathleen McKeown, et al. 2011. Automatic summarization. *Foundations and Trends® in Information Retrieval* 5, 2–3 (2011), 103–233.
- [45] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K. Dey, and Hideyuki Tokuda. 2015. Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, Piscataway, NJ, USA, 96–104. doi:10.1109/PERCOM.2015.7146515
- [46] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shamer Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tamer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kopic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz

- Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O’Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]
- [47] Chunjong Park, Junsung Lim, Juho Kim, Sung-Ju Lee, and Dongman Lee. 2017. Don’t Bother Me. I’m Socializing! A Breakpoint-Based Smartphone Notification System. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) (CSCW ’17). Association for Computing Machinery, New York, NY, USA, 541–554. doi:10.1145/2998181.2998189
- [48] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: designing intelligent prompting mechanisms for pervasive applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp ’14)*. Association for Computing Machinery, New York, NY, USA, 897–908. doi:10.1145/2632048.2632062
- [49] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond Interruption: Predicting Opportune Moments to Engage Mobile Phone Users. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 91 (sep 2017), 25 pages. doi:10.1145/3130956
- [50] Martin Pielot, Karen Church, and Rodrigo de Oliveira. 2014. An In-Situ Study of Mobile Phone Notifications. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) (*MobileHCI ’14*). Association for Computing Machinery, New York, NY, USA, 233–242. doi:10.1145/2628363.2628364
- [51] Martin Pielot and Luz Rello. 2017. Productive, anxious, lonely: 24 hours without push notifications. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) (*MobileHCI ’17*). Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. doi:10.1145/3098279.3098526
- [52] Martin Pielot, Amalia Vradi, and Souneil Park. 2018. Dismissed! A Detailed Exploration of How Mobile Phone Users Handle Push Notifications. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) (*MobileHCI ’18*). Association for Computing Machinery, New York, NY, USA, Article 3, 11 pages. doi:10.1145/3229434.3229445
- [53] Benjamin Poppinga, Wilko Heuten, and Susanne Boll. 2014. Sensor-Based Identification of Opportune Moments for Triggering Notifications. *IEEE Pervasive Computing* 13, 1 (2014), 22–29. doi:10.1109/MPRV.2014.15
- [54] Yara Rizk, Vatche Isahagian, Merve Unuvar, and Yasaman Khazaeni. 2020. A Snooze-less User-Aware Notification System for Proactive Conversational Agents. arXiv:2003.02097 [cs.HC] <https://arxiv.org/abs/2003.02097>
- [55] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-Scale Assessment of Mobile Notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (*CHI ’14*). Association for Computing Machinery, New York, NY, USA, 3055–3064. doi:10.1145/2556288.2557189
- [56] Alison Smith, Varun Kumar, Jordan Boyd-Graber, Kevin Seppi, and Leah Findlater. 2018. Closing the Loop: User-Centered Design and Evaluation of a Human-in-the-Loop Topic Modeling System. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) (*IUI ’18*). Association for Computing Machinery, New York, NY, USA, 293–304. doi:10.1145/3172944.3172965
- [57] Namrata Srivastava, Rajiv Jain, Jennifer Healey, Zoya Bylinskii, and Tilman Dingler. 2021. Mitigating the Effects of Reading Interruptions by Providing Reviews and Previews. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI EA ’21*). Association for Computing Machinery, New York, NY,

- USA, Article 229, 6 pages. doi:10.1145/3411763.3451610
- [58] Dinara Talypova, Alexander Lingler, and Philipp Wintersberger. 2023. User-Centered Investigation of Features for Attention Management Systems in an Online Vignette Study. In *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia* (Vienna, Austria) (*MUM '23*). Association for Computing Machinery, New York, NY, USA, 108–121. doi:10.1145/3626705.3627766
- [59] Eran Toch, Hadas Chassidim, and Tali Hatuka. 2020. Can you Turn it Off? The Spatial and Social Context of Mobile Disturbance. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 91 (Oct. 2020), 18 pages. doi:10.1145/3415162
- [60] Dmitry Ustalov, Natalia Fedorova, and Nikita Pavlichenko. 2022. Improving Recommender Systems with Human-in-the-Loop. In *Proceedings of the 16th ACM Conference on Recommender Systems* (Seattle, WA, USA) (*RecSys '22*). Association for Computing Machinery, New York, NY, USA, 708–709. doi:10.1145/3523227.3547373
- [61] Dave Van Veen, Cara Van Uden, Louis Blankemeier, Jean-Benoit Delbrouck, Asad Aali, Christian Bluethgen, Anuj Pareek, Malgorzata Polacin, Eduardo Pontes Reis, Anna Seehofnerová, Nidhi Rohatgi, Poonam Hosamani, William Collins, Neera Ahuja, Curtis P. Langlotz, Jason Hom, Sergios Gatidis, John Pauly, and Akshay S. Chaudhari. 2024. Adapted large language models can outperform medical experts in clinical text summarization. *Nature Medicine* 30, 4 (01 Apr 2024), 1134–1142. doi:10.1038/s41591-024-02855-5
- [62] Aku Visuri, Niels van Berkel, Tadashi Okoshi, Jorge Goncalves, and Vassilis Kostakos. 2019. Understanding smartphone notifications' user interactions and content importance. *International Journal of Human-Computer Studies* 128 (2019), 72–85. doi:10.1016/j.ijhcs.2019.03.001
- [63] Zijie J. Wang, Dongjin Choi, Shenyu Xu, and Diyi Yang. 2021. Putting Humans in the Natural Language Processing Loop: A Survey. In *Proceedings of the First Workshop on Bridging Human-Computer Interaction and Natural Language Processing*, Su Lin Blodgett, Michael Madaio, Brendan O'Connor, Hanna Wallach, and Qian Yang (Eds.). Association for Computational Linguistics, Online, 47–52. <https://aclanthology.org/2021.hcinlp-1.8/>
- [64] Dominik Weber, Alexandra Voit, Jonas Auda, Stefan Schneegass, and Niels Henze. 2018. Snooze! Investigating the User-Defined Deferral of Mobile Notifications. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) (*MobileHCI '18*). Association for Computing Machinery, New York, NY, USA, Article 2, 13 pages. doi:10.1145/3229434.3229436
- [65] Dominik Weber, Alexandra Voit, and Niels Henze. 2019. Clear All: A Large-Scale Observational Study on Mobile Notification Drawers. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) (*MuC'19*). Association for Computing Machinery, New York, NY, USA, 361–372. doi:10.1145/3340764.3340765
- [66] Dominik Weber, Alexandra Voit, Philipp Kratzer, and Niels Henze. 2016. In-situ investigation of notifications in multi-device environments. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) (*UbiComp '16*). Association for Computing Machinery, New York, NY, USA, 1259–1264. doi:10.1145/2971648.2971732
- [67] Thomas Weber, Heinrich Hußmann, Zhiwei Han, Stefan Matthes, and Yuanting Liu. 2020. Draw with me: human-in-the-loop for image restoration. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) (*IUI '20*). Association for Computing Machinery, New York, NY, USA, 243–253. doi:10.1145/3377325.3377509
- [68] Tilo Westermann, Sebastian Möller, and Ina Wechsung. 2015. Assessing the Relationship between Technical Affinity, Stress and Notifications on Smartphones. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Copenhagen, Denmark) (*MobileHCI '15*). Association for Computing Machinery, New York, NY, USA, 652–659. doi:10.1145/2786567.2793684
- [69] Fengpeng Yuan, Xianyi Gao, and Janne Lindqvist. 2017. How Busy Are You? Predicting the Interruptibility Intensity of Mobile Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (*CHI '17*). Association for Computing Machinery, New York, NY, USA, 5346–5360. doi:10.1145/3025453.3025946
- [70] Haopeng Zhang, Philip S. Yu, and Jiawei Zhang. 2025. A Systematic Survey of Text Summarization: From Statistical Methods to Large Language Models. *ACM Comput. Surv.* 57, 11, Article 277 (June 2025), 41 pages. doi:10.1145/3731445
- [71] Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B. Hashimoto. 2024. Benchmarking Large Language Models for News Summarization. *Transactions of the Association for Computational Linguistics* 12 (01 2024), 39–57. doi:10.1162/tacl_a_00632 arXiv:https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00632/2325685/tacl_a_00632.pdf