Modeling the Intuitiveness of Mobile GUI Navigation by Understanding and Simulating Users' Attention Distribution

Xiaozhu Hu
Hong Kong University of Science and Technology
Division of Integrative Systems and Design
Hong Kong SAR, China
xhubk@connect.ust.hk

Xiaoyu Mo
Hong Kong University of Science and Technology
Hong Kong SAR, China
xmoac@connect.ust.hk

Xiaofu Jin
Hong Kong University of Science and Technology
Hong Kong SAR, China
xjinao@connect.ust.hk

Yongquan 'Owen' Hu
National University of Singapore
Augmented Human Lab
Singapore, Singapore
yongquan@ahlab.org

Mingming Fan
Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
Hong Kong University of Science and Technology
Hong Kong SAR, China
mingmingfan@ust.hk

Tristan Braud
Hong Kong University of Science and Technology
Division of Integrative Systems and Design
Hong Kong SAR, China
braudt@ust.hk

Abstract

Analyzing users' attention distribution is an effective way to evaluate whether the primary UI content captures users' engagement. This has led to the development of eye trackers that gather data on users' attention, as well as computational models that simulate this distribution on individual UI screens. However, there is limited research on how users' attention distribution correlates with the intuitiveness of goal-oriented UI navigation. Additionally, simulating user attention as they reason and identify the correct UI element to navigate to the target screen remains unexplored. To address this gap, we introduce an AI-driven model, the UI Link Transformer (UILT), which predicts users' attention distribution as they navigate from the current UI screen to the target UI screen. This model helps designers to evaluate the intuitiveness of UI navigation. Our initial study aimed to understand how users typically identify the UI element that links two consecutive UI screens and how this identification relates to the intuitiveness of the UI navigation design. The insights gained from this study offer designers actionable recommendations to improve the intuitiveness of mobile GUI navigation, with a focus on users' attention distribution. Moreover, the dataset collected during this study supports the development of the UILT. Building on the insights and data from the initial study, we designed, trained, and evaluated the UILT.

CCS Concepts

• **Human-centered computing → User models**; **Ubiquitous and mobile computing design and evaluation methods**.

Keywords

Intuitive GUI navigation; User attention simulation; Eye-tracking

ACM Reference Format:
Xiaozhu Hu, Xiaoyu Mo, Xiaofu Jin, Yongquan 'Owen' Hu, Mingming Fan, and Tristan Braud. 2025. Modeling the Intuitiveness of Mobile GUI Navigation by Understanding and Simulating Users' Attention Distribution. In *Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25), April 28-May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3701716.3717552

1 Introduction

Navigating through multiple graphical user interface (GUI) screens is crucial for enhancing user experience, engaging users effectively, and often determines the broader success of smartphone applications [41]. Analyzing visual attention patterns from users is an effective method to assess whether the UI design successfully guides users' mental focus [21, 43], which in turn reflects the success of the UI navigation design. Consequently, eye-tracking devices are commonly used in usability testing. Moreover, to speed up the design iteration process and reduce costs, recent studies have employed artificial intelligence (AI) models to simulate users' attention distribution [14, 25, 37, 39, 46]. These advancements enable designers to quickly gain design insights and conserve user study resources.

While previous research focuses on users' attention distribution within individual UI screens to evaluate key content saliency, its ability to address goal-oriented navigation remains limited. However, user navigation actions are sometimes not driven by the current UI content but by the objective to utilize specific functions. The potential of attention distribution analysis to assess whether the UI design supports such goal-oriented navigation remains unexplored.

WWW Companion '25, Sydney, NSW, Australia

ACM ISBN 979-8-4007-1331-6/2025/04
https://doi.org/10.1145/3701716.3717552

Additionally, the development of predictive models to simulate users' attention distribution during goal-oriented navigation is also a valuable yet unexplored area.

Our research investigates users' attention distribution within goal-oriented navigation through two key phases. (1) Formative Study: We aim to understand how users' selection of the UI element (referred to as the *link UI*) for navigating between the current and target screens relates to their attention patterns. We conduct a user experiment during which participants are presented with novel app screen pairs and tasked with identifying the link UI between the screens. We analyze their selections and gaze data to reveal the relationship between the attention distribution patterns and the intuitiveness of UI navigation. This phase also allows us to collect data used to train the model developed during the second phase of our research. (2) AI Model Development: We introduce UI Link Transformer (UILT), an AI model trained and evaluated on the dataset collected from the formative user study to simulate users' attention distribution when seeking the correct link UI.

Our findings indicate that intuitive UI navigation directs users' attention distribution progressively towards the correct link UI, while unintuitive UI navigation results in a dispersed attention distribution. To simulate users' attention distribution when seeking the link UI in goal-oriented UI navigation, supervising the UILT with users' actual attention data, without pre-training the model by predicting the correct link UI, achieves slightly better performance.

This work makes two key contributions: (1) **Formative Study:** Provides insights into user attention patterns and cognitive processes when selecting the *link UI* for navigation. (2) **UI Link Transformer (UILT):** Proposes an AI model to simulate users' attention distribution, advancing our understanding of interaction dynamics in UI navigation design.

2 Related Works

We review prior research from two perspectives: UI navigation design and its evaluation using eye-tracking and saliency analysis.

2.1 Design of GUI Navigation

Graphical User Interfaces (GUIs) introduced link elements to navigate across disparate interfaces, facilitating transitions between pages or sections [41]. Research on GUI navigation design can be categorized into macro and micro levels. Early studies on the World Wide Web focused on macro-level aspects, such as design guidelines [13, 22, 32], conceptual models [10, 44], and interactive techniques [1, 17], aimed at structuring web information hierarchically to improve navigation efficiency. In contrast, micro-level research emphasizes the intuitiveness of navigation relationships between consecutive UI screens, often leveraging semantic-based design concepts like interface metaphors [2, 24]. Design guidelines for crafting intuitive link elements are particularly prevalent in web [4, 13] and mobile applications [35].

2.2 GUI Navigation Evaluation with Eye Tracking

Advances in eye-tracking technology [34, 38, 40, 49] have made it a common tool for usability evaluation across platforms, including web-based GUIs [8, 12, 33], computer games [18], smartphone interfaces [6], and smartwatches [45]. Based on the eye-mind theory [21, 43], visual attention patterns are considered proxies for mental focus and cognitive strategies. Central to this analytical framework are two metrics: the number of fixations and scanpath [8, 45, 50]. The number of fixations indicates the user's attention distribution, commonly used to assess the visual saliency of different UI elements on a GUI [3, 6, 19, 26]. The scanpath reveals how users search for specific information, aiding designers in evaluating the layout of a UI [3]. An auxiliary metric, spatial entropy, derived from the fixation distribution, provides insights into the convergence of user attention [15]. In this work, we explore how the user's performance of identifying the navigation relationship between two UI screens can relate to their gaze patterns, providing more details for evaluating the intuitiveness of GUI navigation semantics.

While eye-tracking experiments provide valuable insights, recruiting participants for these studies is time-consuming. Consequently, there is growing interest in simulating user gaze patterns to streamline GUI evaluations. Existing research primarily focuses on developing gaze datasets across diverse UI types and formulating algorithms to predict the saliency maps of individual UI screens [7, 19, 26]. These predictions aim to replicate users' perception patterns when exposed to various visual elements on individual UI screens, primarily to assess whether the saliency of UI elements aligns with their importance. Our study differs by simulating user attention distribution during goal-oriented UI navigation, offering a novel perspective on evaluating UI navigation intuitiveness.

3 Formative Study: Understanding Users' Attention Distribution Patterns When Identifying the Link UI

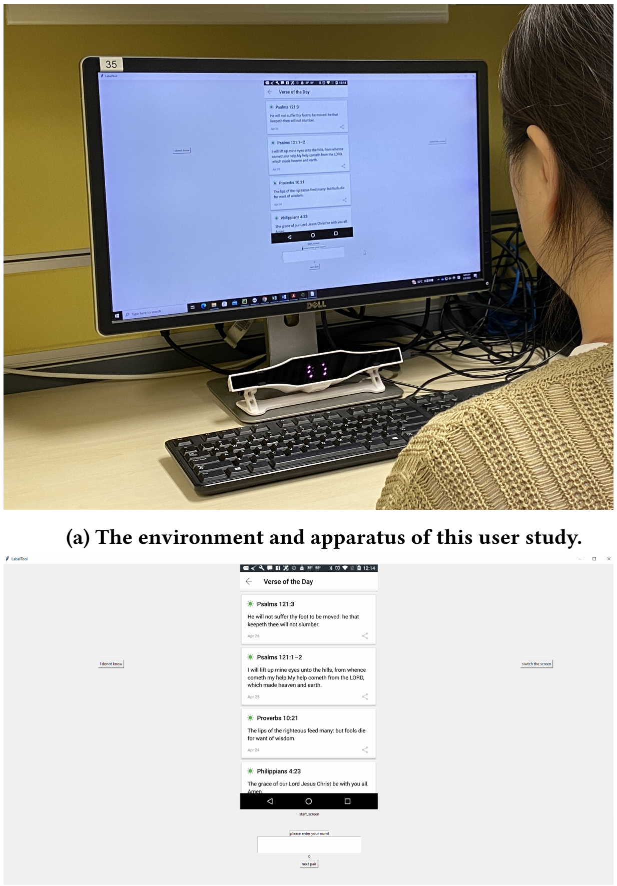
This study addresses two primary objectives through two sessions. The first session (*user research session*) explores the users' attention distribution across UI screens pairs when identifying the link UI between them. The second session (*dataset collection*) aims to collect users' gaze data as the dataset we use in Section 4 to train the AI model for predicting users' attention patterns.

3.1 Participants and Apparatus

20 participants (9 females, 11 males; aged from 23 to 31, M = 26.50, SD = 2.21) were recruited through online social platforms. All participants were college students with over five years of smartphone usage experience. Each participant received an HK$75 bonus for their participation, and their gaze data was collected with informed consent.

The study was conducted in a quiet lab room. As shown in Figure 1a, each participant sat in front of a computer and used a software tool to label the link UI between two UI screens. The labeling interface was displayed on a Dell P2314H monitor (23 inches, 1920x1080 resolution). A Gazepoint GP3sd v2 eye tracker was positioned at the bottom of the monitor to collect the user's gaze data.

The labeling tool, shown in Figure 1b, was developed with Python Tkinter. This tool offers several functions: a "switch" button allows users to toggle between viewing the current screen that needs labeling and the target screen, which provides hints about the goal.



Verse of the Day

Psalms 121:3

He will not suffer thy foot to be moved: he that keepeth thee will not slumber.

Apr 26

Psalms 121:1–2

I will lift up mine eyes unto the hills, from whence cometh my help.My help cometh from the LORD, which made heaven and earth.

Apr 25

Proverbs 10:21

The lips of the righteous feed many: but fools die for want of wisdom.

Apr 24

Philippians 4:23

The grace of our Lord Jesus Christ be with you all. Amen.

(a) The environment and apparatus of this user study.

(b) The user interface of labeling tool.

Figure 1: The environment, apparatus, and labeling tool interface of this user study

Participants label the UI component they believe links both screens on the current screen by left-clicking with the mouse. The software includes an "uncertain" button for cases where participants cannot clearly identify a link UI component between the screens. After labeling a UI component, participants are prompted to enter a justification for their choice.

The UI screens were sampled from the mobile UI animation sub-dataset of the Rico dataset [9], which contains 10811 user interaction traces and 72219 unique UIs from 9772 Android apps. 40 pairs of screens were used in the user research session and 2000 pairs in the dataset collection session, covering 5 common types of navigation patterns (board, tab, list, gallery, and drawer [31]). As our focus was on how users identify the link UI between two UI screens, we only sampled single tap gestures. We identified the ground truth of the link UI according to the coordinates of the tap gestures in the dataset.

3.2 Procedure

Upon entering the experiment room, participants were first given a brief overview of the experiment. They then signed a consent form for the collection of gaze data.

Next, we explained how to use the labeling tool to each participant. They were allowed to practice several times until they became proficient in using it. Assistance was provided whenever needed.

The study began with the user research session, where participants were instructed to label the link UI element across a standard set of 40 UI pairs, consistent for all participants. After a 5-minute break, participants commenced the dataset collection session, during which they needed to label 100 UI pairs that were unique to each participant.

3.3 Data Collection and Analysis

3.3.1 Data Collection. We collected each participant's labeled link UI, with the Rico dataset serving as the ground truth for these link UIs (refer to Section 3.1). Additionally, we recorded participants' gaze data as they viewed the current UI screen to identify the link UI necessary for navigating to the target screen. From the user research session, we gathered 800 data points (40 pairs * 20 participants). From the dataset collection session, we collected 2,000 data points (100 pairs * 20 participants). The data points from the first session were used to identify the relationship between the UI navigation intuitiveness and the attention distribution patterns. The data collected from the second session were used to create the dataset for the AI model (refer to Section 4).

3.3.2 Data Analysis. We introduce the **intuitiveness score** to represent the intuitiveness of the navigation relationship between two consecutive UI screens. This metric corresponds to the ratio of correct labels (aligned with the ground truth) to the total number of users. A higher intuitiveness score implies a more discernible navigation relationship between the two screens, enabling users to make accurate choices more readily.

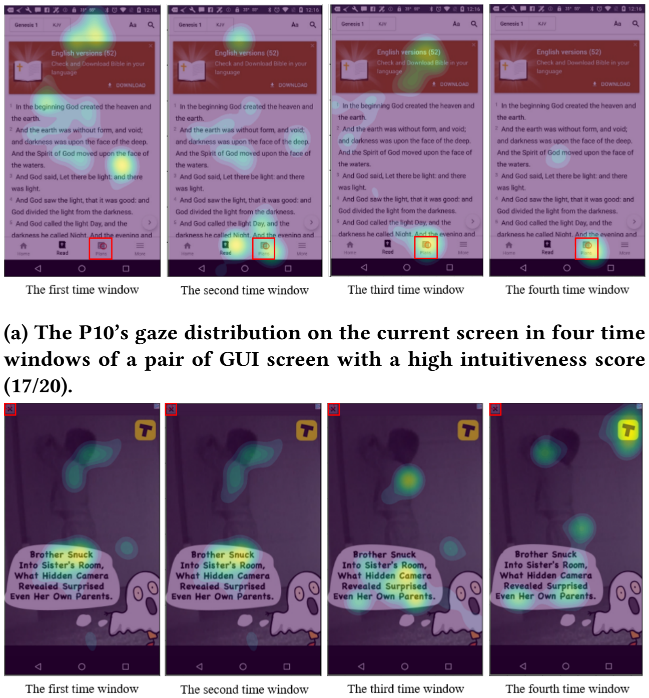
We explore the relationship between the intuitiveness score of different pairs of UI screens and the users' attention patterns. Previous research widely adopts two kinds of metric in analyzing eye-tracking data: fixation and scan path such as viewed sequence of AOI(s) [29]. Our data displays no significant correlation between the scan path and the intuitiveness score. Consequently we consider spatial entropy [28] and the AOI (area of interest) fixation counts percentage [29], to explore gaze patterns. Spatial entropy measures the high-level dispersion of the user's attention across the UI screen, while AOI fixation count percentage determines the user's low-level attention on a specific AOI. Spatial entropy is computed based on the fixation distribution heatmap, given by

$$h(x) = \frac{1}{T}\sum_{t=1}^{T} N(x\left|g_t, \sigma^2 I\right) \tag{1}$$

where T denotes the total time period we compute the heatmap, x denotes the indexing image pixels on the GUI screen, g_t denotes the fixation point at time t, and $\sigma = 5$ pixels. Then, spatial entropy is given by

$$E(h) = -\sum_{x \in h} h(x) \cdot \log_2(h(x)) \tag{2}$$

The heatmap describes the distribution of all fixation points on the UI screen in a certain time period while the spatial entropy measures the degree of exploration of the UI screen by the participant [28, 30, 36]. Higher spatial entropy indicates more exploration. To understand how participants' visual exploration varied in the



(a) The P10's gaze distribution on the current screen in four time windows of a pair of GUI screen with a high intuitiveness score (17/20).

(b) The P10's gaze distribution on the current screen in four time windows of a pair of GUI screen with a low intuitiveness score (6/20).

Figure 2: P10's gaze distributions on a high intuitiveness score screen and a low intuitiveness score screen

process, we divide each participant's labeling duration into four time windows and calculate the corresponding four spatial entropies over time. We analyze the correlation between these four spatial entropies and the intuitiveness score of the UI screen pairs.

Regarding the AOI fixation count percentage, we define the ground truth link UI as the AOI and calculate the AOI fixation count percentage as

$$PoF = \frac{AOI\_fixation}{Total\_fixation} \tag{3}$$

where *PoF* denotes the percentage of fixations in the AOI (ground truth link UI component), *AOI_fixation* denotes fixations that fall into the AOI (the ground truth link UI component), and *Total_fixation* denotes the total fixation numbers in the certain time period. The AOI fixation count percentage measures the weight of the user's visual attention on the AOI during the identifying period. We analyzed the fixation count percentage (PoF) of the whole time duration including the PoF of each time window *e.g.*, $PoF_{4th\_time\_window}$ of the 4th time window.

3.4 Intuitive Navigation Design and Users' Attention Distribution Pattern

The gaze heatmaps and their evolution over the four time windows provide distinct insights on intuitiveness. UI navigation with a high intuitiveness score (17/20) lead to a gradually convergent attention distribution towards the correct link UI, while UI with a low intuitiveness score (6/20) have a spread distribution in all time windows (see Figure 2).

Table 1: Correlation between the average spatial entropy of all subjects in each time window and the intuitiveness scores

Time window	Correlation with intuitiveness score
The 1st time window spatial entropy	-0.019
The 2nd time window spatial entropy	-0.082
The 3th time window spatial entropy	-0.130
The 4th time window spatial entropy	-0.367*

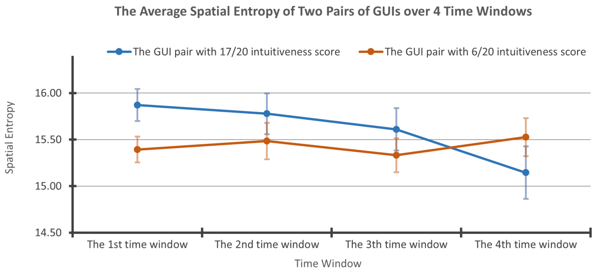


Figure 3: The average spatial entropy of four time windows on two pairs of UI screens, with a 17/20 (see Figure 2a) and 6/20 (see Figure 2b) intuitiveness score, respectively.

Spatial Entropy. Figure 3 plots the average spatial entropy for all participants in each time window for the two screen pairs with an intuitiveness score (17/20) and (6/20), respectively, presented in Figure 2. The average spatial entropy gradually decreases from the first time window to the fourth time window for the GUI with high intuitiveness score. However, entropy remains constant across all time windows in the case of the GUI with a low intuitiveness score (6/20), even displaying a slight increase in the last one.

We also calculate the Pearson's correlation coefficient between average spatial entropy in each time window and the intuitiveness score of the given GUI screen pairs (see Table 1). The negative correlation between the average spatial entropy and the intuitiveness score strengthens over time, with a significant ($p<0.05$) negative correlation observed in the fourth time window. This trend underscores the eye movement behavior of participants on GUIs with high intuitiveness scores, where individuals initially scan the entire interface before gradually focusing on their target UI component.

AOI Fixation Count Percentage. We do not find a significant correlation between the $PoF_{all\_time\_windows}$ and the intuitiveness score. Participants' visual attention distribution is thus not directly related to their understanding of the GUI screens. Since users' gaze tends to converge in the final time window before they label the link UI component, we also calculate the $PoF_{4th\_time\_window}$ of the 4th time window.

Spearman correlation test reveals that the intuitiveness score of given UI screen pairs has a positive correlation with a coefficient of 0.346 ($p<0.05$) with the $PoF_{4th\_time\_window}$. The more visual attention the participants pay to the ground truth link UI component, the more chance they can make the correct selection. However, the coefficient does not indicate a very strong correlation for other reasonable circumstances. For example, some participants may direct their visual attention to multiple UI elements when hesitating among several candidates. They may then exclude the one that they viewed the most and label another one or choose "uncertain".

We conclude that well-designed UI navigation can gradually attract users' attention to the correct link UI. However, the variation of their attention distribution over time on other UI elements can also reveal their mental model. This motivates us to not only predict the users' choices but also simulate their attention distribution.

4 UI Link Transformer: Simulating Users' Attention Distribution When Identifying Link UI

In this section, we propose an AI model, named UI Link Transformer (UILT) to simulate users' attention distribution when identifying link UI to navigate to the desired UI screen. We adopt the data collected during the dataset collection session of the formative study to develop the UILT.

4.1 Problem Modeling

We state our target problem mathematically as follows. The current screen X_c can be denoted as a set of features: $\{x_c^1, x_c^2, \ldots, x_c^s\}$, where s denotes the number of UI elements of the current screen. The target screen X_t can be denoted as a set of features $\{x_t^1, x_t^2, \ldots, x_t^e\}$, where e denotes the number of UI elements of the next screen. The UI element that links these two UI screens can be denoted as X_c^o and the user's gaze distribution on the current screen features are represented as set of probabilities: $\{g_c^1, g_c^2, \ldots, g_c^s\}$. As such, we target to develop a seq2seq model F. The input of our model is two sequences of screen features from the current screen and the target screen, and the output of our model is a sequence of probability $\{P(x_c^1 \mid F(X_c, X_t)), P(x_c^2 \mid F(X_c, X_t)), \ldots, P(x_c^s \mid F(X_c, X_t))\}$. When simulating the users' attention distribution, the model is trained to learn the parameter θ which minimizes the Kullback-Leibler Divergence (KL Divergence) between the predicted probability distribution and users' gaze distribution: $\mathrm{KL}(\{g_c^1, g_c^2, \ldots, g_c^s\} || \{P(x_c^1 \mid F_\theta(X_c, X_t)), P(x_c^2 \mid F_\theta(X_c, X_t)), \ldots, P(x_c^s \mid F_\theta(X_c, X_t))\})$. KL Divergence [20] quantifies the difference between the simulated distributions p and the ground truth distribution q with the formula (4).

$$KL(q||p) = E_q[\log \frac{q(X)}{p(X)}] \tag{4}$$

4.2 Feature Encoding

We first use UIED [48] to detect the UI elements of a given screenshot and reconstruct the bounding box of each UI element. As for the UI elements which do not have text labels, we use LabelDroid [5] to add labels. Since the original view hierarchy parsing from the mobile operating system is not always available, using computer vision techniques to obtain the metadata of each UI element makes our methods more generalizable. To vectorize each UI element on both screens in a pair, we use pre-trained ResNet50 and sentence-BERT to extract the visual features and text features (see Figure 4). We concatenate the bounding box parameter of each UI element with the extracted visual and text features, following the embedding method used in ActionBert [16]. Finally, each UI element is embedded into a 2820-dimensional vector, consisting of 2048-dimensional visual features, 768-dimensional text features, and 4-dimensional bounding box parameters.

Table 2: The KL divergence between the simulated attention distribution and users' attention distribution in four time windows (TW). (The lower the value, the more similar the two distributions are.)

Model variants	KL in TW 1	KL in TW 2	KL in TW 3	KL in TW 4
Random Selector	0.1435	0.1433	0.1431	0.1463
UILT without pre-training	0.0449	0.0473	0.0593	0.0577
UILT with pre-training	0.0497	0.0552	0.0587	0.0702

4.3 Model Structure

Figure 5 provides an overview of the UILT model, which processes the UI features from the current and target screens as separate inputs. The backbone structure of this model follows the Transformer architecture [42], which forwards the input words into a Transformer encoder and the target words into a Transformer decoder. We use the target screen as the input of the Transformer encoder and the current screen as the input of the Transformer decoder. During this process, the features on the target screen are considered together through the self-attention mechanism in the Transformer Encoder, and the attention distribution on the current screen can be decoded through the cross-attention mechanism of the Transformer Decoder layers. The output of the Transformer decoder is a sequence of features, where the length of the sequence is the same as the number of UI elements on the current screen. With MLP layers and a softmax prediction head, the model can output possibility distributions for simulating the human attention distribution which represents the degree of attention given to each UI element.

4.4 Model Training for Simulating Users' Attention Distribution

We first pre-train the UILT model on the Rico dataset for predicting the ground truth of the link UI element between two consecutive UI screens, as the work ActionBert [16] did. Through this step, our model learns foundational knowledge for predicting the link UI element. We then fine-tune the pre-trained model on the human-labeled dataset to simulate users' attention distribution.

For the fine-tuning process, we allocate 90% of human-labeled data samples for training the model and reserve the remaining 10% for model validation. We also evaluate a model trained directly on the human-labeled dataset for simulating users' attention distribution on the current screen in the four time windows, without model pre-training. The dataset splitting strategy adhered to a 90% portion for model training and a 10% portion for model validation.

These models are implemented with Pytorch, and trained with a Nvidia A100 GPU using a batch size of 32 pairs of screens, and the Adam optimizer [23]

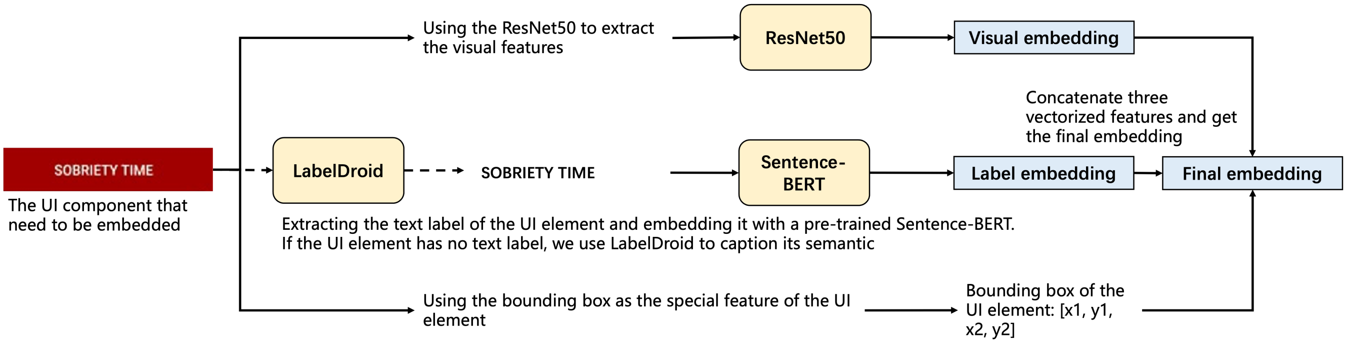

Using the ResNet50 to extract the visual features
ResNet50
Visual embedding
Concatenate three vectorized features and get the final embedding
SOBRIETY TIME
The UI component that need to be embedded
LabelDroid
SOBRIETY TIME
Sentence-BERT
Label embedding
Final embedding
Extracting the text label of the UI element and embedding it with a pre-trained Sentence-BERT.
If the UI element has no text label, we use LabelDroid to caption its semantic
Using the bounding box as the special feature of the UI element
Bounding box of the UI element: [x1, y1, x2, y2]

Figure 4: The pipeline of encoding the UI element features

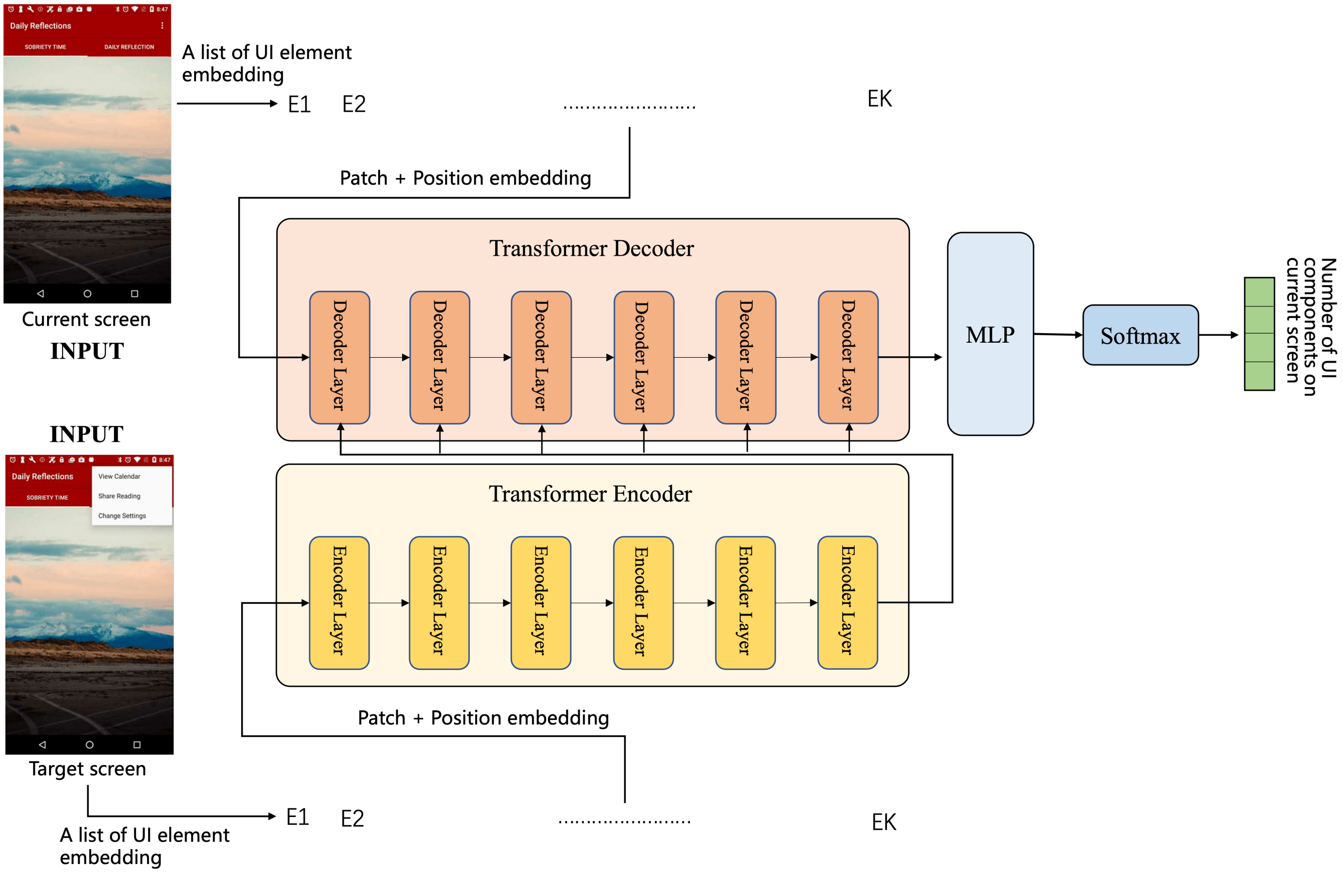


Figure 5: The model structure of UILT

4.5 Evaluation and Interpretion

Table 2 shows the KL divergence between the simulated attention distribution and users' attention distribution in the four time windows. The KL divergence is a common metric for evaluating the difference between the distributions. The lower the value, the more similar the two distributions are. The model without pre-training achieves a KL divergence of 0.0449 in time window 1, 0.0473 in time window 2, 0.0593 in time window 3, and 0.0577 in time window 4. The model that is finetuned on the pre-training model achieves the KL divergence of 0.0497 in time window 1, 0.0552 in time window 2, 0.0587 in time window 3, and 0.0702 in time window 4.

Since no previous work can be used as a baseline, we compare our model with a random selector, which displays a significantly higher KL divergence in each time window compared to both UILT models. UILT can thus simulate users attention better than a random selector.

5 Discussion

In this section, we discuss the importance of designing intuitive GUI navigation, how our method make novel contributions to existing methods, the limitations of this work, as well as the directions to move forward.

5.1 Importance of Intuitive GUI Navigation Design

Users often rely on memory and trial-and-error to navigate mobile apps, which is not ideal. Improving the intuitiveness of UI navigation significantly enhances user experience. According to Dorum and Garland, UI design should enable users to leverage their previous experiences to form accurate mental models [11]. Non-intuitive navigation increases the cognitive load, leading to frustration and potential abandonment, particularly among specific demographics such as older adults [27]. While some studies suggest using interface metaphors to enhance intuitiveness [2, 24], there is also a need for instant design evaluation methods. Our approach uses a data-driven AI model to evaluate various design concepts efficiently, addressing the time constraints associated with user testing in iterative design processes.

5.2 Novel Contributions to Existing Methods

This study introduces a unique approach by modeling GUI navigation intuitiveness through user attention distribution. Subjective experiences like user engagement and brand personality, which are traditionally challenging to quantify [46, 47] are usually modeled with scale values. We predict user attention distribution to understand their interaction with the GUI navigational elements, which provides UI designers with detailed, behavior-based insights rather than abstract metrics, facilitating more informed design decisions.

Prior research on user attention prediction primarily focuses on perceptual characteristics of UI elements like visual saliency. Such a perspective cannot reflect users' thinking process when reasoning the relationship between the current screen and the target screen. Our work differs from previous saliency prediction works by reflecting users attention distribution in a goal-oriented UI navigation process.

5.3 Limitations and Future Directions

5.3.1 Study Design Limitations. Our research currently focuses on single-step navigation tasks, which simplifies data collection and analysis but does not capture the complexities of multi-step navigation. This limitation could narrow the applicability of our findings and the UILT model.

Moreover, the integration of UILT into actual design tools has not yet been tested in practice, limiting our evaluation to theoretical scenarios. Future studies should aim to implement UILT within design workflows to assess its practical effectiveness quantitatively.

5.3.2 Data Collection and Analysis Challenges. The data for training UILT were labeled by individual participants, which may introduce bias. Additionally, instances where participants opted not to make a selection were excluded from the analysis but could provide insights into confusing design elements. Future research will include these data to enhance our understanding of user navigation challenges.

5.3.3 Model Performance and Usage Concerns. Due to the absence of established benchmarks, we compared our model's performance against random selectors after training with various methods. This approach does not fully illustrate UILT's effectiveness. We aim for this study to serve as a foundational baseline for future research in this area. Additionally, this work introduces a novel approach to predicting user attention distribution, shifting the focus from merely understanding users' perceptual characteristics to analyzing their cognitive processes in goal-oriented tasks. This shift is expected to guide subsequent research towards more comprehensive user behavior modeling.

5.3.4 Expanding Research and Application. To enhance UILT's performance, we aim to collect a broader range of data, including from specific demographic groups such as older adults. Additionally, we plan to integrate UILT with large language models (LLMs) to improve its application in design practice. This integration is expected to not only increase the explainability of UILT's predictions but also bolster UI designers' confidence in the model. This collaborative approach will help refine UILT's utility and reliability in real-world design settings.

6 Conclusion

This paper presents the UI Link Transformer (UILT), an AI model designed to predict users' attention distribution on mobile UI screens during navigation tasks. A formative user study was conducted to explore the correlation between the intuitiveness of navigation design and users' attention patterns. This research lays the groundwork for developing predictive models that simulate user attention distribution based on UI screen features when trying to navigate to a target UI screen. The evaluation of UILT establishes a benchmark for this innovative computational task, which focuses on predicting user attention distribution in a goal-oriented context.

Acknowledgments

The work described in this paper was supported by the HKUST Center for Aging Science (project ID: Z1010)

References

[1] Jeongwon Baeg, Atsushi Hirahara, and Yoshiaki Fukazawa. 1994. A development strategy of user navigation systems and GUI applications. In *Proceedings Eighteenth Annual International Computer Software and Applications Conference (COMPSAC 94).* IEEE, 163–168.

[2] Pippin Barr, Robert Biddle, and James Noble. 2002. A taxonomy of user-interface metaphors. In *Proceedings of the SIGCHI-NZ Symposium on Computer-Human Interaction.* 25–30.

[3] Agnieszka Bojko. 2005. Eye tracking in user experience testing: How to make the most of it. In *Proceedings of the UPA 2005 Conference.*

[4] A Burrell and Angela C Sodan. 2006. Web interface navigation design: which style of navigation-link menus do users prefer?. In *22nd International Conference on Data Engineering Workshops (ICDEW'06).* IEEE, 42–42.

[5] Jieshan Chen, Chunyang Chen, Zhenchang Xing, Xiwei Xu, Liming Zhu, Guoqiang Li, and Jinshui Wang. 2020. Unblind your apps: Predicting natural-language labels for mobile gui components by deep learning. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering.* 322–334.

[6] Hwayoung Cho, Dakota Powell, Adrienne Pichon, Lisa M Kuhns, Robert Garofalo, and Rebecca Schnall. 2019. Eye-tracking retrospective think-aloud as a novel approach for a usability evaluation. *International journal of medical informatics* 129 (2019), 366–373.

[7] Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara. 2018. Predicting human eye fixations via an lstm-based saliency attentive model. *IEEE Transactions on Image Processing* 27, 10 (2018), 5142–5154.

[8] Laura Cowen, Linden Js Ball, and Judy Delin. 2002. An eye movement analysis of web page usability. In *People and Computers XVI-Memorable yet Invisible: Proceedings of HCI 2002*. Springer, 317–335.

[9] 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.

[10] Emilia Djonov. 2007. Website hierarchy and the interaction between content organization, webpage and navigation design: A systemic functional hypermedia discourse analysis perspective. *Information Design Journal* 15, 2 (2007), 144–162.

[11] Kine Dørum and Kate Garland. 2011. Efficient electronic navigation: A metaphorical question? *Interacting with Computers* 23, 2 (2011), 129–136.

[12] C. Ehmke and S. Wilson. 2007. Identifying Web Usability Problems from Eyetracking Data. In *British HCI conference 2007*. 119–128. http://ewic.bcs.org/content/ConWebDoc/13298 © Stephanie Wilson et al. The original version of this article is published in the eWiC series at http://ewic.bcs.org/content/ConWebDoc/13298.

[13] David K Farkas and Jean B Farkas. 2000. Guidelines for designing web navigation. *Technical communication* 47, 3 (2000), 341–358.

[14] Camilo Fosco, Vincent Casser, Amish Kumar Bedi, Peter O'Donovan, Aaron Hertzmann, and Zoya Bylinskii. 2020. Predicting visual importance across graphic design types. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 249–260.

[15] Zhenyu Gu, Chenhao Jin, Danny Chang, and Liqun Zhang. 2021. Predicting webpage aesthetics with heatmap entropy. *Behaviour & Information Technology* 40, 7 (2021), 676–690.

[16] Zecheng He, Srinivas Sunkara, Xiaoxue Zang, Ying Xu, Lijuan Liu, Nevan Wichers, Gabriel Schubiner, Ruby Lee, and Jindong Chen. 2021. Actionbert: Leveraging user actions for semantic understanding of user interfaces. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 5931–5938.

[17] Yongquan Hu, Zhaocheng Xiang, Lihang Pan, Xiaozhu Hu, Yinshuai Zhang, and Aaron J Quigley. 2023. Exploring the Adaptation of Mobile GUI to Human Motion Status. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*. 145–147.

[18] A Ezgi İlhan. 2019. Eye-Tracking to Enhance Usability: A Race Game. In *Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference (IntelliSys) Volume 1*. Springer, 201–214.

[19] Yue Jiang, Luis A Leiva, Hamed Rezazadegan Tavakoli, Paul RB Houssel, Julia Kylmälä, and Antti Oulasvirta. 2023. UEyes: Understanding Visual Saliency across User Interface Types. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.

[20] James M. Joyce. 2011. *Kullback-Leibler Divergence*. Springer Berlin Heidelberg, Berlin, Heidelberg, 720–722. doi:10.1007/978-3-642-04898-2_327

[21] Marcel A Just and Patricia A Carpenter. 1980. A theory of reading: from eye fixations to comprehension. *Psychological review* 87, 4 (1980), 329.

[22] James Kalbach. 2007. *Designing Web navigation: Optimizing the user experience*. " O'Reilly Media, Inc.".

[23] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[24] George Lakoff and Mark Johnson. 2008. *Metaphors we live by*. University of Chicago press.

[25] Chunggi Lee, Sanghoon Kim, Dongyun Han, Hongjun Yang, Young-Woo Park, Bum Chul Kwon, and Sungahn Ko. 2020. GUIComp: A GUI design assistant with real-time, multi-faceted feedback. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[26] Luis A Leiva, Yunfei Xue, Avya Bansal, Hamed R Tavakoli, Tuðçe Köroðlu, Jingzhou Du, Niraj R Dayama, and Antti Oulasvirta. 2020. Understanding visual saliency in mobile user interfaces. In *22nd International conference on human-computer interaction with mobile devices and services*. 1–12.

[27] Rock Leung, Charlotte Tang, Shathel Haddad, Joanna Mcgrenere, Peter Graf, and Vilia Ingriany. 2012. How older adults learn to use mobile devices: Survey and field investigations. *ACM Transactions on Accessible Computing* 4, 3 (dec 2012), 1–33. doi:10.1145/2399193.2399195

[28] Congcong Liu, Karl Herrup, Seiko Goto, and Bertram E Shi. 2020. Viewing garden scenes: Interaction between gaze behavior and physiological responses. *Journal of Eye Movement Research* 13, 1 (2020).

[29] Ilia Maslov and Shahrokh Nikou. 2020. Usability and UX of learning management systems: an eye-tracking approach. In *2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. IEEE, 1–9.

[30] Mark Mills, Andrew Hollingworth, Stefan Van der Stigchel, Lesa Hoffman, and Michael D Dodd. 2011. Examining the influence of task set on eye movements and fixations. *Journal of vision* 11, 8 (2011), 17–17.

[31] Theresa Neil. 2014. *Mobile design pattern gallery: UI patterns for smartphone apps*. " O'Reilly Media, Inc.".

[32] Jakob Nielsen. 1999. The top ten new mistakes of Web design. *Retrieved August* 26 (1999), 2001.

[33] Jakob Nielsen and Kara Pernice. 2010. *Eyetracking web usability*. New Riders.

[34] May Phu Paing, Aniwat Juhong, and Chuchart Pintavirooj. 2022. Design and development of an assistive system based on eye tracking. *Electronics* 11, 4 (2022), 535.

[35] Lumpapun Punchoojit, Nuttanont Hongwarittorrn, et al. 2017. Usability studies on mobile user interface design patterns: a systematic literature review. *Advances in Human-Computer Interaction* 2017 (2017).

[36] Ricardo Ramos Gameiro, Kai Kaspar, Sabine U König, Sontje Nordholt, and Peter König. 2017. Exploration and exploitation in natural viewing behavior. *Scientific Reports* 7, 1 (2017), 2311.

[37] Eldon Schoop, Xin Zhou, Gang Li, Zhourong Chen, Bjoern Hartmann, and Yang Li. 2022. Predicting and explaining mobile ui tappability with vision modeling and saliency analysis. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–21.

[38] Immo Schuetz and Katja Fiehler. 2022. Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability. *Journal of Eye Movement Research* 15, 3 (2022).

[39] Amanda Swearngin and Yang Li. 2019. Modeling mobile interface tappability using crowdsourcing and deep learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[40] Mark Thibeault, Monica Jesteen, and Andrew Beitman. 2019. Improved accuracy test method for mobile eye tracking in usability scenarios. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63. SAGE Publications Sage CA: Los Angeles, CA, 2226–2230.

[41] Jenifer Tidwell. 2010. *Designing interfaces: Patterns for effective interaction design*. " O'Reilly Media, Inc.".

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[43] Jiahui Wang, Pavlo Antonenko, Mehmet Celepkolu, Yerika Jimenez, Ethan Fieldman, and Ashley Fieldman. 2019. Exploring relationships between eye tracking and traditional usability testing data. *International Journal of Human–Computer Interaction* 35, 6 (2019), 483–494.

[44] Jane Webster and Jaspreet S Ahuja. 2006. Enhancing the design of web navigation systems: The influence of user disorientation on engagement and performance. *Mis Quarterly* (2006), 661–678.

[45] Yixiang Wu, Jianxin Cheng, and Xinhui Kang. 2016. Study of smart watch interface usability evaluation based on eye-tracking. In *Design, User Experience, and Usability: Technological Contexts: 5th International Conference, DUXU 2016, Held as Part of HCI International 2016, Toronto, Canada, July 17–22, 2016, Proceedings, Part III 5*. Springer, 98–109.

[46] Ziming Wu, Yulun Jiang, Yiding Liu, and Xiaojuan Ma. 2020. Predicting and diagnosing user engagement with mobile ui animation via a data-driven approach. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[47] Ziming Wu, Taewook Kim, Quan Li, and Xiaojuan Ma. 2019. Understanding and modeling user-perceived brand personality from mobile application uis. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[48] Mulong Xie, Sidong Feng, Zhenchang Xing, Jieshan Chen, and Chunyang Chen. 2020. UIED: a hybrid tool for GUI element detection. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 1655–1659.

[49] Pingmei Xu, Krista A Ehinger, Yinda Zhang, Adam Finkelstein, Sanjeev R Kulkarni, and Jianxiong Xiao. 2015. Turkergaze: Crowdsourcing saliency with webcam based eye tracking. *arXiv preprint arXiv:1504.06755* (2015).

[50] Mihai Țichindelean, Monica Teodora Țichindelean, Iuliana Cetină, and Gheorghe Orzan. 2021. A comparative eye tracking study of usability—towards sustainable web design. *Sustainability* 13, 18 (2021), 10415.