Reducing the Search Space on demand helps Older Adults find Mobile UI Features quickly, on par with Younger Adults

Ja Eun Yu
jyu86@uic.edu
University of Illinois Chicago
USA

Debaleena Chattopadhyay
debchatt@uic.edu
University of Illinois Chicago
USA

ABSTRACT

As mobile user interfaces (UI) become feature-rich, navigation gets more complex. Finding features quickly starts demanding information-intensive strategies for decision-making—which can be challenging for older adults. Older adults examine fewer details, requiring fewer cognitive resources, when searching for information with a large number of alternatives. In this paper, we first systematically examine various ways to convey a reduced feature space. Visually emphasizing three relevant options helped older adults find a specific feature more quickly—on par with younger adults. Older users were more efficient when options were highlighted along with their context or with a weighted zoom than when just highlighted, and they also preferred these two the most. We then present *Nav Nudge*, an interaction technique that uses voice input and large language models to visually reduce the feature search space on demand—and discuss how older adults use it within a mobile map application.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in accessibility**; *Accessibility design and evaluation methods*; *Mobile devices*; **Empirical studies in interaction design**; **Accessibility systems and tools**.

KEYWORDS

older adults, mobile interface, accessibility, visual feedback, interaction technique, tech support

ACM Reference Format:
Ja Eun Yu and Debaleena Chattopadhyay. 2024. Reducing the Search Space on demand helps Older Adults find Mobile UI Features quickly, on par with Younger Adults. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3613904.3642796

1 INTRODUCTION

Following the recent global pandemic, it has become undeniably clear that older adults need better technology (tech) support to take full advantage of the digital ecosystem. The ever-growing number of services available *only* online—often via smartphones and tablets—has put older adults at a major disadvantage. Reasons, why older adults do not use most mobile apps, features, and services, neither benefit from them, are aplenty. They may consider them nonessential to their lifestyle [110], avoid them due to perceived risks [28, 112], or ignore them due to a misalignment of values [49]. Even when there is an intention to use, older adults face many barriers to actually using mobile tech over the long term: for example, a lack of onboarding support [84], a lack of continuing tech support from preferred sources [96, 97], frequent updates to the visual layouts of user interface (UI) components [30, 46], and complex app navigation [9, 70, 113, 114]. Specifically, that older adults find it difficult to navigate feature-rich UIs—much more than younger people—has been well-documented in the HCI literature [63, 70, 92, 114, 117].

A common mobile navigation issue for older adults is quickly finding a feature on a UI [84, 113, 114, 117]—which involves first determining what feature to look for and then locating it. The problem worsens when mobile apps have more than four or five functions (i.e., features) and sophisticated navigation designs to organize them. Studies show older adults fixate more on the central part of the screen than the peripheral top and left parts, where navigation cues frequently reside [63, 90]. They may also have difficulty shifting attention to tab menus, identifying scrollable tab menus, and exploring vertical menus, such as the side drawer or springboard to explore other menu options [63]. Because of this difficulty in identifying navigation structures, diffidence in exploring them further, and mixing up neighboring or similar-appearing icons and buttons [10, 54, 63], older users can get stuck—taking the same wrong steps over and over, never locating the intended feature, and ultimately giving up [113].

Suggestions have been made to simplify navigation techniques and design senior-purpose mobile apps [63, 77, 116]. While such "senior-friendly" apps may be adopted if health-related [88], in general, older adults rarely adopt senior-friendly technology variants that are designed only for them. Older adults do not view aging as disabling [52, 110] and may consider themselves "elderly" in certain everyday contexts, but not in all-encompassing life circumstances [8]. Furthermore, segregating the tech used by older adults from that used by younger people runs the risk of perpetuating negative stereotypes of aging [51], discouraging intergenerational use and collaborative experiences [42, 50], and ceasing any accidental tech support during co-use [73]. Other design solutions include offering a heatmap of recent feature selections, tooltips, or a list of past actions, using gaze fixations to cue hidden menu options [70], and employing a voice assistant to locate features [114].

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642796

No past studies, however, have demonstrated an approach that helps older adults find mobile UI features *quicker* than the norm. In a controlled study, where users explored a feature-rich desktop UI, older adults took more time to successfully select features than younger adults (6.5 vs. 4.2 s) as well as made more errors (55.5% vs 43.9% non-unique feature selections) [70]. To assist older adults in mobile navigation, researchers are exploring different types of tech support, for example, offering metaphorical instructions [117], inviting remote human helpers to record and share instructions [43], or asking a voice assistant UI-related questions when lost [114]. Some of these approaches suggest reducing the number of features per page [114, 117] or highlighting the target feature with a red rectangle as identified by a remote human helper [43]. In a Wizard-of-Oz study, we previously found that when older adults are lost exploring a mobile UI if the target feature is visually highlighted, then they mostly recover, reorient, and complete the task at hand [114]. While these are encouraging results, we still do not know if visual cues can make older adults more efficient or accurate in mobile UI explorations. Neither how to design cues that would help older adults the most. We examine these open questions here.

The premise of our work is that older adults find it hard to locate a particular feature on a feature-rich mobile UI because of how aging affects working memory and speed of processing [93, 94]. Of course, there could be many other reasons like users' lack of interest or familiarity, or an app's poor usability [3, 102]. But all other things being equal, older adults' reliance on simpler, information-frugal problem-solving strategies might be why they take more time to locate a correct feature when searching in a large feature space [71, 117]. As we age, knowledge accumulates over time, and general information, vocabulary, and comprehension increase. These reflect *crystallized intelligence* (Gc) [14, 39], which is known to influence tech adoption among older adults [21]. While Gc increases with age (mediated by factors like education, environment, and exposure), *fluid intelligence* (Gf) is known to decline. Gf is the ability to perceive relations and correlate, form concepts, reason, and abstract [14, 39]. For instance, a recent study found that older adults were most efficient in navigating a smartphone app when one or two features were displayed per view, rather than 4, 8, or 16 [117]. But that conflicts with the very nature of feature-rich mobile apps where several functions are organized in sophisticated navigation structures—conveniently accessible from most app views.

So, in this paper, we investigate *how to visually augment feature-rich mobile UIs to assist older adults in feature selection.* Our work contributes the following:

(1) a design exploration of visual cues to communicate a reduced set of UI features to older adults,
(2) empirical evidence from a controlled study comparing UI-feature-selection performance with eight types of visual cues between older and younger adults (n = 60),
(3) an interaction technique, *Nav Nudge*, designed based on the findings from our first study, and using large language models, to display a visually reduced search space (of features) on a feature-rich mobile UI, and
(4) insights from a small quantitative study of how older adults use *Nav Nudge* within a mobile map application (n = 10).

2 RELATED WORK

Last decade's research on older adults' mobile use largely focused on tech adoption [21, 95], the usability of touch gestures [40, 53, 59, 72], text-entry issues [37, 79, 101], and general UI design guidelines, like organizing information in a hierarchy emphasizing depth, ensuring screen-to-screen consistency, and minimizing the number of steps in a task [20]. Mobile tech ownership among older adults has significantly increased since then, from 18% in 2013 [100] to 76% in 2023 [86]. Despite that, the many conveniences that online apps offer now, from checking e-health records to ordering groceries to ridesharing, are rarely usable to older adults, arguably those who could benefit the most. Even after long-term ownership of a mobile device and ongoing regular use, on average, older adults use way fewer apps (3 to 4) and features than younger users [33, 61, 62, 84, 96, 116].

Recent research argues that such limited use may stem from the challenges older adults face in mobile UI navigation [63, 113]. Difficulties can include ambiguous affordances of UI components, like not knowing what parts to press, swipe, or drag [78, 113], unfamiliarity with icons [111], an incorrect or incomplete mental model of how an app works [35, 114], complex information structures [111, 117], or general anxiety during use [96, 97]—all of which can be further worsened by a lack of preferred mobile tech support [97]. Mental models offer organizing structures with which people can interpret the world [34]. Good usability requires helping users develop an accurate mental model of how a system functions [13]. Extending an already-acquired mental model requires fewer cognitive resources than acquiring a new one [34]. Acquiring a new mental model can be taxing for the working memory, and thus, be more difficult for older adults than younger adults [34, 81]. However, once acquired, older adults may be able to use their mental model as well as younger adults, except when limited by their working memory [34].

Most of today's mobile apps, like social media, rideshare, or cloud storage, significantly differ in functionality from any of their predecessor technologies, like email or telephone, that older adults may be familiar with. Indeed, studies have found that older adults may struggle to acquire an accurate mental model of social media and cloud storage apps when compared with younger users [4, 89]. Even when using a new mobile app from a familiar category (e.g., an Amazon app user using the Target app for the first time), older adults had questions about how the app works over 60% of the time [114]. Of course, it should be noted that even mobile apps offering similar functions can differ in UI design substantially, thereby, needing older adults to acquire newer mental models.

Older adults may also find it difficult to rapidly shift attention between different parts of a UI, for example, from the center of the screen (content) to the periphery (navbar, menus) [63] or from somewhere they expect to find a feature to another location (e.g., *"Why are there these things [create notebook/section] at the bottom? I would expect them to be at the top, maybe under View"* [70]). This tendency among older adults to miss UI components outside their immediate locus of attention may be due to the narrowing of their visual field, commonly known as the useful field of view (UFOV) [5, 6, 24, 113]. Furthermore, when presented with many bits and pieces of textual information, like on a feature-rich mobile UI, older adults may

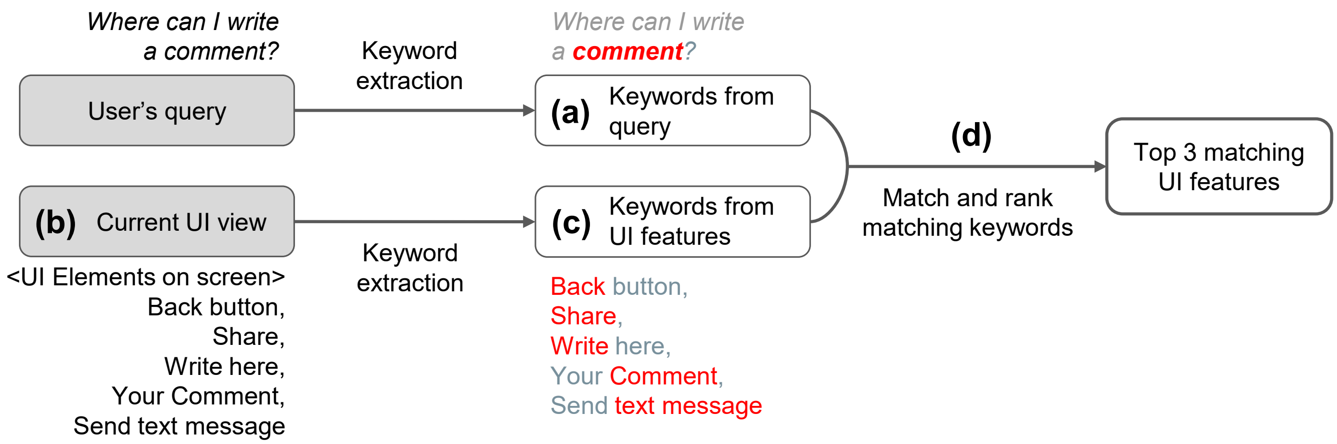


Figure 1: To reduce the number of features that older adults have to search to find one, we created a feature search space reduction technique that reduces the feature search space of any feature-rich mobile UI to three (see Figure 2). It first gathers information about the target function by extracting keywords from the user's query (a). Then it obtains the list of features currently available on their UI view (b) and keywords describing those UI elements (c). Finally, our method compares the two sets of keywords to identify three features on the user's current UI view that are most relevant to them at that moment (d).

struggle to selectively focus on relevant information, while simultaneously ignoring irrelevant information (i.e., inhibitory control) [12, 109]. This may explain why older users can keep using the same incorrect UI components over and over and get stuck in a cycle—even when the correct functions are visible all along (e.g., Figure 3, [113]). Or, why older adults performed the best (most efficient) and preferred the most when menus were narrower and deeper during UI exploration, e.g., only one or two features were present per page of a smartphone app [117]. Older adults are known to look up fewer details, and take longer to process them, than younger adults, and use simpler, less cognitively demanding strategies when decision-making [71].

Despite all the changes in cognitive capacities that may occur with getting older, older adults do not view aging as disabling [51]. They are a diverse demographic and chronological age does not uniformly predict physical and cognitive abilities. Moreover, situational aspects, such as environments, objects, and life changes might affect an older adult's perception of the self as "elderly" or "not" [8]. So, when designing for and with older adults, it is worth identifying how they function differently and when and what support is needed. The retirement age, when people become eligible for government benefits, has been historically a marker to define older adults (e.g., 65 or 67 in the US), but that age varies across the world (e.g., 60 in India and China). Most importantly, when, how much, and how long people work today have shifted largely with the turn of the century. Communities and organizations that serve older adults may even define them as people 50 and older (e.g., AARP [1]). In this work, we consider older adults as people ages 60-plus. Younger older adults may benefit more from using feature-rich mobile apps than the oldest old (80-plus) [87].

Broadly speaking, there are two types of tools that may assist older adults in navigating feature-rich mobile UIs: onboarding support [80, 103] and interactive tutorials [43, 104, 115]. To show the features of an app and how they function—when an app is first launched—overlays could highlight UI components and provide a short instructional snippet, either in general [29, 69, 80, 103] or about a task [55, 103]. But onboarding experiences are a one-time event. On the other hand, interactive tutorials can help older adults navigate a mobile UI during a task, with either instruction provided by a remote helper [43] or generated preemptively from a batch of text instructions about the app and its metadata [115]. However, creating these tutorials and maintaining them following each and every app update is resource-intensive. It requires people available to help out at the exact time of need and provide step-by-step instructions [44] or that correct instructions about how to do a task, in the user's variant of the app and mobile device, are already publicly available beforehand [115]. We explore an alternative design solution to assist older adults in navigating feature-rich mobile UIs: helping them quickly find a feature, at any time, by *visually reducing the search space.*

3 VISUALLY REDUCING THE FEATURE SEARCH SPACE OF A MOBILE UI

Finding a target feature on a feature-rich mobile UI can be hard for older adults [63, 70, 111, 113, 117]. While the exact reasons for that have not been directly investigated, these observations align well with what we know about the common cognitive behaviors of older adults: difficulty in selectively focusing on a few relevant details, while simultaneously ignoring other irrelevant ones [12, 109], and a tendency to problem-solve using fewer details than younger adults [71]. So, we hypothesize that reducing the search space of a feature-rich mobile UI, i.e., decreasing the number of features that an older adult must consider before deciding which one to select next, will make them more efficient and more accurate in mobile UI navigation.

Prior work in this area has tried to offer a reduced search space by either only displaying one or two features per page [117] or highlighting one or multiple features of a feature-rich UI with a red rectangle [43, 114]. The former approach is at odds with feature-rich mobile UIs while the latter requires a remote human helper to identify the feature(s) to highlight. Furthermore, it remains to be seen whether visually highlighting multiple features of a feature-rich UI actually helps older adults improve efficiency and accuracy in mobile UI navigation. In this section, we explore how to reduce the search space of a feature-rich mobile UI: (1) on demand, (2) automatically, without requiring a remote human helper, and (3) visually. First, we present our search space reduction technique, and then, discuss the different visual cues considered to convey the reduced search space to older adults.

3.1 Feature search space reduction

When having issues interacting with technology, older adults tend to verbalize their problems using question words, like 'what', 'why', and 'how' [26]. When the problem is specifically about finding a UI feature on the screen, they often use 'how' and 'where' questions, along with words describing the functions that they are looking for (e.g., "How to find the event?" [114]). Of course, there could be times when older adults do not know what function they are looking for or even how the app works. But we make the following design assumptions here: (1) the user has some prior experience with mobile apps and is not an absolute beginner, (2) their interaction problem is finding a UI feature that performs a particular function, and (3) when that problem occurs, the user provides some words describing what they are looking for. We do not anticipate many older users trying to explore and use feature-rich mobile apps at the very beginning of their mobile tech adoption and ownership [63, 97]. While there could be many other issues in mobile UI navigation [84, 96, 97, 113, 114], we focus on how to help older adults quickly find a feature when using a feature-rich mobile UI.

Our feature search space reduction technique consists of three steps: (1) determining what function the user is looking for, from a self-generated, free-form, verbal query; (2) extracting the list of features available in the user's current UI view, along with their location; and then, (3) identifying what UI features match the most with the user's query (Figure 1). Next, we discuss each of these steps in detail.

3.1.1 Keyword extraction from the user query. We experimented with standard natural-language processing (NLP) methods for keyword extraction from English-language phrases. Both unsupervised and supervised approaches, including embedding-based, frequency-based, feature-based, and graph-based methods, were examined. Because there is currently no public dataset available with example questions about how to use or explore feature-rich mobile UIs, we created a small dataset in-house: QUERY (240 instances; for more details, see Appendix B). Ground truth was established following annotations by the first author and then by two external annotators. Inter-rater agreement was excellent (Fleiss' κ = .97) [58]. A subset of the QUERY dataset was used for fine-tuning the unsupervised models, QUERY-tune (40 instances), and the rest was used for testing, QUERY-test (200 instances).

Other than the general NLP methods for keyword extraction, we tested a newer model, a transformer, fine-tuned specifically on language related to mobile UI: the Action Phrase Extraction (APE) model [64]. APE can extract UI operations and their corresponding UI elements from natural language mobile app instructions (e.g., 'app drawer' from *"Open the app drawer"*, 'settings' from *"Navigate to settings"*). APE was tested using 10-fold cross-validation (90/10 train/test split) and performed the best in comparison with other keyword extraction algorithms (Table 1)—likely because the latter are primarily designed for keyword extraction from much longer text snippets.

	Input type	
Model	*Query (N = 200)*	*UI element (N = 300)*
KeyBERT [36]	79.5%	81.9%
Rake [91]	64.5%	78.1%
Yake! [11]	66%	78.1%
TextRank [74]	71%	80.6%
APE [64]	85%	85.4%

Table 1: Accuracy of each keyword extraction model. The Action Phrase Extraction (APE) model was the most accurate for both the user's query and UI element descriptions.

3.1.2 Keyword extraction from the current UI view. This step required first obtaining a list of features currently available on the user's UI view and then identifying the keywords describing those UI elements. The information about the UI elements on the user's current view was obtained using the Assist API [23] that required platform-level permissions from Android. An *AssistStructure* instance retrieved the UI structure data in the form of text and view hierarchy. By traversing the view nodes in *AssistStructure*, we obtained the details about all the UI elements in the current view, including displayed texts, content labels, and their coordinates. The UI structure data, when retrieved in this way, not only includes visible elements that are rendered on the screen but also elements that are loaded but hidden, e.g., elements that would require scrolling or expanding to become visible. Content labels of a UI element may contain multi-word phrases (e.g., *"Like button. Double tap and hold to react", "Go to profile"*), thus, requiring keyword extraction.

We examined the same keyword-extraction algorithms as discussed in the previous section. Likewise, we created another small dataset in-house: UILABEL (320 instances; for more details, see Appendix B). A subset of the UILABEL dataset was used for fine-tuning the unsupervised models, UILABEL-tune (20 instances), and the rest was used for testing, UILABEL-test (300 instances). The names of the UI elements obtained from the *AssistStructure* were used as ground truth. Similar to user queries, the APE model provided the highest accuracy in extracting keywords from UI labels (Table 1).

3.1.3 Matching between the user query and current UI view. We used the Universal Sentence Encoder (universal-sentence-encoder v4) to determine a semantic similarity score between the two sets of keywords (user's query and UI elements) [15]. The UI element with the highest semantic similarity score would then be the most likely feature that the user is looking for. We evaluated the matching algorithm using an existing dataset (WIDGET CAPTION, [65]) and compared it with two other versions of the Universal Sentence Encoder (universal-sentence-encoder-large v5 and universal-sentence-encoder-lite v2). Each instance in the WIDGET CAPTION dataset provides the label of a primary UI element, one to three keywords about how a human described that UI element, and a list of all UI elements on the mobile screen where the primary UI element was encountered (for more details, see Appendix B). We used 3,712 UI elements and 10,149 human-annotated captions to conduct a 5-fold cross-validation (80/20 train/test split).

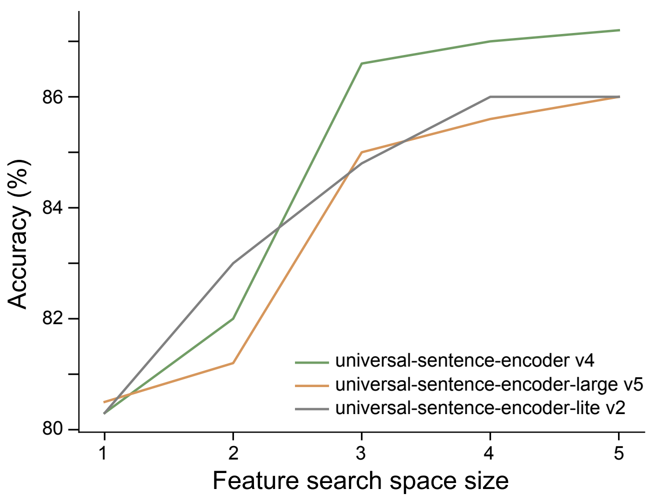


Figure 2: When matching the keywords between the user query and current UI view, we obtained the highest accuracy with universal-sentence-encoder v4 (87.2%). This accuracy did not significantly increase over a search space size of 3.

The purpose of our evaluation was two-fold: to determine the best encoder model and to determine the optimal size of the reduced feature search space. Selecting too many features could tax the working memory of older adults and defeat the purpose, but then again selecting too few could compromise the prediction accuracy. We considered reducing the search space from two to five features per view. In our evaluation, the default model exhibited the highest accuracy (87.2%, Figure 2), in particular, when the search space size was *at least three*. Given the clear elbow in the graph (Figure 2) indicating no significant increase in accuracy with a search space larger than three, we decided to reduce the feature search space to three relevant options.

3.2 Visually communicating a reduced search space

Our search space reduction pipeline determined *three* as the optimal size for the reduced feature space (Figure 2). So, when a user asks for help in finding a particular function (formulated as a verbal question containing words pertinent to what they are looking for, like [114], not [43]), we could list three relevant UI elements on their current view with a high likelihood that those three would include the correct feature (if that function was indeed available on that app page). These three features could be communicated to the user in various ways, e.g., visually on the app page [43, 114], as instructions on a separate app [115], or aurally. We chose to explore a way that would be minimally disorienting to users (e.g., not take them away from their current workflow in any way)—by visually

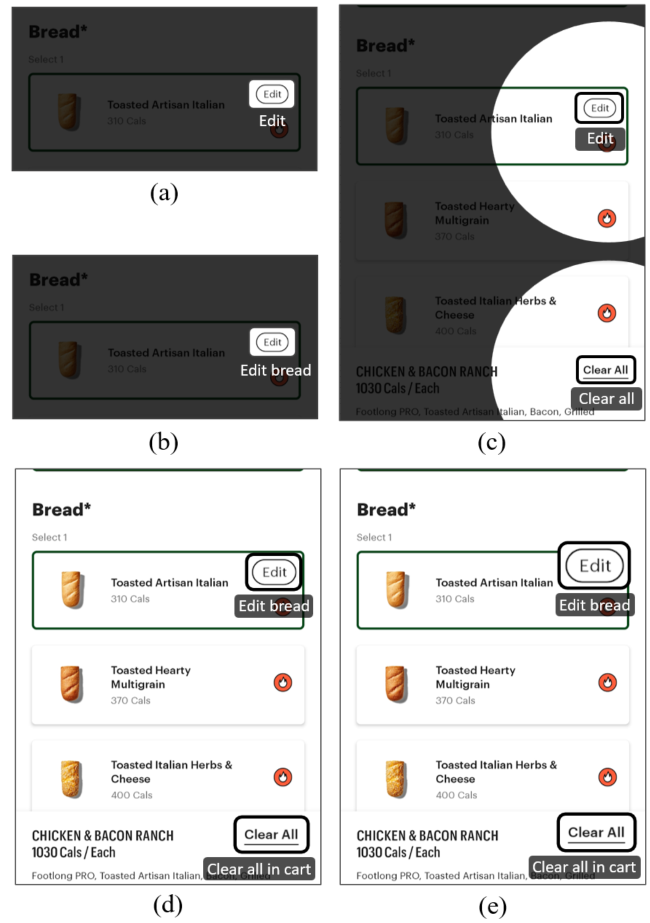


Figure 3: To convey a reduced feature search space to older adults, we designed six ways to emphasize the three UI elements predicted by our feature search space reduction technique (Figure 1): highlight (a,b), highlight with context (c), zoom (d), and weighted zoom (e), along with content labels (a, c) or manually-authored instructions (b, d, e). Any spatial emphasis could be paired with any textual emphasis.

emphasizing the three relevant features predicted by our pipeline (Figure 1) right on the users' current UI view (Figure 3).

Nevertheless, visually augmenting a feature-rich mobile UI can do more harm than good—e.g., add to the cognitive load of older adults or not be salient enough to direct their attention (selective focus). There is no obvious design solution or prior research on how to visually communicate a reduced search space on a feature-rich mobile UI to older adults. The only design used in the past, but not tested, is a red rectangle outlining a UI element [43, 114]. However, the capacity of visual working memory can vary across many aspects, including color, orientation, and shape of the visual cues [7, 38]. While it was suggested back in 1956, that people can remember about seven chunks in short-term memory (STM) [75], later work refined that number to only three to four chunks [19, 67, 108]. Even that number depends on the property of the visual cue

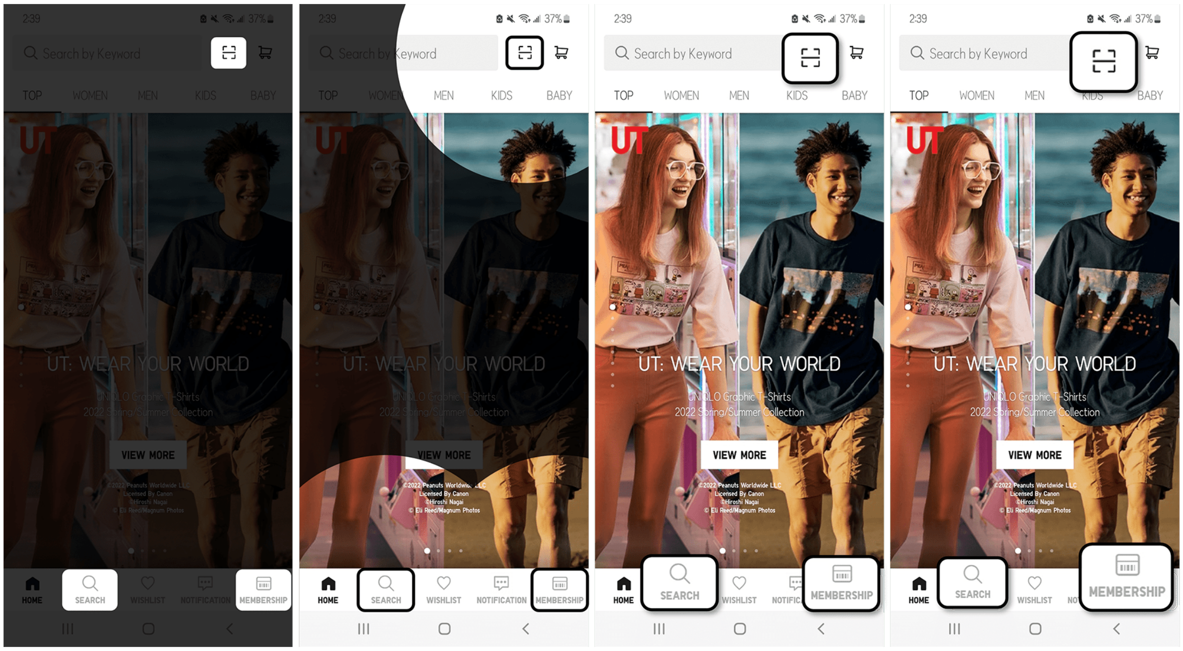


Figure 4: Four types of visual cues, as shown in Phase 1 of the experiment, for the task *"Scan the barcode of a dress you are checking out in the store to get more information about the product."*: (left to right) highlight (H), highlight with context (HC), zoom (Z), and weighted zoom (WZ).

(e.g., simple vs. complex cues) and individual differences [68, 76]. So, visually communicating a reduced search space that would truly help older adults navigate a feature-rich mobile UI is not a straightforward task.

To evaluate what type of visual cues would be helpful, we reviewed similar designs in other HCI tasks [16, 29, 103], and after several iterations, chose six cues (Figure 3): (1) highlight, (2) highlight with context, (3) zoom, (4) weighted zoom, (5) text showing the UI element's content label (from Android's Assist API [23]), and (6) manually generated text descriptions. Option (5) or (6) would accompany one of the options (1) to (4). Highlight (H) darkens the UI with a black overlay (opacity 20%) except for the three UI elements predicted to be relevant to the user's query (Figure 3a). Highlight with context (HC) does not darken the surrounding space (a circle of diameter equal to three times the longest side of the respective UI element) of the three UI elements to display their context (Figure 3c). Zoom (Z) magnifies all three UI elements equally, by a factor of 1.3 (Figure 3d). Weighted zoom (WZ) magnifies the three elements in proportion to their relevance to the user's query (i.e., rank, Figure 1), by 1.6, 1.4, and 1.2, respectively (Figure 3e). While content labels are strongly recommended in most platforms, they may not be of sufficient quality [106]. Thus, we also considered providing manually authored instructions alongside a spatial emphasis on the three features to reduce the search space for older adults, e.g., authored instructions in Figure 3d vs. content labels in Figure 3c.

4 UI NAVIGATION PERFORMANCE OF OLDER ADULTS

We set out to reduce the feature search space of a feature-rich mobile UI with the premise that doing so will make older adults more efficient and more accurate in mobile UI navigation. To test those hypotheses, we conducted a controlled experiment with older and younger adults. The study objectives were: (1) to identify the type of visual cue that helps older adults the most in UI navigation, and (2) to determine if reducing the search space helps older adults find mobile UI features quickly. In particular, we tested the following hypotheses for a feature-rich mobile UI:

H1. Older adults (OA) will be less efficient **(a)** and less accurate **(b)** in finding features than younger adults (YA).

H2. OA will be more efficient **(a)** and more accurate **(b)** in finding features with a reduced search space than without.

H3. There will be no large effect of age (OA vs. YA) on the efficiency **(a)** and accuracy **(b)** of finding features when using a reduced search space.

4.1 Method

This study used a mixed factorial design, with *age* as the between-group factor (old vs. young) and *visual cue* as the within-group factor (four levels, discussed later). The experiment was performed online via a responsive web application, where the task was locating a feature on a feature-rich mobile UI. Participants could use either a mobile device or a personal computer. Irrespective of the device used for the study, the feature-rich UI was always rendered as a mobile screen (Appendix A, Figures 10 and 11). We decided against using a mobile application to run this study because that would require either (1) platform-level permissions on personal mobile devices from all participants (unlikely) or (2) asking participants to use a new, unfamiliar device provided by us (likely to confound results and threaten internal validity, especially because older adults often have their mobile devices customized [48]).

4.1.1 Participants. Participants were recruited via university mailing lists and word of mouth. The study was approved by the university's institutional review board. A power analysis was conducted using R (PWR) for sample size estimation. We only found one relevant study that compared task performance between older and younger adults when using a mobile device and reported an effect size [60]. Although this study, conducted before 2010, did not use a touch-enabled smartphone, it investigated UI navigation. Thus, we used their reported effect size (large) for our power analysis. With a significance criterion of α = .05 and power = 80%, the minimum sample size needed would then be 20 in each group for a one-tailed *t*-test. To allow for instrumentation errors and attrition, we recruited a total of 60 participants, 30 younger (between ages 20 and 29) and 30 older adults (60-plus). The study inclusion criterion was that older participants must have at least one year of experience in using touch-enabled mobile devices (i.e., they cannot be beginner mobile tech users). We measured the mobile device proficiency of our participants using the MDPQ-14 questionnaire (Appendix A, Table 6) [85].

4.1.2 Tasks and Procedure. During the study, participants were provided with a task (e.g., *"You are ordering a Chai Tea Latte. Request cane sugar as an add-on", "Quickly view all the products that you looked at yesterday", or "You are pre-ordering snacks. Change the pickup location."*) and a static image of the corresponding feature-rich UI—with or without any visual cues (Figure 4). To design these tasks, we selected six feature-rich apps from the Android Play Store: Starbucks, Uniqlo, AMC Theatre, Ventra, Subway, and Audible. Four unique screens were captured per app and used to create 24 tasks in total (Appendix, Table 7). The screens were selected such that they represent feature-rich mobile UIs. Task descriptions were carefully worded to avoid any words or names of specific UI elements that were present in the corresponding screen (e.g., "sort by", "edit" or "view all"). For each task, a reduced search space was generated using the method described in Section 3.1 (Figure 1). We created a set of likely queries for each task and used them as input to the pipeline. The output was manually checked by the first author to confirm that it contained the correct UI feature. No adjustments were needed.

The experiment was organized into two phases. In Phase 1, the reduced search space was only emphasized using the spatial cues: highlight (H), highlight with context (HC), zoom (Z), and weighted zoom (WZ). Each task was *first* presented to the participants without any visual cue, i.e., the full search space was shown (baseline). If they successfully selected the correct feature that they would need to interact with to complete the task, the task was never shown again. If no correct selection was made within 13 seconds or participants clicked on the "Give up" button, the task would then reappear with some type of visual cue, randomly chosen from the four options (H, HC, Z, WZ). A time threshold was set to ensure that the correct answer was not chosen as a result of random interactions over an extensive time. We ran a pilot study with three young adults who only viewed the study screens without any visual cue. Their mean response time was 5.2 s. We know that older adults may take up to two times that of younger adults to complete the same task [27, 41]. Thus, we decided on a cutoff of 2.5 x 5.2 = 13 s.

In case participants did not find the correct feature in any task, they would encounter three tasks with each visual cue type—12 in total. If they found the correct feature in all 12 tasks, they would never encounter any reduced search space. We acknowledge that this experimental design did not guarantee that each participant would be exposed to an equal number of tasks per type of visual cue. So, to mitigate this limitation, we adjusted the number of tasks per type of visual cue in run time. For example, if a participant successfully completed eight tasks without any visual cue, each of the remaining four tasks represented a different type of visual cue. During Phase 1, for each participant, the two conditions with the lowest error rates were noted, which would then be used in Phase 2. In case of a tie, the condition with the shorter completion time was selected.

In Phase 2, we combined the two selected types of spatial visual cues with two types of text cues (content labels or manually-authored instructions). For each of the four conditions (two types of spatial cues x two types of text cues), the participant would perform three new tasks—first without any cues and then, if unsuccessful (gave up or timed out), with a cue. For example, if a participant performed the best with HC and WZ in Phase 1, they would receive four different types of cues in Phase 2: HC with content labels, HC with manually-authored instructions, WZ with content labels, and WZ with manually-authored instructions. In total, each participant would complete 12 new tasks in Phase 2.

Prior to the tasks, participants were provided with an introductory video that explained the concept of each mobile app. These videos did not include demonstrations of the app's actual interface or instructions on how to use specific features. For each task, the UI (with or without visual cue) was not displayed right away. The task description was first provided to the participant, along with a "Start" button. Participants were asked to click the "Start" button when ready. When clicked, the UI was rendered. During the task, participants could click unlimited times at unlimited places on the UI image. There was no time limit for tasks with visual cues but a 13-second time limit when no visual cue was provided. In addition, participants could select the "Give up" button at any time during a task to skip the task and proceed to the next one. When participants clicked on a correct UI element, a "Correct" prompt appeared (Appendix A, Figure 11), and the experiment moved on to the next task. After completing all tasks, participants were asked to provide their preference for the different types of visual cue and complete demographic and MDPQ-14 questionnaires.

4.1.3 Measures. We measured participant performance and preference. *Efficiency* was measured as the task completion time, the duration between UI rendered, and the correct UI element selected. *Accuracy* was measured as the error rate, i.e., the number of incorrect clicks divided by the total number of clicks. Participants ranked the four types of visual cue (H, HC, Z, WZ) in order of their preference. We also captured the time to failure (clicking "Give up") and the UI coordinates clicked during a task.

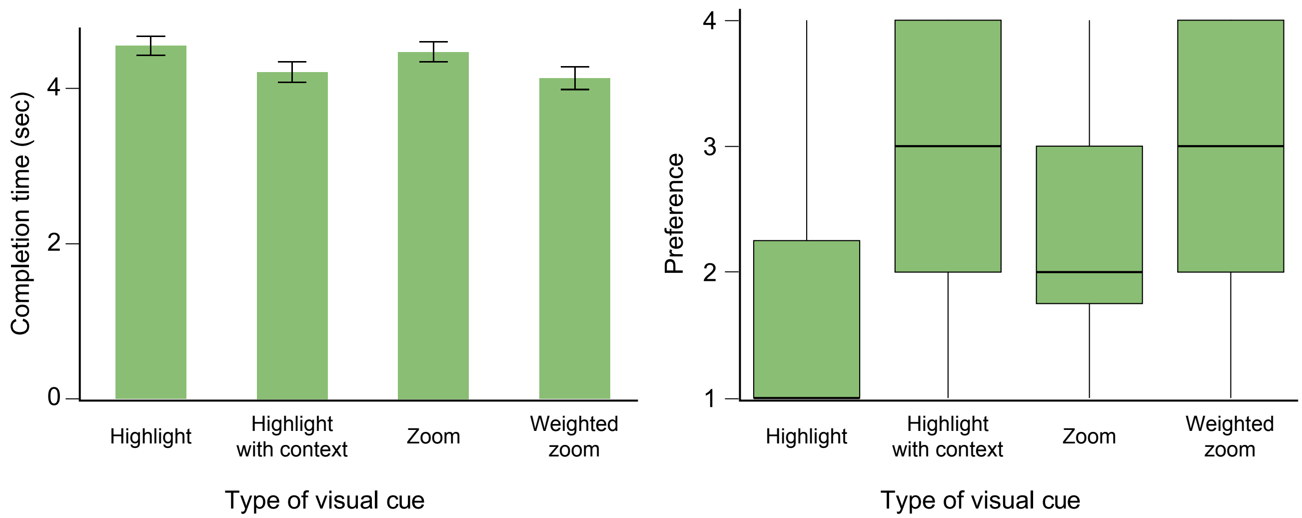


Figure 5: Task completion times (left) and visual cue preferences (right) of older adults from Phase 1 of our study.

4.1.4 *Data Analysis.* We conducted a univariate outlier analysis; trials were eliminated if the values were more than 1.5 times the interquartile range (IQR). Individual performance differences across participants were correlated with the overall performance measures. Completion time was positively skewed and log-transformed; thus, replications of unique experimental conditions were represented by their *median*. A GLMM with standard repeated measures REML technique was used that handled participants as a random factor. For GLMM, the R `lme4` package was used. We report F-statistic using type III ANOVA with *Satterthwaite approximation*, and pairwise comparisons (using pooled variance) with *Bonferroni* correction. The initial level of significance (α) was set to .05.

4.2 Results

Out of 60 participants, 56 completed all of the tasks and were considered for the analysis: 28 younger adults (YA; 10 women, Mdn_{age} = 22, *IQR* = 3, Mdn_{MDPQ} = 70) and 28 older adults (OA; 14 women, Mdn_{age} = 62, *IQR* = 3, Mdn_{MDPQ} = 40). Everyone owned a touch-enabled smartphone. As expected, older adults were less proficient in mobile device use than younger adults (U = 60, $p < 0.001$, r = 0.847). However, they were not absolute beginners (*Mdn* = 40, on a 14 to 70 scale), and thus, appropriate for our study about navigating feature-rich mobile UIs. There were no significant differences in familiarity with the six mobile apps used in our study between older and younger adults (Table 2). Most participants were either not or slightly familiar with the mobile apps. Twenty-three people (OA: 11, YA: 12) used mobile devices, and 33 (OA: 17, YA: 16) used computers to participate in the study. No significant differences in average efficiency or accuracy were found between the two types of devices used in the experiment—neither for the younger nor older cohort. In Phase 1, on average, each older and younger adult encountered 6 and 3.52 trials with a reduced search space, respectively. Table 3 shows a breakdown of trials by each type of visual cue, across Phases 1 and 2.

	Older Adults			**Yonger Adults**		
Mobile app	*NF*	*SF*	*VF*	*NF*	*SF*	*VF*
Starbucks	14	8	6	8	13	7
Uniqlo	22	5	1	17	10	1
AMC Theatre	16	8	4	9	12	7
Ventra	11	8	9	3	14	11
Subway	16	9	3	9	15	4
Audible	20	4	4	14	10	4

Table 2: Participants self-reported their familiarity with the six mobile apps as not familiar (NF), slightly familiar (SF), or very familiar (VF). On average, only 16% of older and 20.2% of younger participants were very familiar with the mobile apps used in the study.

4.2.1 *Visual cue performance.* A linear mixed effect model (LMM) found a significant main effect of spatial cues on task completion time, $F(3, 81) = 21.04, p < .0001, \Omega_0^2 = .91$ (Figure 5, left). Pairwise comparisons found that older adults were significantly more efficient with highlight with context (HC, M = 4.2, SD = 1.13) than zoom (Z, M = 4.74, SD = 1.06), and highlight (H, M = 4.97, SD = .97), ps < .05. Similarly, they were significantly more efficient with weighted zoom (WZ, M = 4.08, SD = 1.05) than Z and H, ps < .001. Error rates were not significantly different. No significant effects were found for the type of text cues (content labels vs. manually-authored instructions). Given these results, for all further analysis, we only considered the data with a spatial cue, that is data from Phase 1.

4.2.2 *Visual cue preference.* Older adults' preference for visual cue varied significantly by type, $x^2(3) = 13.89$, $p = 0.003$. HC (*Mdn* = 3) and WZ (*Mdn* = 3) were preferred significantly more than H (*Mdn* = 1), and Z (*Mdn* = 2), ps < .05 (Figure 5, right).

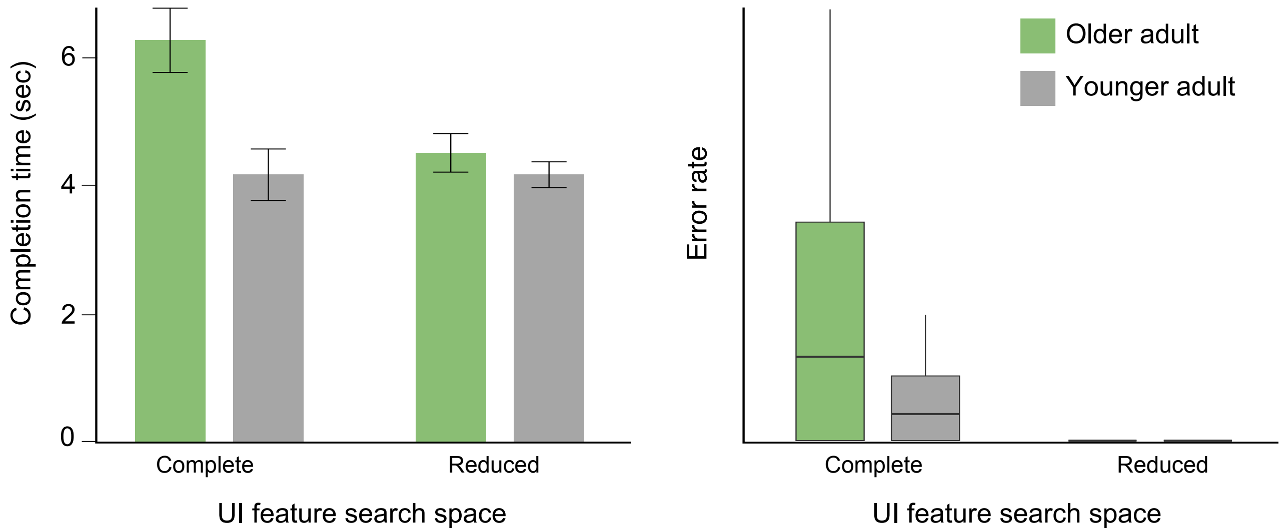


Figure 6: Older adults were significantly less efficient (left) and less accurate (right) than younger adults when finding a target feature using the complete search space. However, when the search space was reduced, those differences were not significant.

		OA	*YA*
Phase 1	Highlight (H)	1.5 (1)	0.86 (1)
	Highlight with context (HC)	1.5 (1)	0.9 (1)
	Zoom (Z)	1.54 (1.5)	0.86 (1)
	Weighted zoom (WZ)	1.46 (1)	0.9 (1)
	No cue	6 (7)	8.49 (8)
Phase 2	Content label	2.28 (2)	2.18 (2)
	Manually-authored cue	2.25 (2)	2.21 (2)
	No cue	7.46 (8)	7.6 (8)

Table 3: Mean (median) number of trials per type of visual cue per participant in Phases 1 and 2. On average, older adults (OA) encountered more trials with visual cues than younger adults (YA).

4.2.3 Navigation performance with a reduced search space. When the feature search space was not reduced—i.e., no visual cue of any kind was provided—older adults took significantly more time (M = 6.24, SD = .99) to find the correct UI feature than younger adults (M = 4.15, SD = .73), $t(54)$= 8.91, p < .001, r = 1.66 (Figure 6, left). Older adults also made more errors (Mdn = .125, IQR = .315) prior to finding the correct feature than younger adults (Mdn = .04, IQR = .09), U = 259, p = .006, r = .65 (Figure 6, right). **Both H1(a) and H1(b) were supported.**

Older adults were significantly quicker in finding the correct UI feature when the search space was reduced (M = 4.49, SD = 1.32) than when not (M = 6.24, SD = 0.99), $t(54)$= -5.57, p < .001, r = 1.2. Older adults also made fewer errors with a reduced search space, (Mdn = 0, IQR = 0) than a complete search space, (Mdn = .125, IQR = .315), V = 18.5, p = .001, r = .82. **Both H2(a) and H2(b) were supported.**

An LMM found a significant main effect of age $F(1, 54) = 27.19, p < .001$ and visual cue $F(1, 54) = 17.55, p < .0001$ on task completion time. Significant interaction effects were found for AGE X VISUAL CUE, $F(1, 54) = 17.30, p < .001$. However, a pairwise comparison found no significant difference in task completion times between older adults and younger adults when the search space was reduced, p = .41. Another LMM found a significant main effect of age $F(1, 54) = 8.42, p = .005$ and visual cue $F(1, 54) = 22.00, p < .0001$ on error rates. Significant interaction effects were found for AGE X VISUAL CUE, $F(1, 54) = 6.28, p = .015$. However, a pairwise comparison found no significant difference in error rates between older and younger adults when the search space was reduced, p = .32.

H3 states that there will be no large effect of age (OA vs. YA) on the efficiency and accuracy of finding features when using a reduced search space. To determine the smallest effect size of interest (SESOI), we followed standard guidelines established in the methods literature [57]. Without any quantifiable theoretical predictions to set an objective SESOI, we based our subjective SESOI on an earlier work examining a similar hypothesis [60]. As recommended in the literature [57, 99], SESOI was set as the effect size that the earlier study would have had 33% power to detect, which was d = .55. To test H3, we tested for the presence of an effect using an equivalence test [56]. A TOST equivalence test with *alpha* = 0.05 was found significant for completion time, $t(160.75)$ = -2.493, p = .007, based on equivalence bounds of -0.55 and 0.55. These equivalence bonds were derived from the SESOI, d = .55. Because sample sizes were unequal, and variances were unequal, a Welch's t-test was used. We can, thus, reject effects more extreme than d = 0.55. No large effect of age was found on the efficiency of finding features with a reduced search space. **H3(a) was supported.** Likewise, a TOST equivalence test with *alpha* = 0.05 was found significant for error rates, $t(152.37)$ = 1.787, p = .038, based on equivalence bounds of -0.55 and 0.55. **H3(b) was supported.**

Participant	Age	Gender	MDPQ-14	Task	Question(s)	# of features emphasized
P1	60	F	40	1	Where to enter the destination address?	3
				2	Hotel?	2
				3	How can I send?	1
				3	Send the current location.	2
P2	66	F	36	1	How can I see the public transit option?	1
				4	How can I see the bookmark list?	1
P3	62	M	48	2	How can I bookmark this place?	1
				3	Where is a sharing button?	1
				4	Where is the bookmark that I have created?	1
P4	62	M	50	3	How can I share my location?	2
P5	63	M	37	1	Where can I find the public transportation option?	2
				4	How can I remove this bookmark?	2
P6	61	F	56	-	-	-
P7	63	F	38	1	I want to add a starting point	2
					Set this as a destination	2
					Can I see the transportation?	2
				4	How can I remove this bookmark?	1
				4	Where is the place I bookmarked?	2
P8	66	M	38	-	-	-
P9	61	M	46	2	Is this star icon a bookmark?	1
P10	60	F	42	1	I need public transit information	2

Table 4: Participant demographics, their MDPQ-14 scores, the questions they asked while using Nav Nudge, and the number of features emphasized as a result. F = female, M = male.

5 NAV NUDGE

Based on the study results in Section 4, we designed Navigation Nudges, or *Nav Nudge*, an interaction technique that reduces the search space of a feature-rich UI using visual cues, on demand—to nudge older adults in the right direction. In our study, older adults performed equally well with weighted zoom (WZ) and highlight with context (HC) and there were no differences in terms of what text cues were provided alongside spatial cues. So we decided to develop an Android app using both WZ and HC and use the default content labels provided via the *AssistStructure*.

5.1 Design and Development

Nav Nudge was designed on the Android 13 platform. We made one significant update to our existing feature search space reduction pipeline (Figure 1). We updated our keyword extraction method, for both the user's query and UI labels. Instead of using APE [64], we used OpenAI's text-davinci-002 model [83], by directly sending prompts to the model from within the app ("Extract a keyword from the following sentence: <INPUT-PHRASE>, Keyword:"). When tested with Query and UILabel datasets, the new method achieved an accuracy of 94% and 93.4%, respectively. The rest of the pipeline remained the same (Figure 1).

Running the entire pipeline on a mobile device would be infeasible, so a dedicated server was established that communicated with the app via a REST API and computed the semantic similarity score. UI elements with a score of more than .3 were selected for visual emphasis, using either weighted zoom or highlight with context. If more than three elements were selected, those with the three highest similarity scores were emphasized. Next, we discuss how older adults used Nav Nudge within a mobile application.

5.2 Evaluation

To evaluate Nav Nudge, we chose a mobile app that would be conceptually familiar to older adults but not commonly used—Organic Maps. Organic Maps is an open-source software, which allowed us to manually add descriptive labels to UI components that weren't labeled. We chose to use speech as the input mode for the user's query, similar to prior work [43, 114].

5.2.1 *Method.*

We conducted a small task-focused study to explore how older adults would use Nav Nudge outside the lab, in an unconstrained setting. Our prior work has found that older adults often struggle to use mobile maps [113]. We adopted our lookup tasks from [113], and designed two additional tasks that would require participants to find a hidden feature on the screen and undo a previous interaction, respectively:

(1) Find out how long it will take to travel from Chicago O'Hare Airport to Chicago Union Station using public transit.
(2) Identify nearby hotels to Chicago Union Station without using a search box and bookmark one of them.

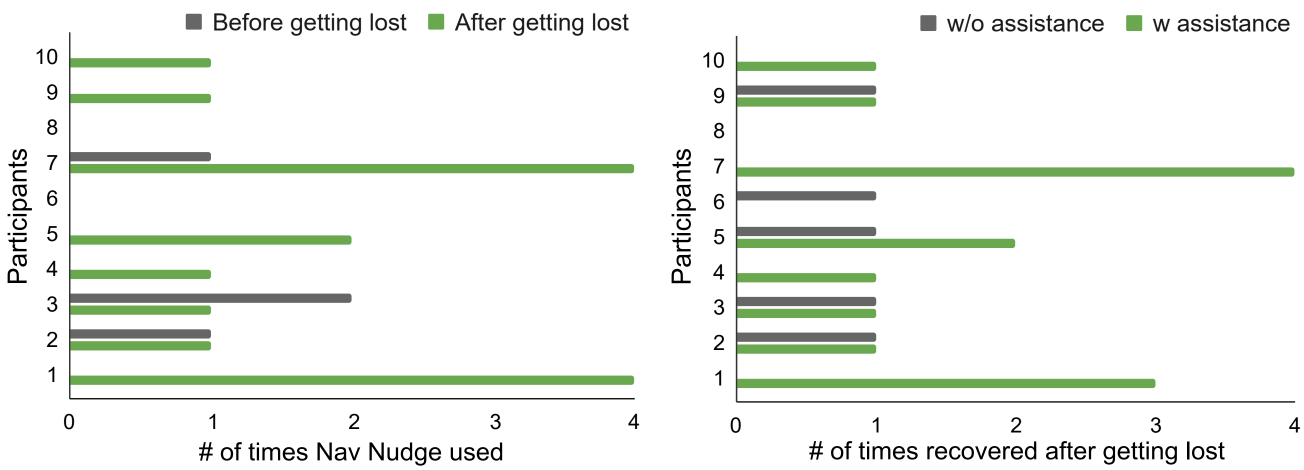


Figure 7: Eight out of ten participants used Nav Nudge during our study. Participants used Nav Nudge significantly more after getting lost than when not lost (left). Among participants who used Nav Nudge at least once, it was used significantly more times than other forms of self-exploration when recovering from getting lost (right).

(3) Share your current location with <first author's email address>.
(4) Remove the hotel that you bookmarked in Task 2 from your saved list.

This study was conducted face-to-face at a place and time chosen by the participant. Participants were asked to complete the tasks using a Samsung Galaxy S23 Android phone. Before the study, Nav Nudge was introduced to them and a practice session was provided. In addition, adjustments were made to the screen brightness and font size according to participants' preferences. This was not a think-aloud study. We recorded the queries participants made to Nav Nudge and the device's screen during the study.

5.2.2 Results. We recruited 10 older adults who were all aged 60-plus (5 women, Mdn_{age} = 62, Mdn_{MDPQ} = 41). Because we wanted evaluations from individuals with no prior experience or knowledge of Nav Nudge, participants who had taken part in the previous study were excluded. All participants owned a touch-enabled smartphone and were familiar with mobile map apps. All of them were Google Maps users but had no prior experience with Organic Maps. They all reported having some experience with voice assistants such as Siri or Google Assistant but did not use them frequently. Table 4 lists participant demographics, their MDPQ-14 scores, and the questions they asked Nav Nudge.

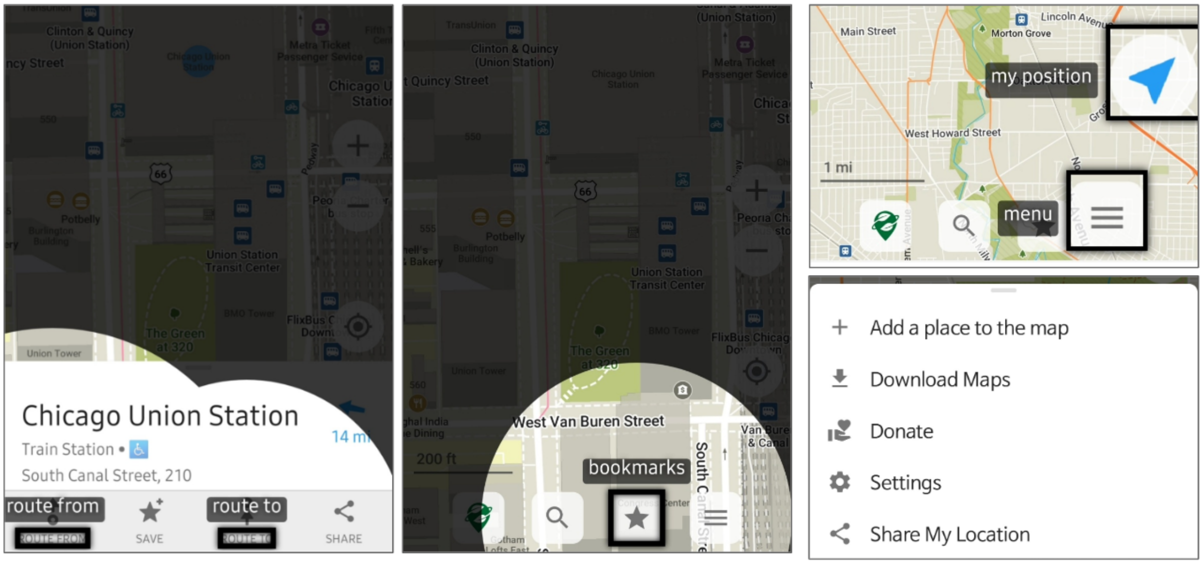
All participants successfully completed the four tasks, and eight out of ten participants used Nav Nudge at least once. In total, Nav Nudge reduced the feature space 19 times. In 18 out of 19 instances, the desired feature was included in the reduced feature space. The only instance of failure occurred when P1 hesitated during a question (*"Where can I send..."*) while performing Task 3, leading to a voice recognizer timeout (1.5 seconds). However, P1 immediately re-invoked Nav Nudge and asked the same question again. On average, participants spend 2min 35s, 1min 52s, 1min 30s, and 1min 4s to complete Tasks 1, 2, 3, and 4, respectively. Table 5 shows individual task completion times.

We further analyzed the video corpus (85 minutes) in two ways. First, we identified instances where a participant was 'getting lost' and then 'recovered' during a task. Second, we looked for common patterns among the 19 times Nav Nudge was used, while examining successes, failures, and unexpected use. Following prior work [70, 114], we labeled participant interactions as 'getting lost' when non-unique selections (choosing a feature already selected before) or off-task selections (choosing a feature unrelated to the sub-task) occurred consecutively. If participants recovered by getting back on track and then successfully completed the task at hand, we considered them 'recovered'. Participants got lost 20 times and used Nav Nudge 15 out of those 20 times. In each of those 15 times, participants recovered following the use of Nav Nudge (Figure 7).

We observed participants exploring more than one feature when multiple features were emphasized in the reduced search space. When P1 invoked Nav Nudge with the query "*send the current location*", she was presented with two features, the "my position" button and the menu icon (Figure 8c). Her target feature could be reached by selecting the menu button, but she selected the "my

Participant	Task 1	Task 2	Task 3	Task 4
P1	2m 50s	1m 42s	1m 14s	37s
P2	3m 20s	2m 15s	2m 15s	1m 10s
P3	1m 46s	2m	1m 21s	55s
P4	52s	45s	52s	57s
P5	1m 57s	1m 18s	30s	1m
P6	4m 20s	4m 14s	3m 45s	2m 10s
P7	2m 10s	1m 4s	1m 15s	55s
P8	1m 45s	1m 25s	42s	48s
P9	3m 40s	2m 50s	1m 54s	1m 12s
P10	3m 10s	1m 5s	1m 13s	1m

Table 5: Task completion times for each participant.



(a) (b) (c)

Figure 8: UI screenshots showing how older adults used Nav Nudge in a mobile map application: (a) P7: *"I want to add a starting point"*, (b) P2: *"How can I see the bookmark list?"* and (c) P1: *"Send the current location"*.

position" button first. After exploring what the button does, and finding that it adjusts her location on the map, she selected the menu button. The exact sequence of interactions also occurred with P4. This shows how older adults remembered the various features when multiple features were highlighted in the reduced feature space, and that an incorrect initial choice did not significantly thwart UI navigation. Figure 8 shows three example uses of Nav Nudge.

Our evaluation did not include a full-length interview, but we asked participants about their preference between the two types of visual cues at the end of the session, highlight with context (HC) and weighted zoom (WZ). Opinions were mixed; some preferred HC (P2: *"It was easy to see at a glance what was marked."*; P7: *"HC is a bit better, because, with WZ, other buttons are covered."*), while others preferred WZ (P1: *"The black part of HC feels like it's blocking the view."*).

6 DISCUSSION

Tech *ownership* does not always imply tech *use* in older adults. More and more older adults are purchasing mobile devices and using them frequently [45, 86], especially after the COVID-19 pandemic. However, they can find it difficult to adjust to the transition from using mobile devices for basic tasks, like calling, messaging, and web browsing, to having basically every other service online [18, 116]. On average, older adults currently use three to five mobile apps [62, 97], suggesting that even after getting a mobile device and intending to use it, they can remain in a stage of limited use in the long term. While non-use or limited use can definitely be a personal choice [49], the most common barrier to mobile tech use, per older adults, are the complexities of mobile apps and little to no tech support [2, 46]. When considering tech support for older adults, most often the focus is on designing tech training programs, like in-person or online classes [17, 105], online learning modules [66], library help desks, or 1:1 community help sessions [31, 32]). This kind of formal support, also called 'cold' support (as opposed to 'warm' or social support), is sought out by older adults mostly when starting to use new technology—but rarely when they have problems during ongoing use and retention [31, 82].

Our work is motivated by this need for ad-hoc tech support during ongoing tech use in older adults. But note that any user may require such tech support, not just older users. The need for tech support does not exclusively stem from performance deficits of aging [107] but can also be due to unfamiliarity, situated impairments, or poor design. Furthermore, different older tech users may prefer different types of tech support, some may prefer social support while others self-exploration [47, 70, 84, 97]. Older adults who try learning or troubleshooting by self-exploration often struggle to find correct features quickly during a task [63, 84], which then leads to confusion, frustration, and ultimately giving up [70, 113].

To address that, we introduced a new interaction technique that could assist older adults just in time and just in place—without requiring them to wait for a person (i.e., social support) or look up a third-party resource, like Google or YouTube [84, 97]. Note that older adults rarely use help options of mobile apps nowadays, thereby eliminating them as a possible source of tech support [84, 97]. With our proposed technique, *Nav Nudge*, users could invoke the default voice assistant of their mobile devices while using an app and ask questions about a feature they are looking for. Using our proposed feature space reduction pipeline (Figure 1), *Nav Nudge* would then identify up to three relevant features on the app's current view and emphasize them visually (e.g., Figure 8).

Figure 9: Interaction sequence (clicks) with a static image of a feature-rich UI when the feature search space was not reduced (by P3, left) vs. when the feature search space was reduced (by P10, right).

6.1 Primary Findings

We found that older adults can locate a target feature on a feature-rich mobile UI more quickly and more accurately when the feature search space is reduced on demand (Figure 6). Unlike prior work, we identified the most relevant UI features on the users’ current app view automatically, using accessibility information from the Assist API—without obtaining the information from a remote helper or scrapping the web [43, 114, 115]. While we constrained the size of the reduced feature search space to three—based on the optimal performance of our proposed algorithm (Figure 2)—it remains to be seen if older adults would retain the performance advantage when more than three features are included in the reduced search space.

Older adults were not only more efficient and accurate when finding a feature in a reduced search space but their performance was also on par with younger adults. However, note that we only tested for a large effect of age with equivalence bounds of .55. So with our data, we cannot claim the presence or absence of any smaller effect sizes. Furthermore, although we compared younger and older adults’ UI navigation performance, we do not argue that older tech users must be as efficient or as accurate as younger users—or that using tech slowly is somehow a problem that needs to be solved (i.e., a performance deficit). Neither do we claim that *all* older adults would benefit from a reduced feature search space or using Nav Nudge. For example, older adults who do not prefer self-exploration when learning or troubleshooting tech or do not prefer using voice assistants would not find Nav Nudge helpful.

Nevertheless, it is exciting that reducing the feature search space can help some older adults focus selectively and guide their locus of attention. For example, consider Figure 9, which shows how two older adults, P3 and P10, interacted with a feature-rich UI in Phase 1 of our first study (Section 4), without (left) and with visual cues (right), respectively. The task was “You are buying a ticket to a movie. Lookup your AMC membership information”. Assuming the interaction coordinates, i.e., where participants clicked on the screen before selecting the correct feature, as points of attention, we see that in the beginning, both P3 and P10 are focused on the upper-right portion of the screen, where multiple UI elements are located. As time passes, P3, working with the entire feature space, does not move their locus of attention, makes several unsuccessful attempts, and then finally shifts to the lower part of the screen where the correct feature resides. In contrast, P10, with a reduced feature space, quickly shifts attention to the lower part of the screen when the first option turns out to be incorrect. Although this study used images of mobile UI screens—which meant selecting an irrelevant feature would not take a participant away from their current screen—we observed similar usage patterns when older adults used Nav Nudge with an actual mobile app (Section 5.2).

Our primary objective here was to help older adults—who prefer self-exploration—navigate a feature-rich mobile UI, particularly in finding a target feature quickly. Quickly, because studies show otherwise older users tend to get confused or frustrated and eventually give up [70, 84, 113, 117]. To communicate a reduced feature search space, we explored different types of visual cues. Among them, older adults were most proficient with highlight with context (HC) and weighted zoom (WZ). They also preferred them the most. HC provides context but obstructs other parts of the UI by making them slightly opaque. WZ makes the UI elements larger, and in doing so hides the immediate neighborhood of the highlighted UI element—but overall, HC obstructs the UI more than WZ (Figures 3 and 4). We did not investigate the reasons behind these results, but having the option to see the spatial context of a UI element seems important to older adults. Interestingly, the quality of text cues accompanying the spatial emphasis did not matter significantly. Content labels without much contextual information about a UI element, e.g., 'edit' instead of 'edit bread' (Figure 3) did not impact older adults' performance. We did not test for spatial cues without any text cues, however.

6.2 Limitations

There are several limitations to our first user study (Section 4). First, it was conducted online which may have confounded our results. However, we recruited via the university and community mailing lists and checked the quality of data, e.g., MDPQ scores. No significant outliers were found. As expected, the average mobile device proficiency of older adults was significantly less than younger adults. Although a web application was used in our online experiment, the UI was always rendered as a mobile screen (Figures 10 and 11). Second, there were some differences in the level of familiarity with the six mobile apps across our participants (Table 2). Moreover, we did not record participants' familiarity with similar mobile apps, which could have affected their performance. However, there were no significant differences between older and younger adults' familiarity with the apps. Third, our first study had a small sample size (28 per group) and was not suited to capture small effect sizes. We only tested for a large effect of age (older vs. younger) on the efficiency and accuracy of finding a correct feature in a reduced feature space—with equivalence bounds of Cohen's $d = .55$. So we cannot claim the presence or absence of any smaller effect sizes than .55 in terms of older and younger adults' performance similarity in a reduced feature space on a mobile UI. Finally, we used screenshots of feature-rich UIs in the first study, which limited the type of immediate feedback participants could receive when interacting with the UI. However, our second user study, evaluating Nav Nudge (Section 5.2), addressed that limitation and demonstrated how older adults would use a reduced feature search space in an actual interactive mobile application.

6.3 Future Work

Post-pandemic, many more services have become available online, if not *only* online. Governments and local community organizations are amping up efforts to provide affordable, high-speed internet, computer devices, and tech support to everyone. Not all older adults seek or prefer the same types of tech support all the time [84, 97, 98]. For older users, who prefer self-exploration, assistive technologies like Nav Nudge can be beneficial. While we explored mobile UI, future work could look into how reducing the feature search space helps older adults effectively use desktop interfaces—which can pack in way more features. Nav Nudge may also be helpful for younger people who are either novices or occasional users of complex, feature-rich apps.

Furthermore, the following four aspects of Nav Nudge can be extended to improve the scope of its support to older adults. First, alternatives to voice as a mode of input can not only circumvent problems with speech recognition but also help people with speech issues or non-native accents—or those who do not prefer to use speech input. Second, our approach to communicating a reduced search space is entirely visual. For people with visual impairments, new forms of output could be explored, such as directional cues using spoken words, like upper left or bottom right, or vibrotactile cues on the surface/sides of a mobile device. Third, users need to have some idea about the feature they are looking for to successfully use Nav Nudge. What if they have no such idea? Can we use foundation models to form a set of questions from an incomplete description of what the user wants to accomplish using an app and offer possible features to use? Finally, our technique is limited to the current app view of the user and can only assist with a single step of an interaction. A future improvement would be to help with multistep interactions and muti-app interactions.

7 CONCLUSION

Traditional initiatives to help older adults learn digital skills heavily lean toward providing education, training, and support via structured courses, IT help desks, and instruction manuals. These current modes of training and education often fail to keep up with the constantly changing digital ecosystem and older adults' immediate troubleshooting needs. Some older adults prefer tech help that is more immediate, personalized, and repetitive, and sometimes to access it independently on demand. To help in self-exploration, particularly to quickly find the correct feature on a feature-rich UI, we introduced Nav Nudge. Nav Nudge is a support tool that uses accessibility APIs, large language models (LLMs), and visual cues to automatically reduce the search space of features on demand. Results from a controlled user study showed that older adults can find a feature more efficiently and more accurately when the search space is reduced to up to three features, based on a verbal query about what they are looking for. Older adults were also able to find their desired features and complete all tasks in a second (unconstrained) user study with a mobile map application.

REFERENCES

[1] AARP. 2021. AARP Foundation - for a future without senior poverty. Retrieved December 12, 2022 from https://www.aarp.org/aarp-foundation/.
[2] Link age Connect. 2021. 2021 Technology Study Older Adults Age 55-100. https://www.linkageconnect.com/wp-content/uploads/2021-Technology-Study-Report.pdf. [Accessed 24-Jul-2023].
[3] Elenia Carrasco Almao and Fatemeh Golpayegani. 2019. Are mobile apps usable and accessible for senior citizens in smart cities?. In *Human Aspects of IT for the Aged Population. Design for the Elderly and Technology Acceptance: 5th International Conference, ITAP 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings, Part I 21*. Springer, 357–375.
[4] Benett Axtell and Cosmin Munteanu. 2019. Back to real pictures: A cross-generational understanding of users' mental models of photo cloud storage. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
[5] Mortimer Mamelak Allison B. Sekuler, Patrick J. Bennett. 2000. Effects of aging on the useful field of view. *Experimental aging research* 26, 2 (2000), 103–120.
[6] Karlene K Ball, Bettina L Beard, Daniel L Roenker, Richard L Miller, and David S Griggs. 1988. Age and visual search: Expanding the useful field of view. *JOSA A* 5, 12 (1988), 2210–2219.
[7] Timothy F Brady and George A Alvarez. 2015. No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41, 3 (2015), 921.
[8] Eva Brandt, Thomas Binder, Lone Malmborg, and Tomas Sokoler. 2010. Communities of everyday practice and situated elderliness as an approach to co-design for senior interaction. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. 400–403.
[9] Nelson Kakulla. Brittne. 2020. 2020 Tech Trends of the 50+. Retrieved October, 2021 from https://doi.org/10.26419/res.00329.001.

[10] Anke M Brock, Jon E Froehlich, João Guerreiro, Benjamin Tannert, Anat Caspi, Johannes Schöning, and Steve Landau. 2018. Sig: Making maps accessible and putting accessibility in maps. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, SIG03.

[11] Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alipio Jorge, Célia Nunes, and Adam Jatowt. 2020. YAKE! Keyword extraction from single documents using multiple local features. *Information Sciences* 509 (2020), 257–289.

[12] Michelle C Carlson, Lynn Hasher, S Lisa Connelly, and Rose T Zacks. 1995. Aging, distraction, and the benefits of predictable location. *Psychology and aging* 10, 3 (1995), 427.

[13] John M Carroll and Judith Reitman Olson. 1988. Mental models in human-computer interaction. *Handbook of human-computer interaction* (1988), 45–65.

[14] Raymond B Cattell and John L Horn. 1978. A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement* 15, 3 (1978), 139–164.

[15] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. 2018. Universal sentence encoder for English. In *Proceedings of the 2018 conference on empirical methods in natural language processing: system demonstrations*. 169–174.

[16] Pei-Yu Chi, Sally Ahn, Amanda Ren, Mira Dontcheva, Wilmot Li, and Björn Hartmann. 2012. MixT: automatic generation of step-by-step mixed media tutorials. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. 93–102.

[17] Adeline Chu and Beth Mastel-Smith. 2010. The outcomes of anxiety, confidence, and self-efficacy with Internet health information retrieval in older adults: a pilot study. *CIN: Computers, Informatics, Nursing* 28, 4 (2010), 222–228.

[18] The Chicago Digital Equity Council. 2023. *The Chicago Digital Equity Plan*. https://www.chicago.gov/content/dam/city/sites/digital-equity-council/pdf/Digital-Equity-Plan-2023.pdf

[19] Nelson Cowan. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences* 24, 1 (2001), 87–114.

[20] Sara J Czaja, Walter R Boot, Neil Charness, and Wendy A Rogers. 2019. *Designing for older adults: Principles and creative human factors approaches*. CRC press.

[21] Sara J Czaja, Neil Charness, Arthur D Fisk, Christopher Hertzog, Sankaran N Nair, Wendy A Rogers, and Joseph Sharit. 2006. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychology and aging* 21, 2 (2006), 333.

[22] Biplab Deka, Zifeng Huang, Chad Franzen, Joshua Hibschman, Daniel Afergan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. 2017. Rico: A mobile app dataset for building data-driven design applications. In *Proceedings of the 30th annual ACM symposium on user interface software and technology*. 845–854.

[23] Android Developers. 2023. Optimizing Contextual Content for the Assistant. Retrieved November 10, 2023 from https://developer.android.com/training/articles/assistant.

[24] Jerri D Edwards, Lesley A Ross, Virginia G Wadley, Olivio J Clay, Michael Crowe, Daniel L Roenker, and Karlene K Ball. 2006. The useful field of view test: normative data for older adults. *Archives of Clinical Neuropsychology* 21, 4 (2006), 275–286.

[25] Hugging Face. N/A. T5 - Hugging Face. https://huggingface.co/docs/transformers/model_doc/t5.

[26] Mingming Fan, Qiwen Zhao, and Vinita Tibdewal. 2021. Older Adults' Think-Aloud Verbalizations and Speech Features for Identifying User Experience Problems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[27] Arthur D Fisk, Wendy A Rogers, Neil Charness, Sara J Czaja, and Joseph Sharit. 2004. *Designing for older adults: Principles and creative human factors approach*. CRC Press, Boca Raton, FL.

[28] Alisa Frik, Leysan Nurgalieva, Julia Bernd, Joyce Lee, Florian Schaub, and Serge Egelman. 2019. Privacy and security threat models and mitigation strategies of older adults. In *Fifteenth symposium on usable privacy and security (SOUPS 2019)*. 21–40.

[29] Michael Fröhlich, Charlotte Kobiella, Albrecht Schmidt, and Florian Alt. 2021. Is it better with onboarding? improving First-Time cryptocurrency app experiences. In *Designing Interactive Systems Conference 2021*. 78–89.

[30] Shihui Fu, Kunkun Xue, Mengya Yang, and Xiaona Wang. 2023. An exploratory study on users' resistance to mobile app updates: Using netnography and fsQCA. *Technological Forecasting and Social Change* 191 (2023), 122479.

[31] Nelly Geerts, Werner Schirmer, Anina Vercruyssen, and Ignace Glorieux. 2023. Bridging the 'instruction gap': how ICT instructors help older adults with the acquisition of digital skills. *International Journal of Lifelong Education* 42, 2 (2023), 195–207.

[32] Nelly Geerts, Werner Schirmer, Anina Vercruyssen, and Ignace Glorieux. 2023. Exploring older adults' ICT support: A mismatch between needs and provision. *New Media & Society* (2023), 14614448231166356.

[33] Helene Gelderblom, Tobie van Dyk, and Judy van Biljon. 2010. Mobile phone adoption: Do existing models adequately capture the actual usage of older adults?. In *Proceedings of the 2010 annual research conference of the south african institute of computer scientists and information technologists*. 67–74.

[34] D Kristen Gilbert and Wendy A Rogers. 1999. Age-related differences in the acquisition, utilization, and extension of a spatial mental model. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 54, 4 (1999), P246–P255.

[35] Miguel Gomez-Hernandez, Stine Willum Adrian, Xavier Ferre, and Elena Villalba-Mora. 2022. Implicit, explicit, and structural barriers and facilitators for information and communication technology access in older adults. *Frontiers in Psychology* 13 (2022), 874025.

[36] Maarten Grootendorst. 2020. KeyBERT: Minimal keyword extraction with BERT. https://doi.org/10.5281/zenodo.4461265

[37] Toshiyuki Hagiya, Toshiharu Horiuchi, and Tomonori Yazaki. 2016. Typing tutor: Individualized tutoring in text entry for older adults based on input stumble detection. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 733–744.

[38] Chuanxiuyue He, Peri Gunalp, Hauke S Meyerhoff, Zoe Rathbun, Mike Stieff, Steven L Franconeri, and Mary Hegarty. 2022. Visual working memory for connected 3D objects: effects of stimulus complexity, dimensionality and connectivity. *Cognitive Research: Principles and Implications* 7, 1 (2022), 19.

[39] John L Horn and Raymond B Cattell. 1967. Age differences in fluid and crystallized intelligence. *Acta psychologica* 26 (1967), 107–129.

[40] Hwan Hwangbo, Sol Hee Yoon, Beom Suk Jin, Young Suk Han, and Yong Gu Ji. 2013. A study of pointing performance of elderly users on smartphones. *International Journal of Human-Computer Interaction* 29, 9 (2013), 604–618.

[41] Tiffany S Jastrzembski and Neil Charness. 2007. The Model Human Processor and the older adult: Parameter estimation and validation within a mobile phone task. *Journal of Experimental Psychology: Applied* 13, 4 (2007), 224.

[42] Bora Jin, Junghwan Kim, and Lisa M Baumgartner. 2019. Informal learning of older adults in using mobile devices: A review of the literature. *Adult Education Quarterly* 69, 2 (2019), 120–141.

[43] Xiaofu Jin, Xiaozhu Hu, Xiaoying Wei, and Mingming Fan. 2022. Synapse: Interactive Guidance by Demonstration with Trial-and-Error Support for Older Adults to Use Smartphone Apps. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–24.

[44] Zhao Xia Jin, Tom Plocher, and Liana Kiff. 2007. Touch screen user interfaces for older adults: button size and spacing. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, 933–941.

[45] Brittne Kakulla. 2021. 2020 Tech Ownership and Use. Retrieved December 12, 2022 from https://doi.org/10.26419/res.00420.004.

[46] Brittne Kakulla. 2021. 2022 Tech Trends and Adults 50-Plus. Retrieved December 12, 2022 from https://doi.org/10.26419/res.00493.001.

[47] Sunyoung Kim, Abhishek Choudhury, et al. 2020. Comparison of older and younger adults' attitudes toward the adoption and use of activity trackers. *JMIR mHealth and uHealth* 8, 10 (2020), e18312.

[48] Sunyoung Kim, Willow Yao, and Xiaotong Du. 2022. Exploring older adults' adoption and use of a tablet computer during COVID-19: longitudinal qualitative study. *JMIR aging* 5, 1 (2022), e32957.

[49] Bran Knowles and Vicki L Hanson. 2018. The Wisdom of Older Technology (Non-) Users. *Commun. ACM* 61, 3 (2018), 72–77.

[50] Bran Knowles, Vicki L Hanson, Yvonne Rogers, Anne Marie Piper, Jenny Waycott, and Nigel Davies. 2019. HCI and Aging: Beyond Accessibility. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, W07.

[51] Bran Knowles, Vicki L Hanson, Yvonne Rogers, Anne Marie Piper, Jenny Waycott, Nigel Davies, Aloha Ambe, Robin N Brewer, Debaleena Chattopadhyay, Marianne Dee, et al. 2020. The Harm in Conflating Aging with Accessibility. *Communications of the Association for Information Systems* (2020).

[52] Bran Knowles, Vicki L Hanson, Yvonne Rogers, Anne Marie Piper, Jenny Waycott, Nigel Davies, Aloha Hufana Ambe, Robin N Brewer, Debaleena Chattopadhyay, Marianne Dee, et al. 2021. The harm in conflating aging with accessibility. *Commun. ACM* 64, 7 (2021), 66–71.

[53] Masatomo Kobayashi, Atsushi Hiyama, Takahiro Miura, Chieko Asakawa, Michitaka Hirose, and Tohru Ifukube. 2011. Elderly user evaluation of mobile touchscreen interactions. In *IFIP Conference on Human-Computer Interaction*. Springer, 83–99.

[54] Junhan Kong, Anhong Guo, and Jeffrey P Bigham. 2019. Supporting Older Adults in Using Complex User Interfaces with Augmented Reality. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 661–663.

[55] Mandy Korzetz, Romina Kühn, Lukas Büschel, Franz-Wilhelm Schumann, Uwe Aßmann, and Thomas Schlegel. 2020. Introducing Mobile Device-Based Interactions to Users: An Investigation of Onboarding Tutorials. In *International Conference on Human-Computer Interaction*. Springer, 428–442.

[56] Daniël Lakens. 2017. Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social psychological and personality science* 8, 4 (2017), 355–362.

[57] Daniël Lakens, Anne M Scheel, and Peder M Isager. 2018. Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science* 1, 2 (2018), 259–269.
[58] J Richard Landis and Gary G Koch. 1977. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* (1977), 363–374.
[59] Roxanne Leitão and Paula Alexandra Silva. 2012. Target and spacing sizes for smartphone user interfaces for older adults: design patterns based on an evaluation with users. In *Proceedings of the 19th Conference on Pattern Languages of Programs*. The Hillside Group, 5.
[60] Rock Leung, Leah Findlater, Joanna McGrenere, Peter Graf, and Justine Yang. 2010. Multi-layered interfaces to improve older adults' initial learnability of mobile applications. *ACM Transactions on Accessible Computing (TACCESS)* 3, 1 (2010), 1.
[61] Qingchuan Li and Yan Luximon. 2016. Older adults and digital technology: A study of user perception and usage behavior. In *Advances in physical ergonomics and human factors*. Springer, 155–163.
[62] Qingchuan Li and Yan Luximon. 2018. Understanding older adults' post-adoption usage behavior and perceptions of mobile technology. *International Journal of Design* 12, 3 (2018).
[63] Qingchuan Li and Yan Luximon. 2020. Older adults' use of mobile device: usability challenges while navigating various interfaces. *Behaviour & Information Technology* 39, 8 (2020), 837–861.
[64] Yang Li, Jiacong He, Xin Zhou, Yuan Zhang, and Jason Baldridge. 2020. Mapping natural language instructions to mobile ui action sequences. *arXiv preprint arXiv:2005.03776* (2020).
[65] Yang Li, Gang Li, Luheng He, Jingjie Zheng, Hong Li, and Zhiwei Guan. 2020. Widget captioning: Generating natural language description for mobile user interface elements. *arXiv preprint arXiv:2010.04295* (2020).
[66] Chicago Public Library. 2023. *Chicago DigitalLearn - Use a computer to do almost anything!* https://chipublib.digitallearn.org/
[67] Steven J Luck and Edward K Vogel. 1997. The capacity of visual working memory for features and conjunctions. *Nature* 390, 6657 (1997), 279–281.
[68] Steven J Luck and Edward K Vogel. 2013. Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in cognitive sciences* 17, 8 (2013), 391–400.
[69] Diana MacDonald. 2019. *Practical ui patterns for design systems: Fast-track interaction design for a seamless user experience*. Apress.
[70] Shareen Mahmud, Jessalyn Alvina, Parmit K Chilana, Andrea Bunt, and Joanna McGrenere. 2020. Learning through exploration: how children, adults, and older adults interact with a new feature-rich application. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
[71] Rui Mata, Lael J Schooler, and Jörg Rieskamp. 2007. The aging decision maker: cognitive aging and the adaptive selection of decision strategies. *Psychology and aging* 22, 4 (2007), 796.
[72] Roberto Menghi, Silvia Ceccacci, Francesca Gullà, Lorenzo Cavalieri, Michele Germani, and Roberta Bevilacqua. 2017. How Older People Who Have Never Used Touchscreen Technology Interact with a Tablet. In *IFIP Conference on Human-Computer Interaction*. Springer, 117–131.
[73] Helena M Mentis, Galina Madjaroff, and Aaron K Massey. 2019. Upside and downside risk in online security for older adults with mild cognitive impairment. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[74] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*. 404–411.
[75] George A Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review* 63, 2 (1956), 81.
[76] Karen J Mitchell, Marcia K Johnson, Carol L Raye, Mara Mather, and Mark D'Esposito. 2000. Aging and reflective processes of working memory: binding and test load deficits. *Psychology and aging* 15, 3 (2000), 527.
[77] Stephanie A Morey, Rachel E Stuck, Amy W Chong, Laura H Barg-Walkow, Tracy L Mitzner, and Wendy A Rogers. 2019. Mobile health apps: improving usability for older adult users. *Ergonomics in Design* 27, 4 (2019), 4–13.
[78] Barbara Barbosa Neves, Rachel L Franz, Cosmin Munteanu, Ronald Baecker, and Mags Ngo. 2015. My Hand Doesn't Listen to Me!: Adoption and Evaluation of a Communication Technology for the'Oldest Old'. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1593–1602.
[79] Hugo Nicolau and Joaquim Jorge. 2012. Elderly text-entry performance on touchscreens. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 127–134.
[80] Paulo Noriega, Diana Carvalho, Igor Correia, Jose Luís Alves, Tiago Oliveira, and Francisco Rebelo. 2020. Effectiveness of Coach Marks or Instructional Overlay in Smartphone Apps Interfaces. In *Advances in Ergonomics in Design: Proceedings of the AHFE 2019 International Conference on Ergonomics in Design, July 24-28, 2019, Washington DC, USA 10*. Springer, 68–78.
[81] Katherine E Olson, Arthur D Fisk, and Wendy A Rogers. 2009. Collaborative automated systems: Older adults' mental model acquisition and trust in automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 53. SAGE Publications Sage CA: Los Angeles, CA, 1704–1708.
[82] Tobias Olsson and Dino Viscovi. 2018. Warm experts for elderly users: Who are they and what do they do? *Human Technology* 14, 3 (2018).
[83] OpenAI. 2023. GPT-3.5. Retrieved Dec 12, 2023 from https://platform.openai.com/docs/models/gpt-3-5.
[84] Carolyn Pang, Zhiqin Collin Wang, Joanna McGrenere, Rock Leung, Jiamin Dai, and Karyn Moffatt. 2021. Technology Adoption and Learning Preferences for Older Adults: Evolving Perceptions, Ongoing Challenges, and Emerging Design Opportunities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.
[85] Andraž Petrovčič, Walter R Boot, Tomaž Burnik, and Vesna Vesna Dolničar. 2019. Improving the Measurement of Older Adults' Mobile Device Proficiency: Results and Implications from a Study of Older Adult Smartphone Users. *IEEE Access* 7 (2019), 150412–150422.
[86] Research Center Pew. 2023. Mobile Fact Sheet. Retrieved January, 2024 from https://www.pewresearch.org/internet/fact-sheet/mobile/.
[87] Anne Marie Piper, Raymundo Cornejo Garcia, and Robin N Brewer. 2016. Understanding the challenges and opportunities of smart mobile devices among the oldest old. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 8, 2 (2016), 83–98.
[88] Alexandra A Portenhauser, Yannik Terhorst, Dana Schultchen, Lasse B Sander, Michael D Denkinger, Michael Stach, Natalie Waldherr, Dhayana Dallmeier, Harald Baumeister, and Eva-Maria Messner. 2021. Mobile apps for older adults: systematic search and evaluation within online stores. *JMIR aging* 4, 1 (2021), e23313.
[89] Kelly Quinn, Renae Smith-Ray, and Kristin Boulter. 2016. Concepts, terms, and mental models: Everyday challenges to older adult social media adoption. In *Human Aspects of IT for the Aged Population. Healthy and Active Aging: Second International Conference, ITAP 2016, Held as Part of HCI International 2016 Toronto, ON, Canada, July 17–22, 2016, Proceedings, Part II 2*. Springer, 227–238.
[90] Jennifer C Romano Bergstrom, Erica L Olmsted-Hawala, and Matt E Jans. 2013. Age-related differences in eye tracking and usability performance: website usability for older adults. *International Journal of Human-Computer Interaction* 29, 8 (2013), 541–548.
[91] Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic keyword extraction from individual documents. *Text mining: applications and theory* 1 (2010), 1–20.
[92] Hasanin Mohammed Salman, Wan Fatimah Wan Ahmad, and Suziah Sulaiman. 2018. Usability evaluation of the smartphone user interface in supporting elderly users from experts' perspective. *Ieee Access* 6 (2018), 22578–22591.
[93] Timothy A Salthouse. 1996. The processing-speed theory of adult age differences in cognition. *Psychological review* 103, 3 (1996), 403.
[94] Timothy A Salthouse and Renee L Babcock. 1991. Decomposing adult age differences in working memory. *Developmental psychology* 27, 5 (1991), 763.
[95] Neil Selwyn, Stephen Gorard, John Furlong, and Louise Madden. 2003. Older adults' use of information and communications technology in everyday life. *Ageing & Society* 23, 5 (2003), 561–582.
[96] Hasti Sharifi and Debaleena Chattopadhyay. 2023. A Cross-Cultural Study of Relational Maintenance in Tech Caregiving. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–10.
[97] Hasti Sharifi and Debaleena Chattopadhyay. 2023. Senior Technology Learning Preferences Model for Mobile Technology. *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction* 7, Article 222 (2023), 39 pages. https://doi.org/10.1145/3447526.3472040
[98] HASTI SHARIFI and DEBALEENA CHATTOPADHYAY. 2023. Senior Technology Learning Preferences Model for Mobile Technology. (2023).
[99] Uri Simonsohn. 2015. Small telescopes: Detectability and the evaluation of replication results. *Psychological science* 26, 5 (2015), 559–569.
[100] Aaron Smith. 2014. Older Adults and Technology Use. *Pew Research Center* (2014).
[101] Amanda L Smith and Barbara S Chaparro. 2015. Smartphone text input method performance, usability, and preference with younger and older adults. *Human factors* 57, 6 (2015), 1015–1028.
[102] Christian Stößel. 2009. Familiarity as a factor in designing finger gestures for elderly users. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–2.
[103] Brendan Strahm, Colin M Gray, and Mihaela Vorvoreanu. 2018. Generating mobile application onboarding insights through minimalist instruction. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 361–372.
[104] Meirav Taieb-Maimon and Elior Vaisman-Fairstein. 2022. Improving older adults' accessibility to the web using real-time online interactive guides. *International Journal of Human-Computer Studies* 164 (2022), 102830.
[105] Hsin-Yi Sandy Tsai, RV Rikard, Shelia R Cotten, and Ruth Shillair. 2019. Senior technology exploration, learning, and acceptance (STELA) model: from exploration to use–a longitudinal randomized controlled trial. *Educational*

Gerontology 45, 12 (2019), 728–743.

[106] Christopher Vendome, Diana Solano, Santiago Liñán, and Mario Linares-Vásquez. 2019. Can everyone use my app? an empirical study on accessibility in android apps. In *2019 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 41–52.

[107] John Vines, Gary Pritchard, Peter Wright, Patrick Olivier, and Katie Brittain. 2015. An age-old problem: Examining the discourses of ageing in HCI and strategies for future research. *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 1 (2015), 2.

[108] Edward K Vogel, Geoffrey F Woodman, and Steven J Luck. 2001. Storage of features, conjunctions, and objects in visual working memory. *Journal of experimental psychology: human perception and performance* 27, 1 (2001), 92.

[109] Maegen E Walker, Jonas F Vibell, Andrew D Dewald, and Scott Sinnett. 2022. Ageing and selective inhibition of irrelevant information in an attention-demanding rapid serial visual presentation task. *Brain and Neuroscience Advances* 6 (2022), 23982128211073427.

[110] Jenny Waycott, Frank Vetere, Sonja Pedell, Amee Morgans, Elizabeth Ozanne, and Lars Kulik. 2016. Not for me: Older adults choosing not to participate in a social isolation intervention. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 745–757.

[111] Chui Wong, Rahimah Ibrahim, Tengku Hamid, and Evi Mansor. 2020. Measuring expectation for an affordance gap on a smartphone user interface and its usage among older adults. *Human Technology* 16, 1 (2020), 6–34.

[112] Cheng-Chia Yang, Shang-Yu Yang, and Yu-Chia Chang. 2023. Predicting Older Adults' Mobile Payment Adoption: An Extended TAM Model. *International journal of environmental research and public health* 20, 2 (2023), 1391.

[113] Ja Eun Yu and Debaleena Chattopadhyay. 2020. "Maps are hard for me": How Older Adults Struggle with Mobile Maps. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. in press.

[114] Ja Eun Yu, Natalie Parde, and Debaleena Chattopadhyay. 2023. "Where is history": Toward Designing a Voice Assistant to help Older Adults locate Interface Features quickly. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[115] Mingyuan Zhong, Gang Li, Peggy Chi, and Yang Li. 2021. HelpViz: Automatic Generation of Contextual Visual Mobile Tutorials from Text-Based Instructions. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 1144–1153.

[116] Jia Zhou, Pei-Luen Patrick Rau, and Gavriel Salvendy. 2012. Use and design of handheld computers for older adults: A review and appraisal. *International Journal of Human-Computer Interaction* 28, 12 (2012), 799–826.

[117] Ziyao Zhou, Jia Zhou, and Fengli Liu. 2022. Fewer steps the better? Instructing older adults' learning and searching in smartphone apps. *International Journal of Human–Computer Interaction* 38, 9 (2022), 789–800.

A RESEARCH INSTRUMENTS

A.1 Mobile Device Proficiency Questionnaire-14 (MDPQ-14)

Hello there, thank you for your participation!

The questions below ask about your ability to perform a number of tasks with a mobile device.
Please answer each question by selecting the option that you feel is most appropriate.
If you have never tried to perform a task with a mobile device or do not know what a task is, please choose "Never tried," regardless of whether or not you think you may be able to perform the task.

Please remember that you are rating your ability to perform each of these tasks specifically using a mobile device (tablet or smartphone).

1. Using a mobile device I can adjust the screen brightness.						
	1	2	3	4	5	
Never tried	O	O	O	O	O	Very easily
2. Using a mobile device I can connect to a WiFi network.						
Never tried	O	O	O	O	O	Very easily
3. Using a mobile device I can open Email.						
Never tried	O	O	O	O	O	Very easily
4. Using a mobile device I can view pictures sent by Email.						
Never tried	O	O	O	O	O	Very easily
5. I can transfer information (files such as music, pictures, documents) on my computer to my mobile device.						
Never tried	O	O	O	O	O	Very easily
6. Using a mobile device I can store information with a service that lets me view my files from anywhere (e.g., Dropbox, Google Drive, Microsoft Onedrive).						
Never tried	O	O	O	O	O	Very easily
7. Using a mobile device I can find information about local community resources on the Internet.						
Never tried	O	O	O	O	O	Very easily
8. Using a mobile device I can find information about my hobbies and interests on the Internet.						
Never tried	O	O	O	O	O	Very easily
9. Using a mobile device I can enter events and appointments into a calendar.						
Never tried	O	O	O	O	O	Very easily
10. Using a mobile device I can set up alerts to remind me of events and appointments.						
Never tried	O	O	O	O	O	Very easily
11. Using a mobile device I can watch movies and videos.						
Never tried	O	O	O	O	O	Very easily
12. Using a mobile device I can read a book.						
Never tried	O	O	O	O	O	Very easily
13. I can update games and other applications on a mobile device.						
Never tried	O	O	O	O	O	Very easily
14. I can upgrade device software on a mobile device.						
Never tried	O	O	O	O	O	Very easily

Table 6: MDPQ-14 was used to assess participants' mobile device proficiency in our two user studies.

A.2 Web Application rendered on a Desktop

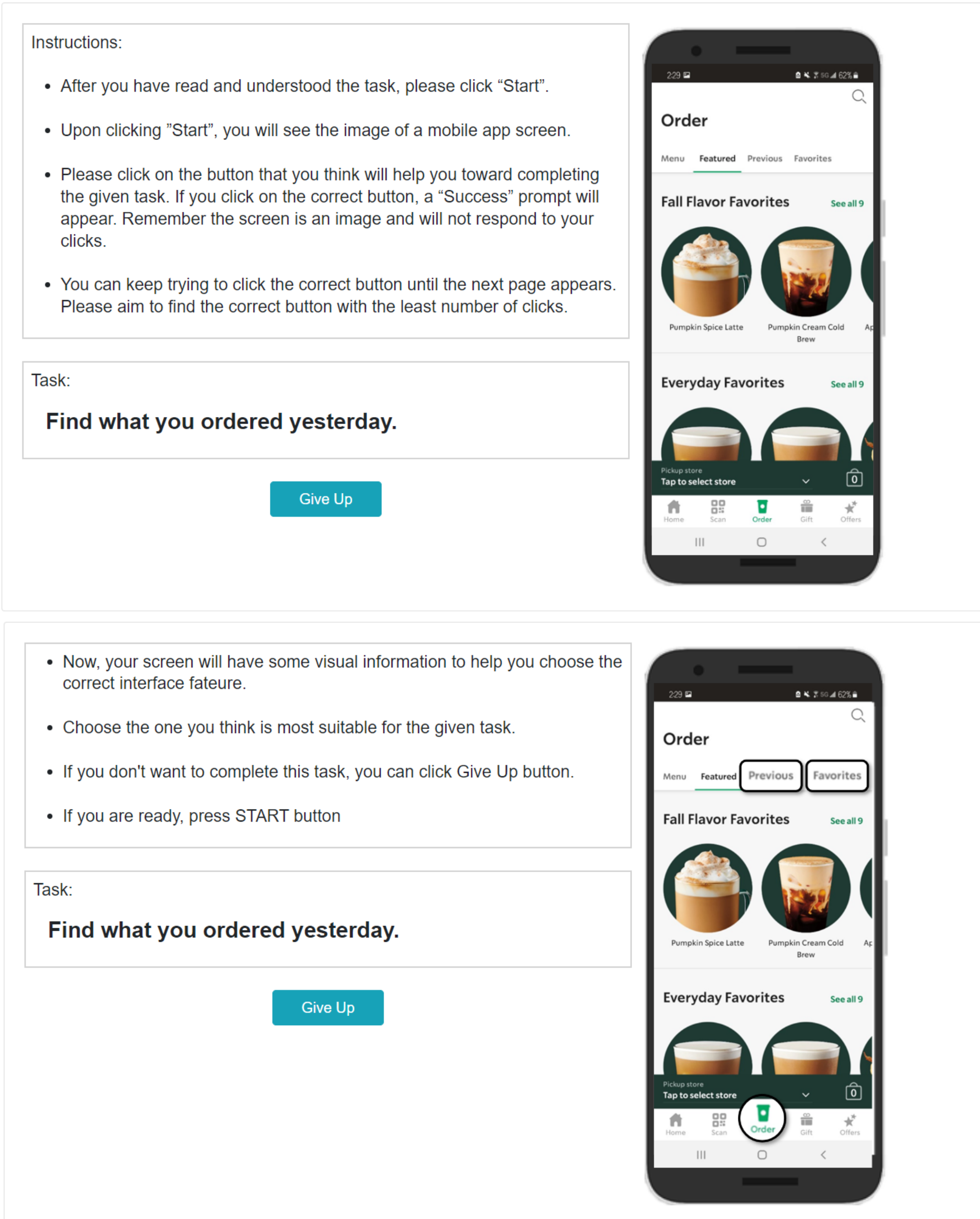


Figure 10: Screenshots of the web application on a desktop. Participants are required to read the task on the left and select the UI element on the mobile interface on the right. If they fail to locate the correct UI element within the given time on the screen with no visual cues (top), a screen with visual cues (bottom) is provided. In this task, the correct answer is the 'previous' button.

A.3 Web Application rendered on a Mobile Device

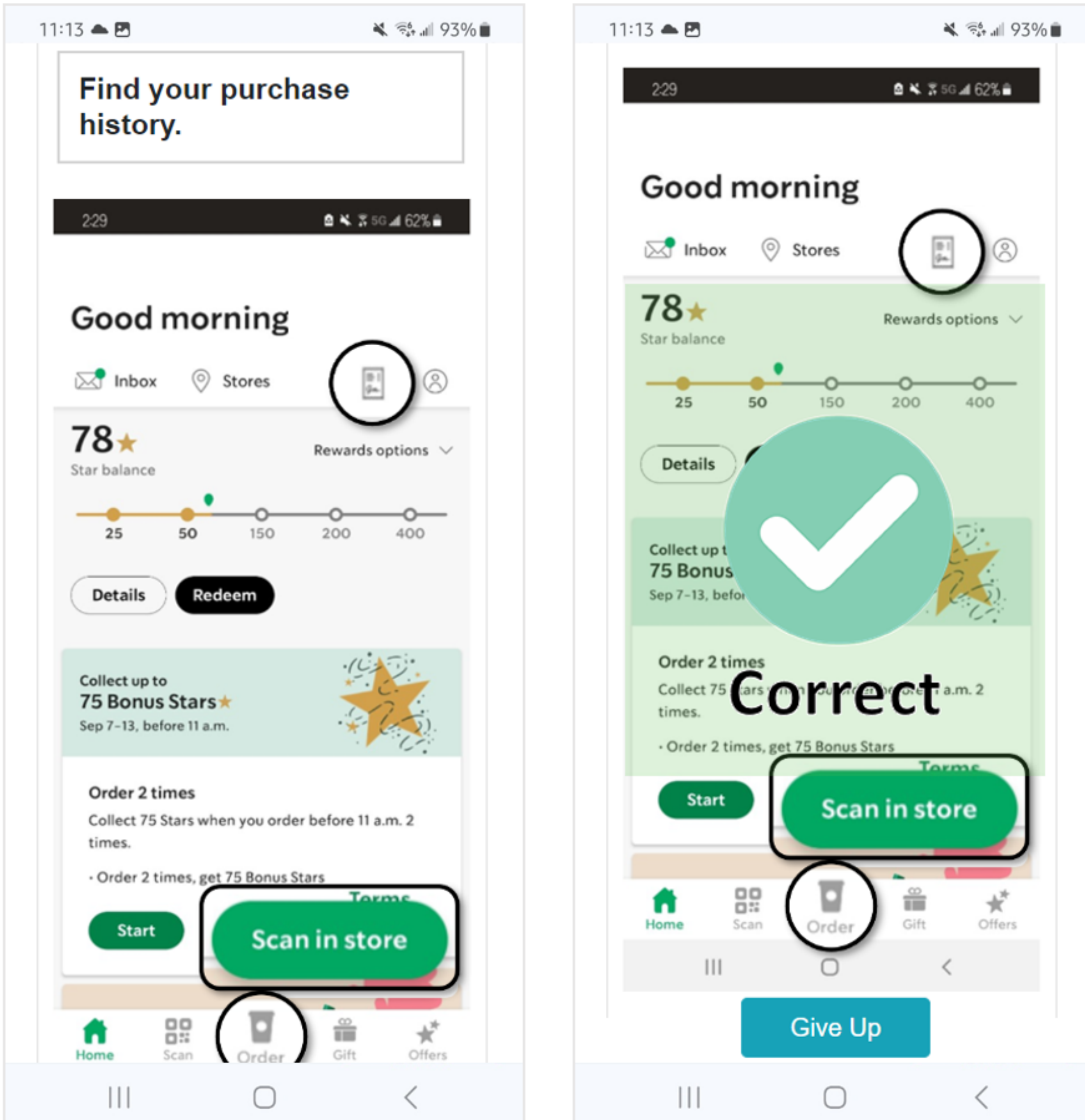


Figure 11: Screenshots of the web application on a mobile device. The web application is designed as a responsive web, allowing the task and mobile interface to be visible on a single screen (left). When participants click on a correct UI element, a brief 'correct' prompt appears. Continuing to scroll down reveals the "Give up" button (right).

A.4 Tasks used in our first user study

Mobile app	*Tasks*
Starbucks	Find your purchase history; Find what you ordered yesterday; Find a Starbucks store near you;You are ordering a Chai Tea Latte. Request cane sugar as an add on.
Uniqlo	Scan the barcode of a dress you are checking out in the store to get more information about the product; The item you want to buy is out of stock. Request to be notified when it becomes available again; Order the products displayed according to their highest customer rating; Quickly view all the products that you looked at yesterday.
AMC	Find AMC theaters that you have bookmarked before; You are buying a ticket to a movie. Lookup your AMC membership information; You are pre-ordering snacks. Change the pickup location; Find out what movies are playing today at the Woodfield Theatre.
Ventra	Find what train station is closest to you; You are planning a trip. Change the time you will depart to customize your search results; Find the train station that you had bookmarked before; You just buy a new Ventra card. Add the new card to the app.
Subway	Find a Subway store near you; The sandwich in your cart contains tomatoes. Remove tomatoes from the sandwich; Your current order is set for pickup. Change your order to delivery; Change your choice of bread from untoasted to toasted.
Audible	Quickly find the podcast content by its name; You want to see other episodes of this channel. Find out where you can find them; You would like to read more detailed descriptions of this channel. Find out where you can read it; You have bookmarked your favorite episodes. Find the list of those episodes.

Table 7: For Study 1, we created a total of 24 tasks across six Android apps.

B DATASETS AND MODELS

B.1 Query

We created a new dataset using two methods. First, we collected English sentences from multiple sources, such as YouTube videos and the Help pages of the mobile app, and then converted them into interrogatives. We scraped 30K English sentences from YouTube how-to videos and 35 mobile apps' help pages [categories: map (13), shopping (12), social media (10)]. Then we used the T5 model from Huggingface's transformers library [25] to generate paraphrased queries. The model was trained using the Quora Question Pairs dataset which contains lists of questions identified as the same question. The basis for using it is that duplicate questions can be considered paraphrased pairs. Five paraphrases were created for each query, and paraphrases that were not relevant were manually removed. This process resulted in obtaining 140 queries. To complement the age-agnostic dataset, we wanted to collect queries from older adults particularly. We used content sourced from prior research in this domain [113, 114], where conducted user studies with older adults and collected verbal queries from them. From the sources, we could get 100 useful queries. We then manually tagged the keywords of each sentence and created a set of 240 query-keyword pairs.

B.2 UILabel

After selecting three mobile apps from different categories, we extracted content labels across multiple pages within the apps. We removed duplicates and eliminated labels containing only numbers or special characters, as well as abbreviated labels (e.g., *"hh"*, *"yyyy"*). Following that, we manually selected content labels containing multi-word phrases to align with the purpose of this dataset. This resulted in the creation of a dataset consisting of a total of 320 UI elements' content labels.

B.3 Modified Widget Caption Dataset

Ground truth	*Human Annotations*
City	enter address, search, location
Review	comment text box, comment/review, write a comment
Help	question, help, help
Alarmclock	play, alarm clock, start timer
Appearance	edit list colors, open categories, paint

Table 8: Examples of ground truth and human annotations in the modified widget caption dataset. While omitted from the table, each row includes a corresponding list of other UI elements present in the UI view.

The original Widget Caption dataset [65] is created based on the extended RICO dataset [22]. RICO is a dataset of Android User Interfaces, comprising app screens extracted from Android apps, along with their associated UI view hierarchy. By combining these two datasets, we could obtain a dataset consisting of *the names of specific UI elements*, *captions provided by human annotators for those UI elements*, and a *list of other UI elements present in the UI view where the element exists*. We filtered out labels consisting only of numbers or special characters and removed non-English labels using the Lingua library. Additionally, we manually removed abbreviated or obscure labels and labels with typos.