Carnegie Mellon University

Online Consent for Participation in Research

Protocol Number:
STUDY2025_00000079

Study Title:
Browsing URL Sequence Study

This study is part of a research study conducted by Jingyuan He (jingyuah@cs.cmu.edu) under the supervision of Professor Chenyan Xiong (cx@cs.cmu.edu) at Carnegie Mellon University and is funded by Meta Platforms, Inc.

Summary:
This study collects browsing URL sequences to investigate the browsing behaviors of Internet participants. Participants will upload their recent browsing history (a sequence of URLs) to a web interface we provide.

Participants should be 18 years or older, present in the U.S., fluent in English in order to fully understand the study consent and URL submission instructions, and able to upload browsing history URLs we provided.

Purpose:
The purpose of this study is to investigate the up-to-date browsing behavior of U.S. adult Internet participants to shed light on recommendation systems, by providing a multi-domain and multi-modality benchmark dataset to evaluate the current recommendation systems' performance over modeling user interest on URL sequences.

Procedures:
A participant is expected to make 10 submissions, each containing 10 different URLs in a day of their browsing history within the past year from when they take the study.
The 10 submissions should each contain URLs from a different day. That is, a total of 10 different days' browsing history shall be submitted.

Participants are expected to spend 2-3 minutes over each submission of 10 URLs. Participants can make a maximum of 10 submissions, each containing URLs from a different day in the past year. Depending on the submissions one makes, the study can take 2-30 minutes in total.

First, they need to navigate to the web interface we provide. Second, they will go to their browsing history, find their visited URL and the corresponding visit timestamp, and copy this information into the submission box of our web interface for a submission.

While submitting browsing history, participants may omit the URLs that contain their personal identifiable information, such as a URL with their username in the query field. Participants are

encouraged to submit URLs over browsing tasks such as entertainment tabs through the frequently asked question sections in our URL collection web interface.

Participant Requirements:
Participation in this study is limited to individuals aged 18 and older. Participation is limited to the U.S.

Internet connection is required, and participants should be familiar with basic website usage such as navigating a website, copying and pasting information.
Participants need access to a web browser that contains their search history, such as Chrome, Edge, etc.

Risks:
The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during other online activities. No identifiable private information will be collected, as the research team will not save URLs linked to non-accessible web pages, which are bound by private / group permissions and are likely to contain personal content. The collected URLs will be further processed to remove any identifiable private information such that only fully de-identified data will be used for study / possible release.

The Prolific.co. platform, the Vercel neon database (where the initial collected data is stored), and the CMU virtual computation clusters (where the data will be further filtered and processed) might have access to the subject's identifiable data and be a possible source of a breach of confidentiality.

Benefits:
Indirect benefit of the study is its influence in the field of recommendation system, serving as a benchmarking test that models real-life user-item interaction on URL sequences.

Compensation & Costs:
The equivalent $4 per subject, for submitting 10 URLs from their browsing histories for 10 times from 10 different days.
A single submission of 10 URLs will be rewarded $0.4.
The compensation will be rewarded in U.S. dollars in cash through Prolific.
There will be no cost to you if you participate in this study.

Confidentiality
By participating in this research, you understand and agree that Carnegie Mellon may be required to disclose your consent form, data and other personally identifiable information as required by law, regulation, subpoena or court order. Otherwise, your confidentiality will be maintained in the following manner:

Carnegie Mellon University

Your data and consent form will be kept separate. Any paper files will be stored in a secure location on Carnegie Mellon property and all digital files will be stored under Carnegie Mellon control. By participating, you understand and agree that the data and information gathered during this study may be used by Carnegie Mellon and published and/or disclosed by Carnegie Mellon to others outside of Carnegie Mellon. However, your name, address, contact information and other direct personal identifiers will not be mentioned in any such publication or dissemination of the research data and/or results by Carnegie Mellon. Note that per regulation all research data must be kept for a minimum of 3 years.]

The URL data you submitted will be kept confidential. If distributed, all identifiable information in the data will be removed.

The sponsor of this study (Meta Platforms, Inc.) may obtain access to the copy of de-identifiable data and research records.

Third-Party Confidentiality: The study will collect your research data through your use of Prolific.co and Vercel Neon database. These companies are not owned by CMU. The companies will have access to the research data that you produce and any identifiable information that you share with them while using their product. Please note that Carnegie Mellon does not control the Terms and Conditions of the companies or how they will use or protect any information that they collect.

Payment Confidentiality: Payment methods, especially those facilitated by third-party vendors (such as Visa, Prolific, Qualtrics, Venmo, Amazon, PayPal, Zelle), may require that the researchers and/or the vendor collect and use personal information (such as your first and last name, email addresses, phone numbers, banking information) provided by you in order for your payment to be processed. As with any payment transaction, there is the risk of a breach of confidentiality from the third-party vendor. All personal information collected by the researcher will be held as strictly confidential and stored in a password-protected digital file, or in a locked file cabinet, until payments are processed and reconciled. This information will be destroyed at the earliest acceptable time. Personal information held by the third-party vendor will be held according to their terms of use policy.

Right to Ask Questions & Contact Information:
If you have any questions about this study, you should feel free to ask them by contacting the Principal Investigator now at jingyuah@cs.cmu.edu.
If you have questions later, desire additional information, or wish to withdraw your participation please contact the Principal Investigator by e-mail in accordance with the contact information listed above.

If you have questions pertaining to your rights as a research participant; or to report concerns to this study, you should contact the Office of Research integrity and Compliance at Carnegie Mellon University. Email: irb-review@andrew.cmu.edu . Phone: 412-268-4721.

Carnegie Mellon University

Voluntary Participation
Your participation in this research is voluntary. You may discontinue participation at any time during the research activity. You may print a copy of this consent form for your records.
By continuing to the web interface of this study, you agree that the above information has been explained to you and all your current questions have been answered. You are encouraged to ask questions about any aspect of this research study during the course of the study and in the future.