



RecGaze: The First Eye Tracking and User Interaction Dataset for Carousel Interfaces

Santiago de
Leon-Martinez*
Brno University of
Technology
Brno, Czechia
santiago.deleon@kinit.sk

Jingwei Kang
University of Amsterdam
Amsterdam
The Netherlands
j.kang@uva.nl

Robert Moro
Kempelen Institute of
Intelligent Technologies
Bratislava, Slovakia
robert.moro@kinit.sk

Maarten de Rijke
University of Amsterdam
Amsterdam
The Netherlands
m.derijke@uva.nl

Branislav Kveton
Adobe Research
San Jose, CA, United States
kveton@adobe.com

Harrie Oosterhuis
Radboud University
Nijmegen, The Netherlands
harrie.oosterhuis@ru.nl

Maria Bielikova
Kempelen Institute of
Intelligent Technologies
Bratislava, Slovakia
maria.bielikova@kinit.sk

Abstract

Carousel interfaces are widely used in e-commerce and streaming services, but little research has been devoted to them. Previous studies of interfaces for presenting search and recommendation results have focused on single ranked lists, but it appears their results cannot be extrapolated to carousels due to the added complexity. Eye tracking is a highly informative approach to understanding how users click, yet there are no eye tracking studies concerning carousels. There are very few interaction datasets on recommenders with carousel interfaces and none that contain gaze data.

We introduce the RecGaze dataset: the first comprehensive feedback dataset on carousels that includes eye tracking results, clicks, cursor movements, and selection explanations. The dataset comprises of interactions from 3 movie selection tasks with 40 different carousel interfaces per user. In total, 87 users and 3,477 interactions are logged. In addition to the dataset, its description and possible use cases, we provide results of a survey on carousel design and the first analysis of gaze data on carousels, which reveals a golden triangle or F-pattern browsing behavior.

Our work seeks to advance the field of carousel interfaces by providing the first dataset with eye tracking results on carousels. In this manner, we provide and encourage an empirical understanding of interactions with carousel interfaces, for building better recommender systems through gaze information, and also encourage the development of gaze-based recommenders.

CCS Concepts

- Information systems → Recommender systems; • Human-centered computing → User studies; Human computer interaction (HCI).

*Also with Kempelen Institute of Intelligent Technologies.



This work is licensed under a Creative Commons Attribution 4.0 International License.
SIGIR '25, Padua, Italy

© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1592-1/2025/07
<https://doi.org/10.1145/3726302.3730301>

Keywords

Carousel interfaces, Browsing behavior, Eye tracking

ACM Reference Format:

Santiago de Leon-Martinez, Jingwei Kang, Robert Moro, Maarten de Rijke, Branislav Kveton, Harrie Oosterhuis, and Maria Bielikova. 2025. RecGaze: The First Eye Tracking and User Interaction Dataset for Carousel Interfaces. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), July 13–18, 2025, Padua, Italy*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3726302.3730301>

1 Introduction

Single ranked lists have been the default manner for presenting search and recommendation results for many years [16]. Different domains and scenarios may, however, require different interfaces. For instance, in web image search results are typically displayed in a 2D grid (i.e., single 2D ranked list), allowing users to inspect the results in a vertical as well as a horizontal direction [38]. *Carousel* and *multi-list*¹ interfaces have become a popular way for e-commerce and streaming services to display items (see Figure 1). These interfaces organize items into topics, where a topic consists of items that share certain characteristics. Multiple topics are shown below each other in a list that can be scrolled vertically and each topic is shown as a row that can be scrolled horizontally [2].

While carousels are an effective means of displaying recommendations through their use and success in practice (e.g. Netflix or Spotify), there is little research on why carousels may be better than other presentations formats. Moreover, there is little research in how users interact with them, which may greatly impact how systems are designed in practice. In contrast to the hundreds of publications on simpler interfaces like single 1D ranked lists and single 2D ranked lists [see, e.g., 6], we have only been able to locate fewer than 30 publications on carousels and multi-lists. We believe that this lack of research is due to:

- (1) a lack of publicly available datasets of interaction data with carousels and multi-lists, and
- (2) a lack of empirical studies of users' browsing behaviors.

¹Carousel interfaces are a subclass of multi-list interfaces with swipeable lists called carousels.

Our goal in this paper is to address the first problem. We leave addressing the second problem as future work. We hope that the release of a dataset of interaction data with carousels enables researchers to advance the field of carousel and multi-list interfaces.

This work extends beyond just carousel/multi-list interfaces and presents the first publicly available recommender dataset with eye tracking. Eye tracking technology has been steadily developing over the years. Examples of this can be seen in the continued development and popularity of virtual reality (VR) and augmented reality (AR) devices (e.g., Apple Vision Pro [1] or Meta Quest 3 [26]), which point to a near future where a VR/AR headset or glasses replace today's devices. Gaze will soon be an accessible online data stream for recommender system designers that can revolutionize recommenders by providing item observations, item dwell times, and gaze interaction sequences before a click. The latter can greatly help in determining what leads to a click or purchase.

This work seeks to achieve the following objectives:

- (1) Provide the first recommendation dataset with eye tracking data (along with cursor and selection explanations), enabling researchers to empirically understand how users browse carousel interfaces and advance the field of gaze-based recommenders;
- (2) Provide the first click feedback dataset in carousels and the second in multi-lists with more than 9 times as many interactions (movie selections) than prior work, enabling research in the underdeveloped field of carousel/multi-list [18]; and
- (3) Present the first gaze visualizations in carousel interfaces supporting golden triangle or F-pattern browsing behavior.

2 Related Work

2.1 User Interactions

Single 1D ranked lists & search engine results pages. In traditional single 1D ranked lists and search engine results pages, top-down browsing behavior was originally assumed. It was confirmed empirically through both click data [19] and eye tracking studies, in both search results [8, 14, 17, 27] and recommender systems [5, 23, 40]. These findings of top-down browsing behavior inspired the popular cascade click model [7]. Additionally, eye tracking studies were not only necessary to truly confirm browsing behavior, but also helped to learn parameters of click models.

Single 2D ranked lists (slates & grids). For single 2D ranked lists, eye tracking studies were also performed to determine browsing behavior. Kammerer and Gerjets [20] were one of the first to find that users did not browse line by line (left-right or top-down) and instead used a mixture of both. Later, Zhao et al. [40] found that in grid-based interfaces, user gaze behavior follows an F-shaped pattern (also known as the “golden triangle”) from top to bottom and from left to right, rather than being concentrated in the center of the screen. However, other studies have found that the F-shaped pattern does not apply to image search and browsing based on grid layouts [33, 34]. Xie et al. [39] further examined user attention on grid-based image search results page and found behavioral patterns such as “Middle bias,” “Slower decay,” and “Row skipping.” As single 2D ranked lists exhibit more complex browsing behaviors than single 1D ranked lists, it is reasonable to expect that carousel interfaces may have even more complex and varied browsing behaviors.

Multi-list/carousel interfaces. There are only a few user studies in carousel/multi-list interfaces and no eye tracking studies. Jannach et al. [18] compared carousels to a single 2D ranked list in item recommendation finding that users were slower and explored longer, while also perceiving the items in carousels as more diverse and novel. Similarly, Starke et al. [35] compared single 1D ranked lists, single 2D ranked lists, and carousels for recipe recommendations finding that single 2D ranked lists and carousels were both easier to use for participants than the single 1D ranked list. The most recent study by Loepf and Ziegler [25] examined user interactions more completely, finding that item position impacts selection probability (similarly to NDCG2D proposed for carousels [10, 11]) and there were noticeable differences in how users perceived and used the carousels. Rahdari et al. [28, 30] simulated users showing the advantages and efficiency of browsing carousels vs. single 2D ranked lists to find movies. Additionally, Rahdari et al. [29] extended the cascade click model to carousels called the “carousel click model,” which assumes that users browse topics until finding one of interest and then examine the items of that topic only.

2.2 Related Datasets

There are several commonly used datasets for recommendation systems research, such as the MovieLens dataset [15] or domain-specific ones, such as the MIND news dataset [37]. They typically contain large amounts of feedback (clicks and/or ratings), but lack other data sources. For example, we know of only one dataset that contains mouse cursor movements in search [22] and there are no existing datasets with gaze data.

In terms of existing carousel/multi-list recommender datasets, Bendada et al. [3] provided a simulation dataset from the music streaming platform Deezer with $n = 974,960$ user embeddings and $n = 862$ playlist embeddings with “ground truth” display-to-stream probabilities. While this is a large dataset, it may not reflect true user behavior and is not a click feedback dataset. To explore the effect of multi-list layout vs. single 2D ranked list in the context of similar-item recommendations, Jannach et al. [18] conducted a controlled study with 380 interactions (movie selection sessions) on multi-list interfaces with 6 unswipeable lists of 5 items each across multiple scenarios varying topics from non-descriptive to “Movies with the same genre” or actor/director among others. This is the only click feedback dataset available for multi-list interfaces and it does not include topic labels. Our dataset is more descriptive, is the only available in carousels (swipeable lists), and provides more than 9 times as many interactions.

3 User Study Methodology

To better understand how users browse carousel interfaces, we designed an eye tracked user study of 40 screens, where participants pick one movie found on a carousel page (or exit without selection). The user study was designed with the following goals in mind:

- (1) Determine any general browsing behaviors across users;
- (2) Determine the impact of carousel topic (genre) preference on browsing;
- (3) Examine the differences in browsing behavior across 3 browsing tasks: free-browsing, semi-free-browsing, direct search;
- (4) Survey users to address open questions in carousel design [24];

Table 1: Task summaries of interaction feedback, the number of movies selected (users can exit without selecting) and average screen time (until selection or exit).

Task	Users	Screens	Interactions	Movie Selections	Avg. screen time
Free-browsing	87	30	2,607	2,432	56.99s
Semi-free	87	5	435	383	43.85s
Direct Search	87	5	435	424	11.57s
All Tasks	87	40	3,477	3,239	49.67s

- (5) Provide an extensive mouse, click, and gaze dataset along with user explanations for future research.

The study consisted of three parts: (i) a pre-survey, (ii) movie selection tasks, and (iii) a post-survey. The survey sections gather user information, particularly their genre preferences, ideas on carousel interfaces, and feedback on the interface. The main part of the study are 40 screens of movie selection tasks. Of these 40 screens, 30 are free-browsing (pick any movie on the screen), 5 are semi-free browsing (pick a movie from your favorite genre), and 5 are direct search tasks (find a specific movie). The same screens and tasks were used for every participant. We used a larger number of free-browsing screens as this scenario is the closest to real-world interactions and while examining behavior in other scenarios may be insightful, we did not want to make the study too long.

Additionally, after the free-browsing and semi-free browsing search tasks, participants were asked to give feedback on their movie selection: familiarity with the movie selected and why they selected it. Along with selection explanations, we gathered mouse movements, clicks, gaze, and screen recordings throughout the movie selection tasks to provide the most complete recommender system dataset that is publicly available. A summary of the tasks and screens with the number of interactions (screens where feedback data was gathered) can be found in Table 1.

3.1 Screen Design & Movie Selection

Due to eye tracking limitations (physical setup, calibration, accuracy problems with movement, etc.) and the challenges of gathering other forms of feedback (clicks, cursor movements, etc.), we designed a user study in a controlled environment. We created custom screens emulating a carousel experience allowing for feedback to be gathered using available software and the Chrome browser.

The screens were made to allow participants to browse and select a movie in a carousel interface. Our design was inspired by the Netflix homepage as it is one of the most well-known examples of a carousel interface. This made it easy for participants to be comfortable with the interface and rely on previously learned browsing habits/behaviors. In Figure 1, we show a sample screen.

We restrict ourselves to movies from the years 2000 to 2023 to avoid any biases against older movies. Also, we did not include any movie duplicates across/within the screens. To provide a suitable number of movies for users to browse and allow at least 2 swipes to be possible (for examining list depth behaviors), we set our carousels to have 15 movies total with 5 movies shown at a time and 3 genre carousels at initial presentation (seen without scrolling down).

Only genres were used as carousel topics, as one of the goals of the study was to determine how carousel topic preference impacts browsing behavior from the gathered preference information. Genres are a simple and familiar topic for users that are much easier to compare between each other.

For the genres themselves, we chose 10 genres available from the International Movie Database (IMDB) genres: Action, Animation, Comedy, Crime, Drama, Fantasy, Horror, Romance, Sci-Fi, and Thriller. These 10 cover a wide range of genres that are familiar to users and can be typically seen in Netflix or other streaming services. Moreover, the list contains genres that are likely to be attractive to different user populations, such as Animation and Horror. Strong genre preferences (positive or negative) may have a greater impact on how users browse carousel interfaces.

Populating carousels with movies. In total, 40 screens were created with 150 movies per screen (15 movies for each of 10 genres). The IMDB non-commercial dataset (extracted on October 3, 2024)² was used for finding top/popular movies to populate our carousels along with The Movie Database (TMDB) API³ for finding movie posters and movie information (short description, year, actors). We used the number of votes from the IMDB dataset to determine the ranking of the movies (from most votes to least). The top 150 movies were removed to avoid very popular movies that users have seen or would know about (and to provide an instruction screen). We detail the whole procedure of populating carousels with movies below:

- (1) Sort IMDB dataset by number of votes.
- (2) Keep only movies from 2000 to 2023.
- (3) Remove top 150 movies.
- (4) Populate 10 ranked genre movie lists:
 - (a) Take the top movie in the dataset and place it in the top, unfilled position (i.e. append) in a genre list to which it belongs.
 - (b) Movies with multiple genres are appended to the least filled genre (i.e. highest ranked position among genres), and if genre lists are equally full, then it is decided by random assignment.
 - (c) Movies annotated as not having English text in their poster image are excluded.⁴
- (5) Populate ranked genre carousels for the 40 screens:
 - (a) For each genre, take the top 40 movies of the genre list.
 - (b) Randomly append these movies to the same top, unfilled position in the 40 genre carousels across all screens (i.e., the 1st movie position for all 40 Horror carousels).
 - (c) Continue with the next 40 top movies and unfilled positions until filling all genre carousels across the 40 screens.

This procedure creates screens with carousels that present 15 movies going from most popular to least popular. Additionally, across the 40 screens each genre carousel is popularity balanced by the random allocation allowing for a fair comparison across screens.

Genre ordering. The order of the genre carousels was performed randomly and separately for each of the 3 browsing tasks. For the 30 screens of free-browsing, genre ordering was determined randomly, while also guaranteeing that: (i) across the 30 screens, each genre is equally present within three initial presentation locations; and (ii) across the 30 screens, each genre is equally present in each

²<https://developer.imdb.com/non-commercial-datasets/>

³<https://www.themoviedb.org>

⁴Despite this, a few posters did not have text (see limitations in Section 6)

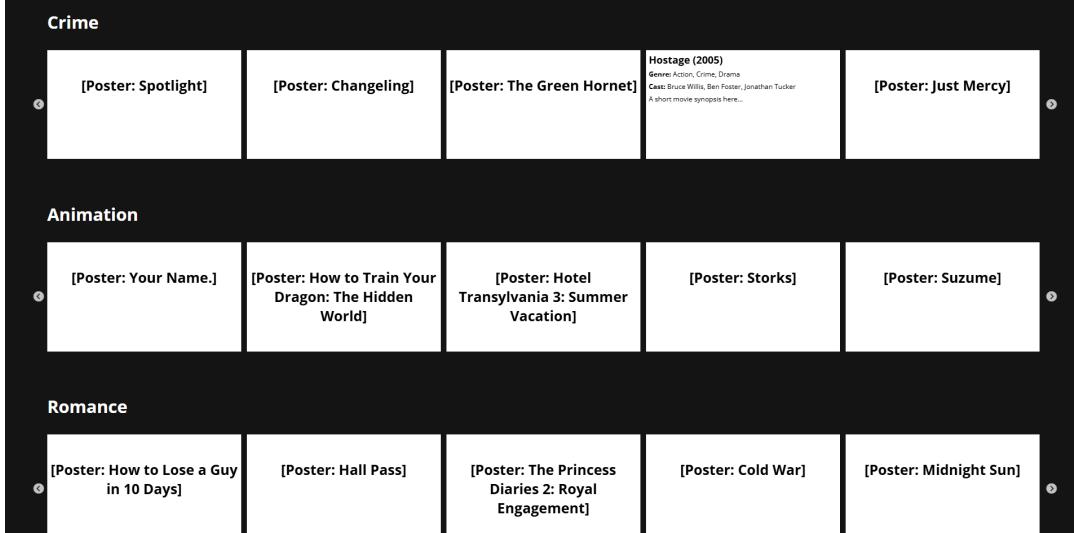


Figure 1: A capture of a free-browsing screen at initial presentation with the first 3 carousels shown and mouse-over details of a movie in the top row. Movies are ranked (left to right) by popularity based on IMDB number of votes. Movie posters and mouse-over details were shown (removed here due to copyright). See GitHub below for a sample screen recording gif.

of the 10 possible positions on the screens. For the 5 screens of semi-free-browsing and 5 screens of direct search, only (i) above is guaranteed across the 5 screens due to the low number of screens.

Movie details. Movie details are displayed by hovering over a poster with the mouse cursor showing: English title, year, genres, top 3 actors, and a short description. The title, genres, and year are from the IMDB Dataset, while the top 3 actors and the short description are from the TMDB API.

3.2 Data Collection & Study Procedure

Setting and apparatus. Data was collected in Bratislava, Slovakia and Amsterdam, Netherlands using identical setups: a Lenovo Thinkpad T490 laptop, Tobii 4C remote bar eye tracker (90 Hz sampling rate and error of less than 1 degree [13, 36]), and a 27 inch external monitor with resolution 1920x1080 and window scale 100%. Gaze, fixation, click, scroll, and cursor data were gathered by Eye Square's in-house testplayer software for eye tracking experiments. In addition, the positions of the browser elements, clicks, and scrolls were also collected through the Chrome browser. Eye tracking data, along with cursor and click data, were analyzed and visualized in Blickshift Analytics Build 2024, sub version 6 [4].

Study sample. Participants were gathered either from the institutions themselves or a coworking space, with those from the coworking being offered remuneration through a 15 euro coupon. The eligibility criteria were: (i) 18 years of age or older and (ii) ability to use mouse and keyboard to navigate. All participants were first given an explanation of the study along with an informed consent for participation in the study and for data sharing and processing.

In total, 87 participants were included in the final dataset, of which 47 were men and 39 woman with the following ages: (18–19): 1, (20–29): 50, (30–39): 29, (40–49): 6, and (60–69): 1. 61 participants were gathered in Bratislava while the remaining 26 were in Amsterdam. Additionally, 68 of the participants reported that they had used Netflix or a similar streaming service before; 19 had not.

Study procedure. All participants consented and fully completed the study in one-sitting composed of 3 parts as described below.

Pre-survey. After informed consent, participants were given a pre-survey that gathered basic information (personal information, age, gender) of the participants. Genre preference was assessed through 3 different questions: marking preferred/non-preferred for each, 1-star to 5-star rating for each, and pick the top or most preferred genre. The 3 different genre preference questions cover a wide range of common feedback, for comparison among each other (ranking genre preference) and to be relatable to other feedback datasets. To avoid any positional bias in survey presentation, the order of the genres were randomized for the preference questions.

Along with genre preference, information on frequency of use of movie/TV streaming services and frequency of movie watching (in any format) were gathered as proxies of user expertise with carousel interfaces and movie expertise.

As a follow-up to a previous user study into carousel interfaces by Loepp and Ziegler [25], we also include a psychological decision-making scale to relate the scores to browsing behaviors in a future work. We also seek to help address the open questions presented in the same user study and survey paper [24]: what are the best type of topics to use for carousels? For 10 different types of carousels (i.e., genre-based, content-based, item-based, etc.), we ask users which they would find most helpful (see Section 5.2).

Movie selection tasks. The movie selection tasks were administered in sets of 10 screens each beginning by calibrating the eye tracker. Re-calibration of the eye tracker can improve the accuracy of the tracking, especially for tasks that can take many minutes to complete. The free-browsing task was separated into 3 sets of 10 screens (for a total of 30 screens) and the 5 screens of semi-free browsing were followed by 5 screens of direct search to complete the last set of 10 screens. Half of the participants completed the 3 sets of 10 free-browsing screens followed by the 5+5 screens of

the semi-free-browsing and direct search, while the other half completed the semi-free-browsing and direct search set first. This was done to control for the effect of tiredness from previous screens.

Across all tasks, participants were asked to imagine that they were at home and had opened Netflix (or a similar streaming service) to find a movie to watch. In the case of the free-browsing task, they were asked to pick any movie that they would be interested in watching. In the semi-free-browsing task, participants were simply limited to picking a movie from their top genre choice from the pre-survey. In the direct search task, participants were instructed to find a specific movie and also provided the genre in which the movie was found (see GitHub for a visualization of search targets). The 3 tasks cover different scenarios/intents a user may have when finding a movie using a carousel homepage (without use of the search); we will examine the differences in the eye tracked browsing behavior in our future work.

With regards to study timing, in general we did not limit the time participants spent on the screens. For the free-browsing task, on average participants spent 56.99 seconds on a screen with a minimum of 1.17s and maximum of 412.09s. However, some participants who spent 2 minutes or more on the majority of free-browsing screens were advised to try to browse a bit faster and that they only needed to pick a movie they would like to watch. These participants that were advised (and when) are marked in the dataset.

Post-movie selection explanation. Following movie selection in free-browsing and semi-free-browsing screens, two questions were asked to determine the users' familiarity with the movie (i.e., already watched, seen a trailer, heard of the movie, unheard of, etc.) and why they picked the movie (i.e., poster, details, already wanted to watch, movie has been recommended, I don't know, etc.).

Post-survey. After completion of the 40 screens of movie selection, participants filled out a post-survey. The post-survey gathered information on the user experience (from 1–10) of the free-browsing task and tiredness/exhaustion after completing the study. Additionally, we gathered feedback on the interface design asking if the participants were overwhelmed by the number of genres or number of movies in each carousel and if they would like to see more movies or genres along with different types of topics for carousels (see Section 5.2). The last two questions provided feedback/suggestions for improving the interface and if participants would be willing to customize their carousel homepages (for both see Section 5.2).

4 Dataset Description

In this section we describe the main features, format, and preprocessing of the RecGaze dataset.

4.1 Dataset Features

We provide two datasets for the community to use due to the sensitive nature of raw gaze and cursor data that could be used to potentially link participants between datasets. The first is a public dataset that can be found on Zenodo⁵ with summarized fixation and mouse information with already timestamped labeling of Areas of Interest (AOIs) for cursor, clicks, and fixations. The supplementary material (stimuli, all screen layouts, etc.) are available in a

GitHub repository.⁶ The public dataset is designed to be an already preprocessed dataset that can be used out of the box for research in recommender systems, information retrieval and human computer interaction. It is organized in 4 dataframes in csv format. The user features, item features, and summary feedback dataframes contain the features (as columns) mentioned in Table 2. We additionally provide a fourth dataframe that is a simple click feedback dataset with only the final movie selection click per screen per user (in total 3,239 selections) and also contains the selection explanation information.

The second is a non-public dataset that contains sensitive information that can be provided by reaching out to the authors. It is better suited for researchers that are more interested in eye/mouse tracking research or would like to do the gaze preprocessing (e.g., AOI linking and fixations) themselves. We provide example use cases of the dataset in Section 4.3. In the following section, we outline our preprocessing procedure for AOI determination.

4.2 Data Preprocessing

Time alignment. All timestamped data (gaze, fixation, cursor, click) is aligned to the screen recording video start time of 0:00. Data before the carousel interface webpage loads as well as after the movie selection click was removed.

Gaze preprocessing. All gaze that was found to be outside of the screen was removed from the dataset and thus did not impact fixation calculation. Fixations were calculated using the Velocity-Threshold (I-VT) algorithm [32]. All fixations under 60ms were discarded, a common practice in processing gaze data that removes eye movement artifacts not linked to cognitive processes [12, 31].

Linking gaze and cursor location with AOIs. One of the most important steps in processing the data to leverage the cursor and eye tracking information, was linking gaze and cursor (x, y) screen pixel location with the corresponding AOI: carousel genre text, movie posters, and swipe buttons. In most eye tracking experiments, AOIs are predetermined using a fixed stimuli (the scene or screen being observed) or computer vision is used for object detection. In our case, we had a moving stimuli (the webpage with carousels), which could be scrolled vertically or swiped (scrolled horizontally).

To adjust for vertical scrolling, a fixed stimuli of the whole webpage was created using the position of the AOIs scraped as elements from the browser (see Figure 5). Then scroll data was used to adjust (x, y) positions for fixations, cursor movement, and clicks.

To adjust for swiping of the carousels (horizontal scroll), click events on the swipe buttons were gathered and then used to signal the change of AOIs to the next or previous set of movies. On the stimuli, 2 more identical pages were added for the 2 possible swipes. Combining the stimuli and the adjustments for vertical scroll and swipes, provides perfect AOI linking for cursor and click data and a possible approach for fixation AOI linking.

Gaze margin of error and fixation AOIs. Unlike cursor data, gaze data always has a margin of error in the (x, y) coordinate position. This is mitigated by the fixation algorithm that has a margin for aggregating gaze within a certain time and position window. Despite this, we provide 2 different methods for fixation

⁵<https://zenodo.org/records/15270518>

⁶https://github.com/santideleon/RecGaze_Dataset

Table 2: Major features of the RecGaze Dataset. Features marked with * are part of the non-public dataset and square brackets [] refer to multiple responses. To be concise, we use shorthand names, combine similar features, and denote different columns using bold. For a detailed explanation of the features refer to the GitHub documentation.

Category	Feature name	Type	Example
User features	UserID	Multi	KInIT_18 (Location_ID)
	Age*	Range	30-39
	Gender*	Text	Man
	Genre preferences	Multi	Q_Top: Action; Q_Preferred: [Action, Comedy]; Q_Ratings 1-5: [Action_5, Animation_1...]
	Interface expertise	Multi	3-4 times per week
	Movie expertise	Multi	1-2 times per month
Feedback & AOI label	Decision-making	Int 1-7	Q_1: 1; Q_2: 3; Q_3: 5; Q_4: 7; Q_5: 4; Q_6: 5;
	Genre/Movie #	Text	Q_1: Yes, I felt overwhelmed by # of Genres; Q_2: No, I didn't feel overwhelmed by # of movies
Item features	MovieID	Int	372058 (ID from TMDB)
	Screen position	Multi	TaskID/Screen: 7; Carousel_position: 1; Genre: Horror; Movie_position: 6
	Title & Year	Multi	Title: Hostage; Year: 2005
	Description	Text	"When a mafia accountant is taken host..."
	Poster	Img	http://[url]
Summary feedback	Gaze (90 Hz)*	Multi	Time: 1.241s; Position: 120, 300 px;
	Fixation (90 Hz)*	Multi	Time: 1.241s; Position: 120, 300 px; AOI: Movie_372058; Duration: 0.891s
	Cursor (event)*	Multi	Time: 2.450s; Position: 100, 600 px; AOI: GenreText_1_Horror
	Click (event)*	Multi	Time: 10.903s; Position: 100, 142 px; AOI: BackSwipe_Genre_1_Horror
Selection explanation	Movie familiarity	Text	I have seen part of the movie
	Why selected?	Text	[Because of the poster, Because of the details, Other: "I like the director"]
Other	Screen recording*	Video	25 fps .avi video recording

AOI processing, one of which assumes that participants are always gazing at an AOI on the screen (if within a distance threshold).

The first is the most simple and uses the exact rectangular bounds of the AOI in the webpage. In this case, fixation data that fall out of the rectangular bounds of the AOI (e.g., a movie poster) is not attributed to any AOI.

The second method works upon the first and allocates all unassigned fixation data to the nearest AOI (movie poster, forward or backward swipe button, or carousel genre text) up to a distance of 60 pixels.⁷ This method assumes that a participant is more likely observing an AOI rather than the empty background. Initial gaze analyses showed little difference between the two methods.

Data quality and missing values. The data gathered by the user study for the RecGaze dataset combines many modalities to make the most complete recommender dataset to date. To ensure the data quality gathered, we manually examined all fixation scan paths for every participant and screen (e.g., Figure 5a) to check for errors.

Across all screens there were only 3 free-browsing screens (0.1%) that failed to collect any feedback (clicks, cursor, gaze, fixation are all missing completely). For cursor data, 55 screens (1.6%) did not gather any mouse movements. For gaze data, 12 screens (0.33%) did not collect gaze data. Moreover, for the non-missing gaze data, we estimated the percentage of missing gaze data (below 90 Hz) was 2.9%. For fixation data, 61 (1.8%) screens had no fixations calculated.

For click data, we used the browser collected clicks to fill missing data from the software. As it is possible to not click, we compared

the click data to the selection explanations to determine screens that were missing the final movie selection click. Only 4 (0.1%) screens were missing the movie selection. Finally, we reviewed the direct search tasks and found that in 7 screens (1.6% of direct search screens), the participant failed to successfully find and click the instructed movie. These screens are not included in the click summary dataset and are noted in the GitHub documentation.

4.3 Use Cases for RecGaze Dataset

The RecGaze dataset is the first recommender dataset to provide a wide range of feedback data (eye tracking, cursor, clicks, and selection explanations). For this reason, it can lend itself to many applications. First and foremost, it allows one to comprehensively analyze and understand possible user behaviors within the carousel interfaces by combining all available modalities as illustrated in Section 5. Secondly, it can be used to fine-tune designs of these interfaces and their underlying models. The following are additional possible use cases for which the dataset may serve:⁸

Click models for carousel interfaces. Designing click models based on observed browsing behaviors like the F-pattern in Figure 4 and using the data to determine model parameters.

Offline evaluation. The browsing behaviors observed in the fixation or gaze data can be used to design better simulations of carousel users for offline experiments. Also, it can be used to understand positional bias in carousels.

⁷Calculated using the 1 degree error and screen dimensions along with the maximum recorded participant distance to screen 76cm.

⁸When applying the dataset in specific settings, it is worth considering the possible limitations of the dataset as discussed in detail in Section 6.

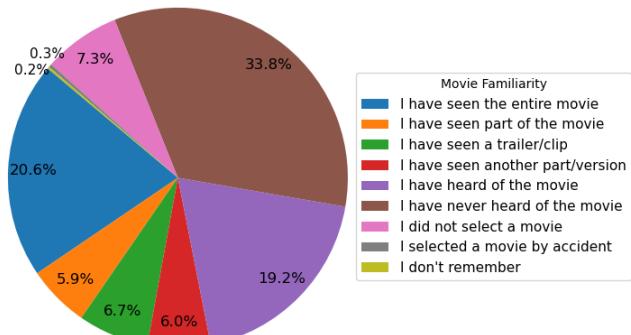


Figure 2: How familiar users are with the selected movie.

Whole page optimization. As an extension of click models or with general insights found in the data, one can better understand how different carousels interact with each other, particularly with the included preference information, and design better page layouts.

Impressions. The dataset can also be used for impressions, as they can be extracted and compared to the “ground truth” gaze data.

5 User Study Results

In this section we describe part of the results of the user study, namely the selection explanations and surveys motivated by the open questions posed by the survey paper of Loepp [24] on how to design carousel interfaces. We also provide the first ever visualizations of gaze data to demonstrate the dataset’s utility for analyzing the users’ browsing behavior. A more extensive analysis of this behavior is a subject of our future work.

5.1 Familiarity and Explanation for Selection

User familiarity with movie. Figure 2 shows the results of the movie familiarity question shown after each of the 3,045 free-browsing and semi-free browsing screens. Only one response was allowed and the question instructed participants to pick the first (topmost) true statement (ordering seen in legend).⁹

Why users selected a movie. In Figure 3, we present the results of the follow-up selection explanation question. Participants were able to select multiple responses and allowed to add their own free-form response (other). It can be seen that in more than half of movie selections, participants attributed it to the poster. This points to the importance of item visual biases in carousel interfaces and particularly for the case of movie recommendation.

5.2 Surveys on Carousel Design

In the only survey paper on carousels and multi-list interfaces, Loepp [24] present the following open questions that we address in this paper: (i) which types of carousels are preferred by which users, (ii) how many carousels they want to explore, and (iii) how many items per carousel ensure a good decision.

What types of carousel topics are the most helpful. In our survey, we defined ten different carousel topics (based on [21, 24] and examples taken from streaming services) for which we asked

⁹It is important to note that users were instructed to pick any movie that they would like to watch, which may include movies already watched.

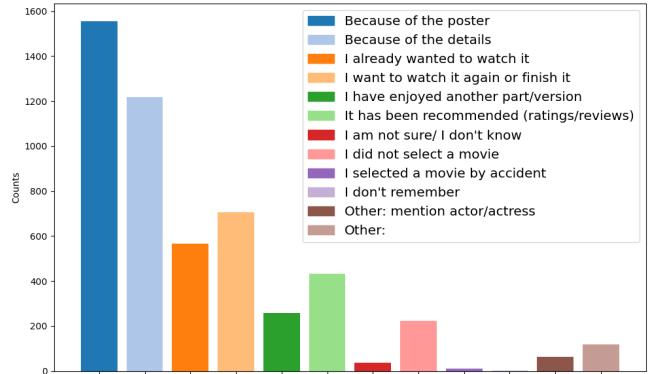


Figure 3: Why users select a movie (multiple responses).

participants how useful on a Likert scale they have been or imagine they would be in finding a movie.

Close to half of the participants (also the highest response among the 10) found the common *genre* topics very helpful. *Personalized* and *item-based* were the next topics that also had many positive responses. *Content*, *user-based*, *expert-based*, and *temporal* topics were all found to be more helpful than not. Finally, *global top/popular* was found to be slightly more helpful than not and the remaining two (*regional top/popular*, *exclusive*) were either neutral or not helpful. When comparing *global top/popular* to *regional top/popular*, it seems that participants would rather trust the global appeal of a movie than their own smaller region, but this may be mediated by culture and may vary by location. On the other hand, users preferred “the smaller group” personalized recommendations over user-based. In terms of unhelpful topics, most participants found that the *exclusive* topic was not helpful in finding movies. While streaming service providers are incentivized to present their own shows and movies, it appears that it is not helpful for the user browsing experience to have an exclusive topic.

How many carousels on a page. In the post-task survey, participants were asked multiple questions for feedback on the design of the carousel, namely number of carousels and items. For the number of carousels, 27.6% ($n = 24$) of the participants answered that they felt overwhelmed by the 10 genre carousels presented. When asked about having more genre carousels on the page (additional to the 10 present), 2.3% ($n = 2$) would have liked to see 5 or more genres, 5.7% ($n = 5$) would have liked to see 3–4 more genres, 14.9% ($n = 13$) would have liked to see 1–2 more genres, and 77.0% ($n = 67$) found that number was sufficient. These results show that 10 (or slightly less) genre carousels could be a good amount to satisfy exploration needs, but not overwhelm.

How many items in a carousel. For the number of movies in the carousels, 19.5% ($n = 17$) of the participants answered that they felt overwhelmed by the 15 movies presented (5 with 2 swipes). When asked about having more movies, 8.0% ($n = 7$) would have liked to see 5–10 more, 9.2% ($n = 8$) would have liked to see 15–20 more, 33.3% ($n = 29$) would have liked to see 25 or more additional movies, and 49.4% ($n = 43$) found that number was sufficient. In summary, 15 items (or more) seem to be a desirable number per carousel. Interestingly, services like Prime Video or Netflix tend to have many more carousel genres displayed (with Prime Video having infinite

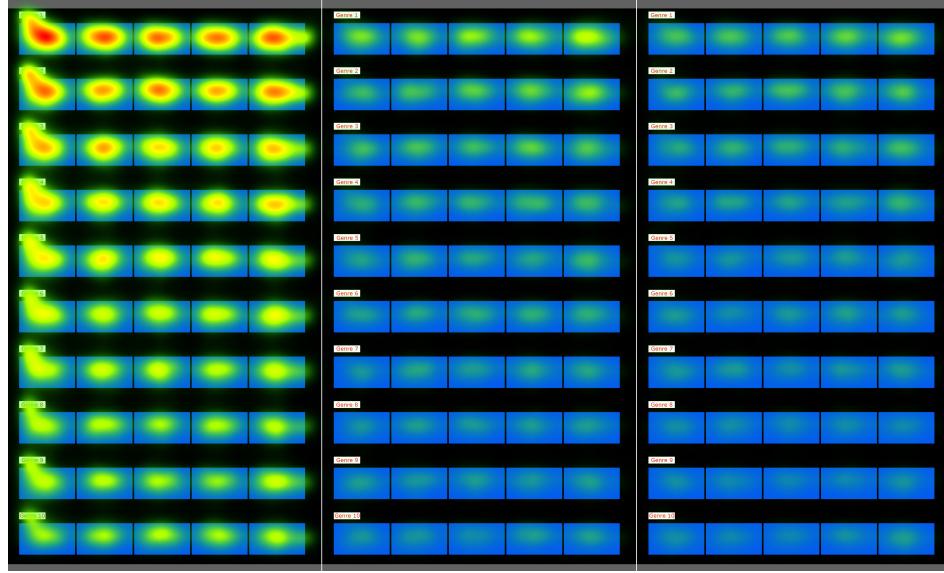


Figure 4: Aggregate fixation heat map for every user on all 30 free-browsing screens with duration weighting shown on a to-scale background stimuli of the movie screens (with movie posters shown as blue and genre text as white boxes). It includes horizontal displacement, so initial 5 movies can be distinguished from second and third set from swiping.

scroll) and many more movies/shows in each carousel. Based on these results, users may seem to expect or be satisfied with the number of movies in each carousel on these services, but also may be overwhelmed by the large number of carousels.

Suggestions for improving the study interface. In addition to the above, participants were asked if they would have liked to see other types of carousels presented, of which 55.2% ($n = 48$) wanted more types in addition to the 10 genres, 27.6% ($n = 24$) wanted more types replacing some of the genres, and 17.2% ($n = 15$) said that the 10 genres were sufficient.

Also, participants were asked to provide any comments or suggestions for improving the carousels with an open response. The most mentioned additions were: implementing IMDB ratings, adding director information, short clips/trailer on hovering on the poster, rearranging carousel topics to have preferred always at the top of the screen, and addition of different carousel topic types. The suggested features are found in popular streaming services and seem to be desired by users when interacting with a similar service.

Interest in customizable carousel homepages. One possible approach to improving the design of carousel interfaces is allowing users more autonomy by being able to customize their interface, which is in line with the suggestion to have preferred topics always at the top of the screen. Therefore, we asked the participants whether they would customize their streaming service homepage if they had the option to do so. Results showed that 70.1% ($n = 61$) would customize their homepage, 16.1% ($n = 14$) would not, and 13.8% ($n = 12$) did not know. This shows that allowing users to customize their homepage could be a feasible method to improve the user experience in carousel interfaces.

5.3 Visualizing Browsing of Carousels

In Figures 4 and 5, we present the first visualizations of gaze data in carousel interfaces. For Figures 4 and 5a, fixation data with (x, y)

pixel position value were used (non-public dataset), and Figures 5b and 5c used position value for clicks and cursor respectively (non-public dataset) rather than AOI allocation (public dataset).

In Figure 4, fixation data from all tasks across every user was combined to generate an aggregated heat map on a to scale stimuli (also included in GitHub). The heat map also takes into account fixation duration. Therefore, the heat is representative of both how long and how often users were fixating on a location. The heat map points to a strong top-down browsing behavior, left-right browsing behavior (at least on page 1), and a bias towards initially presented movies (unswiped). These preliminary results support the golden triangle browsing behavior or F-shaped pattern seen in single 2D ranked lists [40], which supports our original hypothesis that this behavior would also be seen in carousels [9]. However, on pages 2 and 3 more fixations can be seen on the farthest right movies which go against the general trend of F-pattern seen, particularly left-right browsing. We hypothesize that the increase in fixations on the last presented movie is due to how the swiping of the carousels works. However, there may be several different behaviors that could explain this result (e.g., central bias or a bias to the cursor/click position), which we plan to examine in our future work.

For Figures 5a to 5c, we selected one participant and one free-browsing screen to visualize the types of data available. They present scan paths showing the points where data were registered (red circles) with radius proportional to the duration. For this screen, the participant began to fixate on the first, left-most movie of genre 2 and ended on the fourth movie of genre 1, as can be seen in the fixation data from Figure 5a. Figure 5b shows the clicks for two same location forward swipes on genre 6 and the final click to select the fourth movie of genre 1. Finally, the cursor path can be seen in Figure 5c from an initial cursor position hovering the third movie of genre 2. The scan paths provide a detailed summary of

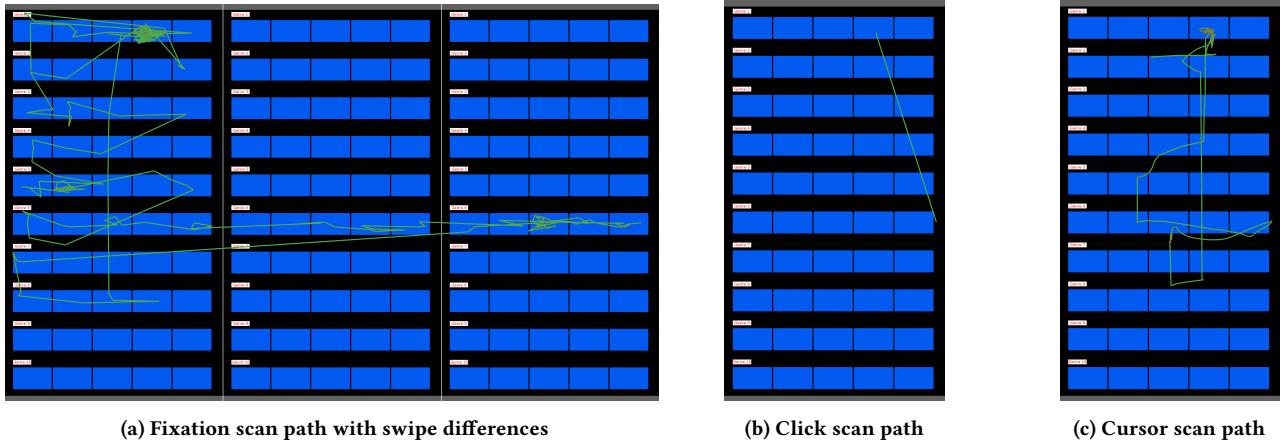


Figure 5: Visualizations of fixations, clicks and mouse cursor movements for one selected user and one free-browsing screen.

the participant’s browsing/interaction sequence. For example, multiple fixations (Figure 5a) and mouse cursor movements (Figure 5c) on the selected fourth movie of genre 1 are indicative of the user reading its description (tracing the text with the cursor).

It is also possible to add user preference information to interpret these results. In Figure 5a, we can see that the participant often examined multiple movies of each genre that they scrolled to. Outside of genre 1, the majority of fixations and all swipes are found in genres 5 (Drama) or 6 (Animation), which are both user’s preferred ones. On the other hand, genre 4 (Fantasy), which was marked as the top genre by the user, lacked fixations when compared to the other ones. We provide this as an example and leave an analysis of the impact of genres on browsing as future work.

6 Limitations

User study limitations. The hardware and software were identical between the two locations of the study, however, the lighting conditions differ, which could impact the accuracy of the gaze data.

Another limitation is related to participants who may be skewed to those of above average socioeconomic level and education, which may not represent the average streaming service users. While this most likely impacts preferences for certain movies, it is unclear if it impacts how one would browse in general.

Interface limitations. The study did not explicitly consider poster visual bias. Color and content may impact browsing behavior. We address it only indirectly through aggregation (over screens/participants), in which the individual screen biases are less impactful.

Secondly, although we only gathered movie posters from the MovieDB API annotated as having English text, there were a few cases that did not have any text, which could possibly encourage (standing out from the rest) or discourage user interest.

Dataset limitations. Due to small inaccuracies in eye tracking, it is possible that some fixations are misattributed to incorrect locations/AOs. This is most likely to happen when (i) a participant is gazing at the border between 2 movies or (ii) moving from the genre text to the movie below. The second case could lead to attributing genre text AOs to the movie below and vice versa. An important limitation is the size of the dataset. While it is more than 9 times larger than the only available click feedback dataset in multi-list

interfaces [18], the 3,477 interactions provided in the dataset are still small when compared to industrial datasets. This may pose difficulties when training recommenders just on this data, but it is possible to connect the movies in our study with other datasets [15]. Finally, results of the dataset may be limited to the movie domain. Browsing behaviors may vary in other domains, especially due to how different carousel topics can be to the movie domain.

7 Conclusion

Although carousel and multi-list interfaces are widely used in commercial services, thus impacting users daily, there is little research that is focused on this type of interface. In this work, we help address the barriers to research in carousels and encourage a return to eye tracking and empirical understanding of user behavior by providing the RecGaze dataset. We demonstrated the utility of the dataset for a range of use cases and provided the first gaze analysis of user behavior in carousel interfaces supporting the golden triangle or F-pattern browsing behavior. By providing this dataset, we hope to enable and advance research in carousel and multi-list interfaces as well as gaze-based recommendation systems.

In future work, we will present an extensive analysis of the gaze data gathered herein for determining how users browse carousel interfaces taking into account preferences, interaction sequences (i.e., swipes), and decision-making scale results.

Acknowledgments

We would like to acknowledge the support of Eye Square for providing hardware and software. Also, we would like to acknowledge Aneta Zugecova in helping administer the user study and Blickshift Analytics [4] for its support with software and eye tracking analysis. This work was supported by Eyes4ICU, a project funded by the European Union under the Horizon Europe Marie Skłodowska-Curie Actions, GA No. 101072410. It is also partially supported by NWO under project numbers VI.Veni.222.269, 024.004.022, NWA.1389.20.-183, and KICH3.LTP.20.006, and the European Union’s Horizon Europe program under grant agreement No. 101070212. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

References

- [1] Apple 2025. Apple Vision Pro – apple.com. <https://www.apple.com/apple-vision-pro/>. [Accessed 03-02-2025].
- [2] Walid Bendada, Guillaume Salha, and Théo Bontempelli. 2020. Carousel Personalization in Music Streaming Apps with Contextual Bandits. In *Proceedings of the 14th ACM Conference on Recommender Systems*. Association for Computing Machinery, 420–425.
- [3] Walid Bendada, Guillaume Salha, and Théo Bontempelli. 2020. Carousel Personalization in Music Streaming Apps with Contextual Bandits. In *Fourteenth ACM Conference on Recommender Systems*. ACM, Virtual Event Brazil, 420–425. <https://doi.org/10.1145/3383313.3412217>
- [4] Blickshift Analytics GMBH 2025. <https://www.blickshift.com/>
- [5] Sylvain Castagnos, Nicolas Jones, and Pearl Pu. 2010. Eye-tracking Product Recommenders' usage. In *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*. ACM Press, Barcelona, Spain, 29. <https://doi.org/10.1145/1864708.1864717>
- [6] Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. 2015. *Click Models for Web Search*. Morgan & Claypool Publishers.
- [7] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An Experimental Comparison of Click Position-bias Models. In *Proceedings of the international conference on Web search and web data mining - WSDM '08*. ACM Press, Palo Alto, California, USA, 87. <https://doi.org/10.1145/1341531.1341545>
- [8] Edward Cutrell and Zhiwei Guan. 2007. What Are You Looking For?: An Eye-tracking Study of Information Usage in Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 407–416. <https://doi.org/10.1145/1240624.1240690>
- [9] Santiago de Leon-Martinez. 2024. Understanding User Behavior in Carousel Recommendation Systems for Click Modeling and Learning to Rank. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining* (Merida, Mexico) (WSDM '24). Association for Computing Machinery, New York, NY, USA, 1139–1141. <https://doi.org/10.1145/3616855.3635734>
- [10] Nicolò Felicioni, Maurizio Ferrari Dacrema, and Paolo Cremonesi. 2021. Measuring the User Satisfaction in a Recommendation Interface with Multiple Carousels. In *ACM International Conference on Interactive Media Experiences*. ACM, Virtual Event USA, 212–217. <https://doi.org/10.1145/3452918.3465493>
- [11] Maurizio Ferrari Dacrema, Nicolò Felicioni, and Paolo Cremonesi. 2022. Offline Evaluation of Recommender Systems in a User Interface With Multiple Carousels. *Frontiers in Big Data* 5 (June 2022), 910030. <https://doi.org/10.3389/fdata.2022.910030>
- [12] Valentin Foucher and Anke Huckauf. 2024. Unveiling Deceptive Intentions: Insights from Eye Movements and Pupil Size. *Proceedings of the ACM on Human-Computer Interaction* 8, ETRA (May 2024), 1–17. <https://doi.org/10.1145/3655612>
- [13] Agostino Gibaldi, Mauricio Vanegas, Peter J. Bex, and Guido Maiello. 2017. Evaluation of the Tobii EyeX Eye Tracking Controller and Matlab Toolkit for Research. *Behavior Research Methods* 49, 3 (June 2017), 923–946. <https://doi.org/10.3758/s13428-016-0762-9>
- [14] Laura A. Granka, Thorsten Joachims, and Geri Gay. 2004. Eye-tracking Analysis of User Behavior in WWW Search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, Sheffield United Kingdom, 478–479. <https://doi.org/10.1145/1008992.1009079>
- [15] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Trans. Interact. Intell. Syst.* 5, 4, Article 19 (Dec. 2015), 19 pages. <https://doi.org/10.1145/2827872>
- [16] Marti A. Hearst. 2009. *Search User Interfaces*. Cambridge University Press.
- [17] Kajita Hofmann, Bhaskar Mitra, Filip Radlinski, and Milad Shokouhi. 2014. An Eye-tracking Study of User Interactions with Query Auto Completion. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, Shanghai China, 549–558. <https://doi.org/10.1145/2661829.2661922>
- [18] Dietmar Jannach, Mathias Jesse, Michael Jugovac, and Christoph Trattner. 2021. Exploring Multi-List User Interfaces for Similar-Item Recommendations. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 224–228.
- [19] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2016. Accurately Interpreting Clickthrough Data as Implicit Feedback. *ACM SIGIR Forum* 51, 1 (2016), 8.
- [20] Yvonne Kammerer and Peter Gerjets. 2010. How the Interface Design Influences Users' Spontaneous Trustworthiness Evaluations of Web Search Results: Comparing a List and a Grid Interface. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) (ETRA '10). Association for Computing Machinery, New York, NY, USA, 299–306. <https://doi.org/10.1145/1743666.1743736>
- [21] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2019. Personalized Explanations for Hybrid Recommender Systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) (IUI '19). Association for Computing Machinery, New York, NY, USA, 379–390. <https://doi.org/10.1145/3301275.3302306>
- [22] Luis A. Leiva and Ioannis Arapakis. 2020. The Attentive Cursor Dataset. *Frontiers in Human Neuroscience* 14 (Nov. 2020), 565664. <https://doi.org/10.3389/fnhum.2020.565664>
- [23] Yixuan Li, Pingmei Xu, Dmitry Lagun, and Vidhya Navalpakkam. 2017. Towards Measuring and Inferring User Interest from Gaze. In *Proceedings of the 26th International Conference on World Wide Web Companion* (WWW '17 Companion). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 525–533. <https://doi.org/10.1145/3041021.3054182>
- [24] Benedikt Loepf. 2023. Multi-list Interfaces for Recommender Systems: Survey and Future Directions. *Frontiers in Big Data* 6 (2023), 1239705.
- [25] Benedikt Loepf and Jürgen Ziegler. 2023. How Users Ride the Carousel: Exploring the Design of Multi-List Recommender Interfaces From a User Perspective. In *Proceedings of the 17th ACM Conference on Recommender Systems*. ACM, Singapore Singapore, 1090–1095. <https://doi.org/10.1145/3604915.3610638>
- [26] Meta 2025. Meta Quest 3 – Meta.com. <https://www.meta.com/quest/quest-3/>. [Accessed 03-02-2025].
- [27] Bing Pan, Helene A. Hembrooke, Geri K. Gay, Laura A. Granka, Matthew K. Feusner, and Jill K. Newman. 2004. The Determinants of Web Page Viewing Behavior: An Eye-tracking Study. In *Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA'2004*. ACM Press, San Antonio, Texas, 147–154. <https://doi.org/10.1145/968363.968391>
- [28] Behnam Rahdari, Peter Brusilovsky, and Branislav Kveton. 2024. Towards Simulation-Based Evaluation of Recommender Systems with Carousel Interfaces. *ACM Transactions on Recommender Systems* 2, 1 (March 2024), 1–25. <https://doi.org/10.1145/3643709>
- [29] Behnam Rahdari, Branislav Kveton, and Peter Brusilovsky. 2022. From Ranked Lists to Carousels: A Carousel Click Model. <http://arxiv.org/abs/2209.13426> arXiv:2209.13426 [cs, math].
- [30] Behnam Rahdari, Branislav Kveton, and Peter Brusilovsky. 2022. The Magic of Carousels: Single vs. Multi-List Recommender Systems. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media* (Barcelona, Spain) (HT '22). Association for Computing Machinery, New York, NY, USA, 166–174. <https://doi.org/10.1145/3511095.3531278>
- [31] Keith Rayner. 1998. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin* 124, 3 (1998), 372. Publisher: American Psychological Association.
- [32] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) (ETRA '00). Association for Computing Machinery, New York, NY, USA, 71–78. <https://doi.org/10.1145/355017.355028>
- [33] Sav Shrestha, Kelsen Lenz, Barbara Chaparro, and Justin Owens. 2007. "F" Pattern Scanning of Text and Images in Web Pages. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 51, 18 (2007), 1200–1204. <https://doi.org/10.1177/154193120705101831>
- [34] Christina Siu and Barbara S. Chaparro. 2014. First Look: Examining the Horizontal Grid Layout using Eye-tracking. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58, 1 (2014), 1119–1123. <https://doi.org/10.1177/1541931214581234>
- [35] Alain Starke, Edis Asotic, and Christoph Trattner. 2021. "Serving Each User": Supporting Different Eating Goals Through a Multi-List Recommender Interface. In *Fifteenth ACM Conference on Recommender Systems*. ACM, Amsterdam, The Netherlands, 124–132. <https://doi.org/10.1145/3460231.3474232>
- [36] Aryaman Taore, Michelle Tiang, and Steven C. Dakin. 2024. (The Limits of) Eye-tracking with iPads. *Journal of Vision* 24, 7 (July 2024), 1. <https://doi.org/10.1167/jov.24.7.1>
- [37] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. 2020. MIND: A Large-scale Dataset for News Recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Stroudsburg, PA, USA, 3597–3606.
- [38] Xiaohui Xie, Yiqun Liu, Maarten de Rijke, Jiyin He, Min Zhang, and Shaoping Ma. 2018. Why People Search for Images using Web Search Engines. In *WSDM 2018: 11th International Conference on Web Search and Data Mining*. ACM, 655–663.
- [39] Xiaohui Xie, Jiaxin Mao, Yiqun Liu, Maarten de Rijke, Yunqiu Shao, Zixin Ye, Min Zhang, and Shaoping Ma. 2019. Grid-based Evaluation Metrics for Web Image Search. In *The World Wide Web Conference* (San Francisco, CA, USA) (WWW '19). Association for Computing Machinery, New York, NY, USA, 2103–2114. <https://doi.org/10.1145/3308558.3313514>
- [40] Qian Zhao, Shuo Chang, F. Maxwell Harper, and Joseph A. Konstan. 2016. Gaze Prediction for Recommender Systems. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) (RecSys '16). Association for Computing Machinery, New York, NY, USA, 131–138. <https://doi.org/10.1145/2959100.2959150>