**UX/UI study – public users**

***Context:***

Museums are producing evermore digital content – especially since COVID-19 had sped up digital strategies and online being the only form of engagement during lockdowns - and offer audiences various forms of engagement beyond their physical walls. Online collections offer a wealth of information, enabling users to query and interact with an abundance of data. However, those systems are just as good and reliable as the data entered, are prone to errors and missing entries, and mostly limited to traditional keyword search, which leaves lay users struggling leading to high abandonment rates. Novel approaches, such as machine learning (ML) algorithms for the creation of art works and digital objects have gained popularity over the recent years, but systems enabling various museum constituents – from professionals to on- and offline visitors – have yet to be trialled and adopted on a big scale.

One of the applications that could facilitate interaction with museum collections are recommender systems. Those systems collect intrinsic and/or extrinsic user feedback to serve recommendations based on certain criteria, e.g., likes or ratings, low-level interactions (click through rates, dwell times, click events etc.), user profiles etc. and can therefore provide personalised pathways through data.

The MuseREC recommendation system is a web app-based recommender engine using museum collection data to serve recommendations based on object choices. The models used to serve recommendations are trained offline and based on metadata and image data vectors, with similarities calculated using the FAISS algorithm (Johnson, Douze, and Jégou, 2017). Recommendations are served based on user clicks on artworks.

***Rationale:***

The use of recommender engines to support systems used by on-site visitors has been the focus of scholarly research so far, their application to support interaction with online collections in museums however is yet to be investigated and properly understood. It is important to collect information about how users can be informed by, and also inform the development and usage of those applications to better comprehend their implications in relation to museum data.

***Hypothesis:***

Recommender engines can improve the user experience compared to random browsing and traditional search interfaces, and can be a novel tool to serve a diverse range of artworks.

***RQs:***

**RQ 1:** What is the difference in subjective recommendation quality between different models?

**RQ 2:** Does a system that provides recommendations enhance the user experience of browsing online collections compared to random suggestions?

**RQ 3:** How satisfied were users with recommendations?

***Expectations:***

Discover users’ preferences on interaction with a museum online collection recommender engine and gather results about the models’ performances in terms of user engagement. Collect interaction data that informs future recommender system developments.

***Research goals:***

* Better understand how users perceive the recommender system
* Focus on user-centric evaluations
* Explore the way social and cultural knowledge and meaning are produced around system
* Identify areas for further development and research
* Add context and professional insight to the system
* Gather feedback and user interaction/experience data
* Investigate ethical considerations
* Compare different user group requirements

**EXPERIMENTAL SETTING:**

**Introduction:**

To answer the above stated research questions, we have designed an experiment to gather quantitative and qualitative user data. The outcome of the study may provide evidence that different models of recommender engines are differently perceived in terms of user experience and engagement, and we further aim to investigate if some models are more suitable to serve meaningful suggestions to users of museum online collections compared to others.

Participants will take part in a controlled online study that investigates the user experience and interaction with a recommender system for museum online collections. They will not be made test aware, however they will be told they are participating in a study about novel methods of browsing museum online collections. Participants will be made aware of time constraints and a post-study questionnaire.

**Apparatus/stimuli:**

We developed a web app that resembles traditional museum online collection interfaces of medium to large sized institutions across the Western world.

We randomly selected 35,000 objects out of Art UK’s database containing 350,000 items altogether. Out of these 35,000 we draw a sample of 28 objects representative of the collections to be shown to participants. Surveying 10 online collection websites of renowned museums, ranging from 6 to 55, with some option to include up to 100 items on the first page, we decided that 28 would be a reasonable number of objects to be displayed.

**Participants**

Participants are recruited via the ArtUK audience panel, at the University of Manchester, and Twitter.

The ArtUK panel consists of around 400 people, who are purposively sampled, but they tend to be older and they are slightly more women signed up to it.

At the university we aim to recruit staff and students by posting flyers throughout buildings.

We further advertise the study on Twitter using the official IAM lab Twitter account.

Sample size: a minimum of 35 participants for any of the three recommender conditions = 105

**Conditions:**

There are 4 different conditions:

*Recommendations:*

* RM: recommendations based on metadata model
* RI: recommendations based on image model
* RMI: recommendations based on both, meta and image data

*Random:*

* R-: Random selection of objects

**Procedure:**

1. PIS and consent form

On landing on the experiment’s webpage users will be supplied with a PIS and are required to give informed consent before proceeding further. In case a user does not give consent, the process abandons and takes the user to the end page of the study.

Should the user agree to take part they may proceed to step 2.

Participants will not be made test aware, however, they will be informed about the nature of the study being held online, the data that we collect, and the approximate duration of the study.

We don’t expect demand effects, even if participants elicit the study’s purpose (Mummolo and Peterson, 2018; Berinsky, Huber, and Lenz, 2018)

1. Demographics and domain knowledge

Participants will be asked questions about their demographics and questions to assess their use patterns and domain knowledge.

*Demographics*

How old are you?

* 18-24 years
* 25-34 years
* 35-44 years
* 45-54 years
* 55-64 years
* 65+
* Prefer not to answer

What gender do you most identify as?

* Woman
* Man
* Non-binary
* Prefer to self-describe:
* Prefer not to say

What is your highest level of education?

* Primary school
* Secondary school up to 16 years
* Higher or secondary school or further education (A-level, BTEC, etc.)
* College or bachelor’s degree
* Master’s degree
* Doctorate
* Prefer not to answer

What is your ethnic group?

* White
* Mixed/multiple ethnic groups
* Asian
* Black/African/Caribbean
* Other:
* Prefer not to answer

Have you got a disability?

* Identify as disabled
* Do not identify as disabled
* Prefer not to answer

*Domain knowledge*

How would you rate your knowledge of art?

* Novice
* I have got some knowledge
* Knowledgeable
* Hobby-expert
* Expert

How often do you visit museum/art gallery websites?

* Most days
* At least once a week
* At least once a month
* At least every two to three months
* At least once a year
* This is my first time

How often do you specifically look at collections online?

* Most days
* At least once a week
* At least once a month
* At least every two to three months
* At least once a year
* This is my first time

How often do you physically visit a museum/art gallery?

* At least once every 6 months
* At least once a year
* Every two or three years
* Every four or five years
* I have not been to a museum/art gallery in 6 years or longer
* I’ve never visited a museum/art gallery in person
1. Study phase

After having finished points 1 and 2, participants will be presented with a website (webpage\_1) as outlined in point 3. Participants see a webpage with 28 images including the name of the artist and the title of artwork. More information can be obtained through clicking on the artwork image, which opens up a single view page containing the artist’s name, artwork title, year of creation and a ‘show me more’ option displaying further information, e.g., themes, topics, and notes.

The single view page can be closed and the subject returns to webpage\_1, where they can either inspect more elements or start to complete task 1 (task\_1), which prompts the user to select up 10 but at least 5 artworks that they want to see more objects of.

Participants always have the option to click on a “Leave this study”-button throughout the experiment, which will take them straight to the debrief page.

*Part I:*

After completing task\_1, participants can click proceed, which will take them to another webpage (webpage\_2) containing another set of 28 objects.

Webpage\_2 is populated by either one of the 3 recommender conditions (RM, RI, or RMI) or by a random sample (R-) without any recommendations. Participants can interact with the newly generated objects and get to choose another 5 to 10 of which they would like to see more. Part I will last 10 minutes. In these 10 minutes participants can repeat the task as many times as they want, resulting in n tasks (task\_n) and n webpages (webpage\_n) at the end of this part.

*Distraction task:*

Participants have to answer a question to be distracted from part I and to ensure that they’re attentive and not just rushing through the study.

Which of the animals is the lightest?

* Duck
* Mouse
* Elephant
* Bumblebee

*Part II:*

If participants were presented with one of the recommendations in part I, they’ll be now presented with a random selection of objects (R-), if they saw the random selection in part I they will be presented with one of the recommender conditions (RM, RI, or RMI). They then get ask to select objects again for another 10mins. After having completed the task, participants are prompted to answer the post-study questionnaire.

1. Post-study questionnaire

On completion of parts I and II participants will be asked to complete a post-study questionnaire, which uses adapted questions from the Knijnenburg et al. (2012) framework.

Answers are entered via 5-point Likert scales ranging from 1 (completely disagree) to 5 (completely agree).

* 1. Please think of **PART I** (the first half of the study) when answering the following questions:

*Perceived recommendation quality*

* + I liked the artworks shown by the system.
	+ The artworks fitted my preference.
	+ The artworks were well-chosen.
	+ The artworks were relevant.
	+ The system showed me too many bad artworks.
	+ I didn’t like any of the artworks shown.

*Perceived system effectiveness and fun*

* I have fun when I am using the system.
* I would recommend the system to others.
* Using the system is a pleasant experience.
* I can find interesting artworks with the system.
* The system showed me artworks I would usually not find.
* The system is useless.
* The system makes me more aware of my choice option.
* I make more informed choices with the system.
* I can find better items without the help of the system.
* The system showed useful items.

*Choice satisfaction*

* I like the artworks I have seen.
* I was excited about the artworks shown.
* I enjoyed seeing the artworks shown to me.
* The artworks shown to me were diverse.
* The artworks shown to me were novel.
* The system offered serendipitous items.
* The artworks I have seen were a waste of time.
* I would recommend some of the shown artworks to family and friends.

*Test awareness*

* I am aware that the system showed me recommendations.
* I am aware that items in this part were especially chosen to suit my choice of artworks.
	1. Please think of **PART II** (the second half of the study) when answering the following questions:

*Perceived recommendation quality*

* + I liked the artworks shown by the system.
	+ The artworks fitted my preference.
	+ The artworks were well-chosen.
	+ The artworks were relevant.
	+ The system showed me too many bad artworks.
	+ I didn’t like any of the artworks shown.

*Perceived system effectiveness and fun*

* I have fun when I am using the system.
* I would recommend the system to others.
* Using the system is a pleasant experience.
* I can find interesting artworks with the system.
* The system showed me artworks I would usually not find.
* The system is useless.
* The system makes me more aware of my choice option.
* I make more informed choices with the system.
* I can find better items without the help of the system.
* The system showed useful items.

*Choice satisfaction*

* I like the artworks I have seen.
* I was excited about the artworks shown.
* I enjoyed seeing the artworks shown to me.
* The artworks shown to me were diverse.
* The artworks shown to me were novel.
* The system offered serendipitous items.
* The artworks I have seen were a waste of time.
* The chosen artworks fitted my preference.
* I would recommend some of the shown artworks to family and friends.

*Test awareness*

* I am aware that the system showed me recommendations.
* I am aware that items in this part were especially chosen to suit my choice of artworks.
	1. **General** questions asked after the part specific ones:

*Intention to provide feedback*

* I didn’t mind having to choose artworks.

*General trust in technology*

* Technology never works.
* I trust the system I have just used.
* Technology should always be explainable.
* I am not fussed about how things work in the background as long as the technology works.
* I am generally questioning what happens to my personal data

*Museum online collection relevant question*

* I prefer a classic keyword search compared to this system.
* The system is not suitable to display artworks.
* Museum online collections are generally boring.
* I do not need museum online collections.
1. Debrief

Thank you for participating, we really appreciate it.

We conduct this experiment to see if users of our museum online collection website prefer artworks that were recommended to them over randomly selected ones. There are three different ways how we generate recommendations: one is based on metadata (textual information held about the objects), image data (digital images of the objects), and lastly a mix of these two. You have just experienced one of the three methods, with either part I or part II being randomly chosen artworks.

Do you have any questions about this? If so, please send an email to lukas.noehrer@manchester.ac.uk and we will get back to you with further information.

Please don’t share the details of this experiment with others as they may participate and cannot know the true aim of this study.

**References**

Berinsky, A. J., Huber, G. A., and Lenz G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk. *In: Political Analysis,* 20 (3), pp. 351-68.

Johnson, J., Douze, M. and Jégou, H. (2017). Billion-scale similarity search with GPUs. *In: arXiv:1702.08734 [cs].*

Knijnenburg et al. (2012). Explaining the user experience of recommender system. *In:* *User Modeling and User-Adapted Interaction*, 22 (4-5), pp. 441-504.

Mummolo, J. and Peterson, E. (2019). Demand Effects in Survey Experiments: An Empirical Assessment. *In: American Political Science Review*, 113 (2), pp. 517-29.