Reranking Social Media Feeds: A Practical Guide for Field Experiments

TIZIANO PICCARDI[*], Stanford University, USA
MARTIN SAVESKI[*], Stanford University, University Of Washington, USA
CHENYAN JIA[*], Stanford University, Northeastern University, USA
JEFF HANCOCK, Stanford University, USA
JEANNE L. TSAI, Stanford University, USA
MICHAEL BERNSTEIN, Stanford University, USA

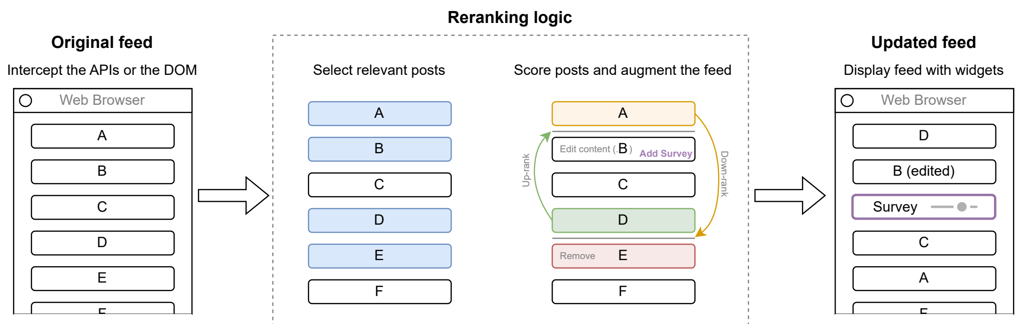

Fig. 1. Diagram of the typical feed reranking steps using a browser extension. First, the extension identifies the posts that need to be rescored (optional if all posts are included). The ranking model assigns a new score to the posts, and finally, the updated data is shown in the browser. The modified feed can be enriched with custom widgets, such as surveys, to assess the impact of the intervention.

Social media plays a central role in shaping public opinion and behavior, yet performing experiments on these platforms and, in particular, on feed algorithms is becoming increasingly challenging. This article offers practical recommendations to researchers developing and deploying field experiments focused on real-time re-ranking of social media feeds. This article is organized around two contributions. First, we overview an experimental method using web browser extensions that intercepts and re-ranks content in real-time, enabling naturalistic re-ranking field experiments. We then describe feed interventions and measurements that this paradigm enables on participants' actual feeds, without requiring the involvement of social media platforms. Second, we offer concrete technical recommendations for intercepting and re-ranking social media feeds with minimal user-facing delay, and provide an open-source implementation [1]. This document aims to summarize lessons learned, provide concrete implementation details, and foster the ecosystem of independent social media research.

1 INTRODUCTION

Social media plays a critical role in forming public opinion and influencing both individual behaviors and societal outcomes [36]. Many rely on social media to follow the news and stay informed [6]. Central to this media consumption behavior is the *feed*, which selects and ranks the social media posts that users are exposed to. It

[*]Equal contribution.
[1]https://github.com/StanfordHCI/FeedMonitor

Authors' addresses: Tiziano Piccardi, Stanford University, USA, piccardi@stanford.edu; Martin Saveski, Stanford University, University Of Washington, USA, msaveski@stanford.edu; Chenyan Jia, Stanford University, Northeastern University, USA, chenyanj@stanford.edu; Jeff Hancock, Stanford University, USA, hancockj@stanford.edu; Jeanne L. Tsai, Stanford University, USA, jltsai@stanford.edu; Michael Bernstein, Stanford University, USA, mbernst@stanford.edu.

is the primary interface through which users engage with content, and it is the lens they use to make sense of the world. Most major platforms, including Facebook, X (Twitter), and Reddit, rely on their feed to give users a centralized view of relevant content. However, in doing so, these platforms must make algorithmic curation choices [1]. In response, researchers have called for special attention to the impact of these feed algorithms, for example the kinds of content and behaviors that they uprank or downrank. These design choices are of substantial interest because they may profoundly shape our experiences and beliefs. Of particular concern are feed algorithms that aim to maximize user engagement and inadvertently amplify undesirable content, which may exacerbate polarization [19, 24], misinformation [33], and antisocial behaviors [26]. These questions have remained largely unanswerable to the group of people without personal access to hundreds of millions of users—a group that probably includes you if you are reading this article.

Investigating the causal impact of the feed ranking in an ecologically valid way is challenging. Despite efforts to open-source some of the key components of the platforms' ranking algorithms [31], their "black box" nature, reinforced by the lack of data [30], complicates attempts to evaluate their causal effects on individual users and society as a whole. While the "logic" behind the algorithm may be open source, the algorithms' outputs inherently depend on closed machine learning models that score aspects of each post and determine what is visible in the feed and in what order. Another challenge is the increasingly limited access to posts that would be considered to be shown in users' feeds, known as the user's *inventory*. Incomplete access to the key components of the feed—algorithms, models, and inventory—inherently limits the validity of any investigation.

This article aims to bridge this gap by offering a practical guide for conducting independent field experiments to investigate the impact of feed ranking in social media. The guidelines focus on facilitating the design of experiments that do not require the direct cooperation of the social media platform, enabling audits of feed algorithms and development of interventions that aim to improve the design of online social spaces. We describe how to implement experiments on the web interfaces of social media platforms using browser extensions, a feature available on all major browsers, by catching and customizing the network calls that populate the feed. We focus on specific interventions that this method supports, including feed re-ranking and content editing, and describe the relative advantages and disadvantages of each. We do not aim to exhaustively list all possible ways to implement a field experiment, but we do provide concrete building blocks for designing custom experiments.

Figure 1 summarizes the main steps of a typical field experiment designed to customize the ranking of a social media feed with a browser extension. This guide focuses on designing interventions that do not affect the users' typical experience on the platform, aiming to limit latency and any interface behavior that can disrupt the user experience. To achieve this goal, the browser extension (1) captures the feed, (2) applies any experimental transformations, then (3) inserts any custom user interface elements. First, it reads the feed to identify the posts that need editing according to the objective of the experiment. For example, the experiment may focus on all the content in the feed, or it may target specific posts such as politics or misinformation. Then, the extension can apply the desired intervention to the selected posts (or those newly inserted), such as changing their rank position, rewriting their text, or editing the social metrics. Finally, the extension can change the design of the webpage or add widgets that enhance the interface and enable new user interactions. Such modifications enable researchers to design custom user interactions or assess the effect of a specific intervention by adding targeted surveys.

This article is organized as follows. In Sec. 2, we summarise the state-of-the-art and highlight studies closely related to our work. In Sec. 3, we present opportunities and challenges in changing the feed ranking with a browser extension, followed by an overview of possible approaches to measuring the outcomes of the field experiment. Finally, Sec. 5 provides key implementation details to develop the feed experiment, ranging from ensuring low-effort user registration to intercepting and editing the feed. We conclude with a discussion of implications.

2 BACKGROUND

In this section, we summarize previous efforts to study social media through experiments and position this article in the current literature. On-platform feed experiments are the gold standard for inferring the causal effects of interventions on social media, as they test interventions in the same setting where users would typically encounter them, maximizing ecological validity. An alternative to running experiments on the platform is creating experimental clones of the platform and simulating a social media environment [10, 16, 34]. Developing a custom social media feed simulator is a meaningful approach, as it allows greater control over the experimental setup, such as the posts visualized by the user and features offered by the websites. This full control of the feed has significant advantages because participants can be exposed to the same content, potentially leading to a more controlled experiment that may help measure small effects. However, the downside is that the individual participants' online experience is generally hard to reproduce, and the generalizability of the findings may be questioned. The participants may be exposed to a different interface and a different type of posts they typically consume, they may miss realistic social signals and dynamics, or they may be required to engage in activities not aligned with their daily habits. The requirements to engage with a custom platform may also exacerbate phenomena such as the Hawthorn effect, potentially altering the participants' behavior that does not fully generalize to a real-world scenario.

Previous work shows the effectiveness of field experiments on social media and that they can bring valuable findings. Some examples include investigating the effect of changing Facebook's feed ranking from engagement-based to chronological during a US election [13], the impact and propagation of emotions [18], the amplification of political discourse [15], the effect of perspective-taking on political polarization [27] or using ads to reduce misinformation [20] and support charitable donations [17].

However, designing a field experiment with live intervention on social media feeds presents many technical challenges, especially when done without the involvement of the platforms [4, 12, 21]. The present work aims to support researchers with practical recommendations for designing field experiments focused on feed ranking.

Previous contributions similar in spirit to this work include practical guides on designing studies on fake news and misinformation [23], analyzing browsing data [8], designing surveys to measure political behavior online [14], sampling survey participants [25], running split tests on Facebook [22], and software and techniques to collect and process social media data [3].

While many field experiments of interest only require changing participants' feeds, some experiments remain out of scope for this approach. For example, this approach cannot perform network bucket testing or other experimental protocols that require changing entire groups of users' experiences together [28, 32], e.g., to understand equilibrium effects rather than individual effects. Our approach also cannot easily reach the user's full inventory unless the experimenter implements a separate application layer to acquire it directly from the platform; typically, in the context of this paper, we will be focused on changes to the items that are already chosen to be shown to the user.

2.1 Key concepts

In this section, we describe the key concepts required to understand this guide.

Browser extensions. This guide focuses on modifying social media feeds using Browser extensions. Browser or web extensions are add-ons typically available via the browser-specific store, designed to enhance the browser's functionality. They support developers in customizing the browser's behavior and modifying websites' content. This capability is advantageous in developing custom social media feed interventions, enabling design control trials or A/B tests. From the perspective of the experimental participants, the only technical prerequisite is to install the extension and accept the permissions. With authorization, browser extensions can modify the pages of interest to achieve the desired intervention. These modifications to the web pages range from updating the content and its visual appearance to the website's behavior. Sec. 5 provides an overview of implementation details.

Platform Inventory. In this guide, we refer to the full database of candidate posts used by the feed algorithm as the participant's *inventory*. This is a crucial concept because, depending on the platform of interest, the level of access to the candidate set can enable or limit some interventions. When the feed is organized chronologically, the algorithm can sort by time the posts of the relevant users (e.g., friends or followed users) and return the most recent ones. When the feed is algorithmically curated, the algorithm needs to evaluate a large set of posts by users inside and outside the focal user's first-degree social network. This process typically relies on multiple steps. For instance, Twitter/X's "For you" algorithm, as described in the open-sourced implementation [31], first selects content using a high-recall retrieval system that scans a large amount of data which is then scored and ranked by a more refined model.

3 FEED RANKING

This section summarises the potential of browser extensions to customize social media feeds. The re-ranking can be a complete shuffle of the feed based on a new objective or more focused, for example up-ranking or down-ranking content the researchers believe should receive more or less exposure. In this guide, we focus on these two use cases, up-ranking and down-ranking, where the content of interest is respectively demoted or promoted in the feed, but these recommendations can be adapted for different types of rank adaptations. These edits can be achieved by combining up-ranking and down-ranking operations, which constitute the building blocks of a feed intervention.

A key functionality of the system handling the experiment is identifying relevant content and scoring it according to the metric of interest. Depending on the research goal, this capability can be achieved with different levels of complexity, ranging from simple keyword matching to more complex logic relying on external AI services like large language models (LLMs). For example, feed posts can be scored and rearranged to demote hostile political content or prioritize posts reflecting positive emotions.

These manipulations offer a unique opportunity to investigate the causal effects of variations of the rank objective in a realistic scenario. We summarise the challenges and limitations of both implementations. Additionally, a browser extension can edit the content of the posts. We conclude this section with a discussion on content editing and how a browser extension can modify text or social metrics. Section 5 offers a detailed description of how these manipulations can be accomplished in practice.

3.1 Down-ranking

The ability to down-rank content allows researchers to investigate alternative moderation strategies. Harmful content can be dynamically penalized in the ranking, making it available only when the user is interested in consuming more content and scrolling further down their feed. Given criteria for down-ranking content, such interventions are generally easy to implement. The extension should select the posts that must be penalized and estimate how far down the feed the item ought to be down-ranked. The new position can be determined based on a fixed offset or based on the score of the content. For example, researchers may decide to add 100 positions to all down-ranked posts or use a content-based offset, making sure that very harmful posts receive less exposure. In practical terms, the extension should keep track of the number of posts consumed by the user and insert the post when the browser's viewport reaches the desired position. This step can be achieved easily on most platforms by inserting the same content removed in the previous position. This intervention offers valuable insights into the impact of ranking. It can provide concrete recommendations for the social media platform on adapting the algorithm to penalize problematic content and achieve the desired outcome. This intervention allows testing interventions that can reduce exposure to some content but do not entirely remove the recommended posts–a topic that may be sensitive for those opposing moderation strategies.

As an alternative, an extreme form of down-ranking is the complete removal of specific content. This logic, similar to how ad blockers work, operates on the definition of an exclusion logic or metric that guides the extension in identifying the posts that should be removed from the user's feed. This intervention can provide valuable insights into the impact of filtering out posts with particular characteristics. This approach allows researchers to investigate research questions such as the impact of stricter moderation methods by reducing the content that has the potential to polarize communities, affect people emotionally, or to investigate the impact of alternative feed designs such as removing all the posts with videos or images.

Finally, when measuring the effect of down-ranking or removing content, it may be important to consider what the user was shown in place of the removed content. To answer these questions, we recommend keeping track of the posts the participant was exposed to during the study.

3.2 Up-ranking

Without access to the complete platform inventory, up-ranking content in the feed presents more challenges than down-ranking content. The content returned by the server is obtained by scoring the inventory, which includes the entire candidate set of potential posts available for display. Without the full set or a meaningful approximation, the intervention can operate only on the pre-selected content delivered to the user, representing the top items the curation algorithm has already prefiltered. For example, when the platform's algorithm selects content based on engagement, the posts available for re-ranking may consist only of high-engagement posts. Access to the full inventory could significantly expand the dynamic range of content available for re-ranking, potentially showing a more prominent effect of an intervention. However, compiling the full or even partial inventory may require more active support from the platforms or data access agreements, which may be difficult or expensive to set up. Possible approaches include up-ranking posts loaded by the website but not visible in the browser's viewport, pre-fetching content with the extension by simulating scrolling or inserting entirely new posts from the platform inventory obtained in other ways.

Inserting new posts provides a mechanism for customizing the feed and exposing the user to content not selected by the curation algorithm. This intervention can help researchers investigate research questions that rely on adding content blended in the context of the original feed. Some examples include adding more content promoting positive emotions, adding out-party posts to counter filter bubbles, or adding posts on a specific topic to change the topic distribution of the feed. This strategy is particularly helpful when the type of posts the researchers are interested in investigating is rare, and the goal is to ensure the feed has enough relevant content.

From a technical point of view, this approach requires careful consideration. The inserted posts could be created in three ways: (i) generating entirely new posts that fit the requirements of the experiment, (ii) monitoring posts by a curated list of accounts that are likely to post suitable content, and (iii) transferring a post that matches these requirements from another feed. The first approach may present challenges because it requires replicating a fully functional social media post, including user interactions, such as likes, comments, and shares. This method may allow full customization but demands careful attention to detail to ensure the post supports real-world user interaction. Failing to recreate all these interactions accurately (i.e., the share button not working) may disrupt the user experience and compromise the study's validity by influencing user behavior in unanticipated ways.

The second approach is monitoring a set of public accounts that are likely to post content that fits the requirements of the intervention. This method requires running a background process that continuously collects and scores the content posted by the selected accounts. There are two challenges with this approach: (i) the content that individual users post varies and curating a list of suitable accounts effectively may be challenging for some interventions; and (ii) the data collection may require additional data access, which may be prohibitively expensive.

The third approach is transferring content from other existing feeds. This approach offers a more pragmatic alternative, allowing researchers to utilize authentic posts. This technique has a downside: researchers must implement a mechanism to find or produce posts that qualify for the intended intervention, especially without full inventory. This can be done by drawing the post from a dedicated account, using tweets from the same user previously removed, or monitoring the feeds of other participants in the experiment. In the former case, this method requires researchers to ensure that the content selected for transfer respects user privacy and does not inadvertently expose private or sensitive information. Finally, researchers employing this intervention must reflect on the potential impact of context when transferring posts between feeds, as the relevance and reception of a post may vary significantly based on its original context.

3.3 Content Editing

In addition to re-ranking the content of the feed, researchers can operate directly on individual posts. This intervention can involve direct manipulation of the post, such as editing the posts' visible social metrics (e.g., likes, comments, and shares), the appearance of the post (e.g., making some aspects of the post prominent), changing the attachments (e.g., replacing a link or a picture), or modifying the text. These interventions may be important in addressing research questions on how the platform's design or content can affect the users. For example, the tone of existing posts can be subtly reframed to reflect a more positive sentiment or mitigate aggressive language. This intervention can allow researchers to explore how language choices in social media posts may influence user behavior, including interactions, affective experiences, discourse patterns, misinformation, or hate speech. Depending on the complexity of the desired edits, the modification can be achieved with simple dictionary substitution or complete reframing, e.g., leveraging LLMs to generate high-quality and contextual text.

Summary. This section summarises the opportunities and challenges associated with down-ranking, up-ranking, and editing feed content. Down-ranking is generally the simplest intervention as it requires operating on content already available in the feed, whereas up-ranking, in the absence of access to the complete inventory, may require a more careful design to find or generate the posts to up-rank. Content editing can be combined with the previous methods and represents an effective strategy for testing the impact of social metrics, language, or post appearance. Furthermore, thanks to browser extensions' capabilities, these modifications can be combined with additional interventions. Some examples include extending the platform's functionality with widgets like the Twitter/X Community Notes labels or adding visual indications on some post characteristics, such as a warning about language toxicity. These interventions can be combined to achieve more complex experimental designs that fit particular research goals.

Finally, all these interventions require researchers to pay attention to the distribution of posts in the feed of the target population, potentially with some repercussions for the recruitment strategy. For example, manipulating a single post may produce minimal exposure to the treatment condition and the effect of the intervention may be too small to detect [12]. This problem is particularly impactful in situations when users interact with only a small fraction of the content [35], typically at the top of the feed [2] or when the types of posts that the researchers intend to edit are rare.

4 MEASUREMENT

In this section, we describe a set of approaches to measuring the feed intervention's impact. We summarise three alternatives that enable different insights into the effect of the feed manipulation: momentary measures, longitudinal surveys, and engagement signals.

4.1 Ecological Momentary Assessments

When researchers are interested in measuring the momentary impact of a feed intervention, a suitable design is to add surveys directly to the feed. These in-feed surveys, inspired by the approach developed in clinical psychology called Ecological Momentary Assessment [29], allow researchers to capture real-time reactions directly in the context where the intervention is applied [1]. In the case of social media, this strategy can elicit direct feedback on the intervention at the right time when users scroll through the social media feed. Researchers can ask highly contextualized questions such as "Are you interested in this post?" or "How does this post make you feel?". This method can be used to measure a momentary impact of the intervention that may be too short to be measured with long-term surveys, or that is hard to measure when the question is posed without context. The insertion of these surveys can be triggered by specific events, such as when a specific post is displayed, added at regular intervals, or shown at random times while the user consumes their feed. The appearance of the survey widget may vary, from attention-grabbing designs such as popups or modal windows that can interrupt the scrolling, to more subtle designs integrated as a special post in the regular feed. If integrated with the regular feed, one practical challenge of in-feed surveys is making them visible and clear enough for participants to know these questions are part of the study. If researchers require that the insertions get noticed, we recommend using colors and animations that stand out from the default design. The design must also be adapted to the different templates (i.e., dark mode theme) that the social media platform supports. Finally, when deciding where to place the in-feed survey, researchers must carefully consider the goals of their measurement. Participants answering the survey will be exposed to the content that precedes it in the feed, and might be influenced by other content visible in their viewport.

4.2 Survey methodology

Researchers interested in measuring the cumulative effects of the treatment can rely on a standard pre-post survey design. In such a setup, participants are presented with the same survey before (pre-) and after (post-) the experiment. Examples of attitudinal shifts that can be caused by feed interventions include changes in affective polarization, often measured using the feeling thermometer, or opinion change, typically measured with Likert-scale surveys. These surveys are assigned at the start and the end of a study and require a seamless integration into the experimental design. Pre-surveys intended to be completed at the beginning of the experiment can also be used to assign participants to specific experimental conditions or decide whether they qualify to be enrolled in the experiment. Attention checks and validation questions can be valuable tools for identifying early participants who may be less attentive and contribute unreliable data.

A similar approach, reassembling a diary study technique [5], consists of regular assessments throughout the experiment. This method relies on administering surveys at regular intervals, e.g., daily or weekly. This design may be recommended for different reasons, such as measuring the progressive change in the outcomes of interest or because of a study design requiring participants to be exposed to different experimental conditions (e.g., crossover or stepped-wedge design).

Implementing a strategy to administer surveys at regular intervals can present additional challenges, especially when the users are not online on the social media platform. Researchers need to establish a notification system through the browser extension, implement regular messaging on the recruitment platform, or may need to distribute a survey link by establishing extra communication channels like email or text messaging. This last approach requires participants' consent to collect potentially identifiable information, which we discuss in more detail in Sec. 5.4. Another challenge of this setup is ensuring that participants consume enough content in time between surveys to meaningfully measure the impact of the intervention. One potential approach, valid in the case of repeated assessments, is to invite the participants to use social media and deliver the survey only after they spent enough time on the platform or have been exposed to a predefined number of posts.

4.3 Engagement signals

Implementing a feed experiment using a browser extension allows researchers to go beyond self-reported measures and make very granular behavioral measurements that can provide further insights into the effects of the intervention. This includes metrics such as time spent, engagement (e.g., clicks, likes/favorites, reports), navigation patterns, conversion rates, and other metrics [9] that are imperative to the business objectives of the platforms. Access to such measures also allows researchers to evaluate the trade-offs the proposed intervention introduces. For instance, removing some content entirely (e.g., all political posts) may have an impact on an outcome of interest (e.g., affective polarization) but may disrupt the engagement in a way that makes the adaption not attractive for social media companies. Reporting these metrics may make social media platforms more receptive to integrating the findings into their products.

Similar to the previous section, these measurement approaches can be combined to measure multiple effects and increase the understanding of the intervention at different time horizons.

5 IMPLEMENTING A BROWSER EXTENSION

In this section, we focus on the implementation details of running social media feed experiments with a browser extension and summarize various technical challenges. We offer practical recommendations and release source code [2]—in Javascript—that can be used as a blueprint to implement feed experiments on X. Our examples are based on Google Chrome but can easily be adapted to other browsers. The rest of this section assumes familiarity with key concepts of web architecture and extension development.

First, we provide implementation guidance on how to intercept and edit the responses from the server containing the feed. Then, we cover guidelines on participant recruitment and registration. Finally, we conclude with technical details on how to notify the user about specific events or deliver surveys.

5.1 Feed interventions

In this section, we present some key details on the implementation of the core functionality of the browser extension. Although these extensions can change many aspects of the experience on a website, here we focus specifically on some edits they can support on social media feeds. Browser extensions are typically written in languages supported by the browser, such as Javascript, HTML, and CSS, and they can employ different methods to apply the edits. A common strategy involves modifying the rendered page by manipulating its Document Object Model (DOM). This method has the advantage of simplifying the identification of the element involved in the desired edit, and it gives the extension access to the content after the frontend scripts have already processed it. However, despite this approach being the most straightforward, for some scenarios, such as augmenting the feed, intervening at the network level may be more effective.

In this case, we recommend adapting XMLHttpRequest to support the customization of the server requests. This technique offers the advantage of intercepting and modifying the original incoming data before its rendering on the page, minimizing the need for direct interaction with the actual page interface. Accessing the communication with the server allows logging many client-side actions shared with the server (e.g., like and share) without adding a listener for every event. The downside is the need to gain a deeper understanding of the communication protocol between the browser and the server. Finally, alternative approaches, such as manipulating the in-memory data structure used to render the content, are possible, but depending on the framework used by the front end (i.e., React), it may introduce implementation-specific complications.

[2]https://github.com/StanfordHCI/FeedMonitor

Implementation details. Pragmatically, to customize the behavior of the XMLHttpRequest native object, we need to run the overriding code as soon as the main page is loaded. Since behavior customization must happen within the scope of the social media page and not in the content script, we need to inject the modification script directly into the main page. This requirement demands a slight change from the common pattern used to develop browser extensions, as the script must be loaded in the *Manifest* file as a resource and not as a content script. Then, the content script can inject the override code into the scope of the main page.

The injected code overrides the *open* and *send* functions of XMLHttpRequest to adapt their behavior. The *open* function can be modified to save properties of the connection, such as the current URL or the HTTP headers, while the *send* function can be adapted to dynamically customize the response callback (*onreadystatechange*) if the URL matches the endpoint of interest. In this function, researchers interested in fetching more data (i.e., to have a large pool of posts to re-rank) can imitate the server requests to act as the users scroll through the feed. The injected script can communicate with the browser extension's content script using the browser's message-passing features by broadcasting the server's raw response. Depending on the intervention logic, when the feed needs to be modified, the customized XMLHttpRequest object can interrupt the execution, waiting for the response from the content script. Once the modified feed is ready, the request can resume by replacing the server response with the updated content. The content can be manipulated on the client side or by calling an external backend. If the logic or model can be distributed with the extension, keeping the manipulation on the client may ensure more participant privacy. Moving the edit to the backend may be preferable in other cases, like when complex logic or large models, such as LLMs, are required. Still, in this case, researchers must pay special attention to data security practices and the potential implications of sharing the data with third-party services.

Finally, it is crucial to ensure that scoring and reranking introduce minimal latency to avoid disrupting the user experience. The backend performance plays a critical role when the intervention depends on server-side logic. The implementation may require careful parallelization when relying on external services (LLM APIs) or using custom hardware (e.g., GPUs) when dedicated models are needed (e.g., fine-tuned BERT models). In specific conditions, when the model can run on the client side, researchers may consider running the models directly in the browser, potentially relying on the performance of WebGL or WASM-based implementations.

5.2 Participant recruitment

The recruitment strategy depends on many aspects, such as availability constraints, specific demographic requirements (e.g., partisans, or users of specific age groups), or length of the study.

Some experiments may aim to recruit a participant sample representative of the platform's user base, which requires adjusting the recruitment based on the platform's user composition [12]. For example, in 2023, Twitter was used by 22% of the adult population with a bias toward young, highly educated, and liberals [7], while Facebook is more balanced on these properties but has a larger active user base of women.

Depending on the requirements and the recruitment platform, it may be desirable to first screen participants through a short survey. This preliminary screening may help exclude individuals who do not meet the eligibility criteria that the recruitment platform does not allow to filter, such as how frequently they use social media. This step is crucial to selecting the population of interest and ensuring that the study includes the right participants to receive the treatment. Researchers investigating interventions on extreme or niche content need to ensure that the recruited participants are typically exposed to the posts of interest. For example, a study found that during the 2016 election, 80% of the fake news on Twitter was consumed by only 1% of the users [11]. If this step is necessary, we recommend keeping the survey minimal to reduce the cost and compensate participants for their time, even if they do not qualify, in accordance with common practices. Alternative methods include using ads directly on the platform of interest [25, 37], direct messages [12], or relying on snowball sampling with other reward incentives, such as gift cards or prizes.

5.3 Registration flow

Once the participants are recruited, the first challenge is ensuring a low-effort and scalable onboarding process. A confusing onboarding procedure can lead to high participant dropout and increase the risk of researchers being overloaded with requests for assistance.

Along with clear instructions, we recommend simplifying the process by designing a registration flow that eliminates common mistakes, such as the need for participants to copy and paste information manually. We also recommend that participants complete all the onboarding steps only by following clickable links to ensure a smooth user experience.

Since browser extension stores like the Chrome Store do not support direct parameter passing during extension installation (e.g., the participant ID), this section presents a solution to address this issue. This approach requires a coordination service, which can be integrated with the extension backend. It allows tracking of the user's registration to assign the user-specific configuration to the extension that can be sourced from the recruitment platform or an initial survey.

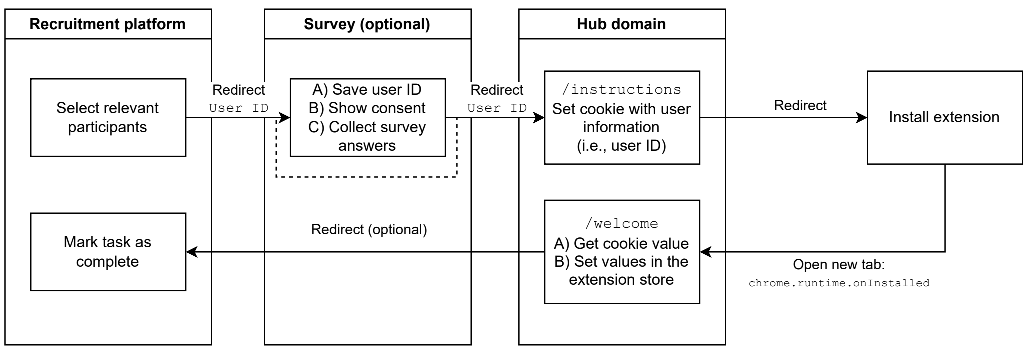

Fig. 2. Diagram of the registration flow that allows for tracking the participant and for passing parameters to the extension.

Fig. 2 summarises the registration flow using this pattern. With this flow, the recruitment platform sends the user to the coordination service directly or after an intermediary survey, which could be a pre-survey used to collect additional information necessary to run the extension (e.g., political ideology, contact information, timezone). Major survey platforms like Qualtrics support transferring information after the survey completion by combining embedded fields and redirecting URLs. The role of the coordination service is to set a persistent entry, such as an HTTP cookie, with the information that the extension must access after the installation. This page precedes the installation of the browser extension and may be used to summarize the instructions before redirecting the participant to the extension store page. Finally, when the extension is installed, opening a new tab to a page on the coordination server gives access to the HTTP cookie data previously saved that can be used to set up the extension for the current participant. This flow is simple and effective, but it is not the only approach that researchers can take. Alternative designs include integrating a custom login or implementing an OAuth flow.

Participant consent. Running an experiment that augments users' feeds requires participant consent. Depending on the registration flow—with or without the initial survey—the participants should be presented with a description approved by an Institutional Review Board (IRB) in the case of a US university or an equivalent record in other cases. It is important to show the consent form as early as possible and interrupt the installation if the participants do not accept the terms. Some participants may perceive installing a browser extension as too invasive. In addition to the consent, we recommend including in the instruction page a clear recap of what data is and what data is *not* collected (e.g., "we do not have access to your direct messages").

Error recovery. Anything (and everything) could go wrong. Especially when the goal is to run a large-scale study, it is important to implement recovery points to minimize the researchers' workload of error fixing. For example, a participant may start the survey using a browser or a computer different from the one they plan to use during the study. It is essential to implement recovery checkpoints where the participant can recover the process. One option consists of collecting contact information such as email addresses and sending a message when they reach the instructions page. The message can contain a copy of the instructions and the link to the instructions page with all the required parameters in the URL. Alternatively, if researchers do not have access to the participants' contact information—due to IRB restrictions or no pre-survey is planned—this step can be automated with a message on the recruitment platform.

5.4 Notifications

Throughout the study, researchers must establish channels of communication with the participants. Participants may need guidance, reminders, or feedback to ensure they engage in the study successfully. For example, in a longitudinal study where it is crucial to visit the social media platform regularly, the study design can employ emails to remind the participants that they are enrolled in the study or to acknowledge that their contribution has been recorded. Similarly, when the study is complete, users may receive a post-survey, instructions to remove the extension, or debriefing notes. This step must be handled carefully to comply with Institutional Review Board (IRB) standards for data protection.

Alternatively, user messages can be delivered directly on the platform by editing the DOM of the social media page. The browser extension can add these messages as in-page banners or modal windows, limiting the need to collect user contact information. In the case of post-surveys showing a pop-up or in-page banner when participants visit the social media platform at the end of the study, researchers can prompt immediate and contextual responses, potentially improving the quality of the data collected.

5.5 Privacy

Web extensions offer a great opportunity to test interventions that social media platforms would never try because of company priorities or because they do not align with the business objective. Nevertheless, despite offering more freedom to researchers, they must commit to ethical standards in designing the experiments and handling the data. Independent oversight, such as Institutional Review Boards (IRBs), should review and approve the study protocol to ensure it meets ethical standards. Participants must be aware of the potential risks and be informed about the data the web extension collects. Designing a registration flow that forces the participants to accept a consent form before collecting any data is essential. It is also crucial to inform participants that their participation is voluntary and that they can withdraw without penalty. If the experiment collects users' data, it is good practice to implement measures to ensure confidentiality and store the data in a secure database. Depending on where the study is conducted, researchers must comply with local regulations, such as the General Data Protection Regulation (GDPR) in the European Union. Unless specified in the consent form, at the end of the study, we recommend that the participants be guided in uninstalling the extension, or the extension should have a mechanism to turn it off. This step ensures that the researchers do not collect data without the participant's consent and that the unused extension does not create potential conflicts with other studies the participant may want to join.

6 CONCLUSION

Social media platforms have recently become less collaborative in supporting independent academic research. This guide offers some practical recommendations on designing field experiments that, using browser extensions, do not require direct involvement of these platforms. Our recommendations are focused on social media feeds, but many aspects generalize beyond this setup. With this guide, we hope to foster a community of independent researchers who aim to investigate social media's impact on society and contribute to designing healthier online spaces.

ACKNOWLEDGEMENTS

Our research was supported by the Hoffman-Yee Research Grants at the Stanford Institute for Human-Centered Artificial Intelligence (HAI), the Swiss National Science Foundation, and the Office of Naval Research.

REFERENCES

[1] Lars Backstrom. 2016. Serving a billion personalized news feeds. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. 469–469.

[2] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[3] Bogdan Batrinca and Philip C Treleaven. 2015. Social media analytics: a survey of techniques, tools and platforms. *Ai & Society* 30 (2015), 89–116.

[4] Leticia Bode, Pamela Davis-Kean, Lisa Singh, Tanya Berger-Wolf, Ceren Budak, Guangqing Chi, Andrew Guess, Jennifer Hill, Adam Hughes, J Bradford Jensen, et al. 2020. Study designs for quantitative social science research using social media. (2020).

[5] Niall Bolger, Angelina Davis, and Eshkol Rafaeli. 2003. Diary methods: Capturing life as it is lived. *Annual review of psychology* 54, 1 (2003), 579–616.

[6] Pew Research Center. 2023. *Social Media and News Fact Sheet.* https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/

[7] Pew Research Center. 2024. *Social Media Fact Sheet.* https://www.pewresearch.org/internet/fact-sheet/social-media/

[8] Bernhard Clemm von Hohenberg, Sebastian Stier, Ana S Cardenal, Andrew M Guess, Ericka Menchen-Trevino, and Magdalena Wojcieszak. 2023. Analysis of Web Browsing Data: A Guide. *Social Science Computer Review* (2023), 08944393241227868.

[9] Tom Cunningham, Sana Pandey, Leif Sigerson, Jonathan Stray, Jeff Allen, Bonnie Barrilleaux, Ravi Iyer, Smitha Milli, Mohit Kothari, and Behnam Rezaei. 2024. What We Know About Using Non-Engagement Signals in Content Ranking. *arXiv preprint arXiv:2402.06831* (2024).

[10] Ziv Epstein and Hause Lin. 2022. Yourfeed: Towards open science and interoperable systems for social media. *arXiv preprint arXiv:2207.07478* (2022).

[11] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on Twitter during the 2016 US presidential election. *Science* 363, 6425 (2019), 374–378.

[12] Andrew M Guess. 2021. Experiments using social media data. *Advances in experimental political science* 184 (2021).

[13] Andrew M Guess, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, et al. 2023. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* 381, 6656 (2023), 398–404.

[14] Andrew M Guess and Kevin Munger. 2023. Digital literacy and online political behavior. *Political science research and methods* 11, 1 (2023), 110–128.

[15] Ferenc Huszár, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. 2022. Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences* 119, 1 (2022), e2025334119.

[16] Chenyan Jia, Michelle S Lam, Minh Chau Mai, Jeffrey T Hancock, and Michael S Bernstein. 2024. Embedding democratic values into social media AIs via societal objective functions. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–36.

[17] S. Jilke, Jiahuan Lu, Chengxin Xu, and Shugo Shinohara. 2019. Using Large Scale Social Media Experiments in Public Administration: Assessing Charitable Consequences of Government Funding of Nonprofits. *Randomized Social Experiments eJournal* (2019). https://doi.org/10.1093/JOPART/MUY021

[18] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National academy of Sciences of the United States of America* 111, 24 (2014), 8788.

[19] Emily Kubin and Christian von Sikorski. 2021. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association* 45, 3 (2021), 188–206.

[20] Hause Lin, Haritz Garro, Nils Wernerfelt, Jesse Shore, Adam Hughes, Daniel Deisenroth, Nathaniel Barr, Adam Berinsky, Dean Eckles, Gordon Pennycook, et al. 2024. Reducing misinformation sharing at scale using digital accuracy prompt ads. (2024).

[21] Mohsen Mosleh, Gordon Pennycook, and David G Rand. 2022. Field experiments on social media. *Current Directions in Psychological Science* 31, 1 (2022), 69–75.

[22] Davide C Orazi and Allen C Johnston. 2020. Running field experiments using Facebook split test. *Journal of Business Research* 118 (2020), 189–198.

[23] Gordon Pennycook, Jabin Binnendyk, Christie Newton, and David G Rand. 2021. A practical guide to doing behavioral research on fake news and misinformation. *Collabra: Psychology* 7, 1 (2021), 25293.

[24] Steve Rathje, Jay J Van Bavel, and Sander Van Der Linden. 2021. Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences* 118, 26 (2021), e2024292118.

[25] Leah Rosenzweig, Parrish Bergquist, Katherine Hoffmann Pham, Francesco Rampazzo, and Matto Mildenberger. 2020. Survey sampling in the Global South using Facebook advertisements. (2020).

[26] Florian Saurwein and Charlotte Spencer-Smith. 2021. Automated trouble: The role of algorithmic selection in harms on social media platforms. *Media and Communication* 9, 4 (2021), 222–233.

[27] Martin Saveski, Nabeel Gillani, Ann Yuan, Prashanth Vijayaraghavan, and Deb Roy. 2022. Perspective-taking to reduce affective polarization on social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 885–895.

[28] Martin Saveski, Jean Pouget-Abadie, Guillaume Saint-Jacques, Weitao Duan, Souvik Ghosh, Ya Xu, and Edoardo M Airoldi. 2017. Detecting network effects: Randomizing over randomized experiments. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 1027–1035.

[29] Saul Shiffman, Arthur A Stone, and Michael R Hufford. 2008. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4 (2008), 1–32.

[30] Rebekah Tromble. 2021. Where have all the data gone? A critical reflection on academic digital research in the post-API age. *Social Media+ Society* 7, 1 (2021), 2056305121988929.

[31] Twitter. 2023. *Twitter's Recommendation Algorithm.* https://github.com/twitter/the-algorithm

[32] Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg. 2013. Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 329–337.

[33] Sebastián Valenzuela, Daniel Halpern, James E Katz, and Juan Pablo Miranda. 2019. The paradox of participation versus misinformation: Social media, political engagement, and the spread of misinformation. *Digital Journalism* 7, 6 (2019), 802–823.

[34] Angelina Voggenreiter, Sophie Brandt, Fabian Putterer, Andreas Frings, and Juergen Pfeffer. 2023. The Role of Likes: How Online Feedback Impacts Users' Mental Health. *arXiv preprint arXiv:2312.11914* (2023).

[35] Lucy Wang, Arthi Ramachandran, and Augustin Chaintreau. 2016. Measuring click and share dynamics on social media: a reproducible and validated approach. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 10. 108–113.

[36] Brian E Weeks, Alberto Ardèvol-Abreu, and Homero Gil de Zúñiga. 2017. Online influence? Social media use, opinion leadership, and political persuasion. *International journal of public opinion research* 29, 2 (2017), 214–239.

[37] Baobao Zhang, Matto Mildenberger, Peter D Howe, Jennifer Marlon, Seth A Rosenthal, and Anthony Leiserowitz. 2020. Quota sampling using Facebook advertisements. *Political Science Research and Methods* 8, 3 (2020), 558–564.