

Assignment I – CS 6474/CS 4803 Social Computing *

<i>Grade</i>	Max 50 points; 5% of overall grade (late policy applies)
<i>Due</i>	January 24, 2024, 11:59pm Eastern Time
<i>What to hand in</i>	A report (as a PDF file) with answers to each question; reports should be approximately 2-3 pages in a single-spaced, single column format with 1-inch margins with a font size of at least 11. You can include any relevant screenshots or excerpts from your data archive to support your observations
<i>Where to submit</i>	Canvas
<i>Evaluation criteria</i>	<ul style="list-style-type: none">- Quality and depth of introspection of the downloaded data- Clarity, organization, and innovativeness of the reflections provided in the answers- Overall presentation with meaningful insights supported by evidence (the data)

The goal of this assignment is to analyze and interpret your own social media data, examining available metadata, your activities on the platform, preferences, and implications for privacy.

Under the European Union’s General Data Protection Regulation (GDPR) [1], social media platforms enable users to access and download their personal data archives. Most major platforms like Facebook, Instagram, and TikTok provide this functionality. These archives contain a wealth of data about a user. They include comprehensive user-generated content (e.g., posts), interactions with other users, and metadata logged during your engagement with the platform (e.g., timestamp of posts, when you friended or followed another user, etc.). This assignment asks you to delve into your own social media data archive, examining various facets of the data collected by the platform and gathering insights into how social media platforms document and utilize user information.

To complete this assignment, begin by picking a platform of your choice that allows you to download your data archive based on the instructions provided by the platform. For instance, [this](#) [2] is a video describing how to download your Instagram archive. **Please note that it can take several days to get a copy of your data, so request the archive early and perhaps select a platform with moderate usage.** In most cases, the archive will be a zipped folder with individual files in an HTML or JSON format. You can pick the format you prefer.

Once you have downloaded the archive, manually browse through the various files. Answer the following questions. Please note that this is an exploratory insight gathering assignment and no programmatic analysis of your data is needed to complete the assignment.

Note: We try very hard to make questions as unambiguous as possible. If confused, send the instructor and TAs a message stating the cause of confusion and your assumptions explicitly. All questions regarding clarifying aspects of the assignment **must** be asked prior to 48 hours before the due date.

1. *Understanding Platform-Gathered Metadata* (10 points): Identify and describe at least three types of metadata present in your data archive (e.g., timestamps, location data, device information). Explain their significance in understanding your online activities.
2. *Discovering Patterns and Trends* (20 points): Manually browse your data to identify any recurring patterns, trends, or a lack thereof. Highlight two interesting findings about your own behavior from your manual exploration of your data.
3. *Privacy Implications* (10 points): Discuss any potential privacy concerns related to the collection and storage of your data, based on your scrutiny of the archive. Enlist any surprising discoveries or aspects that might pose risks to user privacy.
4. *Recommendations* (10 points): Propose at least two recommendations for the platform to better manage user data, balancing the need for comprehensive data collection to improve user experience with the need for protecting user privacy.

[1] https://en.wikipedia.org/wiki/General_Data_Protection_Regulation

[2] <https://www.youtube.com/watch?v=dPPXEefnviI>