# ArtEx: An Interactive Visual Art Recommendation with Diversity and Popularity Control*

Bereket A. Yilma[1,*,†], Rully Agus Hendrawan[2,3,*,†], Peter Brusilovsky[2,*] and Luis A. Leiva[1]

[1]*University of Luxembourg, Luxembourg*
[2]*University of Pittsburgh, USA*
[3]*Institut Teknologi Sepuluh Nopember, Indonesia*

**Abstract**

Recommender Systems (RecSys) have transformed personalized applications by delivering tailored content and experiences. However, modern Deep Learning RecSys often operate as opaque "black boxes," offering users no control over how personalization is shaped. We introduce a novel algorithmic approach to bridge this gap in the context of visual art recommendation by integrating user agency directly into the RecSys engines. By allowing users to dynamically adjust facets such as content diversity and popularity, through the use of hyperparameters implemented as sliders, the system creates a feedback loop where users can actively tune recommendations while also helping the system to learn about their preferences. This approach ensures that personalization is not only algorithmically optimized but also user-driven, fostering a balance between automation and human control. The results of a large-scale user study (n=151) evidenced that sliders enhance engagement and recommendation quality by promoting meaningful exploration.

**Keywords**

Adaptation, Interface Personalisation, Design, Interaction Context,

## 1. Introduction

In recent years, the integration of AI-driven recommendation engines into various applications has led to a revolution in personalized content delivery [1]. These engines leverage sophisticated machine learning models to analyze user behavior and preferences, offering personalized recommendations. This advent in AI-based Recommender Systems (RecSys) has brought about a paradigm shift in personalization across various domains. In particular, the domain of Visual Art (VA) has witnessed a profound transformation, with AI-driven approaches playing a pivotal role in improving the quality of VA recommendations [2, 3]. Current AI-based VA RecSys engines, which leverage advanced Deep Learning (DL) techniques, have demonstrated remarkable capabilities in understanding the latent semantic relationships within artwork collections [4]. By analyzing user preferences, they provide high-quality recommendations that cater to individual tastes [3, 2]. Despite the enormous success of current AI-based RecSys, the lack of transparency and user control raises significant concerns [5]. Particularly, the "black box" nature of these RecSys, which take user data as a starting point and return recommendations, has left users detached from the decision-making process [6].

Users often find themselves at the mercy of opaque algorithms that determine their recommendations without offering insights into how these decisions are made. This opacity can lead to mistrust and a sense of alienation, as users may feel that their personal preferences and inputs are not adequately respected or understood, which is often the case in scenarios such as e-commerce websites where algorithms are optimising to maximize revenue suppressing actual user preferences [7]. Hence, the persuasive power of personalized recommendations can subtly manipulate user choices without their explicit consent or awareness, raising ethical concerns about autonomy and informed decision-making. The detachment from the decision-making process can diminish the overall user experience, reducing

*IntRS'25: Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, September 22, 2025, Prague.*
*Corresponding author.
†These authors contributed equally.
bereket.yilma@uni.lu (B. A. Yilma); ruhendrawan@pitt.edu (R. A. Hendrawan); peterb@pitt.edu (P. Brusilovsky); luis.leiva@uni.lu (L. A. Leiva)

satisfaction and engagement with the system [8]. Another major issue with the "black box" nature of AI-based RecSys is the creation of *filter bubbles*, where users are repeatedly exposed to a narrow set of content that reinforces their existing preferences [9]. This phenomenon can limit the diversity of content that users encounter, stifling discovery and altering users' exposure to alternative viewpoints [10] potentially leading to a monotonous experience. Such an approach can also inadvertently reinforce existing biases and stereotypes [9]. Given these challenges, it is of paramount importance to develop modern AI-based RecSys that offers users a means of control which allows them to fine-tune and adjust various aspects of their recommendations.

We set out to make a step toward user-controllable AI-based RecSys by (1) designing RecSys interfaces that offer users a control over different facets of their personalized recommendations (i.e, we experiment with two of the most commonly valued attributes, popularity and diversity) and (2) developing RecSys algorithms that can observe and leverage user control actions to fine-tune recommendations. In this work, we focus on the domain of visual art recommendation, where the ability to explore diverse and novel content is particularly valuable [11]. Hence, our central research question is: **How does user-controlled tuning of diversity and popularity affect the quality of recommendations and user engagement in VA RecSys?** To answer this question, we developed ArtEx; a user-controllable, personalized visual art exploration interface. ArtEx leverages Bootstrapping Language-Image Pre-training (BLIP) [12]; a state-of-the-art (SOTA) VA RecSys engine as its backbone selected as the best-performing engine across multiple beyond-accuracy measures [2]. ArtEx uses BLIP as a multimodal feature extractor to capture the semantic relationships of artworks. It then extends BLIP's capabilities by introducing user-controllable hyperparameters, enabling users to directly influence different aspects like diversity and popularity and fine-tune their recommendations. By jointly optimizing such personalization policies, ArtEx creates an interactive feedback loop, where user interactions not only adjust the recommendations but also help the system adapt, facilitating richer exploration and uncovering novel and valuable artistic content.

In sum, this paper makes the following contributions:

- We introduce ArtEx; a user-controllable personalized VA exploration interface.
- We develop and evaluate a VA RecSys algorithm with a transformer backbone and tunable diversity/popularity hyperparameters.
- We conduct a large-scale study (n=151) to assess the performance of our approach from a user-centric perspective.
- We contextualize our findings and offer VA RecSys design guidelines to enhance user agency and control.

# 2. Background and Related Work

## 2.1. User Control in Recommendation Systems

User control in RecSys research aims to improve user satisfaction and trust by allowing direct influence on recommendations [13]. Early works have shown the utility of allowing users to choose specific peers in collaborative filtering [14], sort recommendations by item features [15], and adjust preferences at various levels of granularity [16, 17]. Feedback mechanisms, such as explicit ratings and critiquing, have also been shown to improve personalization and engagement [18, 19].

Adjusting the weights of various components in the user profile of interests gradually emerged as the most popular way to control the recommendation process used in numerous projects [17, 20, 21, 22, 23, 24, 25]. Arguably, the second most popular approach was tuning the fusion of several recommendation sources or algoritms in such systems as TasteWeights [26, 20], SetFusion [18], IntersectionExplorer [27], and RelevanceTuner [28]. In most (but not all) cases, the system developers offered user interrace sliders for the tuning of profile and fusion parameters. Slider-based user control was found to be intuitive and efficient in most of the cited studies.

More recently, Wang et al. [29] addressed filter bubbles by allowing users to control recommendation diversity, while He et al. [22] highlighted the importance of balancing control and cognitive load for optimal user satisfaction.

Despite this evidence, most modern RecSys remain black-box models, offering little transparency or direct control to users. This issue is particularly pronounced in the domain of VA RecSys, where DL-based approaches dominate but lack mechanisms for user-driven adjustments. In the following, we review VA RecSys literature and the gap our work addresses.

### 2.2. Visual Art Recommendation

Visual art recommendation systems (VA RecSys) aim to provide personalized suggestions to users exploring art, leveraging computational techniques to match user preferences with artworks. These systems are particularly relevant given the vast and diverse collections available online, which can overwhelm users without effective curation. Recent advances in AI-based RecSys have driven significant progress in this domain, though challenges related to transparency and user agency remain.

SOTA VA RecSys predominantly rely on AI-driven techniques, especially DL models, for effective personalization. Collaborative filtering (CF) and content-based filtering (CBF) methods remain foundational but are increasingly augmented by advanced neural architectures. For example, Convolutional Neural Networks (CNNs) have been widely adopted to extract visual features from artworks [30, 31, 32, 33].

Unimodal VA RecSys engines focus exclusively on visual or textual features of artworks for recommendation by analyzing low-level attributes such as color, texture, and shape [34, 32, 35], as well as high-level features like style and genre [36]. More recently multimodal approaches in VA RecSys started to incorporate multiple data modalities, such as visual, textual, and contextual information, to enhance recommendation quality. For example, models that integrate visual features with metadata (e.g., artist, period, or medium) and user-generated content (e.g., reviews or tags) have demonstrated improved personalization [2]. Multimodal approaches leverage architectures such as multimodal transformers and cross-modal embeddings [37, 12], which align heterogeneous data into a shared representation space. While these methods address the limitations of unimodal systems, they inherit the "black box" nature of DL models, offering limited transparency and user control.

Although such approaches have significantly advanced the SOTA in VA RecSys, their lack of user controllability remains a critical limitation. Current systems primarily optimize for algorithmic accuracy, often overlooking the importance of user agency in shaping recommendations. This calls for strategies that enable users to interact with and influence the recommendation process. In this paper, we propose a novel approach to integrate user control into the SOTA black-box VA RecSys approach; BLIP [2]. The following section presents our proposed VA RecSys algorithms designed to empower users in navigating and refining their recommendations.

## 3. User-controllable Recommendations

### 3.1. Black Box Recommendations

Consider a collection of paintings $\mathscr{P} = p_1, p_2, \dots, p_m$, each represented by an embedding (latent feature vector) $\mathscr{P}^* = \boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_m$ generated by our Black Box model. For a given user $u$, let $\mathscr{P}^u = p_1^u, p_2^u, \dots, p_n^u$ denote the subset of paintings that the user has rated, with $\mathscr{P}^u \subset \mathscr{P}$. The user's ratings are normalized [1] to form a set of weights $\omega^u = \{\omega_1^u, \omega_2^u, \dots, \omega_n^u\}$, where each rating is scaled to the range $[0, 1]$ based on a 5-point rating scale. Once we obtain the embeddings for the entire dataset using our Black-Box model (employing BLIP), we compute the similarity matrix $\mathbf{A}$ for all paintings. The predicted score $S^u(p_i)$ for a painting $p_i$ for user $u$ is then calculated as the weighted average similarity between the paintings rated by the user and all other paintings in the dataset:

[1] In our study, users elicit preferences in a 1–5 point rating scale (higher is better), thus we transform those values into weights $\omega_i^u \in [0, 1]$ for every rated painting $p_i^u$.

$$S^u(p_i) = \frac{1}{n}\sum_{j=1}^{n}\left(\omega_j^u, \mathbf{A}_{ij}\right) \tag{1}$$

Here, $\mathbf{A}_{ij}$ represents the similarity score between the embeddings of paintings $p_i$ and $p_j$ in the similarity matrix. The summation in Equation 1 spans all the paintings rated by the user, where $n = |\mathscr{P}^u|$. After computing the predicted scores for all paintings, we sort them and generate a ranked recommendation list containing the $r$ most similar paintings. The goal of the VA RecSys is to recommend paintings that are most similar to the set of paintings previously rated by the user, based on their elicited preferences. Given the set of paintings $\mathscr{P}$ and a user $u$, the Black Box engines BLIP select the most relevant set $\mathscr{R}$ to recommend, aiming to maximize the following objective:

$$\text{Policy 1: } \arg\max_{\mathscr{R}} \sum_{i=1}^{|\mathscr{R}|} S^u(p_i) \tag{2}$$

This policy ensures that the recommended set $\mathscr{R}$ consists of paintings that achieve the highest predicted scores, thereby closely aligning with the user's profile (i.e., ratings).

## 3.2. Controlling Popularity

In addition to personal preferences, users exhibit varying degrees of interest in the inclusion of popular and well-known content within their recommendations. The desire for exposure to celebrated artworks, such as famous paintings, can significantly differ among users. However, traditional black-box recommendation engines may not adequately account for this nuanced preference. Therefore, it is essential to empower users to actively manage the degree to which popular content is featured in their recommendations. To address this, we introduce a popularity score, denoted by $S^{\text{pop}}(p_i)$, for each painting in our dataset. This score is derived from artwork rankings of popular artists within the SemART dataset, based on public reviews reflecting the collective opinion on notable artworks. This leads to our second policy, which focuses on maximizing the inclusion of popular paintings:

$$\text{Policy 2: } \arg\max_{\mathscr{R}} \sum_{i=1}^{|\mathscr{R}|} S^{\text{pop}}(p_i) \tag{3}$$

To integrate both the user's individual preferences and the popularity aspect of paintings, we develop new recommendation engine BLIP-PoP. This engine maximizes an aggregated preference score $S_+^u(p_i)$ for each painting in the collection by combining the user's personalized scores with the popularity scores, as described by the following equation:

$$\arg\max_{\mathscr{R}} \sum_{i=1}^{|\mathscr{R}|} S_+^u(p_i) = \sum_{i=1}^{|\mathscr{R}|} S^u(p_i) + \beta S^{\text{pop}}(p_i) \tag{4}$$

Here, $\beta$ is a user-provided hyperparameter that controls the level of influence that popularity has on the recommendations. This parameter allows users to adjust their tolerance or preference for popular items, thereby providing a customizable balance between personalized and popular content.

## 3.3. Controlling Diversity and Consistency of Recommendations

There are different semantic categorizations of artworks and in different contexts, users have different tendencies to maintain the consistency or maximize the diversity of content in their recommendation sets. Yet "black box" engines do not allow users to regulate such aspects. For example, the SemArt painting collection [38] features ten different genres, such as religious, landscape or portrait. Each genre constitutes a certain number of paintings from the collection. Thus, if a user wants to increase the number of genres in the recommendation set; i.e., the recommended set contains paintings that are representative or fairly selected from the genre groups. Thus, we define a representative story selection strategy, adopted from Mehrotra et al. [39]. A recommendation set $\mathscr{R}$ is fairly representative if it contains paintings that belong to different story groups in a balanced way, so a representative story selection function $\psi(\mathscr{R})$ is given by:

$$\psi(\mathscr{R}) = \sum_{a=1}^{|K|} \sqrt{\sum_{p_i \in \mathcal{S}_a \cap \mathscr{R}} \gamma_{p_i}} \tag{5}$$

where $K$ is the total number of genre groups, $\mathcal{S}_a$ is the $a^{th}$ genre and $\gamma_{p_i}$ is a count for every painting $p_i$ selected from a genre $i$. The function $\psi(\mathscr{R})$ rewards recommendation sets that are diverse in terms of the different genre groups covered. It can be noticed that the representativeness score for a set that contains paintings belonging to only one or a few of the genre groups is lower than a set that covers all or most of the genre groups, owing to the square root function. For example assuming three genre groups ($K = 3$) a recommendation set $\mathscr{R}$ that chooses 2 paintings from $\mathcal{S}_1$ and 1 painting from each of $\mathcal{S}_2$ and $\mathcal{S}_3$ gets a higher $\psi(\mathscr{R})$ score as compared to a recommendation set that chooses 4 paintings from just $\mathcal{S}_1$, since $\sqrt{2} + \sqrt{1} + \sqrt{1} > \sqrt{4} + \sqrt{0} + \sqrt{0})$. Thus, the goal of finding a recommendation set that features representative paintings from each of the genre groups maximizes Equation 5:

$$\arg\max \ \psi(\mathscr{R}) \tag{6}$$

Note that maximizing over $\psi(\mathscr{R})$ ensures diversity of the recommendations in terms of the genres to be covered. However, minimising over $\psi(\mathscr{R})$ ensures consistency of the recommendations. To combine this aspect of controllability with the black box engines BLIP we create a VA RecSys engine BLIP-Diverse. This engine tries to jointly optimize for black box recommendations (i.e, Equation 2) and representative recommendations (Equation 5) by solving the following Mixed Integer Programming (MIP) problem:

$$\mathrm{MIP}_1 = \arg\max \left( (1-\xi) \sum_{i=1}^{|R|} p_i \, S^u(p_i) + \xi \psi(\mathscr{R}) \right) \tag{7}$$

where $\xi$ is a user-provided hyperparameter, indicating their tolerance to receiving diverse painting recommendations.

### 3.4. Jointly Controlling Popularity and Diversity

To jointly control the aspects of popularity, and consistency/ diversity while benefiting from personalized recommendations of the black box engine we create a recommendation strategy using the black box engine BLIP as a preference scoring model solving the following MIP problem:

$$\mathrm{MIP}_2 = \arg\max \left( (1-\xi) \sum_{i=1}^{|\mathscr{R}|} p_i \, S^u_+(p_i) + \xi \psi(\mathscr{R}) \right) \tag{8}$$

where $S^u_+(p_i)$ combines Policy 1 and 2 and $\xi$ indicates the user's tolerance to receiving diverse painting recommendations.

### 3.5. Multidimensional Diversity

In the above, diversity was primarily defined in terms of the spread across genres, to recommend a balanced set of paintings that reflect different artistic genres. This approach, while effective in capturing a single dimension of diversity, does not fully address the multifaceted nature of diversity in VA. Diversity is inherently multidimensional, encompassing various aspects like, artist, country, style, material, and more. Hence, considering multidimensional aspects of diversity is crucial in VA RecSys where the richness and variety of the content play a significant role in user satisfaction. A user might not only be interested in exploring different genres but also in discovering artworks from different time periods, by various artists, or from distinct cultural backgrounds. For example, a recommendation set that only varies by genre but includes paintings from a narrow time period or a single artist might fail to engage users who seek a broader exploration. Therefore, it is essential to create user-controllable knobs that allow for the adjustment of these multiple dimensions of diversity. Hence, to capture the essence of multidimensional diversity, we extend our previous diversity function to consider multiple dimensions simultaneously.

Consider that each painting $p_i$ can belong to multiple categories across different diversity aspects. Let: $\mathscr{G} = \{g_1, g_2, \dots, g_G\}$ denote the different diversity aspects (e.g., genre, time period, artist, country). For each aspect $g_k \in \mathscr{G}$, there are specific categories or classes $\mathscr{C}_k = \{c_1^k, c_2^k, \dots, c_{N_k}^k\}$; where $N_k$ is the number of categories for aspect $g_k$ (e.g., genre categories like portrait, landscape; time periods like $19^{th}$ century, $20^{th}$ century). Each painting $p_i$ belongs to a specific category $c_j^k$ within each aspect $g_k$.

To account for diversity across all aspects, we generalize the representative story selection function in Equation 5 as follows:

$$\psi'(\mathscr{R}) = \sum_{k=1}^{|\mathscr{G}|} \lambda_k \sqrt{\sum_{j=1}^{N_k} \sum_{p_i \in \mathscr{R} \cap \mathscr{C}_j^k} \gamma_{p_i}} \tag{9}$$

Here:

- $\lambda_k$ is a weight for each diversity aspect $g_k$, allowing control over the importance of each diversity dimension (e.g., genre might be more important than time period for some users).
- $\mathscr{C}_j^k$ represents the paintings in category $j$ of aspect $g_k$.
- $\gamma_{p_i}$ is the count or a relevance score for painting $p_i$ within its respective category.

To combine personalized recommendations with multidimensional diversity, the objective function becomes:

$$\mathrm{MIP}_1^{\mathrm{MultiD}} = \arg\max \left( (1-\xi) \sum_{i=1}^{|\mathscr{R}|} p_i\, S^u(p_i) + \xi \psi'(\mathscr{R}) \right) \tag{10}$$

The objective function which combines Popularity, Personalization, and Multi-Dimensional Diversity becomes:

$$\mathrm{MIP}_2^{\mathrm{MultiD}} = \arg\max \left( (1-\xi) \sum_{i=1}^{|\mathscr{R}|} p_i\, S_+^u(p_i) + \xi \psi'(\mathscr{R}) \right) \tag{11}$$

This formulation allows for the adjustment of diversity across multiple dimensions in addition to the balance between personalization and popularity.

### 3.6. Implementation

We developed three VA RecSys strategies in addition to BLIP, which is considered a baseline black box engine. BLIP offers users no control, as it generates purely personalized recommendations solely based on initial ratings. Following this we introduce a first level controlling mechanism on the baseline engine using the BLIP-PoP to regulate the degree to which popular content is featured in the recommendation set and BLIP-Diverse to tune the diversity of the recommendation set through the hyperparameters $\beta$ and $\xi$ respectively. Finally, we introduce a more advanced strategy which allows controlling specific diversity aspects of the recommendation set such as Genre, Country, Technique, Time period, etc.

By using the proposed VA RecSys strategies, we developed the ArtEx platform, which includes three interfaces corresponding to the levels of control mechanisms described in the following section.

# 4. User Study

In order to assess the proposed approach, we conducted a large-scale user study approved by the Institutional Review Board of [Redacted] presented below. The dataset used in the ArtEx platform is SemArt [38], which comprises 21,384 paintings collected from the Web Gallery of Art (WGA), a repository with over 44,809 images of European fine-art reproductions between the $8^{th}$ and the $19^{th}$ century.[2] Each painting image is accompanied by text-based metadata and artistic comments, which makes it suitable for the multimodal representation learning with BLIP similar to [2]. These paintings are organized into 10 semantic categories: religious, landscape, portrait, mythological, genre, interior, still life, historical, study, and other.

The system has three recommendation interfaces for each experiment condition: *No-Slider*, *2-Sliders*, and *6-Sliders*, illustrated in Figure 2. In the *No-Slider* condition (i.e., baseline), users can influence their recommendations only by providing ratings. New recommendations are generated by clicking the "Get New Recommendation" button. The "Popularity and Diversity Slider" (*2-Sliders*) condition introduces simple sliders that enable users to adjust the balance between popular and diverse paintings. By solving $MIP_2$ (Equation 8) the recommendations automatically refresh each time the sliders are adjusted. Lastly, the "Popularity and Detailed Diversity Slider" (*6-Sliders*) condition introduces additional sliders for controlling specific diversity factors (i.e., Technique, Genre, Country, and Time Period.) For example, users can increase the diversity of genres while keeping the technique consistent or explore artworks from various countries within a particular time period using the different slides as shown in Figure 2.

Figure 1 shows a sample set of 9 paintings (bottom) and a radial chart visualizing their distribution across the diversity aspects: TECHNIQUE, GENRE, COUNTRY, and TIMEFRAME. Each aspect is represented by a different colored region on the chart. The radial axes represent specific categories within each aspect (e.g., "15th century," "17th century" for TIMEFRAME). The numbers on the rings represent the number of items in each specific category from the recommendation set and counts in each category are plotted as points on the axes. The colored areas connect these points, forming distinct shapes that highlight the relative importance or distribution of values across the four diversity aspects in the recommendation set. Notably, although there are 9 items, a single painting can belong to multiple categories across different diversity aspects.

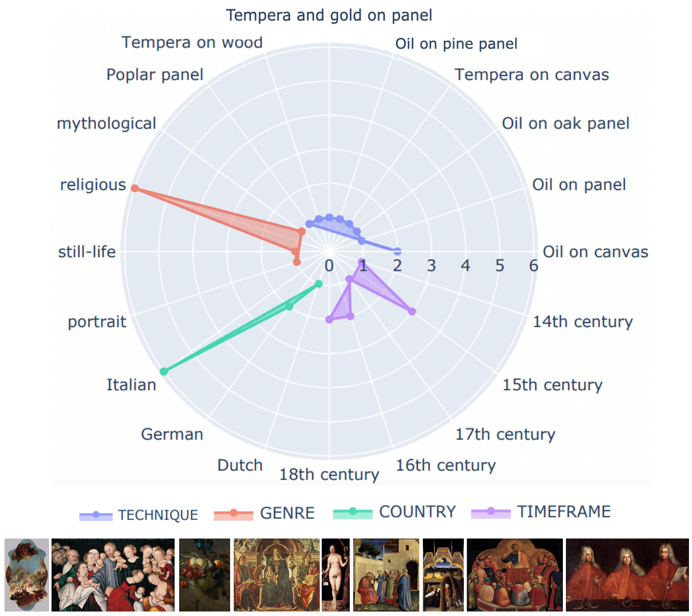


**Figure 1:** Diversity aspects within a recommendation set.

[2] https://www.wga.hu/

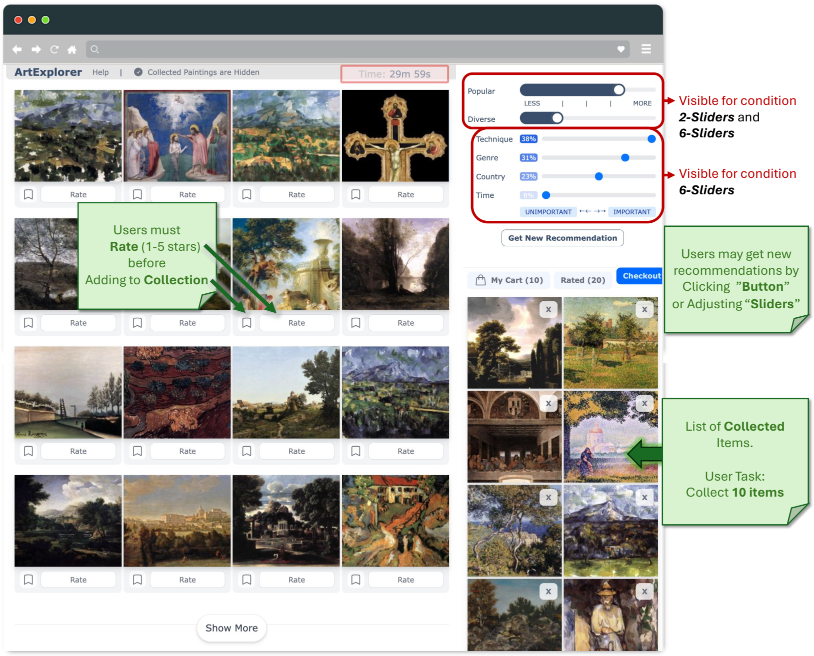


**Figure 2:** ArtEx System Components.

## 4.1. Participants

We recruited a pool of 151 participants from Amazon Mechanical Turk (MTurk)[3] with the following primary screening criteria; a HIT Approval Rate (%) greater than 98% for all tasks they have completed on MTurk; Number of HITs Approved greater than 1,000; Our recruited participants were 52.98% male and 47.02% female, with the majority residing in the USA (92.72%). Participants age is between 23 and 66 years (M=32.87, SD=8.40). The normalized mean of interest in art among participants was high (M=0.88, SD=0.18), between "Moderately interested" and "Extremely interested." They were higher than "Moderately knowledgeable" about art (M=0.75, SD=0.23). The frequency of museum visits was relatively high (M=0.79, SD=0.26), corresponding to "About once per year" or "Once every several years." Participants indicated a strong preference for visiting popular (M=0.84, SD=0.19) and diverse art pieces (M=0.843, SD=0.18).

## 4.2. Procedure

On MTurk, we posted recruitment ads outlining task instructions and participation rules. MTurk workers could only participate once and were required to complete the 15-minute study on a desktop PC within 4 hours upon accepting the task. Since the SemArt dataset may include adult content, participants must confirm their willingness to view such material. Eligibility was determined through a brief 4-question multiple-choice quiz. Participants who successfully passed the initial screening were invited to register on the ArtEx platform. Upon registration, they were presented with an introduction outlining the study's objectives and the steps involved. Participants were then asked to imagine themselves visiting a museum and subsequently being offered an opportunity to select free print reproductions of paintings from the museum's store as a gesture of appreciation for their visit. Their task was to browse through a set of recommended paintings and curate a collection of ten prints that best aligned with their personal preferences. The registration form also collected demographic details such as gender, age, and art-related interests, knowledge, and preferences. They also rated their tolerance towards receiving popular and diverse paintings on a 5-point Likert scale. Finally, they had to agree to the terms and conditions before proceeding as participants.

[3]https://www.mturk.com/

The next step is preference elicitation, where participants rated 10 randomly selected paintings, one from each of the 10 semantic categories in the SemArt dataset, using a 5-star Likert scale. Following this, participants were provided with a brief tutorial on how to use the platform, tailored to their assigned condition (*No-Slider*, *2-Sliders*, or *6-Sliders*). Once familiar with the system, they received personalized recommendations and could explore the platform to identify the 10 paintings that best align with their preferences. During their interaction, participants could also rate additional paintings from the recommendation dashboard to update their preference profile, and click the "Get New Recommendation" button or use the different sliders to receive new recommendations. After finalizing their initial selection, participants had the opportunity to review their collection, add, remove/replace paintings and adjust their ratings before completing the process and submitting their final collection.

In the final step, participants were asked to evaluate the system with a post-study questionnaire. It covered various aspects, including beyond-relevance metrics, usability, satisfaction, initial preferences, and the effectiveness of sliders in adjusting recommendations. Beyond-relevance metrics assessed how well the recommendations matched participant interests, including diversity, novelty, and serendipity. Usability focused on the ease of using the interface and whether participants lost track of time while exploring. Satisfaction metrics included overall satisfaction, willingness to use the system again, and recommending it to colleagues.

For slider conditions, additional aspects were explored, such as the sense of control of the participants, the ability to adjust preferences effectively, and whether the sliders contributed to their satisfaction. Other factors like trust in the sliders, ease of use, and how fun they made the task were also assessed. The study followed a between-subjects design in which a total of 151 participants were randomly assigned to one of the three conditions. After postprocessing the collected data, the sample size is slightly varying between conditions: n=39 in the *No-Slider*, n=52 *2-Sliders*, and n=60 in the *6-Sliders* condition.

## 5. Results

### 5.1. User Interaction with ArtEx

The interactions of participants under the three different conditions are summarized in Table 1. As shown in Figure 2, the system displayed thumbnails of recommended paintings as a 4x4 grid of images per page. "Next Page" refers to when participants clicked the "More" button to view the next 4x4 grid of recommendations. "Lookup" is access to painting details by clicking on a thumbnail and opening the images in a separate window in high resolution with full metadata. "Rate" is adding or changing a rating of a painting, while "Unrate" refers to removing a rating (for a rated item). "Collect" means adding a painting to the personal collection by using a bookmark icon below each painting. "Remove" means removing a painting from the collection, usually to free space for adding a more appealing painting. "Pop Slider" and "Div Sliders" reports the number of times a user changed the position of the popularity slider or any of the diversity sliders (sliders were available in the baseline condition and only one diversity slider was available in 2-slider condition). The table shows no striking differences between the average numbers of user actions in each category across conditions. Kruskal–Wallis found no significant differences between study conditions for any of the tested metrics. Specifically, we found $\chi^2(2) = 2.127$, $p = 0.3452$ for *Next Page*, $\chi^2(2) = 2.269$, $p = 0.3215$ for *Lookup*, $\chi^2(2) = 0.476$, $p = 0.7880$ for *Rate*, $\chi^2(2) = 2.800$, $p = 0.2465$ for *Unrate*, and $\chi^2(2) = 0.511$, $p = 0.774$ for *Collect*.

The lack of significant differences between conditions suggests that the number of actions varied considerably between users. Figure 3 reveals difference between users by showing the distribution of total user actions with the system. As we see, about a third of the users made between 41 and 70 total actions, yet many users made 2-3 times as many actions and 19 users made over 200. The table also shows that 15 users made 40 or fewer actions. Given that 20 actions (10 Rate + 10 Collect) were an absolute minimum, we concluded that these users have not invested expected efforts to achieve the required goal (not unusual in crowd-sourcing platforms). These users were excluded from follow-up statistical analyzes.

**Table 1**
User actions by study conditions

| Action | No-Slider (n=39) | | 2-Sliders (n=52) | | 6-Sliders (n=60) | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Next Page | 5.90 | 5.69 | 8.35 | 8.27 | 9.20 | 11.72 |
| Lookup | 11.64 | 14.47 | 8.52 | 11.28 | 11.10 | 20.22 |
| Rate | 64.13 | 60.19 | 60.15 | 51.76 | 78.47 | 69.73 |
| Unrate | 0.74 | 2.61 | 0.27 | 0.60 | 1.92 | 6.28 |
| Collect | 16.00 | 11.01 | 14.90 | 8.34 | 17.60 | 18.05 |
| Remove | 6.00 | 11.01 | 4.90 | 8.34 | 7.60 | 18.05 |
| Pop Slider | - | - | 2.37 | 3.68 | 1.07 | 1.83 |
| Div Sliders | - | - | 1.88 | 2.48 | 5.17 | 7.21 |
| Total Actions | 97.66 | 59.99 | 96.17 | 51.53 | 122.6 | 77.38 |

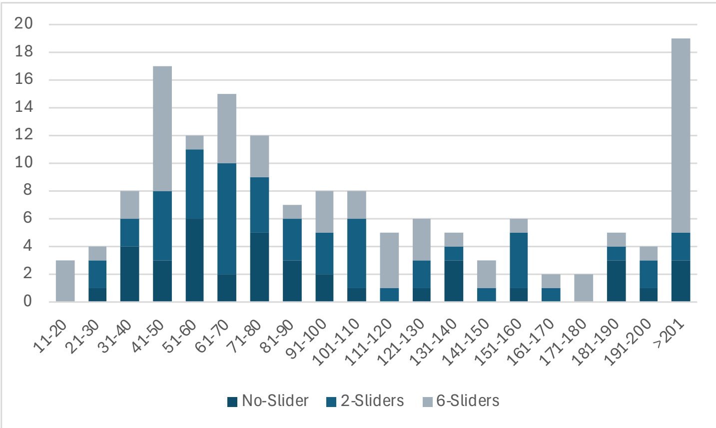


**Figure 3:** Distribution of Participants by Total Actions

The analysis of standard deviation data also shows a considerable difference between users in using each type of actions pointing that not just the total amount, but the distribution of effort was different. The standard deviation data also hint that in the 6-slider condition the difference between users is much larger for most types of activity than in the baseline conditions. Indeed, between-user differences in using interface features are frequently observed in exploratory RecSys, where users are offered new functionalities that they are free to explore or ignore. As the data of several similar studies show [40, 23], in these systems users embrace these functionalities to a different extent, which affects their use of the systems, as well as success and satisfaction.

To explore this trend in ArtEx, we divided 112 users of both slider conditions into three comparable cohorts that differ by the extent to which the cohort users embraced the main additional functionality offered in these conditions, i.e., the use of sliders. The users of the *Low* cohort (n=34) used sliders at most once or not at all. The users of *Medium* cohort (n=42) used sliders between 2 and 5 times. The users of *High* cohort (n=36) use sliders more than 5 times. The usage data in each of these cohort are shown in Table 2. This table reveals several notable differences between the cohorts. Most importantly, the average number of rating actions drops rapidly with the increased users of sliders - from 97.12 in Low cohort to 64.21 in Medium cohort to 51.03 in High cohort. It hints that users who actively used sliders worked more efficiently achieving the same target by rating much fewer items that users who have none to medium use of sliders.

To examine this assumption formally, we merged all users with none to low use of sliders (No-Slider condition + Low cohort) in one group and contrasted it with Medium and High cohorts (Table 3). We excluded 15 participants with 40 or less total actions revealed in Figure 3 (5 participants from *No-Slider*,

**Table 2**
User actions in slider conditions by usage cohorts

| Action | Low (n=34) | | Medium (n=42) | | High (n=36) | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Next Page | 8.32 | 8.44 | 8.83 | 12.43 | 9.22 | 9.13 |
| Lookup | 5.74 | 8.91 | 11.88 | 20.85 | 11.53 | 16.48 |
| Rate | **97.12** | 79.17 | **64.21** | 56.18 | **51.03** | 40.45 |
| Unrate | 0.76 | 1.79 | 0.64 | 2.00 | 2.11 | 7.75 |
| Collect | 18.50 | 20.41 | 16.14 | 13.00 | 14.56 | 7.62 |
| Remove | 8.50 | 20.41 | 6.14 | 13.00 | 4.56 | 7.62 |
| Pop Slider | 0.15 | 0.36 | 1.26 | 0.99 | 3.58 | 4.38 |
| Div Sliders | 0.09 | 0.29 | 1.74 | 1.27 | 9.22 | 7.39 |
| Total Actions | 129.91 | 89.95 | 104.07 | 61.16 | 99.14 | 44.25 |

4 from *2-Sliders*, and 6 from *6-Sliders* conditions). Kruskal–Wallis test with tie correction confirmed significant differences in the number of rated items between these groups $\chi^2(2)Rate = 7.51, p = 0.023$. No differences for other types of actions were significant: $\chi^2(2) = 0.07, p = 0.967$ for *Next Page*, $\chi^2(2) = 0.38, p = 0.828$ for *Lookup*, $\chi^2(2) = 0.08, p = 0.961$ for *Unrate*, and $\chi^2(2) = 0.19, p = 0.910$ for *Collect*.

**Table 3**
Action data by aggregate use of sliders, excluding 15 low-interaction participants

| Action | No + Low (n=62) | | Medium (n=39) | | High (n=36) | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Next Page | 8.02 | 7.21 | 9.46 | 12.68 | 9.22 | 9.13 |
| Lookup | 9.85 | 13.19 | 12.33 | 21.55 | 11.53 | 16.48 |
| Rate | 91.32 | 70.90 | 68.62 | 55.90 | 51.03 | 40.45 |
| Unrate | 0.89 | 2.42 | 0.69 | 2.07 | 2.11 | 7.75 |
| Collect | 18.16 | 17.18 | 16.59 | 13.40 | 14.56 | 7.62 |
| Remove | 8.16 | 17.18 | 6.59 | 13.40 | 4.56 | 7.62 |
| Pop Slider | 0.05 | 0.22 | 1.33 | 0.98 | 3.58 | 4.38 |
| Div Sliders | 0.03 | 0.18 | 1.64 | 1.25 | 9.22 | 7.39 |
| Total Actions | 127.44 | 73.86 | 109.97 | 59.40 | 99.14 | 44.25 |

## 5.2. The Profile of Rating

As the previous section shows, users in a slider condition who choose to use sliders extensively were most efficient in their search for good paintings. They achieved their final goal with fewer actions on average and significantly fewer rating actions. In this section, we attempt to understand how and why the users' choice to use or not to use sliders affected the efficiency of their "hunt" for paintings by examining the rating profiles of users with different levels of slider use. For this analysis, we exclude users from the no-slider condition, since we want to see the impact of choice, and the users in the no-slider condition were not offered a choice. The three groups examined in this section are the Low, Medium, and High slider usage cohorts shown in Table 2. Before starting this analysis, we repeat the Kruskal-Wallis test for these three cohorts to confirm whether the difference between the number of ratings is significant between these cohorts as well. The analysis confirmed a significant difference in the number of ratings between the slider usage cohorts $\chi^2(2) = 13.25, p = 0.001$, highlighting that an increase of slider usage leads to a significant decrease of in the number of ratings required to achieve the goal.

Next we we analyzed rating distributions across these cohorts (Figure 4). The data shows that "additional" ratings that users in lower usage cohorts had to make to achieve their goal are not evenly distributed. The number of items rated with 5 stars was comparable, even a bit lower for participants with low and medium use of sliders. This provides evidence that at the end of their work users of all cohorts were able to achieve their exploration goal. Given that their target was to find 10 appealing paintings, the ability to discover 14 to 17 5-star paintings hints that at the end of their work, users in all cohorts had enough good painting for their collections. However, on the way to this goal, users in the low and medium usage cohorts had to rate approximately twice as many "suboptimal" 4-star items and 2-3 times as many "mediocre" and "poor" 2-3-star items than users who use sliders extensively.

The rating distribution analysis uncovers one possible reason for the higher efficiency of active slider users. At the beginning of our study, all participants found themselves in a *new-user* situatuion. Their initial user profiles were built using a small number of ratings, likely not representing their interests reliably. As a result, items placed by ArtEx on the top of the ranked list at the start of exploration might not be their best choices yet. Some truly appealing paintings could be located deep in the ranked well since users had no chance to express their interest in these kinds of paintings. In a traditional interaction with a RecSys, the path to better recommendations is investment in rating. High ratings to relevant items will help similar items to float up, while low ratings for less relevant and irrelevant items attempt to "sink" poor items in a hope to replace them with more relevant ones. As a prevalence of low ratings shows, users who chose not to use sliders had to do a lot of "sinking", making their "hunt" for good items inefficient. The ability to use sliders offered another opportunity in ArtEx: by manipulating diversity and popularity of recommendations, active slider users had a chance to bring appealing items from the depth of the ranking well to the surface where they can discover and highly rate them, thereby improving their profile. As the data show, using sliders was much more efficient than "sinking" poor items, enabling active slider users to achieve their goal with significantly fewer ratings and a much higher ratio of 5-star ratings.

To compare the differences between cohorts numerically, we calculated the 5-star ratio for each user as a fraction of 5-star ratings among the total number of ratings made by users. The users of High cohort has the highest 5-star-ratio (31.7%) followed by Medium cohort (27%) and the lowest for the low cohort (20.6%). The Kruskal-Wallis test made after exclusion of 10 participants who made less than 40 actions indicated marginally significant differences between slider groups ($\chi^2(2) = 5.186$, $p = 0.074$), suggesting a potential relationship. To clarify this trend, the Spearman correlation test indicated a small but significant positive correlation ($\rho = 0.218$, $p = 0.028$). This confirms that participants who choose not to use sliders actively have to pay a significantly larger rating "toll", i.e., rating suboptimal and poor items to achieve the same goal. cal

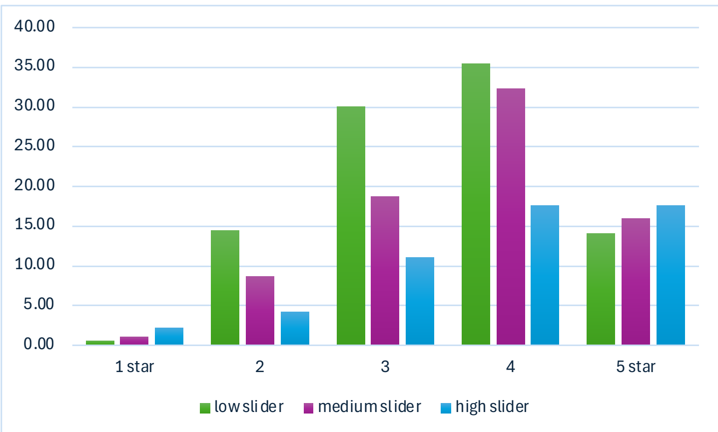


**Figure 4:** Average rating distributions for usage cohorts

## 6. Conclusion and Future Work

We have explored a novel algorithmic approach that gives users control over their VA recommendations via tunable hyperparameters, contrary to the opaque nature of SOTA VA RecSys. We developed the ArtEx platform, which leverages our proposed algorithms to provide real-time control over the diversity and popularity of recommended paintings. Our findings demonstrate that these interactive controls encourage exploration and lead to more thoughtful rating behaviors. We also observed that users provided with the most granular control sliders (*6-Sliders*) had the highest level of interaction and discovery potential. This highlights a promising direction, shifting from black-box algorithmic suggestions to user-driven exploration in VA space. By integrating direct user control over ranking factors, our approach not only improves personalization for more human-centered AI in artistic exploration. Building on our findings, an interesting future research avenue could be refining adaptive slider mechanisms for dynamic preference response, exploring hybrid explicit-implicit control models, and evaluating long-term effects on user trust and satisfaction.

## Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

## References


[1] T. Di Noia, N. Tintarev, P. Fatourou, M. Schedl, Recommender systems under european ai regulations, Communications of the ACM 65 (2022) 69–73.

[2] B. A. Yilma, L. A. Leiva, Together yet apart: Multimodal representation learning for personalised visual art recommendation, in: Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, UMAP '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 204–214.

[3] B. A. Yilma, L. A. Leiva, The elements of visual art recommendation: Learning latent semantic representations of paintings, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany, ACM, 2023.

[4] T. Yemelianenko, I. Tkachenko, T. Masclef, M. Scuturici, S. Miguet, Artwork recommendations guided by foundation models: survey and novel approach, Multimedia Tools and Applications (2025) 1–22.

[5] A. Haque, N. Islam, P. Mikalef, To explain or not to explain: An empirical investigation of AI-based recommendations on social media platforms, Electronic Markets 35 (2025) 1–18.

[6] R. Ramachandranpillai, R. Baeza-Yates, F. Heintz, Fairxai-a taxonomy and framework for fairness and explainability synergy in machine learning, IEEE Transactions on Neural Networks and Learning Systems (2025).

[7] I. Adaji, J. Vassileva, Evaluating personalization and persuasion in e-commerce., PPT@ PERSUASIVE 1582 (2016) 107–113.

[8] P. Briggs, B. Simpson, A. De Angeli, Personalisation and trust: a reciprocal relationship?, Designing Personalized user experiences in eCommerce (2004) 39–55.

[9] G. Einav, O. Allen, T. Gur, Y. Maaravi, D. Ravner, Bursting filter bubbles in a digital age: Opening minds and reducing opinion polarization through digital platforms, Technology in Society 71 (2022) 102136.

[10] P. Liu, K. Shivaram, A. Culotta, M. A. Shapiro, M. Bilgic, The interaction between political typology and filter bubbles in news recommendation algorithms, in: Proceedings of the Web Conference 2021, 2021, pp. 3791–3801.

[11] R. He, C. Fang, Z. Wang, J. McAuley, Vista: a visually, socially, and temporally-aware model for artistic recommendation, in: Proceedings of the 10th ACM Conference on Recommender Systems, 2016, pp. 309–316.

[12] J. Li, D. Li, C. Xiong, S. Hoi, Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation, arXiv preprint arXiv:2201.12086 (2022).
[13] K. Verbert, D. Parra, P. Brusilovsky, E. Duval, Visualizing recommendations to support exploration, transparency and controllability, in: Proceedings of the 2013 international conference on Intelligent user interfaces, 2013, pp. 351–362.
[14] J. O'Donovan, B. Smyth, B. Gretarsson, S. Bostandjiev, T. Höllerer, PeerChooser: visual interactive recommendation, in: the twenty-sixth annual SIGCHI conference on Human factors in computing systems, ACM, 2008, pp. 1085–1088.
[15] B. P. Knijnenburg, N. J. Reijmer, M. C. Willemsen, Each to his own: how different users call for different interaction methods in recommender systems, in: Proceedings of the fifth ACM conference on Recommender systems, 2011, pp. 141–148.
[16] Y. Hijikata, Y. Kai, S. Nishida, The relation between user intervention and user satisfaction for information recommendation, in: Proceedings of the 27th Annual ACM Symposium on Applied Computing, 2012, pp. 2002–2007.
[17] B. P. Knijnenburg, S. Bostandjiev, J. O'Donovan, A. Kobsa, Inspectability and control in social recommenders, in: 6th ACM Conference on Recommender System, 2012, pp. 43–50.
[18] D. Parra, P. Brusilovsky, User-controllable personalization: A case study with setfusion, International Journal of Human-Computer Studies 78 (2015) 43–67.
[19] D. Jannach, S. Naveed, M. Jugovac, User control in recommender systems: Overview and interaction challenges, in: E-Commerce and Web Technologies: 17th International Conference, EC-Web 2016, Porto, Portugal, September 5-8, 2016, Revised Selected Papers 17, Springer, 2017, pp. 21–33.
[20] S. Bostandjiev, J. O'Donovan, T. Höllerer, Linkedvis: exploring social and semantic career recommendations, in: Proceedings of the 2013 international conference on Intelligent user interfaces, 2013, pp. 107–116.
[21] A. Kangasrääsiö, D. Głowacka, S. Kaski, Improving controllability and predictability of interactive recommendation interfaces for exploratory search, in: Proceedings of the 20th International Conference on Intelligent User Interfaces, ACM, 2015, pp. 247–251.
[22] Y. Jin, B. Cardoso, K. Verbert, How do different levels of user control affect cognitive load and acceptance of recommendations?, in: Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS'17) at 2017 ACM Conference on Recommender Systems (RecSys'17), volume 1884 of *CEUR Workshop Proceeding*, CEUR, 2017, pp. 35–42. URL: http://ceur-ws.org/Vol-1884/paper6.pdf.
[23] B. Rahdari, P. Brusilovsky, D. Babichenko, Personalizing information exploration with an open user model, in: 31st ACM Conference on Hypertext and Social Media, ACM, 2020, pp. 167–176.
[24] J. Kunkel, C. Schwenger, J. Ziegler, NewsViz: Depicting and controlling preference profiles using interactive treemaps in news recommender systems, in: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, ACM, 2020, pp. 126–135.
[25] R. Bansal, J. Olmstead, U. Bram, R. Cottrell, G. Reder, J. Altosaar, Recommending interesting writing using a controllable, explanation-aware visual interface, in: Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS'20) at 2020 ACM Conference on Recommender Systems (RecSys'20), volume 2682 of *CEUR Workshop Proceeding*, CEUR, 2020, pp. 77–80. URL: http://ceur-ws.org/Vol-2682/short2.pdf.
[26] S. Bostandjiev, J. O'Donovan, T. Höllerer, TasteWeights: A visual interactive hybrid recommender system, in: 6th ACM Conference on Recommender System, 2012, pp. 35–42.
[27] B. Cardoso, G. Sedrakyan, F. Gutiérrez, D. Parra, P. Brusilovsky, K. Verbert, IntersectionExplorer, a multi-perspective approach for exploring recommendations, International Journal of Human-Computer Studies 121 (2019) 73–92.
[28] C.-H. Tsai, P. Brusilovsky, The effects of controllability and explainability in a social recommender system, User Modeling and User-Adapted Interaction 31 (2021) 591–627.
[29] W. Wang, F. Feng, L. Nie, T.-S. Chua, User-controllable recommendation against filter bubbles, in: Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval, 2022, pp. 1251–1261.

[30] B. A. Yilma, N. Aghenda, M. Romero, Y. Naudet, H. Panetto, Personalised visual art recommendation by learning latent semantic representations, in: 2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA, IEEE, 2020, pp. 1–6.
[31] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[32] P. Messina, V. Dominguez, D. Parra, C. Trattner, A. Soto, Exploring content-based artwork recommendation with metadata and visual features, arXiv preprint arXiv:1706.05786 (2017).
[33] B. A. Yilma, Y. Naudet, H. Panetto, Personalisation in cyber-physical-social systems: A multi-stakeholder aware recommendation and guidance, in: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP'21, ACM, New York, NY, USA, 2021, p. 251–255.
[34] D. Parra, P. Messina, M. Cartagena, F. del Rio, P. Cerda-Mardini, Curatornet: Visually-aware recommendation of art images, in: Workshops on Recommendation in Complex Scenarios and the Impact of Recommender Systems co-located with 14th ACM Conference on Recommender Systems (RecSys 2020), volume 2697 of *CEUR Workshop Proceedings*, CEUR, 2020, pp. 37–44. URL: https://ceur-ws.org/Vol-2697/paper2_complexrec.pdf.
[35] P. Messina, V. Dominguez, D. Parra, C. Trattner, A. Soto, Content-based artwork recommendation: integrating painting metadata with neural and manually-engineered visual features, User Modeling and User-Adapted Interaction 29 (2019) 251–290.
[36] B. Saleh, A. Elgammal, Large-scale classification of fine-art paintings: Learning the right metric on the right feature, arXiv preprint arXiv:1505.00855 (2015).
[37] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., Learning transferable visual models from natural language supervision, in: International Conference on Machine Learning, PMLR, 2021, pp. 8748–8763.
[38] N. Garcia, G. Vogiatzis, How to read paintings: semantic art understanding with multi-modal retrieval, in: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018, pp. 0–0.
[39] R. Mehrotra, E. Yilmaz, Representative & informative query selection for learning to rank using submodular functions, in: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, 2015, pp. 545–554.
[40] M. Millecamp, N. N. Htun, Y. Jin, K. Verbert, Controlling spotify recommendations: Effects of personal characteristics on music recommender user interfaces, in: 26th Conference on User Modeling, Adaptation and Personalization, UMAP'18, ACM, 2018, pp. 101–109.